A Method for Estimating the Capital Cost of Chemical Process Plants
- Fuzzy Matching -

by

Gary John Petley

A Doctoral Thesis
submitted in partial fulfilment of the requirements
for the award of Doctor of Philosophy
of Loughborough University

August 1997

ABSTRACT

The purpose of this thesis is to improve the 'art' of early capital cost estimation of chemical process plants.

Capital cost estimates are required in the early business planning and feasibility assessment stages of a project, in order to evaluate viability and to compare the economics of the alternative processes and operating conditions that are under consideration for the plant. There is limited knowledge about a new plant in the early stages of process development. Nevertheless, accurate cost estimates are needed to prevent incorrect decisions being made, such as terminating the development of a would-be profitable plant.

The published early capital cost estimation methods are described. The methods are grouped into three types of estimate: exponent, factorial and functional unit. The performance of these methods when used to estimate the capital costs of chemical plants is assessed. A new estimating method is presented. This method was developed using the same standard regression techniques as used in the published methods, but derived from a new set of chemical plant data.

The effect that computers have had on capital cost estimating and the future possibilities for the use of the latest computer techniques are assessed. This leads to the fuzzy matching technique being chosen to develop a new method for capital cost estimation. The results achieved when using fuzzy matching to estimate the capital cost of chemical plants are presented. These results show that the new method is better than those that already exist. Finally, there is a brief discussion of how fuzzy matching could be applied in the future to other fields of chemical engineering.

CERTIFICATE OF ORIGINALITY

This is to certify that I am responsible for the work submitted in this thesis, that the original work is my own except as specified in acknowledgements or in footnotes, and that neither the thesis nor the original work contained therein has been submitted to this or any other institution for a higher degree.

GARY JOHN PETLEY

August 1997

ACKNOWLEDGEMENTS

I would like to express my thanks to my supervisor, Dr. D.W. Edwards for valuable guidance and discussion throughout the course of this work.

I would like to thank my colleagues in the Department of Chemical Engineering for their help and encouragement.

This work was funded by the Engineering and Physical Sciences Research Council.

CONTENTS

Chapter 7 FUZZY MATCHING

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

AACE	American Association of Cost Engineers
ACE	Association of Cost Engineering
AEEE	Average Equivalent Estimate Error
AI	Artificial Intelligence
ANN	Artificial Neural Networks
ASEE	Average Standard Estimate Error
AUC	Average Unit Cost
BLCC	Battery Limits Capital Cost
BM	Best Match
CBR	Case-Based Reasoning
CCE	Capital Cost Estimate
CE	Chemical Engineering
CF	Complexity Factor
COP	Cost of Production
CPF	Cost per Functional Unit
CS	Chem Systems
DV	Dependent Variables
EEE	Equivalent Estimate Error
ENR	Engineering News-Record
ES	Expert Systems
EV	Explanatory Variables
FST	Fuzzy Set Theory
FU	Functional Unit
GA	Genetic Algorithms
M & S	Marshall and Swift
MF	Membership Function

MOC	Materials Of Construction
MV	Match Value
N/A	Not Applicable
PE	Process Engineering
PEI	Process Economics International
PID	Piping and Instrumentation Diagram
RS	Reaction Steps
SEE	Standard Estimate Error
SP	Shape Parameter
SSA	Standard Set of Attributes
TMV	Total Match Value

Chapter 1

INTRODUCTION

This chapter introduces the research presented in this thesis. The reasons for and objectives of the research are described. An outline of the thesis completes the chapter.

1.1 Background

This thesis is about estimating the capital costs of chemical process plants in the early stages of their development.

A ***process*** is a set of connected actions which convert raw materials into a chemical product. The ***plant*** is the pieces of equipment for transforming materials (reactors, distillation columns, etc.), connecting pipework and other material transfer devices and infrastructure (laboratories, storage, etc.), that make the process happen. The ***capital cost*** of a chemical process plant is the investment required to design, purchase, build, install and start up its equipment, ancillary facilities and infrastructure. The ***Cost Of Production (COP)*** is the cost of producing one unit of main product, including for example, raw materials, utilities, labour costs, maintenance, tax, insurance, and financial provisions.

In the early business planning and feasibility assessment stages of a project, cost estimates are required in order to evaluate viability and to compare the economics of the alternative processes and the different operating conditions of individual processes that are under consideration for the plant. There is limited knowledge about a new plant in the early stages of process development. Nevertheless, accurate cost estimates are needed to prevent incorrect decisions being made, such as terminating the development of a would-be profitable plant.

1.2 Reasons for the Research

The initial motivation for this research was the inadequacy of existing cost estimation methods for use in the early stages of process development and the lack of new estimating methods in the open literature. The last significant new method to be published was by Klumpar et al (1988). Significant developments in computers and software have been made in the years since then. For example, powerful computers are now commonplace and developments in the computing fields of artificial intelligence have produced new techniques such as neural networks, fuzzy logic and case-based reasoning.

These new artificial intelligence techniques were untested as methods for estimating the capital cost of chemical process plants. Yet, they have properties that make them worthy of consideration as estimating techniques. For example, the ability to simulate experts' reasoning with software that aids decision making, using the same rules as would be used by experts (expert systems). Techniques which generate solutions to problems using the existing data from similar problems (neural networks and case-based reasoning).

Plant data supplied by Chem Systems Limited (see section 1.3 Chem Systems Data) allowed a comparison of the effectiveness of existing estimating methods with the new techniques over a wide range of chemical plants.

1.2.1 Original Brief

The original brief provided for this research as specified by Dr. D. W. Edwards follows:-

ECONOMIC FEASIBILITY ASSESSMENT USING PLANT COST DATABASES

Evaluation of potential processes for chemicals production needs highly qualified and experienced people and it is time consuming and expensive. However, there exists a huge body of data referring to existing plants, which could be used to develop correlations for use in such evaluations - thereby saving time and money.

A collection of existing plant capital and operating cost and process data has been made available to Dr. David W. Edwards. The number of processes is about 230. The project will be:

1) Devise a database structure for the data, load it and run consistency checks.

2) Calculate statistics for publication from the data.

3) Develop correlations for estimating the capital and operating costs of industrial-scale processes from data generated in the laboratory.

4) Highlight potential technical difficulties, safety and environmental problems in the process design.

These aims changed during the course of the research, with the third aim broadened in scope to include new techniques in addition to correlation and the fourth being omitted.

1.2.2 Research Aim and Objectives

The aim of this research is to produce as accurate a method as is possible for the early estimation of the capital costs of chemical plants using data available at the time of compiling such an estimate. The first objective is to find out the capabilities of the existing methods using the Chem System data, the second is to develop new estimating techniques and finally, after comparing the different methods, nominate the best method for early capital cost estimation.

Artificial Intelligence techniques were thought suitable for estimating. The field of AI was reviewed with the objective of finding methods which could be used as novel techniques for the estimation of the costs of chemical process plants. Then techniques which showed promise were adapted for cost estimation and tested using the Chem Systems data.

1.3 Chem Systems Data

Process specifications, capital cost and COP data for 210 different chemical process plants were supplied by the chemical industry consultancy firm Chem Systems Limited. This ***Chem Systems (CS)*** data has two important advantages. Firstly, it refers to a wide range of actual plants with commercially significant products and processes. This is illustrated by the wide range of values found for the plant capital costs and COP, as shown in table 1.1.

Secondly, the data is expressed on a common basis for all the plants: a typical, large, modern plant, forming part of a large chemicals complex in West Germany in mid 1988. This avoids the problem encountered with most sets of plant cost data, that of converting the data from different sources to a common basis. Normally, in order for the cost data to be standardised upon a year and location requires cost indices (see section 2.3) to adapt data from different years and location factors (see section 2.4) for plants in different locations.

The cost data was provided in computer spreadsheets with a common format and the process and plant data were in the form of printed process descriptions and flowsheets showing main equipment items. Process descriptions were available for 166 of the plants. Some descriptions were more detailed than others and therefore there were variations in the number of chemical process plants for which a value was known for the various process specifications.

Table 1.1 Range of Values for the CS Data

	Capital Cost[a] (million $ US)	**Cost of Production** ($/tonne of main product)	**Capacity** (tonnes/year)
Average	52.5	1050.4	112 000
Maximum	330.0	7719.7	990 000
Minimum	2	32.8	2 250
Standard Deviation	56.7	1048.2	144 000

[a] All costs are for plants constructed in West Germany in mid 1988

The values of the following process specifications were provided in the CS data:-

- Number of functional units: The number of significant steps in the chemical plant process.
- Capacity: The amount of product that a plant can produce within a given period, usually a year.
- Throughput: A measure of the amount of material passing through the process.
- Maximum temperature: The maximum temperature reached during the process.
- Maximum pressure: The maximum pressure reached during the process.
- Materials of construction: The materials required to make the pieces of equipment in the plant.
- Phase of process: The physical state of the materials that pass through the plant, for example, gas, liquid, solid or a combination.

The CS data set is large and diverse compared with that used in the derivation of other methods. The published research on chemical plant cost estimation uses the known data of only a few plants, averaging about 40 per method, and for some of the methods the data used is only for plants of a certain type. For example, Gore (1969) provides a method for gas phased processes.

Estimating methods use data from plants constructed at some point in the past. Hence, there is a need to correct for the changes in costs due to inflation and other factors, when estimating the costs of a new plant. In this thesis the cost estimate is calculated for the same year as that of the data used to produce the estimate. So updating is avoided. The important concerns of the research in this thesis are the techniques used for cost estimating and not issues, such as the effects of inflation.

A database was devised for storing the data for the chemical processes as supplied by Chem Systems Limited using the PARADOX relational database management system.

1.4 Outline of Thesis

Chapter 2 introduces capital cost estimation, explains its importance and classifies estimates prepared during the development of a new chemical plant.

Chapter 3 describes published early capital cost estimation methods. The methods are grouped into three types of estimate: exponent, factorial and functional unit.

In chapter 4 the performance of methods used for the early estimation of capital costs of chemical plants is assessed.

A new method for the early estimation of the capital cost of chemical plants is described in chapter 5. The methods are developed from the CS data using standard regression techniques.

Chapter 6 assesses the effect that computers have had on capital cost estimating and then looks at how the latest computer techniques could be used in the future.

The fuzzy matching methodology and its suitability for capital cost estimation are explained in chapter 7.

The results achieved when using fuzzy matching to estimate the capital cost of chemical plants are presented and analysed in chapter 8.

The final chapter discusses why fuzzy matching is a better capital cost estimating technique than those that already exist. There is also a brief discussion of how fuzzy matching could be applied to other fields of chemical engineering.

Chapter 2

COST ESTIMATION

This chapter introduces capital cost estimation, explains its importance and classifies estimates prepared during the development of a new chemical plant.

2.1 Capital Cost Estimation

Developing a new chemical process and then designing and building the plant to implement it takes a long time and costs a lot of money. After the initial idea or identification of business opportunity, the work starts with laboratory experiments and simple design studies, it proceeds through a small-scale or 'pilot' plant and more detailed designs. The whole undertaking culminates in the detailed drawings and specifications from which the plant is built.

The capital cost of the plant represents the expenditure of current wealth in the expectation of future benefits. An accurate ***Capital Cost Estimate (CCE)*** must be determined for a proposed plant in order to gauge the size of investment and to assess whether the plant can generate sufficient returns to make this investment worthwhile. A detailed CCE is also used for project cost control during construction and commissioning.

The total capital cost investment is split into two parts, the ***Battery Limits Capital Cost (BLCC)*** and the ***Offsites***.

The BLCC corresponds to the cost of the manufacturing installation of the plant, including its equipment, preparation of the site and construction. The manufacturing installation is considered to be the part of the plant which imports the raw materials, utilities (such as electricity, water, fuel and refrigeration) and other required chemicals, catalysts and solvents. It exports the manufactured products and any by-products and surplus generated utilities.

The offsites investment covers the cost of: the equipment that produces and distributes the utilities (such as steam boilers and cooling towers), storage facilities for the raw materials and final products, roads, waste disposal facilities, buildings and laboratories.

There is a problem with apportioning the cost of the offsites when the plant is part of a multi-plant site. These type of sites share the offsites costs, for example the cost of the roads from the site is split between all the plants. So the offsites costs for a plant in a multi-plant site is lower than those for an equivalent plant on a single plant site.

Concentrating on estimating the BLCC part of the capital costs avoids the above problem with offsites. Hence, the capital cost estimating methods that are developed later in this thesis are all for the estimation of the BLCC.

Capital cost estimates are made at various stages during the development of a chemical plant and are a key input to the decision whether to continue with the project. In the later stages of design there are well-established and accurate estimation methods. However, the large costs incurred while developing a new chemical process are a powerful incentive for using quick and accurate methods to estimate the capital cost and then distinguish between economic and uneconomic plants at an early stage. Also, methods which produce a cost estimate quickly will cost less to use.

The estimates are made before project completion and with incomplete information, so some error is inevitable. The error decreases as more information becomes available about the plant during its development from an idea to a fully operational plant. The error of any estimate must be minimised without incurring too great an increase in the cost of producing the estimate. This is illustrated by plotting the increase in cost versus accuracy of estimate, see figure 2.1.

The cost of an estimate with a certain accuracy is represented by the relative cost factor. This factor is calculated by dividing the cost of an estimate with a particular accuracy by the cost of developing an estimate with an accuracy of ± 30%. For example, if an estimate with an accuracy expected to be in the range of ±5% costs ten times that of an estimate with an accuracy of ±30%, then its cost factor would be 10.

Figure 2.1 Relationship Between Cost and Accuracy

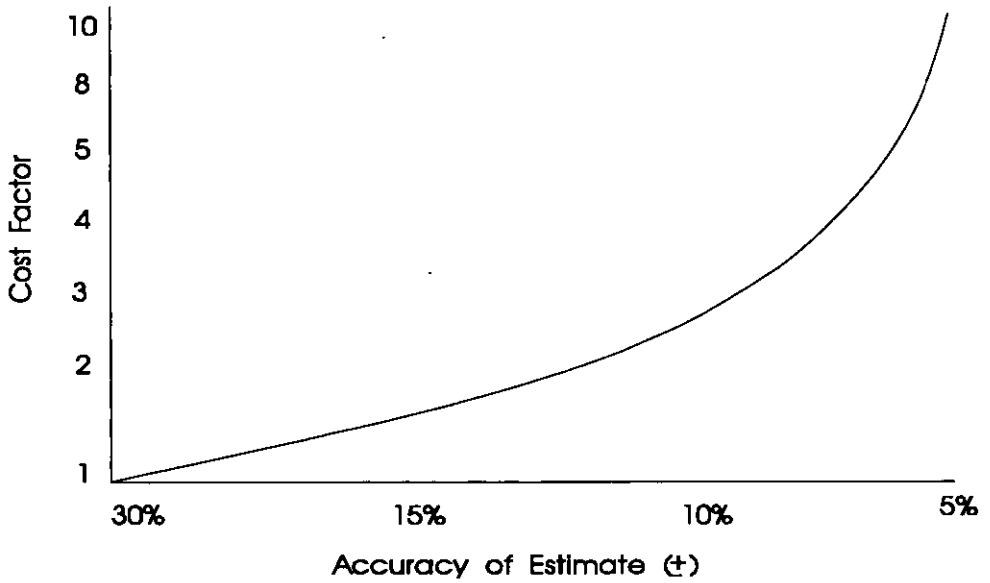


Figure 2.1 shows how the cost of estimating starts to rapidly increase as an accuracy of 10% or better is attained (Kharbanda and Stallworthy, 1988).

2.2 Classification of Capital Cost Estimates

The ***American Association of Cost Engineers (AACE)*** published in 1958 a generally accepted classification of the different types of estimate with the accuracy that should be achievable by professional estimators. The AACE classification of estimates is shown in the first four columns of Table 2.1. Garrett (1989) has added 'realistic' error bands, which he claims should be achievable by the 'average engineer', and his own estimate of the cost of developing the estimates. He argues that his error estimates are more reasonable, particularly in the early stages, when professional estimators would not be involved.

Table 2.1 shows that the 'realistic' accuracy achieved by 'average engineers' in the early stages of process development is poor, while the estimates are costly. Even the first, 'back-of-envelope' calculation can cost $5,000, while the smallest error realistically achievable is ±40%. The expected accuracy for the earlier estimates will in fact vary depending on the requirements and procedures of the company considering constructing the new chemical plant. This is due to the varying amount of information about the process that is required for an early estimate by different companies.

There are numerous ways of classifying the different types of capital cost estimates for chemical plants (Liddle, 1978). This leads to confusion in the terminology, for example, Jelen and Black (1983) call the first two estimate stages the preliminary estimates, but the same name is used for the third estimate in the AACE classification. There would be less confusion if the estimates were classified by their purpose, as opposed to using names which have different meanings for different classifications (Liddle, 1978). The classifications defined by AACE in 1958 are used throughout this thesis.

The techniques used to compile the various types of estimate are well documented in the literature and so only a brief explanation of the five AACE types of estimates follows. The descriptions mention: the names used in other classifications, how the estimate is calculated, the information that must be available to calculate the estimate, and the purpose of the estimate. The explanations are a general outline of each estimate and variations will be found for different companies.

Table 2.1 Characteristics, Accuracy and Cost of Capital Cost Estimation

Type of Estimate	Prepared by	Leads to Approval of	Possible Error[a] Expected	Realistic	Possible Cost of Estimate (89 $'s)
Order of Magnitude	Individual Engineer	Inexpensive Study	40%	40-100%	2,000-5,000
Study	Project Group	Expensive Study	25%	30-50%	11,000-50,000
Preliminary	Contractor, Professional Estimator	'Detailed Design and Market Research	12%	20-35%	50,000-200,000
Definitive	Contractor	Construction of Plant	6%	10-15%	150,000-700,000
Detailed	Contractor	Continue Construction	3%	5-10%	1-5%[b]

[a] ±Percentage of Capital Costs [b]Percentage of Capital Costs

This thesis is concerned with the first two stages of the AACE system. The new methods described in later chapters are alternatives to the existing methods used when making an estimate of the capital cost in the early stages of development, that is the order of magnitude and study estimates.

2.2.1 Order of Magnitude

The Order of Magnitude is also commonly known as a 'back-of-envelope', 'seat of the pants', 'quickie', 'guesstimate', or 'ball park' estimate. The estimate should be inexpensive and quick to prepare, as the names suggest.

Historical data for existing plants is often used to calculate this type of estimate. The estimate is normally an 'exponent estimate', and is calculated by using the 'sixth tenths rule' (Williams, 1947b) or 'two thirds power law'. The capital cost of an existing plant, which uses the same process as the new plant, is multiplied by a factor calculated as the ratio of the capacities of the new and existing plants raised to an exponent which is on average between 0.6 and 0.7. The range of possible values for an exponent is between 0.5 and 1. The method is described in more detail in Chapter 3. However, the sixth tenths rule for overall plant costs can not be used when the proposed plant uses a new process.

Another way to estimate the capital cost at this early stage is for a skilled estimator to find a plant for which they had previously estimated the capital cost or which they know about, and which is 'like' the new one. They then adjust for

any slight differences using various explicit or implicit 'rules' and produce an estimate. For example they might use the sixth tenths rule for the difference in capacity. This estimating method is emulated later in the thesis by fuzzy matching (see chapter 7).

The 'functional unit estimates', first developed by Zevnik and Buchanan (1963), are also used at this stage (they are described in chapter 3). The number of main equipment units can be found from the plant block diagram or process flowsheets and then used to multiply an average unit cost calculated from other process parameters known in the early stages of design.

The estimate is used as a rough screen of further interest in different process proposals. A high contingency of 30%, if not more, must be allowed because the level of error of the estimates at this stage is in the range 40-100%.

2.2.2 Study

The study is alternatively known as conceptual, evaluation, predesign, feasibility, or factored CCE. It uses 'factors' to multiply the equipment costs of the plant to allow for installation and ancillary equipment. Some functional unit methods are also used at the study stage, as they require data which is not known until this later stage of development.

The first method to use factors was developed by Lang (1947b). Lang multiplied the total equipment cost by a single factor to produce an estimate of the capital cost of the plant. In other methods (Hand, 1958) different factors are used for different types of equipment or an overall factor is calculated by combining factors calculated from different process parameters (Hirsh and Glazier, 1960). The complexity involved with calculating the factors varies. These factorial methods are described in more detail in Chapter 3.

In order to estimate the equipment costs a preliminary process flow sheet must be prepared that shows all the major items of equipment. Then material and energy balances are needed to provide the process pressures, temperatures and stream compositions for each major piece of equipment. Next the pieces of equipment are sized and their materials of construction determined. Finally, sometimes a factor

is determined which represents the complexity of the plant. Then with all this information an estimate of the equipment costs can be made.

The sixth tenths rule has been adapted from capital cost estimation to the calculation of equipment costs from previously constructed pieces of equipment costs. An equipment list for the new plant is required and so the estimate produced is a study estimate.

The accuracy of this type of estimate is expected to be around 25%, although Garrett (1989) maintains that their accuracy is more likely to be in the range of 30-50%.

The estimates found at this stage are again used to analyse the economic feasibility of new processes and to compare the capital cost for different processes or process variations, such as a larger capacity. The company must be confident that the investment needed to construct the plant is affordable and will result in a profit. The approval for an expensive study by contractors and professional estimators depends on the results of this feasibility study.

From this point in the sequence of estimates the BLCC is based upon equipment costs derived from quotations and contracts with manufacturers, rather than process data. The estimate shifts from being based on previous plant and equipment costs to actual equipment quotations and estimates of material and labour requirements.

After this stage professional estimators and the companies that will construct the equipment and plant are required for the calculation of the other types of estimates listed in table 2.1. The development needed is of a greater detail and not comparable with the early estimates on which this thesis concentrates. However a brief description of these later stages of estimates follows for the sake of completeness.

2.2.3 Preliminary

A preliminary estimate is also know as a budget, execution or scope estimate.

The professional estimator requires the specifications used in the study estimate but in more detail. The piping and instrumentation diagrams must be available (PIDs). A plot plan is also required, including storage areas, warehouses, etc. A more detailed development of the plant means that the process parameters are more likely to be correct. This means the factors calculated from the process parameters are also better and so there is a higher degree of confidence in the capital cost estimate. Also, the major equipment costs are at least quoted over the phone by the equipment vendors and in some cases provided in written detail.

The improved accuracy is sufficient for approval from the management of the company commissioning the plant for the provision of funds for further detailed design and development.

2.2.4 Definitive

The definitive estimate is also known as the project control estimate.

The plant design details are almost complete and in the majority of cases finalised at this stage of estimate. Nearly all the quotes for major equipment will have been received. The construction schedule and labour costs are also calculated.

With this level of information a definitive estimate is made with an accuracy within ±10%. The funds for plant construction are authorised and the plan for project cost control is set.

2.2.5 Detailed

Detailed estimates are commonly called the final, tender, control, fixed price, firm price and contractors price.

All the design, drawing and specifications are complete. An estimate is made based on the actual equipment costs, labour projections and a scale model.

The accuracy of the estimate should be within ±5%. Allowing the plant constructors to give a 'firm' price for cost of construction to the client. This confirms the cost of constructing the plant.

2.3 Cost Indices

A cost estimate is required before a plant is built, therefore the estimate is needed for the current year or some point in the near future. However, the methods used to estimate the capital cost in the early stages of development will have been derived from data for previous years, therefore producing an estimate for that period in time. ***Cost Indices*** must be used to adjust for the difference in the cost of goods and services at the two different points in time. The cost index will nearly always increase the cost with time. In the case of chemical plant construction this will take account of increases in material prices and erection labour costs (I. Chem. E. and Assoc. Cost Engrs., 1988).

The relationship used by a cost index is:-

$$\text{Present Cost} = \text{Historical Cost} \times \left[\frac{\text{Present value of Cost Index}}{\text{Historical value of Cost Index}}\right] \qquad (2.1)$$

There are different indices for different purposes and different countries. The main ones are:-

United Kingdom cost indices:

- Association of Cost Engineering (ACE) index for erected plant costs
- Process Engineering (PE) index for plant costs

USA cost indices:

- Engineering News-Record (ENR) indices for building and construction costs
- Chemical Engineering (CE) index for plant costs
- Marshall and Swift (M & S) index for installed equipment
  (formally know as Marshall and Stevens index)

International cost index:

- Process Economics International (PEI) index for plant costs

Cost indices should only be used to update the type of costs for which they were specifically designed (Humphreys, 1987). For example, updating the capital cost of a complete plant using the M & S index would be inadvisable, as the index is for installed equipment costs.

The statistics used in the derivation of the cost index also need to be considered (I. Chem. E. and Assoc. Cost Engrs., 1988) as the cost index for capital costs is calculated by looking at the current data on relevant cost elements, such as environmental regulations, equipment and wages. Each of these elements is weighted in the calculation of the new value for the cost index. With the weights dependent upon the sort of plants for which the cost index is being developed. Therefore, applying a cost index to another types of plant will lead to a less accurate cost adjustment. The cost engineer makes a judgement on which of the available cost indices is most suitable by considering which cost elements are most relevant in the calculation of the capital cost for the plant under consideration.

The usually accepted limit of the period over which indices can be used to correct costs is five years (Allen and Page, 1975). There will be a lack of confidence in the value for updates of over five years because indices do not take account of the comprehensive changes in legislation concerning environmental, health and safety standards, altering market conditions, technological advances, and productivity gains. Indices also tend to be updated about every five years and if the basis of the statistics used to derive them changes then misleading results may occur (I. Chem. E. and Assoc. Cost Engrs., 1988).

2.4 Location Factors

Location Factors are used to adjust the capital cost of a plant constructed in one part of the world to the equivalent capital cost of an identical plant constructed in some other part. Location factors are available for countries and regions within countries.

The location factor is for a set point in time, but a factor for a different date can be calculated using cost indices and exchange rates (I. Chem. E. and Assoc. Cost Engrs., 1988). However the use of cost indices means that changes for periods of longer than five years are open to doubt.

Chapter 3

EXISTING METHODS FOR EARLY CAPITAL COST ESTIMATION

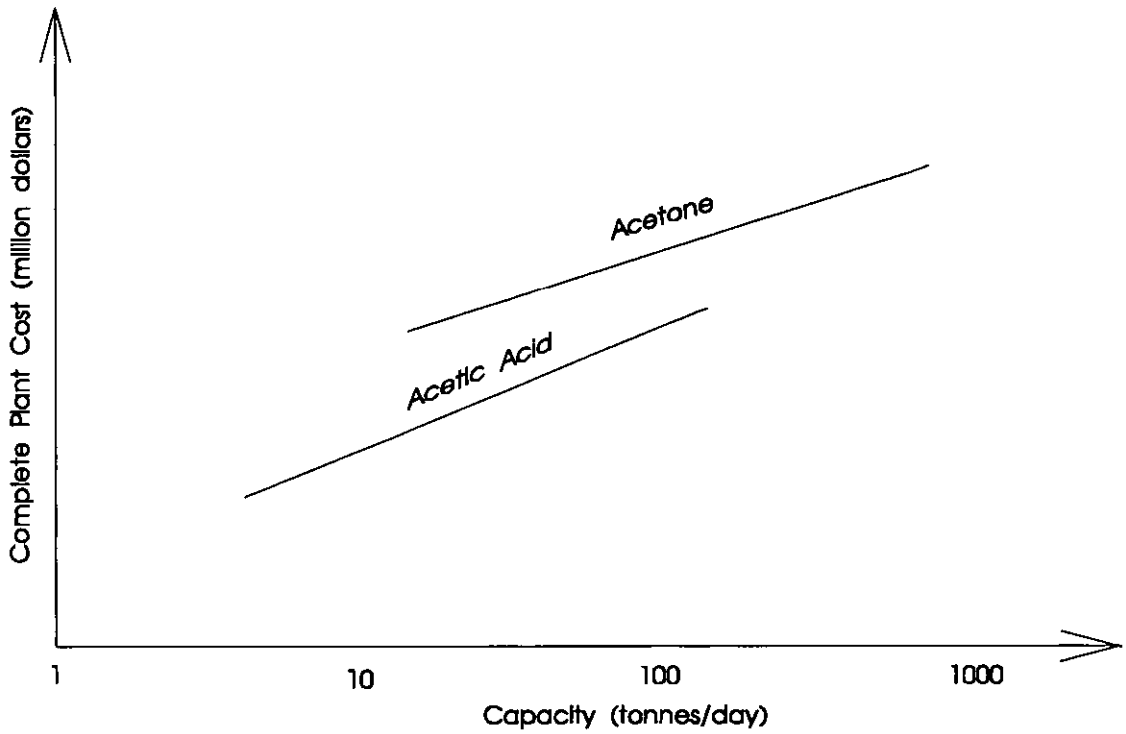
This chapter describes published order of magnitude or study capital cost estimation methods. A literature survey found numerous papers on capital cost estimation methods. These methods can be split into three types. Those that estimate the capital cost of a new plant with data from operational plants, adapted by taking the ratio of capacities and raising by an exponent, these are called ***exponent estimates***. In the second type the estimating methods use factors to multiply certain plant costs to produce a cost estimate for the overall plant, these are called ***factorial estimates***. Thirdly, methods that use equations whose variables are the plant parameters known in the early stages of development and with the number of functional units included as one of the variables are called ***functional unit estimates***. A review of the CCE methods found for each type of estimate follows.

3.1 Exponent Estimates

A simple sixth tenths exponent or two thirds power law method was introduced by Williams (1947b) for estimating the cost of equipment for a chemical plant, using the known cost of existing pieces of equipment and the ratio of the equipment capacities raised to the exponent. Williams mentioned that the exponent method would be suitable for estimating the capital cost of plants and this was shown to work by Chilton (1950).

Exponent estimates are used in the very early stages of a process plants design and are therefore always an order of magnitude estimate.

Figure 3.1 Complete Plant Cost Estimating Chart



There are two different representations of the method, one uses charts with capital cost versus capacity lines for particular processes (see figure 3.1). The other uses equations similar to equation 3.1, where the equations correspond to the lines on the chart.

$$\begin{pmatrix}\text{Capital cost}\\ \text{of new plant}\end{pmatrix} = \left(\frac{\text{Capacity of new plant}}{\text{Capacity of old plant}}\right)^{n} \times \begin{pmatrix}\text{Capital cost}\\ \text{of old plant}\end{pmatrix} \qquad (3.1)$$

The value of the exponent n varies between about 0.5 and 1, with the value corresponding to the slope of the line.

The method has been updated for charts to 1970 (Guthrie, 1974) and to 1987 (Garrett, 1989) and for exponents to 1990 (Remer and Chai, 1990). Sixth tenths is a value of the exponent claimed to have general applicability, with specific exponents available for individual chemical processes (Remer and Chai, 1990), see table 3.1.

Table 3.1 Example Process Exponents

Process	Product	Size Range (1 000 tons/year)	Exponent
Acrylonitrile	Acrylic Fibre	3-24	1.02
Methanol	Acetic Acid	2-64	0.59
Hydrocarbons	Acetylene	3-45	0.65
Bauxite	Alumina	34-365	0.54

Recent analysis (Remer and Chai, 1990) has found the average exponent value over many processes to be around 0.67. Liddle (1992) has theoretically justified this two thirds power law for the cost of cylindrical vessels which keep their dimensional similarity by finding that the amount of material needed to construct the vessel increases in proportion to the two thirds power of the capacity.

An exponent of 0.6 means that there is an economy of scale. That is, as the capacity of a plant is increased then the resulting increase in capital cost is not as great. For example, doubling the capacity would result in the capital cost increasing by a factor of $2^{0.6}$ or 1.6.

The variation of the value of the exponent for specific chemical plants is illustrated by considering two plants at the different ends of the possible range of exponent values. A sulphuric acid plant has major plant items which increase in cost at a slower rate than the capacity, due mainly to the amount of material needed to build the equipment not increasing as quickly as capacity. Therefore it has a very low exponent of 0.54. At the other extreme, a polyethylene plant has a number of parallel production lines and a large increase in size requires further production lines. There is no economy of scale and hence such plants have a very high exponent of 0.97.

However, the capacity of a plant can not be increased indefinitely and result in an economy of scale. The plant reaches a capacity where the pieces of equipment can not be fabricated or transported due to their size. Before this point is reached costs start to increase rapidly. Kharbanda and Stallworthy (1988) suggest that one exponent value is not adequate over all the possible capacities of a plant and there should be two or three different exponents for different capacity ranges.

The disadvantage of the exponent method is that it requires the capital cost to be known for an existing plant with a process identical to that of the new plant. Therefore, the method is no good for estimating the capital cost of plants with new processes. Also, the exponents for a specific process are calculated from the gradient of the line between two or more points, plotted on a graph of capital cost versus capacity, for operational plants using the same process. An exponent can not be calculated for a process when there is only one operational plant using that particular process. An assumption can be made that the process is average and therefore the exponent needed is the average of all other processes where a exponent has been calculated (Garrett, 1989).

Uppal and Van Gool (1992) have found that the best accuracy obtainable from the exponent method is ±40%.

Two other methods, which are similar to the exponent method, replace the capacity with the cost per unit of product or the yearly value of product sold (the turnover ratio) (Garrett, 1989). The relationship between these quantities and the capital cost of the plant is found from data on existing plants, as in the exponent method described above. These methods are only mentioned briefly as they have an accuracy, at best, of ±100% (Hill, 1956). The accuracy of the methods is poor because a chemical product which is cheap to produce will be estimated to have a low capital cost. However, the product will often be sold in large quantities and therefore be produced by a plant with a large capacity and a higher than expected capital cost.

3.2 Factorial Estimates

Lang (1947b) introduced the first factorial method for capital cost estimating, and proposed a relationship between the cost of plant equipment and the capital cost (Lang, 1948), see equation 3.2.

$$C = f \cdot E \tag{3.2}$$

Where:

- C = Capital cost.
- f = Factor.
  - 3.10 for solid processes.
  - 3.63 for solid and fluid processes.
  - 4.74 for fluid processes.
- E = Delivered equipment costs.

The equipment costs may be established from a mixture of quotations from vendors, previous equipment costs and any published data. The costs can be scaled up using exponents, in an identical fashion to the method described in the last section. The way in which the equipment cost is found will determine at which stage of estimation the Lang method can be applied. Quotations will mean using the method at the study stage, but costs from previous plants enables an order of magnitude estimate to be made, when the equipment list is known .

The overall factors are in fact made up of four parts, one for foundations, supports, chutes, vents, insulation and installation of equipment, another part for piping costs, a third for construction costs, that is the civil engineering and buildings and the fourth for overheads. The overheads costs are for contingency, temporary constructions, engineering expense and engineer/contractor fees (Lang, 1948). The factors used were calculated from the analysis of 14 preliminary estimates for chemical plants.

Lang (1948) classified processes into three types:-

- Solid
- Solid and Fluid
- Fluid

Factors were then determined for each type of process. Cran (1981) suggests using a universal factor of 3.45, instead of classifying the plants. This factor is the average of the factors for the 14 plants analysed by Lang. The average factor is acceptable due to the statistically low number of plants making the difference in factors insignificant.

The factorial method has been developed by many authors. A brief discussion follows, but a detailed investigation was not possible as the methods were complicated and are used more at the preliminary and later stages of estimation, rather than in the early estimating stages researched in this thesis. Also, the CS data does not include any equipment costs and so the methods can not be tested.

Hand (1958) introduced the idea of considering the plant one piece of equipment at a time and using factors to multiply the cost for each piece to produce the capital cost. Guthrie (1969) also developed factors which could be used to multiply the base cost of equipment in order to get the installed cost. The factors were split into two parts, one set of factors for multiplying the total equipment cost and the other set consisted of factors specific to particular pieces of equipment, for example columns, furnaces, heat exchangers etc. Comprehensive data on 42 existing plants was used to derive the specific factors.

Chilton (1960) used factors for 8 different components of a chemical plant to produce an estimate of plant costs from the total installed equipment costs. The factors were for the following components:-

- Piping
- Instrumentation
- Manufacturing buildings
- Auxiliary facilities
- Outside production lines
- Engineering and Construction
- Contingencies
- Size

The value for each factor was determined using information about the plant. For example, the auxiliary facilities factor would require a knowledge about where the plant was to be built, the facilities already provided at the site and the new

facilities required by the new plant. The detailed nature of this method meant that the estimate could only be made at an even later stage in development than the other factorial methods.

Hirsch and Glazier (1960) used more component factors than Lang and introduced the concept that the size of the factors depends on the average cost of the basic equipment. With an increase in average cost leading to a decrease in the factor. The reason for this is that the cost of some of the components, such as the manufacturing of buildings cost, that make up the capital cost will not increase in cost due to more expensive equipment but will be a relatively smaller value.

Miller (1965) states that the factors depend on the size of the equipment, the materials of construction and the operating pressure. Which are in effect accounted for by the value for the average cost of pieces of equipment in the process. This theory of the average equipment cost being linked to size, construction materials and pressures in the process is the theoretical basis of the functional unit methods described in the next section. The method developed by Miller becomes very detailed when calculating the ten parts that make up the overall factor. With the factor used depending on process details and the average cost of equipment. The ten parts to the factor were:-

- Field erection of basic equipment
- Foundations and structural supports
- Piping
- Insulation of equipment
- Insulation of piping
- All electrical equipment
- Instrumentation
- Miscellaneous
- Architecture and structure of buildings
- Building services

The actual values for the factors should be developed by each company that uses the method. Taking into account the preferred requirements, that are specific to a company, for the construction of a chemical plant. For example, the requirements for the plant layout, instrumentation techniques, degree of oversizing and provision of installed spares will all effect the value of the factors.

The stage at which the factored method is used and its resulting accuracy depend on how the equipment cost is established and how much detail is needed for determining the factors. The experience of the estimator is also important in selecting the best values for the factors. If Lang factors are used and the equipment costs are calculated from a basic flow sheet of the major equipment pieces then an order of magnitude estimate is possible. However, using a complicated method, such as that devised by Miller (1965), and equipment quotations from the manufacturer will provide estimates in between the study and detailed estimation stages.

3.3 Functional Unit Estimates

Functional unit estimating methods use the values of the fundamental process parameters known in the earliest stages of process development to predict the capital cost of the corresponding plant. Examples of these parameters are the number of functional units, product capacity or maximum throughput (for processes with a low conversion reaction), reaction pressure, temperature, and materials of construction. These process parameters are known as the ***Attributes*** of the plant for the rest of this thesis. An explanation of these attributes and their usage, for example maximum, minimum or average temperature, are best illustrated by describing how they are used in each of the existing functional unit methods.

Functional unit estimation can be an alternative to factorial methods, but is normally used to produce estimates at an earlier stage than is possible with factorial methods. With estimators tending towards one or the other of the methods, for example, Garrett (1989) prefers the factorial methods. The major difference between the two methods is in the way equipment costs are considered. The factorial method requires a value for the cost of each piece of equipment in the plant. Whereas, functional unit methods in effect use the values of the attributes for a particular plant in the derived equations to estimate the average cost of a unit. With the unit consisting of one or more pieces of equipment and auxiliaries.

The stage at which the functional unit methods can be used to estimate depends, like the factorial method, on the parameters used in the equation(s). Most methods require data available at the study stage.

The methods have been derived by statistical analysis of existing plant data and are usually expressed by a combination of equations, charts, tables and plant classifications. Most of these methods assume that the process route and outline flowsheet, that is raw materials and sequence of significant process steps to the product(s), has been worked out.

The idea that the number of functional units influenced the costs of a chemical plant was introduced by Wessel (1953), who used the 'number of steps' in calculating the labour costs. Functional units (Zevnik and Buchanan, 1963) have

also been termed significant steps (Wilson, 1971), operating units (DeCicco, 1968), major operating steps (Viola, 1981), process steps (Taylor, 1977), and process modules (Klumpar et al, 1988), but they all represent the same basic concept. The term ***Functional Unit (FU)*** will be used in this thesis.

The definition for a functional unit varies, but the following is given as an example (I. Chem. E. and Assoc. Cost Engrs., 1988):

> a functional unit is a significant step in a process and includes all equipment and ancillaries necessary for operation of that unit. Thus the sum of the costs of all functional units in a process gives the total capital cost.

The functional units split the process up into parts where a change occurs to the material passing through the plant, for example, a reaction or separation. However, the specific pieces of equipment which are used to achieve the change are not important as far as the method is concerned. Hence, these methods require a broad specification of the process and avoid the need to determine precise details about every piece of equipment.

Determining the value for the number of functional unit attribute presents a difficulty. Counting the number of functional units in a process is subjective, because it depends on what the cost engineer takes to be a significant step in the process. There is no precise definition of a functional unit, and the methods use different definitions.

Twelve different examples of functional unit estimation methods are described in their chronological order. Methods that have been updated to nearer the current year using cost indices have the most recent equation shown and the year of the original derivation stated. A cost index would be used when an estimate is required for a year different to that for which the method was derived, see section 2.3.

An analysis was made of the occurrence of the different attributes in the twelve functional unit methods for early capital cost estimation that are described later. This showed which chemical plant attributes were occurring most frequently and

Table 3.2 Number of Occurrences of Attributes in Functional Unit Methods

Attribute	Occurrences
Number of Functional Units	11
Pressure	10
Temperature	9
Materials of Construction	9
Capacity	7
Phase of Process	6
Throughput	5
Process Type	1

are therefore considered by other authors as the most important attributes to include when capital cost estimating, see table 3.2.

The number of functional units is the most common as would be expected, because this is the name given to the methods after all. However, capacity and throughput are alternative ways of representing the volume of material passing through the plant and when added together total 12, making this the most popular attribute.

Some process parameters in the CS data are not used in the functional unit estimating methods. These unused attributes were the workforce and the number of reaction steps, a description of these attributes is given in section 5.1.1.1.

3.3.1 Hill

Hill (1956) was the first to present a short cut capital cost estimating method that uses something akin to functional units. However, it differs slightly from the rest, because a major piece of equipment in a process can be counted as one unit (stripper, absorber), two units (compressor) or as even more units, depending on the relative cost of a particular piece of equipment. The method is similar to factorial methods except that the installed equipment costs are estimated by equation (3.4). The factors developed by Chilton (1949) give the value for F. This method was developed for fluid processes. The equation used is:-

$$C = F \,.\, IEC \tag{3.3}$$

C = Capital cost in 1954 US dollars.

F = Chilton factor to allow for piping, instrumentation, manufacturing buildings, auxiliaries, outside lines, engineering and construction, contingencies and size factor.

IEC = Installed Equipment Cost in dollars.

$$= N \,.\, \left[\frac{Q}{10}\right]^{0.6} . 30\,000 \tag{3.4}$$

N = Total number of units.

Q = Capacity in million lb per year.

3.3.2 Zevnik and Buchanan

The description of this method (Zevnik and Buchanan, 1963) was the first to use the term functional units. The method also involves the plant capacity and factors for the maximum temperature, maximum pressure and materials of construction. It was devised for use with fluid processes. The ***Cost Per Functional unit (CPF)*** is read from a graph of CPF versus capacity in million lbs per year. Different graphs are used depending on the ***Complexity Factor (CF)***, see below. DeCicco (1968) and Ward (1984) developed equations for calculating the CPF (equation 3.7), which could replace the graphical method.

The capital cost is calculated using:-

$$C = 1.33 \cdot N \cdot CPF \tag{3.5}$$

C = Capital cost in 1963 million US dollars.
N = Number of functional units.
CPF = Cost per functional unit, determined from a graph using the complexity factor CF.

$$CF = 2 \cdot 10^{(F_t + F_p + F_m)} \tag{3.6}$$

F_t = Maximum temperature factor is determined from a graph of temperature °K versus temperature factor.
F_p = Maximum pressure factor is determined from a graph of pressure atm. versus pressure factor.
F_m = Materials of construction factor is read from a table of construction materials and factors.

The method with the graphs replaced by equation is:-

$$C = k \cdot N \cdot Q^{0.6} \cdot 10^{(F_t + F_p + F_m)} \qquad (3.7)$$

C = Capital cost in 1992 pounds sterling.

Q = Capacity in tons per year.

k = 6270 $\quad Q \leq 4464$ tons per year

or

= 4400 $\quad$ 4464 tons per year $< Q$

N = Number of functional units.

F_t = $1.80 \times 10^{-4}\ (T_{MAX} - 300)$ Temperature above ambient (300K).

T_{MAX} = Maximum temperature in K.

or

= $0.57 - (1.9 \times 10^{-2} \cdot T_{MIN})$ Temperature sub ambient.

T_{MIN} = Minimum temperature in K.

F_p = $0.1 \cdot \log (P_{MAX})$ Pressure above ambient (1 atm).

P_{MAX} = Maximum pressure in atm.

or

= $0.1 \cdot \log \left(\frac{1}{P_{MIN}} \right)$ Pressure sub ambient.

P_{MIN} = Minimum pressure in atm.

F_m = Read from a table, as in the original method.

3.3.3 Gore

Gore (1969) developed his method using data from gas-phase processes. The attributes used are: number of functional units, throughput, maximum temperature, maximum pressure and materials of construction.

Gore used throughput in place of capacity and calculated this by multiplying the capacity by an empirically derived recycle factor (see the discussion of the methods developed by Bridgwater for more details on the calculation of throughput). The reason for using throughput is that the size of the plant, and therefore the cost, is more closely related to the throughput as it represents the amount of material passing through the plant. Whereas, the capacity only measures the amount of product produced.

Gore used a volumetric basis since the size of the equipment is dependent on the volume and not the mass passing through the process in gas-phased processes. So the throughput was in lb moles per year instead of a more normal lb per year, because moles are proportional to the volume of a gas. The reason for this not being a more common unit for throughput is the difficulty in expressing the throughput in molar terms, because of complex minerals and materials having unknown molecular weights (Bridgwater, 1976).

Although a material factor is shown in equation (3.8), the dependence was never evaluated because the plants from which the method was developed were considered to be not significantly different in this attribute. The equation that Gore derived was:-

$$C = 4680 \cdot N \cdot Q^{0.62} \cdot T^{x} \cdot P^{0.395} \cdot F_m \qquad (3.8)$$

C = Capital cost in 1967 US dollars.
N = Number of functional units.
Q = Throughput, million lb moles per year.
T = Temperature factor.

$$= \left(\frac{T_{max} - 300}{300}\right)$$

T_{max} = Maximum temperature in Kelvins.

$$x = \frac{Q^{0.206}}{2.52}$$

P = Maximum pressure in atmospheres.
F_m = Materials of construction factor.

3.3.4 Stallworthy

Stallworthy (1970) developed a sophisticated method from the Zevnik and Buchanan (1963) method. In previous methods the flow had been assumed to be the same throughout the whole of the process. Stallworthy modified the method to take account of the different conditions for side streams, such as recycles. The temperature, pressure and materials of construction must be known for each side stream, resulting in the method requiring a lot of information in order to produce an estimate. The attributes used are capacity, number of functional units, pressure, temperature and materials of construction in each stream, and the ratio of stream flow to the flow of the main output of product.

The equation is:-

$$C = \frac{0.0075}{A} \cdot \sum_{i=1}^{S} (N_i \cdot F_{t_i} \cdot F_{p_i} \cdot F_{m_i} \cdot R_i) \qquad (3.9)$$

C = Capital cost in 1970 pounds sterling.

A = $6.2 \times 10^{-5} \cdot Q^{-0.65}$

Q = Capacity, long tons per year.

S = Number of main product and process side streams.

N_i = Number of functional units in stream i.

F_{t_i} = Temperature factor for stream i.

F_{p_i} = Pressure factor for stream i.

F_{m_i} = Materials of construction factor for stream i.

R_i = Ratio of flow for stream i to the flow of the main product stream.

3.3.5 Wilson

Wilson (1971) uses a quite simple method , with the attributes: capacity, number of functional units, temperature, pressure and materials of construction. An investment factor is also used, which is determined from a graph of the ***Average Unit Cost (AUC)*** versus the investment factor. The AUC is based on the dominant phase of the process and the capacity. The investment factor varies between about 1.3 and 4.1 and is analogous to a Lang factor (1947b). The material of construction factor, F_m, is read from a table of factors and materials of construction. With the factors value varying between 1 (mild steel) and 2 (titanium). The pressure factor, F_p, applies when the pressure is outside of the range 1-7 bars and the temperature factor, F_t, applies when the temperatures is outside of the range 0-100°C. F_p and F_t are determined from a graph of the factor versus the pressure (psia) or temperature (°C).

$$C = 10 \cdot f \cdot N \cdot AUC \cdot F_m \cdot F_p \cdot F_t \qquad (3.10)$$

- C = Capital cost in early 1987 pounds sterling.
- f = Investment factor.
- N = Number of functional units.
- AUC = Average unit cost.
  - $= 21 \cdot V^{0.675}$
  - V = Capacity, tons (long) per year.
- F_m = Materials of construction factor.
- F_p = Pressure factor.
- F_t = Temperature factor.

3.3.6 Bridgwater

Bridgwater has been the most prolific publisher of papers in early capital cost estimation. All of his methods have been for liquid and/or solid phase processes. His first methods were published in the early seventies and were based on work done by Gore (1969).

Method 1 (Bridgwater, 1974) states that the capital cost is independent of the number of functional units and only depends on throughput. This method was derived from effluent treatment processes. However all his subsequently published methods including the most recent involve functional units.

Method 2 (Bridgwater, 1974) and method 3 (Bridgwater, 1978) were last mentioned in papers published in 1976 and 1981 respectively; neither are now used, even by Bridgwater. Method 3 correlated the capital cost to the number of functional units, throughput, temperature and pressure. However, the method only provided an equation for plants with a throughput over 60,000 tonnes/year.

Bridgwater uses $\left(\frac{Q}{s}\right)$ to represent the throughput of the plant, where Q is the plant capacity and the s is the 'conversion' factor. This conversion factor represents the efficiency at which the reactor is converting raw materials into product. A value of one would mean that all of the input into the reactor is converted to product. This is unlikely in a chemical process and so the efficiency will normally be lower than one and therefore the throughput value will be higher than the capacity.

The current Bridgwater method is method 4 (Bridgwater, 1978). Which is still used and has been updated from 1978 too 1992 by Bridgwater (Bridgwater, 1994). The method uses less attributes than the previous two methods and is split into two equations. The one that is used depends on the throughput of the plant. He also used the same variables in a linear equation but the results were not improved.

Method 1

$$C = 39 \cdot G^{0.83} \tag{3.11}$$

C = Capital cost in 1974 pounds sterling.
G = Throughput, gallons per hour.

Method 2

$$C = 50.26 \cdot N \cdot \left(\frac{Q}{s^{0.5}}\right)^{0.85} \cdot \left(\frac{T \times n}{N}\right)^{-0.17} \cdot \left(\frac{p \times n'}{N}\right)^{0.14} \tag{3.12}$$

C = Capital cost in 1969/70 pounds sterling.
N = Number of functional units.
Q = Capacity, long tons per year.
s = Reactor "conversion".
$= \left(\frac{\text{weight of desired reactor product}}{\text{weight of reactor input}}\right)$
T = Maximum temperature, °C.
n = Number of functional units operating at a temperature above $\frac{T}{2}$.
p = Maximum pressure, atm.
n' = Number of functional units operating at a pressure above $\frac{p}{2}$.

Method 3 **(only valid for $\frac{Q}{s}$ > 60 000 tonnes per year)**

$$C = 193 \cdot N \cdot \left(\frac{Q}{s}\right)^{0.665} \cdot e^{(2.58 \times 10^{-7} \cdot Q)} \cdot T^{-0.022} \cdot p^{-0.064} \qquad (3.13)$$

C = Capital cost in first quarter 1975 pounds sterling.
Q = Capacity in tonnes per year.
e = Natural logarithm base, 2.71
T = Maximum temperature, °C.
p = Maximum pressure, atm.

Other symbols are the same as in method 2.

Method 4

$$C = k \cdot N \cdot \left(\frac{Q}{s}\right)^{X} \qquad (3.14)$$

C = Capital cost in 1992 pounds sterling.
Q = Capacity in tonnes per year.
k = Constant.

k = 133 300		$\frac{Q}{s}$	≤ 60 000 tonnes per year
k = 1520	60 000 tonnes per year <	$\frac{Q}{s}$	
X = 0.3		$\frac{Q}{s}$	≤ 60 000 tonnes per year
X = 0.675	60 000 tonnes per year <	$\frac{Q}{s}$	

Other symbols are the same as in method 2.

3.3.7 Allen and Page

This is a complicated estimating method (Allen and Page, 1975) and is hard to use in the early stages of plant design because the method requires a lot of plant data, some of which is unlikely to be known at this point in design. The method sets out to simplify the Stallworthy method by not requiring data about each stream. Throughput, number of functional units, maximum temperature, maximum pressure, and material of construction are used in the estimation of the capital costs. This method can be used to estimate the capital cost of a process for any of the material phases that are possible for a chemical plant.

$$C = f \cdot DEC \tag{3.15}$$

C = Capital cost in 1972 US dollars.
f = Factor allowing for other costs outside DEC.
= 4.76 for fluid processes.
DEC = Delivered equipment cost.
= N . SF . BIC
N = Number of functional units.
SF = State factor
= $F_t \cdot F_p \cdot F_m$
These three factors are calculated using Wilson's techniques.
F_t = Maximum temperature factor.
F_p = Maximum pressure factor.
F_m = Mean materials of construction factor, Wilson factors used.
BIC = Basic item cost, read from graph using TP value.
TP = Throughput, lb mol/year.
= CAP . FF . PF
CAP = Total plant feed, lb mol/year.
FF = Flow factor.

$$= \frac{\sum_1^N \left(\begin{array}{c} \text{Number of input and output streams} \\ \text{for each functional unit} \end{array} \right)}{N}$$

PF = Phase factor.

$$= 0.0075 + \frac{VI}{N}$$

VI = Volume items - number of main plant items which operate with material in the gas phase.

3.3.8 Taylor

Taylor devised a method for ICI using "Process Step Scoring" between 1972-75 (Taylor, 1977). For each significant process step, basically another term for a functional unit, a complexity score is calculated. The process steps are actually defined by example in the paper. The complexity score for each process step is the sum of scores: 0, 1, 2 or 3 for the following parameters:

- Relative throughput
- Reaction time
- Storage time
- Temperature
- Pressure
- Materials of construction
- Explosion
- Dust
- Odours
- Toxicity
- Reaction in fluid bed
- Distilling materials
- Tight specifications
- Film evaporation

The complexity score for each step is summed to get a costliness index and from this the capital cost. Therefore the method requires a lot of data to be known about the process. The constants K and p (equation 3.16) are derived from the regression between capital cost and the plants costliness index and capacity:-

$$C = K \cdot I \cdot Q^{p} \qquad (3.16)$$

C = Capital cost.
K, p = Constants
I = Costliness index.
$= \sum_{i=1}^{N} 1.3^{y_i}$
N = Number of process steps.
y_i = Complexity score calculated for process step i.
Q = Capacity, in 1000 tonnes per year.

Equation for continuous process plants

$$C = 298\,500 \cdot I \cdot Q^{0.39} \qquad (3.17)$$

C = Capital cost in 1992 pounds sterling.

Equation for batch process plants

$$C = 298\,500 \cdot I \cdot Q^{0.6} \qquad (3.18)$$

C = Capital cost in 1992 pounds sterling.

3.3.9 Timms

Timms (1980) gives two equations for inorganic and organic gas-phase processes, a simple one with just capacity and number of functional units and a more complicated one with the extra attributes: maximum temperature, maximum pressure and a materials of construction factor, F_m. The F_m is quoted for steel only and varies between 1.0 (carbon steel) and 1.3 (high grade stainless steel) these factors are identical to the material factors developed by Wilson (1971).

Method 1

$$C = 8\,300 \cdot N \cdot Q^{0.615} \tag{3.19}$$

Method 2

$$C = 3\,860 \cdot N \cdot Q^{0.639} \cdot F_m \cdot T_{max}^{0.066} \cdot P_{max}^{-0.016} \tag{3.20}$$

C = Capital cost in 1992 pounds sterling.
N = Number of functional units.
Q = Capacity, tonnes per year.
F_m = Materials of construction factor, Wilson factors used.
T_{max} = Maximum temperature, degrees K.
P_{max} = Maximum pressure, bars.

3.3.10 Viola

This method was published in 1981 (Viola, 1981) and uses a plant complexity factor to account for the variability of plant costs due to plant complexity, overall operating conditions, materials of construction and plant throughput. The capital cost estimate is determined from a graph of capital cost versus the complexity factor K, with a set of curves for different capacities.

Viola states that the major operating steps of the plant, which are equivalent to functional units, are the key influence on the capital cost. The other attributes used are pressure, materials of construction, throughput and the fraction of steps in the process that are solid/fluid mixtures. The correction factor for average throughput and the correction factor for pressure and materials of construction are found by reading off the factors value from a graph of correction factor versus the relevant process attributes.

This method is applicable to solid/fluid and fluid phase only plants.

Solid/Fluid plant method

$$K = N \cdot S \cdot \phi\left(\frac{I}{O \times N}\right) \cdot (1 - 0.6 \cdot f_S) \tag{3.21}$$

Fluid plant method

$$K = N \cdot S \cdot \phi\left(\frac{I}{O \times N}\right) \tag{3.22}$$

C = Capital cost is read from graph using K and capacity, graphs derived from 1981 data.

K = Complexity factor.

N = Number of major operating steps.

S = Correction factor for pressure and materials of construction, read from graph.

$\frac{I}{O}$ = Input to output ratio over the whole plant, raw material to product.

$\phi\left(\frac{I}{O \times N}\right)$ = Average throughput correction factor, read from graph.

f_S = Fraction of major operating steps handling solid/fluid mixtures.

3.3.11 Klumpar, Brown and Fromme

The method (Klumpar et al., 1988) was designed to cater for a wide range of processes and capacities. The plant capacity is replaced by the average throughput in this method. A lot of work has also gone into the description of the process modules, which is another phrase for functional units.

$$C = k \cdot F \cdot N \cdot G^{v} \qquad (3.23)$$

C = Capital cost, the year and currency will depend on the data from which the k is calculated, the value given below for k is for 1981 dollars.

k = Constant factor, given as 1.1×10^{2}.

F = Complexity factor.

$= 2 \cdot \left(10^{T + P + M}\right)$

$T = 1.8 \times 10^{-4} \cdot (t - 27)$ for $t \geq 27°C$ (hot process)

or

$= 2.0 \times 10^{-3} \cdot (27 - t)$ for $t < 27°C$ (cold process)

t = Temperature in °C.

$P = 0.1 \cdot \log p$ for $p \geq 1$ atm (pressure process)

or

$P = 0.1 \cdot \log \left(\frac{1}{p}\right)$ for $p < 1$ atm (vacuum process)

p = Pressure in atmospheres.

M = Material of construction factor, Zevnik and Buchanan factors used.

N = Number of process modules.

G = Average throughput, lbs per hour.

$$= \frac{\text{(Sum of all module throughputs)}}{N}$$

$$\text{Module Throughput} = \frac{\left(\begin{array}{c}\text{Sum of all the inlet and outlet} \\ \text{stream flowrates in the module}\end{array}\right)}{2}$$

ν = Exponent

Constant exponent = 0.57

$$\text{Variable exponent} = 0.83x_1 + 1.05x_2 + 0.59x_3 + 0.47x_4 + 0.59x_5 + 1.07x_8 + 0.6x_9 + 0.83x_{11} + 0.4x_{12} \quad (3.24)$$

x_i = Number of process modules of type i divided by the total number of process modules.

The constants for x_6, x_7, x_{10} were found to be zero, meaning that the value of exponent ν was unaffected by the number of modules of types 6, 7, and 10.

The different types of process modules or functional units are defined in Table 3.3:-

Table 3.3 Different Types of Functional Units

Type	Parameter Changed	Example of Operation	Example of Equipment
1	Temperature	Heat exchange	Heater, cooler, exchanger
2	Pressure	Compression	Compressor, blower
3	Location of solids	Receiving, storing	Conveyor system, reclaimer
4	Location of liquids	Liquid collection	Systems of tanks, pumps
5	Particle size	Comminution	Crusher, ball mill
6	Particle size distribution	Screening	Screen, cyclone
7	Solid body shape	Casting, briquetting	Extruder, press
8		Storage	Tank farm, pond
9	Number of streams	Liquid mixing	Stirred tank, kneader
10	Phase	Melting, evaporation	Furnace, evaporator
11	Phase distribution of components	Absorption	Column, dryer
12	Composition of isotope	Chemical or nuclear reaction	Reactor

3.3.12 Tolson and Sommerfeld

This method, published in 1990 (Tolson and Sommerfeld, 1990), is the most recent method and the most simple, using only one attribute, the capacity. The author's idea was to develop a method which could be used without a process flowsheet, but still have an accuracy in the range expected for an order of magnitude estimate. Therefore, necessitating the omission of all the attributes used in the previous methods, with the exception of capacity.

The method is not a functional unit method. However, it is included in this section as the equation is of a similar form to the functional unit methods and is not an important enough method to merit a section of its own.

$$C = 0.75 \cdot V^{0.677} \tag{3.25}$$

C	=	Capital cost in 1987 million US dollars.
V	=	Capacity in million lbs per year.

Chapter 4

PERFORMANCE OF EARLY CAPITAL COST ESTIMATION

In this chapter the performance of methods used for the early estimation of capital costs of chemical plants is assessed.

4.1 Accuracy

The performance of the estimating methods will be assessed by the accuracy of their estimates. Two metrics are used in this thesis to measure the estimate accuracy. The first metric uses a standard percentage error, ***Standard Estimate Error (SEE)***, calculated by:-

$$\text{SEE (\%)} = \left(\frac{\text{Estimated capital cost - Actual capital cost}}{\text{Actual capital cost}}\right) \times 100 \qquad (4.1)$$

The metric is the ***Average Standard Estimate Error (ASEE)*** of the capital cost estimates for a set of plants. It is calculated by taking the arithmetic mean of the absolute values of the SEE, calculated with equation 4.1.

Allen and Page (1975) pointed out that equation 4.1 assigns the same significance to errors that are above or below the actual value by equal amounts. However, there is no limit to the size of an overestimate, but there is a limit to an underestimate, because a chemical plant can not have a negative capital cost. This means that the possible error range is from -100% to an infinitely large positive percentage. Therefore, if the two extremes are judged to be equally significant then it follows that an underestimate is a worse estimate than an overestimate of the same amount.

In order to assign greater significance to underestimates Allen and Page (1975) recommended the use of equation 4.2 to convert the SEE for an underestimate to

the SEE that would be obtained for an equivalent overestimate, ***Equivalent Estimate Error (EEE).***

$$\text{EEE (\%)} = \left[\frac{-100 \times \text{Underestimate SEE}}{100 + \text{Underestimate SEE}}\right] \quad (4.2)$$

However, an alternative equation, see equation 4.3, can be used to calculate the same EEE directly from any estimated capital cost which is less than the actual capital cost.

$$\text{EEE (\%)} = \left[\frac{\text{Actual capital cost} - \text{Estimated capital cost}}{\text{Estimated capital cost}}\right] \times 100 \quad (4.3)$$

The second metric is defined as the average of the SEE for the overestimates and the EEE of the underestimates. This is the ***Average Equivalent Estimate Error (AEEE).***

The application of this equation results in a sizeable increase in magnitude for the negative errors. For example, an underestimate of -50% (half the actual cost) would be changed to an equivalent overestimate of +100% (twice the actual cost). The AEEE more accurately reflects the effectiveness of the method and indicates when an estimating method is achieving a low ASEE by consistently underestimating.

When looking at the error of an estimate there must be some consideration of the error in the 'actual' costs. The actual capital cost of a plant is the capital cost calculated once the plant has been built and started up, but it is only as accurate as the accounting systems used in its calculation. Therefore, the error calculated from the actual cost will itself have a certain amount of error.

The rest of the chapter analyses the accuracy of the types of estimating methods: exponents and functional units. The accuracy of the methods claimed by the original author is compared to that found by others who have tested the accuracy of the methods and to that found when the methods are used to estimate the capital cost of the process plants in the CS data. Then the methods are adapted by normalising and re-correlating using the CS data.

4.2 Accuracy Claimed by Authors

The actual accuracies claimed for functional unit estimates are listed in table 4.1. Early capital cost estimates should have an accuracy in the range of ±30% (Wilson, 1971), but a more realistic figure is ±50% (Garrett, 1989). Most authors of methods claim that the achieved average accuracy is within the range ±30% to ±50%. The *N/A* stands for information 'not available' and all the errors are calculated using the ASEE metric.

The resulting accuracy of methods tested by independent sources is shown in addition to the accuracy claimed by each author for their own method, see table 4.1. Allen and Page (1975) tried four different methods and found that the estimates produced had a lot larger error than claimed when tried on 4 different processes (8 plants). The Stallworthy (1970) method produced the best estimates with an ASEE of under ±30%, a low ±18%. Ward (1984) found the accuracy of the two methods he tested on 8 processes to be slightly higher than claimed, but still close to the desirable ±30%.

Table 4.1 Functional Unit Methods - Quantity of Data and Claimed Accuracy

Method	No. of Processes	No. of Plants	Accuracy Claimed by Author	Ward	Allen & Page
Viola	28	N/A	±15%	±33%	
Gore	11	65	±20%		
Bridgwater (3)	16	24	±20%		
Stallworthy	N/A	N/A	±20%		±18%
Hill	N/A	N/A	±25%		±500%
Zevnik & Buchanan	N/A	N/A	±25%	±29	±47%
Bridgwater (4)	16	24	±25%		
Allen & Page	N/A	8	±25%		±25%
Klumpar et al.	N/A	20	94% within ±30%		
Wilson	16	N/A	81% within ±30%		±43%
Taylor	N/A	45	95% within +36% & -26%		
Tolson & Sommerfeld	N/A	40	±37%		
Timms	103	N/A	No Accuracy Claimed		

Some of the functional unit methods were tested by Bridgwater (1978), but for estimating the cost of just one plant. This does not give any confidence in a method as it might be very accurate for the one plant but highly erroneous for all others.

A more complete analysis of the accuracy of some of these methods is found in the next section, where the ability of the methods for estimating the capital cost of chemical plants is tested on the CS data.

4.3 Accuracy Obtained with CS Data

The CS data provided a large sample of chemical plant data for testing the published methods. Moreover, all the data was on the same basis. The results of the tests undertaken are presented and discussed in this section.

4.3.1 Exponent Estimates

Garrett (1989) published a set of charts for estimating the capital cost of chemical plants using the exponent method. The charts were assembled from four sources: Guthrie (1974), Chemical Engineering (1973/1974), Kharbanda (1979) and construction notices in Construction Alert, published by the journal, Chemical Engineering. The CS data allowed the accuracy of the charts to be tested for 79 plants, which is a large sample size in this field of research.

The charts were for the estimation of the capital cost of a plant in April 1987; this differs by only one year from the CS data, which was for 1988. Cost indices were used to correct for the difference in cost year. The use of indices is acceptable for differences of less than five years (Allen and Page, 1975).

The capacity was stated in tonnes per day for the charts, whereas the CS data was in tonnes per year. The capacity quoted in the CS data was divided by 333.33 days in order to convert into tonnes per day, because the CS data was for plants that operated for 8000 working hours per year.

The charts being generated from various different data sources led to problems. In some cases the offsites had been included with the capital cost, but in others it was left out of the cost and it was not possible to distinguish between the two. This meant that on average the charts should over estimate the BLCC of the CS plants and underestimate the sum of offsites and BLCC. This was found to be the case when the results were compared. However, the average error for estimates of the capital costs of the CS plants made using the charts provides a general idea of the potential accuracy of the exponent method.

The exponent is found by taking the gradient of the line through two or more points plotted on a logarithmic graph of capacity against capital cost. Therefore, the capacity and capital cost must be available for two or more plants with the

same process for an exponent to be calculated. An assumed exponent was used in charts by Garrett when the data was known for only one plant. The assumed exponent had a value of 0.64, an average found from charts developed by Guthrie (1974). The plants with this assumed exponent would be expected to produce less accurate estimates.

The lowest ASEE was ±70% which was found using the charts to estimate the sum of capital cost and offsites. This was achieved with the majority of estimates (81%) being underestimates of the sum of capital cost and offsites. This high proportion of underestimates is confirmed by the AEEE 226% for the same estimates. The lowest value found for the AEEE was 154% (with a corresponding ASEE of ±81%). This was achieved when the estimates of the capital cost without the offsites had been updated using a cost index. The percentage of underestimates was then a more normal 63% of estimates.

The conclusion from these results is that the exponent method does not produce accurate estimates for the capital costs, even allowing for the fact that the charts might be producing an estimate for the capital costs plus the offsites in some cases. The best AEEE of 154% is a large average error and its ASEE of ±81% is near to the ±100% expected by Hill (1956).

4.3.2 Factorial Estimates

The factorial method estimates the capital cost by multiplying equipment costs by factors which allow for the other costs. The CS data had no equipment costs and so could not be used for testing the factorial methods.

4.3.3 Functional Unit Estimates

The functional unit methods were tested in three ways. Firstly, the equations were used as stated, apart from a realignment for the difference between a method and the CS data in the year of construction, location and unit of currency. The second and third tests made allowances for the fact that the equations had been derived from a different set of plant data and type of processes to that of the CS data.

The second test normalised the methods to remove any systematic error by correcting the estimates so that their average is the same as that for the actual capital costs. While in the third test the equations and methodology remained the same, but the equations were re-derived using the CS data.

Table 4.2 shows how some methods are only applicable to processes with specified phase(s). The phase is the state of the majority of the materials as they pass through the plant, that is gas, liquid, solid or a combination of these. Fluid means the ability to flow and therefore represents gas and liquid phases. Such methods should be inaccurate when applied to the whole of the CS data, which includes chemical plants exhibiting all the different process phases and combinations thereof. For a method to be applied effectively to all of the plants in the CS data requires a realignment of the method so it represents all of the possible process phases and combinations. This is assuming that the method is not of a form which is only suitable for a certain phase.

Table 4.2 Applicable Process Phases for Functional Unit Methods

Method	Applicable Process Phase
Hill	Fluid
Zevnik & Buchanan	Fluid
Gore	Gas
Stallworthy	All Phases
Wilson	Fluid and/or Solid
Bridgwater (1)	Liquid
Bridgwater (2)	Liquid-Solid
Bridgwater (3 & 4)	Liquid and/or Solid
Allen & Page	Fluid
Taylor	All Phases
Timms	Gas
Viola	Solid-Fluid, Fluid
Klumpar	All Phases
Tolson	All Phases

4.3.3.1 Testing the Methods

Some of the functional unit methods were too complicated to be used for early capital cost estimation. Others required data not included in the CS data and so could not be tested. For example, the Stallworthy (1970) method needs data on the process streams. The following methods were not tested for the reasons listed:-

Hill	Hill provided no definition of which types of equipment to count as 1, 2 or 3 units.
Gore	The recycle data required for this method was unavailable. Also, Gore requires the throughput in moles per year, which could not be calculated from the CS data for liquid and solid phase plants because their number of moles is unrelated to volume.
Stallworthy	Certain values were needed for each of the process streams, which were not in the CS data.
Bridgwater	Method 1: The throughput attribute was unavailable in gallons per hour. Method 2: The temperature and pressure were needed for each functional unit. These were not in the CS data.
Allen and Page	The throughput in moles per year and the number of functional units in the gas phase were not included in the CS data.
Taylor	The complexity score was impossible to calculate because some values were not in the CS data. Also, the required values would not normally be known in the early stages of estimation.
Viola	The materials of construction factor requires the percentage of the equipment which is fabricated from the special material, this was not included in the CS data.
Klumpar et al.	This method required the throughput for each module, which was missing from the CS data.

This left the following seven functional unit methods to be tested:-

1)	Zevnik and Buchanan	(Method 2)
2)	Wilson	
3)	Bridgwater	(Method 3)
4)	Bridgwater	(Method 4)
5)	Timms	(Method 1)
6)	Timms	(Method 2)
7)	Tolson and Sommerfeld	

4.3.3.2 Methods used as Stated by Author

The CS data refers to plants constructed in West Germany in mid 1988, with the costs stated in US. dollars. This meant that the accuracy of methods could only be tested by estimating the capital cost on this basis. Most of the methods had been updated (I.Chem.E. and Assoc. Cost Engrs., 1988) to estimate the cost in pounds sterling of a plant constructed in the UK in early 1987.

The methods were updated to allow for the difference in construction years by using a factor derived from an appropriate cost index for the country in which the method was developed. These were a cost index published by Gerrard (1994) for the UK, and the Chemical Engineering (CE) index for plant costs for the USA. Then an approximate factor of two was used as a combined factor which took into account the change of location and the currency exchange rate in 1988. The factor resulted from multiplying the value 1.1 for changing the location of the plant from the UK to West Germany (I.Chem.E. and Assoc. Cost Engrs., 1988) by the value 1.78 for the pounds to US. dollars exchange rate (Cost Engineer Journal).

A standard set of 79 plants was used to compare the accuracy of the methods. Values for all of the attributes needed for each of the methods being tested were known for each plant in the set. The results are presented in figure 4.1.

Figure 4.1 Accuracy of Methods used as Stated by Authors on 79 CS data Plants

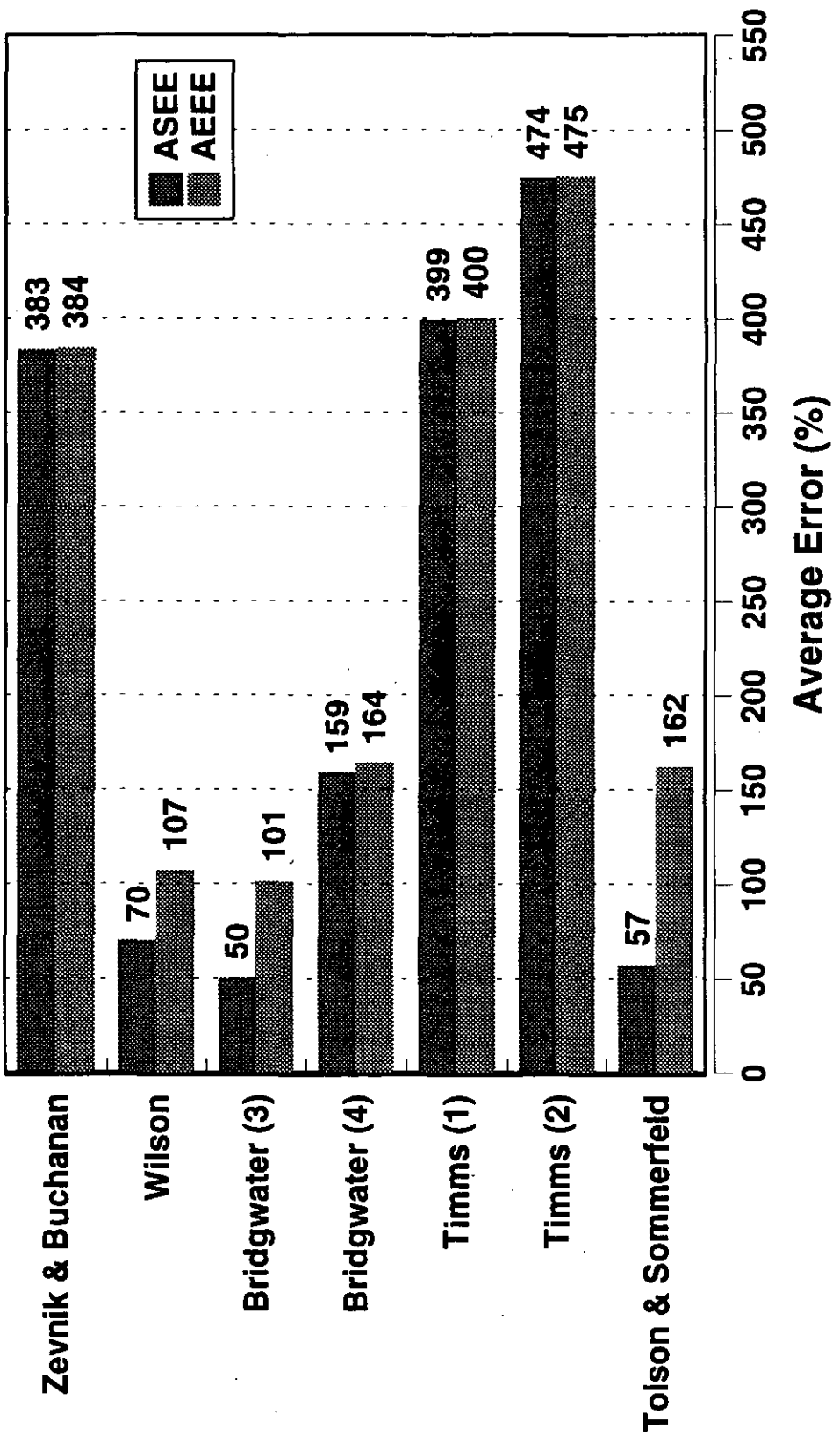


Table 4.3 Accuracy of Methods used as Stated by Author on CS data for plants with Appropriate Phase

Method	Phase	Number of Plants	Error	
			ASEE (±%)	AEEE
Zevnik & Buchanan	Fluid	73	391	393
Wilson	Fluid &/or Solid	79	70	107
Bridgwater (3)	Liquid &/or Solid	38	**43**	**84**
Bridgwater (4)	Liquid &/or Solid	38	143	143
Timms (1)	Gas	22	532	532
Timms (2)	Gas	22	621	621
Tolson & Sommerfeld	All Phases	79	57	162

The accuracy obtained when using the methods as specified by the authors, but adjusted for year and location was very poor. The best standard error (ASEE) was ±50% when using Bridgwater (3). However, a lot of these estimates were underestimates, as shown by the significant increase in the value for the equivalent error (AEEE), 101%. Bridgwater (3) also had the best AEEE.

There are three reasons why the existing methods give such high errors.

1) The methods have been updated to refer to the same cost year as the CS data by Bridgwater and the author. The period between derivation of the methods and the CS cost year (1988) is in many cases greater than 5 years. There is also the possibility of inaccuracies in cost indices. This will be corrected by normalising.

2) The CS data set is different to the data sets used to derive the existing methods. With the data for different years, types of plant and from different sources. Re-correlating the equations will correct for this.

3) The methods were developed for a certain type of plant (fluid, gas, solid etc.). This problem was tackled by classifying the types of the CS plants and then using each method on only the type of plants for which they were developed. The results are displayed in table 4.3.

The results did improve, but were still not within the desired error range. The Bridgwater(3) method produced the best AEEE of 84% and the lowest ASEE of ± 43%. The improvement was about 20% for the methods using fluid phases, but led to an increase in error for the methods designed for vapour phase plants. This might be caused by the vapour phase plants being more diverse in their attribute values and capital costs. Therefore, making the representation of the capital cost of vapour phase plants by an equation more erroneous than other phase types.

4.3.3.3 Normalised Methods

This test normalised the method estimates, using equation (4.4), so that the average of the estimates is the same as the average of the actual capital cost values in the CS data. This procedure removes any systematic error.

$$\text{Normalised Estimate} = \frac{\text{Average Actual Capital Cost}}{\text{Average Estimated Capital Cost}} \times \text{Initial Estimate} \quad (4.4)$$

For example, a method is used to estimate the capital cost and then requires a cost index to be used to correct for the difference in the year for which the estimate was produced and the year for which it is required. However, the cost index value might not be applicable for the plants being estimated and is say only two thirds the value required. Therefore, all the estimates are too low and need to be corrected. Normalisation produces a multiplying factor of 1.5 and this corrects the problem with the initial estimates. The normalising factors obtained for each method are listed in table 4.4 and the resulting accuracies shown in figure 4.2.

Table 4.4 Factors for Normalising Methods

Method	Factor
Zevnik & Buchanan	0.22
Wilson	0.85
Bridgwater (3)	1.23
Bridgwater (4)	0.54
Timms (1)	0.24
Timms (2)	0.19
Tolson & Sommerfeld	2.07

Figure 4.2 Accuracy of Normalised Methods on 79 CS data Plants

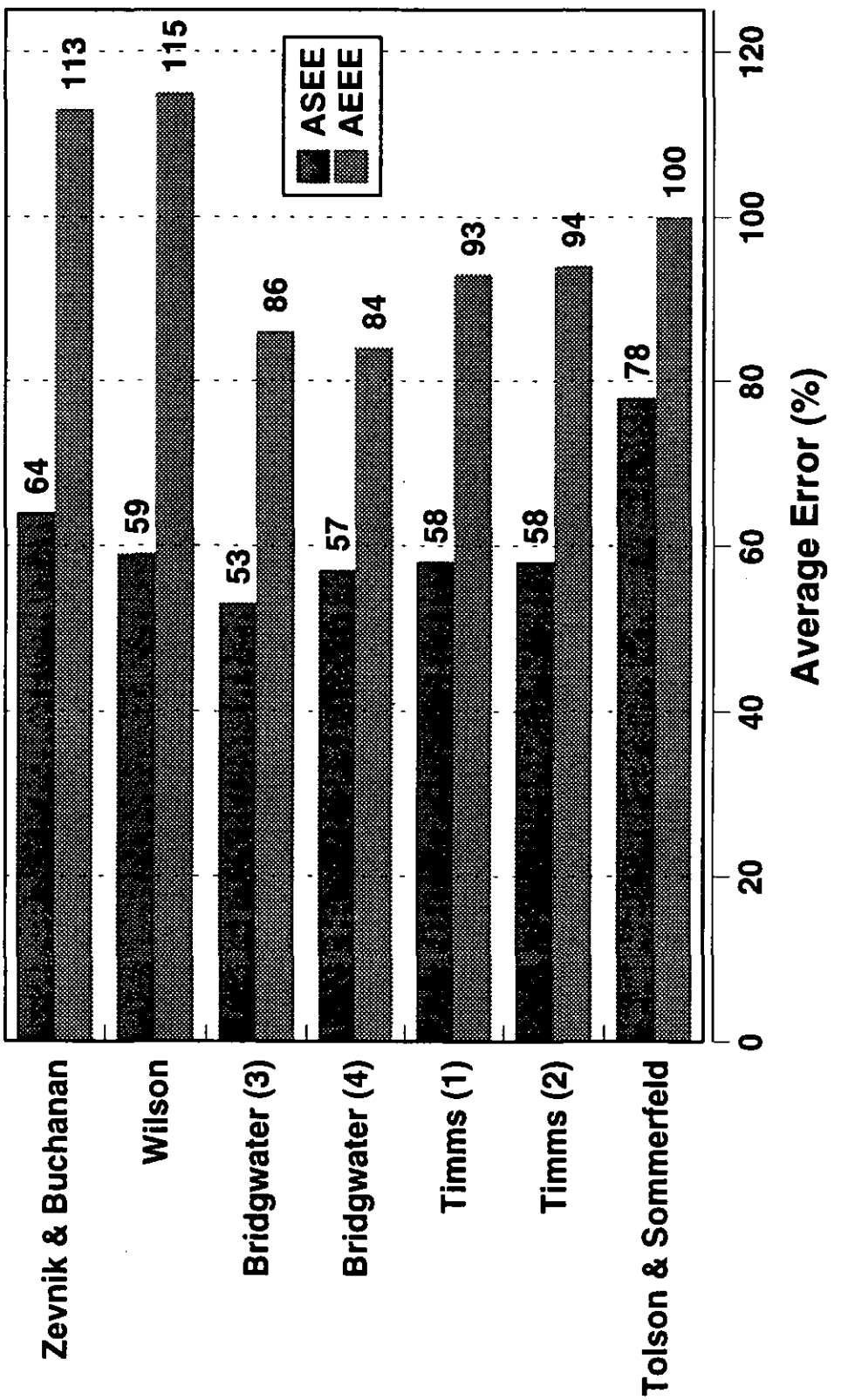


Table 4.5 Accuracy of Normalised Methods used on CS data for Plants with Appropriate Phase

Method	Phase	Number of Plants	Factor	Error	
				ASEE (±%)	AEEE
Zevnik et al.	Fluid	73	0.22	65	115
Wilson	Fluid &/or Solid	79	0.85	59	115
Bridgwater (3)	Liquid &/or Solid	38	1.13	**43**	73
Bridgwater (4)	Liquid &/or Solid	38	0.48	48	**60**
Timms (1)	Gas	22	0.22	80	103
Timms (2)	Gas	22	0.20	90	122
Tolson et al.	All Phases	79	2.07	78	100

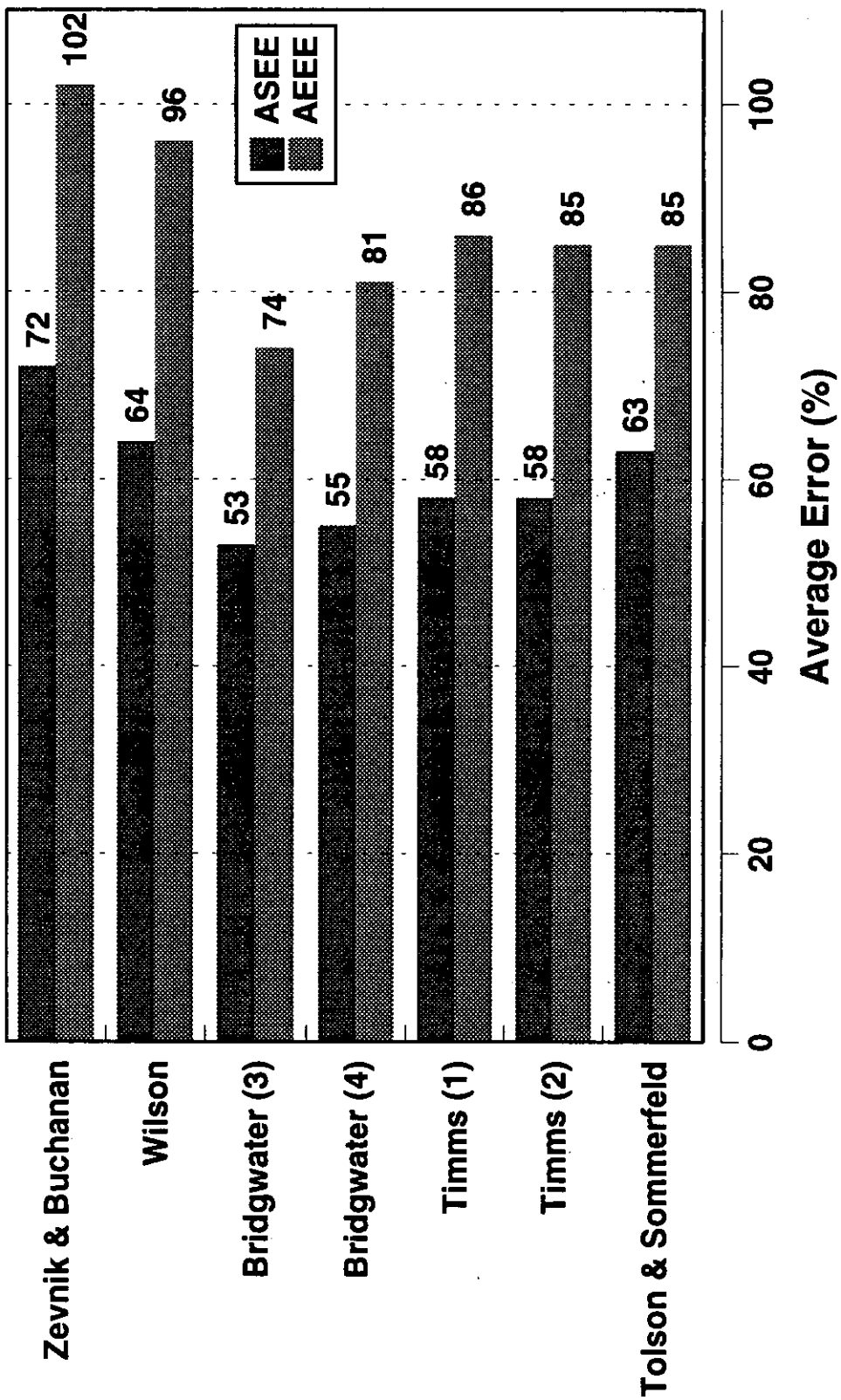
The factors used to normalise the methods varied from 0.19 to 2.07. This wide range of values for the factors indicates that the methods were derived from very different sets of data to that of the CS data, with some closer than others.

The performance of the methods is now much more comparable to that claimed. For all 79 plants an ASEE of ±53% was the best achieved, the method used was Bridgwater(3). With 84% the lowest AEEE, using Bridgwater(4), an improvement of 17% on the previous lowest equivalent error achieved when using the methods exactly as published.

Applying the method to only the plants with the phase type for which the methods were designed, again led to an improvement (ASEE of ±43%, AEEE of 60%), see table 4.5. This improvement was a significant 24% on the best AEEE when using the methods exactly as published and for the correct phase.

The conclusion is that normalising the methods did considerably improve the accuracy of the existing methods examined.

Figure 4.3 Accuracy of Methods Re-correlated with 79 CS data Plants



4.3.3.4 Re-correlated Methods

The third way of checking the methods was to use the same attributes and methodology, but to re-correlate the equations using the CS data. The accuracy of the re-correlated methods is shown in figure 4.3. The re-correlation was made using the statistical facilities available in the spreadsheet Quattro Pro.

Comparing the results for the re-correlated methods to the normalised methods showed the ASEE to be improved by less than 1% on average over all the methods. Whereas, the AEEE was on average improved by over 10%. The reason for this difference is that normalisation improves the estimate errors, but still leaves some large underestimates, which have a large error in the AEEE. The re-correlated equation is a better fit to the data and avoids large errors and therefore has a much better AEEE.

Bridgwater (4) uses one of two possible equations, depending on the capacity of the plant be estimated. This method was simplified to one equation by re-correlating and resulted in a minimal loss of accuracy.

The lowest AEEE for all 79 plants was 74% using the Bridgwater (3) method, compared with 101% for the same method used exactly, and the lowest AEEE of 84% after normalisation for the Bridgwater(4) method.

Re-correlating the methods using only the CS plants with the appropriate phase for the method resulted in an average improvement of ±2% for the ASEE, and 12% for the AEEE, see table 4.6. However, there was an improvement of only one percent for the lowest AEEE, to 59%, and an increase of ±4% was found for the lowest ASEE, when using either of the Bridgwater methods.

Methods Timms (1) and (2) were found to be better when re-correlated for all plants than for the gas phased plants for which it had been developed. Again, this is probably due to the gas phased processes being varied and difficult to represent with an equation derived by regression.

Table 4.6 Accuracy of Methods Re-correlated with CS Data for Plants with Appropriate Phase

Method	Phase	Number of Plants	Error	
			ASEE (±%)	AEEE
Zevnik & Buchanan	Fluid	73	73	104
Wilson	Fluid &/or Solid	79	64	96
Bridgwater (3)	Liquid &/or Solid	38	**47**	**59**
Bridgwater (4)	Liquid &/or Solid	38	**47**	**59**
Timms (1)	Gas	22	77	101
Timms (2)	Gas	22	75	96
Tolson and	All Phases	79	63	85

4.3.3.4.1 New Equations for Methods

The new equations derived by re-correlating are listed for each of the methods. Every equation which was re-correlated using the CS data is shown. The subsidiary equation, such as the ones for producing temperature factors F_t, were not re-correlated as the procedure for deriving them was unknown.

The following equations will provide the capital cost in US. dollars of a chemical plant constructed in West Germany in 1988:-

Zevnik & Buchanan

$$C = 9981 \cdot N \cdot Q^{0.436} \cdot 10^{(F_t + F_p + F_m)} \quad (4.5)$$

Wilson

$$C = 11025 \cdot N \cdot V^{0.445} \cdot F_m \cdot F_p \cdot F_t \quad (4.6)$$

Bridgwater (3)

$$C = 70227 \cdot N \cdot \left(\frac{Q}{s}\right)^{0.393} \cdot e^{(1.66 \times 10^{-6} \cdot Q)} \cdot T^{-0.157} \cdot p^{-0.06} \quad (4.7)$$

Bridgwater (4)

$$C = k \cdot N \cdot \left(\frac{Q}{s}\right)^{X} \quad (4.8)$$

$k = 17\,816$ $\quad \frac{Q}{s} \leq 60\,000$ tonnes per year

$k = 3\,214$ $\quad 60\,000$ tonnes per year $< \frac{Q}{s}$

$X = 0.443$ $\quad \frac{Q}{s} \leq 60\,000$ tonnes per year

$X = 0.586$ $\quad 60\,000$ tonnes per year $< \frac{Q}{s}$

The method was simplified to one equation for all throughputs with the AEEE increasing by only three quarters of a percent:-

$k = 51\,730$

$X = 0.331$

Timms (1)

$$C = 20\,967 \cdot N \cdot Q^{0.434} \quad (4.9)$$

Timms (2)

$$C = 6\,187 \cdot N \cdot Q^{0.5} \cdot F_m \cdot T_{max}^{0.057} \cdot P_{max}^{-0.06} \quad (4.10)$$

Tolson & Sommerfeld

$$C = 4.565 \cdot V^{0.429} \quad (4.11)$$

4.3.3.5 Methods with Materials Of Construction Omitted

The ***Materials Of Construction (MOC)*** attribute was used in three methods (Zevnik and Buchanan, Wilson and Timms (2)). There is an amount of uncertainty about the MOC data. The reason for this being that the CS data MOC were too

general or, in most cases, absent. Therefore, the MOC had to be found in available encyclopaedia of chemical processes which were again vague about the details. Also, the best method (Bridgwater (3)) did not use the MOC attribute. Therefore, the methods that did use the MOC were re-correlated without the MOC as an attribute and the results are presented in table 4.7. These results are for re-correlation using all the CS plants.

However, the results are on average better by ±4% for the ASEE and 3% for the AEEE, when without the MOC. Indicating that either the materials of construction gathered from the CS data is not accurate or that the MOC is an unnecessary attribute in a method for estimating the capital cost.

The results were very similar for the re-correlation using plants of the desired phase for the methods.

Table 4.7 Accuracy of Methods without MOC attribute

Method	With MOC		Without MOC	
	ASEE (±%)	AEEE	ASEE (±%)	AEEE
Zevnik et al.	72	102	67	98
Wilson	64	96	58	92
Timms (2)	58	85	**56**	**83**

4.4 Existing Methods Conclusions

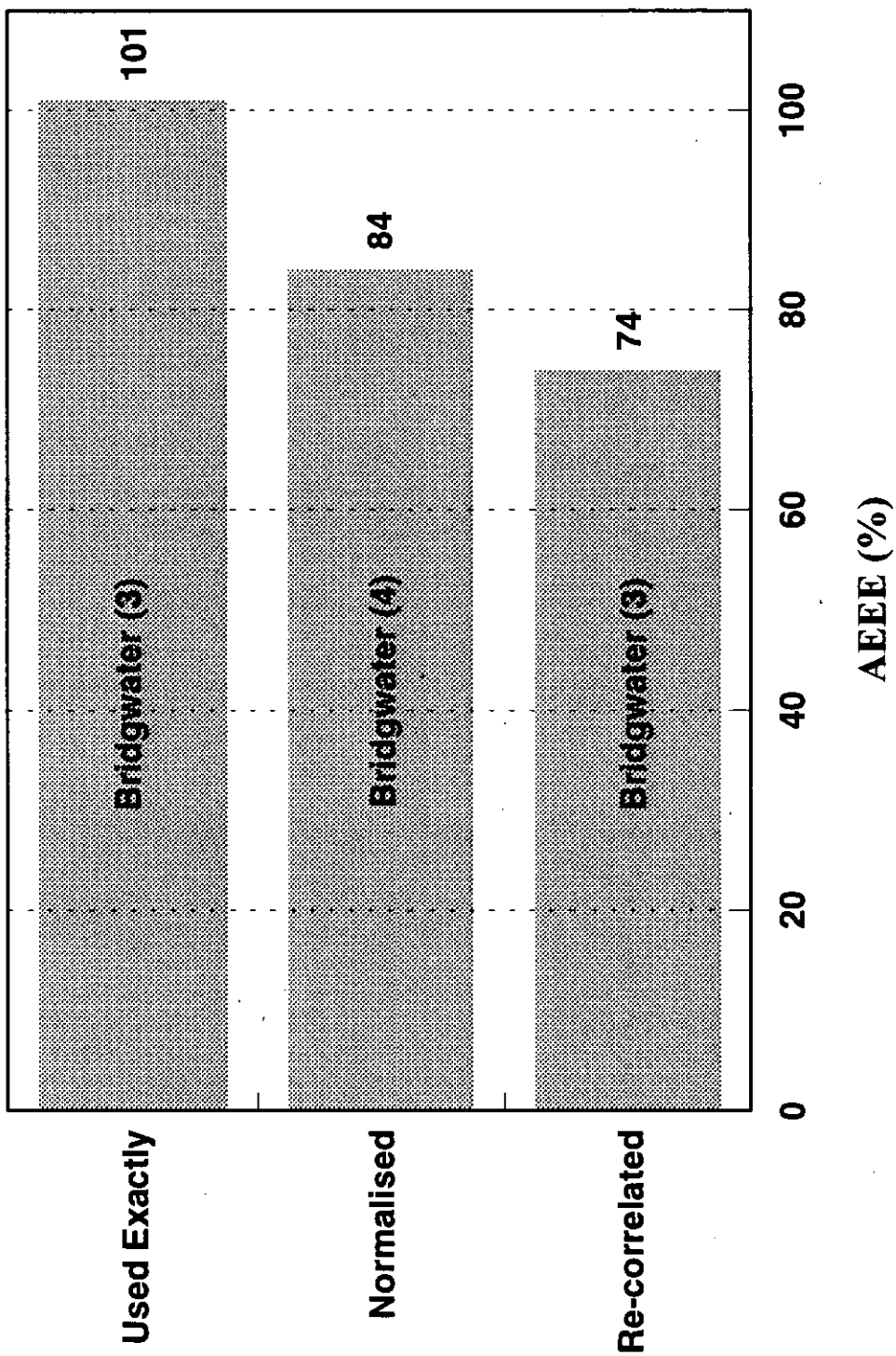
Pre-design methods have generic disadvantages. The existing methods use equations which have been derived from the data available to the researcher and the lack of published, up-to-date plant cost data is a serious problem. According to Tolson and Sommerfeld (1990) data has been published by Zimmerman (1965), Haselbarth (1967), Guthrie (1969, 1970), and Drayer (1970), but none has appeared since. An indication of how hard it is to find plant data was given by Tolson and Sommerfeld (1990) who estimated the costs for certain processes and then asked for the reader to tell them if the estimates were correct.

Most of the authors have used plant cost indices to update the cost data that they used to the year in which their method was published. The models that they derived must be further updated to the present day in order to use them now. However, the usual limit to the period for using indices to update is five years (Allen and Page, 1975); it is not possible to update costs by 20 to 30 years and be confident in the results. This is because indices do not take account of the comprehensive changes in legislation concerning environmental, health and safety standards, altering market conditions, technological advances, and productivity gains. Indices also tend to be updated about every five years and if the basis of the statistics used to derive them changes then misleading results may occur.

Even when the information available to the researcher is up to date, it will usually only be for a small number of processes, of a similar type, normally depending on the phase of the process. There is often too little data for the correlation to be statistically significant - most methods are derived from data for less than forty plants, see table 4.1.

Many of the methods require arbitrary decisions about the process, where the basis for making the decision is not clear. These decisions are often about so-called 'investment' or 'complexity' factors. Other methods cannot be used in the initial stages of design, which is the purpose of these methods, because they require design data that is not available then. For example, Stallworthy (1970) and Taylor (1977) developed methods that required a block flowsheet with a mass balance and the temperature, pressure and materials of construction for each block.

Figure 4.4 Comparison of Accuracy and Method of Best Result for each Test



Some methods have several equations, with the one used depending on the value of a particular process attribute. For example, Zevnik and Buchanan (1963) used different equations depending on the capacity, the change over point was at a capacity of 4464 tons per year. At this capacity, and with identical values for the rest of the attributes, two different values for the estimate are produced, with the higher estimate being 1.45 times that of the lower. This begs the question, which equation is best when the capacity is around the value at which the equation changes.

Another problem is that some methods use graphical methods to provide factors, for the pressure, temperature and materials of construction. This is more time consuming than putting the figures into an equation. Although Ward (1984) changed the graphical factors used by Zevnik and Buchanan (1963) into equations.

The average error of the methods were found to be a lot higher than that claimed by the authors. It seems that the results for the methods are good when tested with the data they were derived from, but when applied to new data the results achieved are disappointing.

The average errors for the existing methods when estimating the capital cost of all the CS plants were high. The best AEEE of 74% was obtained by re-correlating the third Bridgwater method with the CS data, and achieved in the same way was the lowest ASEE of ±53%. The Bridgwater methods were found to produce the best results in all three of the different tests performed on the existing methods, see figure 4.4.

A reason for the inaccuracy of the methods when estimating the capital cost of the chemical plants in the CS data is that the methods were applied to all types of plants. Whereas, most of the methods strictly apply to specific types of chemical plants, for example Zevnik and Buchanan (1963) developed their method for fluid processes and so care must be exercised to only use the methods for estimating the costs of plants with the same type of process. Estimates made for plants which are of a different type must be treated with some scepticism. Analysis of the methods accuracy when using the correct phase did produce a lower ASEE of ±47% and an AEEE of 59% when using the Bridgwater methods, but was for only 39 plants.

The results presented in this chapter leads to the following recommendations:-

1) The third Bridgwater method should be used.
2) The equation used should be the one developed by re-correlating with the CS data.
3) Do not use the materials of construction part in existing methods.
4) The error of the estimate is expected to be high.

Chapter 5

NEW REGRESSION BASED ESTIMATION METHODS

New methods for the early estimation of the capital cost of chemical plants are described in this chapter. The methods are developed from the CS data using standard regression techniques.

5.1 Development of the New Methods

Equations relating capital cost to the values of plant attributes were derived from regression analysis on the CS data, using the Minitab and Quattro Pro computer programs. Regression produces an equation that can be used for estimating the capital cost by finding the equation with the best fit for the available data. This is achieved by looking at the deviation of the lines and curves represented by an equation from the data points, and taking the equation with the least total square of the deviations.

The resulting equations were of the following form for multiple linear regression:-

$$\text{Capital cost} = a_0 + a_1x_1 + a_2x_2 + \ldots\ldots\ldots + a_nx_n \qquad (5.1)$$

and for multiple non-linear regression:-

$$\text{Capital cost} = a_0 \cdot x_1^{a_1} \cdot x_2^{a_2} \cdot \ldots\ldots\ldots \cdot x_n^{a_n} \qquad (5.2)$$

In these equations the a_0, a_1, ..., a_n are the constants or exponents found by the regression analysis. The ***Explanatory Variables (EV)*** are the variables used in the regression equation to predict the value of the ***Dependent Variable (DV)***. In this chapter the DV is the capital cost and the EV are chosen from the attributes in the CS data. The EV are represented by x_1, x_2, ..., x_n and therefore, the number of different EV used in the regression is n.

The EV used in the regression are an important consideration when developing a new method for estimating the capital cost. The accuracy and usability of a method will depend on the EV that are chosen and the point in chemical plant development at which they are known.

In previous methods the number of functional units has not been included in the non-linear regression, in that it has always had a power of one, it was included in the EV in this analysis.

Some of the previous methods use graphs to produce the values for inclusion in the method equations. The value included in the equation is found from a graph where the attribute is the abscissa. However, these were the older methods and with computerisation of methods this approach has become a lot less attractive. Regression is used to produce methods with equations that do not require the use of graphs.

5.1.1 Explanatory Variables

An analysis of the occurrence of the different attributes in the twelve functional unit methods, described in chapter 3 (table 3.3), gives an idea of which chemical plant attributes other authors have decided are suitable and important EV.

Capacity and throughput are alternatives for the representation of the amount of material passing through the plant and one or the other is used in all 12 of the functional unit methods. Making this in effect the most popular attribute, above the number of functional units.

New capital cost estimating equations were derived by using the values in the CS data for the attributes used in the existing methods. New attributes were also tried as EV in the development of the new methods. These new attributes are discussed in the next section.

Forward selection was used to derive the new regression based estimation methods in this chapter. Forward selection is a technique for the selection of the attributes to be used as EV in the equation for estimating the capital cost.

Forward selection starts with a regression using only one EV. A regression is done for every possible variable as the sole EV. A variable is then selected using the statistical information (described later in the chapter) provided by the regressions or by looking at the accuracy of the capital cost estimation of the resulting equation. Next, each of the remaining unselected EV is added to the selected EV in the regression and the results are again analysed to decide which pair of EV is chosen. The addition of EV continues until there is no significant improvement in the regression.

5.1.1.1 New Attributes

The CS data included some attributes which had never been used in previous methods for capital cost estimation of chemical plants. These new attributes were tried because they were available and untried.

The new attributes were:-

- Workforce
- Number of Reaction Steps

The workforce is a figure representing the number of operators and supervisors on a plant, with higher paid workers counting pro rata to their salary. The formula used to calculate the total workforce for mid 1988 is shown (5.3). The figures 23 300 , 28 200, and 54 700 are the annual salaries of the labourers, forepersons and supervisors. For example, every foreperson counts as 1.2 labourers due to their higher salary.

$$\text{Workforce} = \text{No. of Labourers} + \left\lceil \frac{28\,200}{23\,300} \right\rceil \times \text{No. of Forepersons} + \left\lceil \frac{54\,700}{23\,300} \right\rceil \times \text{No. of Supervisors} \qquad (5.3)$$

However, there must be some doubt about using this attribute because the stage in plant development at which the workforce would be known is not clear. If the workforce numbers are unknown in the initial project stages, then this attribute is of no use for early capital cost estimation.

The number of ***Reaction Steps (RS)*** was also considered for use as an attribute. A reaction step corresponds to a distinct chemical reaction in the route from raw material to product.

An important consideration in the calculation of the number of reaction steps in each plant are cases of two or more identical reactors in parallel. These cases were counted as one reaction step because the identical reactors are required to increase the plant capacity due to size limits on the equipment, and are not another stage in the reactions of the process. Therefore, the number of reaction steps are for some plants lower than the actual number of reactors.

The number of reaction steps were tried as a new attribute because reactors are normally the most expensive components of a chemical plant and more reaction steps indicates greater complexity. Therefore, the capital cost should be related more to the number of reaction steps than to the number of functional units. For example, two plants are constructed from the same equipment apart from one plant having two agitators and the other a reactor. The cost of including a reactor in a chemical plant would be greater than for two agitators. Therefore, the plant with the reactor will have the highest capital cost, but an estimating method using the number of functional units will produce a higher estimate of the capital cost for the plant with the agitators.

The number of reaction steps attribute was tried as an alternative to the number of functional units and also in combination.

5.1.2 Analysis of Regression

The effectiveness of the prediction of the DV by the EV for a regression equation was analysed using four statistical methods.

Firstly, the **coefficient of determination** (R^2) was calculated from the sum of the squares of the differences between the estimated values and the average values of the DV, divided by the sum of the squares of the differences between the DV values and the average values of the DV (Spiegel, 1972). The coefficient evaluates how much the variability of the DV is explained by the EV. If the coefficient of determination is close to zero then the equation is a poor representation, but if the value is close to one then the fit is good. The coefficient of determination is 'large'

and impressive when the value is greater than 0.8 (Coates, 1993). However, this can be achieved by simply having a lot of EV. This trap is avoided by using an adjusted coefficient of determination which takes into consideration the number of EV.

Secondly, the **f-test** is used to see whether the DV is dependent on the EV. This can not be assumed from the coefficient of determination which indicates that the regression equation is a good model of the variation. The f-test involves finding the f value and checking that it is in the range expected when there is a dependency between the EV and DV.

Thirdly, a **plot of the residuals** is made by plotting the residual values against the values of an EV. A plot of the residuals is required for each of the EV used in the regression. The residual is the difference between the capital cost as estimated by the equation and the actual value. A pattern or structure shows that a relationship between the EV and the DV is not being used to full effect, no pattern or structure means a good model.

Finally, the importance of each particular EV in the regression equation is provided by the **t-ratio**. A table of the t distribution provides the range of t-ratio values that an EV will have when not needed in a regression. The t-ratio is calculated for each EV and if it is within the range then that EV is discarded from the equation. This method is used especially when deciding during forward selection whether an EV has a significant influence or should be omitted.

5.2 New Regression Equations

The results of the regressions using the EV described earlier are presented in this section. The regression equations were derived using the same 79 processes as in chapter 4. Firstly, the type of regression to use was selected, the choice being between linear or non-linear.

5.2.1 Linear Regression versus Non-linear Regression

Linear regression was found to produce equations whose representation of the DV was not as good as those produced with a non-linear regression using the same EV. The reason for this is that linear regression tries to represent the relationship between the DV and the EV with a straight line and in most situations this is not the best possible model of the relationship. However, non-linear regression uses the natural logarithm of the dependant variable and the EV. This results in the relationship being modelled by a curve, which in most cases is a better representation.

To examine whether linear or non-linear regression equations are the most suitable for capital cost estimation required all of the attributes in the CS data to be regressed in turn, both linearly and non-linearly, as the sole EV for the capital cost DV. The results for linear regression are shown in table 5.1 and non-linear regression in table 5.2.

Table 5.1 Linear Relationship of Capital Cost to Individual Attributes

EV	Code	R^2 (%)	ASEE (±%)	AEEE (%)
Workforce	W	27	82.7	94.0
Capacity	Q	41	**79.3**	**91.3**
Number of Functional Units	N	3	110.7	130.1
Materials of Construction	F_m	2	114.7	135.2
Maximum Temperature	T_{max}	4	116.4	135.2
Maximum Pressure	P_{max}	0	119.4	139.5
Number of Reaction Steps	R	0	122.7	142.6
Throughput	TP	0	121.6	142.0
Throughput (Process Water)	TP_{PW}	0	120.8	141.1

Table 5.2 Non-Linear Relationship of Capital Cost to Individual Attributes

EV	Units	Code	R^2 (%)	ASEE (±%)	AEEE (%)
Workforce	-	W	38	**59.4**	**83.7**
Capacity	tonnes/year	Q	36	62.8	85.2
No. of FU	-	N	14	73.3	108.7
MOC	-	F_m	3	75.5	121.1
Max. Temp.	K	T_{max}	3	76.3	120.4
Max. Press.	Bars	P_{max}	1	82.5	122.0
No. of RS	-	R	0	78.5	124.4
Throughput		TP	0	81.4	124.6
Throughput (PW)		TP_{PW}	0	81.3	124.5

Comparing the average accuracy of the regression equations presented in tables 5.1 and 5.2 reveals that the capacity results in the regression equation with the most accurate estimates when using a single EV. For the linear regression the ASEE was ±79% and the AEEE was 91%. Non-linear regression resulted in an ASEE of ±59% and an AEEE of 84% for the workforce.

The comparison between the accuracy of the estimates for each attribute's equations derived by linear regression and non-linear regression revealed that non-linear regression always produced better equations. This is further illustrated by Figure 5.1.

These results justified no further linear regression analysis, and the development of the new methods by concentrating on the use of non-linear regression.

Figure 5.1 Comparison between Linear and Non-linear Regression

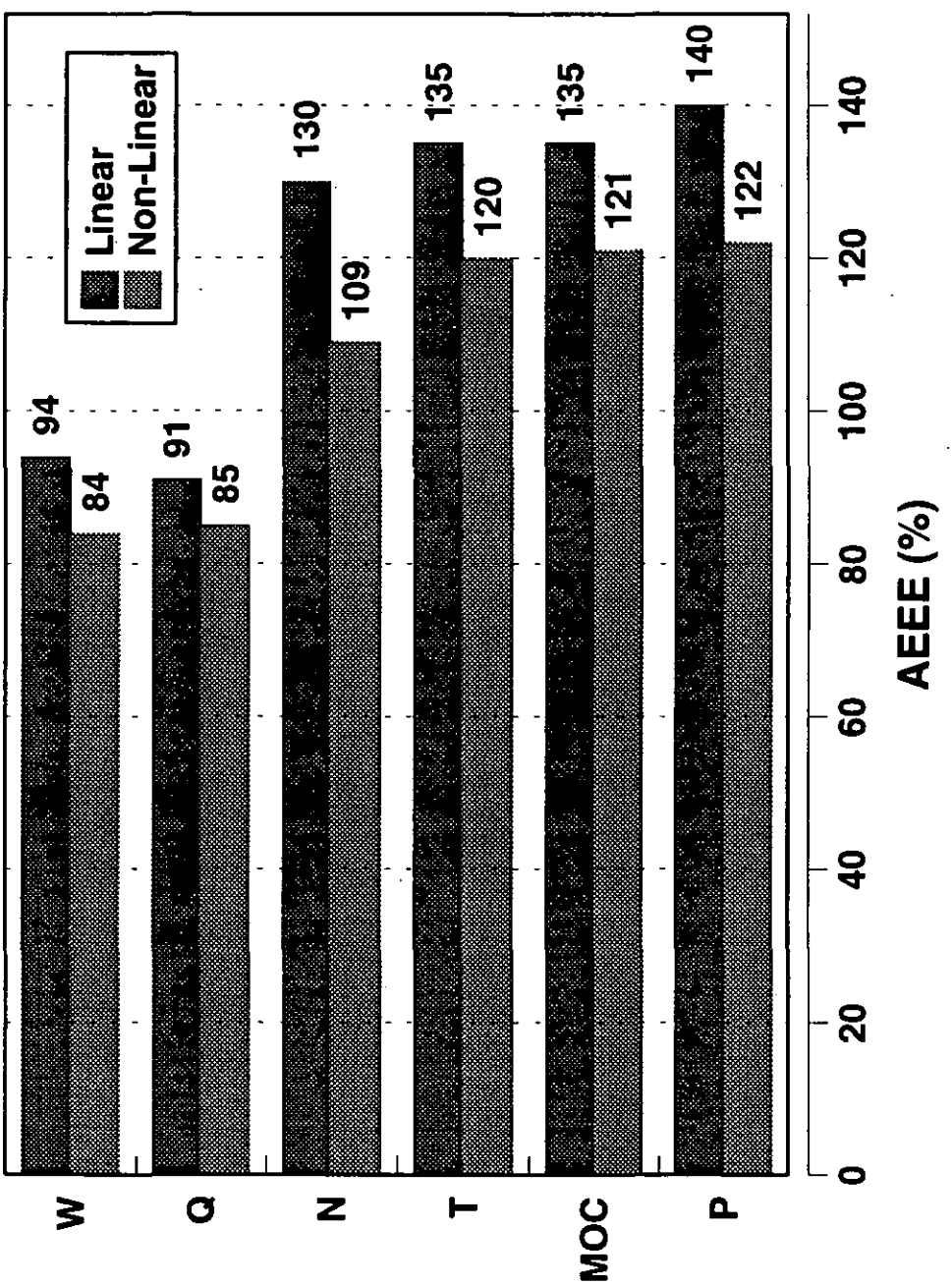


Table 5.3 New Regression Accuracy for CS Data Attributes including Workforce

EV	No. of Attributes	ASEE (±%)	AEEE (%)
Q, W	2	36.9	50.0
Q, W, N	3	**35.5**	46.8
Q, W, N, F_m, R	5	35.6	**46.1**

5.2.2 New Method using Non-linear Regression

The next stage in the development of the new methods was to use more EV in the regression. Forward selection started with capacity and then other attributes were added one at a time, with the effects analysed in a manner identical to the earlier analysis. All the attributes were tried, but only the cases where the addition of another EV resulted in an improvement are shown in table 5.3.

The effect of using T_{max} and P_{max} in the regression analysis was to slightly increase the error. The equation for the most accurate regression for the ASEE (±36%) used capacity, workforce, and number of functional units, see equation 5.4.

$$\text{Capital Cost} = 14765 \cdot Q^{0.4} \cdot W^{0.807} \cdot N^{0.254} \quad (5.4)$$

The most accurate equation for the AEEE (46%) used the three attributes in equation 5.4, plus the MOC factor and the number of reaction steps, see equation 5.5.

$$\text{Capital Cost} = 20952 \cdot Q^{0.38} \cdot W^{0.757} \cdot N^{0.26} \cdot F_m^{1.29} \cdot R^{(-1.45)} \quad (5.5)$$

The exponent value for the reaction steps in equation 5.5 means that the capital cost decreases as the number of RS increases. This is surprising and difficult to explain.

An equation was also developed that only used the attributes occurring in existing methods. This was to make sure that an estimator using one of the existing methods could use one of these new methods. For example, the workforce would

Table 5.4 New Regression Accuracy for CS Data Attributes without Workforce

EV	No. of Attributes	ASEE (%)	AEEE (%)
Q, N	2	48.9	69.1
Q, N, P_{max}	3	48.7	68.5
Q, N, P_{max}, T_{max}	4	48.6	68.5
Q, N, P_{max}, T_{max}, F_m	5	**46.4**	**65.8**

probably not be available to an estimator, especially when it was not required for the method that they currently used. The results of the analysis when not including workforce and number of reaction steps are shown in table 5.4.

Without the workforce a considerably higher error was found, approximately 20% for the equivalent accuracy. The reason for the effectiveness of workforce could be due the values being calculated from the capital costs, but the CS data is from actual plants and therefore the workforce should be known. Also, statistical analysis showed that the workforce was not as strongly related to the capital cost as would be expected for a value calculated from the capital cost. Another possibility is that higher value plants have a larger workforce. The reasons for this could be:

1) An expensive plant is complicated and requires a large workforce when operating.

2) The stream-time of the plant is increased by taking more preventative maintenance.

3) Provision against costly failures.

The main problem with using the workforce in capital cost estimating techniques for the early stages of design is the likelihood that it would be unknown, and for this case the lack of a method for estimating the workforce values.

The equation for the most accurate new regression method (equation 5.6) that does not use either of the workforce or number of reaction steps is shown below:

$$\text{Capital Cost} = 55882 \cdot Q^{0.44} \cdot N^{0.486} \cdot T_{max}^{0.038} \cdot P_{max}^{-0.02} \cdot F_m^{0.341} \quad (5.6)$$

5.3 New Methods Conclusions

Two new regression based methods are proposed. One incorporates the attributes previously used in the existing methods (equation 5.6) and the other uses new attributes (equation 5.4 and 5.5). This precaution is taken in case anyone wishing to use these methods does not have the data needed for the new attributes.

Equations derived by non-linear regression gave better estimates, as expected, which meant that the new equations are of the same form as existing methods. However, the proposed methods in this chapter differ from existing methods in the following three ways:-

- The equations are derived from 79 plants, which is more than has been used in previous methods.
- The number of functional units does not have to use an exponent value of one.
- New attributes are used in the regression. For example, reaction steps and workforce.

The best accuracy attained by the new estimating equations was ±36% for the ASEE and 46% for the AEEE when including the new attributes, and ±46% for the ASEE and 66% for the AEEE when only using attributes that have previously appeared in published methods. The accuracy obtained with the inclusion of the new attributes was on average 23% better than the best accuracy achieved when using the existing methods in chapter 4, and the average improvement was 9% when only using attributes found in existing methods.

Chapter 6

CAPITAL COST ESTIMATION USING COMPUTERS

This chapter assesses the effect that computers have had on capital cost estimating and then looks at how the latest computer techniques could be used in the future.

6.1 Current Use of Computers

Capital cost estimating packages for computers have been under development since the late sixties and are now commonplace in industry. Computerisation of capital cost estimation speeds up the estimating process and improves the accuracy of the calculations for the estimator. The ready access to a database of existing plant costs, preferably the company's own records, is important because knowledge about previously constructed plants can greatly improve the accuracy of early estimates.

ICI produced the first major commercially available program, FACTEST. The program applied a series of factors to each main plant item, accounting for the level of process and construction sophistication. Different factors were used for erection, piping, instrumentation, electrical, civils and lagging. The factors were found from historical data for about 100 projects. The leading commercially available packages are now produced by Icarus; QUESTIMATE, which was the market leader in 1991, costs items and avoids the factorial approach for bulk items (Gerrard, 1991). However, there are a lot of other commercially available products which mostly use the factorial method.

However, the usefulness of the available computerised methods is questioned by Kharbanda and Stallworthy (1988), because the currently commercially available programs do not try to emulate the 'feel' and judgements made by a skilled estimator. There is no great improvement in the accuracy of the estimates over the

manual techniques (Baldwin and Oteifa, 1991). Skilled estimators, operating in areas where they have experience and easily available data, can achieve the accuracy required in the early stages of plant development with little effort. An estimator tends to evaluate the importance of the figures for each attribute subconsciously, relying on a 'feel' for the situation under consideration (Pearce, 1989). Hence, capital cost estimating is considered more of an 'art' than a science. For example, the estimator would use the information about the costs for existing plants that have similar values for the key attributes to infer the capital cost of the proposed plant. Some of these key attributes are shown in table 6.1.

Imitating the way a skilled estimator approaches estimating requires the inclusion of their expertise in the method. This is difficult for a computer to emulate and leads to an expectation of computers being unsuccessful in improving the accuracy of estimates. Liddle and Gerrard (1975), and Kharbanda and Stallworthy (1988) state that a computer may be able to achieve this with the development of modern ***Artificial Intelligence (AI)*** techniques. The AI possibilities are looked at in the remainder of this chapter.

Table 6.1 Key Attributes

Capacity	Operating conditions
Materials of construction	Safety considerations
Materials processed	Control and operation
Unit operations	Environmental impact considerations

6.2 Estimating and Artificial Intelligence

The name AI was first suggested by John McCarthy during the late fifties. There is no clear cut definition of AI, but it is fundamentally a field of science aimed at conferring intelligence on machines (computers) - a concept introduced by Turing (1947). In this context intelligence is best summarised by a list of its characteristic features (Parmar, 1986): judgement, comprehension, reasoning, concept formation, response selection, adaptation, creativity.

AI techniques aim to represent knowledge in some form and then use it intelligently to solve problems. The knowledge can be acquired from human experts and then consists of heuristics, which are rules developed as a result of experience. An AI system which uses heuristics provided by experts to solve problems is called an expert system. Other AI systems solve problems by using knowledge that is represented by standard formats for data. Examples of these AI techniques are neural networks and case-based reasoning. These two ways of providing solutions can be combined. The initial solution is determined from the knowledge in the standard data and this is then modified with the aid of heuristics due to experts. The type of knowledge needed (data, heuristics or a combination) and the mechanism that uses this knowledge to solve the problem depends on the AI technique.

There is a lot of current research into the various methods covered by the field of AI. A survey of the most relevant methods and their suitability for capital cost estimation follows. Each AI method is described separately. However, when developing an AI system the best results are nearly always achieved by constructing a system that combines a number of techniques. For example, systems developed in this thesis combined different methods. Therefore, an understanding of the methods available is important.

6.2.1 Expert Systems

Expert Systems (ES) attempt to computerise experts' experience and how it is used. They are normally used to offer advice or make a diagnosis about a problem.

The first ES were DENDRAL (Feigenbaum, 1971) and MYCIN (Shortliffe and Buchanan, 1975), which diagnosed medical problems and prescribed therapy in a

very specific problem area. For example, MYCIN was for bacterial infections. They use hundreds of rules gathered from experts such as:-

IF	(condition,	infection is primary-baceremia) AND
	(condition	site of culture is sterile)
THEN	(diagnosis,	the identity of the organism is bacteriods)

Each rule is a semi-independent packet of knowledge. These rules once developed can be used to solve new problems. This requires an inference engine. There are different possibilities for the mechanism of the engine. One example is forward chaining, which searches through the rules in the knowledge base, in order to find which ones have conditions like those in the new problem under consideration and from these produces a diagnosis.

ES are used for classification and diagnosis. The fact that they use rules provided by human experts means that they are often useful in areas where experts are scarce, such as cost estimating. However, for capital cost estimation the required final output is a number, rather than the diagnosis which an ES usually generates.

ES have been used in costing for providing advice. Noble and Tanchoco (1990) developed a prototype ES that shows the potential economic implications of different design alternatives. ES have helped to reduce the amount of data required by a computer package to produce a cost estimate, since an expert system can infer some of the necessary data. This results in a time saving rather than improving the accuracy, and enables estimates to be made earlier. ES have been used in this fashion for estimating construction costs of buildings, ELSIE (Ashworth, 1988) and Oil and Gas installations, TOPEX (Greffioz et al., 1993).

ES are complicated to set up and require the co-operation of experts to produce the rules for capital cost estimating. This would require help from expert cost estimators, who are in short supply. Even if the experts were available to help, it would be difficult to explain their 'feel' for decision making and each expert would have their own methods.

6.2.2 Artificial Neural Networks

Biological nervous systems are believed to consist of neurons interconnected in a network. The neurons are believed to have no central control and are activated by chemicals released between neurons. The activation of the neurons enables the nervous system to work in the way that it does. This set up of neurons is called a neural network.

Artificial Neural Networks (ANN) are attempts by computer scientists to copy the working of the nervous system, by creating a simulation of the structure of the brain on a computer. However, because no one fully understands how the neurons work and due to the incredible complexity in their elementary structure an ANN is a simplistic representation of the nervous system. McCulloch and Pitts (1943) started neuro-computing with their paper about a neural network that simulated the visual system of a frog in 1943. Rosenblatt (1958) stimulated the interest in ANN with the development of the PERCEPTRON system for recognising patterns. However, the interest in this field dwindled after Minsky and Papert (1969) showed that the PERCEPTRON could only recognise a limited classes of patterns. Hopfield (1982) inspired the current research into ANN by showing them to have emergent features when PERCEPTRONs were combined together. That is features that were not built into the system, but arose out of the basic configuration of an ANN. These were associative memory and error correction, indicating that perhaps ANN were approaching the working of the human brain.

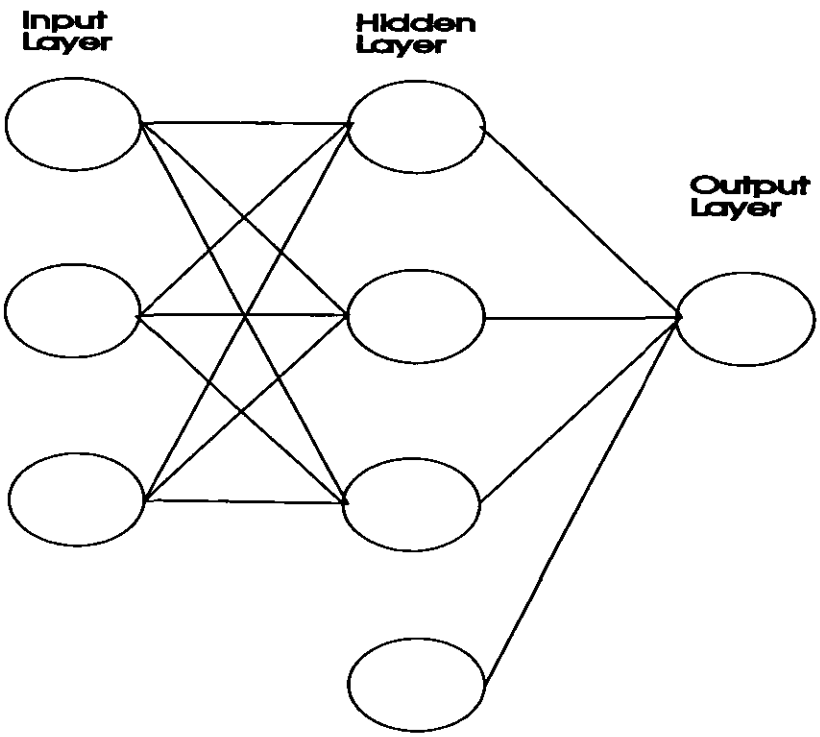
A basic description of ANN is given by the neural network news group (1995):

> An ANN is a network of many very simple processors ("units"), each possibly having a (small amount of) local memory. The units are connected by unidirectional communication channels ("connections"), which carry numeric (as opposed to symbolic) data. The units operate only on their local data and on the inputs they receive via the connections

In most applications the neural network is given a set of input patterns from which an output is produced, these are called the input and output layers. Hidden layers of nodes are used between the input and output layers, all of which are connected

in a network, with the interconnections weighted. An example of such an ANN is shown in figure 6.1.

Figure 6.1 Three Layer Artificial Neural Network



The actual and desired outputs are compared, and a more accurate set of weights for the interconnections is learned using iterative mathematical processes, such as back-propagation, driven by the error resulting from the current set-up. This training continues until a minimum error is achieved for the set of input patterns.

The main disadvantages of the learning in ANN is when the error-surface contains local minima and the training finishes at one of these points, missing the global minimum.

The use of ANN is advantageous when there is a lot of example data, and rules are unavailable or hard to successfully apply to the problem. Neural networks have the ability to teach themselves through training on the input data and do not require complicated mathematical models. This makes neural networks suitable for cost estimation when a set of historical data is available for training, thereby removing the need for regression equations. For example, in the estimation of the capital

cost of a chemical plant the input pattern values would be for the capacity, number of units, operating temperature, and pressure and the output would be the capital cost (Gerrard, 1994).

Even though ANN produces good estimates, which is the major concern of the cost engineer, there are reservations about using an ANN to produce a successful estimating system. One of these is that the explanation for how the ANN has produced its estimate is not clear. For example, in figure 6.1 the top input node in the input layer has three weights on its interconnections. Comparing these weights with those for the second input would not give any indication of which input has the biggest influence on the output value, due to the complicated nature of the combined effect of the weights. A cost estimator would have more confidence in a method with an explicit methodology and, even better, exhibiting properties that they would expect. For example, the capacity input influencing the capital cost output more then the operating temperature.

6.2.3 Case-Based Reasoning

The first ***Case-Based Reasoning (CBR)*** system was developed by Kolodner (1983a, b) and was called CYRUS. CBR was applied to planning the travel and meetings of ex-US secretary of state Cyrus Vance. Since then many systems have been developed to solve various problems.

The generally accepted definition of CBR is:

> A case-based reasoner solves new problems by adapting solutions that were used to solve old problems (Riesbeck and Schank, 1989).

CBR is a cyclical process comprising of the four 'REs' (Aamodt and Plaza, 1994):

1 RETRIEVE the most similar case(s)
2 REUSE the case(s) to attempt to solve the problem
3 REVISE the proposed solution if necessary
4 RETAIN the new solution as a part of a new case

The first stage relies on the development of a standard case representation, with a case normally comprising of the problem, the solution and the result of the

solution. Also, there needs to be an effective matching algorithm for finding the most similar case to the new problem. The best matching case that is retrieved is used as a solution. With a large number of cases, a high similarity should exist between the retrieved and new cases. However, when there is a prominent difference between the values for the new problem and the retrieved case, then an adaptation is made to the solution provided by the retrieved case. The adaptation of the solution is accomplished by applying rules or formulae made available by the system. The new solution is retained in the case database.

Instead of trying to come up with rules, formulae or the weights in an ANN to calculate a value for an unknown variable from input data, CBR finds the existing case that is most similar to the new problem and uses this as the basis for producing the estimate. For example, the combination and arrangement of plane parts being heated in a large autoclave changes the heating characteristics of the autoclave and may lead to parts being ruined. The effect of different loads of parts on the autoclave are too complicated to be modelled. However, using CBR to select a set up of parts which was similar to previously successful loads led to a higher success rate for loads (Hennessy and Hinkle, 1992).

An advantage of CBR is that experts' knowledge is not required. CBR systems are very simple to develop and bypass the need to derive equations to model the problem; this is especially useful for very complicated problems which are not susceptible to mathematical modelling.

There are no published examples of CBR applied to capital cost estimating, even though the advantages of using CBR for capital cost estimating seem to outweigh the disadvantages. There is no need for the elicitation of information from experts, as problems can be solved without a full understanding. There is no need to make a decision about which of the many formulae for early capital cost estimation should be used. For the best results, data for numerous plants is required and the effectiveness of the CBR improves with an increase in the number of past cases in the database. This means CBR is dependent on the number of cases available. In capital cost estimating the data is scarce and the number of new plants built each year is small and so the database of cases can not be built up over a short period of time. Meanwhile, the existing data is gradually getting out of date.

CBR is similar to the fuzzy matching method described in chapter 7.

6.2.4 Fuzzy Set Theory

Fuzzy Set Theory (FST) was first proposed by Zadeh (1965), and was initially an adaptation of standard crisp set theory. Fuzzy sets allowed a value to have partial membership of a set. This allowed the treatment of imprecisely defined knowledge in an exact mathematical way. FST is described in more detail in the introduction to fuzzy matching in chapter 7.

Initially, the mathematical properties of FST were investigated, and eventually applications of this new area of AI were tried. FST was developed into fuzzy logic which interprets statements using fuzzy sets. For example, what is the truth of a statement such as Bob is tall and old, given Bob has a height of 7ft and is 37 years old? Fuzzy sets have to be developed to assign the membership of 7ft in the set tall and 37 years old in the set old. The two values then have to be combined to give a truth value to the statement about Bob.

Fuzzy logic is used with fuzzy rules in fuzzy expert systems. Fuzzy rules are of the form:

IF (condition, fuel input is high)
THEN (diagnosis, engine speed is much greater than 100rpm)

There are three parts to a fuzzy expert system:

1 Fuzzification
2 Inference
3 Defuzzification

The first part takes the input values, such as Bob is 7ft, and then assigns a fuzzy value by using a fuzzy set for height. For example, the fuzzy value for Bob being tall might be 0.99 and would be calculated by using a fuzzy subset of height that gives a fuzzy value for tallness. Then the inference engine evaluates the combined effect of the fuzzy valued inputs and the fuzzy rules, using fuzzy logic, and produces a fuzzy-valued output. The third part of the system then uses a defuzzification method to produce useful output, for example a fuzzy expert system controlling the speed of an engine would produce an output such as 82rpm.

Another area of FST is fuzzy matching, and is a method for assessing the degree of similarity between two entities given a set of facts about each of them. Fuzzy matching is discussed in detail in chapter 7.

Applications of fuzzy expert systems are found in control and pattern recognition. However, the main commercial application of fuzzy expert systems is in fuzzy control for systems which are complicated and difficult to model, such as cooling systems (Hakata and Masuda, 1990). Fuzzy controllers are now used in everyday household items such as washing machines, air conditioning, video camcorders, toasters, and many others. FST has also been used as a tool for handling uncertainty in AI systems (Henkin and Harrison, 1988). FST has been used in chemical engineering for simulation (Dohnal and Hartmann, 1981), reliability (Babinec and Dohnal, 1981) and commercial applications are available for process control (Leigh and Wetton, 1983).

Applications of FST in cost engineering include using fuzzy mathematics to allow for uncertainty in values used for the calculation of cash flow (Buckley, 1987) and investment (Lin et al., 1990). FST has also been applied to capital cost estimating by Turunen (1984, Turunen et al., 1984). This involved developing the standard factorial and functional unit methods using fuzzy expert systems. The process parameter values normally inserted into the formulae were allowed to be fuzzy values. Therefore, 'linguistics' could be used to describe the process parameters when there was uncertainty about exact values in the early stages of design. For example, a cost estimator might look at the design and decide that the operating temperature will be low for the process. Fuzzy rules developed from experts were used in the inference to produce an output. The fuzzy expert system then defuzzified the output to the estimate for the capital cost. No conclusion was made on the accuracy of the method for capital cost estimation due to insufficient testing of the method.

The advantages of using FST for capital cost estimating are its ability to handle uncertainty in the process parameters during the early stages of a process design and for the representation of the complicated and ill-defined relationships between process parameters and capital cost. However, the knowledge of experts was vital in preparing the fuzzy rules and, as for the traditional methods a lack of cost data hindered development.

6.2.5 Genetic Algorithms

The initial idea for ***Genetic Algorithms (GA)*** is credited to Holland (1962) and the actual term was given to an application for game playing developed by Bagley (1967). GA are a method for computers to develop a solution to a problem using nature's principle of survival of the fittest.

A GA starts with a population of randomly created sets of values, or 'chromosomes', which could be used as a possible solution to a problem. Each chromosome is tested by a fitness function which indicates its effectiveness as a solution to the problem. A second population of chromosomes, 'the next generation', is developed from this first population. The chromosomes from the first generation are selected at random, but with the probability of a chromosome being selected being related to the quality of its solution as judged by the fitness function. The selected chromosomes are then paired randomly and a certain percentage recombined, using various techniques for the crossing over of their values, to create two new chromosomes. There is also a chance of a randomly selected value being inserted in to the chromosome, this is called a mutation. Also, a few of the best chromosomes from the first generation are kept unchanged in the next generation. The new population is tested and then in an identical fashion another generation of chromosomes is produced. This continues until a specified number of generations has been tried or the chromosomes have converged on a solution to the problem.

GA have found applications in solving problems which are difficult to solve with traditional methods, such as scheduling, sequencing, machine learning and optimisation. They are particularly good for finding global optima in a very hilly space (Corne et al., 1993).

GA have not been applied in the cost estimating field, but could have potential in finding the optimum values for the constants and powers of the equations used by the functional unit methods.

6.3 Conclusion

A review of AI techniques and their application to estimating is provided in this chapter. Some methods have been used previously for cost estimating, but others appear to be untried as estimating methods.

A decision was made to develop fuzzy matching for cost estimating for reasons explained in the next chapter. Genetic algorithms were also used during the research in an attempt to improve the time taken to develop the optimum set-up for the fuzzy matcher.

It is also worth pointing out that there is a similarity between fuzzy matching and case based reasoning. Both are based on finding the existing case that is most like the new problem. In fact, Fuzzy matching could be used for the retrieval part of CBR.

Chapter 7

FUZZY MATCHING

The fuzzy matching methodology and its suitability for capital cost estimation are explained in this chapter.

7.1 Methodology and Suitability

An estimate for the cost of a chemical plant can be made by using the principle that similar plants have similar costs. Ideally an existing plant would be found that has the same capacity and process, and therefore identical pieces of equipment, as the chemical plant under development for which a capital cost estimate is required. The cost of the existing plant would be multiplied by the relevant cost index and location factors to produce a capital cost estimate for the new plant. However, in the majority of cases there will not be such an existing plant and so the estimator must determine which of the existing plants is the closest 'match' to the new plant. Then the cost of the available existing plant is used as a basis for the estimate for the new plant.

An estimator assesses the similarity of plants by considering the closeness of the values for the attributes of the new and existing plant and the relative importance of each attribute to the capital cost. The attributes used must influence the capital cost of a plant, such as design specifications, operational data, capacity, equipment, materials of construction, etc.

The experience of the estimator, aids in the judgement of which plants are similar, and in the adjustment made to the capital cost value of the existing plant to allow for the different characteristics of the new plant. Fuzzy matching provides a way of computerising the method used by an estimator to decide which chemical plants are similar.

Fuzzy matching is a development of fuzzy set theory, and is a method for assessing the degree of similarity between two entities given a set of facts about each of them. Fuzzy matching defines a continuous scale of similarity between the two extremes of an exact match and no match at all.

For example, Chung and Inder (1992) used fuzzy matching to assess how similar the geological formations in an area under exploration are to those formations in an area where petroleum had already been discovered. Further, more costly exploration would only be sanctioned in cases where there was a sufficient similarity.

The ability of fuzzy matching to assess the degree of similarity between two entities gives it an advantage over conventional database operations, which can only find an exact match between entities. An exact match is unlikely over a number of attributes with known values, especially for entities with a wide range of possible values for each of their attributes, such as chemical plants. Very few plants would have identical values for each of the different attributes. For example, only one case of two plants having exactly the same values for five attributes was found amongst 90 plants. An exact match might be found on one or even two of the attributes, but not on all of them.

Also, to manually work through a lot of plants and find the best match for a new plant would be slow and the quality of the decision would depend on the skill of the expert. Fuzzy matching allows the computerisation of this decision making procedure.

Fuzzy matching has not been used for estimating before, but the arguments for using this technique as a new method for estimating are strong.

7.1.1 Fuzzy Matching Example

To illustrate fuzzy matching an example from every day life follows. The data used is hypothetical. Someone might decide to sell their house, but for what price should it be sold? In order for them to decide the sale price of their house they would have to look at the prices of other houses on the market. Finding a house of the same size in terms of the number of rooms and garden area should give a good idea of the value of the house. However, when a house is put on the market and it

attracts no interest or gets five offers in a week, then it becomes apparent that the wrong price must have been assigned. This implies that some of the details available for a house that are influential to the cost have been omitted. A likely candidate is location, which, for fuzzy matching purposes, is represented by the insurance band. Then, with this attribute having a numeric value, like the other attributes, allows an assessment of the 'closeness' between values.

The data available consisted of three attributes: the number of rooms, garden size, and the insurance band. This information and the price was known for four sold houses. This and the details of the house for sale are shown in table 7.1.

Fuzzy matching does not require the person trying to sell the house to know anything about house prices. The method determines which already sold house is most like the one for sale. A membership function is used to assign a value which measures how similar the match is between two values for an attribute. The similarity value varies between zero for no similarity and one for a perfect match; the details of how this works are explained later. Table 7.2. shows that the match value for each attribute and the total of these values for the fuzzy matching between the four sold houses and the house for sale.

Table 7.1 House Data

House	No. of Rooms	Garden Size (m^2)	Insurance Band	Cost (£)
For Sale	10	100	1	?
Sold 1	5	100	4	**30 000**
Sold 2	8	400	1	**65 000**
Sold 3	6	200	2	**45 000**
Sold 4	9	0	2	**60 000**

Table 7.2 Results of Fuzzy Matching

House	No. of Rooms	Garden Size	Insurance Band	Match Value
	Match Value			Total
Sold 1	0.83	1.00	0.00	**1.83**
Sold 2	0.93	0.00	1.00	**1.93**
Sold 3	0.86	0.66	0.66	**2.20**
Sold 4	0.96	0.66	0.66	**2.30**

The fourth sold house has the highest match value total and so its sale price gives an initial idea for the price of the house for sale, that is 60 000 pounds. The match value total of 2.3 is quite close to the maximum of three and so the sold house is similar to the one for sale.

With no knowledge about what affects the price of a house all three attributes will be viewed as having the same effect. However, it is likely that some attributes will have a far larger effect on the price of the house than others. For example, maybe the number of rooms would be a lot more important than the size of the garden. Weighting the match values for the attributes according to their influence on the price is one way to allow for the relative importance of each attribute.

7.2 Application of Fuzzy Matching to Capital Cost Estimation

In the following sections the fuzzy matching method is described in detail and the important development issues are discussed. Attention is paid to the application of fuzzy matching to capital cost estimation, with the specific results of this application shown in the next chapter, chapter 8.

Fuzzy matching attempts to imitate the techniques used by a skilled estimator when estimating the capital cost of a chemical plant. Fuzzy matching for capital cost estimation finds the existing plant that is the closest match to the new plant and then uses the cost of this closest match as the estimate.

7.2.1 Membership Functions

The essential concept of fuzzy matching is the quantifying of closeness or goodness of fit between two attribute values. This value is called the ***Match Value (MV)*** and is calculated by the ***Membership Function (MF)***, which is sometimes also known as the characteristic function.

In classical set theory an element is either a member of a set or it is not a member, that is, it is a non-member. However, this representation can not be used in many situations. The classic example is the concept of tallness, which is a linguistic variable. If someone is 7 ft in height they are tall, while if their height is 5 ft they are not tall. However, what about the range of heights in between? A representation is needed for a value which is partly in a set, this is the basis of fuzzy set theory.

Fuzzy set theory introduces degrees of membership. A member of a set is given a value 1 and a non-member 0 and values in between represent an intermediate degree of membership.

In fuzzy matching the theory is used in a slightly different form. The problem is not vagueness in human terminology, for example tallness, but the uncertainty of the relationship between two values for the same attribute. For example, is the capital cost of an existing plant with a capacity of 50 000 tonnes per year worthy of consideration as the basis for a capital cost estimate for a new plant of capacity 70 000 tonnes per year.

The goodness of fit for an attribute value of the new plant, the ***target***, to values of the same attribute for known plants, the ***data***, must be quantified. This requires a membership function for rating the goodness of fit between data and target. The match value is on a scale of between one and zero, with one indicating a perfect match and zero showing that the values for the attribute are so far apart that no relation can be inferred between capital costs of the two plants on the basis of this attribute.

The membership function, MF(d, t, b), is defined by a set of equations. When comparing numeric attributes, d is the data value, t is the target value, and b is the ***Shape Parameter (SP)***. The shape parameter defines the range of data values which have a MF value higher than zero for a given target value.

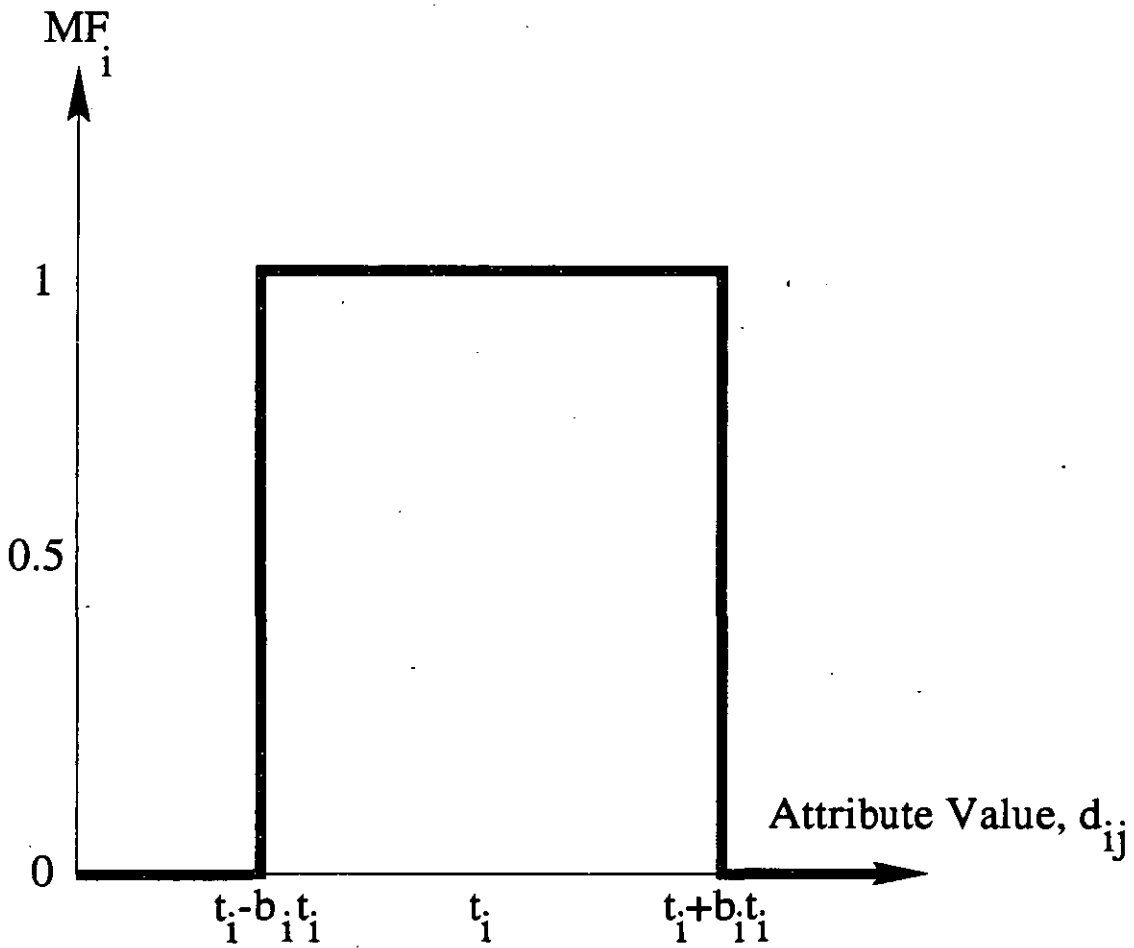
Let there be p attributes, identified by a subscript i and let there be a match value, MF_i, for each attribute i. Furthermore, let the database contain data for q existing plants, identified by a subscript j. Typical MF are piece-wise, continuous functions, $MF_i(d_{ij},t_i,b_i)$, with values on the interval [0,1]. $MF_i(d_{ij},t_i,b_i)$ measures how close the value of attribute i for existing plant j, that is d_{ij}, is to the value of attribute i for the new plant, that is t_i. The shape parameter, b_i, has a subscript, because in general the value is different for each attribute.

7.2.1.1 Possible Membership Functions

There are many different membership functions, and some are mentioned by Chung and Inder (1992), and Zadeh (1975). Generally, the function will approach one when the data is close to the target and decrease to zero as the data deviates above or below the target value.

The following figures show three typical membership functions and the equations that describe them. The function in figure 1 is called a flat MF in this thesis. The general shapes of the two other membership functions shown in figures 7.2 and 7.3 are known as the ramp and curve shaped MF. The curve MF was defined by Zadeh (1975).

Figure 7.1 Flat Membership Function

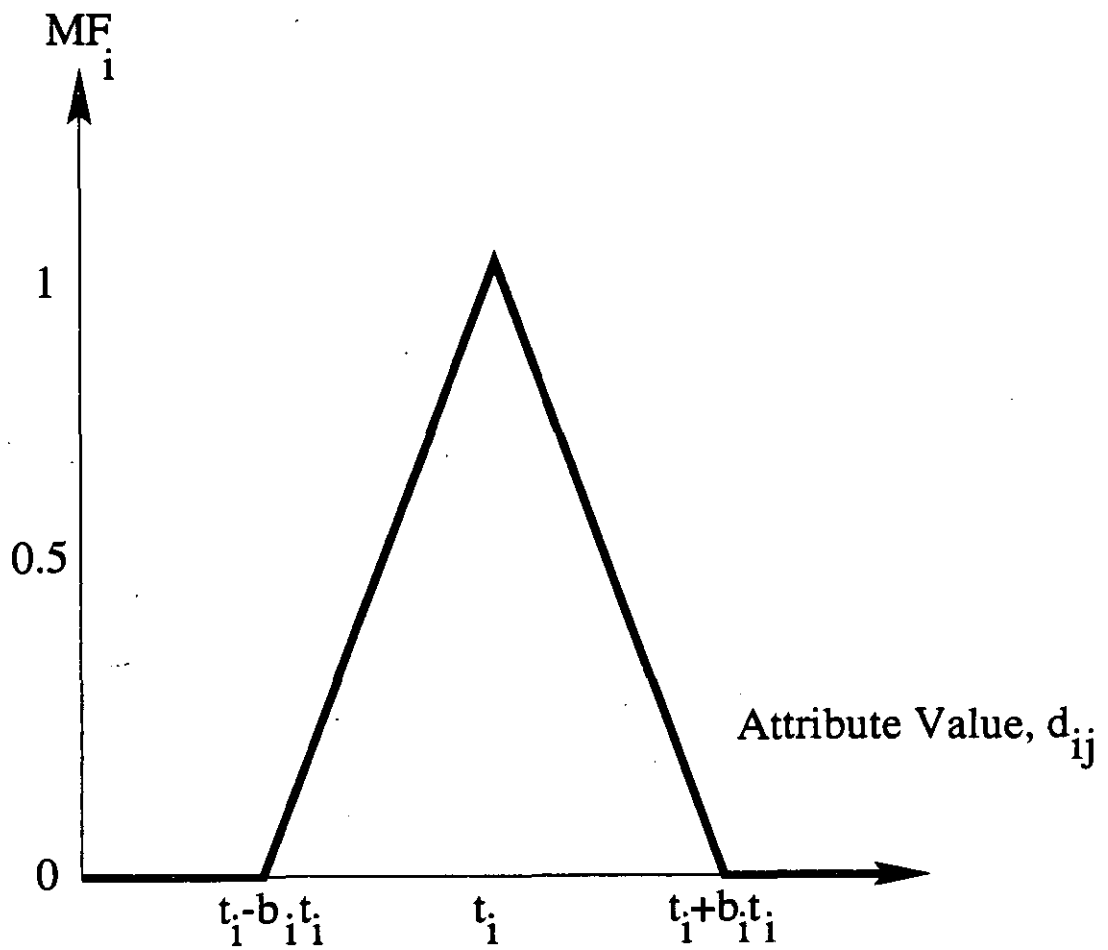


$MF_i = 0$, $d_{ij} < t_i - b_i t_i$

$MF_i = 1$, $t_i - b_i t_i \leq d_{ij} \leq t_i + b_i t_i$

$MF_i = 0$, $t_i + b_i t_i < d_{ij}$

Figure 7.2 Ramp Membership Function



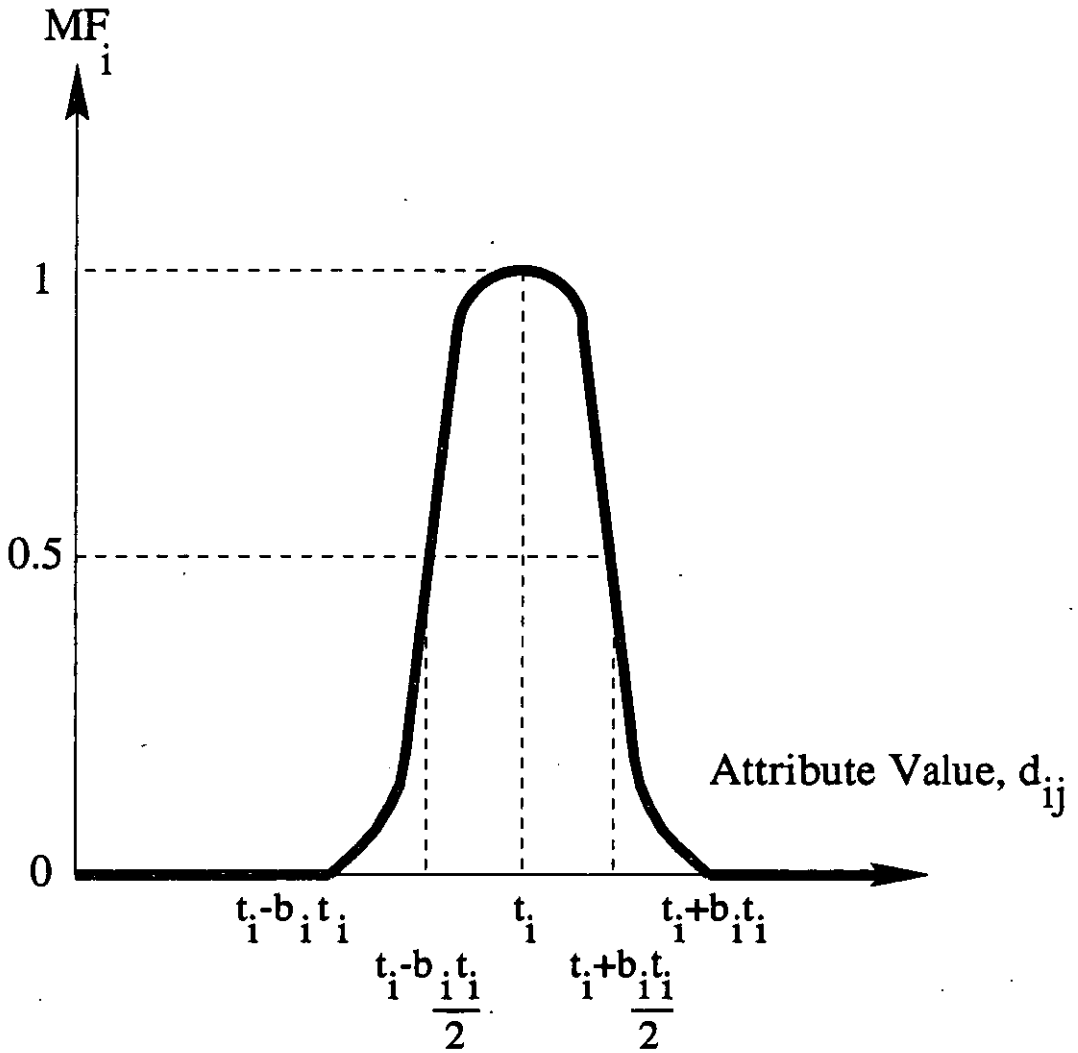
$MF_i = 0,$ $\quad d_{ij} < t_i - b_i t_i$

$$MF_i = \frac{d_{ij} - t_i + b_i t_i}{b_i t_i}, \quad t_i - b_i t_i \le d_{ij} < t_i$$

$$MF_i = \frac{d_{ij} - t_i - b_i t_i}{-b_i t_i}, \quad t_i \le d_{ij} \le t_i + b_i t_i$$

$MF_i = 0,$ $\quad t_i + b_i t_i < d_{ij}$

Figure 7.3 Curve Membership Function



$MF_i = 0$, $d_{ij} < t_i - b_i t_i$

$$MF_i = 2 \times \left(\frac{d_{ij} - t_i + b_i t_i}{b_i t_i} \right)^2 , \qquad t_i - b_i t_i \le d_{ij} < t_i - \frac{b_i t_i}{2}$$

$$MF_i = 1 - 2 \times \left(\frac{d_{ij} - t_i}{b_i t_i} \right)^2 , \qquad t_i - \frac{b_i t_i}{2} \le d_{ij} \le t_i + \frac{b_i t_i}{2}$$

$$MF_i = 2 \times \left(\frac{d_{ij} - t_i - b_i t_i}{b_i t_i} \right)^2 , \qquad t_i + \frac{b_i t_i}{2} \le d_{ij} < t_i + b_i t_i$$

$MF_i = 0$, $t_i + b_i t_i < d_{ij}$

7.2.2 Shape Parameter Optimisation

A method that uses fuzzy matching must be 'tuned' using the existing plants before being used to estimate the capital cost of new plants, if accurate estimates are to be achieved. The tuning is accomplished by finding the fuzzy matching set-up that provides the minimum average error when 'estimating' the capital cost for the existing plants using the other plants in the data set. An important part of the fuzzy matching set-up is the SP values which define the range of each MF. A technique for finding the set of shape parameter values that produces the optimum estimates is needed. The following sections describe the methods that are available.

7.2.2.1 Optimising Algorithm

The problem was to find the global minimum value of the average error. The average error behaves as a function whose variables are the shape parameters for each attribute.

E04JAF is a optimising routine in the NAG FORTRAN library (National Algorithm Group Ltd., 1986). This routine starts at a point defined by the user, which is in this case a set of values for the shape parameters. Then, by looking at the gradient and curvature of the function, points are generated using a quasi-Newton algorithm that are intended to converge on a local minimum. This method only requires the value of the function at various points.

7.2.2.2 Combinations

Another technique was adopted in which the average error was evaluated for as many different combinations of shape parameter values as possible and then the combination with the lowest average error selected. These combinations were obtained by selecting a minimum and maximum value for the shape parameters and an increment for varying between these two values. The size of the increment will alter the number of combinations tried. A compromise is required between getting an accurate estimate and the number of combinations evaluated, because the running time of the program increases dramatically with a large number of combinations.

The effect of the size of the increment value is shown by the number of combinations and the running time as shown in table 7.3. A minimum of zero and maximum of three were chosen for the shape parameters limits and four attributes were used.

Table 7.3 Effect of Step Increment

Increment	No. of Combinations	Running Time (minutes)
1.5	81	<1
1	256	<1
0.5	2 401	1
0.25	28 561	10
0.2	65 536	25
0.1	923 521	300

Note how a small increment value results in a long run time. An increment for the shape parameter of between 0.2 and 0.25 is practical, with a good balance between on one hand as small an increment as possible and the resulting lower average error, and on the other, a practical limit on the time taken by the computer to process all the calculations. The running time of 10 to 25 minutes may seem short, but the addition of only a few more attributes to the fuzzy matching results in the running time increasing to days.

The number of calculations required to find the ASEE for one set of shape parameter values is large when using 90 plants and four attributes. Each plant comparison requires the calculation of 4 match values when there are four attributes, and these values are then summed equalling 5 calculations. This means that, in order to find the best match and the error of the estimate for one plant 446 (89 × 5 + 1) calculations are needed. Therefore, for all 90 plants and the calculation of the ASEE there are 40 141 (90 × 446 +1) calculations, and this is for just one set of shape parameter values.

7.2.2.3 Genetic Algorithms

Exhaustively trying all the combinations of shape parameters was a very simplistic and slow way of optimising the fuzzy matcher. Genetic algorithms were another optimisation technique that was considered. This method tries a set number of randomly generated combinations of shape parameter values. The combinations with the lowest average error are then used in producing new combinations. GA were tried because they can be effective at finding optimum set-ups and avoiding local minima.

7.2.2.4 Asymmetric Membership Functions

Asymmetric MF were developed in pursuit of a lower AEEE. Such functions are steeper to the left of the target value than to the right. The reason for this approach is that having a data value less than the target value is more likely to result in an underestimate when the capital cost increases as the attribute value increases. The asymmetric MF assigns proportionally higher match values to data values which are greater than the target value. Therefore, existing plants with attribute values above the target value are more likely to be the best match, and so underestimates will be avoided and the AEEE improved.

7.2.3 Selecting the Best Match

Determining the best match for a new plant requires that the match values for the attributes be combined to produce a number that can be used to rank the match between each existing plant and the new plant, this number is called the ***Total Match Value (TMV)***.

7.2.3.1 Unweighted Total Match Value

The simplest way to quantify how closely an existing plant matches the new plant is to sum the $MF_i(d_{ij},t_i,b_i)$ for all the attributes and take the existing plant with the

highest total of MF_i values as the best match (7.2). In this method the best match has been determined using unweighted attributes.

$$\max_{j=1,..,q} \sum_{i=1}^{p} MF_i(d_{ij}, t_i, b_i) \qquad (7.2)$$

7.2.3.2 Weighted Total Match Value

Unweighted evaluation of the best match does not take into account attributes having a different degree of influence on the capital cost. For example, the capital cost might be affected more by the capacity than the maximum temperature - intuitively, capacity will affect every part of a plant, but the maximum temperature only applies to certain pieces of equipment. Consider a new plant found to have identical attribute values to those of two existing plants for all but one attribute in both cases. With the total match value the same in each case, but the non-identical attribute is capacity in one case and maximum temperature in the other. Then the existing plant with the identical capacity is the best match, but the fuzzy matcher views them as identical best matches.

One way to account for the different influence that attributes have on the capital cost is to weight the membership function values in the sum thereof. With larger weights, w_i, indicating greater importance. The equation that now calculates the match value total for an existing plant using a weighted sum of the MF_i is (7.3).

$$\max_{j=1,..,q} \sum_{i=1}^{p} w_i MF_i(d_{ij}, t_i, b_i) \qquad (7.3)$$

Ways of assigning values to the weights, w_i, are discussed in the following sections.

7.2.3.2.1 Expert Assessment

One possible method is to ask experts to assign values. This was not possible in this research because no expert estimators were available.

7.2.3.2.2 Assessment of Attribute Influence on Estimation Performance

Another method is to assess the relevant importance of each attribute by trying fuzzy matching with only one attribute. The resulting average estimate gives the strength of the relationship between that attribute and the capital cost. The average estimate achieved can be used to assign a value for the weight of each attribute, with the lower the average estimate when matching with one attribute resulting in the higher the weighting for that attribute.

7.2.3.2.3 Combinations

For the best estimate the weights must be optimised to produce the lowest average error. One option for finding the weights was to try different combinations of weights and assess which produced the best results. The combinations were generated in the same way as for the shape parameter optimisation. For example, weight combinations could be chosen from 0 to 1 in steps of 0.1, allowing an attribute to out-weigh another attribute by up to ten times. The resulting weightings for the minimum average error could be capacity, 0.7, maximum temperature, 0.3, etc.

7.2.3.2.4 Genetic Algorithms

The use of genetic algorithms to generate the weights was also tried. The same program that was developed for the generation of the shape parameter values using genetic algorithms was used. The results are discussed in the next chapter.

7.2.4 Attributes

A database containing the details of existing plant attributes and capital costs was required in order to develop fuzzy matching techniques for capital cost estimation.

Attributes included in the database should be ones that distinguish plants and have a bearing on plant cost. For example, the colour of the plant is irrelevant - unless some colours are a lot more expensive than others! Whereas, the capacity of a plant will have a major influence. However, the attributes must be available to a cost estimator at the stage of plant development when the estimate is required.

The decision on which attributes to use in the fuzzy matching requires a knowledge of their influence on the capital cost of a chemical plant. The frequency of the use of attributes in existing methods for early capital cost estimation shows the importance assigned by others to different attributes, see table 5.1.

Estimates with large errors will be investigated to see if there is a common cause, which may lead to the consideration of an attribute not previously anticipated as being important in estimating the capital cost.

The format of the attribute must also be considered. Numeric attributes, such as capacity, can be fuzzy matched using equations. However, linguistic attributes need to be assigned a numeric value by experts or by experimentation. For example, the materials of construction for a new plant is stainless steel and for an existing plant carbon steel. It is not possible for a computer or indeed a human to judge similarity without knowledge of these terms.

The attributes that were used in fuzzy matching, and the reasons why, are discussed in the next chapter.

7.2.5 Match Values

The match value represents the similarity between the data and target values for attributes. The total match value is the overall combination of match values for all of the attributes. The meaning of these match values needs to be considered.

7.2.5.1 Minimum Attribute Match Value

Firstly, the possibility of not allowing a plant to be a best match when one of the membership function values was below a certain level, for example 0.25. It could be better to ignore a plant as a possible match when any of its attribute match values are under a defined level. This would especially be the case for important attributes, such as capacity.

7.2.5.2 Total Match Value Versus Average Error

The TMV quantifies the closeness of the plant used for the estimate to the new plant. The TMV value shows the quality of the match between the two plants, and should therefore provide some confidence in the estimate. Correlating the TMV for the best match against the accuracy will show if higher match values result in lower errors.

7.3 Capital Cost Estimating Example of Fuzzy Matching

To clarify the method consider a simplified example. An estimate of the capital cost for a new plant is needed and the following attribute values are known: number of functional units, capacity, temperature, and pressure. The fuzzy matcher was used to find the two best matching plants with the results presented in table 7.4. The membership function used was a ramp, see figure 7.2. The matches were ranked by non-weighted summation of match values, according to equation 7.2. The shape parameter values were:

1.75	Number of functional units
1	Capacity
0.5	Temperature
1	Pressure

The highest total of match values was 3.13, out of a possible maximum of 4. The capital cost estimate is the capital cost of this best match and has an error of 7%. The size of the difference between the total match value of the best match and the maximum possible total match value indicates how similar the new process plant is to previously constructed plants. A large difference means that none of the existing plants have similar process details to the new plant.

Table 7.4 Example of Best Match

Attribute	**Target**	**Best match**		**2nd Best match**	
(units)	**Value**	**Value**	**MV**	**Value**	**MV**
Functional unit	24	38	.67	11	.69
Capacity (tonnes/y)	50000	40000	.80	65000	.70
Temperature (K)	643	753	.66	473	.47
Pressure (bar)	1.0	1.0	1.00	1.0	1.00
TOTAL MV			**3.13**		**2.86**
Capital Cost (10^6 1988)	68.00		73.00		25.74
Error (%)			7.40		-62.10

Chapter 8

PERFORMANCE OF FUZZY MATCHING

The results achieved when using fuzzy matching to estimate the capital cost of chemical plants are presented and analysed in the following chapter.

8.1 Implementation

A database of existing chemical plants with known attributes and capital costs was provided by the CS data and was used to test fuzzy matching. One of the plants in the database is taken to be a target (new) plant for which a cost estimate is required. Fuzzy matching then finds which of the remaining (existing) plants in the database is the best match. The capital cost of the best matching existing plant is taken as the estimate for the new plant. This procedure was followed for each plant in the database and the overall accuracy of the capital cost estimates was assessed using the metrics described in chapter 4, that is the average standard estimate error (ASEE) and the average equivalent estimate error (AEEE).

The first fuzzy matching programs were authored in the application language (PAL) of a relational database, PARADOX. The program was found to run for days when there were many combinations of shape parameters (SP) and weights tried in the optimization of the fuzzy matching method. In an effort to reduce the running time, the method was converted from PAL to PC FORTRAN, but the high number of calculations meant it still took twelve or more hours for the programs to run on a PC. However, the run time for the FORTRAN program was found to be a lot shorter on a Hewlett-Packard 9000, especially when a limited access machine was used. The final running time was reduced from days to hours by using these more powerful computers.

8.2 Initial Fuzzy Matching Set-up

One of the purposes of this chapter is to describe the development of the fuzzy matching technique. In order to accomplish this an initial set-up for fuzzy matching is defined in this section. Then the development of the fuzzy matching method from this starting point is shown in the following sections of this chapter.

The first decision for the initial set-up was which database of plants to use. There were two alternatives to choose from, a database of 79 plants which had already been used for testing existing methods and deriving new regression methods in chapters 4 and 5, or a larger database of 90 plants. The database that included the largest number of plants from the CS data was chosen for testing fuzzy matching, and consisted of 90 process plants. The values were known for the following attributes: capacity, throughput, number of functional units (No. of FU), maximum temperature (Max.T), maximum pressure (Max. P), workforce, number of reaction steps (RS), process phase and materials of construction (MOC). From this set of attributes a ***Standard Set of Attributes (SSA)*** were chosen: capacity, number of functional units, maximum temperature and maximum pressure. This particular set of four attributes were chosen as they are the most frequently occurring attributes in the existing methods for early capital cost estimation. The SSA are the initial set of attributes used in the fuzzy matching, with the attributes that are not included in the SSA investigated after the fuzzy matching technique has been refined. Four attributes were chosen because the run time for the fuzzy matcher was then hours rather than days.

An initial set-up for the combinations of SP values to be tried when searching for an optimum average error was required. The variables that define the combinations of SP values are the minimum and maximum for the values, and the increment that sets the values to be tried between the minimum and maximum. The initial set-up for the SP values was 0-5(0.25), with the 0-5 showing the minimum and maximum, and the value in brackets the increment. This set-up was chosen to be used with the SSA due to the resulting run time for the fuzzy matching program, which was around an hour. Also, when the SP values are shown in a table they are the set of values that produced the lowest average estimating error, with the first figure the SP value for the ASEE and the value in brackets for the AEEE value, for an example see table 8.1.

The best match was initially selected by unweighted addition of the match values.

Any alteration to this initial set-up will be mentioned in the discussion of the fuzzy matching experiments that follows. Any alteration to this initial fuzzy matching set-up that will be used from the point it is mentioned and onwards will be in italics.

8.3 Membership Functions

The membership function that fuzzy matching will use for the rest of this thesis was the first requirement. The MFs defined in chapter 7, section 7.2.1.1, were compared using the initial set-up for the fuzzy matcher. The fuzzy matching was carried out using the flat, ramp and curve membership functions. The results are shown in table 8.1.

Table 8.1 Comparison of Estimating Accuracy for Different MF

Membership Function	Best Performance SP				ASEE (±%)	AEEE (%)
	Capacity	No. of FU	Max. T	Max. P		
Flat	0.75 (0.25)	0.75 (0.50)	0.25 (0.50)	0 (0)	62	123
Ramp	1.00 (0.75)	1.75 (1.75)	0.50 (0.75)	0 (0)	47	73
Curve	1.50 (0.25)	2.00 (1.25)	0.75 (0.50)	0 (0)	48	78

The flat MF produced poor estimates. Fuzzy matching with this MF is similar to finding an exact match, but there are ranges of values over which attributes are considered to be identical. Hence, the number of existing plants that have attribute values that are deemed to be 'identical' to the target attribute value of the new plant is increased. The problem with this membership function is that two different values for the same attribute that are within the range of the flat MF will be assigned the same match value, which is one. Even though one of the values will normally be closer than the other to the target value.

The fact that the ramp MF gave slightly better results than the curve MF is a significant finding, because the curve MF is computationally more intensive than the ramp. The curve MF is defined by four equations over different portions of the data range, whereas the ramp only uses two. This means that the simpler ramp MF can be used, with shorter running time than the curve MF and yet with no loss of accuracy.

Therefore, the ramp is used as the MF in all the remaining fuzzy matching experiments in this chapter.

The ramp and curve MF would be expected to have similar SP values for the combinations that produced the near optimum average error, due to the relatively

similar shapes of the MF. This is seen to be true for the SP values shown in table 8.1.

The best performance SP values are said to be near optimal because the combinations of SP values that are tried search a grid of points, and so the global optimum is missed when it is in a space between the grid points.

The set of SP values that produced the best performance for the fuzzy matcher lead to some interesting conclusions. The maximum pressure SP value is zero for each of the membership functions in table 8.1, yet the SP value for the number of functional units is 1.75, when using the ramp MF. This indicates that maximum pressure is not as an important attribute as the number of functional units and the other attributes when fuzzy matching plants for CCE, because it only contributes to the MV total when the target and data values for the maximum pressure attribute are identical. This conclusion is proved later when the maximum pressure was not included in the set of attributes used for fuzzy matching, and the resulting increase in the average error was the smallest obtained by leaving one of the SSA out of the fuzzy matching.

8.4 Shape Parameter Optimization

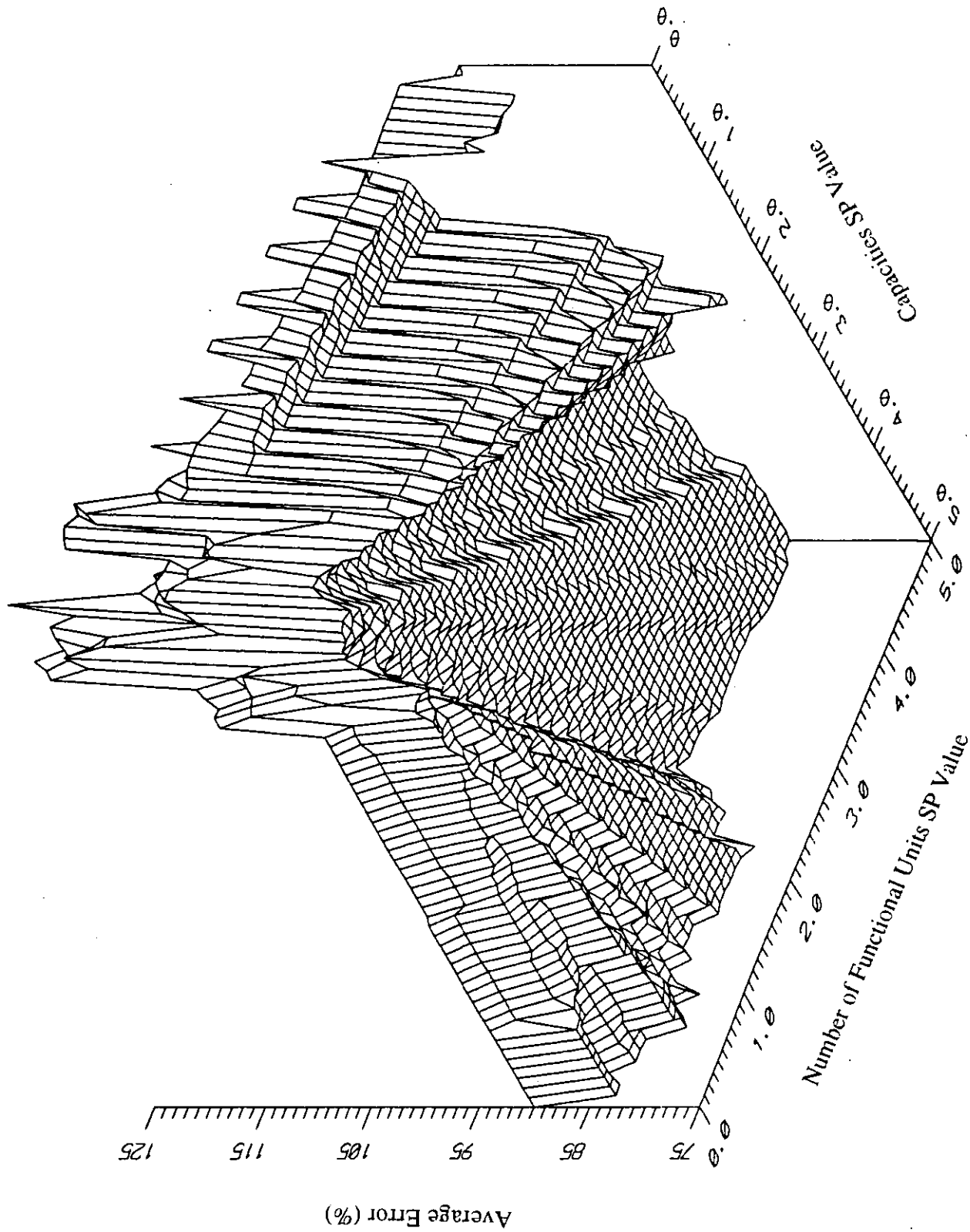
The most effective method for Optimising the SP values for the fuzzy matcher was investigated after the selection of the ramp as the membership function.

The significance of the SP value is not that clear. Increasing the SP value for an attribute leads to an increase in the match value for a particular data point, but the difference in the match values for two data values within the range covered by the MF will decrease. Therefore the differentiation between plants due to this attribute will decrease, but the match value total will increase. This situation is complicated by data values with a zero match value, where the difference in match value will also increase relative to a data value within the range of the MF. The point is that increasing the SP value will make the MV higher for this attribute, probably giving the attribute more effect on the selection of the best match, but it must be mentioned that the attribute's MV for other plants will also increase and so the effect is not as significant as you might think. Also, an SP value of 0 means that when the data value is the same as the target value there is a strong influence for this attribute, as the match value equals one, but all the other data values are viewed in the same way, as completely different and with a match value of 0. Therefore, a SP value of 0 does not mean that an attribute can be left out of the fuzzy matching, unlike a value of 0 for a weight which does mean this.

8.4.1 Optimising Algorithm

The E04JAF optimising routine was used to try and find the global minimum for the average error. However, this method did not improve upon the lowest average error found by the combination method. The reason being that the average error function is discontinuous with many local minima. Therefore, when the E04JAF optimising routine is given a randomly generated starting point for the SP values, then the method will move to a point that is a local minimum for the average error. The discontinuous nature of the average error function is illustrated by a three dimensional plot of the average error versus values for the shape parameters for two attributes, capacity and the number of functional units, see figure 8.1.

Figure 8.1 Three Dimensional Plot of Average Error versus SP Values for Two Attributes



8.4.2 Combinations

The membership functions size is defined by the SP value. This means that the MF will assign a match value to attribute values that are further from the target value as the SP value increases, and the MV will increase for attributes that already have a MV. The values that must be considered for the set-up that generates the different combinations of SP values are the minimum, maximum of the range of values tried and the increment to be used in the selection of the values in between.

As discussed in chapter 7, the SP values are chosen to minimise the average error of the estimates. A wide range for the SP values was used, in order to reduce the chance of missing the global minimum for the average error and the corresponding best performing combination of SP values, as a wide range ensures that the combinations tried cover a large area of the average error function, and therefore include the region where the global minimum exists. Also, the use of as small an increment as is possible for the SP values when creating the different combinations will improve the chances of finding the global minimum. However, there does have to be a limit to the increment and range, because otherwise the number of combinations becomes very large and the time required for the calculations takes too long for the computer to accomplish in a reasonable time scale. A compromise has to be sought between the running time and the number of combinations tried.

Firstly, the range of the SP values to be used was investigated by varying the maximum possible value for the SP, while keeping the increment at the same value of 0.25. Table 8.2 presents the near optimum accuracy achieved with different ranges when using the standard set-up for the fuzzy matcher. The maximum value found for a SP, out of the values for the attributes in the SSA for the optimum

Table 8.2 Shape Parameter Ranges

SP Range	Maximum SP Value in Optimum	ASEE (±%)	AEEE (%)	Time (minutes)
0 to 1	1.00	49	80	1
0 to 2	1.75	47	73	2
0 to 3	1.75	47	73	10
0 to 5	1.75	47	73	60
0 to 10	1.75	47	73	920

accuracy is also shown.

The minimum error for the ASEE and AEEE is found with a range of 0-2, further increases in the range only result in an increase in the run time. A decision was made from the results in table 8.2 to decrease the maximum value for the SP range from 5 to 3 as the standard set-up for the fuzzy matcher. The maximum SP value in a combination for a minimum average error was 1.75, which was too close to 2 for this to be the maximum SP value ever attainable. However, when the number of matched attributes increases, which dramatically increases the number of combinations, then lowering the maximum from 3 to 2 should be considered.

The set-up for fuzzy matching will now have a SP range of 0-3.

The next attempt to refine the fuzzy matching involved the increment used in the generation of the SP combinations. The SP range was still 0-5 for this fuzzy matching, but the highest SP value in an optimum case was 2.1, which agrees with the decision to change the SP range to 0-3.

Table 8.3 shows that the lower the increment the lower the minimum average error, but the run time increases rapidly as the increment gets smaller. However, using an increment of 0.1 or 0.2 resulted in too long a run time. Therefore, the SP increment for the fuzzy matcher set-up was kept at 0.25. However, the lowest possible average error would be found by using the lowest possible increment.

Table 8.3 Shape Parameter Increments

Increment	ASEE (±%)	AEEE (%)	Time (minutes)
1	54	82	4
0.5	48	80	9
0.25	47	73	60
0.2	45	73	150
0.1	45	71	3600

The SP set-up is 0-3(0.25) for the following experiments.

8.4.3 Genetic Algorithm

A genetic algorithm was developed and used, in place of the combination method, for finding the best performing set of SP values. The results are shown in table 8.4.

Table 8.4 Genetic Algorithms versus Combinations

Method	Best Performance SP				ASEE	AEEE
	Capacity	No. of FU	Max. T	Max. P	(±%)	(%)
GA	1.38 (0.91)	1.99 (2.16)	0.64 (0.81)	2.63 (2.08)	48	83
Combination	1.00 (0.75)	1.75 (1.75)	0.50 (0.75)	0 (0)	47	73

Clearly, with the ASEE ±10% higher, the results are not as good as for the combinations. However, when a larger number of combinations need to be tried later, due to the addition of weighting and extra attributes, then the GA could be used in place of the combination method. The argument for this is that the run time will not increase as rapidly as for the combinations and the results produced were not that far off the combinations minimum average error. The time taken to find the GA results was ten minutes, as compared with an hour for the combination method.

The SP values were similar for the first three attributes in the SSA. However, the SP values for the maximum pressure was not zero, as previously, and was in actual fact the highest SP value.

8.4.4 Asymmetric Membership Function

The membership function used was altered from the standard ramp MF to an asymmetric MF, but with the rest of the standard set-up for the fuzzy matcher unaltered. The results are shown in table 8.5.

Table 8.5 Average Errors for Asymmetric MF

MF	ASEE (±%)	AEEE (%)
Ramp	47	73
Asymmetric Ramp	56	78

It is apparent that biasing the membership function does not result in an improvement in the accuracy. The ASEE was expected to increase, but the AEEE should have improved, with the asymmetric membership function more likely to over estimate the capital cost. However, table 8.5 clearly shows that there was no improvement for the AEEE, with in fact a 5% increase in the error that occurred.

8.5 Selecting the Best Match

The fuzzy matcher had been selecting the best match by totalling the match values without weights. There now follows a description of the investigation into the use of weights in the selection of the best match.

8.5.1 Weighted Total Match Value

The MV of the attributes are weighted to reflect the relative importance of different attributes. A method has to be developed for finding the weights, that optimises the error of the fuzzy matching estimates.

Three methods for finding the weights were tried. Firstly, looking at the accuracy when fuzzy matching with just one attribute, in order to give an insight into the relative importance of each attribute. Secondly, trying different combinations of weights to produce the best estimates. Thirdly, using genetic algorithms to find the weights.

8.5.1.1 Assessment of Attribute Influence on Estimation Performance

Fuzzy matching was tried with just one attribute to gauge an idea of the relative importance of each attribute. Cases will often occur, when using one attribute, where none of the other processes attribute values are within the range set by the SP, especially for low SP values. This problem is avoided by taking the best match to simply be the existing process with the attribute value nearest to the target value.

The weights required for equation 7.3, which was described in chapter 7 for allowing weights to be used in fuzzy matching, are shown in table 8.6. The results

Table 8.6 Single Attribute Matching and Derived Weights

Attribute	ASEE		AEEE	
	(±%)	w_i	(%)	w_i
Capacity	88	1.14	110	0.91
No. of FU	110	0.91	150	0.67
Max Temperature	118	0.85	198	0.51
Max Pressure	121	0.83	198	0.51

clearly show that the capacity attribute has the lowest average error when fuzzy matching with a single attribute, and is therefore the most closely related to the capital cost. The weights shown in table 8.6 were calculated by dividing 100 by the ASEE and the AEEE for the attribute. The reason for dividing by the average error was an arbitrary decision to make the weights inversely proportional to the average error, making the lowest averages have the highest weighting. Dividing by one hundred produced weights with values of around one. Then the weights were used with the SSA, the ramp MF and the set of SP values that produced the lowest average estimates for the combinations from the SP set-up of 0-3(0.25).

The derived weights resulted in an increase in the average errors when compared to unweighted fuzzy matching, as seen in table 8.7. The increase was ±8% for the ASEE and 15% for the AEEE. This method for deriving the weights was not used after these results.

Table 8.7 Effect of Derived Weights

Weights	ASEE (±%)	AEEE (%)
None	47	73
Derived	55	88

8.5.1.2 Combinations

The next method was to find the near optimum weights by enumerating combinations in the same way as for the SP values optimization.

Firstly, the range for the values of the weights to be used was investigated. This involved varying the maximum possible value for the weight while keeping the increment at the same value of 0.25, in an identical fashion to that used to test the ranges for the shape parameter values. Table 8.8 presents the optimum accuracy achieved with different ranges and using the standard set-up for the fuzzy matcher, apart from the SP value being set to unity for all attributes.

Table 8.8 Weight Ranges

Weight Range	Maximum Ratio	ASEE (±%)	AEEE (%)	Time (minutes)
0 to 1	4	53	87	1
0 to 2	8	53	81	4
0 to 3	5	49	76	40
0 to 5	4.67	49	73	120
0 to 10	5.2	48	73	1500

Table 8.8 shows that the use of weights, with an SP value of unity, produces as nearly as good a match as that produced by trying different combinations of SP values, ±1% higher for ASEE, but the same for the AEEE. The larger the range the better the estimate, but on the other hand the run time becomes longer. As for the SP, a range of 0-3 seems to result in the right balance between accuracy and run time.

The interpretation of the value of a weight is much clearer than that of a SP value. The ratio between the values of the weights for attributes shows the effect of one attribute in relation to another in the selection of the best match. This difference in value is called the ratio for the following discussion. For example, an increment of 0.25 with a maximum of one for the range allows an attribute to have weights between 0.25 and 1, and so the maximum ratio is 4, and for a 0.05 increment the maximum ratio increases to 20. The largest ratio for a near optimum set of attribute weights was 8, and so long as the increment and range allow this figure to be possible then the combinations of values considered are suitable. For example, 0-2(0.25) or 0-1(0.1) are suitable set-ups for the range and increment of the weights.

In fact, the maximum of the range can be viewed as a constant and the increment used as the only variable for the set-up. This is achieved by keeping the range set at 0-1, and then varying the increment. For example, using 0.05 would in effect result in the same combinations of weights being tried as for a weights set-up of 0-5(0.25). Therefore, for the increment experiments the range of values for the weights was kept at 0-1. The rest of the set-up for the fuzzy matching was standard, with the SSA, a ramp MF and with the SP values all set at one.

The range of possible values for the weights when fuzzy matching is from 0 to 1.

The results for the variation of the weight increment are shown in table 8.9. Again, the lower the increment the lower the average error. However, as mentioned already, a balance between accuracy and run time must be made and so an increment of 0.1 for a range of 0-1 is best. This also allows a maximum ratio of 10 which encompasses the greatest difference found in a near optimum case.

Table 8.9 Weight Increments

Step	Maximum Ratio	ASEE (±%)	AEEE (%)	Time (minutes)
1	1	78	118	<1
0.5	2	56	87	<1
0.25	4	53	87	1
0.2	5	53	84	2
0.1	8	49	81	20
0.05	4.67	49	73	120

Looking at the values of the weights for the lowest average errors showed that the values of the weights for the pressure attribute to be the lowest and therefore reinforced the view of it being the least important attribute of the SSA. The maximum temperature had the highest value when only varying the weights.

The next fuzzy matching enumerated combinations of both SP and weight values. The set-up for the fuzzy matching used the SSA, ramp MF, and a SP set-up of 0-3(0.25) and weights set-up of 0-1(0.2), with the resulting running time being 2 days.

Table 8.10 shows that the average error is reduced by a combination of weighting the attribute MF values in the summation and by optimising the attributes SP values. The addition of weighting improved the average accuracy of the ASEE by ±5% and the AEEE by 7% when compared to equal weighting for each attribute. Thus, showing the importance of weighting in fuzzy matching.

Table 8.10 Effect of Both SP and Weights on Average Error

Variant		Attribute				ASEE	AEEE
		Capacity	No. of FU	Max. T	Max. P	(±%)	(%)
None	SP	1	1	1	1	78	118
	W	1	1	1	1		
SP	SP	1.00 (0.75)	1.75 (1.75)	0.50 (0.75)	0.00 (0.00)	47	73
	W	1	1	1	1		
	SP	1	1	1	1	48	73
Weights	W	0.275 (0.225)	0.25 (0.175)	0.65 (0.35)	0.125 (0.075)		
SP &	SP	1.25 (1.50)	2.00 (2.00)	0.50 (0.50)	0.25 (1.25)	42	66
Weights	W	0.80 (1.00)	0.80 (0.80)	1.00 (0.60)	0.60 (0.20)		

8.5.1.3 Genetic Algorithms

Genetic algorithms (GA) were used in an attempt to replace the combination method that has been discussed previously. The GA approach for the selection of the weight values for the attributes is based on the approach used for the SP values.

The range for the SP values was 0-3, and 0-1 for the weights, so that a comparison between GA and combinations could be made.

The results for the GA approach are not as good, as seen in table 8.11, with the ASEE ±7% worse and the AEEE 13%. However, the run time for the combinations was 2 days as compared to 1 hour for the GA. The GA used was only a trial one written by the author, with a commercially available package the average error achieved would be improved upon.

Table 8.11 Comparison of Combinations and Genetic Algorithms

Variant		Attribute				ASEE	AEEE
		Capacity	No. of FU	Max. T	Max. P	(±%)	(%)
Comb	SP	1.25 (1.50)	2.00 (2.00)	2.00 (2.00)	0.25 (1.25)	42	66
	W	0.80 (1.00)	0.80 (0.80	0.80 (0.80	0.60 (0.20)		
GA	SP	2.84 (2.59)	2.81 (2.00)	0.85 (1.29)	1.59 (0.53)	49	79
	W	0.72 (0.48)	0.46 (0.22)	0.58 (0.65)	0.24 (0.02)		

8.6 Attributes

The attributes used when fuzzy matching must be selected with the production of the estimate and its accuracy in mind. Attributes that are not available in the early stages of plant design, and those that have no influence or even a small influence on the capital cost should be omitted from the fuzzy matching method. For example, the workforce is viewed as inappropriate because it is not known in the early stages of process development. Also, pressure is a possibility for omission from fuzzy matching as it was found to be less influential than the other SSA. This is in contradiction to the existing methods which all include pressure.

Some attributes are process classifications such as batch/continuous and materials of construction. These attributes required different methods for calculating their MF values as their values are linguistic, in contrast to the numeric values of the other attributes considered. Linguistic means that the value is a word with no numeric value and is an assignment to a class derived for that attribute. For example, the material of construction for an ammonia plant is stainless steel.

The effects of each attribute in the data supplied by CS when fuzzy matching are discussed in the following sections.

The set-up used when fuzzy matching was of the standard form, with the new attributes added to the SSA. The ramp membership function was used. The set-up for the shape parameters was 0-3(0.25), unless stated otherwise. However, weights as discussed in the last section were left out of the fuzzy matching to prevent excessive run times, with the set of attributes that produced the best average error used with weights in the final results presented in the conclusion.

8.6.1 Capacity

Existing capital cost estimation methods have not always included the capacity, replacing it with a more complicated, but related quantity, throughput. Throughput is a measure of the amount of material passing through the process, whereas the capacity is the amount of product leaving the plant.

There are two reasons that the throughput might be a better attribute for matching plants. Firstly, some plants will require several tonnes of raw materials to produce

a tonne of product, while in others a tonne of raw material can produce a tonne of product. Two such plants could have the same capacity and so be assigned the same match value for the capacity attribute. However, the first plant would have larger pieces of equipment due to the larger volume of material moving through at least some parts of the plant, and therefore have a higher capital cost. Secondly, the size of equipment can be affected by a large recycle. The throughput is a way of representing both these situations.

Throughput was calculated from the CS data by dividing the amount of raw material entering into the plant by the amount of products and useful co-products coming out of the plant. The throughput can not be lower than one. However, there were cases where the throughput was calculated to be less than one, normally when air was a major raw material and the consumption was not specified in the CS data. For these cases the throughput was assumed to be one.

One point to note is that there were two different sets of throughput figures, one considered raw material throughput and the second included the amount of process water. This was necessary as the process water is used in the reactions, and so should be considered a raw material.

A comparison of the performance of fuzzy matching for these three different attributes is shown in table 8.12. The SSA were used for the capacity results and then each of the two different throughputs were used in turn in place of the capacity.

Table 8.12 Capacity versus Throughput

Attribute	ASEE (±%)	AEEE (%)
Capacity	47	73
Throughput	76	138
Throughput (process water)	57	108

Capacity was found to give the best accuracy, by ±10% for the ASEE and 35% for AEEE, and so was used in all further fuzzy matching. The capacity gives better results because it is a more accurate figure, while the throughput was susceptible to errors in its calculation from the CS data because some details were missing, such as the amount of air required and size of recycles in the process.

The importance of each attribute in the SSA needed to be considered. This assessment was achieved by removing each of the four SSA from the fuzzy matcher in turn, and then analysing the effect on the lowest average error which was then found. The results are shown in table 8.13.

Table 8.13 Effect of Attributes in SSA on Fuzzy Matching

SP Value for Attribute				ASEE	AEEE
Capacity	No. of FU	Max. T	Max. P	(±%)	(%)
1 (1.75)	1.75 (1.75)	0.5 (0.75)	0 (0)	47	73
-	1.25 (1.5)	1.75 (1.00)	1.75 (1.75	77	135
2.00 (0)	-	0.25 (0.25)	0.75 (0)	53	103
2.75 (0)	1.00 (2.75)	-	1.25 (2.50)	65	119
2.00 (1.25)	2.00 (2.00)	0.50 (0.50)	-	52	82

Leaving out the capacity resulted in the largest increase in the average error. The increase for the ASEE was a very large ±30% and an even larger increase of 62% for the AEEE.

The capacity should be included in the set of attributes for fuzzy matching.

8.6.2 Number of Functional Units

The definition of a functional unit varies from one person to the next, but in this work the number of functional units were calculated for the CS data by counting the pieces of equipment in the process diagrams that were of the type as defined by the I. Chem. E. Assoc. Cost Engrs. (1988), see chapter 3 for the definition. The number of functional units for each process was calculated from the CS data by the author and also during the course of an undergraduate final year project. This allowed a check to be made on the figures by subtracting one set from the other and looking closely at any plants where there was a large difference. This resulted in a few changes and a new amalgamated and more representative set of values. This combined set of values for the number of functional units was found to produce the lowest average error.

The importance of the number of functional units attribute when fuzzy matching is

shown in table 8.13. This attribute has the third largest effect on the average error of the four SSA, with the ASEE relatively unaffected with an increase of ±6%, but a large increase of 30% for the AEEE.

The number of functional units should be included in the set of attributes for fuzzy matching.

8.6.3 Maximum Temperature

The temperature attribute for a chemical plant represent the temperatures encountered during the chemical process. The value of the temperature attribute was determined in two ways. Firstly as the maximum process temperature and secondly as the mid point of a range if one was quoted in the CS data. Taking the maximum value was found to give the best results. This was expected because equipment has to be constructed to cope with the maximum possible temperature.

Again, table 8.13 shows the importance of the maximum temperature attribute when fuzzy matching. This attribute results in the second largest increase in the best average error. The ASEE increased by ±18%, and the AEEE by 46%, which are large increases.

The maximum temperature should be included in the set of attributes for fuzzy matching.

8.6.4 Maximum Pressure

The pressure attribute represents the pressures found within the process plant. The maximum pressure was the best way to determine the attribute values, as was the case for the temperature.

However, it was noticeable that the SP value for the pressure attribute was often zero. The interpretation of this result was that the pressure attribute was not so important. This was backed up by the accuracy found when fuzzy matching was undertaken with the omission of the pressure attribute from the SSA, see table 8.13. Leaving out the pressure attribute resulted in the smallest increases of ±5% and 9% for the ASEE and AEEE respectively. Therefore, the pressure attribute is not so

important when using fuzzy matching for capital cost estimation. This was surprising, given that pressure determines the type of vessel and its wall thickness, both of which are directly related to capital cost. The reason for its insignificance is unclear.

The use of the pressure attribute when fuzzy matching is just about worthwhile.

8.6.5 Workforce

The correlation results in chapter 5 implied that the capital cost was strongly related to the size of the workforce, with the reasons for this discussed at that point.

However, it is not possible to use the workforce in a method for early capital cost estimation because the labour requirements are not normally known in the early stages of plant development, which is when the fuzzy matching estimation technique is designed to be used. The success of the attribute as an important parameter when estimating the capital cost of a plant is worth remembering. Even though there is some doubt about the stage in the process development at which the workforce is known.

Table 8.14 shows the effect of adding the workforce to the SSA. The improvement in the average errors is large, ±11% for the ASEE and 15% for the AEEE.

Table 8.14 Workforce

Attributes	ASEE (±%)	AEEE (%)
SSA	47	73
SSA+Workforce	36	58

The workforce attribute should be used, if available, when fuzzy matching.

8.6.6 Number of Reaction Steps

A reaction step in a process is a distinct chemical reaction that requires a dedicated reactor in the corresponding chemical plant. The important consideration when

counting the number of reactors is that some plants have a reaction step that uses reactors in parallel in order to increase the plant capacity due to size limits on the equipment. One reaction step might have two or more identical reactors, which should not be counted separately.

The number of reaction steps in each process were counted, with parallel reactors counted as one reactor. The number of reaction steps was then used instead of the number of functional units and also in addition to the SSA. The ASEE and AEEE obtained when matching with these sets of attributes are presented in table 8.15. The number of functional units attribute produced slightly lower average errors than when replaced by the number of reaction steps, by 2% for the AEEE. However, when both attributes were used a reasonable improvement in the average errors of ±6% and 5% for the ASEE and AEEE respectively resulted.

Table 8.15 Functional Units versus Reaction Steps

Attribute	ASEE (±%)	AEEE (%)
No. of FU	47	73
Reaction Steps	47	75
Both	41	70

These results led to the number of reaction steps being included as a fuzzy matching attribute.

8.6.7 Process Phase

The process phase is the physical state of the materials that pass through a plant, for example, vapour, liquid, solid or a combination. Plants may be classified by the dominant phases. There were five different types of dominant process phase in the CS data plants: vapour, liquid/vapour, liquid, liquid/solid, solid.

The membership functions described in chapter 7 are only suitable for numeric values and are therefore unsuitable for the linguistic values of the process phase. Two slightly different methods were tried in place of these membership functions. Firstly, when an exact match was made between the process phase values for the

data and target plants, then a value was added to the match value total. In effect the value used weighted the importance of the phase attribute. The value added was varied as for the combinations and weights. Also considered were the phases nearest to the target phase. For example, a liquid phase process has something in common with both a liquid/vapour and a liquid/solid. Therefore some sort of match value is warranted, but of a lesser value than an exact match. The value added for the nearest phase was varied in a similar way to the value for the exact matches.

The set-up for the values tried for the exact match was 0-3(0.25), as used previously for the shape parameters values. The set-up for the variations in values tried for the nearest phase was 0.2-0.8(0.2), this used a smaller range to prevent long run times.

The process phase was added to the SSA used in the fuzzy matching. Table 8.16 shows the results for the cases where the phase is not considered, then with only an exact phase match, and thirdly with the nearest phases given a match value. The process phase attribute did lead to a small improvement of 3% for both the ASEE and the AEEE when the exact and nearest phases were considered.

Table 8.16 Process Phase

	Phase	Nearest Phase	ASEE (±%)	AEEE (%)
SSA	0	0	47	73
SSA + Exact Phase	0-3(0.25)	0	47	71
SSA + Exact Phase & Nearest Phase	0-3(0.25)	0.2-0.8(0.2)	44	70

However, the improvement was too small to merit the addition of the process phase attribute to fuzzy matching, and the extra run time that would result.

8.6.8 Materials of Construction

The ***materials of construction (MOC)*** usually appears as a plant parameter in the traditional methods. However, the MOC was often missing from the CS data. The following chemical engineering encyclopaedias: Sittig (1967), Ullmann (1988), and Kirk-Othmer (1992) were used to provide the MOC for the plants in the CS data.

The MOC attribute has linguistic values such as: Hastelloy C, stainless steel and carbon steel. The membership function had to be replaced by another method to quantify the similarity between the attribute values for the purpose of cost estimation. This was possible in two ways, firstly as for the process phase attribute, or secondly by the conversion of the linguistic values into numeric values.

Zevnik and Buchanan (1963) and Wilson (1971) derived factors for the MOC, for use in their correlation-based estimating techniques. Wilson (1971) developed a method that used the MOC as a multiplying factor, with a carbon steel plant providing the base cost. For example, when estimating the cost of two plants which are identical to each other, except for one being constructed from titanium and the other of carbon steel, the Wilson method would use all the other necessary plant attribute values in an equation to produce the base cost estimate and then multiply by the relevant MOC factor. Therefore, the titanium plant would cost twice as much as the carbon steel plant. Table 8.17 shows the Wilson factors.

Table 8.17 The Wilson Factors

Wilson Factor	Material of Construction
1.00	Carbon steel
1.28	Stainless steel(400 series)
1.5.	Stainless steel(300 series)
1.54	Hastelloy C
2.00	Titanium, Tantalum

These factors were used as numeric values for the MOC attribute, which is then suitable for fuzzy matching in the same way as the other attributes. The attribute values are for the relative cost of building a plant containing that particular MOC.

When there was more than one MOC for a plant, then a decision had to be made on how to combine the factors for the different MOC to produce the attribute value. The possibilities are to either choose one of the MOC attribute values (factors), or calculate a composite value. The first method selected the largest MOC factor out of those for the MOC used in the plant. For the second method the factor values for all the MOC used in the plant were totalled, and thirdly a mean factor value was calculated by totalling the factors and then dividing by the number of different materials. Allen & Page (1975) state that in their traditional method, taking the

mean MOC factor led to an improved accuracy. This is a different result than for temperature and pressure, where the maximum is best. The results are shown in table 8.18.

Table 8.18 MOC Results

Attributes	ASEE(±%)	AEEE(%)
SSA	47	73
SSA + Largest Wilson MOC	46	74
SSA + Total Wilson MOC	46	74
SSA + Mean Wilson MOC	46	74
SSA + Largest Zevnik & Buchanan MOC	52	80
SSA + Largest Zevnik & Buchanan MOC	53	80
SSA + Mean Zevnik & Buchanan MOC	51	79

The factors developed by Wilson (1971) produced the more accurate estimates, and made a very slight improvement on the ASEE for the SSA. The overall improvement in accuracy by the addition of the MOC attribute was ±1% for the ASEE, but increased the AEEE by 1%. The method for calculating an attribute value when there was more than one MOC made little difference. In fact 87% of the plants had only one quoted MOC, and so the value for the attribute is normally unchanged by taking the maximum, mean or total.

The best way of calculating the MOC attribute would have been to have used the fraction of the plant made up of each material to calculate a weighted mean. However, this was impossible with the CS data process descriptions, because the proportions of each material from which the plants were constructed was unknown.

The MOC attribute was not used when fuzzy matching.

8.7 Match Values

The match value represents the similarity between data and target values of an attribute when fuzzy matching. This section investigates the meaning of a match value for a particular attribute and for the overall match value, which is the total of the match values for a combination of the attributes.

8.7.1 Minimum Attribute Match Value

A desired condition when selecting the ***Best Match (BM)*** between a new plant and a data plant is that the values for each attribute are close. This could be achieved if the fuzzy matcher only selects a plant as the best match when each of the attribute match values is above a certain minimum value. However, there then might not be an existing plant that matches sufficiently on all the attributes to satisfy the constraint, and so no matching plant is found. This is especially true when low SP values are used in conjunction with high minimum match values.

A minimum match value was used for four attributes, with the results shown in table 8.19. The fuzzy matcher used the SSA, with a ramp MF, a SP set-up of 0-3(0.25), and no weights. The columns titled No. with BM show the number of plants that had a best match for the minimum MV constraints. The average error is only calculated for the plants with a best match.

Table 8.19 Minimum Match Value for All Attributes

Min. MV	ASEE (±%)	No. with BM	AEEE (%)	No. with BM
0	47	90	73	90
0.25	49	83	87	84
0.50	46	84	86	83
0.75	65	81	105	80

The table shows that the error increased with each increase in the minimum MV, apart from for the ASEE when the minimum MV was 0.5. However, the improvement in the ASEE was only ±1%, and was for only 84 plants.

This approach did not improve the fuzzy matching technique. The fact that a

minimum MV constraint makes the results of fuzzy matching worse, indicates that the SP value allows for the importance of each attribute. For example, a high SP value for an attribute will mean that most of the target plants will have a high MV for that attribute.

There is a possibility that a plant with a low match value for capacity, but with a high MV for all the other attributes will be selected as the best match. However, if capacity is the most important attribute, then this might not be a good plant to select as a best match. Therefore, to avoid this the MV for a particular attribute has to be above a designated value. This idea uses a similar methodology to weighting, which would prevent this situation by increasing the relative size of a match value for an attribute by increasing the size of its weight. The set-up for the fuzzy matcher was the same as for the minimum MV for all of the attributes, with the results shown in table 8.20.

Table 8.20 Minimum Match Value for Attribute in SSA

Attribute	Minimum Match Value							
	0.25				0.5			
	ASEE	No. BM	AEEE	No. BM	ASEE	No. BM	AEEE	No. BM
Capacity	46	90	79	90	47	90	79	90
No of FU	46	90	74	90	46	90	73	90
Max T	45	89	72	89	44	89	71	89
Max P	51	90	85	90	51	90	83	90

The average error improved slightly for the number of functional units, by ±1% for the ASEE. The biggest improvement in the average error was for the maximum temperature attribute, with a minimum MV of 0.5. The ASEE improved by ±3%, form ±47% to ±44%, and the AEEE improved by 2%, from 73% to 71%. However, the number of plants for which a best match was found was reduced from 90 to 89. Therefore, this method was left out of the final fuzzing matching methodology because the small improvement was outweighed by the failure to estimate the cost for all plants.

8.7.2 Total Match Value Versus Average Error

The total match value quantifies the similarity between the plant used for the estimate to the new plant, therefore it should provide an indication of the accuracy of the estimate. Regressing the best match value against the accuracy of the estimate should show if a higher match value means a better accuracy, and hence fuzzy matching is working as hypothesised. The total match value can then be used for assigning a degree of confidence to the estimate. If a perfect match is found over all the attributes and the estimate is nevertheless inaccurate, then some attributes must be missing which have a bearing on the quantity being estimated; this was expected in this case because only attributes known in the early stages of process development were being used.

The best fuzzy matching results for the 90 plants, which is described in the final section, was used to investigate the relationship between the total match value and the error for an estimate.

Non-linear regression produced the equations that best represented the relationship. However, the coefficient of determination (R^2) value was 0.2 for the relationship between the total match value and the SEE and the EEE. This means that there is a lack of confidence in the relationship derived by the regression. The equations were:

$$SEE = \frac{7331}{TMV^{5.83}} \tag{8.1}$$

$$EEE = \frac{13620}{TMV^{7.04}} \tag{8.2}$$

According to both equations the error will decrease as the TMV increases, and so the conclusion is that the TMV does give an indication of how good the estimate will be. However, the representation of the relationship is questionable due to the low coefficient of determination for these equations.

8.8 Conclusion & Comparison of Accuracies

In this section the best results for fuzzy matching are discussed. A comparison is made between the accuracy obtained for the existing, new correlations and fuzzy matching methods.

The lowest average error achieved with fuzzy matching using 90 processes, not using the workforce attribute, was for the SSA plus the number of reaction steps. The set-up details for the fuzzy matcher are that the membership function was a ramp, with SP set-up 0-2(0.25) and weights set-up 0-0.99(0.33). This set up was chosen to reduce the running time, which was still two weeks, to calculate the ASEE and AEEE for all possible combinations of SP and weights. The resulting accuracy was ±39% for the ASEE, and 64% for the AEEE. The complete details for the SP and weight values that produced the best fuzzy matching results for the 90 plants are shown in table 8.21. These results put the fuzzy matching estimates somewhere between an order of magnitude and a study estimate, based on the realistic estimate figures shown in Table 2.1.

Table 8.21 Best Fuzzy Matching Result

	Attribute					ASEE	AEEE
	Capacity	No. of FU	Max. T	Max. P	No. of RS	(±%)	(%)
SP	0.75 (1.00)	1.25 (1.25)	0.75 (0.50)	0 (0)	0.50 (0.75)	39	64
W	0.99 (0.99)	0.99 (0.99)	0.99 (0.66)	0.99 (0.66)	0.33 (0.33)		

Table 8.22 compares existing methods, new correlations and fuzzy matching. The existing estimating methods developed by Bridgwater produced the best accuracy of ±61% and 89% for the ASEE and AEEE respectively for the 90 processes. A correlation equation, derived using regression on the CS data was also used for estimating the capital cost and the average error over all the processes was ±53% for the ASEE and 77% for the AEEE, which meant that the fuzzy matching

Table 8.22 Lowest Errors Obtained For Different Techniques

Technique	ASEE (±%)	AEEE (%)
Existing methods	61	89
New Correlation	53	77
Fuzzy Matching	39	64

technique produced better estimates by ±14% for the ASEE, and 10% for the AEEE.

Fuzzy matching is not always the best estimation technique. For example, when using two variables, the correlation equations perform better than fuzzy matching with two attributes. For three attributes/variables the estimates are better when using fuzzy matching for the ASEE, but are worse for the AEEE by 2%. Fuzzy matching does much better than regressions for four or more attributes, see table 8.23. In the columns headed fuzzy of table 8.23 are the average errors obtained when fuzzy matching with the various combinations of attributes are presented. The columns headed regression list the average errors obtained using equations in terms of the same attributes for the capital cost that have been derived by regression. The set-up for the fuzzy matching used a ramp MF, SP value combinations, and weight combinations.

Table 8.23 Average Errors for Fuzzy Matching and New Correlations

Attributes		ASEE (±%)		AEEE (%)	
No.	**Used**	**Fuzzy**	**Regression**	**Fuzzy**	**Regression**
2	Capacity & No. of FU	71	56	113	79
3	SSA without Max P	48	55	80	78
4	SSA	42	55	66	78
5	SSA & RS	39	53	64	77

Finally, the number of plants in the database is an important consideration. Intuitively, there is a greater chance of finding a close match, and hence a more accurate estimate, when fuzzy matching with as many plants as are available in the database. The results for the 90 plant test database have already been discussed, but a 79 plant database was used to test existing methods and new regressions in chapters 4 and 5. The average estimate errors obtained for this database are presented in table 8.24, and compared with the database of 90 plants.

Table 8.24 Number of Plants Effect on Fuzzy Matching

Number of Plants	ASEE (±%)	AEEE (%)
79	45	77
90	39	64

The average error was improved by ±6% for the ASEE and 13% for the AEEE when using the larger of the two databases.

Fuzzy matching should use the maximum number of existing plants that are available.

Chapter 9

CONCLUSIONS AND FURTHER APPLICATIONS

This final chapter has a brief discussion of how fuzzy matching could be applied in other fields of chemical engineering. Finally, there is a discussion on the conclusions from the application of fuzzy matching to capital cost estimation.

9.1 Further Applications of Fuzzy Matching

A brief discussion follows of two possible further applications of fuzzy matching in chemical engineering. Fuzzy matching could also be applied to many other fields of research.

9.1.1 Plant Design and Modifications

Most designers copy and modify what is known to have worked. This is why many chemical processes have a similar arrangement of equipment. This methodology is not surprising when the amount of money that would be wasted on a plant that was non-operational is considered, and the potentially dire health, safety and environmental consequences of design faults. The human copying and modification process requires selection, from memory and/or documentation, of a flowsheet or flowsheet fragment which either achieves the current design aim or which may be modified to achieve it. Fuzzy matching could provide a tool for helping a designer find suitable designs to use when developing a chemical plant.

Also, a chemical company is now more likely in the UK to redesign (retrofit) an existing chemical process rather than construct a new plant. The reasons for this are the

large amount of capital required to construct a new plant, especially when compared with plants constructed abroad and the current economic situation.

Retrofitting a plant might to be necessary for various reasons:-

1) New regulations regarding the performance of plants. For example, a new environmental regulation on emissions of SO_2.

2) A change in the requirements of the output from the plant. The most common being an increase in the amount of main product. The flowrates through the plant are increased and result in some of the current equipment being inadequate - the 'debottlenecking problem'.

3) In order to upgrade the plant equipment. This is normally a consequence of the changes required to the plant by the two previous points. However, an upgrade could also reduce the operating costs.

The redesign of the plant and its process will require a modification of the existing equipment and/or installation of new pieces of equipment. Consideration of the alterations is difficult and involves:-

1) An understanding of the performance of the existing process and equipment.

2) Selecting the equipment that needs altering for the required changes in the plant to be possible.

3) A simulation of the effects of the redesign on the rest of the process.

The records for previous plant modifications are kept, because an alteration can have a significant effect on the operation of a plant and its safety. These records include all the details of the redesign, its effect on the plant operations, and the cost.

Access to details of an identical or similar redesign from a database of past retrofits of plants could assist by providing the starting point for redesign. This could be achieved

by using fuzzy matching. The recorded modifications made, the results and costs of the past retrofit found to be the most similar could then be used when considering the new retrofit.

9.1.2 Fault Diagnosis

When a chemical plant operator realises that there is a problem in the way the plant is operating they try to correct this with a set of actions. An operator could produce a solution to the problem more quickly and more successfully if they have experience of the problem or a similar problem in that plant or other similar plants.

If a database of previous cases of operating problems and their solutions were available then fuzzy matching could be used to find which cases in the database were most like the current predicament (similar to case-based reasoning). The selected cases are then used to aid the operator in their decision on what set of actions to take. This is a different approach to that of expert systems which use rules to infer solutions from the characteristics of the problem.

9.2 Conclusion

The results obtained using fuzzy matching for capital cost estimation are very promising. The accuracy obtained is better than that obtained using both regression equations derived from the CS data and traditional pre-design methods. The accuracy obtained is within the band that a professional estimator would expect to achieve in the early stages of plant design.

The cost of a study estimate in 1989 was between $11,000 and $50,000. This is the class of estimate that has an expected accuracy which is comparable to the best average error achieved by fuzzy matching of ±39%. Since fuzzy matching involves merely entering data into a computer program, it would be a lot cheaper to use when estimating in the early stages of design.

The accuracy improves as a match is found using more plant attributes. However, some attributes are more important than others, for example capacity seems more influential than pressure. Weighting of the attributes to reflect their importance as predictors of capital cost also improves accuracy. A ramp shaped characteristic function used to quantify match closeness produced the best results.

An advantage of fuzzy matching is that there is no need to analyse a set of data to develop correlation equations for the capital cost, nor to decide how to group the data in order to refine the equation for particular types of process. The fuzzy technique may be thought of as finding a specific capital cost formula for each process.

As Uppal and Van Gool (1992) state, a method that uses the information available in the idea stage of capital cost estimation as well as the details that become available in the late stages of capital cost estimation would be valuable as it would avoid the discontinuities of cost estimating at different stages. Fuzzy matching can handle the inclusion of new plant specifications when they are found during the development of the plant. With the new plant specification included as an extra attribute in the fuzzy matching. Obviously using a larger number of different attributes will increase the time taken to find the optimum set-up for the fuzzy matcher, but this will be compensated for by the increased accuracy.

Furthermore, the total of the match values for the best match plant quantifies its closeness to the target process and therefore provides an indication of the confidence in the accuracy of the estimate.

Once fuzzy matching has selected the existing plant most like the new plant then its capital cost is taken as the capital cost estimate for the new plant. However, an expert would modify the capital cost of the existing plant in line with any significant process differences in order to make it a better estimate for the capital cost of the proposed plant. Clearly the capacity is an important determinant of cost, and there is a widely accepted rule for making adjustments, that is the exponent rule. Williams (1947) well known 0.6 rule is commonly used to estimate the capital cost of a new plant that uses the same process as an existing plant, but has a different capacity. When the best match has a different capacity from that of the new process, this rule was used to adjust the capital cost of the existing plant. However, this technique was found to make the estimates worse when tried for the CS data plants.

Fuzzy matching is eminently suitable for computer implementation. Fuzzy matching works with a conventional database of actual costs and process parameters. All the usual database functions are available: new data is simply added and out of date or incorrect data can be deleted or revised. Such changes do not require re-correlation of equations or re-plotting of graphs, as would be the case with existing pre-design methods. The method requires minimal estimating effort, once historic plant cost data is to hand.

Fuzzy matching makes no assumptions about relationships between capital cost and process parameters and no arbitrary decisions, apart from the number of functional units, must be made about the plant, it merely mirrors the professional estimators technique of using past experience. Therefore, fuzzy estimates are more easily justified. Fuzzy matching is consistent with our intuitive view of the influence of plant attributes on the capital cost.

If its potential is realised, fuzzy matching could be an important new method for capital cost estimation in the initial stages of process design

REFERENCES

Aamodt, A. and E. Plaza (1994). Case-based reasoning: foundational issues, methodological variations, and system approaches. *AI Communications*, 7(1), 39-59.

Allen, D. H. and R. C. Page (1975). Revised technique for pre-design cost estimating. *Chemical Engineering*, 82(5), 142-150.

Ashworth, G. (1988). ELSIE plans her budget. *Building*, 253, 60-61.

Babinec, F. and M. Dohnal (1981). Fuzzy reliability of industrial processes. In *Proceedings of 6th International Conference on Production Research*, Novi Sad, Yugoslavia.

Bagley, J. D. (1967). *The Behaviour of Adaptive Systems Which Employ Genetic and Correlation Algorithms*. PhD Thesis, University of Michigan, Ann Arbor MI.

Baldwin, A. and S. A. Oteifa (1991). Application of protocol analysis in the development of knowledge based systems for contractors' resource based estimating. In *Proceedings of 2nd International Conference on the Application of Artificial Intelligence to Civil Engineering*, Oxford, 3-5 September 1991, 65-70.

Borland International (1990). *Paradox Relational Database v. 3.5*. Borland International, Scotts Valley.

Brass, J., A. M. Gerrard and D. Peel (1994). Estimating Vessel Costs via neural networks. In *Proceedings of 13th International Cost Engineering Congress*, London, 9-12 October 1994, CE.4.1-CE.4.13.

Bridgwater, A. V. (1974). The functional unit approach to rapid cost estimation. *The Cost Engineer*, 13(5).

Bridgwater, A. V. (1978). Development of step counting methods for capital cost estimating. In *Proceedings of 5th International Cost Engineering Congress*, 47-54.

Bridgwater, A. V. (1981). Step counting methods for preliminary capital cost estimating. *Cost Engineering*, 23(5), 293-302.

Bridgwater, A. V. (1994). Short cut methods of estimating capital costs. *An I.Chem.E. Course on Techniques of Cost Estimation*, Darlington, 27-29 November 1994.

Buckley, J. J. (1987). The fuzzy mathematics of finance. *Fuzzy Sets and Systems*, 21(3), 257-273.

Chem Systems International Ltd (1988). *Chemical Process Economics*. Chem Systems, London.

Chilton, C. H. (1949). Cost data correlated. *Chemical Engineering*, 56(6), 97-106.

Chilton, C. H. (1950). 'Six tenths factor' applies to complete plant costs. *Chemical Engineering*, 57(4), 112-114.

Chilton, C. H. (1960). *Cost Engineering in the Process Industries*. McGraw-Hill, New York, 121-180.

Chung, P. W. H. and R. Inder (1992). Handling uncertainty in accessing petroleum exploration data. *Revue De L'Institut Francais Du Petrole*, 47(3), 305-314.

Coates, D. (1993). *Regression Modelling*. Loughborough University Staff Training and Development.

Corne, D., H. Fang and C. Mellish (1993). Solving the modular exam scheduling problem with genetic algorithms. In *Proceedings of 6th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, Edinburgh, 1-4 June 1993, 370-373.

Cran J. (1981). Improved factored method gives better preliminary cost estimates. *Chemical Engineering*, 88(7), 65-79.

DeCicco, R. W. (1968). Economics evaluation of research projects - by computer. *Chemical Engineering*, 75(12), 84-90.

Dohnal, M. and K. Hartman (1981). Fuzzy models of unit operations and chemical plants. In *Proceedings of 7th International congress CHISA 81*, Prague.

Drayer, D. E. (1970). How to estimate plant cost-capacity relationship. *Petro/Chem Engineer*, May, 10-38.

Edwards, D. W. (1991). Chemical process cost estimate statistics. *The Cost Engineer* 29(6), 4-7.

Edwards, D. W. and G. J. Petley (1994). Chemical plant capital cost estimation using fuzzy matching. In *Proceedings of Escape 4*, Dublin, 28-30 March 1994, 391-398.

Feigenbaum, E. A., B. Buchanan and J. Lederberg (1971). On generating and problem solving: A case study using the DENDRAL program. *Machine Intelligence*, 6, 165-190.

Gaines B. R. (1976). *International Journal Man-Machines Studies*, 8, 623-668.

Garrett, D. E.(1989). *Chemical Engineering Economics*. Van Nostrand Reinhold, New York.

Gerrard, A. M. (1991). The use and availability of computer programs for teaching progress economics. In *Proceedings I.Chem.E. PENCE and Education Subject Groups Meeting*, Swansea, 12 April 1991.

Gerrard, A. M. (1994). International cost indices. *The Chemical Engineer*, January 18th, 18.

Gerrard, A. M. and M. J. Davenport (1993). The preliminary costing of environmental projects. In *Proceedings of 1993 I.Chem.E. Research Event*, 2, 451-453.

Gerrard, A. M., J. Brass and D. Peel (1994). Using neural nets to cost chemical plants. *I.Chem.E. Symposium Series 133*, 475-478.

Gore, W. H. (1969). M.Sc. Thesis, University of Aston.

Greffioz, J., A. J. Olver and P. Schirmer (1993). TOPEX: An expert System for Estimating and Analysing the Operating Costs of Oil and Gas Production Facilities. In *Proceedings of Asia Pacific Oil and Gas Conference*, Singapore, 8-10 February, 61-68.

Guthrie, K. M. (1969). Data and techniques for preliminary capital cost estimating. *Chemical Engineering*, 76(6), 114-142.

Guthrie, K. M. (1970). Capital and operating costs for 54 chemical processes. *Chemical Engineering*, 77(12), 140-156.

Guthrie, K. M. (1974). *Process plant estimating, evaluation and control.* Craftsman book company of America, California.

Hakata, T. and J. Masuda (1990). Fuzzy control of cooling systems utilising heat storage. In *Proceedings of IIZUKA 1990*, 77-80.

Hand, W. E. (1958). From flow sheet to cost estimate. *Petroleum Refiner*, 37(9), 331-334.

Haselbarth, J. E. (1967). Updated investment costs for 60 types of chemical plants. *Chemical Engineering*, 74(December 4th), 214-215.

Henkin, S. J. and M. C. Harrison (1988). An analysis of four uncertainty calculi, *IEEE Transactions on Systems Man. and Cybernetics*, 18(5), 700-714.

Hennessy, D. and D. Hinkle (1992). Applying case-based reasoning to autoclave loading. *IEEE Expert*, 7(5), 21-26.

Hill, R. D. (1956). What petrochemical plants cost. *Petroleum Refiner*, 35(8), 106-110.

Hirsch, J. H. and E. M. Glazier (1960). Estimating plant investment costs. *Chemical Engineering Progress*, 56(12), 37-43.

Holland, F. A., F. A. Watson and J. K. Wilkinson (1974). Manufacturing costs and how to estimate them. *Chemical Engineering*, 81(8), 91-96.

Holland, J. H. (1962). Outline for a logical theory of adaptive systems. *Journal of the Association for Computing Machinery*, 3, 297-314.

Hopfield, J. J. (1982). Neural Networks and physical systems with emergent collective computational abilities. In *Proceedings of the National Academy of Sciences*, 79, 2554-2558.

Humphreys, K. K. and P. Wellman. (1987). *Basic cost engineering*. Marcel Dekker, New York.

I.Chem.E. and Association Cost Engineers (1988). *A Guide to Capital Cost Estimating*. I.Chem.E., Rugby, UK.

Jelen, F. C. and J. H. Black (1983). *Cost and Optimization Engineering*. McGraw-Hill, New York.

Kharbanda, O. P. (1979). *Process Plant Estimating, Evaluation and Control*. Craftsman Book Co., California.

Kharbanda, O. P. and E. A. Stallworthy (1988). *Capital Cost Estimating for the process industries*. Butterworth, Sevenoaks, UK.

Kirk-Othmer, (1992). *Encyclopaedia of chemical technology*. John Wiley, New York.

Klumpar, I. V., R. F. Brown and J. W. Fromme (1988). Rapid capital estimation based on process modules. *Process Economics International*, 7, 5-10.

Kolodner, J. L. (1983a). Maintaining organisation in a dynamic long-term memory. *Cognitive Science Society*, 7(4), 243-280.

Kolodner, J. L. (1983b). Reconstructive memory: a computer model. *Cognitive Science Society*, 7(4), 281-328.

Lang, H. J. (1947a). Engineering approach to preliminary cost estimates. *Chemical Engineering*, 54(9), 130-133.

Lang, H. J. (1947b). Cost relationships in preliminary cost estimation. *Chemical Engineering*, 54(10), 117-121.

Lang, H. J. (1948). Simplified approach to preliminary cost estimates. *Chemical Engineering*, 55(6), 112-113.

Leigh, R. and M. Wetton (1983). Thinking clearly with fuzzy logic. *Process Engineering*, 64(11), 36-37.

Liddle, C. J. (1978). Estimation or Procrastination? *Engineering and Process Economics*, 3, 165.

Liddle, C. J. (1992). Capital cost estimation. *An I.Chem.E. Course on Techniques of Cost Estimation*, Stockton on Tees, 6-8 December 1992.

Liddle, C. J. and A. M. Gerrard (1975). *The Application of Computers to Capital Cost Estimation*. I.Chem.E., London.

Lin S., Q. Sun and X. Yang (1990). Fuzzy Economic assessment for offshore oil fields. China Ocean Engineering, 4(1), 97-108.

McCulloch, W. S. and W. Pitts (1943). A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.

Miller, C. A. (1965). New cost factors give quick, accurate estimates. *Chemical Engineering*, 72(19), 226-236.

Minsky, M. and S. Papert (1969). *Perceptrons*. MIT Press, Cambridge MA.

National Algorithms Group Ltd. (1986). *NAG FORTRAN Workstation Library*. NAG, Oxford, UK.

Noble, J. S. and J. M. A. Tanchoco (1990). Concurrent design and economic justification in developing a product. In *Proceedings of International Journal of Production Research*, 28(7), 1225-1238.

Parmar, J. C. (1986). *A Method for Computer-Aided Hazard Identification of Process Plants*. Ph.D. Thesis, Loughborough University.

Pearce, D. (1989). Statistical/heuristic approach to estimating mould costs. In *Proceedings of Annual Technical Conference - Society of Plastic Engineers*, New York, 1-4 May 1989, 364-366.

Petley, G. J. and D. W. Edwards (1993). Data-based cost estimation for chemical process plants. In *Proceedings of I.Chem.E. Research Event*, 2, 714-716.

Petley, G. J., D. W. Edwards and P. W. H. Chung, (1993). In *Proceedings of 6th International Conference on Industrial and Engineering Applications of Artificial Intelligence and Expert Systems*, Edinburgh, 1-4 June 1993, 480-483.

Remer, D. S. and L. H. Chai (1990). Estimate costs of scaled-up process plants. *Chemical Engineering*, 97(4), 138-175.

Riesbeck, C. K. and R. C. Schank (1989). *Inside Case-Based Reasoning*. Lawrence Erlbaum Associates, New Jersey.

Rodriguez, A. L. M. and F. J. M. Coronel (1992). A better way to price plants. *Chemical Engineering*, 99(3), 124-127.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organisation in the brain. *Psychological Review*, 65, 386-408.

Shortliffe, E. H. and B. G. Buchanan (1975). A model of inexact Reasoning in Medicine, *Mathematical Biosciences*, 23, 351-379.

Sittig, M. (1967). *Organic chemical process encyclopaedia*. Noyes Development Corp., New Jersey. .

Spiegel, M. R. (1972). *Theory and Problems of Statistics*. McGraw-Hill, New York.

Stallworthy, E. A. (1970). The viewpoint of a large chemical manufacturing company. *The Chemical Engineer*, June, 182-189.

Taylor, J. H. (1977). The 'process step scoring' method for making quick capital cost estimates. *Engineering and Process Economics*, 2(4), 259-267.

Timms S. R. (1980). M. Phil. Thesis, Aston University.

Tolson K. W. and J. T. Sommerfeld (1990). Chemical plant costs from capacity. *Cost Engineering*, 32(8), 17-21.

Turing, A. M. (1947). Computing machinery and intelligence. *Mind*, 59, 433-460.

Turunen, I. (1984). Modern methods for cost estimation and decision-making in preliminary plant design. *Acta Polytechnica Scandinavica, Chemical Technology and Metallurgy Series*, 155.

Turunen, I., M. Jarvelainen and M. Dohnal (1984). Fuzzy approach to factorial cost estimation of chemical plants. *Engineering Costs and Production Economics*, 7, 279-292.

Ullmann (1988). *Encyclopaedia of industrial chemistry*. VCH., Weinheim.

Uppal, K. B. and H. V. Gool (1992). R&D Phase; Capital cost estimating. *Transactions - American Association of Cost Engineers*, A.4.1-A.4.3.

Viola, J. L. (1981). Estimate capital costs via a new, shortcut method. *Chemical Engineering*, 88(7), 80-86.

Ward, T. J. (1984). Predesign estimating of plant capital costs. *Chemical Engineering*, 91(9), 121-124.

Wessel, H. E. (1953) How to estimate costs in a hurry. *Chemical Engineering*, 60(1), 168-200.

Williams, R. (1947a) Standardising cost data on process equipment. *Chemical Engineering*, 54(6), 102-103.

Williams, R. (1947b) 'Six-tenths factor' aids in approximating costs. *Chemical Engineering*, 54(12), 124-125.

Wilson, G. T. (1971). Capital investment for chemical plants. *British Chemical Engineering and Process Technology*, 16(10), 931-934.

Wood, T. L. and S. J. Culley (1993). The utilisation of fuzzy set theory in standard component selection. *6th International Industrial and Engineering Applications of Artificial Intelligence and Expert Systems Conference*, 57-60.

Zadeh, L. A. (1965). Fuzzy Sets. *Information Control*, 8, 338-353.

Zadeh, L. A. (1975). *Fuzzy Sets and Their Applications*. Academic Press Inc., London.

Zevnik, F. C. and R. L. Buchanan (1963). Generalised correlation of process investment. *Chemical Engineering Progress*, 59(2), 70-77.

Zimmerman, O. T. (1965). Process plant costs. *Cost Engineering*, 10(3), 14-19.

APPENDIX

Listing of Fuzzy Matching Program

C Fuzzy Matching Program

```
C Define the number of chemical plants used when fuzzy matching (NoP)
      PARAMETER (NoP=90)

C Define the number of attributes in the data file (NoAF)
      PARAMETER (NoAF=23)

C Define the number of attributes used when fuzzy matching (NoAU)
      PARAMETER (NoAU=5)

C Define which attributes are to be used (n*)
      PARAMETER (n1=9, n2=14, n3=12, n4=13, n5=6)

C Define the size of the increment for shape parameter combinations (SI)
      PARAMETER (SI=0.2)

C Define the maximum value for a shape parameter (SM)
      PARAMETER (SM=2)

C Define the size of the increment for weight combinations (WI)
      PARAMETER (WI=0.33)

C Define the maximum value for a weight (WM)
      PARAMETER (WM=0.99)

C Define the membership function used (MF) (1 - FLAT, 2 - RAMP, 3 - CURVE)
      PARAMETER (MF=2)

C Set up the arrays and parameters
      REAL TotS, TotE
      DOUBLE PRECISION SetSA, SetEA
      DOUBLE PRECISION w(NoAF), d(NoP,NoAF)
      DIMENSION  MP(NoP,NoP), PEr(NoP)
      DIMENSION  targ(NoAF), rang(NoAF), p(NoAF)
      DIMENSION  SEr(NoP), EEr(NoP)

C Read the data from an ASCII data file and into an array
      OPEN (UNIT=10,
     *FILE='CSdata.txt',
     *STATUS='OLD')
```

```
    WRITE (6,637) CHAR(13)
637  FORMAT (/' Reading data file, please wait !',A1,/)

    i = 1

70   READ (10,*,ERR=60) ( d(i,j), j=1,NoAF)

    i = i + 1

C Check that the data file has the expected number of records
    IF (i.LE.NoP) GO TO 70

    PRINT*,'Only ',NoP,' data points allowed'
    GOTO 99

60   PRINT*,'Somethings Wrong'

99   CLOSE (unit=10)

C Set up some initial variable values
    NoLSA=1
    NoCom=0
    BSA=500
    BEA=500

C Define the weights
    DO w1=0,WM,WI
     w(1)=w1
    DO w2=0,WM,WI
     w(2)=w2
    DO w3=0,WM,WI
     w(3)=w3
    DO w4=0,WM,WI
     w(4)=w4
    DO w5=0,WM,WI
     w(5)=w5

C Define the shape parameters
    DO sp1=0,SM,SI
    DO sp2=0,SM,SI
    DO sp3=0,SM,SI
    DO sp4=0,SM,SI
    DO sp5=0,SM,SI
```

```
C Initialise the variable values for the set of shape parameters and weightings
      NoCom=NoCom+1
      TotS=0
      TotE=0
      NoWBM=NoP

C For each plant find which of the other plants is the best match when Fuzzy
Matching
      DO k=1,NoP

C Initialise variable values for the plant
            NoBM=1
            BMV=0
            MP(k,NoBM)=0
            PEr(k)=0

C Store the plants attribute values, the target values
            targ(1)=d(k,n1)
            targ(2)=d(k,n2)
            targ(3)=d(k,n3)
            targ(4)=d(k,n4)
            targ(5)=d(k,n5)

C Set up the range of the membership function using the shape parameters
            rang(1)=(sp1)*targ(1)
            rang(2)=(sp2)*targ(2)
            rang(3)=(sp3)*targ(3)
            rang(4)=(sp4)*targ(4)
            rang(5)=(sp5)*targ(5)


C Calculate the match value for every other process
         DO m=1,NoP

C Check that you are not matching the same processes
            IF (m.NE.k) THEN

C Initialise the variable values for the plant being matched
            PMV=0
```

```
C Store the attribute values for the plant being matched to the target values
        p(1)=d(m,n1)
        p(2)=d(m,n2)
        p(3)=d(m,n3)
        p(4)=d(m,n4)
        p(5)=d(m,n5)

C Calculate the match value for each attribute
        DO n=1,NoAU

C Calculate match value for Curve membership function
          IF (MF.EQ.3) THEN
            IF (targ(n).EQ.p(n)) THEN
           AMV=1.0
          ELSE
              IF (targ(n).GT.p(n)) THEN
               IF (p(n).LE.(targ(n)-rang(n))) THEN
                AMV=0
               ELSE
              IF ( ( p(n).GT.(targ(n)-rang(n)))
   *.AND. ( p(n).LE.(targ(n)-((rang(n))/2)) ) ) THEN
                  AMV=2*(((p(n)-targ(n)+rang(n) )/ rang(n) )**2)
                ELSE
                    AMV=1-2*(((p(n)-targ(n))/ rang(n) )**2)
                END IF
                 END IF
                ELSE
               IF ( p(n) .GE. (targ(n)+rang(n)) ) THEN
                   AMV=0
               ELSE
                   IF ( (p(n).GT.(targ(n))).AND.
   *(p(n).LE.( targ(n)+( (rang(n))/2 ) ) ) ) THEN
                  AMV=1-2*( ( (p(n)-targ(n))/ rang(n) )**2)
                ELSE
                 AMV=2*(((p(n)-targ(n)-rang(n))/rang(n))**2)
                END IF
             END IF
            END IF
           END IF
          ELSE
```

```
C Calculate match value for Ramp membership function
            IF (MF.EQ.2) THEN
             IF (targ(n).EQ.p(n)) THEN
              AMV=1.0
             ELSE
              IF (targ(n).GT.p(n)) THEN
               IF (p(n).LT.(targ(n)-rang(n))) THEN
                AMV=0
               ELSE
                AMV=(p(n)/rang(n))+((rang(n)-targ(n))/rang(n))
               END IF
              ELSE
               IF (p(n).GT.(targ(n)+rang(n))) THEN
                AMV=0
               ELSE
                AMV=((-1)*p(n))/(rang(n))+((rang(n)+targ(n))/rang(n))
               END IF
              END IF
             END IF
            ELSE

C Calculate match value for Flat membership function
             IF ((p(n).GE.(targ(n)-rang(n))).AND.(p(n).LE.(targ(n)+rang(n)))) THEN
              AMV=1.0
             ELSE
              AMV=0
             END IF
            END IF
           END IF

C Calculate total match value
           PMV=PMV+(w(n)*AMV)
          END DO

C Check if total match value is highest and record best match
         IF (PMV.GT.BMV) THEN
          NoBM=1
          BMV=PMV
          MP(k,NoBM)=m
         ELSE
```

```
C Calculate the average error when two or more best matches
            IF (PMV.EQ.BMV) THEN
              NoBM=NoBM+1
              MP(k,NoBM)=m
            END IF
          END IF
        END IF
      END DO

C Initialise the variable values for the errors

      SEr(k)=0
      EEr(k)=0

C Calculate average error of best matches for new plant
      DO l=1,NoBM

C Calculate standard estimate error
        PEr(k)=((d(MP(k,l),8)-d(k,8))/d(k,8))*100
        SEr(k)=((SEr(k)*(l-1))+abs(PEr(k)))/l

C Calculate equivalent estimate error
        IF (PEr(k).LT.0) THEN
          EER(k)=((EER(k)*(l-1)) + ((100*(-1)*PEr(k))/(100+PEr(k))))/l
        ELSE
          EER(k)=((EER(k)*(l-1))+PEr(k))/l
        END IF
      END DO

C Count the number of plants without any best match
      IF (BMV.EQ.0) THEN
          NoWBM=NoWBM-1
      END IF

C Total the error for estimates
      TotS=TotS+SEr(k)
      TotE=TotE+EEr(k)
    END DO
```

```
C Calculate the average error for combination
      IF (NoWBM.EQ.0) THEN
        SetSA=1000
        SetEA=1000
      ELSE
        SetSA=TotS/(NoWBM)
        SetEA=TotE/(NoWBM)
      END IF

C Save the weightings and shape parameters for lowest ASEE
      IF (SetSA.LT.BSA) THEN
          BSA=SetSA
          bws1=w1
          bws2=w2
          bws3=w3
          bws4=w4
          bws5=w5
          bsps1=sp1
          bsps2=sp2
          bsps3=sp3
          bsps4=sp4
          bsps5=sp5
      ELSE

C Count the number of combinations that produce the lowest ASEE
        IF (SetSA.EQ.BSA) THEN
          NoLSA=NoLSA+1
        END IF
      END IF

C Save the weightings and shape parameters for lowest AEEE
      IF (SetEA.LT.BEA) THEN
        BEA=SetEA
          bwe1=w1
          bwe2=w2
          bwe3=w3
          bwe4=w4
          bwe5=w5
          bspe1=sp1
          bspe2=sp2
          bspe3=sp3
          bspe4=sp4
          bspe5=sp5
```

```
      ELSE

C Count the number of combinations that produce the lowest AEEE
        IF (SetEA.EQ.BEA) THEN
          NoLEA=NoLEA+1
        END IF
       END IF

      END DO
      END DO
      END DO
      END DO
      END DO
      END DO
      END DO
      END DO
      END DO
      END DO

C Write to file the results of fuzzy matching

      OPEN(7, FILE='result.txt',
     *STATUS='new', FORM='FORMATTED')

      WRITE(7,90) NoCom, NoWBM, NoBM, NoLSA,
     *BSA, bws1, bws2, bws3, bws4, bws5, bsps1, bsps2, bsps3, bsps4, bsps5,
     *BEA, bwe1, bwe2, bwe3, bwe4, bwe5, bspe1, bspe2, bspe3, bspe4, bspe5

90    FORMAT(I7, 3(',',F3.0), ',',F7.3, 10(',',f6.3), ',',F7.3, 10(',',f6.3))

      CLOSE(7)

C End program
      STOP
      END
```

Variables

NoP	Number of chemical Plants in the data file
NoAF	Number of Attributes in the data File
NoAU	Number of Attributes Used when fuzzy matching
n* (n1, n2, ...)	Numeric position of attribute in data file
SI	Shape parameter Increment
SM	Shape parameter Maximum value
WI	Weight Increment
WM	Weights Maximum value
MF	Membership Function used
NoLSA	Number of different SP and W combinations that have the Lowest Standard error Average
NoLEA	Number of different SP and W combinations that have the Lowest Equivalent error Average
NoCom	Number of different SP and W **Com**binations
BSA	Best Standard error Average
BEA	Best Equivalent error Average
w* (w1, w2, ..)	Weight values
sp* (sp1, ...)	Shape Parameter values
NoBM	Number of Best Matches
BMV	Best Match Value total
PMV	Plants Match Value total
AMV	Attributes Match Value
NoWBM	Number of Plants With Best Match
TotS	**Tot**al Standard error for a set of SP and W
TotE	**Tot**al Equivalent error for a set of SP and W
SetSA	**Set**s Standard error Average
SetEA	**Set**s Equivalent error Average
bws* (bws1,..)	Best weights for standard error
bsps* (bsps1,.)	Best shape parameters for standard error
bwe* (bwe1,..)	Best weights for equivalent error
bspe* (bspe1,.)	Best shape parameters for equivalent error

Arrays

w(a)	**W**eights
MP(a,b)	Position of the **M**atching **P**lant in data file
PEr(a)	**P**ercentage **Er**ror of best match(es)
targ(a)	**Targ**et values for the plant finding a match
d(a,b)	**D**ata in the data file
rang(a)	**Rang**e of the membership function
p(a)	Data for the **p**lant matching
SEr(a)	**S**tandard **Er**ror
EEr(a)	**E**quivalent **Er**ror