A PRACTICAL EVOLUTIONARY METHOD FOR THE MULTI-OBJECTIVE OPTIMISATION OF COMPLEX INTEGRATED ENERGY SYSTEMS INCLUDING VEHICLE DRIVETRAINS

PAR

Adam MOLYNEAUX

M.Sc. in Composite Materials, Imperial College, London, Royaume-Uni
et de nationalité britannique

Lausanne, EPFL
2002

Abstract

Concern for the environment has been steadily growing in recent years, and it is becoming more common to include environmental impact and pollution costs in the design problem along with construction, investment and operating costs. To further complicate matters governmental controls on emissions are still changing and the effect of increased emissions taxes may be critical in choosing a particular design solution.

Thermo-economic and environmental analysis has been used previously at LENI to model investment, operating and pollution costs and aggregate them into a single objective that can be minimised. This has the drawback of requiring multiple optimisations in order to determine the sensitivity of the optimum solution to the presumed pollution costs. In addition, designers usually prefer a choice of *different* technological solutions as well as a clear idea of the trade-offs between multiple objectives without the need to define a common indicator (for example cost). Frequently, the thermodynamic and economic simulation of an energy system is non-linear, disjoint and with multiple local optima making it difficult to optimise with derivative based optimisation methods.

This work presents the development of a multi-modal, multi-objective optimisation tool to respond to this need, and then demonstrates it on two complex problems — the design of a district heating system and the configuration of an advanced vehicle drivetrain.

Multi-objective optimisation techniques aim to find the trade-off between two or more conflicting objectives. For example, if a design must be both efficient and low cost, then a multi-objective optimisation will find a range of solutions between the lowest cost but least efficient solution and the most efficient but most expensive solution, including some solutions that are fairly efficient and reasonable cost. Multi-modal optimisation techniques aim to keep different local optima.

The optimisation tool developed here is the clustering pareto evolutionary algorithm (CPEA). As with other evolutionary algorithms (EAs) it works with a population of solutions, each individual representing a different trade-off between objectives. New solutions are produced using real variable crossover and mutation techniques and the population is ranked and thinned to avoid excessively large populations and maintain convergence pressure. The algorithm uses statistical clustering techniques on the independent variables to keep multiple different local optima simultaneously. The clusters maintain diversity in the population and

identify local optima. In problems with many variables multi dimensional scaling (MDS) is used to reduce the number of variables before clustering.

Applying the CPEA to the problem of designing a district heating system for minimum costs with and without pollution costs, it was possible to repeat previous work in a fraction of the time. For the same effort it was also possible to produce complete trade-off curves for cost and pollution, showing the dramatic change in optimal solution when pollution costs were included. The clustering and in particular the MDS were found to be key factors in the solution of this problem — without them the best overall solution was not found. A three objective problem was solved and the results compared favourably to a combined two objective problem, although convergence was found to be slower.

A parallel version of the CPEA was also applied to the optimisation of a vehicle drivetrain simulation with respect to performance, emissions and costs over a test cycle. The multi-modal nature of the CPEA allowed the simultaneous solution of multiple hybrid and conventional solutions at no extra cost and improving overall convergence. The pollution costs calculated using the same level of taxation as in the district heating problem were found to be of little influence compared to the operating and investment costs, suggesting that pollution costs from the energy domain are unlikely to promote hybrid vehicle technology.

Résumé

Les questions environementales ont pris, ces dernières années, de plus en plus d'importance, c'est pourquoi il est devenu courant de considérer en plus des coûts d'investissement et d'opération, ceux liés à la pollution. Par ailleurs les réglementations gouvernementales en matière d'émissions se renforcent constamment et conduisent à une importance croissante de ces aspects dans les prises de décision économics.

Les techniques d'analyse thermoéconomique et environnementale sont en constant développement au Laboratoire d'Energétique Industrielle avec la modélisation des coûts de pollution, et leur prise en compte essentiellement jusqu'alors dans le cadre d'une seule fonction objectif. Cette méthode a le désavantage de nécessiter plusieurs optimisations afin de déterminer la sensibilité de la meilleuse solution aux changements des coûts de pollution. En plus, les ingénieurs préfèrent avoir un choix de solutions techniques différentes ainsi qu'une idée claire du compromis entre plusieurs objectifs sans nécessairement recourir à un indicateur commun (par exemple coût).

Cette thèse traite du développement d'un outil d'optimisation multi-modal, multi-objectifs pour répondre á ces besoins, puis de son application à deux problèmes complexes—la conception de systèmes de chauffage urbain et la configuration de systèmes d'entraînement de voitures.

Les algorithmes multi-objectifs vise à déterminer le compromis entre deux (ou plusieurs) fonctions objectifs. Par exemple, si un produit doit être bon marché et efficace, une optimisation multi-objectif vise à trouver toutes les solutions entre une solution très bon marché mais peu performante et une solution très chère mais très performante. Les algorithmes multi-modaux en plus permettent de trouver plusieurs solutions différentes.

L'outil developpé est nommé le "clustering pareto evolutionary algorithm" (CPEA). Comme les autres algorithmes évolutifs l'algorithme part d'une population de solutions, chaque individu (solution) représentant un compromis entre les différents objectifs. Les nouveaux individus sont produits avec des techniques de croisement et de mutation et lorsque la population devient trop grande les individus qui sont considérés comme moins nécessaires sont progressivement éliminés.

L'algorithme utilise une technique statistique de groupement sur la base des variables indépendantes afin de garder plusieurs optima locaux. Cette technique a comme effet complémentaire de garder la diversité de la

population.

Dans les problèmes avec beaucoup de variables le nombre de variables déterminantes est réduit sur la base d'une technique de "multi-dimensional scaling" (MDS).

En appliquant le CPEA au problème de minimisation des coûts (avec ou sans coût de pollution) d'un réseau de chauffage urbain, il a été possible de réduire substantiellement le temps de résolution. Pour le même travail, il a été possible de produire les courbes de compromis pour mettre en relation les coûts d'investissement/operation avec ceux de pollution. Le groupement statistique et en particulier le MDS étaient essentiels pour résoudre ce problème rapidement. Un problème avec trois objectives a aussi été considéré avec succès, mais le temps de convergence a été alors très important.

Une version parallèle du CPEA a été appliquée au problème d'optimisation des composants d'un système d'entraînment de vehicules afin de réduire les émissions et les coûts. La nature multi-modale du CPEA à permis la résolution simultanée de plusieurs configurations de systèmes d'entraînement de vehicules dont les systèmes hybrides. Les coûts de pollution ont été calculés comme dans le cas du chauffage urbain, mais ils se sont avérés n'avoir que peu d'influence par rapport aux coûts d'investissment et d'utilisation.

Acknowledgements

I would like to express my thanks to thank Daniel Favrat for initially inviting me to come to Lausanne, and to my brother Alex for creating the opportunity. I would also like to thank David Wallace, for encouragement when needed as well as the chance to visit Boston, meet lots of friendly people, get an great recipe for rotten cabbage and see Star Wars (oh, and work in a lab with no windows).

I would like to thank all of my colleagues at LENI for their support and comments, and for interesting discussions in the cafe, some of which were work oriented.

In addition I would like to thank all the people who have made Lausanne fun, starting with my first contact with Vinicio Curti. Special thanks go to Geoff Leyland, both for being a friend and for helping with the thesis.

Last but by no means least I would like to thank my wife Brigitte, and all of my family and friends who have had to put up with the writing of this thesis for the past year.

Contents

List of Figures

List of Tables

Abbreviations and Symbols

Abbreviations

EA	Evolutionary Algorithm
EP	Evolutionary Programming
ESS	Sum of Squared Errors
FCM	Fuzzy C-Means Clustering
GA	Genetic Algorithm
MDS	Multi-Dimensional Scaling
MNV	Mutual Neighbourhood Value
MOEA	Multi-Objective EA
MOO	Multi-Objective Optimiser
MOP	Multi-Objective Problem
NDS	Non-Dominated Set
NDF	Non-Dominated Frontier
POF	Pareto-Optimal Front
SOEA	Single Objective EA
TCS	Touring Club Suiss

Multi-Objective Evolutionary Algorithms

CPEA	Clustering Pareto Evolutionary Algorithm
NPGA	Niched Pareto Genetic Algorithm
NSGA(-II)	Non-dominated Sorting Genetic Algorithm
MOGA	Multi-Objective Genetic Algorithm
MOSGA	Multi-Objective Struggle Genetic Algorithm
PAES	Pareto-Archived Evolutionary Strategy
SGA	Struggle Genetic Algorithm
SPEA	Strength Pareto Evolutionary Algorithm
VEGA	Vector Evaluated Genetic Algorithm

Symbols

Clustering

clustering

$\alpha_A, \alpha_B, \beta, \gamma$	Coefficients for calculating new proximity when merging two clusters in hierarchical clustering
μ_{ij}	Fuzzy memebership of jth data point in the ith cluster
v_i	Centre of fuzzy cluster i
$\vec{c}_i$	Central point of cluster i
$\vec{x}$	Decision variable values of population members
M_T	Parameter to decide nearest neighbours in MNV clustering
$p(R, Q)$	Proximity of two clusters R, Q in hierarchical
r_a, r_b	Constants for the Chui initial centre estimation method
i, j, k	Indices for decision variables, objectives, and generations

Optimisation Parameters

$p_{initial}$	Initial Population size
$n_{clusters}$	Number of Clusters
$p_{max\_cluster\_size}$	Maximum cluster size
$p_{min\_cluster\_size}$	Maximum cluster size
n_{eval}	Maximum Number of evaluations

Vehicle Terminology

CI	Compression Ignition
FC	Fuel Converter
HEV	Hybrid electric vehicle
ICE	Internal Combustion Engine
SI	Spark Ignition
SOC	Battery State of Charge

Chapter 1

Introduction

1.1 Background

1.1.1 Sustainability and the Environment

In recent years more emphasis and concern has been expressed over the damage done to the environment by pollution. It has been speculated that in the next 100 years, temperatures worldwide will increase by the same amount as they have in the last 20,000 years. An increase such as this is extreme and will carry with it significant global implications. Climate warming is directly related to fossil fuel use and resulting emission gases that absorb radiation[44] and contribute to the "greenhouse effect".

The production, conversion and utilization of energy represents a principal source of local, regional and global pollution, and they rely mainly on non-renewable resources. At the same time, industrial and economic process in general are still heavily energy dependent. In the European Union[18] it is estimated that electricity production accounts for 29% of total CO_2 emissions. The industrial sector that includes electricity generation for its own use is responsible for another 17.6%, the transport sector is responsible for 22.4% and the residential sector 19.5%.

Mass-produced consumer goods affect the planet's water, land and air quality through their development, manufacture, use and disposal. Designers should be conscious of these environmental implications from the beginning and take effort in including them in their designs. The (United States) National Research Council estimates that at least 70 percent of the costs associated with product development, manufacture and use are committed during the initial design stages[72]. Therefore, in order to reduce the environmental impact of consumer goods and plan for global sustainability, it is essential to consciously design the complete life-cycle of products early in the design process.

In order to find a good design for a modern, complex energy system it is necessary to consider the system as a whole rather than individual components. Putting together a series of components each of which has been individually optimised will not, in general, in a complex system result in an optimal system. Far from it, as is clear in examples as diverse as cogeneration plants and America's cup yacht designs.

Rather it is necessary to optimise the full system and that requires a methodology that is computationally practical.

1.1.2   Holistic Design

Over the past 20 years powerful computational methods for aiding product design have been developed. Typically, the methods and tools are specialized for different design domains (e.g. solid geometry, surface geometry, mould flow analysis, design for assembly, etc.). These 'traditional' analytical tools are isolated and that makes the exchange of data and evaluation of tradeoffs between different domains difficult and time consuming. However, this inefficiency has been offset by dramatic local improvements in design capabilities and productivity.

In recent years there has been an explosion in analytical frameworks and tools related to environmentally conscious design such as design for recycling and life-cycle inventory assessment. Like the traditional computational design aids, these methods have helped designers improve a specific attribute or set of related design attributes, but they are typically divorced from the larger product system context.

Fundamentally, sustainable product design must be at the system level and will require that the interactions between a diverse range of design characteristics (including economic, performance, environmental and social factors) are understood and appropriately balanced. A framework for holistic, but distributed design modeling, evaluation and optimization is needed.

Holistic design is a multidisciplinary activity involving the cooperation and integration of many designers and tools in a distributed environment. Members of the design team usually have specific areas of expertise and may develop models that are useful, yet proprietary. Frequently economic models, marketing models and engineering models are built by different people in different locations. In order to have a holistic view of the project a decision maker is often faced with the problem of puting these separate analyses together, and this can take some time. If it is necessary to change the engineering design it may be necessary to re-analyse the marketing or economic or life cycle analysis. A means of dealing with distributed models called DOME (Distributed Object Modeling Environment) has been developed at MIT[99] in order to address this problem and allow different models in different locations, and using different interfaces, to be connected via network. This allows the diverse models to be run remotely under the control of a decision maker without the need for intervention by each team, thereby greatly increasing the potential to explore tradeoffs.

It is clear that this kind of technique is necessary for reducing the lead time of designs, and this has been the

framework into which this work is seen to fit.

The introduction of environmental concerns to the design problem is becoming more common, since it no longer suffices to worry about environmental impact afterwards (indeed permission for a plant may only be given if adequate proof can be provided of the environmental impact). Consequently the design problem has become more complicated - it is no longer enough to consider construction, investment and operating costs. Now it is necessary to consider pollution costs (perhaps even separately for different pollutants) during construction, operation and maybe even decommissioning. To complicate matters further (for the design engineer) governmental control of pollution is still changing, and the impact of increased pollution *taxes* may be of extreme interest in the choice of design, particularly where plants may last more than twenty five years and penalisation of certain types of pollution may bias new technology.

Simulation Model Complexity

Consumer demands are also in most cases not uniform, for example a district heating network will be heavily used in winter but may be used very lightly in summer. A vehicle will have a completely different set of requirements in town and on the highway.

This variation introduces the problem of finding an optimum solution when the load changes periodically (but slowly, hence quasi-static), or continually as in a vehicle. The overall best (*cheapest*) solution is unlikely to be the best solution at any one time period, but is more likely to be a compromise. Many traditional approaches consider a *representative* load for design purposes and solve the optimisation problem for this point with constraints that the maximum load can be achieved. This approach may miss important and interesting solutions, but has been necessary to reduce the problem to a manageable size.

The components used and the fashion in which they are connected frequently result in a system that is nonlinear and noncontiguous in the design space. The complexity and nonlinearity of these systems mean that optimisation requires many simulation evaluations, requiring heavy processing power (or long solution times) to find optimal solutions. In the case of the waste incineration cogeneration model[68,70,71,69] that included a time dependency (stocking of waste for the incinerator), several days of CRAY supercomputer time were required to produce one optimised solution, for one combination of pollution costs.

1.1.3 Thermoeconomic and Environmental Analysis

In previous projects the Laboratoire d'Energétique Industrielle (LENI) has looked at the optimisation of complex systems such as a distributed heating network[21,22], a cogeneration power plant[76,77] and a waste incineration plant[68,70,71,69]. The methodology used has been based on thermoeconomic analysis as developed by[94],[95],and more recently[37] and[38], as well as numerous others, and conists of combining thermodynamic,

economic and environmental analyses by *internalising* the costs of equipment, operation, economic fore-
casting and environmental impact.

The approach adopted was to consider a model *superconfiguration* containing all possible components and
designs and then to use an optimisation to choose both configuration (size and type of component) and
operating parameters for the chosen configuration. The chosen configuration can be used to determine
capital costs (including investment, maintenance and construction), and the model used to simulate the
thermodynamic behaviour of the chosen configuration. The resulting predictions for fuel use and pollution
(and to some extent maintenance and redundancy - components tend to need maintenaince or replacement
at a rate determined by there use) can then be used to calculate operating costs (for example fuel costs),
pollution costs and acceptability (does the design meet the requirements).

These costs can be aggregated as shown in (1.1).

$$C_{total} = \sum C_{capital} + \sum C_{operating} + \sum C_{pollution} \qquad (1.1)$$

where $C_{capital}$ $C_{operating}$ $C_{pollution}$ are the capital, operating and pollution costs respectively, and C_{total} is the
total operating cost.

While capital costs are generally well known, and operating costs can be estimated with a good degree of
accuracy based on simulation results, pollution costs are heavily dependent on pollution factors (see (1.2)).

$$C_{pollution}^i = c_{pollution}^i f^i p^i \qquad (1.2)$$

where :

$C_{pollution}^i$	is the total cost of emission of pollutant *i*.
$c_{pollution}^i$	is the specific cost per unit of pollution produced expressed as a cost in this case.
f^i	is a pollution factor that modifies the specific cost of the pollution taking into account the state of the environment and the current emission levels.
p^i	is a mesure of pollution representing the intensity of the emissions, in line with the definition of specific cost $c_{pollution}^i$.

In order to find the best solution the total cost, C_{total} may be then minimised subject to the thermodynamic
(physical) constraints.

1.1.4 Drawbacks of Aggregated Environomic Analysis

While the internalisation of costs allows the designer to consider pollution and thermodynamic considera-
tions at the same time as economic concerns, it also introduces a lot of uncertain parameters. For example

the capital costs depend on estimates of the interest rates and other governmental policies, and so will be known only within limits. Operating costs depend on fuel costs that may change. Similarly the pollution costs depend on a unit cost that is subject to change and that at the present time is very uncertain.

In order to deal with these problems a designer must consider many different options in order to understand the sensitivity of the problem to the uncertain parameters. If costs are internalised and aggregated to form an objective function that can be minimised the resultant optimum solution is dependent on the parameters chosen in the aggregation process. The decision maker is then faced with questions such as :

- How sensitive is the optimum solution to changes in pollution costs ?

- At what unit pollution cost does a change in configuration occur ?

- How sensitive is the optimum solution to basic configuration ?

- What effect does utility cost have on the optimum solution ?

- Are the weights chosen to aggregate the objective reasonable ?

In order to answer such questions with a single objective optimisation algorithm many optimisations must be performed, each with a new value for the uncertain parameter or a new aggregation weight, and this proves costly both in time and computational effort.

In addition, decision makers and engineers both prefer a range of solutions from which they can make a decision taking into account criteria that may not easily be introduced quantitatively in an optimisation problem.

In the author's opinion the true problem is a multi-objective problem, and should be posed as such. For example capital costs, operating costs and pollution costs should each be kept separate and dealt with in a multi-objective optimisation. Since local optima may also be of interest both for understanding the problem domain and to provide alternative configurations a multi-modal optimisation is preferable.

Aggregation with environomic techniques can then be performed afterwards in "real time" allowing tradeoffs between different solutions to be evaluated using traditional economic multi-criteria analysis methods.

1.2 Aim of this Work

The aim of this work has been to bring together work developed from previous projects and further develop the new techniques that are needed to design a modern, optimal, integrated energy system. This involved addressing the shortcomings of the existing methods and developing a method for treating complex

"real-world" problems, where it is important to get as much information as possible from each simulation evaluation.

The task has not simply been one of creating a cost and thermodynamic simulation and optimising it. Rather it has been necessary to consider practical aspects such as speed of resolution, simulation precision, and the quality of available input data—how to get more information from each simulation evaluation ?

This has meant :

- Developing a highly efficient multi-modal (to keep more than one solution), evolutionary multi-objective optimisation method, that is practical to use with no need to tune parameters. This allows uncertain data to be kept as a separate objective (for example CO_2 pollution rather than aggregating an uncertain CO_2 cost into the objective.

- Dealing with computationally intensive and hence long simulations, necessitating the use of parallel evaluation techniques to provide optimisation results in a user acceptable time frame.

In order to develop and demonstrate these ideas two real world test cases have been considered:

1. A distributed heating network. This was previously considered in [21] using the internalised environomic approach, so optimum values were known in advance and could be used to validate the methods. This is treated in Chapter 5.

2. A vehicle drive train simulation that considers the optimum choice and size of drive train components necessary to meet a given driving cycle. The choice of motive power and vehicle drivetrain is large and is made up of many individual components, some of which are themselves complex systems. In order to evaluate a proposed vehicle configuration, and compare it to an alternative the performance of the vehicle must be simulated (or tested) through a test cycle which represents the proposed use of the car, and results in a quasi-dynamic system model—a process that is computationally very intensive. This is treated in Chapter 6.

Chapter 2

Optimisation

2.1 Introduction

The goal of optimisation in the real world of integrated energy systems design is not necessarily to find the *global* optimum of a numerical model of the real system, but rather to find several *good*, but *different*, solutions, ideally including the global optimum, that can be presented to a design engineer for discussion. This allows the designer or decision makers to choose a solution taking into account limitations that may exist in the numeric model and uncertainty in the available data, as well as personal preferences and criteria that cannot be included quantitatively in a simulation.

The optimisation problems typically encountered in integrated energy systems design (as well as many other domains) are typically:

- non-linear - both because they include non-linear components such as heat pumps and because of the non-linear nature of the system connectivity;

- non-contiguous - components can be turned on and off, and may have a limited range of operating conditions;

- mixed integer and real - choice of component size may be discrete - corresponding to a manufacturer's range of products - but operating conditions will tend to be continuous real values;

- non-differentiable - frequently systems include iterative solutions that can not be directly differentiated;

- large - the number of variables makes it difficult to visualise results and understand the interactions;

- involve unknown interactions in the sytem - frequently the *best* solution is not known beforehand, so the initial search domain must be large - the simulation is frequently a *black box*;

- possess many local optima - frequently the domain is fairly flat with little difference between the better local optima and the *global* optimum. Consequently more than one solution is of interest, particularly if each solution represents a different design solution.

- inherently multi-objective - for example minimisation of the investment cost, the operating costs and more recently pollution.

This class of problem,MINLP (Mixed Integer Non-Linear Programming), forms the most difficult class of problems to solve[10]. Various techniques exist to resolve this class of problem, but the most promising in terms of robustness are based on genetic or evolutionary algorithms (EAs).

Historically nearly all researchers in the field of EAs have taken natural evolution as a blueprint to design optimisation algorithms as though this is in itself a good thing to do. Recently however, the trend has been towards describing EAs in traditional mathematical terms by modeling the process of selection, crossover, mutation and replacement as probability density functions describing a new set of solutions based on the existing distribution[67,91]. The benefit of this analysis is that proofs of convergence may be found to satisfy mathematicians and improve the credibility of EAs, and more importantly, a whole new range of statistical ideas and tools have become available.

In earlier work at LENI, some success was achieved with a single objective Genetic Algorithm termed *Struggle*[83], developed at MIT and applied to the problems of Olsommer[68], Curti[21] and Pelster[76]. However, the Struggle GA proved to have several drawbacks, as did all of the GA's studied, and these will be discussed later in this section.

Existing multi-objective optimisation algorithms suffer from the need to tune too many parameters, as is the case with many GAs. Nearly all focus on finding a non-dominated front to approximate the global Pareto front, rather than finding multiple "local" non-dominated fronts.

In view of this, and in an attempt to better meet LENI's requirements, the ideas of cluster analysis and multi-dimensional scaling from statistical analysis were brought together with ideas from the EA world and developed into a new multi objective multi-modal algorithm termed the *Clustering Pareto Evolutionary Algorithm* (CPEA).

In this chapter some basic ideas of GAs will be introduced and the original *Struggle* algorithm will be described before reviewing its disadvantages and introducing the new ideas. Multi-objective optimisation will be briefly reviewed and the CPEA, will be presented. The clustering and multi-dimensional scaling techniques, essential for the solution of the real world problems tackled later, will be described and the results of several validation test problems will be presented.

2.2 Genetic Algorithms

2.2.1 Brief Introduction to *Classical* GAs

In this section we briefly examine the basic operation of a GA and then look in more detail at niching algorithms, in particular the Struggle GA, developed at MIT. For a more complete introduction to GAs and GA theory the reader is directed to Fogel[34] or Deb[26].

Genetic algorithms were probably first proposed in the 60's by John Holland while working on machine learning at the University of Michigan, and the term Genetic Algorithm (or GA) became popular after the publication of his book in 1975[47]. In 1989 Goldberg[42] published a book that provided a solid basis for traditional GAs and included many sucessful applications of the GA. In recent years interest has continued to grow with an international journal, regular conferences and much available software. Indeed GAs, and more generally Evolutionary Algorithms (EAs), have become widely used as robust optimisation tools for combinatorial problems and those problems that are not amenable to other forms of solution.

A traditional definition of a GA was given by Koza[57]:

> The *genetic algorithm* is a highly parallel mathematical algorithm that transforms a set (*population*) of individual mathematical objects (typically fixed-length character strings patterned after chromosome strings), each with an associated fitness value, into a new population (i.e., the next **generation**) using operations patterned after the Darwinian principle of reproduction and survival of the fittest and after naturally occurring genetic operations (notably sexual recombination).

An EA for a particular problem can be seen as having the following 5 elements[16]:

1. A representation for potential solutions to the problem.

2. A way to create an initial population of potential solutions.

3. An evaluation function that plays the role of the environment, rating solutions in terms of their "fitness".

4. Genetic operators that alter the composition of children.

5. Values for various parameters that the GA uses (population size, probabilities of applying genetic operators, etc)

A single objective problem can be seen as a black box simulation that maps a vector of decision variables to a real valued objective value. To apply the GA to the problem the decision variables are mapped to a binary

bit string representation as illustrated in Fig. 2.1, and the objective mapped to a *fitness* value that is some function of the objective (scaled or inverted for example, to ensure all values will be positive).

Any constraints on the optimisation problem must be incorporated in the objective function using a penalty function. Typically this transforms the objective function (2.1) to one of the form (2.2). Solutions proposed that are infeasible for some reason may be given a very low fitness to eliminate them from the population or may be *corrected* to be feasible (for example values modified after crossover and mutation in order to produce a viable individual).

$$
\begin{aligned}
&\text{minimise} \qquad \vec{f}(\vec{x}) \\
&\text{subject to} \qquad g_i(\vec{x}) \;\leq\; \qquad\qquad 0, i = 1,\ldots,m, \vec{x} \in \Omega. (2.1)
\end{aligned}
$$

$$
\text{minimise} \quad \vec{f}(\vec{x}) + P_g \tag{2.2}
$$

where P_g is the sum of the penalties for each constraint i:

$$
P_g = \sum_i P_{gi} \tag{2.3}
$$

While many penalty functions are possible, historically logarithmic functions have been used in much of the optimisation work at LENI.

$$
P_{gi} = 1 + \log(1 + |E_i|) \tag{2.4}
$$

where E_i is the error in the constraint i.

A fixed size *population* of potential solutions is maintained and basic GA operators such as selection, crossover and mutation are applied to *evolve* towards an optimum solution. Each *generation* is replaced by a new generation created by applying these operators (this scheme is often referred to as a *simple* GA).

There are many variations on this scheme, typically involving some kind of *elitism* (eg steady state GA) where the best members of the current polulation are retained from one generation to the next. In general these tend to perfrom better than simple generational GAs. The many variations will not be explained here (see Goldberg[42] for example) - rather a typical *steady state* GA will be described in order to understand the ideas behind the Struggle GA, and later the CPEA, both of which have very strong elitism.

The basic scheme of a Steady State GA is shown in Fig. 2.2.

A population size is chosen and initialised randomly*. The main loop then begins, consisting of selection

---

*The initial population may laternatively be produced in some other way that is representative of the search domain
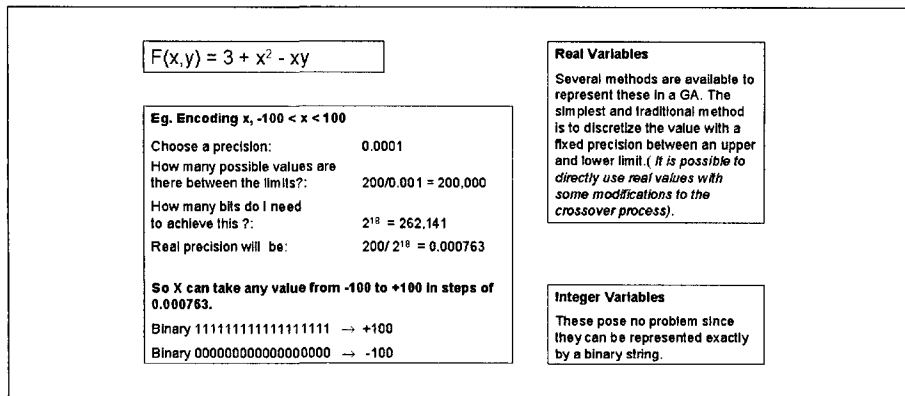
Fig. 2.1: Simple Steady State GA - Binary Encoding

of parents, crossover of the chosen parents, mutation of the resulting offspring and replacement of the new individual into the population. Selection may be performed in a wide variety of fashions but usually involving a bias towards choosing parents with better *fitness* scores. A common method is *proportional selection*, where individuals are chosen to breed in proportion to their fitness values.
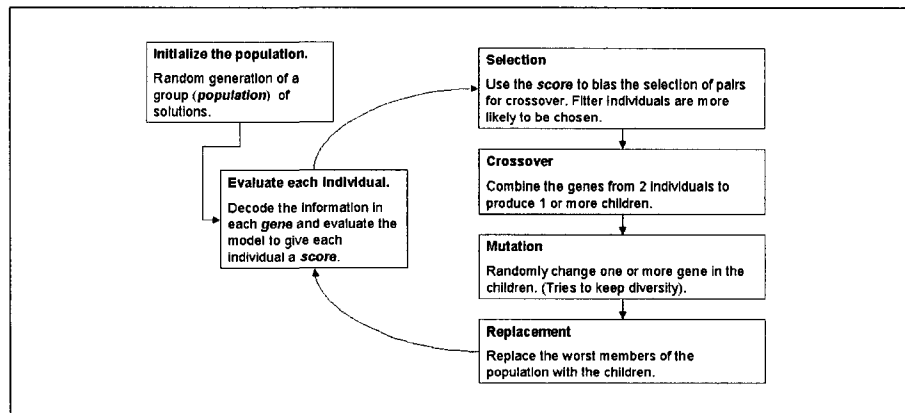
Fig. 2.2: Simple Steady State GA Evolution Scheme

Similarly there are many alternative crossover methods - one commonly used method of crossover for binary representations of solutions - single point crossover - is illustrated in Fig. 2.3.

Mutation is used to increase diversity. In typical binary encoded GA's this is achieved by *flipping* each bit of the binary string with a small probability.
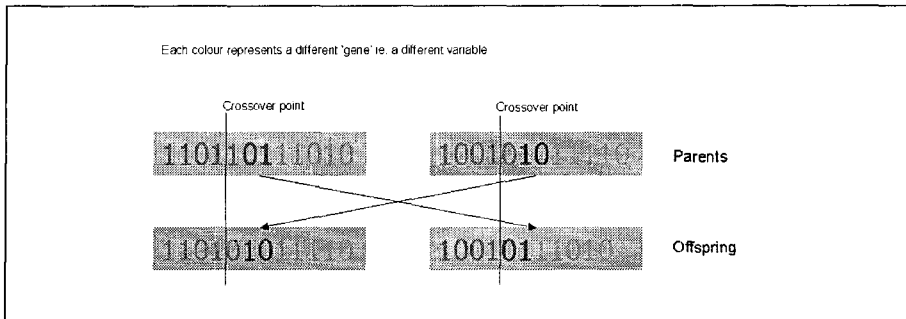
Fig. 2.3: Simple Steady State GA - Single Point Crossover

The loop shown in Fig. 2.2 is continued until some *stopping* criteria is reached or the whole population has converged to identical copies of a single individual. In practice finding a non-arbitrary stopping criteria is extremely difficult. Typically the optimisation is stopped after a *certain* number of evaluations (or generations) which is detemined by either the computer time available or when things *seem* to have stopped evolving - however experience shows that populations may appear to stop evolving for a long time, before restarting.

Simple schemes such as the one described above are surprisingly successful, however they have several drawbacks. The rate of convergence depends on the selection pressure and the mutation rate - higher selection pressure increases *exploitation* (climbing local hills for example) while higher mutation rates increase *searching* through the solution domain. However, mutation rates must be kept very small to avoid introducing too much random noise (after all the GA should try to be better than random search) and consequently they tend to converge to local optima. If this convergence happens before the domain has been adequately covered then the true global optimum may never be found since no diversity remains in the population. However, if convergence is too slow then no optima may found at all (the algorithm has effectively degenerated to than random search[†]).

Much of the research into this type of GA has been on ways to keep population diversity and hence search capacity while ensuring eventual convergence to the global optimum. In addition due to their stochastic nature, and the initial random population, a traditional GA must be run several times (at least) in order to determine whether the same solution is being found repeatedly, and even then there is no indication of whether the solution is the global optimum or a strongly attractive local optimum.

Should the GA converge to a local optimum it is still difficult to determine the quality of the result obtained — there is nothing to compare it with and repeated runs of the GA will produce different results when the problem has many local optima.

[†] It should be noted that random search only works because it is elitist.

2.2.2   Multi-Modal (Niching) Algorithms

In order to resolve some of these issues and to better meet the requirements of design engineers much development has been centred on multi-modal algorithms, often referred to as niching GAs. These make an attempt to find the global optimum by keeping multiple local optima in order to preserve diversity for longer while keeping selection pressure high.

A number of methods have been proposed for niching, including Cavicchio's (1970, Mahfoud 1992) *pre-selection* (offspring can only replace one of the parents), De Jong's[24] *crowding* (new individuals replace similar but less fit solutions), and several more recent methods including *fitness sharing*[43].

Fitness sharing accomplishes niching by degrading the objective fitness of an individual according the the presence of nearby, similar individuals. This type of niching requires a distance metric, as well as a parameter often referred to as the radius of the estimated niches.

Fitness sharing tends to spread the population out over multiple peaks in objective space in proportion to the height of the peaks (ie not really niching - just spreading). GAs with proportional selection and carefully tuned fitness sharing have been sucessfully used in solving a variety of multimodal test functions, although frequently they require "finishing" with local hill climbing routines[23] since the scaling of the objectives tends to favour the edges of the niches.

De Jong crowding works on replacement by first selecting a sub- population at random (size controlled by a parameter, the *Crowding Factor* and then finding the most similar (measured using a simple bit-by-bit comparison) member of this sub-population to replace with the child.

However crowding and fitness both introduce more parameters into an optimisation process that already has enough (rate of mutation, rate of crossover, selection parameters, size of population etc), all of which may need to be *tuned* to solve a particular problem.

A modified version of the crowding algorithm, the Struggle GA, was developed at MIT[45,83], and this has been successfully used at LENI within several projects. This algorithm incorporates niching by replacing the most similar member of the complete existing population with the new child only if the child has a better fitness score.

The key differences are:

- real variable encoding - no longer necessary to discretise variables;
- blend crossover—incorporates an exploration component as well as an exploitation component;
- comparison between new individual and the complete population;
- niching - keeps multiple *different* solutions, which in turn helps to preserve diversity;
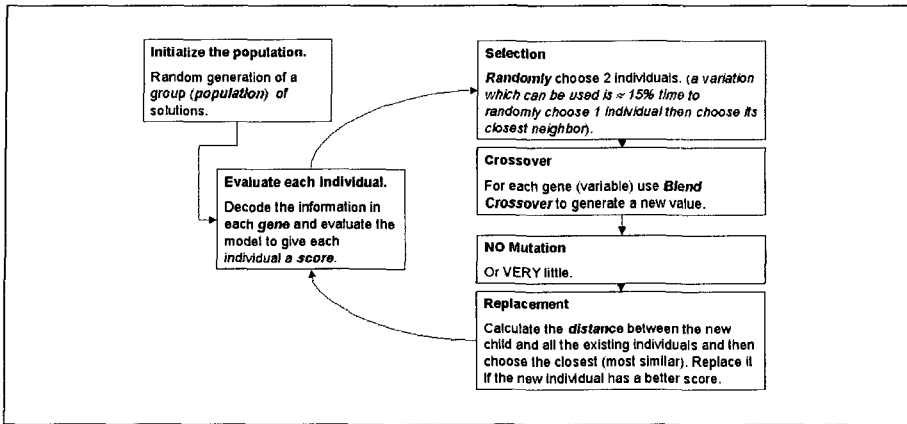
Fig. 2.4: Struggle GA - Evolutionary Scheme

The basic scheme for the Struggle GA is shown in Fig. 2.4 and an explanation of real variable encoding and blend crossover is given in Fig. 2.5 and Fig. 2.6.
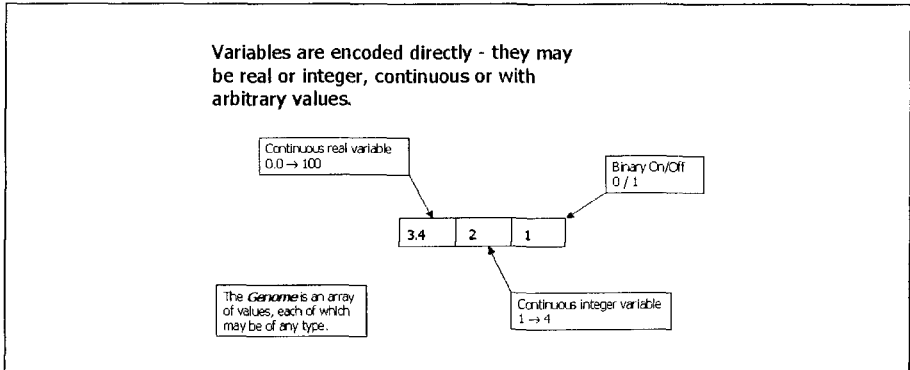
Fig. 2.5: Real Variable Encoding

The key component of the Struggle (and equally of all the niching algorithms) is the idea of a *similarity* measure between solutions (for example the euclidean distance between two points). This allows selection pressure and more importantly replacement pressure to be applied to the process - only *similar* individuals compete for existence. It does however raise the question of how to calculate a representative similarity (or disimilarity). To date at LENI the distance has been calculated as a Euclidean distance between two solution vectors as per (2.5).

Fig. 2.6: Blend crossover

$$d_{rs} = \sqrt{\sum_{i=1}^{N} w_i (x_{si} - x_{ri})^2} \tag{2.5}$$

where N is the dimension of the vectors $\vec{x}_r$ and $\vec{x}_s$, d_{rs} is the distance between them and w_i is a weight.

While this is simple to calculate it becomes less representative as the number of dimensions increases. Indeed, in order to solve the problems mentioned earlier it was necessary to weight a small number of the independent variables—which introduces the problem of deciding which variables to weight, and by how much.

2.2.3   Advantages and Disadvantages of the Struggle GA

The use of real or integer variables to represent the decision variables removes the need to encode variables into binary representations, removing the need to decide what precision is needed and allowing the direct manipulation of values in the crossover and mutation process.

The struggle uses strong elitism—only one individual is replaced at a time, and then only if the proposed new solution is better. This tends to promote well separated niches, and makes the population stable.

However when the problem has many variables the definition of similarity becomes more difficult—this has been termed the *Curse of Dimensionality*[7]. Any two randomly picked vectors in a high-dimensional hypercube tend to have a constant distance from each other, no matter what the measure. So it seems counter-productive to calculate distances using high dimension spaces. The effect of this was apparent from the earlier work[21] on a 38 variable prolem — when distance calculations were performed without heavily

weighting a small number of *important* variables the Struggle GA performed poorly.

2.3   Multi-Objective Optimisation

2.3.1   Introduction

Multi-objective optimisation can be defined as the problem of finding (quoted from Coello Coello[16] after Gero[40]):

> a vector of decision variables which satisfies constraints and optimizes a vector function whose elements represent the objective functions. These functions form a mathematical description of performance criteria which are usually in conflict with each other. Hence, the term "optimise" means finding such a solution which would give the values of all the objective functions acceptable to the designer. Formally, this can be stated as follows:
>
> Find the vector $\vec{x}^* = [x_1^*, x_2^*, \ldots, x_n^*]^T$ which will satisfy the m inequality constraints:

$$g_i(\vec{x}) \geq 0 \ i = 1, 2, \ldots, m \tag{2.6}$$

> the p equality constraints

$$h_i(\vec{x}) = 0 \ i = 1, 2, \ldots, p \tag{2.7}$$

> and optimizes the vector function

$$\vec{f}(\vec{x}) = [f_1(\vec{x}), f_2(\vec{x}), \ldots, f_k(\vec{x})]^T \tag{2.8}$$

> where $\vec{x} = [x_1, x_2, \ldots, x_n]^T$ is the vector of decision variables. In other words, we wish to determine from among the set $\mathcal{F}$ of all numbers which satisty (2.6) and (2.7) the particular set $x_1^*, x_2^*, \ldots, x_k^*$ which yields the optimum values of all the objective functions.

In order to satisfy most design problems a compromise must usually be found among a series of conflicting criteria. This is frequently done using some form of aggregating function to combine the multiple criteria into one objective. Where the problem lends itself to this treatment then a single solution is produced and no more analysis needs to be performed. In practice, though, this is rare, and decision makers prefer to have a range of possible solutions to choose from, rather than be presented with one best solution.

There is also the evident problem that we must somehow scale the objectives in order to avoid one objective dominating the others, and this in turn implies knowledge of each of the objective functions—something that in most real world applications is often computationally too expensive.

One of the most simple, and hence popular, aggregation methods is a weighted sum of the form shown in (2.9), but this raises the question of choosing the weights (for example what weights do you give to pollution and running costs in order to add them to investment costs?).

$$\min \sum_{i=1}^{k} w_i f_i(\vec{x}) \tag{2.9}$$

where $w_i \geq 0$ and

$$\sum_{i=1}^{k} w_i = 1 \tag{2.10}$$

2.3.2 Pareto Optimisation

In the 19th century Vilfredo Pareto[75] formulated the idea of a *Pareto optimum*. A point $\vec{x}_* \in \mathcal{F}$ is Pareto optimal if for every $\vec{x} \in \mathcal{F}$ either

$$\bigwedge_{i \in I} \left(f_i(\vec{x}) = f_i(\vec{x^*}) \right) \tag{2.11}$$

or, there is a least one $i \in I$ such that

$$f_i(\vec{x}) > f_i(\vec{x}^*) \tag{2.12}$$

In other words, $\vec{x^*}$ is Pareto optimal if there exists no feasible vector $\vec{x}$ which would decrease some criterion without causing a simultaneous increase in at least one other criterion. The pareto front is formed by the set of solutions that are Pareto optimal and may be continous or disjoint, convex or concave depending on the optimisation problem.

The Pareto front is usually formed by a set of solutions called the non-dominated solutions or non-dominated front (NDF), although it should be noted that if the objectives are *non-conflicting* then there will only be one solution - not a Pareto optimal front.

The practical goal of a multi-objective optimisation is to find a suficient number of solutions on or near the Pareto front such that the form of the tradeoff surface is discernible, and the decision maker can take an informed decision. The surface actually found by a multi-objective optimisation is a non-dominated front approximating the Pareto optimal front.

2.3.3 Dominance and Ranking

If multiple objectives are not to be aggregated in some way to reduce the problem to a single objective then a means must be found to decide if one solution is *better* than another, and indeed what better actually means. This is frequently done in terms of *dominance*.

The concept of dominance is key to multi-objective optimisation, and is defined by Deb[26] as:

A solution $\vec{x}_1$ is said to dominate the other solution $\vec{x}_2$ if both the following are true:

1. The solution $\vec{x}_1$ is no worse that $\vec{x}_2$ in all objectives.

2. The solution $\vec{x}_1$ is strictly better than $\vec{x}_2$ in at least one objective.

Solution $\vec{x}_1$ can be said to dominate $\vec{x}_2$. Fig. 2.7(a) shows a simple population from a hypothetical two-objective optimisation, illustrating the points that are *non-dominated*, and those that are dominated. The set of points that are non-dominated form the *non-dominated front* (NDF) and this is shown in Fig. 2.7(a) by a dotted line joining the non-dominated points.

However, simply dividing points into dominated and non-dominated sets is not sufficiently fine grained to be of use in a multi-objective optimisation algorithm. Consequently a means of ordering points based on this domanince is needed. The method is usually termed *ranking* and several alternative schemes have been proposed by Goldberg[42], Fonseca and Fleming[35] and more recently that proposed by Zitzler[103].

Goldberg's ranking has been used exclusively in this work since it does not suffer the tendency[25] to favour points in the middle of the optimal front and was found to perform adequately in the tests.

The non-dominated points in the population are found and given the rank 1. Then these points are temporarily removed from the population and the new non-dominating points are found and given rank 2. This continues until all the points have a rank (or until a maximum number of ranks have been found, depending on the requirements of the algorithm). Fig. 2.7(b) shows Goldberg's ranking applied to the example points from Fig. 2.7(a).

This rank can then be used either in a selection method, replacement method or, as in the CPEA, in a killing strategy.

In the case of a single objective function this ranking scheme reduces to a simple sort on the objective value.

(a) The non-dominated front.   (b) Goldberg's Ranking
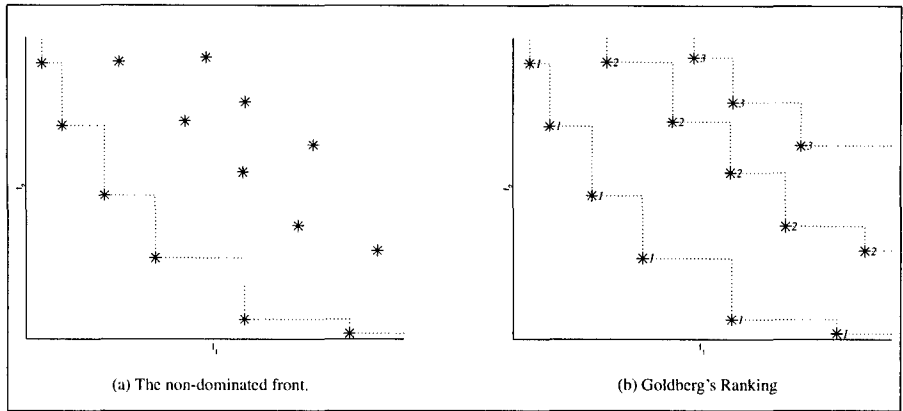
Fig. 2.7: Dominance with two objectives, using Goldberg's Ranking.

2.4 Finding Pareto fronts with EAs

Evolutionary algorithms are an obvious choice for finding Pareto fronts, since they operate with a population of solutions. A survey of the literature reveals that multi-objective optimisation (MOO) is an active area with several different approaches to Pareto-EA based optimisation.

There are several good reviews of the state of research in MOO, such as those produced regularly by Coello Coello[16],[17] and that by Van Veldhuizen and Lamont[97], or earlier by Fonseca and Fleming[36]. Recently Zitzler[103] proposed a new algorithm, the SPEA (Strength Pareto Evolutionary Archive), and compared this favorably with the earlier published algorithms, on a series of test problems proposed by Deb[25] referred to in the literature as Zitzler's T1–T6.

Goldberg[42] first proposed pareto based fitness assignment based on the ranking described in section 2.3.3. His idea was taken up by Fonseca and Fleming[35] and ,independently, by Horn et al.[48]. Fonseca used a slightly modified ranking where an individual's rank was equal to the number of individuals that it dominates.

Horn et al.[48] developed a "Niched Pareto GA" making use of domination tournaments and sharing on the non-dominated frontier. While they found that this was initially successful they also found that the performance was very sensitive to the settings of certain paramters, in particular population size, sharing radius and the tournament size.

Knowles and Corne[56] proposed the Pareto Archived Evolution Strategy (PAES) which is an alternative strategy based on local search with an archive of non-dominated solutions. Local search methods such as

tabu were used to create new candidate solutions, and the objective space subdivided into a grid to encourage a good distribution along the non-dominated front. This algorithm was found[55] to be superior to the NSGA or Horn and Nafpliotis, and the SPEA for all but a highly deceptive problem (Zitzler's T5). As such this has been suggested as a baseline comparison algorithm, and also as an alternative where local search is appropriate.

Andersson and Wallace[4] proposed a modification of the Struggle GA for multi-objective optimisation, termed the *Multi Objective Struggle Genetic Algorithm* (MOSGA), that incorporated the struggle crowding algorithm as described previously with the Pareto ranking used by Fonseca and Fleming[36].

New individuals were produced with blend crossover and the most similar solution in objective space was replaced by the new individual if the new individual dominates. A variation was also proposed in which the crowding algorithm was performed on an equally weighted combination of variable and objective space, and this was found to improve retention of multiple solutions. There was no need to tune parameters, and the EA seemed to perform better than many of the other published methods.

Recently Deb et al.[27] proposed the "NSGA-II", an update of their non-dominated sorting algorithm, that makes use of the idea of ranking and crowding in objective space, together with tournament selection and elitism in order to promote a distribution of solutions on the non-dominated front. Compared to PAES this seems to produce a better distribution of points along the non-dominated front, and better convergence to the Pareto front.

A non-generational, elitist MOO has been proposed by Borges and Barbosa[11] that ranks each point on the number of other points dominated and the local solution density in objective space, $F(i) = (1 + D(i))^\beta (1 + N(i))$ where $D(i)$ is the domination measure for solution i and $N(i)$ is the neighbourhood density. New points are generated from parents selected randomly. The worst point in the population is replaced with the new point only if its domination score or its density is better. This idea is similar to the MOSGA[4].

Deb[25] proposed a series of two objective problems designed to test different aspects of a MOO's performance - ability to find diverse pareto-optimal (PO) fronts, ability to converge to the global PO front and ability to handle convex, non-convex or discontinuous PO fronts. An important observation made by Deb is that assigning fitness proportional to the number of points dominated as done in Zitzler[103] and Fonseca and Fleming[35], favors solutions in the middle of the pareto front if the pareto front is concave.

Coello and Christiansen[15] propose a method based on aggregating different weighted objectives and performing multiple optimisations.

While there is evidently a great deal of variation in the approaches, it is worth summarising the following points:

- All except the MOSGA appear to use EA schemes with binary coded variables, in contrast to using

real variables.

- Most use a population-replacing generational approach, where at each generation an entire new population, is generated from the previous. To avoid losing non-dominated solutions, some algorithms use "elitism", or marking the non-dominated solutions as a "special" population that will persist from generation to generation. The algorithms with elitism consistently do better than those without.

- They try to find only the global Pareto-optimal set of solutions for the problem, and tend to lose other local optima.

- Many use the concept of "fitness sharing", in the objective domain to thin the points on the non-dominated surface. Unfortunately, while similar points in parameter space will usually produce similar results in objective space, the converse is not usually true—hence diversity in parameter space is lost.

There is little discussion on multi-modality in the MOO literature. In the practical world there is a need for a technique that will find several "locally" non-dominated fronts, representing different technological solutions to the problem at hand. There is a similar need for a robust technique that does not require tuning to each new problem.

2.5 Clustering Pareto Evolutionary Algorithm

The motivation behind the CPEA was to develop a MOO method that required little or no tuning, and behaved well on a wide range of problems, producing a good distribution of solutions along the non-dominated frontier, while simultaneously keeping a distribution of solutions in variable space. In some problems this means finding multiple locally non-dominated fronts, while in others, different regions of the variable domain may contribute to the global Pareto front.

2.5.1 Description of Algorithm

In the CPEA, multi-objective optimisation is integrated with clustering and multi-dimensional scaling concepts from statistical multivariate analysis. In contrast to many of the algorithms mentioned above, clustering is in parameter space to try to preserve local optima, and a "breed and die" population control is employed which allows the population to expand over the non-dominated fronts.

There are two separate processes which operate on the population of solutions. The first produces new potential solutions, clusters and ranks the population. The second tries to remove individuals that are deemed poor, and when absolutely necessary (due to over population) thins the population.

The algorithm begins as follows:

1. An initial population of p_i individuals is randomly generated and the objectives for each individual are evaluated[‡].

2. The population is clustered into n_{cl} clusters as described below. If $n_{cl} = 1$ then there is no clustering[§].

3. A dominance matrix is determined to identify which points are dominated by which others. If there is already a dominance matrix it is updated.

4. The individuals in each cluster are ranked using Goldberg's ranking scheme.

5. Each cluster is given the chance to produce n_{ch} children. A single child is created by the following process:

   - A parent is chosen at random from the cluster in question.
   - A decision is made as to whether the second parent will come from this cluster, or from another. Throughout this work 10% of the time the second parent was chosen from another cluster[*].
   - A child is created from the two parents using crossover and mutation as described below.

6. Each new individual is inserted into the existing population provided that it is not identical to an existing point.

7. The new individuals are evaluated.

8. If the population of a cluster is greater than a pre-defined limit, $n_{max\_cluster\_size}$, then a killing and thinning strategy is applied.

9. Steps 2 to 8 are repeated until a maximum number of problem evaluations $n_{evaluations}$ has been reached.

Clustering Policy.   Clustering could be applied to all or any fixed subset of the variables. Before clustering, the dimensions found to have zero standard deviation were removed from the set of variables to cluster. The remaining variables were then adjusted to have a zero mean, and could be left in original units or normalised either by the standard deviation or by the range of the data.

Following this step a Multi Dimensional Scaling (MDS) (see 3.2) technique could, if desired, be applied to reduce the number of dimensions chosen for clustering to a manageable size, typically in this work chosen as two so that plots of the two principal components could be produced.

Where the clustering technique permitted, the automatic detection of "natural" clusters in the data this could also be performed, and the number of clusters set to be one more than apparently present in the data. This

---

[‡] In the CPEA the initial population is chosen to cover the search domain, since the first killing phase will reduce the population to $p_{min\_cluster\_size}$

[§] The value for n_{cl} may be determined automatically.

[*] An alternative idea was also considered, making use of the fuzzy degree of membership of the individual in each of the clusters. This was applicable only when using fuzzy c-means clustering and has not been fully investigated in light of the success of the simpler method, although preliminary results indicated similar performance on the test problems.

was done to promote the generation of new clusters, and hence exploration. Alternatively the number of clusters could be fixed.

Crossover and Mutation. Two types of crossover were implemented - real variable blend crossover, described previously in Fig. 2.6, and real variable uniform crossover (also referred to as naive crossover by Deb[26], pg 108). A real variable mutation operator was implemented as shown in (2.13) (see Deb[26], pg 119).

$$y_i^{(1,t+1)} = x_i^{(1,t+1)} + N(0, \sigma_i) \tag{2.13}$$

where $N(0, \sigma_i)$ represents a zero mean normal probability distribution with variance σ_i. The value of σ_i is chosen as 10% of the variation in the cluster, $\sigma_i = 0.1 \max x_i - \min x_i$ and then trimmed to be within the limits for each variable.

The blend crossover produces children in the hypercube formed by the two parents, and in order to search the domain space relies on a large variation in the population from which the parents are chosen. If the two parents have identical values for one of the variables, say $x_i = c$, then the offspring will similarly have $x_i = c$.

Fig. 2.8(a) and Fig. 2.8(b) show the distribution of 80 offspring produced by blend crossover with and without mutation. The two parents were chosen from the sample set of 20 points used earlier in this section, with all the points assumed to be in the same cluster.

Fig. 2.8(c) and Fig. 2.8(d) show the distribution produced by uniform crossover with and without mutation, from the same two parents.

The uniform crossover produces new points spread around each of the vertices of the hypercube formed by the two parents. This does not cover the domain adequately on its own and relies heavily on the mutation operator for exploration.

Fig. 2.8 also shows that the mutation operator can greatly increase the domain explored with blend crossover when the variables are aligned along one of the principal axes. The blend crossover produces points uniformly in the hypercube determined by the parents. The mutation effectively increases the size of the hypercube.

Clearly it is necessary to use mutation with the uniform crossover, but it is not so clear with the blend crossover. Tests performed by Leyland et al.[61] showed that the uniform crossover with mutation consistently outperformed blend crossover on the set of problems used by Zitzler[103].

On Zitzler's T4[9] the blend crossover with mutation took on average twice as long to converge to the optimum non-dominated front as uniform crossover with mutation. Blend crossover without mutation did not

---

[9]see later discussion of results in section 4.1.5 for a description of this problem
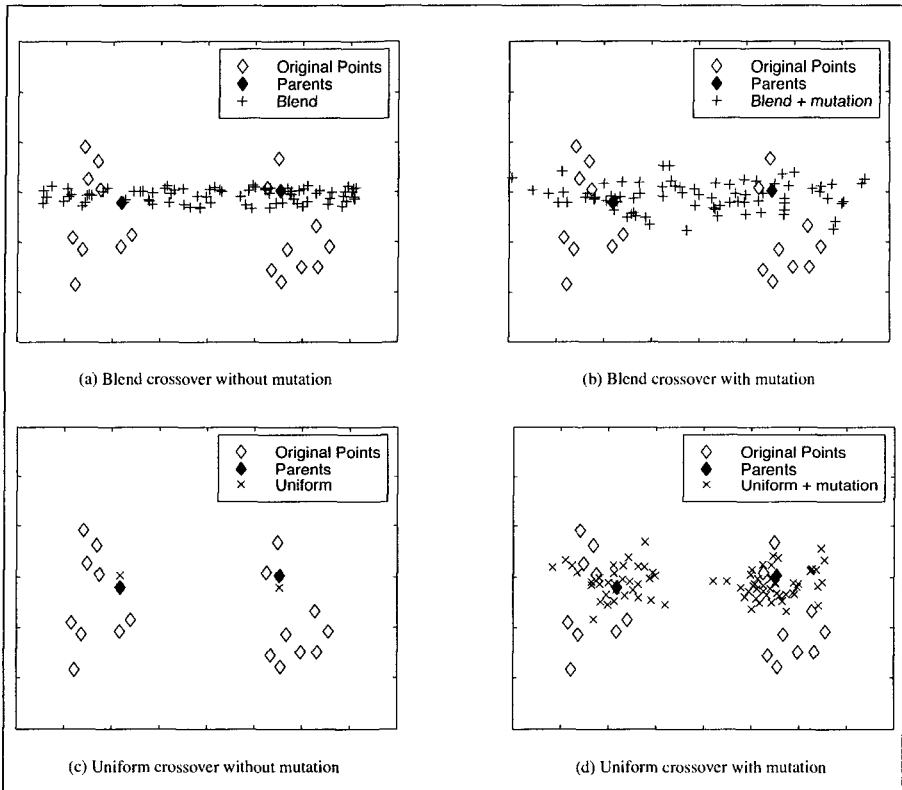
Fig. 2.8: Illustration of the distribution of offspring when using real variable blend crossover, and real variable uniform crossover, together with mutation.

consistently find the global Pareto front, even when allowed to run for 4 times the maximum number of iterations taken by the uniform crossover with mutation.

However, the preliminary analysis suggests that this may be due to inherent properties of the problem, and that the best crossover and mutation method may be problem dependent — an idea that has lead to the idea of evolving operator choice[60,63].

In light of this both crossover methods were tested in conjunction with clustering, and the results are discussed later.

Killing Strategy. It is the killing strategy that applies the majority of the selection pressure. If a cluster has more than $n_{max\_cluster\_size}$ points :

1. If the maximum rank r in a cluster is greater than r_{max} (usually this will be set to 1) then all of the points with the highest (worst) rank are removed. This is repeated until the number of points in the cluster is less than $n_{max\_cluster\_size}$ or until $r = r_{max}$.

2. If the number of points is still greater than $n_{max\_cluster\_size}$ and a thinning mechanism is active then it is applied.

Without thinning the population grows continually once all the individuals are ranked r_{max} or better. While this can be tolerated to a certain extent, when populations become larger that than approximately 500 individuals the ranking and clustering take an unacceptable time[||].

The question of how to thin is clearly very important since it directly affects the non-dominated front. three dimensionalZitzler[103] used clustering in the objective space to thin points from the non-dominated front. The population was clustered into $n_{clusters}$ clusters, the individual closest to the centroid of each cluster was considered representative and all the others removed. However the Pareto front will clearly lie at one edge of each cluster and *not* at the centroid - so the best points (closest to true Pareto front) are systematically removed and this is clearly not ideal.

Two Dimensional Thinning. The thinning used in the CPEA on two-dimensional problems was proposed by Leyland et al.[61].

In each cluster that has more than $n_{max\_cluster\_size}$ points :

1. Sort the points on the objectives.

2. Calculate the distance between each pair of points where a pair of points is separated by $n_{quad} - 2$ intermediate points, and find the shortest section. This represents the shortest section of length n_{quad} points on the non-dominated front.

3. Regression fit a quadratic to the n_{quad} points.

4. Find all the points *behind* the quadratic. From these points find and remove the point who's two neighbours are the closest together.

5. Repeat steps 1 to 4 until the cluster contains $n_{max\_cluster\_size}$ points.

[||]The current implementation is in MATLAB, an interpreted vector oriented language. A dramatic speed increase could be achieved by re-implementing the CPEA in a compiled language such as C or C++ but there would still be a limiting population size although it would be much bigger.

This procedure adds selection pressure since the point removed is chosen from among the points farthest from the non-dominated front, and improves coverage since it removes the point that is in the densest region.

Three Dimensional Thinning. In three dimensions or more this method becomes impractical as the population can no longer be sorted along one objective, and it takes an exponentially increasing number of points to fit a quadratic surface. However the over-population problem becomes even worse with more objectives since it becomes easier to find new non-dominated points.

The following thinning stategy, based on that used by Zitzler[103], was used for the three objective problem considered in chapter 5.

Within each cluster that has more than $n_{max\_cluster\_size}$ points:

1. Cluster the objectives into $n_{max\_cluster\_size}$ sub-clusters.

2. In each sub-cluster with more than 1 point, find the point closest to the cluster centre*. This point is taken as the representative point for the cluster.

3. Remove all the points except the representative point from the sub-cluster.

4. repeat 2 and 3 for all of the sub-clusters.

This technique has demonstrated some limited success, although is far from ideal. Unfortunately this method favors points behind the POF**.

and further work in this area is needed.

Parallel Implementation. The time to run the simulation models used in the later test cases of Chapters 3 and 4 varied from 0.2 seconds per evaluation for the district heating problem to 20 seconds for the vehicle drive train (on a 700MHz PC).

In order to optimise these problems in a reasonable time it was necessary to implement a parallel simulation evaluation strategy.

This is a simple PVM[79] based master slave configuration in which:

1. The CPEA master node produces a series of new solutions to evaluate, and placed them in a queue, to be sent out to available slaves.

---

*The cluster centre is available from the fuzzy clustering algorithm

**This is a problem with Zitzler's thinning.

2. Whenever an evaluated solution was returned from a slave the population was ranked, killed and thinned.

3. When a slave became available it was sent the next solution in the queue, to evaluate and return.

The number of slaves that can effectively be used with this structure is limited by the time to rank, kill, thin and reproduce, all of which is handled by the master. Of these the ranking and thinning are the most time consuming.

There is equally a network overhead — if the problem is extremeley quick to solve then the network overhead becomes important with most of the time sent sending and receiving information between nodes. Consequently, for this sytem to be of applicable the simulation must take longer than the ranking, killing, thinning and reproduction stages.

For the majority of real world problems this will be the case, even with the relatively slow implementation of CPEA in MATLAB.

Initially the computers used as slaves were not dedicated machines, but rather were individual computers in the laboratory. A system was implemented to launch slaves as computers became available (for example when turned on) and at set times (overnight) and the scheduling system was made robust — if a computer was turned off while processing a job, the job was automatically re-scheduled on an available machine. This was found necessary in the laboratory environment where machines were periodically turned off by users.

An added benefit of this mechanism was observed in the vehicle drive train simulation. When certain infeasible solutions were presented to the ADVISOR model the slave crashed. In this event the job was re-submitted three times and if it crashed each time the solution was marked as infeasible and effectively removed from the population. This was found necessary as the complexity of the problem made it unable to identify all the possible solutions.

During the work a cluster of 22 Linux machines became available, and the parallel system was modified to work with these — these were dedicated, robust machines and permitted a dramatic increase in solution times.

2.5.2 Notes on the CPEA

- The clustering is a key element of the algorithm, and several techniques were investigated before finding a method that provided a good clustering behaviour along with satisfactory computational performance. These are discussed in more detail below in 3.

- The number of clusters could be pre-defined or, if the clustering algorithm permitted, could be determined automatically. If chosen automatically, then the number of clusters was set to one more than

the number of "natural" clusters identified, to encourage the development of new solutions.

- In one iteration, each cluster can breed several times. Practically, this is because the reclustering and ranking at the beginning of each iteration is very slow. Perhaps ideally, the reclustering could be performed after each birth and death.

- The algorithm gives an equal chance of breeding to each *cluster*, and not to each individual. This is very important for the performance of the algorithm, as otherwise large clusters (which may have already finished evolving) get far more process time than smaller clusters (which are probably evolving more rapidly).

- The choice of the second parent favours a parent in the same cluster as the first, and small clusters are more likely to chose a local parent than large ones, thus favouring the growth and evolution of small clusters. If the parent is non-local, the parent's cluster is chosen randomly.

- Because more than one point can be removed by thinning, the population can, and occasionally does, decrease. The long term trend, however, is for a growing population limited by a predetermined maximum — generally as large as computing resources can handle.

Chapter 3

Clustering

3.1 Introduction

Cluster analysis is a way to partition a set of objects into groups, or *clusters* in such a way that the profiles of objects in the same cluster are very similar, and the profiles of objects in different clusters are quite distinct. The more distinct the clusters the "better" or more "crisp" the clusters. The goal of introducing clustering was to deal with the difficulties encountered while optimising problems with many variables using algorithms that have a concept of distance between solutions, such as the StruggleGA.

Exactly what constitutes a cluster is not well defined, and in many cases clusters are not in fact well separated from one another. Considering Fig. 3.1 the intuitive answer (allowing nested clusters) is that there are two clusters, each with three sub clusters.

However, it is equally reasonable to classify them as only two clusters, or as six clusters, or even as four. The preference for the human visual system to classify them in one manner is not necessarily justified. In general it is extremely difficult to determine the true number of clusters present in a data set.

The clustering in the CPEA is one of the key issues. Many different methods of clustering data were found in the literature, and in order to find a suitable method for the clustering phase of the optimisation algorithm several promising methods were implemented and tested on sample problems. The methods considered are detailed below.

(a) Initial 20 points.

(b) Two clusters.

(c) Six clusters.

(d) Four clusters.
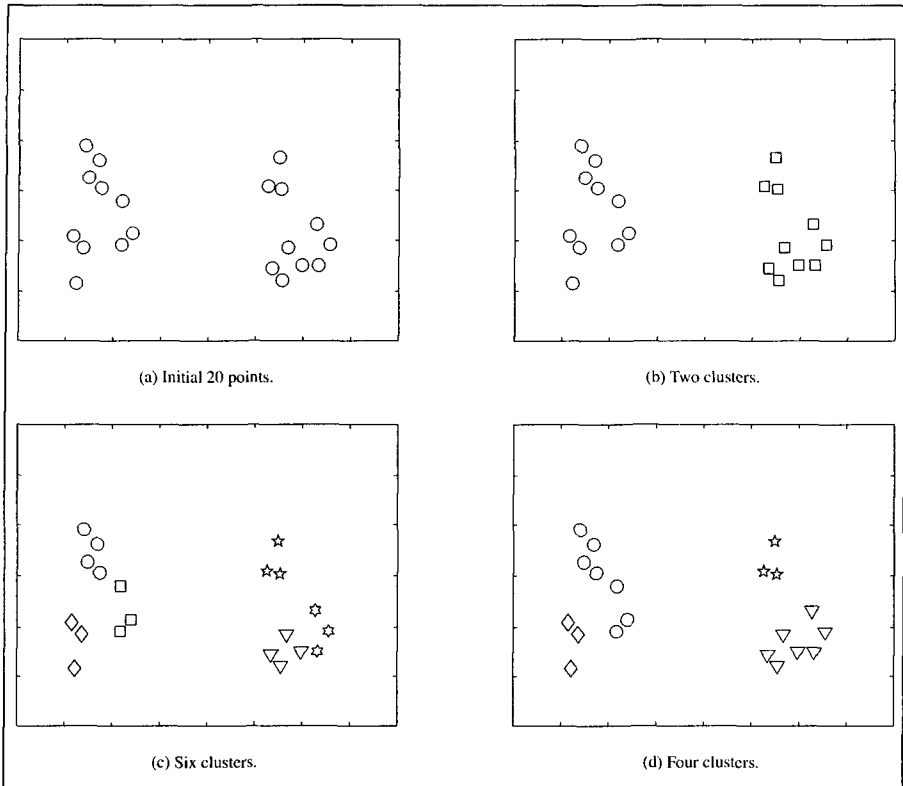
Fig. 3.1: Example of 20 points showing 4 different ways they may be clustered.

3.2 Multi-Dimensional Scaling

The problems typically encountered in system integration have many very different variables. This makes it difficult to visualise and inspect the results of an optimisation, even without introducing multiple objectives. The *curse of dimensionality*[7] makes it counter-productive to cluster on high dimension spaces, as well as time consuming. It was thus essential to find a means of dealing with the dimensionality of problems in the clustering, and hence in the similarity measurement[52].

Multi-dimensional scaling[19] consists of a variety of statistical techniques aimed at reducing the dimension of a set of data points with a minimum of information loss. Classical multi-dimensional scaling is the principal method used here. This aims to find a mapping from N dimensional space to P dimensional space where $P < N$ and typically P is 2 or 3 so that results can be shown graphically. The mapping attempts to keep the

distance between each pair of points in mapped space as close as possible to the distance in the original space. In addition the order of distances should be retained. Although there are several alternative methods for doing this, including optimisation[81] and non-linear neural net approaches[9], a simple and effective method is principal component analysis (PCA), also referred to as the Karhunan-Loeve transform (see Jackson[50] or Geva[41]). This is the method that is used in this work and allows the reduction of dimension from N to P with a minimum squared error. It also permits the reconstruction of an estimate of the N original dimensions from the reduced set P.

3.3 Distance Measures

In order to perform any kind of clustering a similarity or disimilarity metric is needed. The attributes of the objects (in our case the variables forming the solution) can be of three types :

1. binary - two values

2. discrete - a finite number of values

3. continuous - real values

and can be measured on four different scales:

1. Nominal - qualitative where the values are just different "names"

2. Ordinal - qualititive where the values can be ordered good, better, best etc.

3. Interval - quantitative where the difference is meaningful e.g. objects rated on a scale of 1 to 10

4. Ratio - quantitative where the scale has an absolute zero so that ratios are meaningful.

The problems we are interested in typically have a mixture of variables that are on an *ordinal* scale (represent the choice of components) or *ratio* scale (component operating parameters), and consequently the choice of distance measure must be made with care.

Similarity of Binary Vectors   Many possible measures, or similarity coefficients, exist for binary vectors, usually having a value between 0 and 1, where 1 indicates vectors that are identical and 0 completely disimilar. In addition many rationales exist for deciding which is appropriate (see[13],[53] for more details).

The comparison of two binary vectors, p and q, leads to four quantities:

- M_{01} = the number of positions where p was 0 and q was 1

- M_{10} = the number of positions where p was 1 and q was 0

- M_{00} = the number of positions where p was 0 and q was 0

- M_{11} = the number of positions where p was 1 and q was 1

The simplest similarity coefficient, *SMC*, is the simple matching coefficient:

$$SMC = \frac{M_{11} + M_{00}}{M_{01} + M_{10} + M_{11} + M_{00}} \tag{3.1}$$

while another common measure is the Jaccard coefficient (also referred to as the Hamming distance [19]) :

$$J = \frac{M_{11}}{M_{01} + M_{10} + M_{11}} \tag{3.2}$$

The *SMC* equates similarity with the total number of matches, while *J* considers only matches on one's as important. The relevance depends on the case in question.

Similarity of Real Variables in the Solution Vector   Several similarity measures exist for real values - the most common one being the Euclidean distance. This can in fact be considered as special case of the Minkowski metric, defined as :

$$p_{ij} = \left(\sum_{k=1}^{d} |x_{ik} - x_{jk}|^r \right)^{\frac{1}{r}} \tag{3.3}$$

where r is a parameter, d is the dimension of the data object and x_{ij} and x_{jk} are respectively the kth components of the ith and i^{th} objects, x_i and x_j.

The parameter r frequently takes one of the following:

- $r = 1$. The *City Block* (Manhattan, L_1 norm) distance.

- $r = 2$. Euclidean distance - the most commonly used measure.

- $r = \infty$. The L_∞ norm distance is equivalent to the maximum difference between any component of the vectors.

Similarity of Integer Variables in the Solution Vector. Frequently a variable is an integer that represents a choice of component type, for example vehicle engine type may be diesel, petrol, gas or electric. However, in this example the difference between any two of these types of engine should be equivalent. Consequently the dissimilarity between two values can be measured only on a *nominal* scale and care must be taken to ensure this.

Aggregation of different variable types. A common problem is the mixture of different variable types in the same problem. How does a dissimilarity measure cope with this? One approach that was adopted for an earlier project[68] is to group the real parameters together with the controlling integers. For example engine type would be grouped with engine size, and power. A solution vector representing a vehicle drive train with a gas engine may then be compared with a vehicle containing an electric motor using a *nominal* scale, while two alternative electric motor vehicles would be compared on a *ratio* scale using the relevant parameters.

A drawback of this approach is the added complication of calculating dissimilarities — each pair of object types may need a different dissimilarity function, together with the implicit creation of clusters — two individual deemed different on a nominal scale must be in different clusters.

An alternative is to weight the importance of the integer variables in the solution vector so that they dominate the continuous parameters in the dissimilarity calculation.

3.4 Clustering Techniques

3.4.1 Hierarchical Clustering

Hierarchical clustering methods may be either divisive or agglomerative. Divisive methods start with one cluster and repeatedly split it into sub clusters, and agglomerative methods start with every point as a singleton cluster then merge pairs of clusters. They typically make use of a proximity graph - the matrix of dissimilarities between each pair of points.

A simple agglomerative clustering algorithm, known as the Lance-Williams[59] algorithm is:

1. Create *n* singleton clusters, one for each of the *n* points and calculate the initial proximity matrix as the disimilarity matrix.

2. Merge the closest (most similar) two clusters.

3. Update the proximity matrix to reflect the proximity between the new cluster and the original clusters.

4. Repeat steps 2 and 3 until only a single cluster remains.
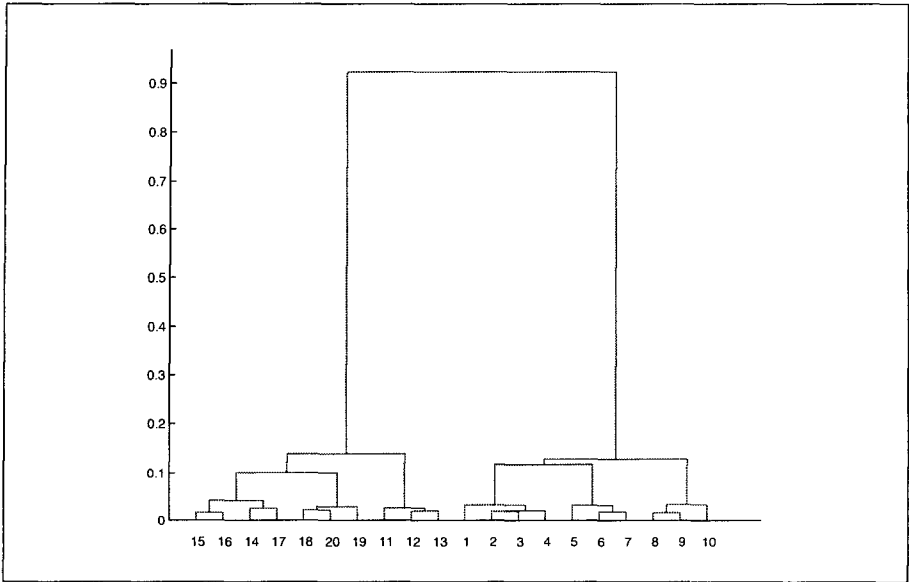
Fig. 3.2: Sample dendrogram for the sample 20 point dataset obtained with Ward linkage

The key element is the calculation of the distance between clusters, or *linkage*, and this is where techniques differ. Any of the cluster proximity calculations used may be defined using the Lance-Williams formula for the proximity between clusters Q and R, where R has been formed by merging clusters A and B:

$$p(R,Q) = \alpha_A p(A,Q) + \alpha_B p(B,Q) + \beta p(A,Q) + \gamma |p(A,Q) - p(B,Q)| \tag{3.4}$$

where α_A, α_B, β and γ depend upon the type of linkage calculation, as described below and in Table 3.1. The result of this linkage process can best be described on a *dendrogram*, such as that in Fig. 3.2 for the data from Fig. 3.1. once the linkage has been established the clusters can be obtained by cutting the tree where desired.

Single Linkage. This is the nearest neighbour or minimum distance method, where two clusters are merged based on the nearest two points. This is very quick to calculate but tends to result in the merging of ill-defined clusters. In particular this method tends to find "line" clusters.

Linkage Type	α_A	α_B	β	γ
Single linkage	$\frac{1}{2}$	$\frac{1}{2}$	0	$-\frac{1}{2}$
Complete linkage	$\frac{1}{2}$	$\frac{1}{2}$	0	$\frac{1}{2}$
Ward linkage	$\frac{n_R+n_P}{n_R+n_P+n_Q}$	$\frac{n_R+n_Q}{n_R+n_P+n_Q}$	$\frac{-n_R}{n_R+n_P+n_Q}$	0

Table 3.1: Several linkage methods tested with hierarchical linkage

Complete Linkage. Also called farthest neighbour or maximum distance method. Two clusters are merged based on the farthest two points in the clusters. Again this is quick to calculate but tends to favor smaller, tighter clusters.

Ward Linkage. This method corresponds to merging the two clusters which minimise the information loss in terms of a the sum of the squared errors, defined as

$$ESS = \sum_{k=1}^{K} \sum_{x_i \in C_k} \sum_{j=1}^{D} (x_{ij} - \bar{x}_{kj})^2 \tag{3.5}$$

where K is the number of clusters, x_i is point i in cluster k and D is the number of dimension of x.

This is the same objective that is used in k-means clustering as described later, however the objective is applied at each merge step rather than on the overall population. There is a tendency to favor spherical clusters.

Clustering the initial example of 20 points demonstrates some of the differences in the methods. Asking for 6 clusters gives the clusters shown in Fig. 3.3(a) for Ward and complete linkage but Fig. 3.3(b) for single linkage. Asking for 4 clusters gives Fig. 3.4(a) for ward and single linkage, but Fig. 3.4(b) for complete linkage.

Ward linkage appeared the most promising of the three hierarchical methods, although it requires the largest computational effort.

Choosing the Number of Clusters with Hierarchical Methods. It is not simple to automatically decide the number of clusters when using hierarchical clustering methods. A measure of *inconsistency* is suggested in the MATLAB statistics toolbox[90] that compares link length of adjacent links. Clusters are then formed by breaking links which have a higher inconsistency than a given value. While this may seem intuitive it is based on a comparison of link length over a relatively small number of link levels and was later found to make clusters that were unstable when adding points.

(a) Clusters from Ward and complete linkage for 6 clusters.

(b) Clusters from single linkage for 6 clusters.
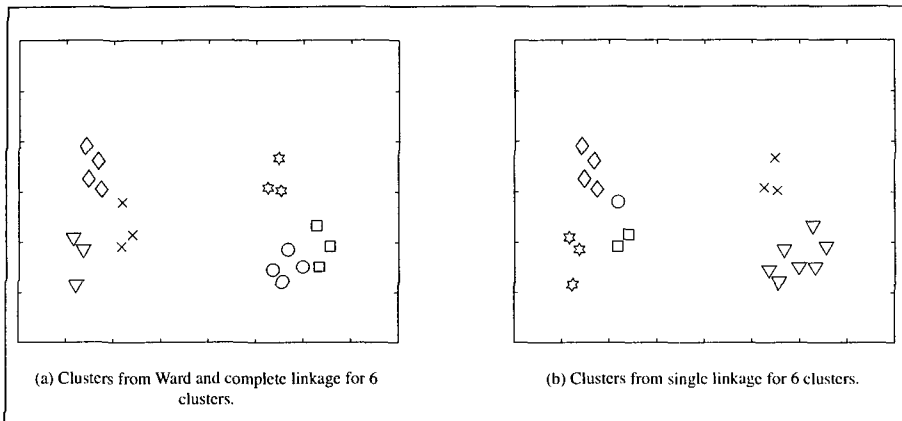
Fig. 3.3: Original 20 sample points clustered with ward and complete linkage compared to single linkage hierarchical clustering, merging each in each case until 6 clusters are left.

(a) Clusters from Ward and single linkage for 4 clusters.

(b) Clusters from complete linkage for 4 clusters.
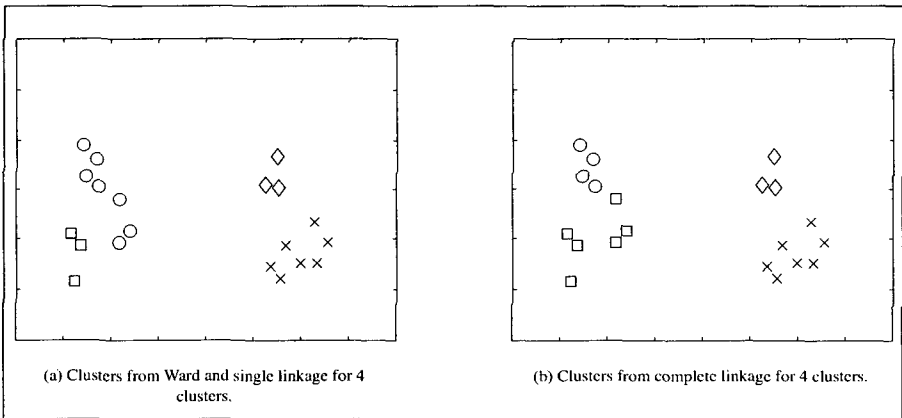
Fig. 3.4: Original 20 sample points clustered with ward and simple linkage compared to complete linkage hierarchical clustering, merging each in each case until 4 clusters are left.

3.4.2   Mutual Neighbourhood Value Hierarchical Clustering

This is a computationally efficient method of unsupervised multidimensional hierarchical clustering proposed by Dugad and Ahuja[29] that is deterministic as well as able to detect non-spherical, non-compact and density based clusters. This means it is independent of the order points are presented, does not require starting centroids and will find, for example, donut shaped clusters.

The algorithm considers the *mutual neighbourhood value* (MNV) between two points. If P and Q are two points in a given data set. If P is the mth nearest neighbour of Q and Q is the nth nearest neighbour of P, then the MNV between P and Q is defined as $m + n$.

A parameter M_T is used to decide if points are neighbours using the following conditions:

1. $mnv(Q,P) \leq M_T$
2. There exists no point K such that $mnv(Q,K) < mnv(Q,P)$ but $d(Q,K) \geq d(Q,P)$. A point P violating this constraint is called *invalid* with respect to Q.
3. Q is not to be invalid with respect to P.

All points Q such that $mnv(P,Q) \leq M_T$ are said to belong to the neighbourhood of P.

The method was implemented in c++ for speed using the algorithm as described in [29]:

1. Sort the distance pairs
2. Find the mnvs
3. Sort the mnvs
4. Make neighbours
5. Find connected components
6. Postprocess

By changing the value of M_T and graphing the results the natural number of clusters in the data set can be identified by "flat" regions in the graph, as illustrated in Fig. 3.5.

The MNV clustering algorithm is sensitive to both the density and the distribution of points.

3.4.3 K-Means Clustering

This is a common partition clustering method that tries to create an *optimum* partioning of the set of data points into K clusters. There are a number of variations of this method (K-Medoid for example) but the K-Means will be described here since it is important in understanding the final method chosen (see 3.4.4). This is the method used by the SPEA to archive solutions, and so is worth describing here. The technique is based on the idea that a cluster can be represented by a central point, that in most cases will not be an existing data point. The basic algorithm is :

1. Select K points as the initial centroids.

(a) Example set of points clustered with MNV clustering.

(b) Changing the value for M_T shows up regions where clusters stay stable.
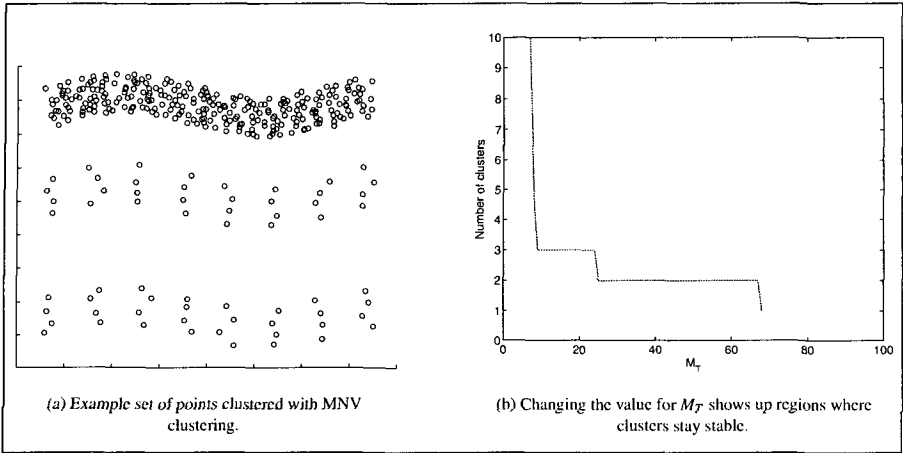
Fig. 3.5: MNV Clustering applied to an example problem.

2. Assign all points to the closest centroid.

3. Recalculate the centroid of each cluster.

4. Repeat steps 2 and 3 until the centroids no longer change.

This process will always converge, but in general will converge to a local optimum. The number of iterations required to converge is typically small (5-10 iterations[58]). The process minimises the sum of the squared distances of each point from the "center" of the cluster.

$$Error = \sum_{i=1}^{K} \sum_{\vec{x} \in C_i} |\vec{x} - \vec{c}_i|^2 = \sum_{i=1}^{K} \sum_{\vec{x} \in C_i} \sum_{j=1}^{d} (x_j - c_{ij})^2 \tag{3.6}$$

By differentiating the *Error* and setting it to zero, the minimum for a cluster p can be shown to be ((3.7))

$$\vec{c}_p = \frac{1}{n_p} \sum_{\vec{x} \in C_p} \vec{x} \tag{3.7}$$

the mean of the points in the cluster C_p.

The biggest difficulty with this algorithm is the choice of the K initial starting centres. Although the simplest method is to randomly choose the K centres, if they are not spread satisfactorily across the clusters some may be missed. This also means that re-clustering with k-means may find different clusters.

K-means clustering is ecomonic in terms of computational effort but as with ward clustering tends to favor well separated, equally sized spherical clusters. It also has the advantage that it can be started from the existing cluster centres, greately improving cluster stability.

3.4.4 Fuzzy C-Means Clustering

In common with the K-means algorithm, the fuzzy c-means algorithm also requires initial cluster centroids.

The fuzzy clustering algorithm used was first proposed by Dunn[30] and Bezdek[8] and is an iterative process that minimizes the function :

$$J = \sum_{j=1}^{n} \sum_{i=1}^{K} \mu_{ij}^{m} \left| \left| x_j - v_i \right| \right|^2 \tag{3.8}$$

where n is the number of data points, K is the number of clusters, x_j is the jth data point, v_i is the ith cluster centre, μ_{ij} is the degree of membership of the jth data in the ith cluster, and m is a constant (fixed at 1.4 for all of this work).

The degree of membership is defined as

$$\mu_{ij} = \frac{1}{\sum_{k=1}^{K} \left(\frac{\left| \left| x_j - v_i \right| \right|}{\left| \left| x_j - v_k \right| \right|} \right)^{2/(m-1)}} \tag{3.9}$$

Given a number of clusters c and and the initial centres this will converge to a local minimum.

In practice this process is quite quick for a reasonable size (less than 300 individuals) population, and produces results equivalent to k-means clustering with the added information about membership*.

Cluster Center Estimation. In order to find reproducible and stable clusters it was necessary to choose the initial centroids in a better than random manner.

This was done using a technique proposed by Chiu[14]. This is a simple but effective algorithm for estimating the number and, more importantly, the initial location of cluster centers. The method is relatively quick and does not use an iterative non-linear optimisation, so is of linear complexity with the dimension but square with the number of points (requires the calculation of distances between all the points).

---

*This was originally thought to be of use elsewhere in the algorithm for choice of parents for preproduction, but the adequacy of the simpler strategy meant that this possibility has not been fully investigated.

The method operates on a normalised space and considers each data point as a potential cluster center, and defines a measure of the potential P_i of data point x_i, from the set of n data points $\{x_1, x_2, \ldots, x_n\}$ as :

$$P_i = \sum_{j=1}^{n} \exp^{-\alpha||x_i-x_j||^2} \qquad (3.10)$$

where

$$\alpha = \frac{4}{r_a^2} \qquad (3.11)$$

and where r_a is a positive constant, set to 0.25 in the current work. Using the square of the distance clearly reduces the work done.

After calculating the potential for each point the point x_1^* with the largest potential P_1^* value is chosen as the first cluster centre, and the potential of each remaining point updated as :

$$P_i = P_i - P_1^* \exp^{-\beta||x_i-x_1^*||^2} \qquad (3.12)$$

where

$$\beta = \frac{4}{r_b^2} \qquad (3.13)$$

and r_b is again a positive constant, set to $1.25 r_a$ as in [14].

The points near the first cluster center will have greatly reduced potential and will be unlikely to be chosen as the next cluster centre. This process is repeated until the remaining potential of all the data points are below a fraction of the potential of the first cluster center P_1*. A value of 0.15 was chosen as the fraction for this work as per Chiu [14], and in addition to this a stopping criterion acceptance and rejection criteria were applied as explained in Chiu [14].

Determining the number of clusters.   A method of determining the most valid number of clusters was still needed and a survey of the literature produced several methods for use with fuzzy clustering [101,8,80].

Several of these methods were implemented and the most stable was found to be a simple ratio of compactness and separation, S proposed by [101] and defined as:

$$S = \frac{\sum_{i=1}^{c} \sum_{j=1}^{n} \mu_{ij}^2 ||v_i - x_j||^2}{n \min_{i,j} ||v_i - v_j||^2} \qquad (3.14)$$

It should be noted that $||v_i - x_j||^2$ is the Euclidean distance between the cluster centre and each point, and

this also suffers when the number of dimensions increases.

In order to identify the correct number of clusters the population was clustered with number of clusters c taken from 2 to c_{max} (fixed at 12 for this work). The validity criterion S was then calculated for each c. The minimum value of S indicates the most likely number of clusters.

3.5 Comparison of Clustering Stability

Three of the most promising clustering methods were evaluated in a simple test to determine the stability of the clusters when adding points:

- Ward linkage hierarchical clustering

- MNV clustering

- c-means fuzzy clustering

The c-means algorithm was chosen in favor of the kmeans simply because the added membership information for each point in each cluster was considered of potential value in the selection strategy of the CPEA*.

The initial set of 20 sample points was clustered with each of the methods, then three additional points were added, one at a time, re-clustering after each addition.

Fig. 3.6,Fig. 3.7 and Fig. 3.8 show the initial clusters and the subsequent new clusters after each point has been added, for each of the algorithms.

It should be remembered that the goal of introducing clustering in the CPEA was twofold — to preserve diversity, allowing solutions in "difficult" areas to develop with a degree of protection, and to find multiple different solutions. For this it is better to favour more clusters rather than less, provided that the clusters remain stable.

By these criteria both the Ward and the c-means clustering algorithm produced acceptable, although different, results.

However, the MNV algorithm, despite being promising on paper, demonstrated a lack of stability of the clusters, and it proved difficult to establish the likely number of clusters that were present.

If clusters are periodically merged because of the addition of a new point the CPEA will rank them together and remove the dominated points, never allowing "weaker" clusters the chance to develop. In problems

---

*The kmeans algorithm in conjunction with a cluster validity measure based on the ratio of inter to intra cluster distance would be expected to produce the same clusters as the cmeans algorithm and S measure used here.
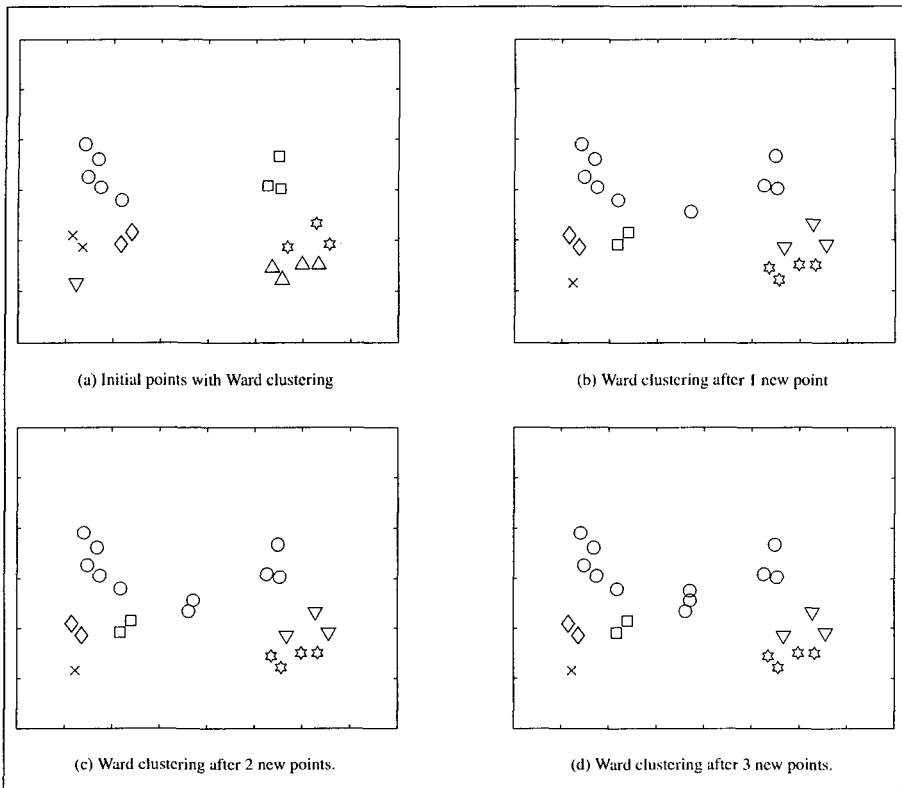
Fig. 3.6: Original 20 sample points clustered with Ward hierarchical clustering, followed by the impact of adding new points one at a time and reclustering.

where the eventual best solutions are difficult to find this would be expected to result in very poor performance, and in the problems with multiple local pareto optimal fronts corresponding to clusters would result in instability of the fronts and probable loss of some altogether.

Examining Fig. 3.7(b) and Fig. 3.7(c) it can be seen that part of the large cluster on the left of the figure is merged with the cluster on the right when the first new point is added to the set, but then separated again when the the second point is added. Addition of the third point result in a separation into three clusters.

Inspection would suggest that this is not desirable, and indeed on later tests on the Himmelbau problem (see section 4.1.3 below) the MNV algorithm did not find all four clusters. Periodic insertion of a new point led to the merging of clusters and subsequent deletion of the weaker, local optimum altogether.

(a) Initial points with MNV clustering

(b) MNV clustering after 1 new point.

(c) MNV clustering after 2 new points.

(d) MNV clustering after 3 new points.
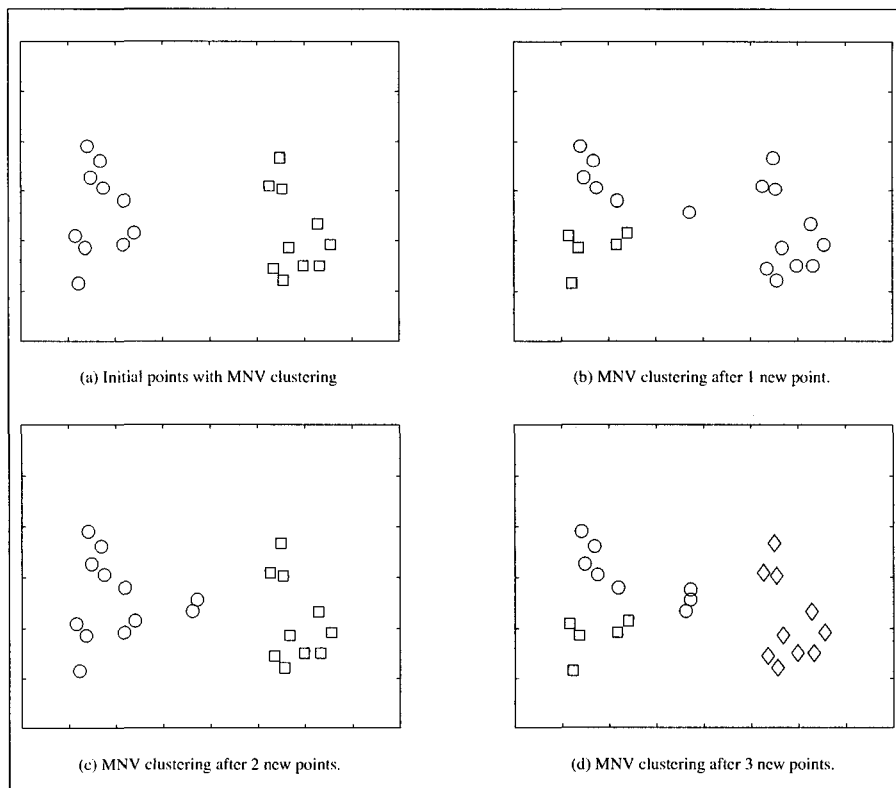
Fig. 3.7: Original 20 sample points clustered with MNV clustering, followed by the impact of adding new points one at a time and reclustering.

(a) Initial points with C-means clustering

(b) C-Means clustering after 1 new point

(c) C-means clustering after 2 new points

(d) C-means clustering after 3 new points.
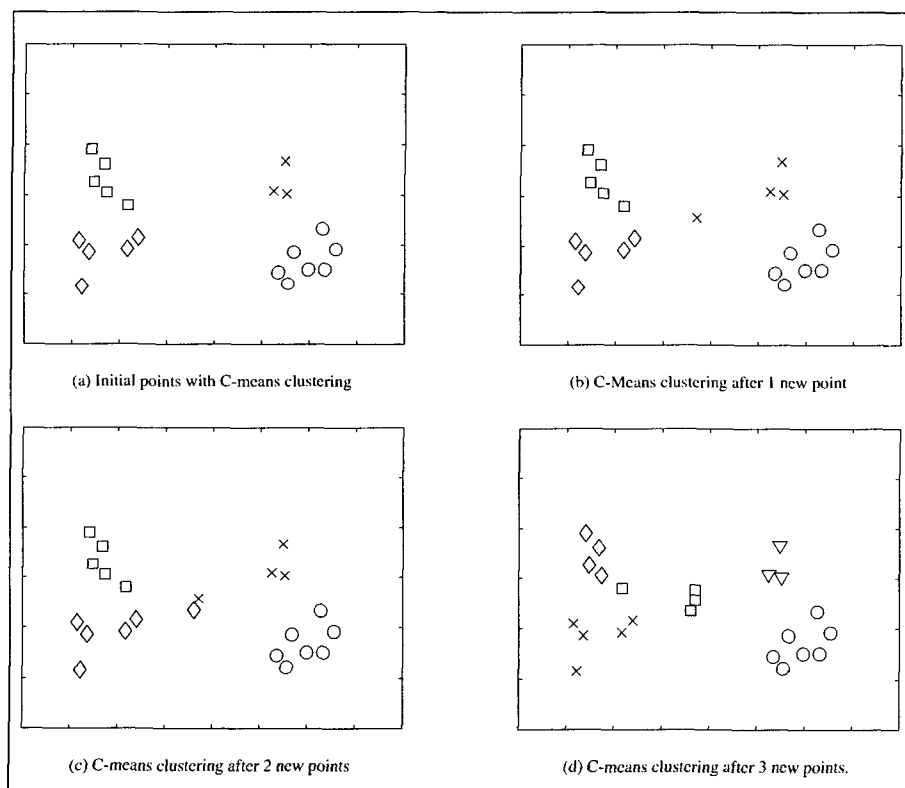
Fig. 3.8: Original 20 sample points clustered with Fuzzy C-means clustering, followed by the impact of adding new points one at a time and reclustering.

Chapter 4

Test Problems

4.1 Introduction

An earlier version of the CPEA without thinning was applied to several test problems and the results presented in Molyneaux et al.[66].

The CPEA *without* clustering has since been applied to more test problems taken from the literature[103,104,25] in order to compare in more detail its performance with existing MOO algorithms. In particular the tests used by Zitzler[103] in his thesis (referred to as T1–T6) were used heavily since the results for the SPEA were available online[†], and because it is among the best algorithms currently published.

The results of these tests[61] show that the CPEA outperforms the SPEA, both in speed and in convergence to the Pareto front. A speed increase of a factor of > 4 was achieved, conservatively measured as the mean number of evaluations for the worst point in the CPEA population to be better than the best point in the SPEA population. These preliminary results would suggest that the CPEA is clearly better than the current existing algorithms on these types of problem, which do not inherently require clustering to solve.

Most of the test problems found in the literature deal with problems of finding a well distributed solution set approximating the global Pareto optimal front, and introduce difficulties such as deception, multiple locally non-dominated fronts, an uneven distribution of solutions along the Pareto front or a bias towards solution density near a local optima.

In order to test the clustering capability of the CPEA specifically multi-modal problems were needed.

The following tests were used to investigate the different aspects of the CPEA.

[†]http://www.tik.ee.ethz.ch/ zitzler/

1. **Schaffers 1D** [82] - This is a simple example to demonstrate the basic behaviour of the CPEA.

2. **Sin-Cos** - A simple one dimensional, two objective problem that was constructed to demonstrate the behaviour of the CPEA with multiple non dominated fronts corresponding to multiple local minima in each objective.

3. **Himmelbau** - A two objective problem constructed from the Himmelbau function, used previously to test single objective niching and multi-modal algorithms [64,45]. This was used extensively to develop and test different clustering techniques.

4. **Deb's bi-modal** - A difficult two objective problem from Deb [25] used to compare the multi-modal capability of the clustering technique to the existing MOO methods. This was also treated by Andersson et al. [3].

5. **Zitler's T4** - A difficult problem with 10^9 local optima and 10 variables, already used to compare the performance of the CPEA without clustering [61]. This was chosen in order to test the dimension reduction technique, and to observe the effect of clustering on performance on a difficult problem.

The test problems were limited to one and two objective minimisations, both to aid visualisation and because very few three or more objective test problems have been treated in the literature.

The following sections discuss these problems in more detail and present the optimisation results.

All of the tests were run with blend crossover and real normal mutation unless specifically stated otherwise.

4.1.1   Schaffers 1D - Single Variable, Single Pareto Front

This is a simple one dimensional test with 2 objectives, as proposed by Schaffer [82] and used in [48,103,84]. The problem is :

$$\begin{aligned} \text{Minimise} \quad f_1 &= x^2 \\ \text{Minimise} \quad f_2 &= (x-2)^2 \end{aligned} \tag{4.1}$$

within the domain $-1000 < x < 1000$.

Parameters used were: $p_{initial}=10$, $p_{min\_cluster\_size}=10$, $p_{max\_cluster\_size}=70$, $n_{clusters}=1$, hence no clustering.

The results from a typical run after 1000 evaluations are shown in Fig. 4.1. Fig. 4.1(a) shows the distribution of points and the corresponding f_1 and f_2 values. Fig. 4.1(b) shows the Pareto front in objective space.

(a) f_1 and f_2 vs x (b) The Pareto front and the final population.
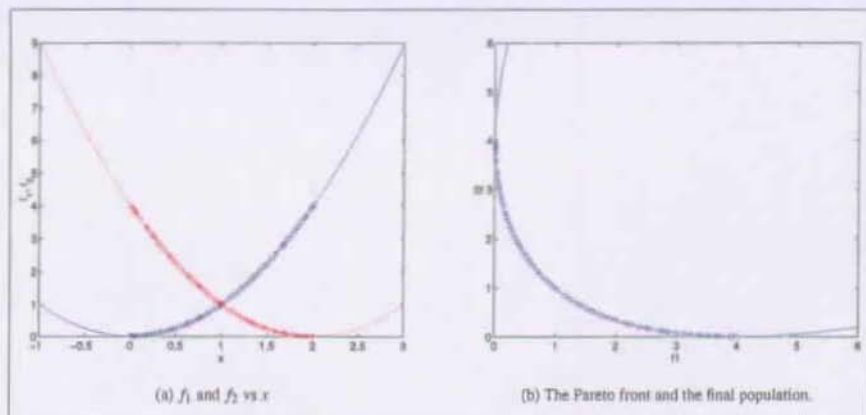
Fig. 4.1: Schaffer's one dimensional problem.

4.1.2 Sin-Cos - Single Variable, Multiple Local Pareto Fronts

To demonstrate the local optima preserving features of the algorithm, the one dimensional function below was constructed.

$$\text{Minimise} f_1 = \sin(x) * (1 * x/20) \tag{4.2}$$
$$\text{Minimise} f_2 = \cos(x) * (1 * x/20)$$

This function has three local Pareto-optimal regions in the domain $0 < x < 20$, the rightmost completely dominating the other two. The CPEA was run with the parameters: $p_{initial}=10$, $p_{min\_cluster\_size}=10$, $p_{max\_cluster\_size}=25$, $n_{clusters}=3$, c-means clustering.

The results from a typical run after 2000 evaluations are shown in Fig. 4.2.

All three local Pareto-optimal frontiers are well defined after only 250 evaluations and remain stable even after 20000 evaluations - the dominant frontier does not overwhelm the other two.

If the CPEA is run without clustering, only the dominant, global Pareto front is found and maintained.

To examine the behaviour of the CPEA with automatic detection of the number of clusters the same problem was re-run letting the c-means validity measure determine the number of clusters as described in section 3.4.4.

This produced results identical to those produced by setting a fixed number of cluster, with $n_{clusters}=4$. The

(a) Final population in variable space

(b) Final population in objective space
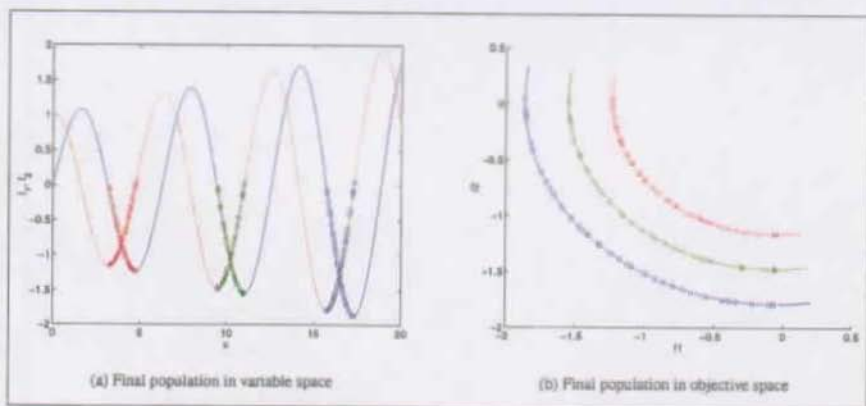
Fig. 4.2: The Sin-Cos Function

natural clusters in the data were identified, but since CPEA algorithm deliberately requests one more cluster than is apparent in the data one of the clusters is split into two. This was consistently found to be the middle of the three local optima.

The results show the same three locally non-dominated fronts.

4.1.3   Himmelbau function - Two dimensional Multi-Modal

A problem was constructed from the Himmelbau function as:

$$f_1 = \frac{(x_1^2 + x_2 - 11)^2 + (x_1 + x_2^2 - 7)^2}{200} - 5 \qquad (4.3)$$

$$f_2 = \frac{(4x_1^2 + 2x_2 - 11)^2 + (2x_1 + 4x_2^2 - 7)^2}{200} - 5 \qquad (4.4)$$

in the interval $-50 \le x \le 50$.

The parameters used were: $p_{lmini} = 10$, $p_{min.cluster.size} = 10$, $p_{max.cluster.size} = 25$. The problem was run with fixed cluster number, $n_{clusters} = 4$, and also with automatic clustering.

The problem has three locally optimal regions and one global optimum, and was used extensively to test the CPEA with clustering.

(a) Surface plot of f_1

(b) Surface plot of f_2.

(c) Contour plot of f_1.

(d) Contour plot of f_2.
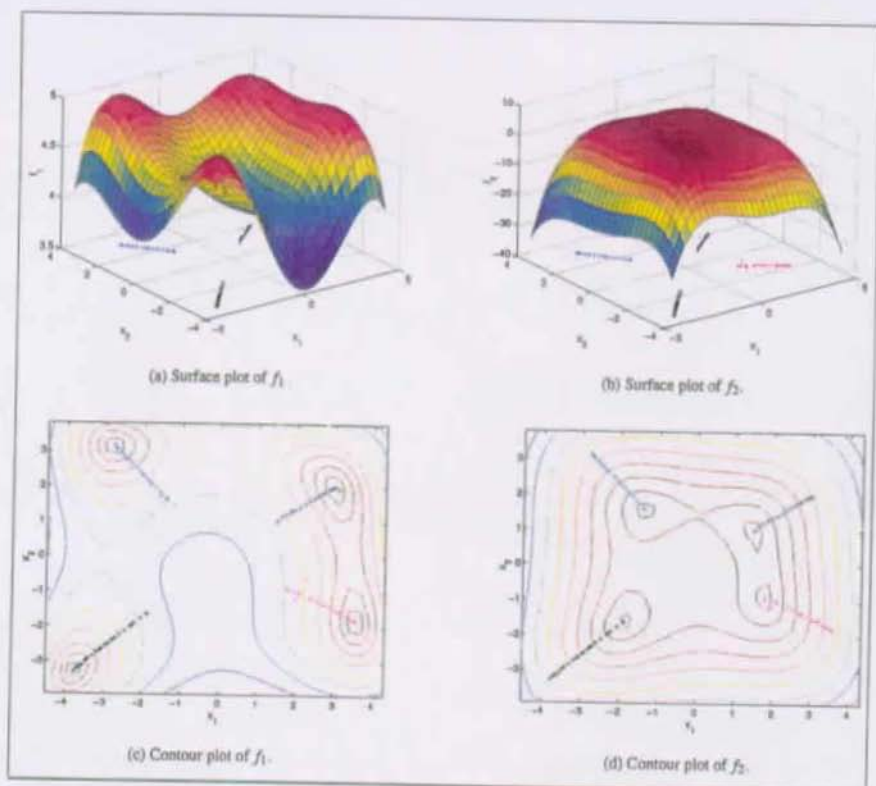
Fig. 4.3: Results from a typical run of the Himmelbau multi-modal function. Clusters are shown in different point styles and colours.

The population after $10,000$ function evaluations, with fixed cluster number and c-means custering is shown in Fig. 4.3, and the resulting objective space in Fig. 4.4.

The test was repeated with MNV clustering, again with fixed number of clusters, and this proved unable to find all four optimal regions. Periodically the four optimal regions would be found and then lost again as two clusters "merged" because of an intermediate point. This confirmed the behaviour found in the earlier simple clustering tests.

The Ward hierarchical clustering behaved very similarly to the c-means clustering when the number of clusters was fixed, but required a higher computational workload. With a consistency measure to automatically identify clusters the Ward clustering failed to find all four clusters or conversely found many clusters,
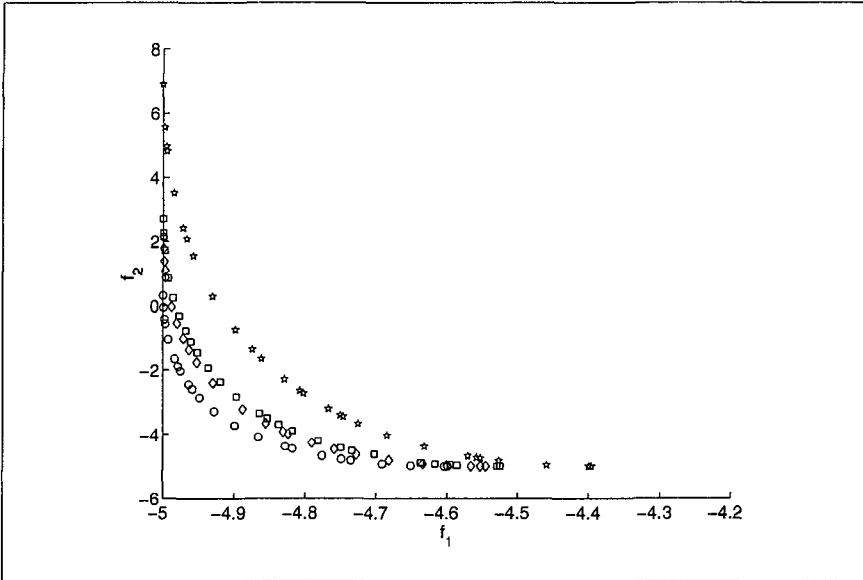
Fig. 4.4: Multi-modal Himmelbau function. Pareto fronts for each of the cluster. Clusters are shown in different point styles.

depending on the setting of the consistency parameter.

Tests were carried out with no scaling before clustering, with normalisation by standard deviation and with normalisation by maximum range and no significant difference was observed in the behaviour or the CPEA for this problem.

Without clustering the CPEA consistently finds the global Pareto front* but evidently does not find the local Pareto fronts in objective space.

---

*When run without clustering the population size was increased to 70.

4.1.4 Deb's bi-modal Test — Two dimensional Multi-Modal

This is bimodal problem defined using Deb[25]'s notation as:

$$\text{minimise } f_1(x_1, x_2) = x_1 \tag{4.5}$$

$$\text{minimise } f_2(x_1, x_2) = \frac{g(x_2)}{x_1}, g(x_2) > 0, x_1 > 0 \tag{4.6}$$

where:

$$g(x_2) = 2 - \exp\left\{ -\left(\frac{x_2 - 0.2}{0.004}\right)^2 \right\} - 0.8\exp\left\{ -\left(\frac{x_2 - 0.6}{0.4}\right)^2 \right\} \tag{4.7}$$

and $0 \le x_2 \le 1$ and $0 \le x_1 \le 1$. The function g is shown in Fig. 4.5(a), and has a global optima at $x_2 = 0.2$ and a local optima at $x_2 = 0.6$. The function is biased—the solution density is much higher near and above the local optima, as shown in a Fig. 4.5(b).

The function is expected to test the ability of the CPEA to keep the two distinct solution areas as well as consistently identify the two Pareto fronts on a difficult problem.

(a) The function $g(x_2)$ has a global and local minimum

(b) A random set of 10000 solutions showing the low solution density at the global Pareto optimal front.
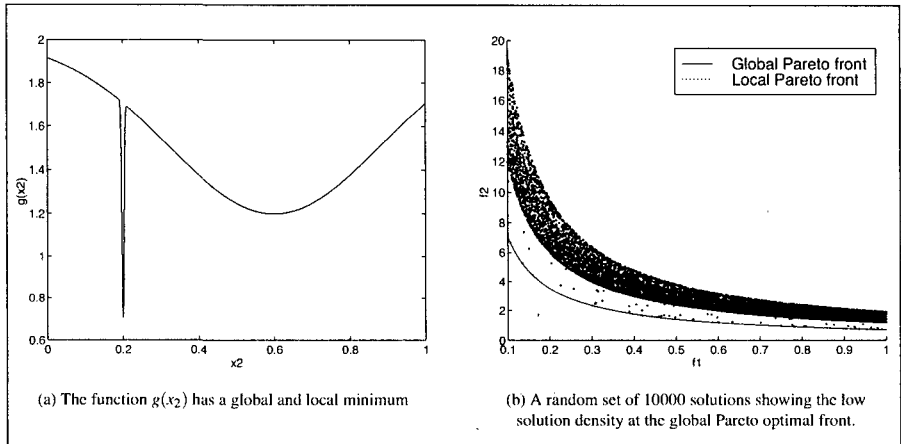
Fig. 4.5: Debs bi-modal test problem.

The parameters used were: $p_{initial}=60$, $p_{min\_cluster\_size}=10$, $p_{max\_cluster\_size}=25$, and the maximum number of evaluations limited to 10,000 (comparable to the 12,000 evaluations used in Andersson and Wallace[4], Andersson et al.[3]). The number of clusters was decided automatically using the c-means validity measure, and the clustering was done by the c-means algorithm.

The results from a typical run after 10,000 evaluations are shown in Fig. 4.6(a) and Fig. 4.6(b), and show very good coverage, as well as very good convergence.

This problem was also used to investigate the effect of objective scaling and crossover/mutation operators. The optimisation was repeated 80 times for each combination of:

- scaling - normalisation with standard deviation or no scaling.

- crossover - blend crossover+uniform mutation, blend crossover alone, and uniform crossover+uniform mutation.

Normalising Objectives by Std. Dev.   When run with uniform crossover and mutation, and with blend crossover with mutation, both the local non-dominated front and the global Pareto front are found quickly and reliably—80 out of 80 tests.

However, with blend crossover and no mutation the solution was only found in 65 out of 80.

No Objective Scaling.   This produced very similar results as scaling—uniform crossover with mutation found both fronts in all tests, blend crossover with mutation found both fronts in 77 out of 80, the remaining times being blocked at the local front. With blend crossover and no mutation the solution only found the two fronts 57 out of 80.

This would suggest that scaling the objectives may be advantageous before clustering, although this is probably problem related.

Once the two non-dominated fronts were found they remained stable when left to run for 100,000 evaluations.

In contrast with these results Deb[25] reported that the NSGA found the global Pareto front in only 41 out of 100 trials. Andersson and Wallace[4] similarly observed that the MOSGA also found only one of the two Pareto fronts per run when using a replacement strategy based on similarity in the objective space, but that when including both objective and variable space in the similarity metric both fronts could be found.

This problem can equally well be solved with clustering on only one dimension, x_2, since the ranking and thinning algorithms promote the distribution of solutions along the non-dominated fronts (which are aligned on x_1).
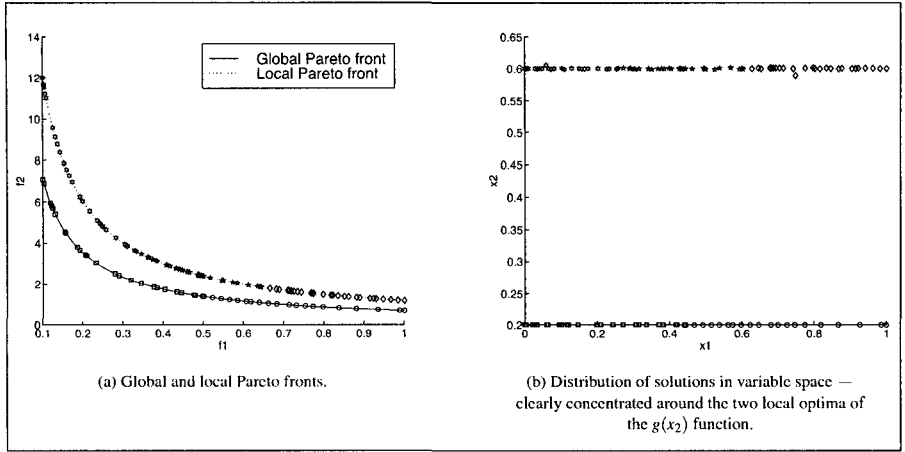
Fig. 4.6: Debs bi-modal test problem. Results of the optimisation with automatic clustering.

4.1.5 Zitzler's T4 — Two objective Multi-modal

This was one of the tests used by Leyland et al.[61] to establish the performance of the CPEA relative to the other MOOs in the literature. It has been used here to allow a comparison of the performance of the CPEA with and without clustering, and to examine the behaviour of the dimensional reduction technique, since the problem is posed with 10 variables.

$$\text{minimise } f_1(x_1, x_2) = x_1 \tag{4.8}$$

$$\text{minimise } f_2(x_1, x_2) = g(x_2, \ldots, x_n).h(f_1, g) \tag{4.9}$$

where:

$$g(x_2, \ldots, x_n) = 1 + 10(n-1) + \sum_{i=2}^{n} (x_i^2 - 10\cos(4\pi x_i)) \tag{4.10}$$

$$h(f_1, g) = 1 - \sqrt{\frac{f_1}{g}} \tag{4.11}$$

and $n = 10$, $0 \leq x_1 \leq 1$, and $-5 \leq x_2, \ldots, x_n \leq 5$. The global Pareto front is formed with $g = 1$ and the best local front with $g = 1.25$. Not all the local Pareto optimal fronts are distinguishable in the objective space. The function g is 1 when all of $x_2, \ldots, x_n$ are 0, but there are an infinite number of solutions that give $g = 1.25$.

Initially the problem was run with the parameters: $p_{initial}=100$, $p_{min\_cluster\_size}=10$, $p_{max\_cluster\_size}=25$, $n_{clusters}4$ with the c-means clustering algorithm.

The results from this were not too encouraging—neither the global nor the best local optima were found, either with or without dimensional reduction before clustering.

Typically, after 150,000 evaluations four local optima (not including the best local optima) were found as illustrated in Fig. 4.7. The local fronts found, however, were fairly well defined.
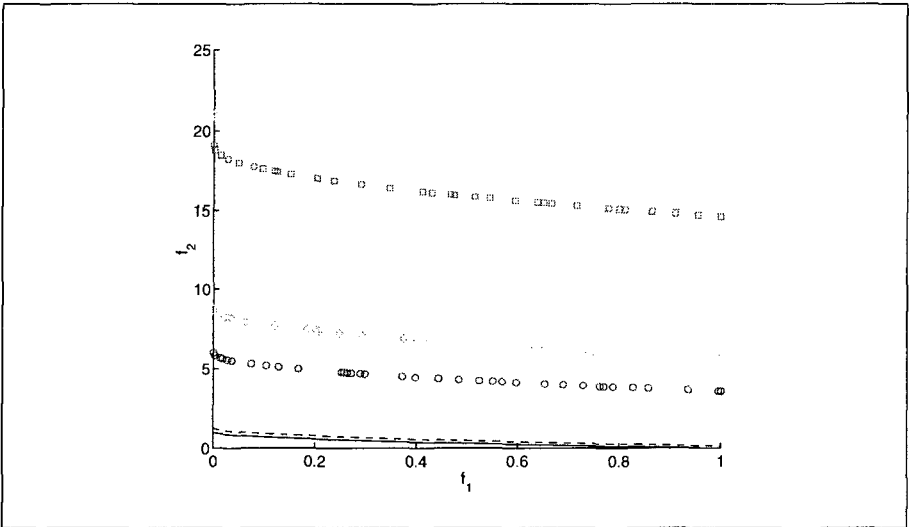
Fig. 4.7: Test results for T4 with clustering and blend crossover after 100,000 evaluations. Neither the local nor global front is found, although several other local fronts seem well defined.

In the earlier comparison work[61] without clustering a difference had been observed between blend crossover and uniform crossover on this test problem. Based on results of 100 repetitions, the uniform crossover with mutation took an average 6537 evaluations (std. dev 2285) to find the Pareto front, in 100/100 trials. In contrast the same test with blend crossover and mutation took on average 12,432 evalations (std. dev 4100) — nearly twice as long, although the Pareto front was still found.

With this in mind the crossover operator was changed to uniform crossover with mutation and the clustering problem re-run 100 times, 10,000 evaluations each time. Typical results from this second series of optimisations are shown in Fig. 4.8. The Pareto front was found in all of the tests, on average after 8031 evaluations (std. dev 2057), and the second best non-domindated front in 90% of the solutions. Repeating the tests with dimension reduction set to reduce the number of variables for clustering from 10 to 2 made no significant difference. The two principle components for the final results from Fig. 4.8 are shown in Fig. 4.9.
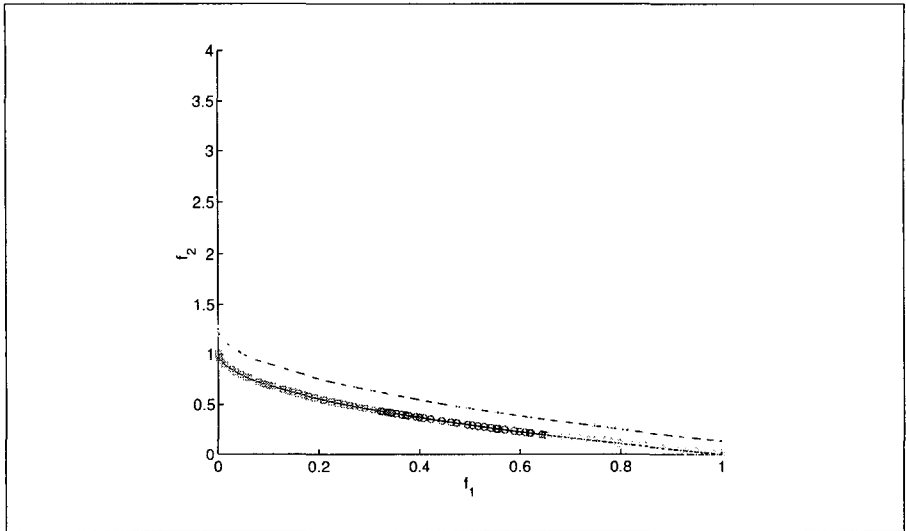
Fig. 4.8: Test results for T4 with clustering and uniform crossover after 10,000 evaluations. Only the global Pareto front is found.

However, when the optimisations were allowed to run for 100,000 evaluations, only the global Pareto front was kept in 90% of the runs.

The results were considered strange, and it was suspected that the the lack of variability in the domain (remember the Pareto front is found with $x_2 = 0, \ldots, x_n = 0$), and the lack of one fixed solution in variable space corresponding to the second best non-dominated front may be causing unstable clusters, hence the domination of all by the Pareto front.

The problems were re-run with no scaling of the variables before clustering.

This had a dramatic effect on the results. With uniform crossover the CPEA found the next best local optimum as well as finding the global Pareto front, as before in less than 10,000 evaluations, but now the second front was stable in more than 90% of the runs, even after 100,000 evaluations.

Using dimension reduction did not seem to affect this performance. However, the fact that the optimal solutions are found with most of the x_i values set to 0, and the second can be found by varying any one of these from 0, would seem to correspond with this behaviour, although it also highlights some problems with this test problem, that have been addressed recently by Leyland[63].
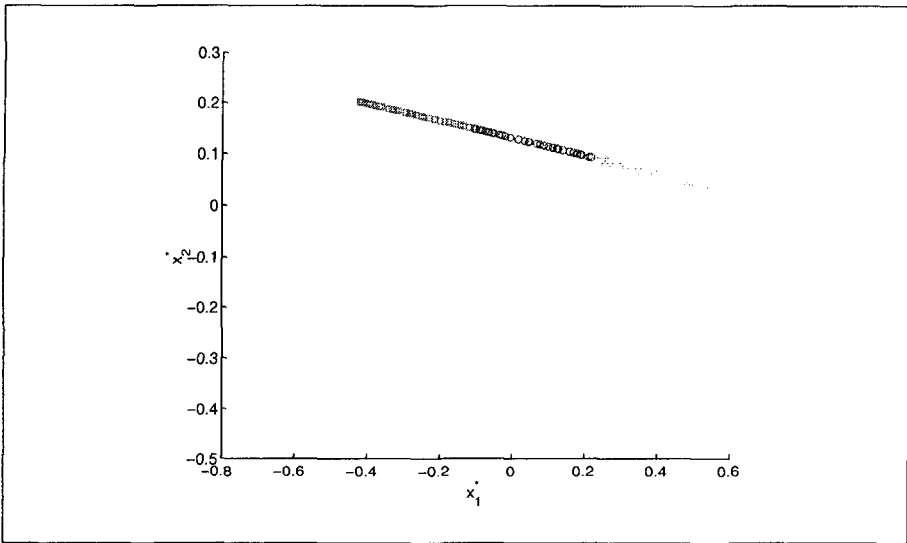
Fig. 4.9: Test results for T4 with clustering and uniform crossover after 10,000 evaluation showing the two principle components used to cluster.
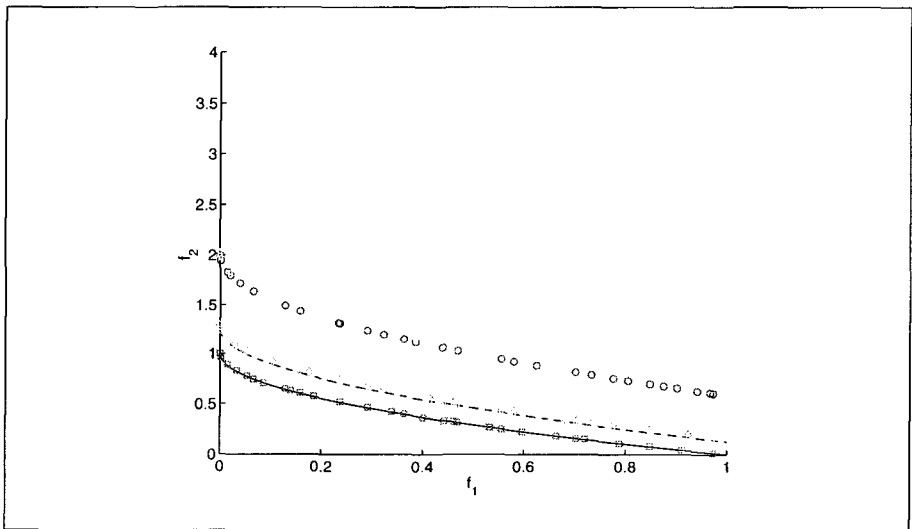
Fig. 4.10: Test results for T4 with clustering and maximum range scaling, and uniform crossover after 15,000 evaluations. Both the global Pareto front and the next best local front are found.
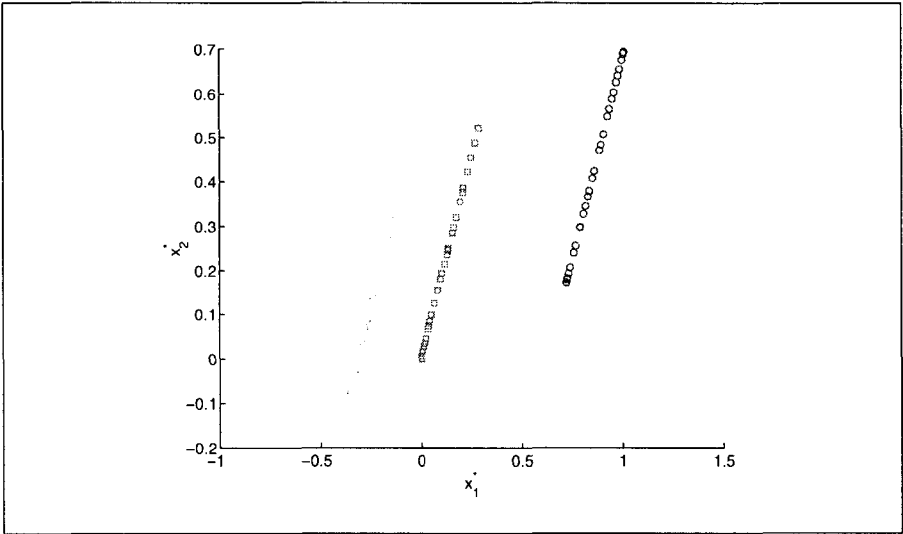
Fig. 4.11: Test results for T4 with clustering and maximum range scaling, with uniform crossover after 15,000 evaluations, showing the two principle components used to cluster.

4.2   Conclusions

The CPEA without clustering has previously[66,61] been shown to be an effective multi-objective optimisation algorithm on test problems from the literature.

In this chapter the use of clustering has been shown to be effective in preserving diversity and keeping multiple "different" solutions, where different refers to the variable space, and allowing multiple local non-dominated fronts to develop in objective space.

Several clustering techniques have been investigated and one of these, the c-means fuzzy clustering algorithm, has been found to be successful at identifying and keeping well defined, stable clusters. A technique for identifying "natural" clusters has been tested with c-means clustering and found to be adequate, although the tendency is to produce too many clusters, and adds a computational overhead to the clustering process. The use of this automatic clustering does not, however, seem to affect the number of evaluations required to solve a problem.

A multi-dimensional scaling technique has been used to reduce the number of dimensions before clustering, and this seems to have no adverse effects on the convergence time in the problems tested, although this should be examined in more detail in the future, and on the test applications.

The scaling used before clustering was found to have a dramatic affect on the convergence time and on the eventual solutions found with Zitzler's T4. Normalising with the standard deviation was found to reduce time to convergence but resulted in no secondary optima being found. Normalising with the current maximum range resulted in slower convergence but led to eventual results with multiple locally optimal fronts.

The use of clustering on problems that do not require it results in an increase in the number of evaluations required to converge to the global Pareto optimal front. In the tests performed here the increase was found to be a factor of approximately two.

A significant difference in the time to converge was observed when using uniform crossover instead of blend crossover, both with normal mutation. This was also observed in work on the non-clustering version of the CPEA[61]. It should, however, be noted that the test problems possess special characteristics favoring uniform crossover that are probably not present in real problems.

The custom built two objective problem based on the Himmelbau function is solved quickly. Use of the autoclustering does not hinder the evolution, and correctly identifies the clusters, although the process of determining the number of clusters is relatively time consuming*.

---

*It should not be forgotten that the test problems are trivial to calculate whereas generally the real-life problems will not be.

Things to look at in the future include:

- impact of ranking scheme.

- alternative thinning schemes.

- thinning in three dimensions.

- alternative crossover mechanisms - impact on convergence of uniform crossover instead of blend crossover.

- impact of scaling before clustering on convergence

- a two objective function similar to the Himmelbaus function but with locally non-dominated front that cross would be an interesting test case.

Chapter 5

Distributed Heating Network

5.1 Introduction

Heating represents an important fraction of modern society's energy needs and the replacement of existing heating systems with advanced, integrated solutions has the potential to significantly reduce environmental pollution. The continuing concern for energy efficient solutions has renewed interest in distributed heating and cooling networks and these were the subject of an earlier research project by Curti[21] at LENI.

The earlier work at LENI considered the methodology of designing a highly efficient urban heating systems using heat pumps and cogeneration units taking into account simultaneously the life cycle and environmental impact while meeting the required heating demand. Traditionally it has been the job of an experienced engineer to find a good solution, a process essentially of guided trial and error, that tends to favor conventional solutions. Curti's work improved upon this by calculating a total cost and minimising this with the Struggle GA.

This current work will take the process one step further by using multi-objective optimisation techniques to simultaneously optimise the separate cost and pollution criteria without combining them.

5.1.1 Aim of this Work

Using the same simulation and superstructure (described below) as in the earlier work the current project has several aims :

- reproduce the results of the previous work and demonstrate the reduced solution time required with the new multi-objective optimisation algorithm, the CPEA.

- investigate more throroughly the solution domain using a byproduct of the CPEA.

- demonstrate the advantages of post processing low level optimisation results, rather than incorporating uncertain cost parameters in the optimisation process (and hence requiring multiple optimisation runs to investigate the importance of these parameters).

- look in more detail at the interaction of pollution with system configuration and the tradeoffs between different technology types.

- look in detail at the importance of the cost of utilities (gas and electricity) with a three objective optimisation.

- demonstrate the need for, and effectiveness of, clustering in order to maintain solution diversity without introducing problem dependent heuristics.

The following section explains in detail the cases considered and presents results comparing the performance of the new algorithm on the original single objective problem, followed by results and discussion on the results from the CPEA.

5.2   Description of the District Heating Problem

Clearly, a district heating system must meet the demands of its clients. However, with a heat pump based central plant, adapting the delivery temperature to the most demanding client is frequently detrimental to the overall performance of the system when using centralised heat pump based plant. To avoid this a model was designed that could make use of decentralised heat pumps on the supply or return line of each user. Fig. 5.1 shows the overall schematic network layout.

The model *superconfiguration*, consisting of a central plant with a heat pump, a cogeneration system and an auxiliary furnace is shown in Fig. 5.2. The *users* are connected to the main network either with a heat exchanger or a local heat pump as shown in Figs. 5.3 and 5.4. The local heat exchangers and heat pumps may be connected either to the outbound or return lines introducing the potential for further reducing relative network costs.

The model considers thermodynamic, economic and environmental aspects associated with the entire life cycle of a distributed heating system, beginning with the manufacture of equipment and energy sources, continuing with operation and ending with equipment removal. Environmental characteristics of the system are internalised through the use of pollution factors that adjust the costs of damage due to pollutant emissions in construction, operation and decommissioning. Unfortunately the pollution factors depend in turn on constants that are either subjective or involve a high degree of uncertainty.
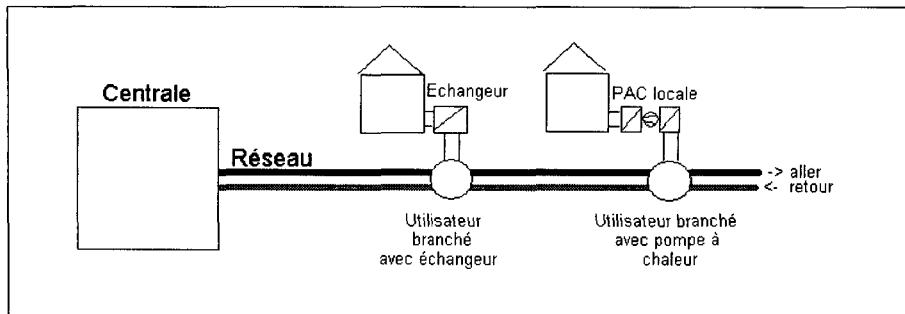
Fig. 5.1: Schematic layout of the overall network

5.2.1 Central Plant

The central plant superconfiguration consists of a combination of :

- a heat pump to upgrade heat from a low temperature heat source such as a lake

- a gas turbine cogeneration system to produce both heat and electricity

- an internal combustion gas engine cogeneration system producing both heat and electricity

- an auxiliary boiler directly producing heat.

To simplify the problem the network is arranged with users in series, with no branches, and is driven by an electric pump.

The pollution calculations are taken to include the pollution produced during the fabrication of the components, the preparation and transport of the primary fuel sources and the production of the electricity. Any mixture of pollution production during electricity may be introduced into the model but for consistency with Curti this chapter deals with the *Swiss mix*[*].

Electricity could be bought from the national grid, and produced by central plant for use internally or to supply users of the heating network but could not be sold back to the electricity grid.

5.2.2 Users

The users have a superconfiguration as shown in Fig. 5.3 for the intermediate users and Fig. 5.4 for the end user. They consist of:

---

[*]The European mixture was considered briefly and is presented in Appendix A.1.2.
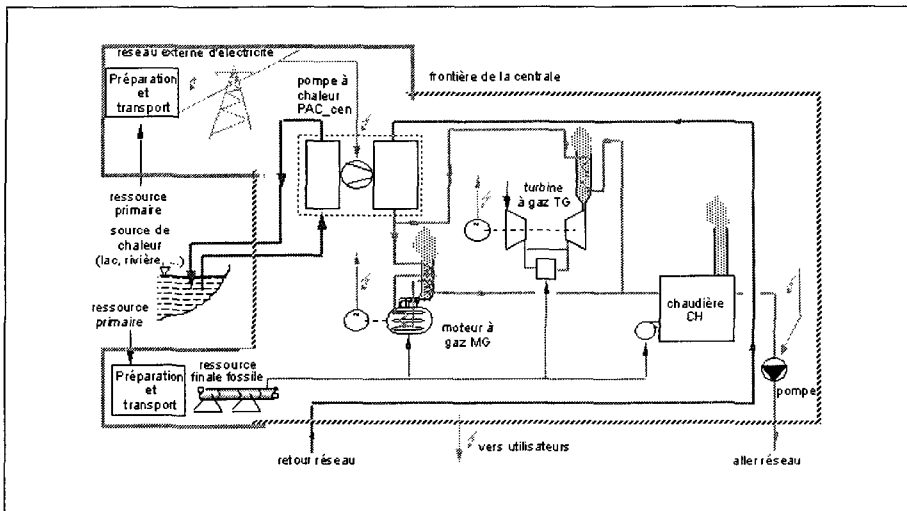
Fig. 5.2: The superconfiguration of components forming the central plant[21]

- a heat exchanger on the supply line

- a heat pump on the supply line

- a heat exchanger on the return line (except for the last user)

- a heat pump on the return line (except for the last user)

- two heat exchangers for domestic hot water at different temperature levels

- an auxiliary electric water heater for the domestic hot water demand that cannot be met by the hot water exchangers.

The last user is considered as a special case since the supply and return lines are effectively the same. Consequently the choice is limited to a single heat pump or heat exchanger for the heating demand as shown in Fig. 5.4.

5.2.3 Simulation

The network thermodynamic behaviour is solved for a nominal steady state operating regime - dynamic effects are not considered. Demand variation throughout the year is modeled in a simplified manner as an adjustment to the nominal operating regime.
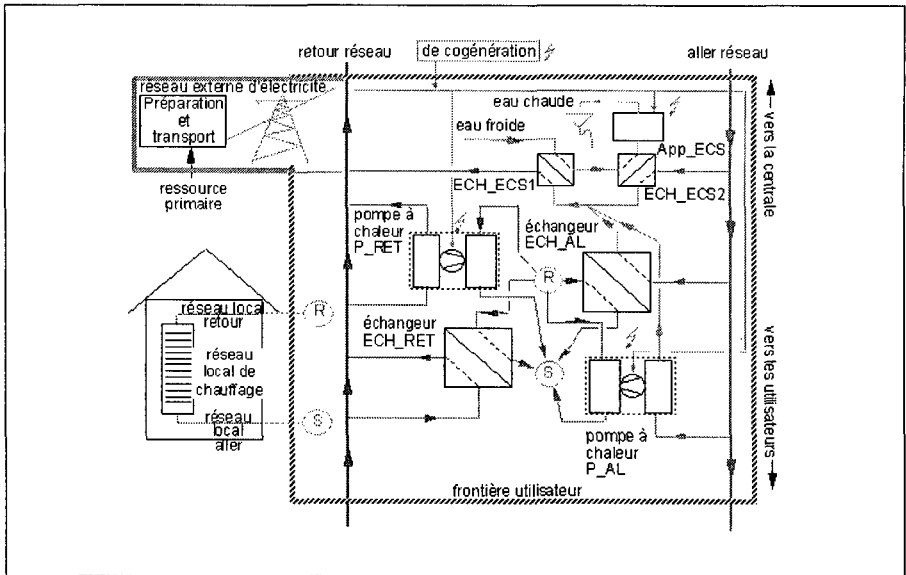
Fig. 5.3: The superconfiguration of components forming an intermediate user in the network[21]

The cases considered had four user *classes*, where a class represented a grouping of users with the same requirements.

The resulting model is both disjoint (different solutions can have different basic elements in the final configuration) and non linear (due to thermodynamic non-linearities and also non-linear cost and pollution factors) and in the cases considered has 37 independent variables (depending on the chosen optimisation options some of these may be inactive) and almost as many constraints. In addition there are a number of fixed parameters such as gas and electricity costs, lake temperature, etc.. The simulation was implemented in Fortran 77 independently of operating system and hardware, and a simple free format data file was used to set the fixed parameters and limits for the independent variables and constraints.

In the previous work the Struggle GA (see Chapter 2) was used to simultaneously optimise the configuration and operating conditions with the aim of minimising overall cost. Constraints were taken into account with penalty functions added to the overall cost.

While this approach allowed the inclusion of environmental costs the drawback was that the optimisation was performed for one chosen set of pollution factors and had to be repeated in order to investigate the sensitivity of proposed solutions to the pollution factors.
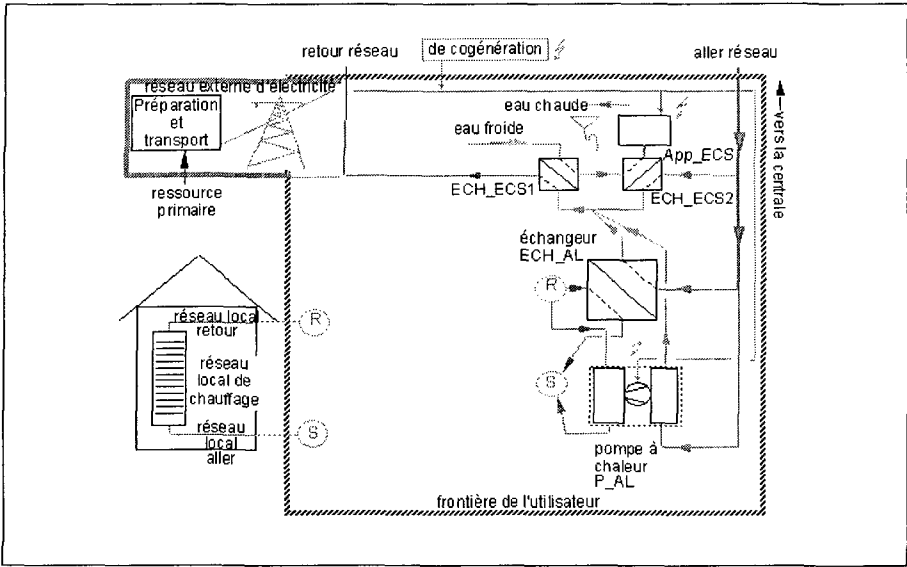
Fig. 5.4: The superconfiguration of components forming the last user in the network[21]

The approach taken in this work has been to make use of a multi-objective optimisation technique to op-
timise pollution costs, and investment plus running costs separately and then to post-process the results to
investigate the effect of pollution factors, allowing a comparison with the results found in Curti et al.[22].

In this Chapter the following notation will be used:

C_{total} refers to the overall cost including pollution.

$C_{pollution}$ refers to the sum of pollution costs.

C_{total}^- refers to the cost without pollution, $C_{total} - C_{pollution}$.

C_{elec} refers to the cost of electricity bought from the grid.

C_{fuel}^+ refers to the cost of gas and maintenaince.

5.2.4   Choice of Demand Case

In order to verify the original model, Curti[21] treated a reference case corresponding to an existing demand
in part of the city of Lausanne, and then considered a network corresponding to a new quarter in which
the majority of buildings would be equipped with (for example) under floor heating, typically working
at low temperature, while still including some older buildings and equipment with a higher temperature
requirement. This demand case was referred to as *demand B* in the original work and was chosen for the

current work.

The demand is for 62.7MW overall, with an initial user that takes 85% of the total power followed by 3 more classes that each take 5%. Fig. 5.5 shows this together with the imposed temperature requirements. The form of the demand case suggests strongly that the initial user will have a dramatic effect on the optimal solution and this does indeed prove to be the case.
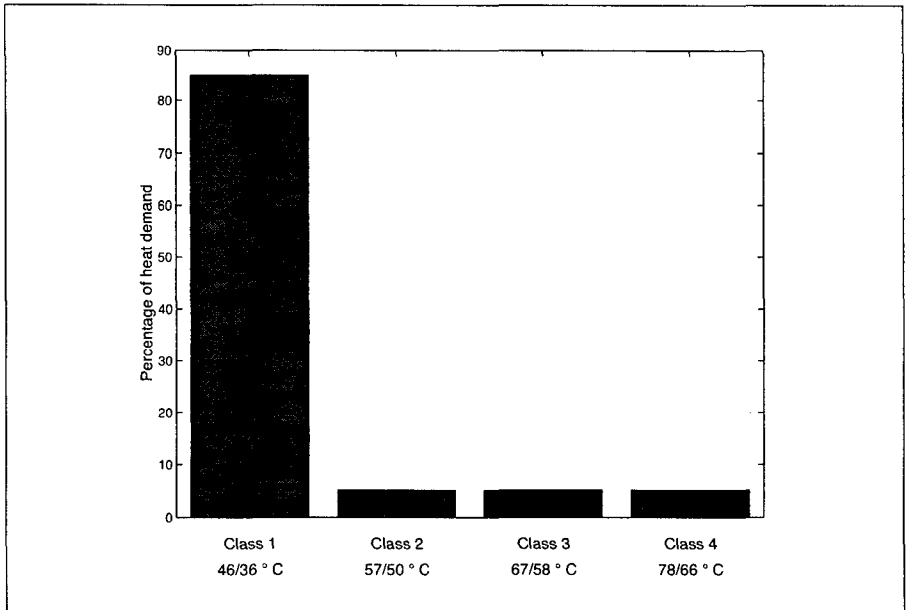
Fig. 5.5: Demand Case B: This represents a new installation in a new quarter with overall thermal require-mement of 62.7MW. User class 1 covers 85% of the total power, each of the others takes 5%.

5.3 Clustering Applied to the District Heating Problem

This problem is one of the cases which requires clustering in order to prevent the population converging to a very small subset of solutions very quickly. Once diversity has been reduced too far, and the domain is poorly represented in the population of solutions, new optima become difficult to find.

The clustering algorithm used in the CPEA was previously described in chapter 2, together with a method of multi-dimensional scaling to improve the clustering make the results more amenable to graphical inter-pretation.

Two approaches were used in this problem. Initially the clustering was performed on a reduced set of transformed independent variables using the method described in section 3.2, and fixing the number of transformed variables at two. This allowed simple visualisation of the clusters. After several initial optimisations using this technique a post processing of the distribution of the solutions showed that the dominant independent variable was clearly the network supply temperature.

In light of this the clustering was changed to use only the network temperature, economising the computational effort needed to calculate and apply the multi-dimensional scaling transform.

The results produced in this manner were largely the same as those with the multi-dimensional scaling of 37 variables (as was to be expected since the variable that accounted most strongly for the variance in the final population was temperature), and proved simpler to analyse as well as marginally faster to optimise[†].

On reflection, these observations are not so surprising. It is *nearly* possible to decompose the problem into two isolated sub-problems - one being the central plant and the other the users - since the only common independent variable is the network supply temperature. However, the matter is complicated because the electricity produced by the central plant cogeneration systems can be used in the remote heat pumps.

5.4   Single Objective Optimisation - The Original Problem

The original optimisation problem, to minimise overall cost, was repeated using the CPEA and the original internalized cost model for the two cases: with pollution and without pollution. The optimisation was run repeatedly and the best solution taken from each final population after 50,000 simulation evaluations. In fact it was observed that the algorithm converged to the best solution on average after 10,000 evaluations, as compared with the original work with the Struggle GA which took 400,000 evaluations (4,000 generations with a population of 100) to achieve comparable results. This is thought to be due to the better clustering technique and better thinning methods maintaining diversity while still applying convergence pressure[‡].

The optimisation was repeated ten times to verify the consistency of the solution and no alternative optimum solutions were found. In all cases the final results obtained were *better* (lower cost) or equal to those found previously and the number of evaluations reduced by a factor of approximately 40.

Without pollution costs.   The optimum solution for the case without pollution costs occurs at 89.5°C. The configuration of the central plant at this point is shown in Table 5.3 and the breakdown of costs as a percentage of total cost in Table 5.2. The central plant consists of a heat pump, a gas turbine and an auxilliary boiler. The high network temperature means that the hot water demand for the users may be met completely

---

[†]An optimisation for pollution cost and and total cost (see section 5.6) using dimension reduction is given in the Appendix A.1.1

[‡]It would be interesting to note the effect of using the MDS in the Struggle GA

with heat exchangers on the supply line. The heating demand may also be met by heat exchangers on the supply line for class 1,3 and 4, and by a heat exchanger on the return line for user class 2. The mass flow rate for users 3 and 4 is high enough to allow the second user to make use of the return line heat exchanger. The solution is comparable to that obtained in the earlier work as shown in Table 5.3.

With pollution costs. [§] The optimum solution for the case with pollution costs occurs at a network supply temperature of 52.5°C. The configuration of the central plant at this point is shown in Table 5.5 and the costs as a percentage of the total in Table 5.4. The central plant consists of a heat pump that supplies 100% of the heating demand. The network temperature is hot enough to allow a heat exchanger to be used on the supply line for the first (largest) user. The remaining users use heat pumps on the supply line to meet the heating requirements. The network is just hot enough to meet the hot water demand (45°C) with a heat exchanger[¶].

The behaviour of these optima with and without pollution costs is analysed in more detail in the following section.

5.5 Pseudo Multi-Objective Optimisation - Temperature and Cost

5.5.1 Introduction

During the development of the CPEA an idea was conceived which proved both interesting and useful. In problems where there is only one objective but the solution domain is complex (many problems, including the district heating problem as originally posed) it is very important to understand the sensitivity of the optimal solution to changes in one (or several) of the more important independent variables.

Traditionally a sensitivity analysis would be performed by fixing the chosen variable and performing an optimisation. Repeating this a series of optimum solutions, one for each fixed value may be found. This process can prove costly, and so the number of different values chosen will typically be limited.

However, the new idea was that rather than optimise a single objective, a multi-objective optimisation could be performed using one of the independent variables that is of interest as an **artificial** second objective.

Two optimisations need to be performed :

1. Minimise the objective function and minimise the chosen variable.

2. Minimise the objective function and maximise the chosen variable.

[§] Pollution costs used were 13.8 sfr/kg for NO_x and 0.03 sfr/kg for CO_2.

[¶] the hot water requirement for each user was fixed at 10% of the user's energy requirement, and at a temperature of 45°C .This assumes that once per week the water temperature will be raised to 60°C to avoid problems of legionella's disease

This produces the equivalent of a parametric study where the optimal solution is found for **many** fixed values of the independent variable, at the cost of only two optimisations - hence a dramatic saving in time and computational effort.

This works because in the first stage while minimising the independent variable the CPEA finds the part of the curve with a negative slope, the area to the left of the vertical line in Fig. 5.6. In the second optimisation the maximisation of the independent variable finds the positive sloping part, the area to the right of the vertical line in Fig. 5.6.

In the event that the function that describes the minimum values is concave with multiple minima then clustering is necessary to find various the various parts of the function.
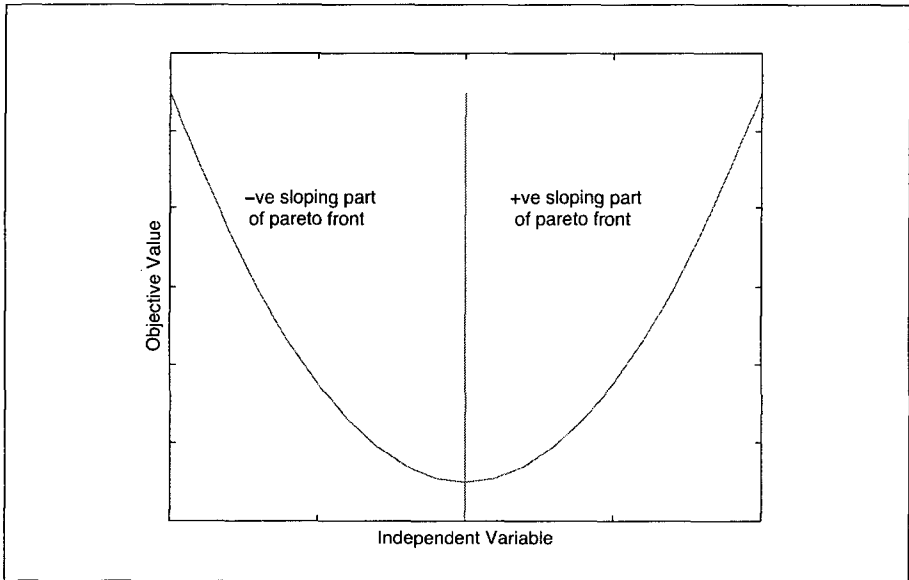
Fig. 5.6: Pseudo multi-objective optimisation. The negative part of the pareto front is found by the minimisation of the independent variable and the positive part by the maximisation.

5.5.2 Temperature and Cost Results

To demonstrate this idea the district heating model was optimised for overall cost taking network temperature as the artificial second objective. This was done for the case with pollution and without pollution, and to minimise and maximise network temperature, requiring a total of four runs.
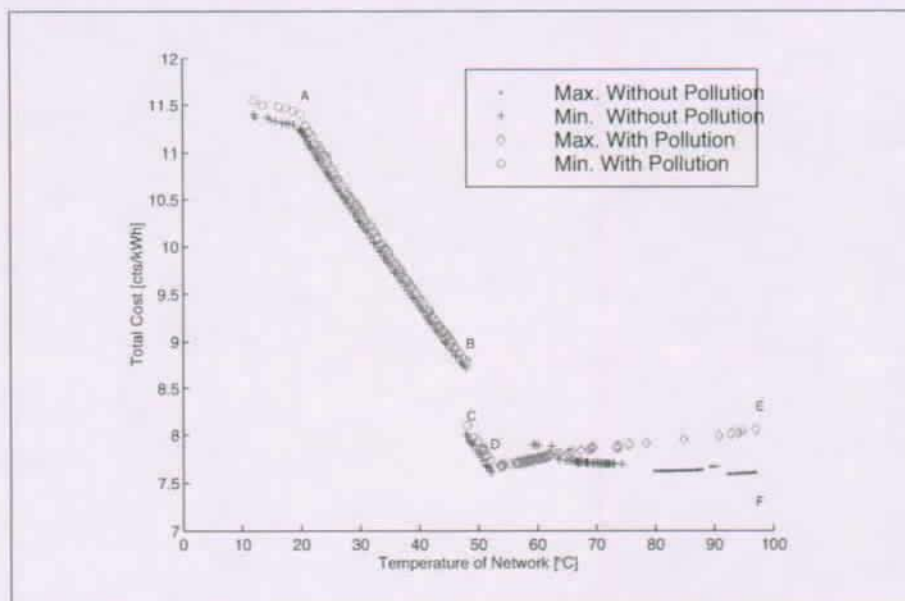
Fig. 5.7: Multiobjective Optimisation with Pseudo Objective Temperature and Cost

Fig. 5.7 presents the combined results for these four runs with several features labeled *A-E*, and Fig. 5.8, Fig. 5.9 and Fig. 5.10 show the configuration and costing breakdowns against network temperature. The characteristics of each region are as follows.

Left of A   To the left of the point labeled **A** on Fig. 5.7 is a region that corresponds to a network temperature less than 20°C. At this low temperature the hot water requirement of the users must be met completely by an electric water heater, since the inlet water temperature (12°C)is too close to the network temperature to allow the heat exchangers to work. The central plant consists of a heat pump that provides 100% of the network heating, and each of the users has a heat pump on the supply line. At point **A** the network temperature becomes high enough to begin to use the hot water heat exchangers to preheat the domestic hot water.

Section A to B   To the right of **A** and up to the point labeled **B** the network temperature increases and hence so does the temperature to which the hot water can be pre-heated. Consequently the electricity required for the electric heaters decreases as shown in Fig. 5.8 and also visible in Fig. 5.9 as the energy cost. The central plant still consists of a heat pump that meets the network heating requirements, and each user has a supply line heat pump. The central plant configuration is indicated in Fig. 5.10 by the percentage of the network energy requirement provided by each component.
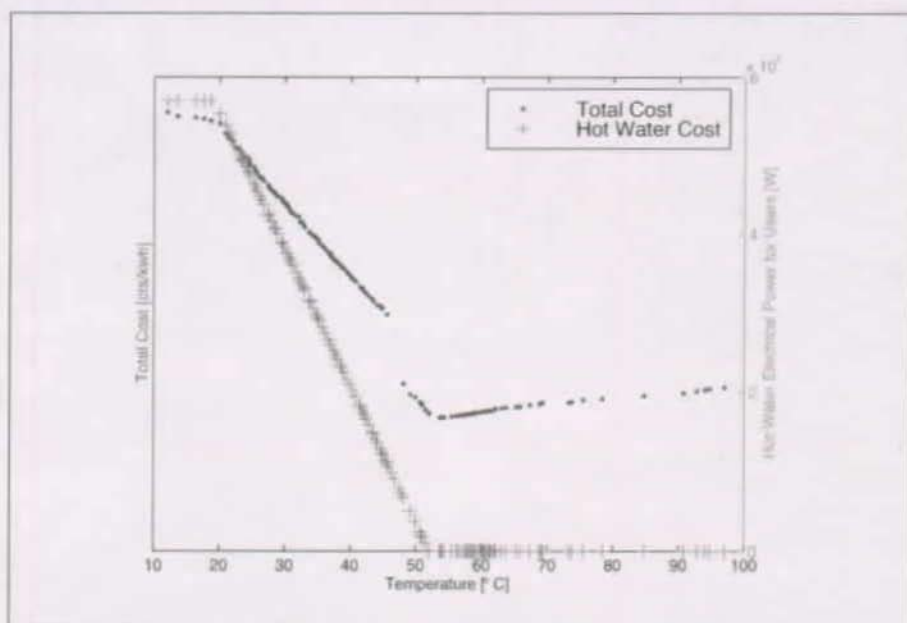
Fig. 5.8: Multiobjective Optimisation with Pseudo Objective Temperature and Total Cost considering pollution costs. The electrical power consumed by the users in order to meet the hot water requirements can be clearly seen to be responsible for the changes in slope of the pareto frontier.

As the network temperature increases the local heat pumps need to do less work, while the central heat pump needs to do more. Since the central heat pump is significantly more efficient there is a net reduction in energy usage (electricity) and hence cost. However, the dominant influence during this section of the graph is clearly the electricity needed to meet the hot water requirement as illustrated in Fig. 5.8.

Points B and C   Between the points labeled **B** and **C** in Fig. 5.7 there is a sudden drop in the cost, that occurs between 47°C and 48°C. This is caused by a change in the optimal configuration of the first user (that makes up 85% of the load) from a heat pump on the supply line (point **B**) to a heat exchanger on the supply line (at point **C**). The increase in network operating temperature makes it possible to meet this user's requirements directly with a heat exchanger, saving both in energy (see Fig. 5.9) and investment cost. The central plant still consists of a heat pump to meet the network requirements.

Section C to D   The decrease in overall cost continues until the point labeled **D**, at a temperature of approximately 52°C which is a sharp minimum, suggesting that a constraint or limit has become active. At
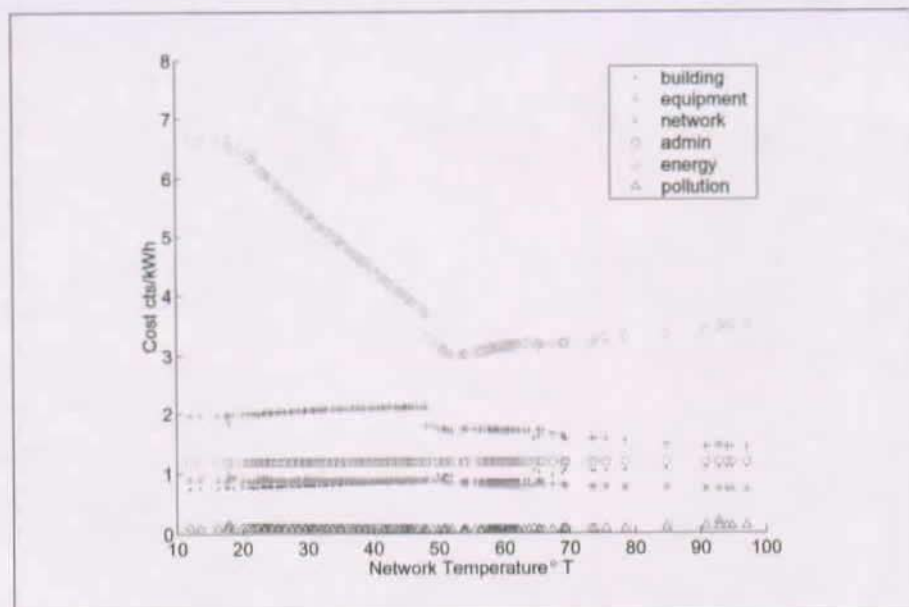
Fig. 5.9: Total cost breakdown from the Multi-objective optimisation of temperature and cost including pollution costs.

this temperature the users' hot water requirement can be met completely by a heat exchanger, and the total cost is no longer dominated by the electric water heaters. The effect of the hot water requirement is more clearly visible in Fig. 5.8, which shows a graph of total cost together with the electrical power consumed by the de-centralised electric water heaters[‡]. To the left of this point (point **D**) the effect of pollution on the optimal configuration and overall running costs are negligible - the best solutions are the same whether or not pollution is considered. In fact the costs of pollution are minimal throughtout this region and are due partly to fabrication but mainly to the production of the imported electricity[**].

Section D to E with pollution costs Continuing to the right of point **D** towards the point labeled **E** the heat pump becomes less efficient due to higher operating temperatures, and the optimal choice moves towards a mixture of heat pump to preheat the network supply and an auxialliary boiler to augment the temperature, as shown on Fig. 5.10. The cost continues to rise because of the cost of fuel and electricity required to raise

---

[‡]It was found that removing the hot water demand resulted in an optimum network temperature when considering pollution costs at around 48°C , which is the temperature at which the first user can make use of a heat exchanger.

[**]This is because at these low temperatures the ideal solution without pollution costs is a heat pump in the central plant. Using the Swiss mix of electricity generation produces relatively low pollution, favoring the importation of electricity over local generation, still favoring the heat pump in the central plant. A different mix of electricity production such as the European mix changes this see Appendix A.1.1

the network supply temperature.
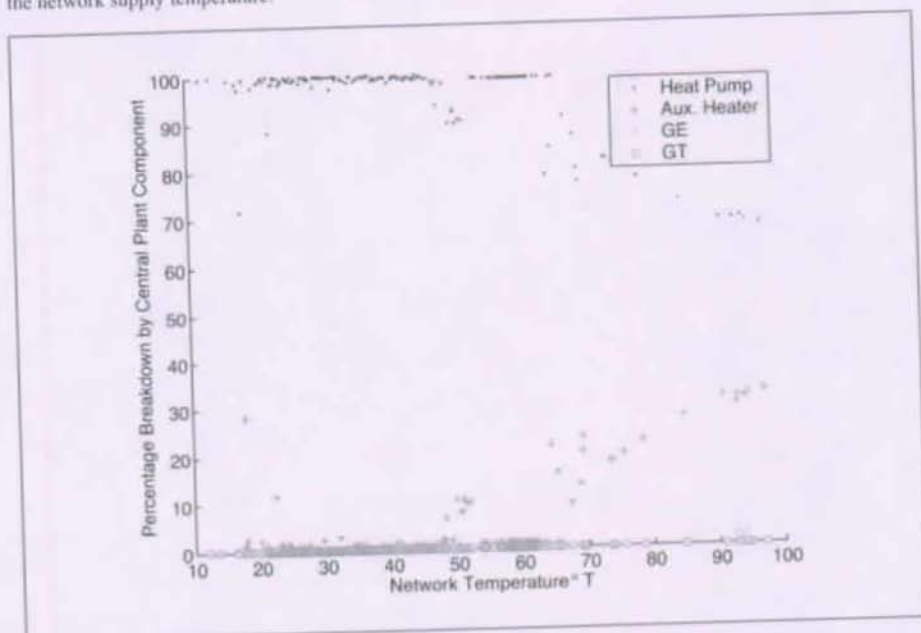
Fig. 5.10: Composition of the central plant. The percentage of the total heat produced by each compo-
nent of the central plant is shown against the network temperature.This is from the pseudo-multiobjective
optimisation of temperature and cost considering pollution costs

Section D to F without pollution costs   When pollution costs are neglected then the tendency from **D** to
F remains more or less flat, with an optimum at around 92°C. There are some small discontinuities that
correspond to the changes in user configuration from supply line heat pumps to supply line heat exchangers
and return line heat pumps, and then towards return line heat exchangers where possible. The optimal
central plant configuration is a heat pump, a gas turbine cogeneration system and an auxiliary boiler. The
gas turbine produces electricity that is then used locally for the heat pump. The network is pre-heated by
the heat pump and then further heated by the gas turbine before passing through the auxiliary boiler. The
gas turbine produces more pollution than the imported electricity (for the swiss electrical pollution mixture)
hence it is only chosen when pollution costs are neglected. The lowest cost solution was found to be very
similar to the solution given in the previous section (HP 42%, AuxGH 39.5% and GT 18.5%) - the small
variations are due to the relative insensitivity to network temperature around this point.

Fig. 5.10 shows the composition of components in the central plant when considering pollution and from
this it is clear that the optimum solution is heavily influenced by the cost of the pollution due to the gas
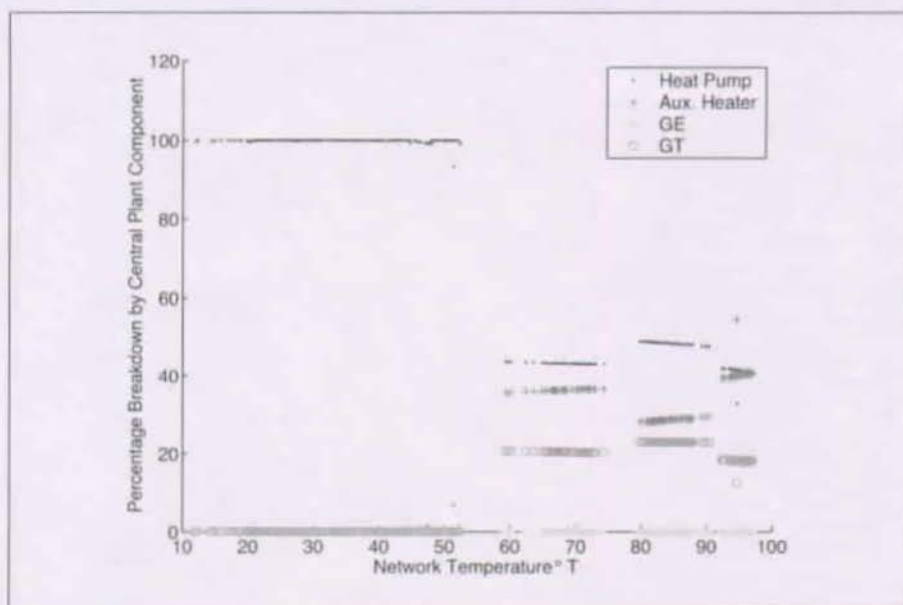
Fig. 5.11: Component breakdown as a percentage of overall energy supplied to the network against temperature from the pseudo-multiobjective optimisation of temperature and cost neglecting pollution costs. The step changes in heat pump use are due to changes in local user configuration as the network temperature is increased.

turbine and auxiliary heater.

Fig. 5.11 shows the composition of components in the central plant when pollution costs are neglected. The major configuration change occurs at approx 59°C when it becomes advantageous to include a gas turbine cogeneration system in the system, along with an auxiliary boiler. The gas turbine is one of the major sources of pollution although it also produces electricity which is used to supply the heat pumps, which turns out to be cheaper than buying electricity provided pollution costs can be neglected. As the temperature increases there are step changes where local user configurations change to use heat exchangers where possible (including heat exchangers on the return line), changing in turn the electricity requirements and hence the optimum combination of gas turbine, heat pump and auxiliary heater.

Miscellaneous Values	
Overall Demand	62.7MW
Electricity Cost	13cts/kWh
Gas Cost	5 cts/kWh
Pollution Cost NO_x	13.8 sfr/kg
Pollution Cost CO_2	0.03 sfr/kg

Table 5.1: Specific costs

Breakdown of Cost as a % of the Total		
	Struggle	CPEA
Buildings	15.8	15.9
Equipment	21.6	21.1
Network	9.7	9.8
Administration	15.6	15.5
Energie	37.2	37.7
Pollution	0.0	0.0
Total [cts/kWh]	7.59	7.59

Table 5.2: Cost breakdown for the optimum solution not taking into account pollution.

Thermal Power as a % of the Total		
	Struggle	CPEA
Network Temperature °C	89.5	89.3
Heat Pump	50.5	47.3
Gas Turbine Cogeneration System	23.3	20.9
Gas Engine Cogeneration System	0.0	0.0
Boiler	26.2	31.8

Table 5.3: Breakdown of power supplied by each component of the central plant for the optimum solution not taking into account pollution.

Breakdown of Cost as a % of the Total		
	Struggle	CPEA
Buildings	11.3	11.4
Equipment	22.7	22.7
Network	10.9	10.9
Administration	15.4	15.4
Energie	39.0	39.0
Pollution	0.7	0.6
Total [cts/kWh]	7.66	7.66

Table 5.4: Cost breakdown for the optimum solution taking into account pollution.

Thermal Power as a % of the Total		
Network Temperature °C	52.5	52.5
Heat Pump	100	100
Gas Turbine Cogeneration System	0.0	0.0
Gas Engine Cogeneration System	0.0	0.0
Boiler	0.0	0.0

Table 5.5: Breakdown of power supplied by each component of the central plant for the optimum solution taking into account pollution.

5.6   CPEA - Pollution Cost and Overall Cost

In his thesis Curti aggregated the cost of pollution with the investment and operating costs into a single objective function, as was done in section 5.4.

The pollution costs were calculated using a representative cost per kilogram together with a pollution factor calculated from the baseline local level of that pollutant. Both CO_2 and NO_x were considered during the fabrication and construction of components and buildings and during the production and transport of primary energy source (for example natural gas) and production of electricity consumed from the grid.
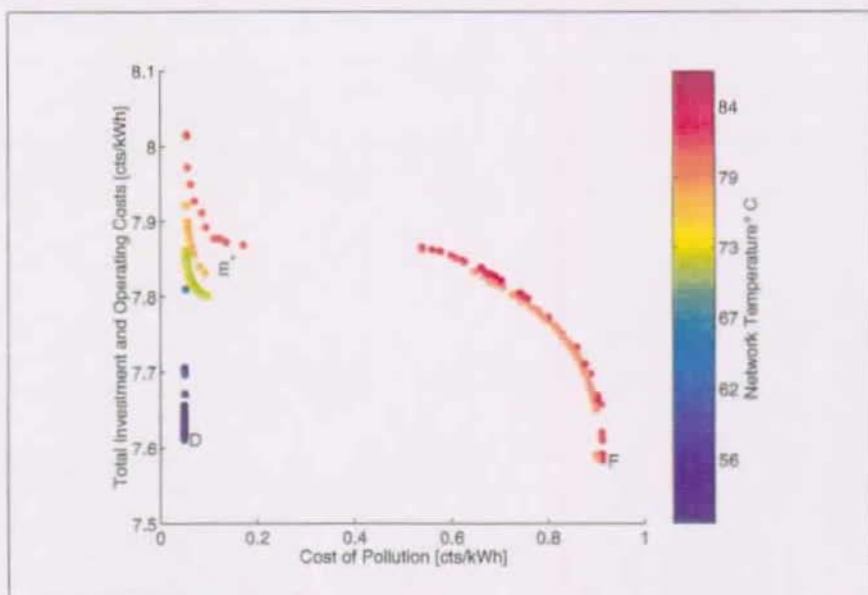
Fig. 5.12: Pareto Surface showing Cost vs Cost of Pollution

For this section the total cost, C_{total}^- was taken as the overall cost (including pollution) minus the pollution costs (these values were readily available from the simulation model), and a two objective optimisation of C_{total}^- and $C_{pollution}$ performed. The pollution costs were calculated as in the previous work. This is the multi-objective equivalent to the two cases (with and without pollution) that have been considered separately as single objective optimisation problems, and used the same fixed CO_2 unit cost as the earlier work.

The optimisation problem was run with the parameters: $p_{initial}$=400, $p_{min\_cluster\_size}$=10, $p_{max\_cluster\_size}$=30, $n_{clusters}$=4, clustering with c-means fuzzy clustering on the network temperature. The optimisation was

continued for 50000 evaluations as in the previous single objective work, using clustering on the network temperature.

Fig. 5.12 is a graph of $\overline{C_{total}}$ vs $C_{pollution}$ with network tempearture indicated by colour. The results show several well defined non-dominated fronts, each one representing a different cluster. Each of the clusters represents a *local* non-dominated front (as described in chapter 2).

The point labeled **F** on Fig. 5.12 corresponds to the same solution marked with an **F** on Fig. 5.7, and represents the best solution if ignoring pollution. This solution consists of a gas turbine, heat pump and auxiliary heater, and if it was built and pollution costs were introduced afterwards then it would clearly not be a good solution. The point labeled **D** represents the minimum solution taking into account pollution, and corresponds to point **D** on Fig. 5.7. The minimum of the sum of the two objectives, $\min(C_{pollution} + \overline{C_{total}})$, is 7.66 cts/kWh, the same as found in the earlier single objective work.

Note that the point **E** on Fig. 5.7 corresponds to a point somewhere in the region of E^* on Fig. 5.12.
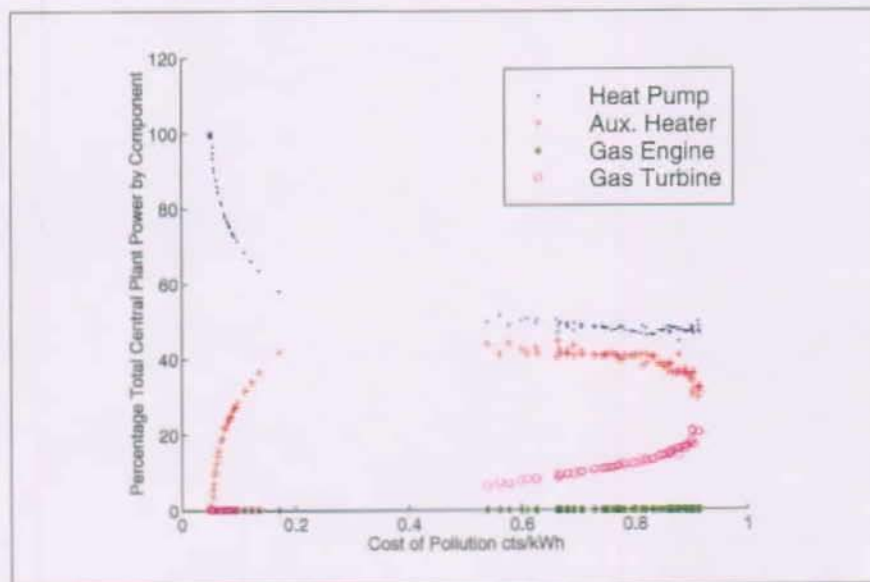
Fig. 5.13: Multi-objective optimisation of Total Cost and Cost of pollution. Central Plant composition as a percentage of total heat produced vs pollution cost.

The composition of components in the central plant is shown in Fig. 5.13, which shows the fraction of energy each composant produces against the cost of the pollution produced by that configuration. Looking more closely at Fig. 5.13 several interesting trends are visible. Moving along the pollution cost axis (pollution

costs increasing) the optimal solution changes progressively from just a heat pump (the optimum solution when pollution costs are considered) to an approximate 50/50 mix heat pump and auxiliary boiler. There is an apparent gap in the graph at this point that corresponds to the change in configuration of the central plant, and the introduction of a gas turbine cogeneration system. Continuing further up the pollution scale the gas turbine produces an increasing percentage of the energy requirement (together with the electricity) but also produces an increasing amount of pollution.

Clustering was the key factor in solving this problem. When run without clustering only the low network temperature solutions are found, since the high temperature, high pollution alternatives are more difficult to find and require sufficient diversity.

5.7 MOO - Overall Cost and Quantity of CO_2 produced

This was the logical extension from the previous section and treated the two objective optimisation of C_{total} (not including pollution costs) and M_{CO_2}, the quantity of CO_2 produced per second (averaged over the year).

The principal difference with the previous case is that the pollution costs due to NO_x are neglected completely.
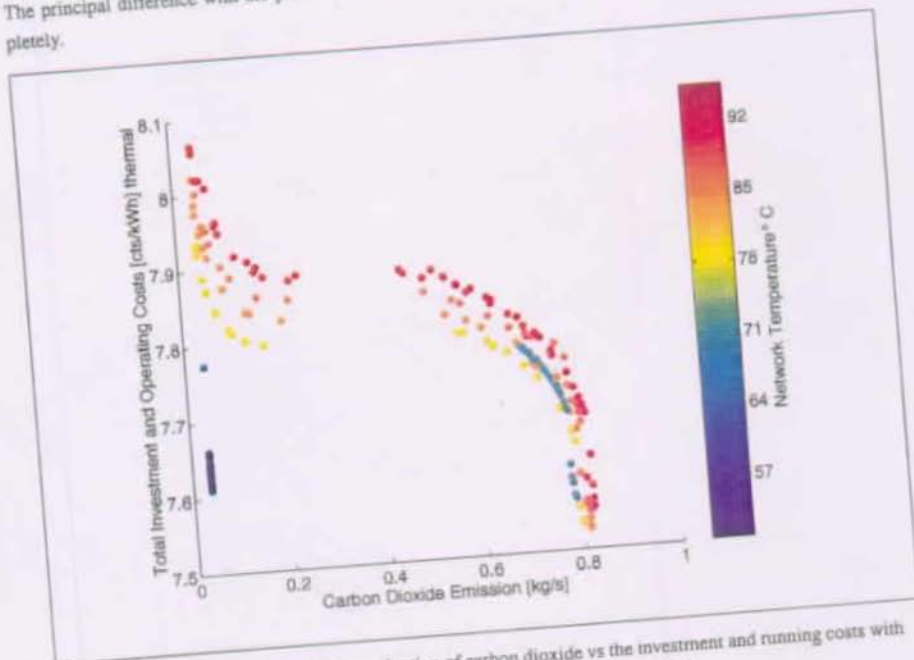
Fig. 5.14: Pareto frontier showing the production of carbon dioxide vs the investment and running costs with the network temperature indicated by colour (for a 62.7MW thermal rated plant)

Fig. 5.14 shows the results of this optimisation on a plot of C_{total} vs M_{CO_2} after 100,000 evaluations, with the network temperature indicated by colour. Fig. 5.15 shows the corresponding central plant configuration as the percentage each component provides of the total power.

The results of this show an understandable similarity to the results found for the case previously considered of C_{total} and $C_{pollution}$ suggesting that the influence of NO_x pollution is negligible in determining the optimum solutions [11]. The optimum solution neglecting CO_2 production is, as before, a mixture of gas turbine, heat pump and auxiliary heater in the central plant, producing a large amount of CO_2.

---

[11] In fact this may be due to NO_x emission following the CO_2 emission, or due to a too low NO_x pollution cost, but this analysis did not form part of the work.
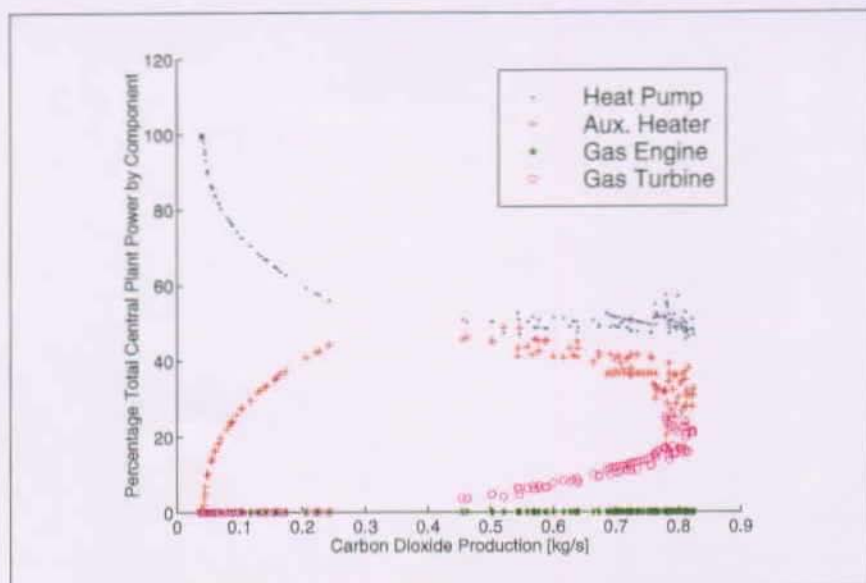
Fig. 5.15: Breakdown of Central Plant Composition for the case of carbon dioxide vs investment and running costs

5.8 Application of Post Processing to Cost, CO_2 Optimisation

To illustrate the idea of post processing, the results from the optimisation of C_{Total}^- and CO_2 were post processed to determine the sensitivity of the best solution to the nominal unit cost of CO_2 pollution. The results of this are shown in Fig. 5.16 which was produced by calculating the cost of the CO_2 produced over a range of unit pollution costs, adding the remaining costs, then choosing the solution with the lowest overall cost.

The graph shows the linear variation due to the increasing specific pollution cost up to the point where a radical change of configuration produces a marked change of slope. This point corresponds to the change in configuration of the network from a central plant with heat pump, gas turbine and auxiliary heater, to a central plant with only heat pump. This involves a change in network supply temperature from approximately 80°C to approximately 52°C and a corresponding change in the first (major) user configuration from a heat exchanger to a heat pump.

The critical value for CO_2 unit pollution cost in this case is approximately 5.3 sfr/tonne CO_2. This may be interpreted as meaning any pollution tax will need to be above this value in order to encourage a minimise
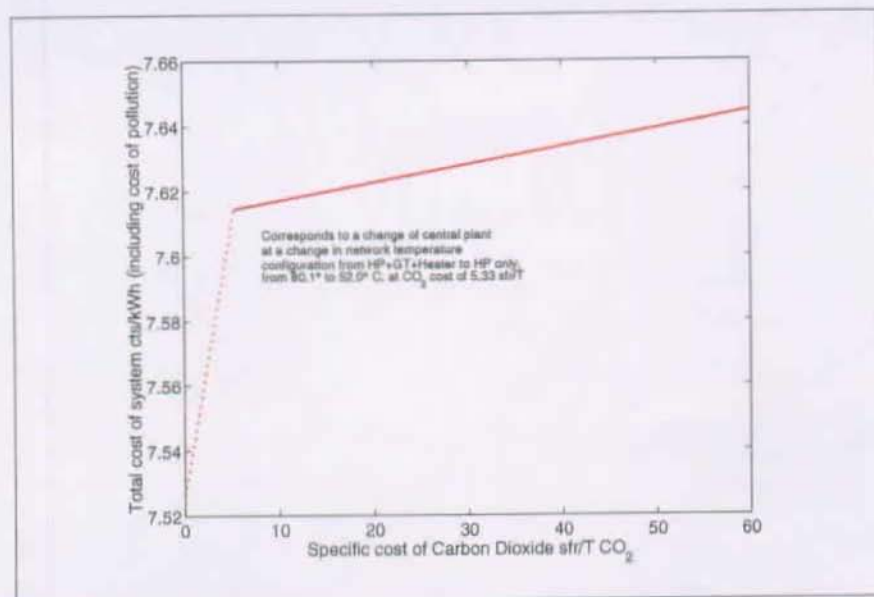
Fig. 5.16: Post processed results of the cost, CO_2 optimisation showing the sensitivity of the solution to an imposed specific carbon dioxide cost

change in the technology used[‡‡].

‡‡Curti used 30sfr/TonneCO$_2$ in his work, based on then accepted values in the literature.

5.9 Three objective problem - Operating, Electricity and Gas Costs

This involved a three objective optimisation of investment cost, electricity (bought - quantity of imported electricity), and the remaining operating costs, composed largely of gas costs, but including maintenance.

These were calculated from readily available values in the simulation, as:

$$C_{inv} = C_{total} - C_{operating} \tag{5.1}$$

$$C_{fuel+} = C_{operating} - C_{elec} \tag{5.2}$$

where C_{elec}, C_{inv} and $C_{operating}$ were directly available, and C_{fuel+} consists of the gas costs together with the remaining operating costs (such as maintenance that could not be easily isolated in the simulation). The problem was posed like this in order to make it directly comparable with previous results, and to minimise the changes to the simulation model. In order to verify the results a further two objective optimisation was performed to find the non-dominated front for the two objective optimisation problem of investment costs, C_{inv} and operating costs $C_{operating}$, defined in the same way. This could then be compared to results of the three objective problem by summing the C_{fuel+} and C_{elec} costs to reproduce the $C_{operating}$ cost.

The problem was run without clustering, since the three objectives were expected to introduce enough diversity. The other parameters were: $p_{initial}=400$, $p_{min\_cluster\_size}=70$, $p_{max\_cluster\_size}=160$.

As has been mentioned previously the major problem with multi-objective optimisations with more than two objectives is the rapid growth of the population.

5.9.1 Results and Analysis

Verification - Investment and Operating costs

The problem was run both with and without pollution costs, but for brevity and clarity only the results without pollution will be presented. A plot of each objective against the network supply temperature is shown in Fig. 5.17.

It is difficult to present clearly the three dimensional objective space, however, the three combinations of two objectives are shown in Fig. 5.18 together with a three-dimensional representation of the objective space.

Closer investigation of the three dimensional results reveals that the points lie very nearly in a plane, sug-
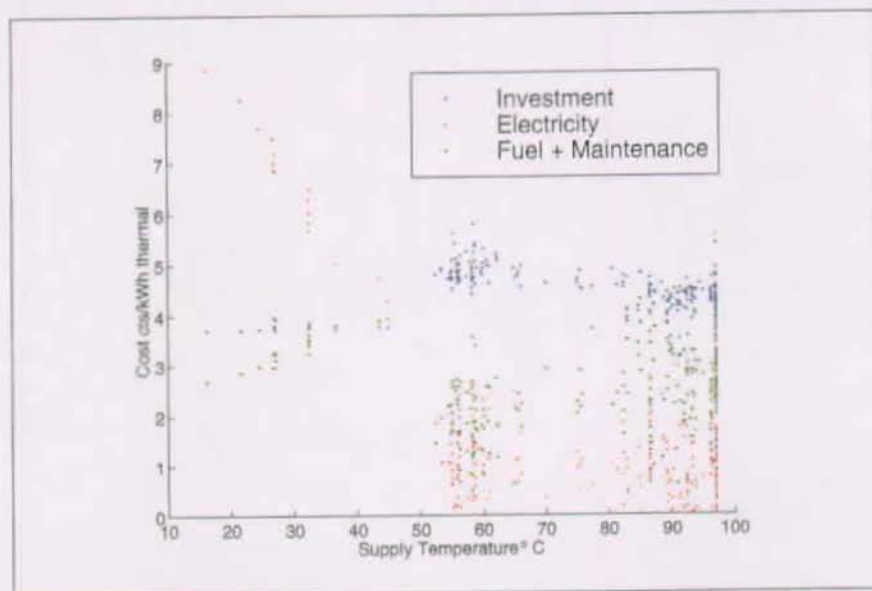
Fig. 5.17: Results from the optimisation of investment cost, electricity cost and the remaining operating costs shownagainst temperature

gesting the existence of a linear (or near linear) relationship between one of the objectives and the other two.

There is an apparent "tail" visible in the three dimensional plot Fig. 5.18 that corresponds to the low network supply temperature solutions more easily visible in Fig. 5.17. These are solutions with very low fuel costs and low investment costs, but high electricity costs. They correspond to configurations which have only the auxiliary heater in the central plant and remote heat pumps in each of the user—since the network temperature is very low the fuel costs are low, the heating being provided by the local heat pumps that have a high electricity demand.

Treating the results of the three objective problem as a response surface and combining them to form two dimensional graphs it was possible to compare the results to the two objective optimisation problem. Fig. 5.19 shows the results of the two objective C_{inv} and $C_{operating}$ case with results of the three objective optimisation.

Looking at the results of the investment vs operating in Fig. 5.19 it can be seen that the three objective optimisation has not yet converged to the same degree as the two objective optimisation, although they both show the same trends. This is a due to the shortcomings of the 3D thinning algorithm.

(a) Investment Costs vs Electricity Costs

(b) Investment Costs vs Fuel Costs

(c) Electricity costs vs Fuel Costs

(d) Three dimensional view.
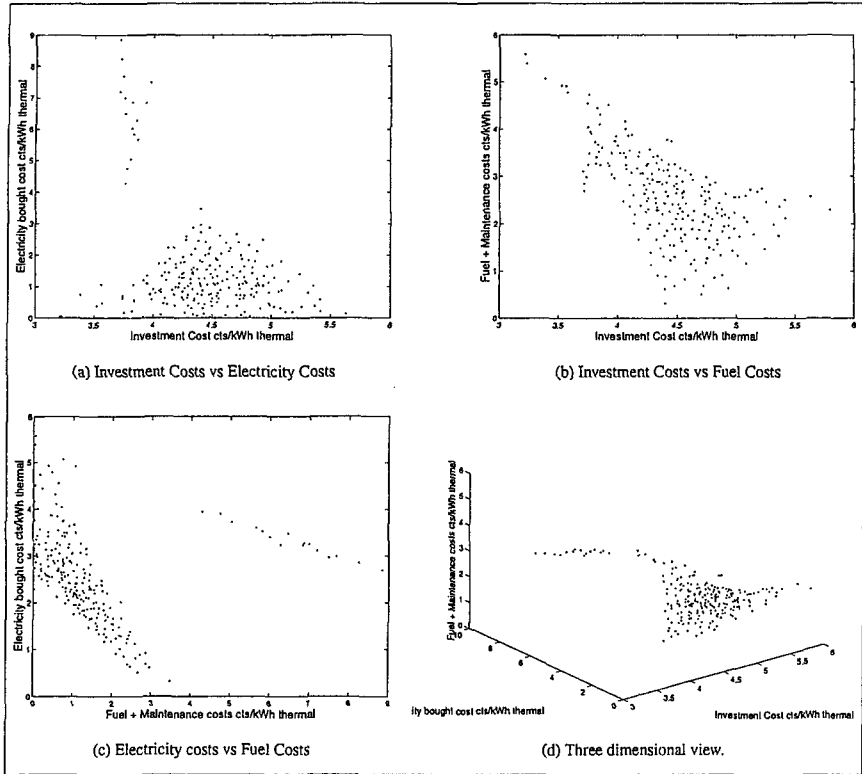
Fig. 5.18: Results of the 3 objective optimisation of Investment cost, electricity bought cost and fuel costs.

However, it is still interesting and informative to continue the post processing of the results. As electicity costs are varied from 30% below to 30% above the nominal cost of 13 cts/kWh, keeping fuel costs fixed, the non-dominated front of operating cost and investment cost changes as indicated in Fig. 5.20 by the lines joining non-dominated points (points themselves have been omitted for clarity).

The best overall solution is defined by summing the the separate costs and finding the solution with the lowest total cost C_{total}, as per equation (5.3).

$$C_{total} = \sum C_{inv} + C_{fuel+} + \alpha_{elec}C_{elec} \qquad (5.3)$$

where α_{elec} was varied from 0.7 to 1.3 to represent changes in electricity prices from 9.1 cts/kWh to 16.9
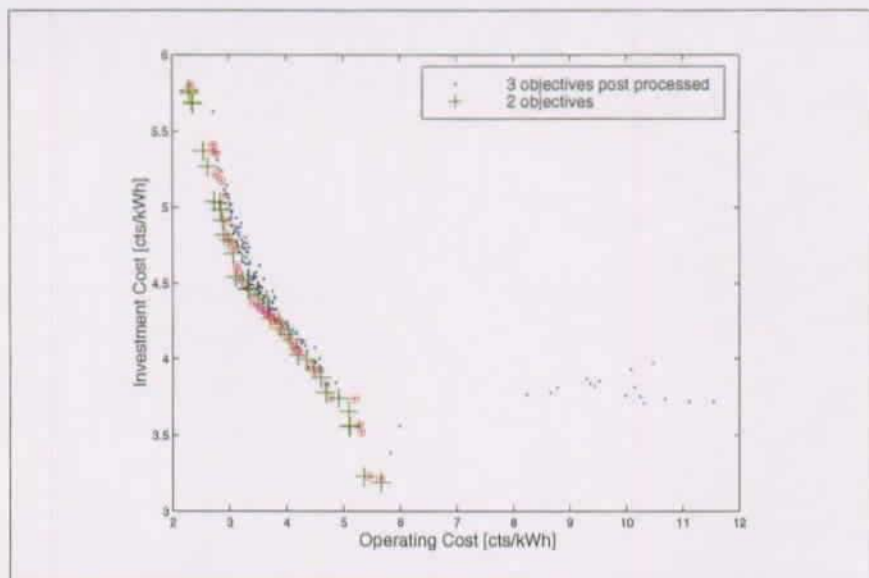
Fig. 5.19: The results of an optimisation of investment cost and operating cost from a two objective optimisation, compared to the same results calculated by post processing an optimisation of investment cost, electricity cost and the remaining operating costs. Results are for the case neglecting pollution costs.

cts/kWh. These solutions are marked with •'s in Fig. 5.20. The nominal 100% solution is marked with a o.

As might be expected, the changes in electricity cost do not directly influence the investment cost until there is a change in configuration. This is apparent on the graph (Fig. 5.20) and is due to a change of central plant configuration. The change occurs at an approximate 10% decrease in electricity prices, and corresponds to a change from heat pump, gas turbine and auxiliary heater at the nominal electricity price of 13cts/kWh to a heat pump and auxiliary heater as the electricity price is dropped. The solution configuration stays more or less constant below and above this point. This is interesting in that it suggests that the gas turbine is only beneficial when the electricity price becomes high enough (as might be expected). It is important to realise that this information **can not** be found from the two objective optimisation without re-optimisation since this represents results for the fixed ratio of $\alpha_{elec} = 1.0$.

While the results are not fully converged they are close enough to correctly show the trends and to demonstrate the advantage of keeping costs separate until after the optimisation process. This information was obtained with 200,000 evaluations, representing a significant gain in information.
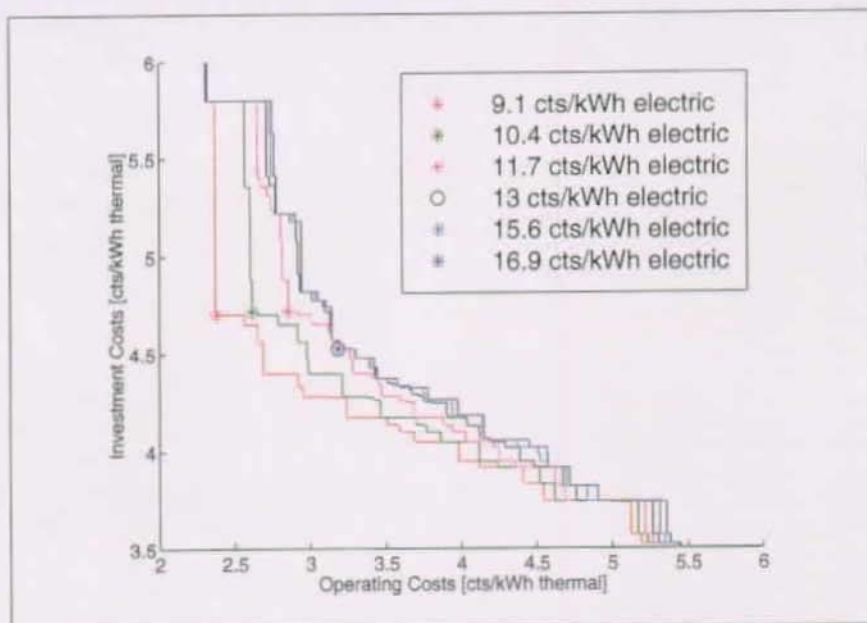
Fig. 5.20: Part of the investment cost vs operating cost graph, showing the effect of varying electricity price on the non-dominated front, obtained by post processing the results of the 3 objective optimisation of investment cost, electricity cost and fuel costs. The sudden jump corresponds to a change in configuration, with the network supply temperature changing from 56.3 to 91.3 °C. Costs are in Suisse cents.

5.10   Conclusions

The district heating network problem has been re-examined and the effectiveness of the new CPEA demonstrated. The single objective version outperforms the algorithm used in earlier work with an apparent forty fold reduction in function evaluations.

A new method of performing an optimised parametric study has been presented and used to examine the solution domain with respect to temperature, giving a more complete view of the solution domain without increasing the computational work load.

The multi-objective CPEA for two objective pollution and cost optimisations has been shown to reproduce the same results as the two single objective optimisations, in a single optimisation process, again halving the function evaluations. There are additional *free* benefits, including the removal of uncertain pollution costing parameters from the optimisation process, and of a more thorough understanding of the solution domain.

The potential benefits of three objective optimisations have been explored and the current limitations high-lighted. The technique still represents a powerful tool in the solution of such integrated problems, reducing even further the computational work load and allowing a greater portion of the uncertainty to be examined in a post processing step.

The effectiveness of clustering has been demonstrated, and it has been found that multi-dimensional scaling of the independent variables in this problem, to reduce the clustering space to two dimensions, produces equivalent solution spaces to clustering on network supply temperature alone.

Chapter 6

Hybrid Vehicle Drivetrain

6.1 Introduction

Transport presents a major problem for sustainability, and energy use for transport is rising faster than in any other sector. In Switzerland transport is responsible for around a third[88] of the man-made CO_2 produced, as well as many other directly harmful pollutants including NO_x, SO_x, hydrocarbons (HCs) and ozone. In high concentrations ozone can damage lung tissue, reduce lung function and sensitize the lung to other irritants, as well as damaging crops[73]. NO_x and particulate emissions have been linked[20,28] to severe health problems and even premature death.

To address this, new, energy-efficient and less polluting transport technologies must be evaluated against traditional solutions (and advances in traditional solutions) in a wide variety of situations to aid decisions on new policies.

The simulation of vehicle drivetrains is a difficult task - useful simulations tend to be complex and hence computationally expensive, even with many simplifying assumptions in the modeling. In addition there is a great deal of uncertainty and variability in the component data, and in the evaluation of potential objective critera.

To be acceptable to a large public a vehicle must have a competitive performance* as well as a good fuel economy, and as seen in earlier work[62] there is a tradeoff between fuel economy, performance and emissions. When investment, operation and potential pollutant costs are introduced the overall problem becomes even more complex. The CPEA optimisation technique is attractive in this kind of problem because it allows more information to be gathered about the solution domain for the same effort as a single point optimisation, allowing costs and emissions to be considered separately.

---

*The trend in the demand in the US is still towards faster 0-60mph acceleration times[73]

The work presented here aims to demonstrate the feasibility of perfoming MOO in this highly demanding area, and to produce some initial results as to the potential for some of the newer vehicle technologies.

6.1.1   Hybrid Vehicles

In a conventional vehicle, the vehicle's final drive (differential and wheels) is driven directly by the internal combustion engine (ICE) through a clutch and gearbox. The gearbox means that the ICE can be run near its optimal speed over a wide range of vehicle speeds. However, engine efficiency and emissions are strongly dependent on load, as well as speed, as illustrated in Figs. 6.1 and 6.2.
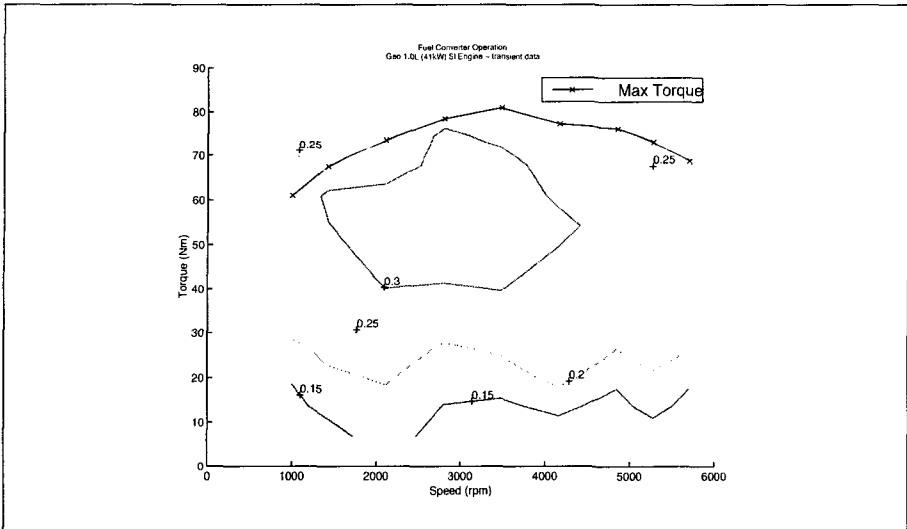
Fig. 6.1: ICE Efficiency countours for a 41kW gasoline SI engine, shown against speed and torque, together with the maximum torque envelope.

Consequently, while an ICE's peak efficiency can be over 30%, overall vehicle efficiencies can be lower than 15%[46]. Most of these losses are due either to the engine running away from its most efficient point, or to the engine running while the vehicle is stopped (more than half of the time in most homologation cycles). Conventional gearboxes and differentials, for example, are extremely efficient, and drivetrain losses are generally only on the order of 5%. Hybrid vehicles are an attempt to reduce these losses by flattening the demands on the ICE. In addition, they can improve efficiency by recovering some of the energy dissipated in braking, which otherwise only serves to heat and wear out the braking system.

A hybrid vehicle, in the broadest sense, is a vehicle that contains several power sources, and attempts to use those power sources in order to maximise overall efficiency. In this and earlier work[62] the emphasis has so
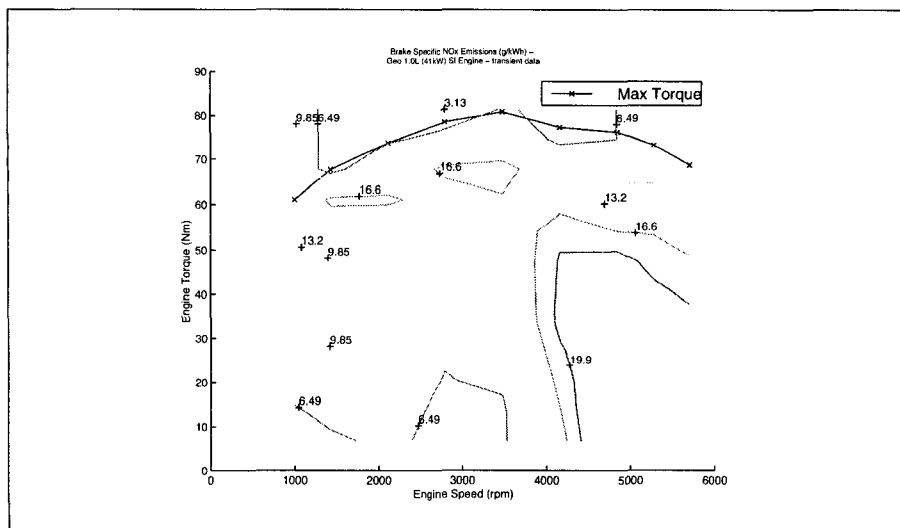
Fig. 6.2: ICE NO_x production in g/kWh for a 41kW gasoline SI engine, shown as contours against speed and torque, together with the maximum torque envelope.

far been on thermal-electric hybrids with an internal combustion engine and an electric motor[†]. These may be configured in a series or parallel configuration. A hybrid vehicle superconfiguration, showing a series hybrid with bold arrows, is shown in Fig. 6.3.

In a series hybrid, an internal combustion engine is connected to a generator, which either charges a battery or powers an electric motor. The motor powered either by the battery or the generator, then powers the wheels. This means that the ICE can, if the control system wishes, always be run at its best operating point [‡], while the electric motor, with better characteristics over a wider load/speed range, copes with demand variations. Series hybrids are relatively simple compared to parallel hybrids (described below). However, in situations where the vehicle operating point would allow the ICE to run well, such as at constant speed on a motorway, there is nothing to gain from the hybrid series configuration, but one still has to pay the losses of the complex mechanical-electrical-mechanical energy chain. There are also additional problems[73] such as the limited capacity of the batteries to absorb the energy produced by the engine operating at an optimal point, and the potential for increased emissions when repeatedly stop starting the engine.

Parallel hybrids solve this problem by allowing the ICE to either drive the generator or power the vehicle final drive directly. This involves the use of a torque coupling device (for example, the Toyota Prius has an

---

[†]This is the most likely next step to improve emissions and fuel economy and to meet demand for short term acceleration. Fuel cells for small cars are still very much in the prototype stage, and information on performance is currently limited.

[‡]Clearly the "best" is not clear - best for pollution or best for economy?

Fig. 6.3: Hybrid Vehicle Superconfiguration in its broadest sense. The links showing a series hybrid are shown in bold.

epicyclic gearbox controlled by the generator[92]) between the ICE, electric motor, generator and final drive.

However parallel hybrids suffer from increased complexity and potentially higher maintenance costs than conventional solutions[§].

Generally, battery capacity in a hybrid is low (the Prius has 50 kg of batteries while the General Motors EV1 *electric* car has 520 kg[39]) and the ICE needs to be run frequently to top up the battery charge, as well as whenever the vehicle has high power demands. In order to reduce charge cycle losses and improve battery life, batteries are usually kept within a narrow range of charge situated around half full. Consequently the batteries are usually larger and heavier than strictly necessary. If, however, the battery capacity is high, the ICE can be turned off completely for short periods, which may be interesting, for example, for crossing a town centre where only non-emitting vehicles are allowed.

The best choice of battery capacity, as with most vehicle design choices, is not evident - the Toyota Prius[92] has twice as many battery modules as the Honda Insight[102].

One reason for the slow growth of the hybrid car market is the cost of the electric components (since these

---

[§]In a survey published by the TCS[89] the cost of maintenance of conventional vehicles is shown to vary dramatically - from as little as 400 sfr for 100,000km up to 4200 sfr.

are not in mass production in the same way that IC engines are) and the doubt over the operational benefits. In order to address this it is necessary to compare the cost of different options as well as emissions and performance, in particular to identify the sensitivity of the best choice to the relative cost of the electrical components.

6.2 The Vehicle Simulation Model

6.2.1 Problems in Simulating a Vehicle

Modelling a vehicle poses several particular difficulties:

- To evaluate a vehicle design it must be "driven" around a test cycle, equivalent for example to a car following a test cycle on a rolling road. If unable to follow the cycle what should be done?

- A vehicle is a highly dynamic system. Considering even quasi-static behaviour means evaluating the performance at many hundreds if not thousands of points around the test cycle. To model detailed dynamic behaviour (for example what happens in the combustion chamber as the throttle is opened) is not feasible as part of a system-level simulation.

- Many of the sub-components that make up a vehicle are themselves complicated, may be non-linear and may themselves require optimisation to run efficiently. During a drive cycle many components will be running "off-design*", so simulation models must be detailed enough to predict off-design losses, or follow an optimal control map.

To help deal with these complications several ideas were considered:

- Adaptive simulation precision - time consuming models of sub-components could be replaced by simpler, faster approximate models, derived from training artificial neural nets or using lookup tables, based on the results of off line sub-optimisations.

- Wherever possible model connectivity "intermediate" result values that are not directly used to calculate objective values should not be calculated.

As part of the Alliance for Global Sustainability's [†] Holistic Design project[99], LENI developed simulation models of several vehicle powertrains, some in co-operation with the University of Tokyo[49]. These

---

*A component will usually have one optimal operating state. When used away from this state the component is being used "off-design".

[†]The Alliance for Global Sustainability (AGS) is a joint venture between the Swiss Federal Institutes of Technology, MIT and the University of Tokyo.

models allow the engineer to specify the layout of a vehicle drivetrain, specify component sizes and con-
trol strategies, and then run the vehicle through any number of test cycles, measuring performance (how
well the vehicle managed to follow the cycle), emissions, fuel economy, battery charge, catalytic converter
performance and many other variables. The models were largely based on data from tests of petrol and
compressed natural gas engines and catalytic converters performed in LENI's engine laboratory. Models of
other components were derived from first principles or from the literature.

This system was in turn interfaced to "DOME"[74] (Distributed Object Modeling Environment), developed
by MIT to create a common interface to many different programs, and allow the interconnection between
analysis and other programs such as LCA (Life Cycle Analysis) systems. Details of this original simulation
are available in the final report of the Holistic Design project[99], and the simulation was used (via DOME)
as input to an LCA and emissions impact study by Amundsen[2].

Despite a large effort in developing this in-house framework it became apparent that resources were not
available to develop enough vehicle components to allow a thorough study[‡]. A search of the literature
identified a MATLAB Simulink vehicle simulation model called ADVISOR[100], with a functionality very
similar to the system developed at LENI.

6.2.2   ADVISOR Vehicle Simulation

ADVISOR was developed by the Vehicle Systems Analysis Team (VSAT) of the National Renewable En-
ergy Lab (NREL)'s Center for Transportation Technologies and Systems (CTTS)[100], and currently has ten
engineers involved directly in the development of component models and testing of components and vehi-
cles, as well as the support of many industrial collaborators and universities (with more than 600 users in
May 1999[12])

It is freely available for research purposes and has an open, extensible, modular structure with a user friendly
graphical interface, shown in Fig. 6.4, primarily aimed at allowing direct user interaction. Use of ADVISOR
has greatly increased the number of component models available and allowed the work to concentrate on
optimisation[§].

The ADVISOR simulation toolbox allows a vehicle to be built up from a series of components, each of
which publishes a set of controlling parameters that may be modified from an initial default value. Once
specified, a given configuration (a car) can be *driven* through a drive cycle using a specific strategy and fuel
economy, emissions and many other quantities monitored throughout.

However, the ADVISOR Simulink model is computationally expensive - requiring several tens of seconds

---

[‡]A model of a CNG and petrol ICE, catalyser and electric motor were developed and comparisons between simulated perfor-
mance and measured performance on the ECE-EUDC test cycle were presented in Wallace et al.[99].

[§]Future work may integrate the previously developed compressed natural gas engine and other models from the AGS project.
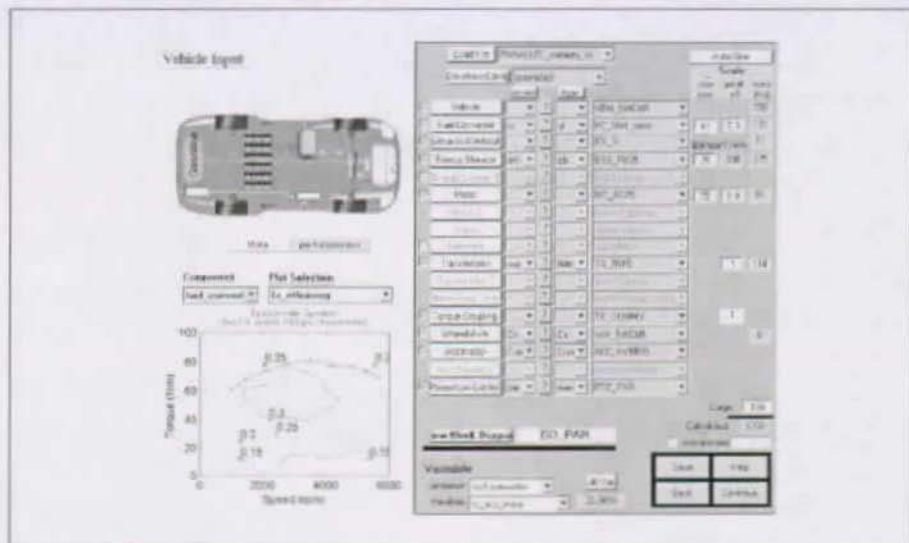
Fig. 6.4: Screen shot of ADVISOR interface showing initial values for a parallel SI hybrid vehicle.

to run on a modern PC. To optimise all but the most trivial of examples requires thousands of simulations. In order to cope with this a parallel version of the CPEA was implemented as described earlier in section 9 which allowed multiple computers to run the simulation concurrently[65] on either a loose collection of Windows 2000 machines, or more recently on a dedicated Linux cluster[93] of 22 machines. The Linux cluster has allowed the previously published work[62] to be greatly enhanced and developed.

The models are a combination of tabularised test data (for example engine emissions maps) and thermodynamic calculations. A vehicle "structure" is chosen (for example parallel hybrid) that defines specific components and default values for parameters. Model parameters may be changed, and the component models will reflect these changes where appropriate - for example scaling the maximum power of an IC engine will scale the emissions test data and fuel consumption and also the engine mass. This in turn will be reflected in the overall vehicle mass.

6.2.3 ADVISOR Vehicle Models

The work presented here made use of 6 standard vehicle models, and 2 specific vehicle types for comparison.

- Conventional SI - A conventional vehicle with a spark ignition 41kW gasoline engine (based on a Geo Metro 1.0L gasoline engine) and 5 speed manual gearbox. This engine was also used in the SI

hybrids.

- Conventional CI - A conventional vehicle with a compression ignition 60kW diesel engine (based on a Mercedes 1.7L Diesel engine) and 5 speed manual gearbox. This engine was also used in the CI hybrids.

- Series SI - A series hybrid with a 41kW gasoline engine, a Unique Mobility 32kW permanent magnet motor, and a second 32kW permanent magnet motor as the generator. Nominally 50 Ovonic NiMh 28Ah 6V battery modules.

- Series CI - A series hybrid with a 60kW diesel engine, a Unique Mobility 32kW permanent magnet motor, and a second 32kW permanent magnet motor operating as the generator. Nominally 50 Ovonic NiMh 28Ah 6V battery modules.

- Parallel SI - A parallel hybrid with 41kW gasoline engine and Unique Mobility 32kW permanent magnet motor. The electric motor was also used as the generator. Nominally 50 Ovonic NiMh 28Ah 6V battery modules.

- Parallel CI - A parallel hybrid with 61kW diesel engine and Unique Mobility 32kW permanent magnet motor. The electric motor was also used as the generator. Nominally 50 Ovonic NiMh 28Ah 6V battery modules.

Each of the standard vehicles made use of a typical small car body with a glider mass (without engine, gearbox, exhaust system, drivetrain, motor etc.) of 592kg, frontal area of 2m^2 and the coefficient of drag (Cd) of 0.33. With a conventional 41kW SI engine and powertrain, together with 5 speed gearbox this resulted in a total vehicle mass of 1192kg. A stoichiometric close coupled catalyser was fitted to all vehicles.

The series configuration used is shown in Fig. 6.5 and the parallel configuration in Fig. 6.6.

In addition certain of the optimisations were run with ADVISOR models of the Toyota Prius (Japanese version)(see Fig. 6.7) and the Honda Insight. The Honda Insight and the Toyota Prius both have a lightweight chassis, low Cd and low rolling resistance wheels. The Insight is a "thin" parallel electric assist (similar to Fig. 6.6) with a 10kW electric motor taking the place of the flywheel and with the clutch between the engine/motor assembly and the gearbox. The engine is an advanced VTEC 1.0 litre three cylinder gasoline engine with variable valve timing. Unfortunately no emissions data was available for the Insight, so it could not be included in the optimisations for pollution or NO$_x$.

The Prius uses a planetary gear set as a power split device with the electric motor speed directly proportional to the wheel speed, and the IC engine speed controlled by the torque applied by the generator.

Both the Prius and the Insight seem to use the same spiral wound NiMh 6.5 Ah batteries, with a nominal 7.2V per module. The Insight has 20 modules and the Prius 40.
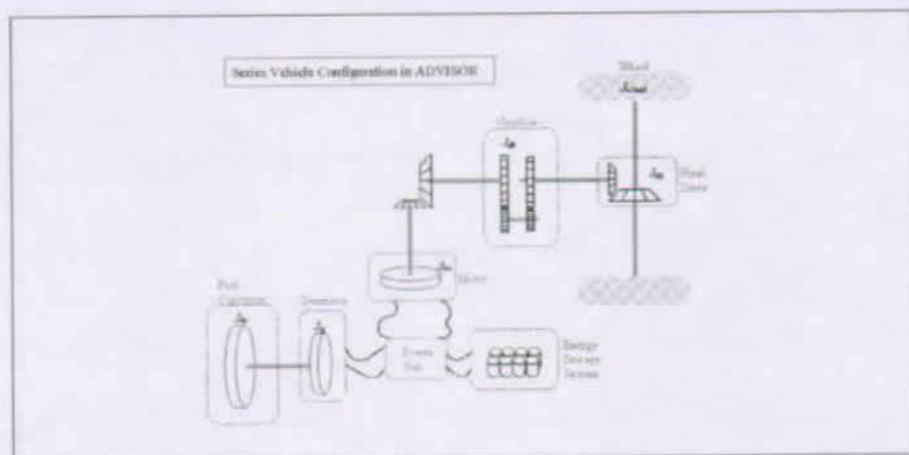
Fig. 6.5: Series Hybrid Configuration (from Advisor documentation).
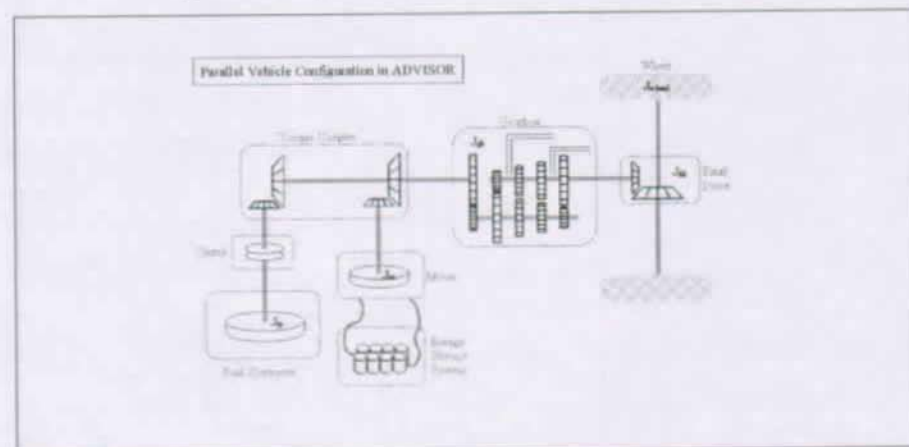
Fig. 6.6: Parallel Hybrid Configuration (from Advisor documentation).

Hybrid Control Strategies

There are many possible strategies for controlling the IC engine and electric motor interaction, and this can have a dramatic impact both on fuel economy and on emissions.

For this work the Advisor control strategies outlined below were used for the parallel and series hybrids. These all made use of the maximum and minimum state of charge (SOC) of the battery pack as parameters
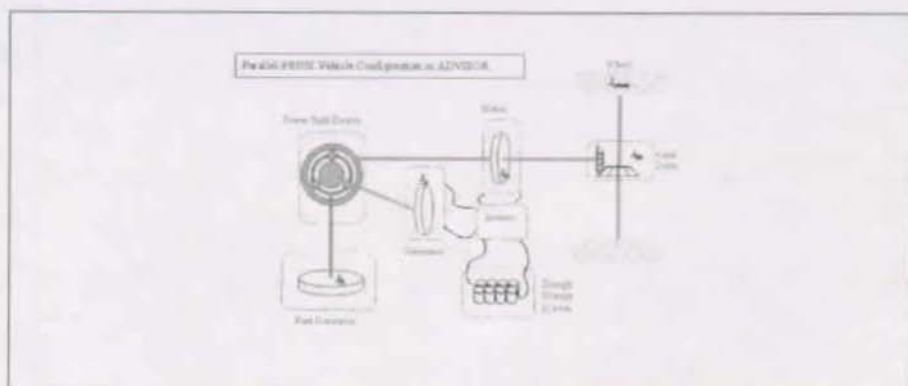
Fig. 6.7: Toyota Prius Configuration (from Advisor documentation).

to the control system.

Parallel Electric Assist The parallel electric assist control strategy uses the motor for additional power when needed by the vehicle and maintains charge in the batteries.

A relatively simple parallel assist strategy[51] was used that controlled the electric motor in a variety of ways:

- The electric motor can be used for all driving torque below a certain minimum vehicle speed.

- The electric motor is used for torque assist if the required torque is greater than the maximum that can be produced by the engine at the engine's current operating speed.

- The motor charges the batteries by regenerative braking.

- When the engine would run inefficiently at the required engine torque at a given speed, the engine will shut off and the electric motor will produce the required torque. Clearly there is an energy requirement to restart the engine.

- When the battery SOC is low, the engine will provide excess torque which will be used by the motor to charge the battery.

Series Electric Control Strategy The series strategy controls the IC Engine as follows:

- The IC engine may be turned off if the battery pack SOC gets too high.

- The IC engine may be turned on again if the power required by the generator (to meet the demand from the motor) gets high enough.

- The IC engine may be turned on again if the SOC gets too low.

When the IC engine is on, its power output tends to follow the power required by the motor, accounting for losses in the generator so that the generator power output matches the motor power requirement. However,

- The IC engine output power may be adjusted by the SOC, tending to bring the SOC back to the center of its operating range.

- The IC engine output power may be kept above some minimum value.

- The IC engine output power may be kept below some maximum value (which is enforced unless the SOC gets too low).

- The IC engine output power may be allowed to change no faster than a prescribed rate.

Prius Control Strategy   The Prius has a planetary gear system to control the flow of power (torque) between the motor, the IC engine, the generator and the wheels (Fig. 6.7).

For a given vehicle speed, and a desired output power (determined by drive cycle, or driver inputs)

- determine the desired operating point of the engine (based on max efficiency curve)

- determine the generator speed (which is controlled by generator torque) to have engine at the desired operating point

- determine motor torque (power or regeneration) to provide necessary power to the wheels (or recapture energy from wheels)

- batteries provide additional power when needed or take back extra charge provided by generator or motor in regeneration.

There are several heuristics:

- Below a SOC of 0.5 the engine is always on.

- If the SOC is $> SOC_{high}$ then the batteries are not charged.

- If the SOC is $< SOC_{low}$ then the batteries are charged if the engine can produce enough torque.

- The batteries are charged on braking via the generator.

In addition the IC engine is started if the temperature drops below a certain limit - by keeping the engine near operating temperature, emissions at restart can be minimised.

Interestingly this strategy can result in the engine running to charge the batteries even when the vehicle is stationary.

Insight Control Strategy   The Honda Insight control strategy in ADVISOR is based on test data collected at NREL[54].

The demand is translated to a required torque at the clutch. Based on this value and the vehicle speed, the electric motor torque contribution is calculated. The remaining torque is demanded from the ICE. The electric motor torque is decided based on the following criteria:

- When accelerating, based on the torque and rate of acceleration, the electric motor assists the ICE, producing around 10 Nm of torque.

- During regeneration (in reality, when the brake is depressed), the electric motor regenerates a portion of the negative torque available to the driveline. Regeneration can only take place if the clutch is engaged.

- At low vehicle speeds, typically below 10 mph, the braking is primarily only the friction brakes.

- There is no electric assist in the first gear.[¶]

6.2.4   Comparison of ADVISOR Models with Published Test Data

As part of a student project[6] several vehicles were modeled and compared to data available from a review article[5] : the Peugot 307, the Ford Focus 1.8 TDCi, and the European version of the Prius.

Simulations were run over the ECE-EUDC cycle in order to compare the three vehicles, which were later used as a basis for the cost estimates. Table 6.1 shows the results of the comparison.

It is clear that there are discrepancies between simulation and experimental test data. Indeed it is interesting to note that in the published tests the Prius has better fuel economy in the urban cycle than on the inter urban cycle. Other researchers have found similar discrepancies in the emissions simulations of ADVISOR[78], since the emissions behaviour is based on limited tests available from IC engine experimental tests typically performed at static speed/torque points. Consequently they are extremely sensitive to catalyst temperature and behaviour, as well as dynamic effects. However the emissions simulations are considered adequate for comparative purposes[31] even though they may not be considered absolute.

[¶]This is the model currently implemented in Advisor but represents a very early model Insight. In newer models electric assist is clearly active[98].

	Fuel Economy (l/100km)	Emissions (g/km)				0-60mph (s)
	ECE-EUDC	HC	CO	NO$_x$	PM	
European Prius	5.4	0.214	0.210	0.080	-	14.09
European Prius (Experimental)	6.3	0.03	0.4	0.05	-	13.8
Peugot	8.7	0.082	0.28	0.079	-	10.2
Peugot (Experimental)	8.6	0.06	0.3	0.06	-	10.6

Table 6.1: Comparison of Experimental data and ADVISOR simulation by Bauman[6].

6.3 Drivetrain Optimisations

6.3.1 Optimisation Variables, Conditions and Parameters

The ADVISOR simulation gives access to an extremely large number of variables that could be optimised, including the structure of the vehicle - choice of catalyser, battery type etc.

However, with the already heavy overhead of the ADVISOR simulation it was decided to limit the number of variables optimised. In addition to a choice of basic vehicle configuration, seven variables were chosen, although not all are operative for all of the configurations.

Four drive cycles were used in the optimisations. The ECE cycle representing a urban European city[1] (Paris, Rome), the EUDC cycle representing inter urban use with a maximum speed of 120km/h, and the US06HWY cycle representing an American short highway driving cycle.

The ECE-EUDC is a combined cycle consisting of 4 ECE segments followed by an EUDC cycle. It is currently used for homologation in Switzerland and was used in the majority of the optimisation work as representative of typical mixed use.

The US06HWY, ECE, EUDC and ECE-EUDC cycles were used for the multi-cycle optimisation of fuel economy.

The drive cycles are shown in Figs. 6.8 and 6.9. During the optimisations all of the cycles were run with initial conditions set to standard ambient conditions - i.e. cold start.

The vehicle configuration problem posed particular difficulties not confronted in the other problems. Notably, with the complete change of vehicle configuration between conventional and hybrid drivetrains, some components are added and removed from the configuration, and the optimal dimensions of others differ greatly.

Independent Variable	Limits
ICE Size	2-74kW for SI,12-108kW for CI)
Final Drive Ratio	0.5 to 5 (needed for Prius which has final drive ratio of 3.94)
SOC_{high}	0.3 to 0.85
SOC_{low}	0.3 to 0.85
($SOC_{high} < SOC_{low}$ was considered infeasible)	
Electric Motor	4.8 to 58 kW
Electric Generator	4.8 to 58 kW
No. of Battery Modules	1 to 60

Table 6.2: Independent Variables and Limits

In order to deal with this the similarity comparison used to identify duplicates in the population was modified to take into account only those variables that were relevant for the configuration — for example the value of battery SOC for a conventional vehicle is clearly meaningless. The variables were implemented as scaling factors, so were equally appropriate for the vehicles with different basic components.

The unused variables do, however, contribute to the evolution of the population as a whole - when crossing two different vehicle configurations all the variables are taken into account and hence the variation in the population is preserved.
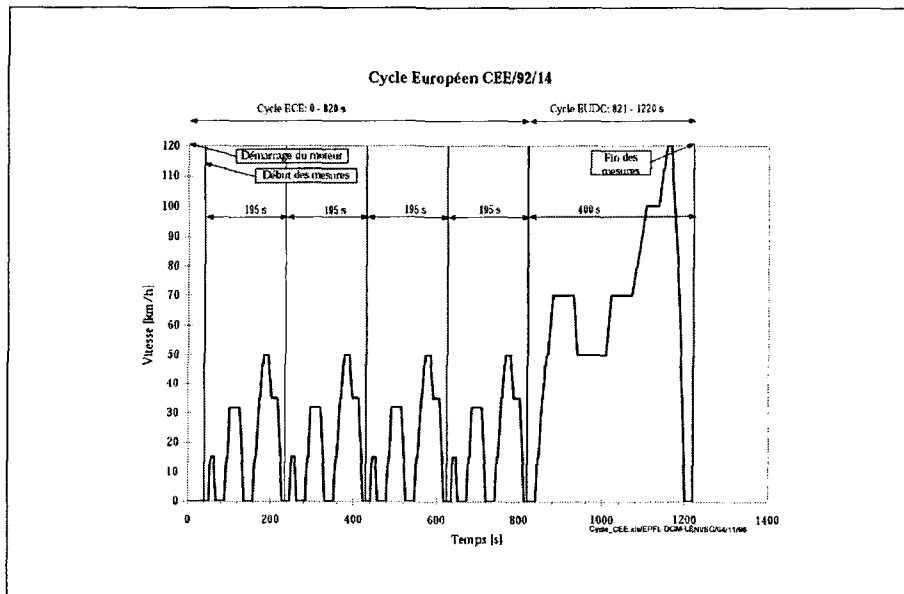
Fig. 6.8: The ECE-EUDC and ECE drive cycles.
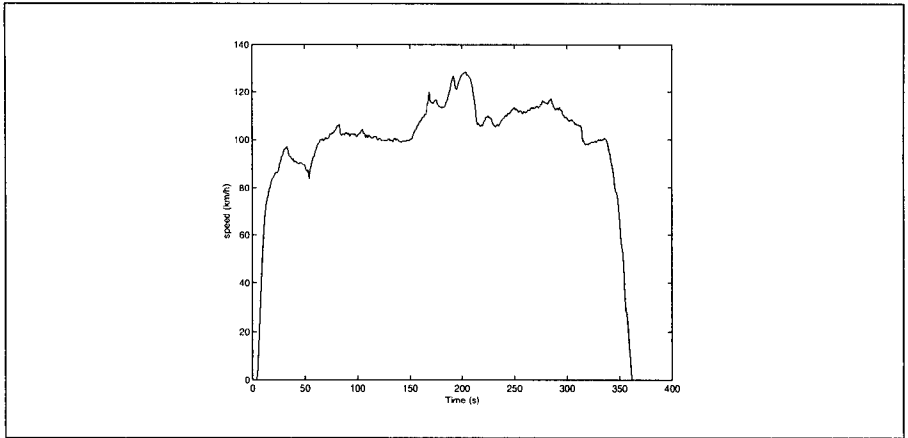
Fig. 6.9: The US06-HWY Drive Cycle.

In order to avoid unfairly biasing hybrid configurations with large battery capacities, the state of charge of the vehicle batteries at the end of the cycle was required to be within 0.5% of its initial state. To achieve this the cycle had to be run iteratively, adjusting control parameters until the required final charge state could be achieved. Hybrid vehicles which failed to satisfy this requirement were also removed from consideration.

A vehicle design was also considered infeasible if it could not follow the chosen cycle within 0.01km/h. This was chosen to be small since it was found that optimum solutions will tend to take advantage of the minimum acceleration required.

The vehicle was required to accelerate from 0 to 60mph in less than 12 seconds*.

This meant that the Prius (Japanese and European versions) could not meet the requirement as it stood.

Though the CPEA can perform optimisations with more than two objectives, results from these optimisations are difficult to visualise and interpret (and do not converge so well, Chapter 5). Consequently the work was restricted to a two-objective optimisation, where the results could be seen as simple trade-offs between the two objectives.

Unless otherwise indicated in the following work the CPEA was run with the parameters: $p_{initial}$=250, $n_{clusters}$=8 corresponding to one cluster for each vehicle type and with clustering limited to the vehicle type†. The relatively large initial population was used to ensure a good spread of solutions over the vehicle types, since it proved more difficult to find feasible solutions randomly for certain vehicle configurations‡

---

*The US Council for Automotive Research (USCAR) proposed the PNGV[96], which suggests this value.

†Where the Insight was not included the number of clusters was adjusted to 7.

‡The initialisation code could alternatively have been changed to produce an equivalent number of each vehicle type.

6.3.2   Economy vs NO_x

Earlier work[62] considered a conventional drive-train vehicle and optimised the size of the ICE and final drive ratio to minimise NO_x and fuel economy around the ECE-EUDC cycle, and this was repeated here. The results are by no means surprising but provide a simple example of the methodology. The NDF (Non Dominated Front) gives a clear illustration of the trade-off between NO_x and fuel economy, as shown in Fig. 6.10.
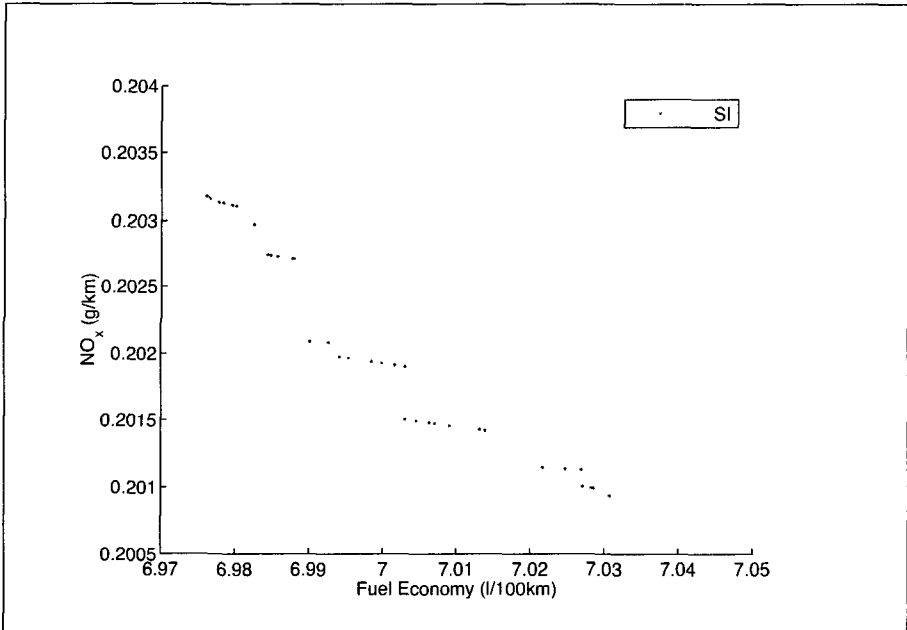
Fig. 6.10: Results from the optimisation of NO_x emissions and fuel economy over the ECE-EUDC cycle. Only the results for a conventional SI vehicle are shown.

Since fuel economy is nearly proportional to engine size this behaviour is to be expected - as the ICE gets smaller, the car gets lighter, the mean ICE regime approaches wide open throttle and engine pumping losses decrease[46].

Thus the best fuel economy is obtained at a very small engine size, below which the vehicle is no longer capable of following the drive cycle, or has to use a very high speed regime, pushing it into the high NO_x area on Fig. 6.2. As the ICE gets larger, fuel economy gets worse for the same reasons, but the NO_x emissions drop, because the engine is operating in a regime where cylinder gas temperatures are low, and less NO_x is formed. Effectively, the engine regime can move into the lower-left corner of Fig. 6.2 while still generating

sufficient power.

6.3.3   Fuel Economy Over Multiple Drive Cycles

The aim of this optimisation was to identify whether different vehicle configurations were, as might be expected, preferable for different driving patterns. Earlier work[62] had considered the US06HWY and ECE-EUDC cycle and had produced a slightly surprising set of results favoring parallel hybrid designs. With ADVISOR running on a Linux cluster with 22 machines the optimisation could be repeated in a fraction of the time taken in previous work. This quickly highlighted some faults in the earlier work that had been penalising conventional vehicles - indeed most were being marked as infeasible.

The objectives chosen were fuel economy over the drive cycle, where fuel economy for diesel engine vehicles was converted to the equivalent gasoline value. Two different combinations were run: the US06 HWY cycle (see Fig. 6.9) against the ECE-EUDC mixed cycle (see Fig. 6.8), and the US06 HWY against just the ECE urban cycle (4 cycles as shown in Fig. 6.8).

The initial results for the Prius were surprising — much worse fuel economy on the ECE cycle than on the other cycles. Closer inspection of these results showed that the Prius ran the ICE continuously. The control system will run the ICE until it reaches operating temperature[§], which happens more quickly on the more aggresive US06HWY cycle, and which happens on the ECE-EUDC cycle because it is longer.

In order to alleviate this fact the problem was run again with hot initial starting conditions.

Fig. 6.11 shows the NDFs from the current work using the US06HWY cycle and mixed ECE-EUDC cycle, and Fig. 6.12 shows the equivalent for the US06HWY and ECE only cycles. Table 6.4 gives values for the points labeled in Fig. 6.12 and Table 6.3 does the same for Fig. 6.11.

		ICE (kW)	Final Drive	SOC_{high}	SOC_{low}	Motor (kW)	Battery Modules
A	Insight	96.9	0.7	0.8	0.3	8.8	36
B	Parallel CI	35.8	0.9	0.7	0.5	41.6	28
C	Parallel CI	51.6	0.6	0.8	0.5	43.7	28
D	Prius	60.2	4.7	0.5	0.4	36.1	60
E	CI	68.2	1.0	-	-	-	-
F	CI	95.6	0.7	-	-	-	-
G	Parallel SI	32.2	1.2	0.6	0.5	50.8	35
H	Parallel SI	45.5	1.1	0.6	0.4	52.7	39
I	SI	74.1	1.5	-	-	-	-
J	SI	77.2	1.4	-	-	-	-

Table 6.3: Variable values for points labeled in Fig. 6.11 from the optimisation of fuel economy over the US06HWY and ECE-EUDC cycles.

[§]Practical experience with the Prius[33] confirms this behaviour.

Fig. 6.11: Two cycle fuel economy optimisation results for US06HWY and ECE-EUDC mixed cycle with hot initial starting conditions. The fuel consumption for diesel has been adjusted to an equivalent gasoline value. The vehicles must also meet the 0-60mph in 12s acceleration test.

	config	ICE (kW)	Final Drive	SOC_{high}	SOC_{low}	Motor (kW)	Battery Modules
A	Insight	98.2	0.7	0.7	0.3	8.9	38
B	Parallel CI	38.9	1.0	0.7	0.5	34.4	25
C	Parallel CI	53.8	0.6	0.7	0.5	44.9	24
D	Prius	59.3	5.0	0.8	0.3	35.5	58
E	CI	68.2	1.0	-	-	-	-
F	CI	105.2	0.7	-	-	-	-
G	Parallel SI	29.6	1.4	0.7	0.6	49.1	33
H	Parallel SI	48.2	1.1	0.8	0.6	46.9	28
I	SI	70.3	1.6	-	-	-	-
J	SI	76.8	1.4	-	-	-	-

Table 6.4: Variable values for points labeled in Fig. 6.12 from the optimisation of fuel economy over the US06HWY and ECE-EUDC cycles.

Overall Behaviour   The results are not quite as expected—notably there are no conventional drivetrains at all in the Pareto-optimal set, and these had been expected to perform well in the US-06 HWY cycle. The results show that there is a clear difference between performance on an urban cycle and a highway cycle, and that in general parallel hybrids are favoured over the other vehicle types.
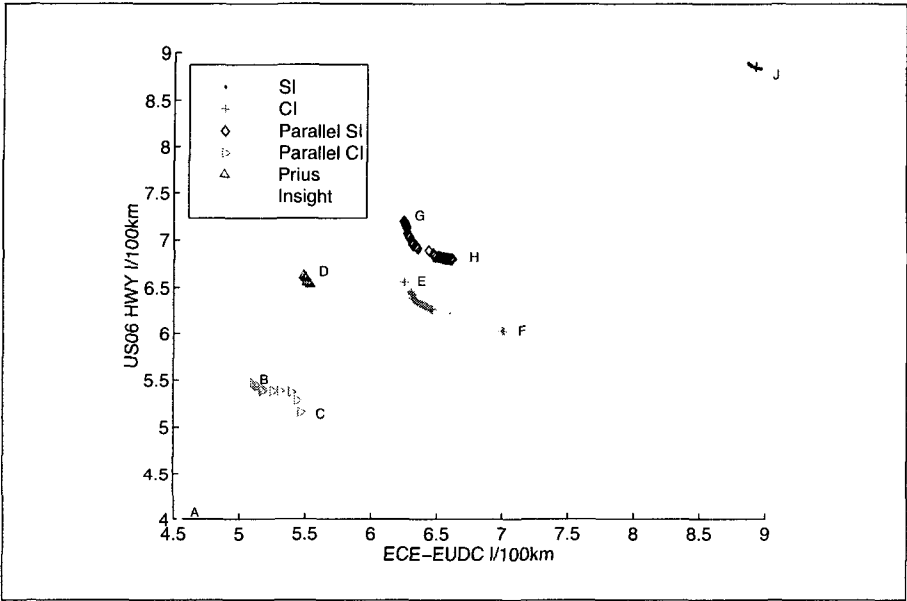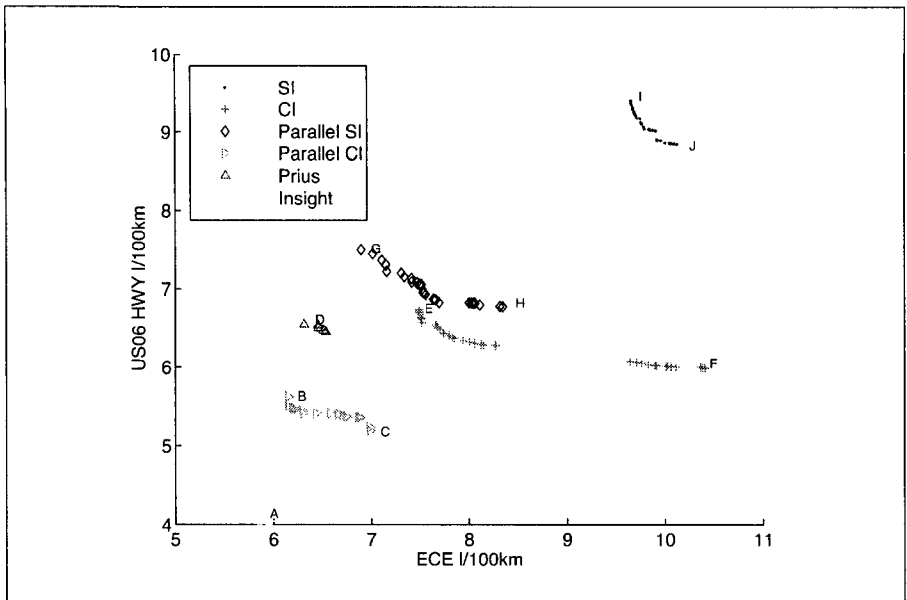
Fig. 6.12: Two cycle fuel economy optimisation results for US06HWY and ECE urban cycle with hot initial starting conditions. The fuel consumption for diesel has been adjusted to an equivalent gasoline value. The vehicles must also meet the 0-60mph in 12s acceleration test.

The series vehicles are heavily punished since the ICE must still be dimensioned for the acceleration, so that when cruising the ICE is not running at maximum efficiency.

It was speculated earlier that conventional drivertains should perform well on highway driving, as a hybrid drivetrain, at first view, offers no advantages in this mode. However, while the US-06 HWY cycle is for the most part at a constant high speed, it also contains a hard acceleration at the start of the cycle. Thus, in a conventional vehicle, the ICE must be sufficiently large to provide the acceleration and so is not dimensioned for economy.

The parallel hybrid configurations show a marked reduction in ICE size and consequently marked reduction in fuel consumption achieved by using an electric assist for the accleration phase.

As expected they still show the same basic trend.

The battery capacities in the cars that perform well on the US-06 HWY are used to provide acceleration, rather than storage — thus it is the battery's *power* density, rather than its *energy* density that is important.

All of the vehicles show the same trends with respect to ICE size. Vehicles with smaller ICEs do better on the ECE and ECE-EUDC cycles than on the US06HWY cycle and this is more noticable on the ECE cycle as might be expected. Here the ICE is being sized for the aggresive acceleration of the US06HWY cycle.

Note that the fuel economy results for the Prius do not show much in the way of a trade-off, and this is thought to be due to the control strategy that results in the ICE turning on and off much more frequently than with simple strategy in the other parallel hybrids.

The Insight had better economy on the ECE-EUDC than on the ECE, and better still on the US06HWY. Since it is much lighter than the other vehicle types (see Fig. 6.15 and Section 6.5.2) it has a distinct advantage for the acceleration. The small electric motor assists in order to further boost the performance on the acceleration phase of the US06HWY cycle, and the efficient ICE then supplies the required cruising ability.

6.4   Vehicle Costing

A very simple cost model was introduced to estimate operating costs, $C_{operating}$ and investment costs, C_{inv} as:

$$C_{inv} = c_{ice}P_{ice} + c_{elec}P_{elec} + C_{fix} \qquad (6.1)$$

where c_{ice}, c_{elec} are respectively the cost per kW of the IC engine and the electrical components, and P_{ice}, P_{elec} are the maximum rated power output in kW.

In order to account for parallel hybrid designs that have no generator the P_{elec} was taken defined as :

$$P_{elec} = P_{motor} + P_{generator} \qquad (6.2)$$

It was considered that a diesel engine, due to the higher compression ratio and high pressure common rail (or direct injection) would be 20% more expensive.

$$C_{CI} = 1.2c_{SI} \qquad (6.3)$$

The C_{fix} is taken to include the bodywork and all the anciliary components, and is assumed to be fixed and the same for a hybrid or conventional vehicle. In reality it is clear that this is a greatly simplified costing, since as engine power varies so does the cost of many associated components such as braking systems, suspension systems and tyres.

Operating costs were calculated as:

$$C_{operating} = c_{gasoline}M_{gasoline} + c_{diesel}M_{diesel} \qquad (6.4)$$

where $c_{gasoline}, c_{diesel}$ are respectively the cost per litre of gasoline and diesel, and $M_{gasoline}, M_{diesel}$ are the volume of fuel used over the assumed life of the vehicle. Values of $c_{gasoline} = 1.35sfr/l$ and $C_{diesel} = 1.36sfr/l$ were used.

The average distance driven in Switzerland (per vehicle) is given as 15,000km by the TCS (Touring Club Suisse)[89], and maintenance and running costs are typically calculated over 100,000km (ie a life of 7 years). However, in light of the 5 year / 100,000km guarantee offered on the hybrid system of the Prius, this was modified to assume 20,000km per year over 7 years as a more reasonable estimate, and to allow the expected better economy of hybrids to be more apparent.

For the purposes of this analysis the maintenance costs have been neglected - this is valid if the maintenance cost is more or less similar for the hybrid and conventional cars (see footnote on p94).

This suggests that maintenance costs are not very closely tied to technology but are much more dependent upon good design and marketing choices.

The electrical system cost was not broken up into individual components partly because detailed costs were not readily available, but also to keep the model simple. It would, however, seem reasonable to assume that the two major elements in the electrical system of a hybrid are the electric motor/controller assembly, and the batteries. The cost of the electric motor/controller can be expected to increase with rated power. The cost of batteries can be considered to be accounted for indirectly since they form a significant mass disadvantage that is included in the vehicle mass and hence apparent both in the acceleration test and the fuel use. Consequently solutions with fewer batteries will have lower operating (fuel) costs, and hence be better than solutions with an excessive number of battery modules.

To estimate initial values for c_{elec} and c_{ice} data was taken from a comparative review[5] of the performance and cost of the Prius, Peugeot 307 (gasoline) and the Ford Focus (diesel). Additional data was taken from tests performed by TCS[85,86,87]. The major characteristics of each are given in Table 6.5.:

Characteristic Data			
	Ford Focus	Peugeot 307	Toyota Prius
Fuel	Diesel	Gasoline	Gasoline
Drive Type	Conventional	Conventional	Hybrid
IC Engine Power	85kW	80kW	53kW
Electric Motor	-	-	Siemens 33kW
Electric Generator	-	-	15kW
Cost	27450chf	24000chf	38800chf

Table 6.5: Characteristic vehicle data used for cost model.

This resulted in initial values for $c_{SI} = 157$, $c_{CI} = 188$, $c_{elec} = 396$ sfr/kW. The calculated value for the cost of electric components, c_{elec}, was varied to determine its impact on optimisation results. Values of c_{elec} were chosen as 396, the nominal calculated value and 198, half the nominal value representing a major reduction in costs that might be imagined due to increased numbers being produced.

6.5 Investment Cost vs Operating Cost

This aim of this was to identify the relationship between investment costs and operating costs (as defined by fuel costs, since maintenance costs have been neglected for this study[‡], and to investigate the impact of reducing the cost of the electric components, C_{elec}.

The vehicles were evaluated over the ECE-EUDC drive cycle, initially with the nominal $c_{elec} = 396$sfr/kW, and the results are presented below. The value of c_{elec} was then halved and the optimisation rerun.

6.5.1 The Impact of Clustering

This optimisation was run both with and without clustering, to determine the impact of the clustering. The NDF for the case without clustering is shown in Fig. 6.13 with the NDF from the clustered optimisation superimposed and the NDF's from the clustered optimisation are shown in Fig. 6.14 . Several points are labeled on Fig. 6.13 and the corresponding independent variable values are shown in Table 6.6.

From the algorithm standpoint the NDF found with the clustering is slightly better than without - the results shown are for approximately 20000 function evaluations for each case. The overall best solution, as defined by $C_{inv} + C_{operating}$, (labeled at point **C**) is slightly better than the best point found without clustering (point **B**) although the difference is small (approximately 200sfr).

A close inspection of the absolute best values shows that a conventional SI vehicle has the miniumum investment cost (point **G**), while the minimum operating cost is a modified Insight. The lowest combined cost, $C_{operating} + C_{inv}$, is similarly an Insight based vehicle (point **A**).

More important than the very small difference in overall costs, the clustered case provides a lot more useful information. With the optimisation being run without clustering no information could be obtained about the vehicle configurations not on the POF. Either individual optimisations for each vehicle type would be needed or the investment cost / operating cost would need to be artificially modified to penalise the winning vehicles. Either of these solutions implies many more evaulations (up to 8 times as many to run each vehicle type separately).

6.5.2 Discussion

A complete set of graphs showing the variables for each of the vehicle configurations is given in A.2.4.

[‡]as previously noted the TCS figures demonstrate that maintenance costs are not easily predictable.
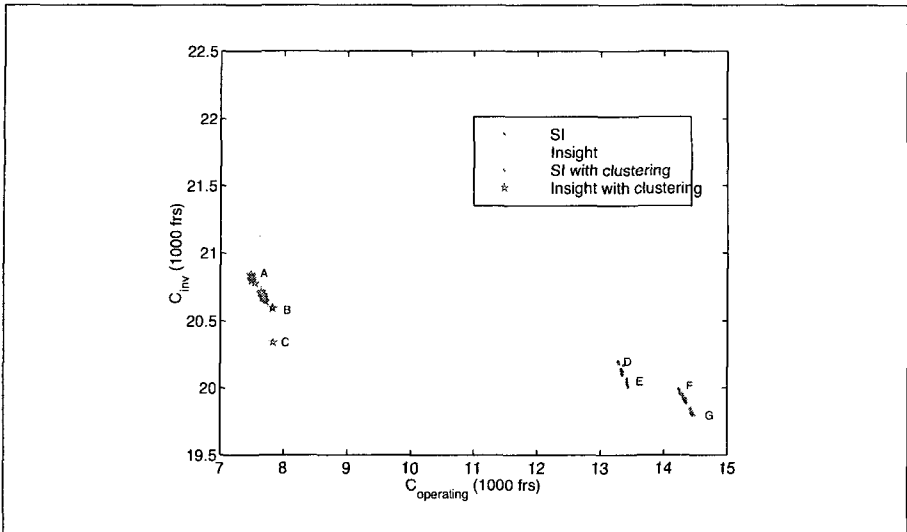
Fig. 6.13: Investment vs Operating Cost (without Pollution) over the ECE-EUDC cycle with 0-60mph 12s acceleration. Comparison of the NDFs between the optimisation with and without clustering, showing slightly better performance with clustering.

	Config.	ICE (kW)	Final Drive	SOC_{high}	SOC_{low}	Motor (kW)	Battery Modules
A	Insight	51.2	1.1	0.7	0.6	3.4	15
B	Insight	52.8	1.1	0.5	0.4	2.1	10
C	Insight	52.7	1.1	0.6	0.3	1.5	11
D	SI	55.6	1.1	-	-	-	-
E	SI	54.7	1.2	-	-	-	-
F	SI	54.4	1.4	-	-	-	-
G	SI	53.1	1.5	-	-	-	-

Table 6.6: Variable values for points A-G in Fig. 6.13 from the optimisation of investment cost vs operating cost over the ECE-EUDC cycle.

Overall Behaviour   Fig. 6.14 seems to show the clear dominance of the Insight configuration. The Insight and the Prius both have advanced engines [*], so can be argued to have comparable ICE costs. However, the Insight is a two door, two person vehicle, in contrast to the other vehicles which are taken as small four person vehicles, and the Prius which is a four door sedan. In addition to this the Insight has a light weight Aluminium structure.

Fig. 6.15 shows the vehicle mass shown plotted against the sum of investment and operating cost, C_{inv} + $C_{operating}$, and clearly shows the Insight to be the lightest and the Prius the heaviest, even more so than the

---

[*]VTEC in the case of the Insight and Atkinson cycle for the Prius
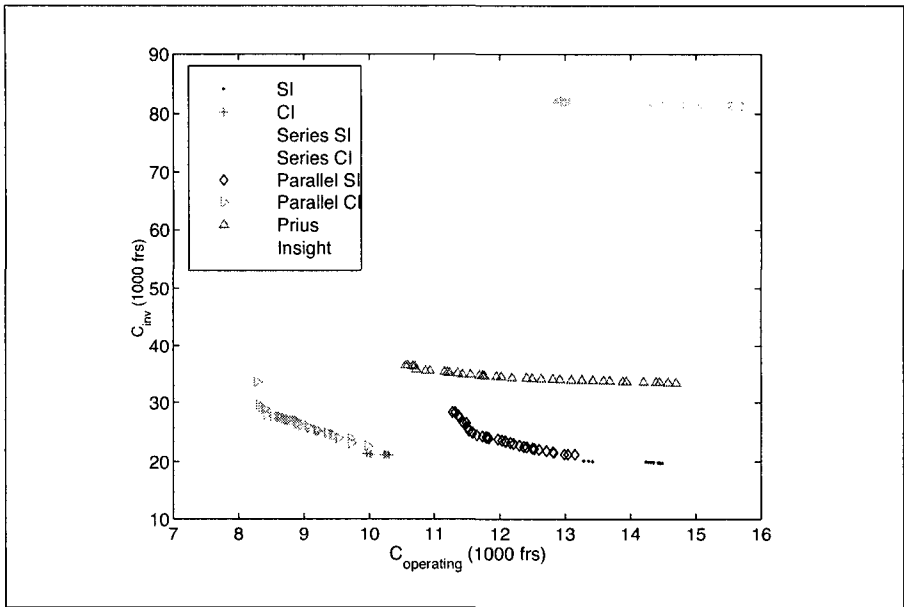
Fig. 6.14: Results of the optimisation of investment vs operating cost (no pollution) over the ECE-EUDC cycle with 0-60mph 12s acceleration.

series hybrids.

The Prius is probably unfairly punished in that no weight was added to the structure of any of the other vehicle types to account for an increase in drivetrain component weight - so the series hybrids are unrealistically light. Since fuel use and hence operating costs are directly related to vehicle mass [†] it is to be expected that a lighter vehicle will outperform a heavier one. This is alleviated to some extent in the hybrid vehicles since some of the energy can be recuperated, but still poses problems on continual climbs.

In general the parallel hybrids are heavier than the conventional vehicles due to the electric motor and batteries. Similarly the series hybrids are heavier than the parallel hybrids because of the extra need for a generator.

The lowest investment cost is for a conventional SI vehicle, the lowest operating cost for the Insight configuration. The Insight also has the minimum overall cost ($C_{inv} + C_{operating}$). The Insight does in fact have a very low level of hybridization - it is nearly a lightweight conventional car with advanced engine - the electric motor helps out with maximum acceleration.

[†] A reduction of n kg in vehicle weight can be considered to translate directly to a fuel economy of m *see OTA report - cant remember figures*

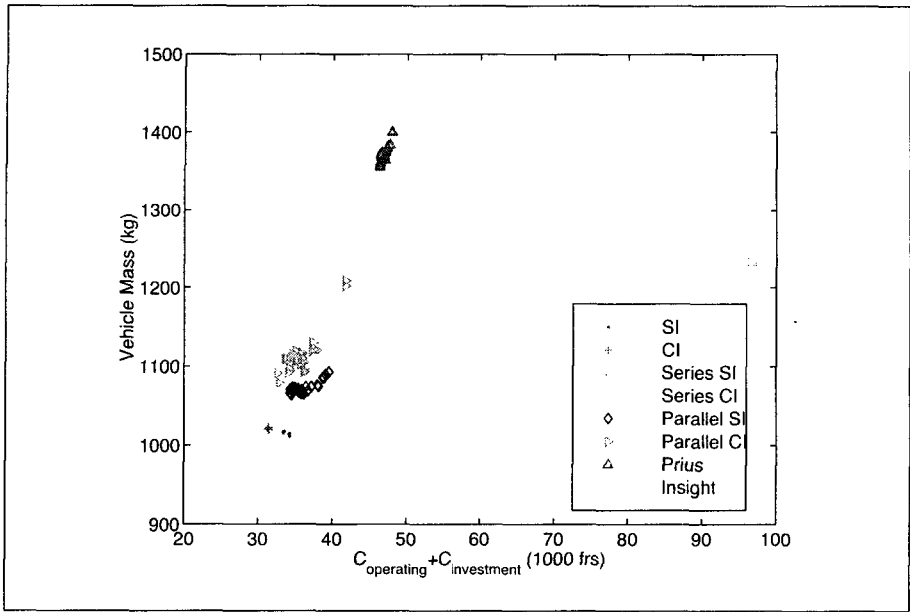Fig. 6.15: Vehicle Mass from optimisation of investment vs total costs (no pollution costs)

Conventional Vehicles   The conventional vehicles show a simple relationship between engine size and operating cost (see points **D,E,F,G** in Fig. 6.13). A smaller ICE must (in general) operate at a higher speed in order to generate the same power, with a corresponding increase in final drive ratio. This results in a reduced C_{inv} but at the expense of fuel economy and hence $C_{operating}$.

Series Hybrids   Fig. 6.14 shows that the series based hybrids are considered unfavorable, with much higher investment cost than any other vehicle configuration, and with operating costs higher than the equivalent alternative configurations. In general the series hybrids have a small (around 12-14kW) ICE with similarly sized generator, and a large motor. The ICE runs most of the time and the large battery storage is used to meet the intermittent loads. As the motor power decreases the generator power drops and hence the ICE power drops accordingly - all of these contribute to reducing C_{inv}. As with the conventional vehicles smaller ICE means higher ICE speed and so higher fuel consumption and hence higher $C_{operating}$.

Parallel Hybrids   The parallel hybrids are clearly more favorable, as might be expected since they do not suffer from the drivetrain losses experienced by the series hybrids. The ratio of electric motor power to ICE power (the degree of hybridization) is shown against the operating cost for each of the parallel hybrid

configurations in Fig. 6.16.

This shows clearly the tradeoff between investment cost and operating cost. A large motor with appropriately high number of battery modules and small ICE produce a high C_{inv} and low $C_{operating}$, since the ICE is operated at high efficiency with the motor meeting the acceleration requirements. As the motor size is reduced, the number of battery modules can also be reduced but the ICE power must increase to meet the demand and cope with the acceleration, hence increased fuel consumption and $C_{operating}$. As the motor size becomes very small (very small degree of hybridization), the NDFs for the parallel hybrids approach those for the conventional vehicles.

The final drive ratio drops as the motor size increases, since this reduces engine speed and hence reduces fuel use.

The SOC_{max} and SOC_{min} show in general a less smooth trend than the other variables,although both the CI and SI based configurations follow the same trend.

Prius The Prius is slightly odd in that it makes use of a generator to control the power flow between the batteries and wheels. The ICE and motor show the same trend as the parallel hybrids - increasing ICE power, decreasing motor power results in lower C_{inv} but higher $C_{operating}$ as described above. The generator power drops approximately in line with the motor, although less smoothly, as does the number of battery modules. There is a clear change of operation at small motor power (high ICE power), with the generator size dropping suddenly and the number of battery modules jumping to near the maximum (see Fig. 6.17). This will allow the ICE to charge the batteries for much of the working part of the cycle, with the batteries then providing extra torque via the motor when needed at high loads. The Prius shows slightly better operating costs than the simple parallel SI hybrid and better than the series hybrids.

6.5.3   Reduced Costs for Electric Components

Reducing the value of c_{elec} will lower the investment costs of all of the hybrid vehicle configurations. It was not necessary to re-optimise the problem - by decomposing the investment costs and recalculating the new costs the new curves could be reconstructed, and the resulting set of solutions ranked to provide a new set of NDFs.

The Insight remains the most favorable configuration (mild hybrid) until a drop in cost of electric components to around 25% of the current value, at which point the CI parallel hybrid takes over, due to the greater degree of hybridization.

If the Insight is removed from contention then the best solution becomes the conventional CI vehicle, and the parallel hybrid CI vehicle will become favourable if the cost of electric components drops to around 55%

.
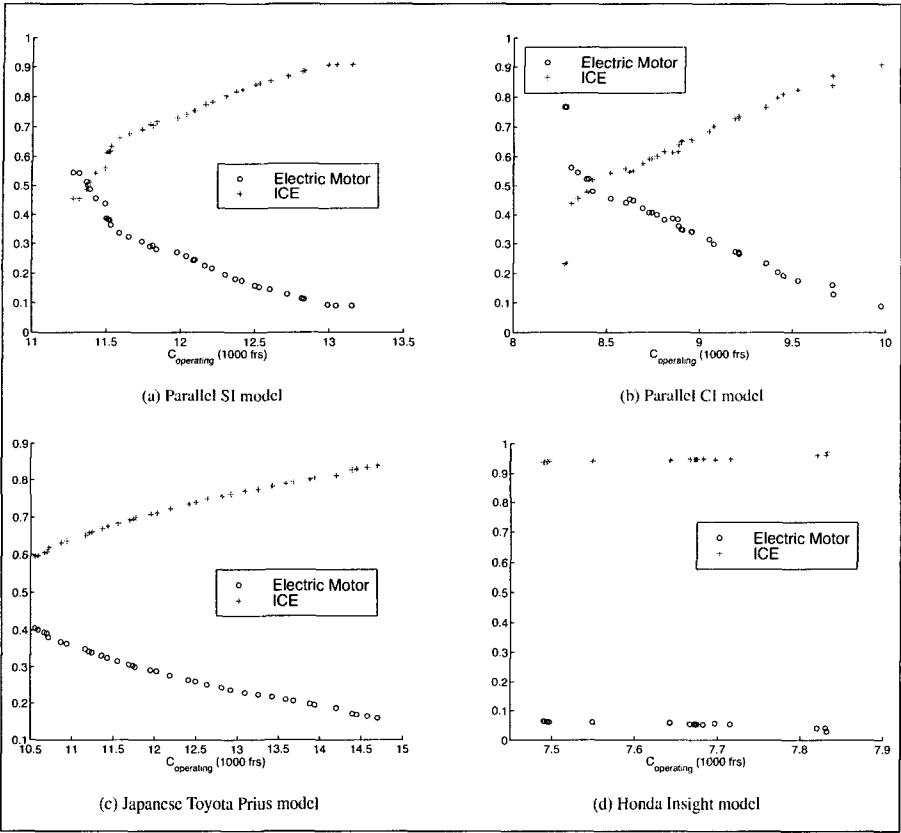
Fig. 6.16: Degree of hybridisation shown against operating cost, from the optimisation of investment vs operating cost.
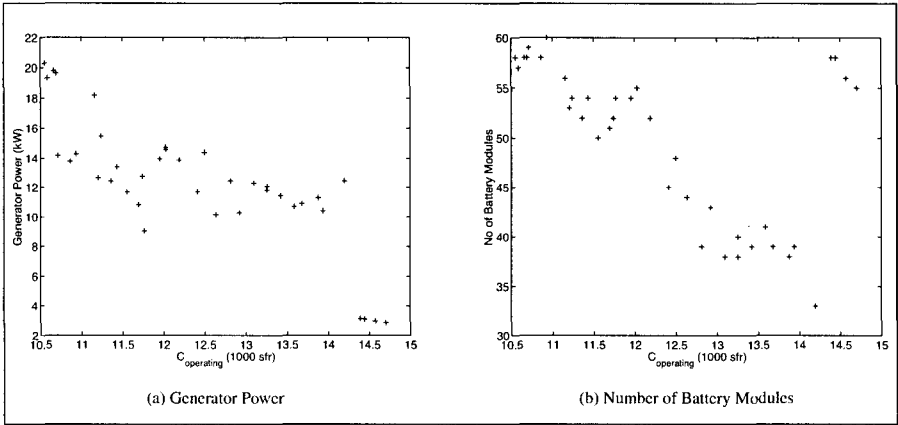
of the current value.

(a) Generator Power   (b) Number of Battery Modules

Fig. 6.17: Generator and number of battery modules for Prius configuration from optimisation of investment cost vs operating cost

6.6   Total Cost vs Pollution Cost

This was to investigate the impact of applying the pollution costs used in the district heating model to the domain of transport. The Insight is not included in this optimisation because engine emissions data was not available.

6.6.1   Pollution Cost Calculations

Pollution costs were limited to the cost of CO_2 and NO_x and the values used were the same as in the district heating network optimisation, notably 13.8 sfr/kg for NO_x and 0.03 sfr/kg for CO_2.

While the NO_x emissions are produced directly from the ADVISOR model, the CO_2 emissions were calculated from the fuel use as:

$$M_{CO_2} = (0.640 * 44/12)M_{gasoline} \tag{6.5}$$

where 0.640 represents the carbon content (in kg) of one litre of gasoline.

6.6.2   Results

A complete set of graphs showing the variables for each of the vehicle configurations vs the pollution cost is given in A.2.5.

Fig. 6.18 shows the NDFs for the optimisation of $C_{inv} + C_{operating}$ and C_{pol}. Fig. 6.19 shows the same information without the series hybrids with labels at certain sites. Variable values at the labeled points are given in Table 6.7.

		ICE (kW)	Final Drive	SOC_{high}	SOC_{low}	Motor (kW)	Battery Modules
A	CI	52.6	0.8	-	-	-	-
B	Parallel CI	46.3	1.0	0.6	0.4	6.2	14
C	Parallel CI	43.3	1.0	0.6	0.4	10.4	16
D	SI	54.3	1.2	-	-	-	-
E	Parallel SI	49.8	1.1	0.6	0.3	6.6	14
F	Parallel SI	25.7	1.1	0.6	0.5	34.5	30
G	Prius	50.8	5.0	0.6	0.6	26.3	54
H	Prius	43.7	4.8	0.6	0.6	33.0	56

Table 6.7: Variable values for points labeled in Fig. 6.19 from the optimisation of fuel economy over the US06HWY and ECE-EUDC cycles.
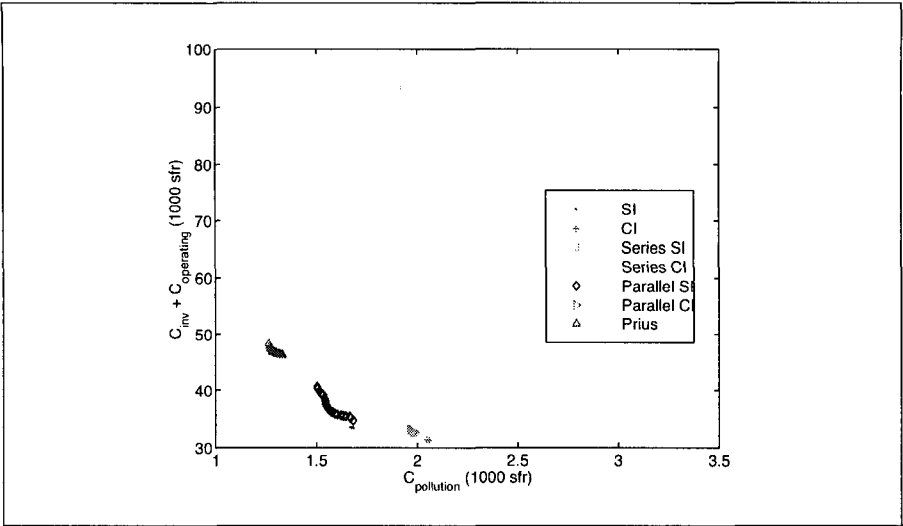
Fig. 6.18: Results of the optimisation of investment and operating cost vs pollution cost
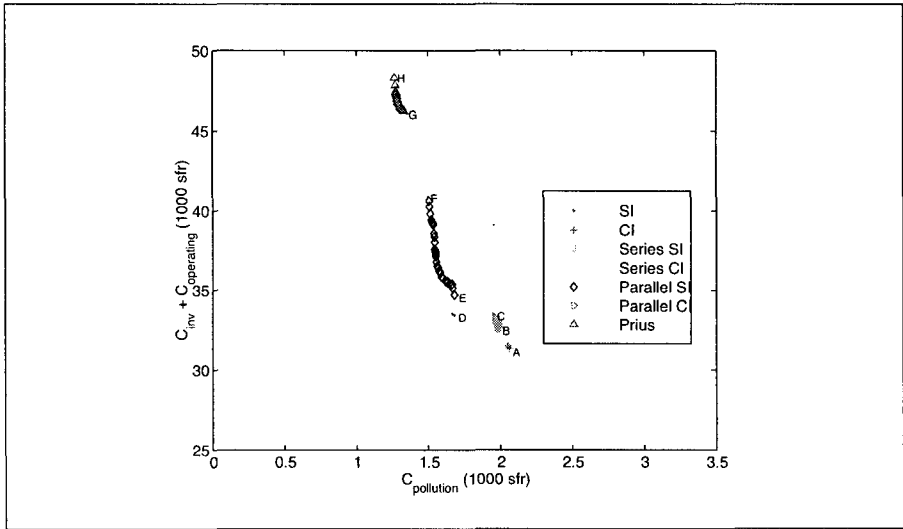

Fig. 6.19: Results of the optimisation of investment and operating cost vs pollution cost

For the specific pollution costs considered the calculated pollution costs are so low as to be insignificant compared to the overall investment and operating cost over the supposed life time — indeed comparing the results to the operating cost shows that it can be expected to have only a limited impact. This is in contrast to the heating domain where the optimum solution is very sensitive to pollution costs. This is largely due to the already high level of taxes imposed on the fuel prices, that already acts to improve fuel economy and swamps the direct pollution costs.

The cheapest, but most polluting solution is a conventional diesel car, while the least polluting but most expensive is the Prius* configuration. Indeed, only the diesel series hybrid configuration is more polluting than the conventional diesel configuration, as might be expected since it makes use of the IC diesel engine continually during the cycle.

The overall best solution as defined by $C_{inv} + C_{operating} + C_{pol}$ is also the conventional CI vehicle. The pollution cost must be increased five fold before there is a change towards conventional SI vehicles, and nearly thirty fold before there is a change towards a hybrid CI vehicle. It should be stressed, however, that the pollution costs do not include particles or hydrocarbons or any of the many other potentially harmful emissions[32].

6.6.3   Discussion

Conventional vehicles   Both the SI engine and CI engine vehicles show very little tradeoff between pollution cost and total cost. In the case of the SI engine vehicle approximately 40% of the investment and operating cost is operating cost (fuel cost), and this contributes to approximately 65% of the pollution costs (CO_2). The small tradeoff is due to a small (0.5%) change in engine size that allows the engine to be more fuel efficient while keeping the NO_x nearly constant. This is due to the specific characteristics of the engine efficiency and NO_x maps and can only happen over a very small change in engine size.

For the CI engine vehicles only 30% of the investment and operating cost is the operating cost and this contributes to only 35% of the pollution costs, the majority being due to NO_x. The more significant NO_x costs dominate the pollution cost and account for the tradeoff - a bigger engine produces more NO_x. This is again due to the specific characteristics of the engine maps.

Series Hybrids vehicles   The series vehicles suffer from high cost due to the generator and large motor costs. This aside the pollution costs are higher than the equivalent conventional vehicles since the ICE spends more time running at high efficiency. The SI series hybrid shows almost no tradoff. As for the conventional equivalent the pollution costs are dominated by CO_2 and so follow closely the operating costs (the two objectives are very nearly the same). The CI series hybrid pollution costs are due mainly to the

---

*The Insight was not considered here since no emissions were available.

NO_x costs - small ICE gives low operating costs and investment costs but high NO_x costs, and conversely. The electric motor size is fixed by the cycle requirements, and the generator more or less fixed by the ICE size, so there is little room for manoeuvre.

Parallel Hybrid vehicles   Both the SI and the CI parallel hybrids show the same trend - smaller ICE and bigger electric motor gives improved fuel economy and lower pollution, and this is clear in Fig. 6.20. It is more noticeable in the case of the SI engine since the CI engine hybrid is penalised by the increased NO_x production of more efficient CI engines. The CI hybrids tend to have bigger ICE (and a lower degree of hybridization), since the diesel ICE are themselves more efficient. Consequently they have fewer battery modules.

Prius   The Prius configuration follows the same trends as the SI parallel hybrids. Reducing the ICE size improves fuel economy and so reduces CO_2 but at the cost of increased NO_x - however the CO_2 effect is dominant, hence reduced engine size leads to reduced pollution cost. Investment increases because of the larger proportion of power from the electric motor (see Fig. 6.20) hence larger investment.

6.6.4   Effect of Decreased c_{elec}

The NDFs for the optimisation with $c_{elec} = 198$sfr are shown in Fig. 6.21. Halving the c_{elec} used in the calculation of C_{inv} brings the cost of the electrical components down to nearly the same as the cost of the ICE engines.

This results in a change of overall best solution from a conventional CI vehicle to a parallel hybrid CI (although the extremely low pollution costs still have a minimal influence). Increasing pollution cost by five fold moves the overall best solution to a SI parallel hybrid, and an increase of 30 fold favors the Prius configuration.

The range of parallel hybrid solutions is reduced since the difference between c_{elec} and c_{ice} was the major driving force in the tradeoff between size of motor and size of ICE. This is apparent in the graphs showing the ratio of P_{elec} and P_{ice} in Fig. 6.22.

(a) Parallel SI model

(b) Parallel CI model
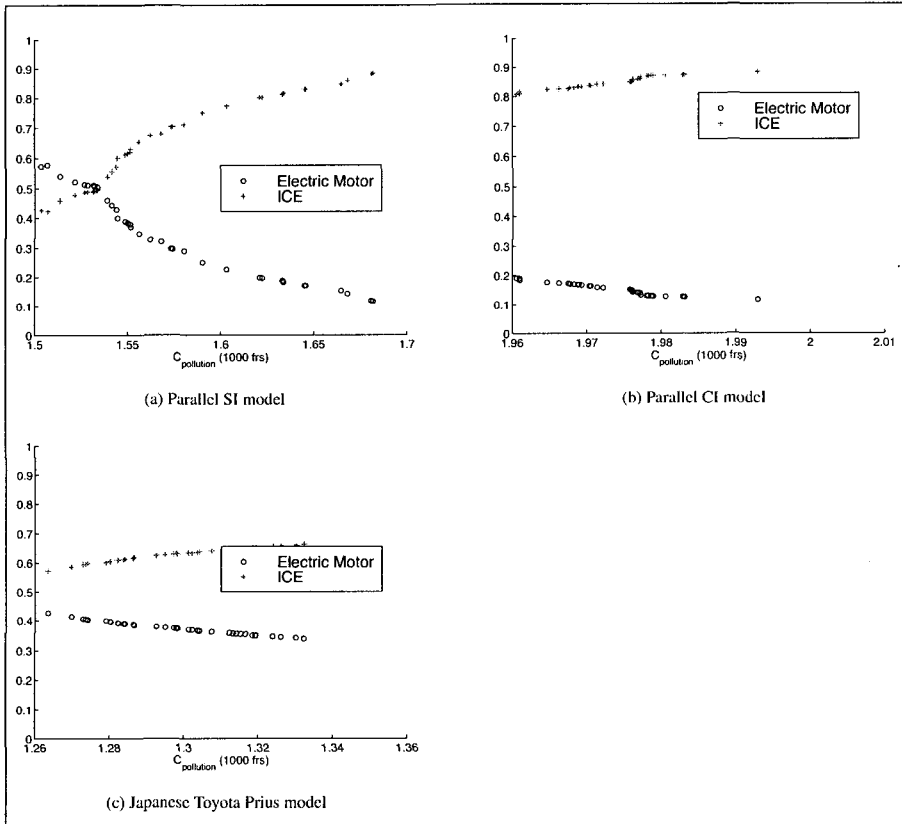
(c) Japanese Toyota Prius model

Fig. 6.20: Degree of hybridisation shown against operating cost, from the optimisation of investment vs operating cost.
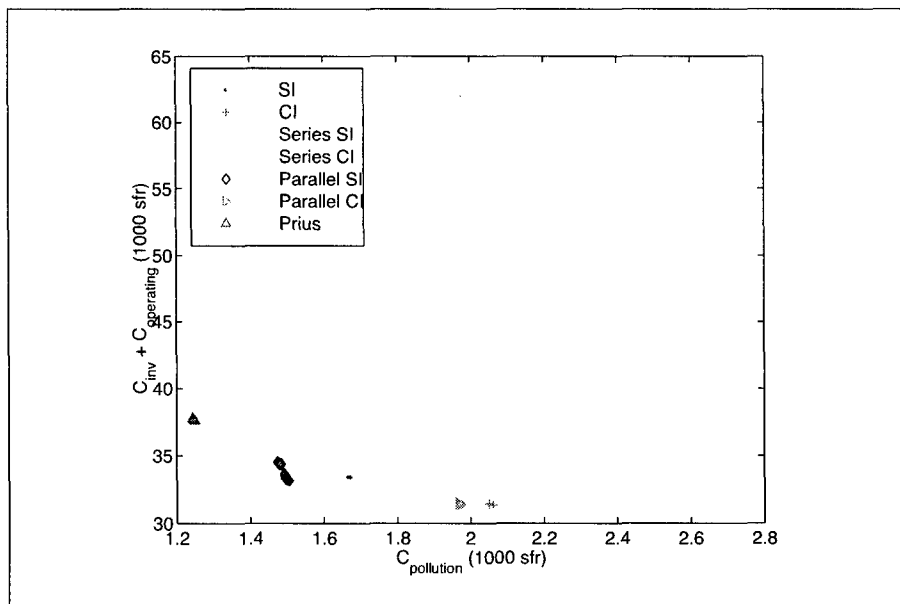
Fig. 6.21: Results of the optimisation of investment and operating cost vs pollution cost over the ECE-EUDC with electric component cost reduced to 198sfr/kW.

(a) Parallel SI model

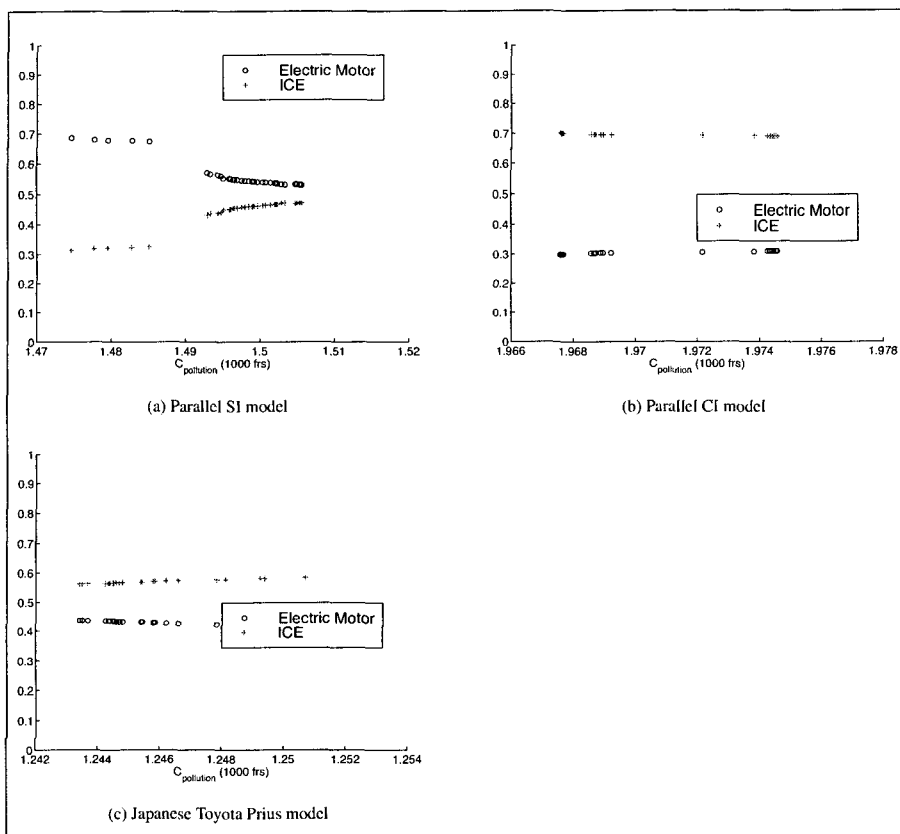(b) Parallel CI model

(c) Japanese Toyota Prius model

Fig. 6.22: Degree of hybridisation shown against operating cost, from the optimisation of investment vs operating cost with electric component costs reduced 198sfr/kW .

6.7 Total Cost vs Quantity of NO$_x$

NO$_x$ emissions have been shown to have a direct impact on human health [20,28] and vehicle NO$_x$ emissions are of particular importance because they produce NO$_x$ in areas where it impacts directly on people. Considering NO$_x$ vs operating and investment cost provides a clearer picture of the hybrid behaviour, particularly since the operating costs are fuel costs and so are directly proportional to the CO$_2$ produced.

Fig. 6.23 shows the NDFs from the optimisation with the series hybrid not shown for greater clarity. The ratio of ICE to electric motor power are given in Fig. 6.24. Table ?? shows values of the independent variables at the points labeled in Fig. 6.24.
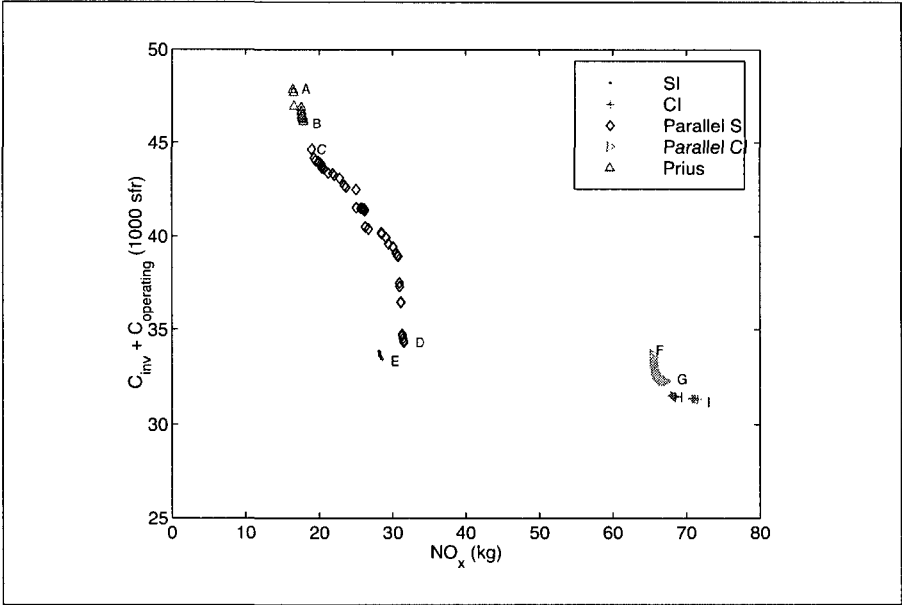

Fig. 6.23: Total Cost vs Quantity of NO$_x$ output over the ECE-EUDC cycle.

The best economy and lowest investment cost hybrid can be expected to use the IC engine at its full capacity for as long as possible to reduce fuel consumption by running the IC engine at its most efficient, and to reduce the size of the IC engine. However this means high NO$_x$ emission for the IC engine.

6.7.1 Results

A complete set of graphs showing the variables for each of the design configuration are given in A.2.6
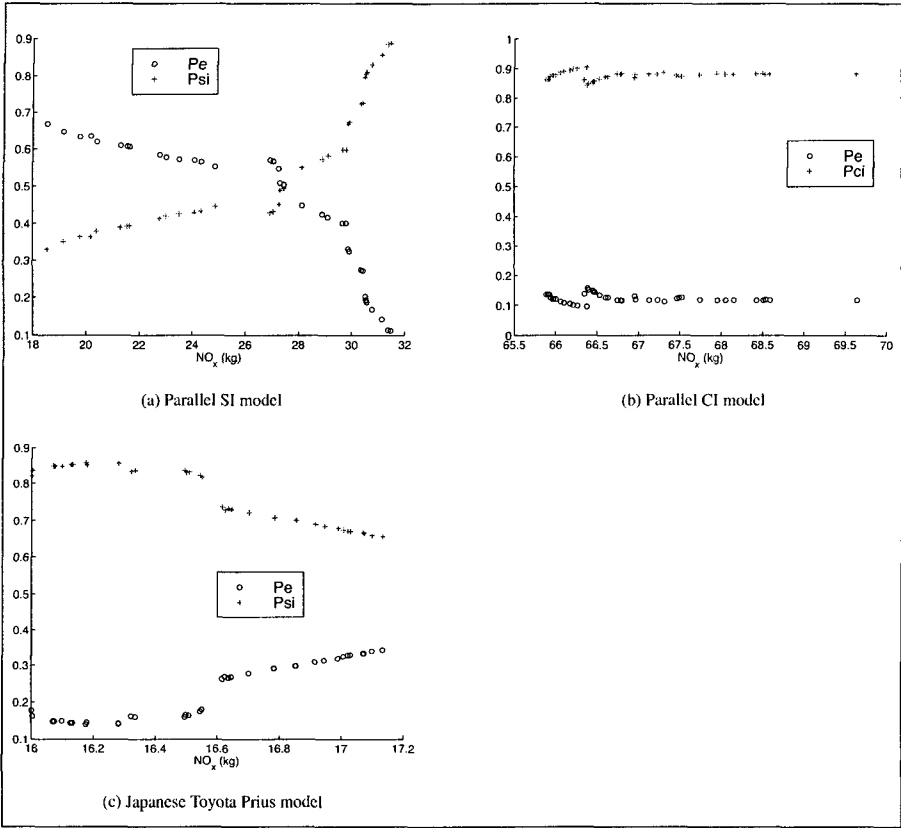
Fig. 6.24: Degree of hybridisation shown against quantity of NO_x output, from the optimisation of total cost vs NO_x.

The global Pareto front contains hybrid gasoline vehicles (**C,D**), hybrid diesel vehicles (**F,G**), conventional gasoline vehicles (**E**) and conventional diesel vehicles (**HI**). This will change if the investment cost of the hybrid electric components is reduced. The conventional vehicles tends to disappear as c_{elec} is reduced, as does the diesel hybrid.

Conventional vehicles   The SI and CI conventional vehicles both follow the same trend as the total cost vs pollution cost optimisation. The SI engine vehicle follows the expected trend - as the ICE size reduces the engine runs more efficiently but produces more NO_x. The CI engine vehicle similarly follows the same trend as the total cost vs pollution cost optimisation, with a small increase in CI engine size leading to an

	config	ICE (kW)	Final Drive	SOC_{high}	SOC_{low}	Motor (kW)	Battery Modules
A	SI	54.3	1.2	-	-	-	-
B	Parallel SI	47.0	1.2	0.7	0.6	6.7	14
C	Parallel SI	53.7	1.2	0.6	0.4	7.9	14
D	Parallel SI	27.5	1.2	0.7	0.7	30.0	23
E	Prius	53.7	5.0	0.7	0.6	24.8	59
F	Prius	64.0	5.0	0.7	0.6	23.2	58
G	Parallel CI	47.4	1.0	0.5	0.4	8.2	14
H	Parallel CI	46.9	1.0	0.8	0.3	5.1	14
I	CI	51.9	1.0	-	-	-	-

Table 6.8: Variable values for points A-I in Fig. 6.23 from the optimisation of overall cost vs quantity of NO_x. produced over the ECE-EUDC cycle.

increase in NO_x. In both cases the changes are fairly small.

Series Hybrids   Again these follow the same trends as the total cost vs pollution cost optimisation.

Parallel Hybrids   The CI parallel hybrid shows a reversed trend for the ICE, as does the Prius.

6.8   Conclusions

The most striking result is the inadequacy of the pollution costs for CO_2 and NO_x. These result in an overall pollution cost for all the vehicles that is negligible compared with the investment cost and the cost of fueling a vehicle over its life, without even considering the maintenance costs.

This is in contrast to the district heating case where the best result is highly sensitive to the inclusion of pollution costs, and even a small pollution cost causes the solution to radically change technology.

From a political point of view, the idea of a CO_2 tax accross all sectors (heating, transport etc) would encourage solutions such as the fully electric vehicle, where it would be advantageous to produce energy at the central power station. The current situation with high taxation already on fuels used for transport overwhelmes the additional direct cost of pollution taxes, which would need to be increased five fold to have any significant effect.

It should be noted that the direct effect of many other pollutants (for example hydrocarbons, carbon monoxide, and particulates) on health have been ignored in this study, and a health tax might be more effective, linked to the toxicity.

It would be interesting to examine in more detail the control strategies, since the SOC_{min} and SOC_{max} vari-

ables used in the optimised appeared to have only a secondary impact on the hybrid behaviour (although were probably necessary in order to find feasible solutions).

It should be noted when comparing the results from the Honda Insight model to the Toyota Prius and other simple vehicles that the Honda Insight is a two person vehicle with a light weight chassis, in contrast to the other vehicles.

Chapter 7

Conclusions

7.0.1 Clustering Pareto Evolutionary Algorithm

An evolutionary multi-objective optimisation algorithm with clustering, the CPEA, has been developed to solve fundamentally multi modal problems. The CPEA has been developed to be robust, and to produce a well defined non-dominated front, approximating the Pareto front, quickly and without the need for problem specific tuning. The CPEA sucessfully finds multiple local Pareto fronts and demonstrates that multi-objective optimisation is a viable alternative to aggregation of conflicting objectives - giving better use of each function evaluation as well as extra information at no cost. In addition the CPEA has beed shown to be practical on large computationally intensive problems.

Initial testing on two objective problems from the literature shows that the CPEA is capable of solving a wide range of problems, as well as specifically multi-modal problems. Preliminary comparisons with existing algorithms indicate that the CPEA finds the non-dominated front in fewer function evaluations than the best of the competitors, and produces an equally well or better defined front.

The use of clustering in the independent variable space has been shown to be an effective method of preserving diversity in the evolving population, and has shown itself able to identify multiple different solutions in multi-modal problems. It is considered a key element in the CPEA strategy, directly responsible for the success in solving the district heating test case.

Several clustering techniques have been investigated, and one of these, the c-means fuzzy clustering algorithm, has been adopted as a successful compromise between computational effort and stable, well defined clusters. A technique for determining natural clusters with c-means clustering has been found in the literature and shown to be adequate at keeping clusters, although with a computational overhead. Determining the number of clusters automatically does not seem to adversely affect the number of evaluations to converge to the Pareto front, although there is a large overhead in computation. Clustering on non multi-modal problems

129

would seem to slow the convergence by a factor of approximately two.

The use of clustering requires calculating distances or differences between solutions, which poses a problem in a many dimensional space. To deal with this method of multi-dimensional scaling, principal component analysis, has been used to reduce the number of dimensions before clustering. This has been shown to work on a 10 variable problem from the literature and on a real test case, the district heating network. Indeed the multi-dimensional scaling before clustering proved to be essential to the solution of the many variable district heating problem. As a side effect the reduction of dimension speeds up the clustering process (dramatically for problems with many variables).

7.0.2   District Heating Problem

The CPEA was successfully applied to the problem of designing a district heating network around a central heat pump and considering pollution, and new insight has been gained into the behaviour of the system as a whole.

Originally this problem was tackled using environomic analysis to internalize the costs associated with pollution, and optimised with a single objective GA, showing that the optimal solution considering pollution was completely different to the optimal solution without pollution, demonstrating the sensitivity of the optimal solution to imposed pollution costs in this domain.

The results of the original work were verified and the CPEA proved effective in greatly reducing the number of evaluations to solve the original single objective problem. The original problem was transformed into a two objective optimisation problem by isolating the costs associated with pollution from all other costs. The non-dominated front from this optimisation included the optimum solutions both with and without pollution, along with the form of the intermediate tradeoff surface, all for significantly less than the number of evaluations necessary for a single solution in the original study. The non-dominated front was disjoint with several locally non-dominated fronts.

Multi-dimensional scaling and clustering were necessary in order to fully solve this problem. Without them the CPEA only found the parts of the non-dominated front corresponding to low pollution cost configurations, since the high pollution cost configurations, while having lower capital and operating costs, were difficult to produce. The clustering also produced locally non-dominated fronts corresponding to different central plant configurations. Further post analysis of the reduced number of variables indicated that network temperature was clearly the dominant variable.

In the earlier work a simple parametric study of the network supply temperature was performed, by fixing its value and performing an optimisation to find the optimal solution. This could only be repeated at a few different temperatures since it required many function evaluations per optimisation. An alternative to this procedure was proposed, using the CPEA to first maximise, then minimise the value of network supply

temperature while minimising the original single objective. This produced a much more detailed parametric optimal study, allowing a detailed investigation of the solution domain that was not possible with a single objective optimisation algorithm, since it would require too many optimisations and hence a prohibitive computational cost. Several interesting features of the problem were revealed, that were not noticed in the original study, such as the importance of the domestic hot water supply in determining the optimal configuration, and identifying clearly the energy cost as the critical factor.

Multi-objective optimisation was shown to be a valuable alternative to environomic internalisation of costs, allowing uncertain parameters to be kept out of the optimisation process. The environomic analysis could then be applied afterwards in "real" time, allowing the uncertainty in calculating pollution costs to be investigated quickly without re-optimisation. This provides a dramatic saving in computational effort and elapsed time.

A three objective version of the district heating network problem considering electricity costs, investment costs and gas costs was optimised with limited success. Convergence to the non-dominated front was found to be slower and not as good as in the two objective cases due largely to the problems of thinning with three or more objectives.

7.0.3 Vehicle Drivetrains

The CPEA has also been applied to the problem of vehicle drivetrain analysis, which posed several specific difficulties, notably the time dependent simulation and requirement to model complicated sub-components efficiently. A simulink model of a vehicle, ADVISOR, was used to evaluate conventional, series electric hybrid and parallel electric hybrid drivetrain configurations over the ECE-EUDC and US06HWY drive cycles, and evaluate performance in terms of CO_2, NO_x, fuel economy, estimated investment, operating and pollution costs. To optimise this problem it was necessary to parallelise the CPEA, and develop a robust problem handling mechanism to deal with infeasible proposed solutions that caused the modeling system to crash.

The results demonstrated once again the advantage of multi-modal multi-objective optimisation in providing the maximum information per evaluation in a situation where solving the simulation is time-consuming, and showed the feasibility of optimising such complex models.

The parallel hybrids were shown to be preferable to series hybrids in all cases studied, but were frequently beaten by conventional vehicles. In particular diesel engine vehicles prove difficult to beat in terms of operating cost, or even overall investment and operating cost, since they have very good fuel economy. Even including pollution costs as calculated in the district heating problem was not sufficient to change the overall best solution. Indeed it was found that a five fold increase in pollution costs would be needed before any difference would be observed.

The Honda Insight, a mild hybrid with ultra light weight structure, was found to be a highly attractive configuration. However, the Insight is a two person vehicle and care is needed in comparing the performance with the Toyota Prius and other simple four person vehicle configurations.

It is interesting to note that the pollution costs were found to be insignificant in comparison with the operating and investment costs of a vehicle. This is in contrast with the heating domain where the same pollution costs per kg of CO_2 and NO_x prove to be extremely significant, changing the overall best solution significantly. A large part of this effect is suspected to be due to the high level of tax already applied to fuel in the transport domain, and hence the already high pressure to reduce fuel consumption.

Politically it would seem that applying a uniform CO_2 cost accross the board to all energy domains will have a limited effect on the transport industry. Perhaps more interesting would be to attribute more penalties to NO_x and other pollutants that are known to be directly harmful to human health.

7.1   Future Work

The Advisor models currently take into account the changes in vehicle mass as the engine and other components are scaled. However the models do not take into account secondary changes - for example increasing the engine size may require uprating the braking and suspension system. This in turn may require increased chassis rigidity adding even more mass. The cost model used could be improved, for example to consider the cost of the batteries separately from the electric motor and controller.

It would be interesting to consider the impact of other pollutants such as particulates on vehicle choice, and also to consider the impact on human health directly as an objective rather than indirectly via pollution.

More work is needed on solving problems with three and more objectives. The current thinning algorithm (for three or more objectives) favors points just behind the pareto optimal front and results in too many points in the population. This in turn slows convergence and overall performance. The impact of variable scaling on clustering and optimisation, particularly the interaction with crossover on certain problems needs more investigation.

References

[1] www, 2002. URL http://www.dieselnet.com/standards/cycles/ece_eudc.html.

[2] E. Amundsen. A distributed object based analysis of environmental automobile policies. Master's thesis, MIT, Jun 1999.

[3] J. Andersson, P. Krus, and D. Wallace. Multi-objective optimization of hydraulic actuation systems. In *Proceedings of the ASME DT Conferences, DETC/DAC-14512,Baltimore*, 2000.

[4] J. Andersson and D. Wallace. Pareto optimization using the struggle genetic crowind algorithm. submitted, under review, 2000.

[5] R. Automobile. Essence, diesel ou hybrid? *Revue Automobile*, 39, September 2001.

[6] R. Bauman. Drive-train optimisation for hybrid vehicles. Technical report, LENI, EPFL, Switzerland, June 2002.

[7] R. Bellman. *Adaptive Control Processes: A Guided Tour*. 1961.

[8] J. Bezdek. Cluster validity with fuzzy sets. *Cybernetics*, 3(3):58–71, 1974.

[9] C. M. Bishop. *Neural Networks for Pattern Recognition*. Oxford University Press, ISBN 0-19-853864-2.

[10] C. Bliek, P. Spelluci, L. Vicente, A. Neumaier, L. Granvilliers, E. Huens, P. Hentenryck, D. Haroud, and B. Faltings. Algorithms for solving nonlinear constrained and optimization problems: The state of the art. Technical report, Deliverable D1,COCONUT IST Project Funded by the EU, June 2001. URL http://www.mat.univie.ac.at/~neum/glopt/coconut.

[11] C. C. H. Borges and H. J. C. Barbosa. A non-generational genetic algorithm for multiobjective optimization. In *Proceedings of the 2000 Congress on Evolutionary Computation CEC00*, pages 172–179, La Jolla Marriott Hotel La Jolla, California, USA, 6-9 2000. IEEE Press. ISBN 0-7803-6375-2. URL citeseer.nj.nec.com/borges00nongenerational.html.

[12] S. Burch. Trading off hev fuel economy and emissions through optimization. Presentation, see URL, 1999. URL http://www.ctts.nrel.gov/analysis/reading_room.html.

[13] R. c. Dubes and A. K. Jain. *Algorithms for Clustering Data*. Prentice Hall, 1988.

[14] S. Chiu. Fuzzy model identification based on cluster estimation, 1994.

[15] C. A. C. Coello and A. D. Christiansen. Two new ga-based methods for multiobjective optimization. *Civil Engineering Systems*, 1998. URL http://www.lania.mx/~ccoello/moo.html.

[16] C. A. Coello Coello. A Comprehensive Survey of Evolutionary-Based Multiobjective Optimization Techniques. *Knowledge and Information Systems*, 1(3):269–308, 1999. URL citeseer.nj.nec.com/coello98comprehensive.html.

[17] C. A. Coello Coello. An updated survex of evolutionary multiobjective optimization techniques: State of the art and future trends. *Knowledge and Information Systems*, 1(3), 1999.

[18] E. Commision. European commission,directorate general for energy, annual energy review, 1996.

[19] T. F. Cox and M. A. A. Cox. *Multidimensional Scaling*. Chapman and Hall, 1994.

[20] P. Crettaz, K. Brand, L. Rhomberg, and J. O. Effects on human health of compounds causing non cancer toxicity. *International Journal of LCA*, 1999.

[21] V. Curti. *Modélisation et Optimisation Environomiques de Systèmes de Chauffage Urbain Alimentés par Pompes à Chaleur*. PhD thesis, École Polytechnique Fédérale de Lausanne, 1998.

[22] V. Curti, M. von Spakovsky, and D. Favrat. An environomic approach for the modeling and optimization of a district heating network based on centralized and decentralized heat pumps, cogeneration and/or gas furnace (part I: Methodology). *International Journal of Thermal Sciences*, 39:721–730, 2000.

[23] P. J. Darwen and X. Yao. A Dilemma for Fitness Sharing with a Scaling Function. In *Proceedings of the Second IEEE International Conference on Evolutionary Computation*, Piscataway, New Jersey, 1995. IEEE Press. URL citeseer.nj.nec.com/darwen95dilemma.html.

[24] K. A. De Jong. *An Analysis of the Behaviour of a Class of Genetic Adaptive Systems*. PhD thesis, University of Michigan, Ann Arbour, MI, 1975.

[25] K. Deb. Multi-objective genetic algorithms: Problem difficulties and construction of test problems. *IEEE Transactions on Evolutionary Computation*, 7(3):205–230, 1999. URL citeseer.nj.nec.com/deb99multiobjective.html.

[26] K. Deb. *Multi-Objective Optimization using Evolutionary Algorithms*. Wiley, Chichester, 2001.

[27] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II, 2000. URL citeseer.nj.nec.com/deb00fast.html.

[28] Dockery. Health. *ota*, page 59.

[29] R. Dugad and N. Ahuja. Unsupervised multidimensional hierarchical clustering. In *International Conference on Acoustics Speed and Signal Processing*, May 1998.

[30] J. Dunn. A fuzzy relative of the isodata process and its use in detecting compact, well separated clusters. *Cybernetics*, 3(3):32–57, 1974.

[31] M. Duoba. Operational characteristics of the first mass-produced hev - the toyota prius. Presentation-Center for Transportation Research, Energy Systems Division, Argonne National Laboratory,Illinois, May 1999. URL http://www.doe.gov/bridge.

[32] EMPA. Empa. Technical report, Bienne, 1998.

[33] D. Favrat. personal communication, 2002.

[34] D. B. Fogel. *Evolutionary Computation*. IEEE Press, New York, second edition, 2000.

[35] C. M. Fonseca and P. J. Fleming. Genetic algorithms for multiobjective optimization: Formulation, discussion and generalization. In S. Forrest, editor, *Genetic Algorithms: Proceedings of the Fifth International Conference*, pages 416–423, San Mateo, CA, USA, July 1993. Morgan Kaufmann. URL citeseer.nj.nec.com/fonseca93genetic.html.

[36] C. M. Fonseca and P. J. Fleming. An overview of evolutionary algorithms in multiobjective optimisation. *Evolutionary Computation*, 3(1):1–16, 1995. URL http://www.lania.mx/~ccoello/ EMOO/fonseca94.ps.gz.

[37] C. Frangopoulos. Introduction to environomics. In R. G. M. Reistad, M. J. Moran, W. J. Wepfer, and N. Lior, editors, *Symposium on Thermodynamics and the Design, Analysis and Improvement of Energy Systems*, volume 25 of *ASME AES*, pages 49–54, 1991.

[38] C. A. Frangoupoulos and M. R. von Spakovsky. The environomic analysis and optimization of energy systems (part i). In *Proceedings of the International Conference on Energy Systems and Ecology: ENSEC'93*, volume 1 of *ASME*, pages 123–132. ASME, July 1993.

[39] General Motors. EV1 electric vehicle website. http://gmev.com, 2001.

[40] J. S. Gero, editor. *Design Optimization*, pages 193–227. Academic Press, 1985.

[41] A. B. Geva. Hierarchical unsupervised fuzzy clustering. *IEEE Transactions on Fuzzy Systems*, 7(6), 1999.

[42] D. E. Goldberg. *Genetic Algorithms in Search, Optimization, and Machine Learning*. Addison-Wesley, Reading, MA, 1989.

[43] D. E. Goldberg and J. Richardson. Genetic algorithms with sharing for multimodal function optimization. In *Genetic Algorithms and Their Applications: Proc. of the 2nd international Conference on GAs*, pages 41–49, Cambridge MA, July 1987. Lawrence Erlbaum.

[44] T. E. Graedel and B. R. Allenby. *Industrial Ecology*. Englewood Cliffs, NJ: Prentice Hall, 1996.

[45] T. Grüninger. Multimodal optimisation using genetic algorithms. Master's thesis, MIT and Universität Stuttgart, 1996.

[46] J. B. Heywood. *Internal Combustion Engine Fundamentals*. McGraw Hill, Singapore, 1988.

[47] J. Holland. 1975.

[48] J. Horn, N. Nafpliotis, and D. E. Goldberg. A Niched Pareto Genetic Algorithm for Multiobjective Optimization. In *Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence*, volume 1, pages 82–87, Piscataway, New Jersey, 1994. IEEE Service Center. URL citeseer.nj.nec.com/horn94niched.html.

[49] H. Ishitani, Y. Baba, A. Molyneaux, and D. Favrat. An optimization tool of hybrid type EV systems. In *The 15th International Electric Vehicle Symposium and Exhibition (EVS-15)*, Brussels, Sept. 1998.

[50] J. E. Jackson. *A User's Guide to Principal Components*. John Wiley and Sons, 1991.

[51] V. H. Johnson, K. B. Wipke, and D. J. Rausen. Hev control strategy for real-time optimization of fuel economy and emissions. *NREL, Society of Automotive Engineers, Inc.*, 2000.

[52] K. V. R. Kanth, D. Agrawal, A. E. Abbadi, and A. Singh. Dimensionality reduction for similarity searching in dynamic databases. *Computer Vision and Image Understanding: CVIU*, 75(1–2):59–72, / 1999. URL citeseer.nj.nec.com/kanth98dimensionality.html.

[53] L. Kaufman and P. J. Rousseeuw. *Finding Groups in Data: an Introduction to Cluster Analysis*. Wiley and Sons., 1990.

[54] K. Kelly. Test results and modeling of the honda insight and toyota prius. Presentation, see URL, 2001. URL http://www.ctts.nrel.gov/analysis/reading_room.html.

[55] J. Knowles and D. Corne. Local search , multiobjective optimization and the pareto archived evolution strategy. Unknown, 1999. URL http://www.reading.ac.uk/˜ssr97jdk.

[56] J. Knowles and D. Corne. The pareto archived evolution strategy : A new baseline algorithm for parteo multiobjective optimisation. Unknown, 1999. URL http://www.reading.ac.uk/˜ssr97jdk.

[57] J. R. Koza. *Genetic Programming: On the Programming of Computers by Means of Natural Selection*. MIT Press, 1992.

[58] V. Kumar. An introduction to cluster analysis for data mining. Technical report, University of Minnesota, Dept. of Computer Scirnce and Engineering, 2000. URL http://www.cs.umn.edu/˜mjoshi/csci5980/index.html.

[59] G. N. Lance and W. T. Williams. A general theory of classificatory sorting strategies 1. Hierarchical systems. *The Computer Journal*, 9(4):373–380, 1967.

[60] G. Leyland and A. K. Molyneaux. Evolving combination and mutation operator choice in multi-objective evolutionary algorithms. In preparation, 2001.

[61] G. Leyland, A. K. Molyneaux, and D. Favrat. A fast multi-objective evolutionary algorithm applied to industrial problems. In *EUROGEN*, Athens, Sept. 2001.

[62] G. Leyland, A. K. Molyneaux, and D. Favrat. A new multi-objective optimisation technique applied to a vehicle drive train simulation. In *ECOS—Efficiency, Costs, Optimization, Simulations and Environmental Impact of Energy Systems*, Istanbul, Turkey, July 2001.

[63] G. B. Leyland. *Multi-Objective Optimisation Applied to Industrial Energy Problems*. PhD thesis, École Polytechnique Fédérale de Lausanne, Lausanne, Switzerland, 2002.

[64] S. Mahfoud. *Niching Methods for Genetic Algorithms*. PhD thesis, University of Illinois, 1995.

[65] A. K. Molyneaux. A simple parallel MATLAB implementation using PVM. Technical Report 0101i, École Polytechnique Fédérale de Lausanne, Jan. 2001.

[66] A. K. Molyneaux, G. Leyland, and D. Favrat. A new, clustering evolutionary multi-objective optimisation technique. In *International Symposium on Adaptative Systems*, pages 41–47, Havana, Cuba, Mar. 2001.

[67] H. Muehlenbein and T. Mahnig. Mathematical analysis of evolutionary algorithms for optimization. In *International Symposium on Adaptative Systems*, Evolutionary Computation and Probabilistic Graphical Models, Havana, Cuba, March 2001.

[68] B. Olsommer. *Méthode d'Optimisation Thermoéconomique Appliqué aux Centrales d'Incinération d'Ordures à Cogénération avec Appoint Énérgétique*. PhD thesis, École Polytechnique Fédérale de Lausanne, 1998.

[69] B. Olsommer, D. Favrat, and D. Comte. Time-dependent thermoeconomic optimization of the future waste incineration power plant in posieux, switzerland. In *5th World Congress on Integrated Resources Management*, Toronto, Canada, June 2000.

[70] B. Olsommer, D. Favrat, and M. von Spakovsky. An approach for the time-dependent thermo-economic modeling and optimization of energy system synthesis, design and operation (part I: Methodology and results). *International Journal of Applied Thermodynamics*, 2(3):97–114, Sept. 1999.

[71] B. Olsommer, D. Favrat, and M. von Spakovsky. An approach for the time-dependent thermo-economic modeling and optimization of energy system synthesis, design and operation (part II : Reliability and availability). *International Journal of Applied Thermodynamics*, 2(4):177–186, Dec. 1999.

[72] OTA. Green products by design: Choices for a cleaner environment. ota-e-541, washington, dc: Us government printing office. Congress of the United States, Office of Technology Assessment, 1992.

[73] OTA-ETI. Advanced automotive technology:visions of a super-efficient family car, gpo stock 052-003-01440-8. Technical report, OTA-ETI-639, 1995.

[74] F. Pahng, N. Senin, and D. R. Wallace. Distributed object-based modeling in design simulation marketplace. *accepted for publication ASME Journal of Mechanical Design*, 2000.

[75] V. Pareto. *Cours D'Economie Politique*, volume I and II. F. Rouge, Lausanne, 1896.

[76] S. Pelster. *Environimic Modelling and Optimization of Advanced Combined Cycle Cogeneration Plants Including CO_2 Separation Options*. PhD thesis, École Polytechnique Fédérale de Lausanne, 1998.

[77] S. Pelster, D. Favrat, and M. von Spakovsky. Optimisation thermo-économique et environomique de nouvelles centrales de production d'électricité à cycles combinés. *GWA Gas Wasser und Abwasser*, Aug. 2000.

[78] S. Plotkin, D. Santini, A. Vyas, J. Anderson, M. Wang, J. He, and D. Bharathan. Hybrid electric vehicle technology assessment: Methodology, analytical issues, and interim results. Technical report, Center for Transportation Research, Energy Systems Division, Argonne National Laboratory,Illinois, 2001. URL http://www.doe.gov/bridge.

[79] PVM, 2002. URL http://www.netlib.org/pvm3/.

[80] S. Ray and R. H. Turi. Determination of number of clusters in k-means clustering and application in colour image segmentation.

[81] M. Raymer, W. Punch, E. Goodman, L. Kuhn, and A. Jain. Dimensionality reduction using genetic algorithms. *IEEE Transactions on Evolutionary Computation*, 4(2), July 2000.

[82] D. Schaffer. Multiple objective optimization with vector evaluated genetic algorithms. In *Genetic Algorithms and their Applications: Proceedings of the First International Conference on Genetic Algorithms*, pages 93–100, 1985.

[83] N. Senin, D. Wallace, and N. Borland. Object-based design modeling and optimization with genetic algorithms. In *GECCO-99: Proceedings of the Genetic and Evolutionary Computation Conference*, Orlando, FL, USA, July 1999.

[84] N. Srinivas and K. Deb. Multiobjective optimization using nondominated sorting in genetic algorithms. *Evolutionary Computation*, 2(3):221–248, 1995. URL citeseer.nj.nec.com/srinivas94multiobjective.html.

[85] T. C. Suisse. Touring 18/01, 2001.

[86] T. C. Suisse. No. 2833, 2001.

[87] T. C. Suisse. Touring 13/01, 2001.

[88] TCS. *Consommation de carburant*, page 32. Touring Club Suisse, 14 edition, 2001.

[89] TCS. *Cout des services 2002*, page 32. Touring Club Suisse, 21 edition, 2002.

[90] The Mathworks. *MATLAB Reference Manual*. The Mathworks, 1998.

[91] D. Thierens and P. Bosman. Multi-objective optimization with iterated density estimation evolu-
tionary algorithms using mixture models. In *International Symposium on Adaptative Systems*, pages
129–135, Havana, Cuba, Mar. 2001.

[92] Toyota. Prius hybrid system. CD-ROM, Jan. 2001.

[93] T. Trach-Minh and R. Gruber. Grappe de pc linux pour le calcul a haute performance. *Flash Infor-
matique*, 02(6), July 2002.

[94] M. Tribus. Thermodynamic and economic consideration in the preparation of fresh water from sea
water. Technical Report 56-16, Dept of Engineering,UCLA, 1956.

[95] M. Tribus and R. B. Evans. Thermoeconomics. Technical Report 62-63, Dept of Engineering,UCLA,
1962.

[96] USCAR, 1993. URL http://www.uscar.org/pngv/.

[97] D. A. Van Veldhuizen and G. B. Lamont. Multiobjective Evolutionary Algorithm Test Suites. In
J. Carroll, H. Haddad, D. Oppenheim, B. Bryant, and G. B. Lamont, editors, *Proceedings of the 1999
ACM Symposium on Applied Computing*, pages 351–357, San Antonio, Texas, 1999. ACM. URL
citeseer.nj.nec.com/david99multiobjective.html.

[98] D. Wallace. Personal communication, August 2002.

[99] D. Wallace, D. Favrat, T. Tomiyama, and H. Ishitani. A framework for holistic life-cycle design: the
integration of performance, economic, manufacturing and environmental measures. Technical report,
Massachusetts Institute of Technology, 1999.

[100] K. Wipke, M. Cuddy, and S. Burch. ADVISOR 2.1: A user-friendly advanced powertrain simulation
using a combined backward/forward approach. *IEEE Transactions on Vehicular Technology*, 1999.

[101] X. Xie and G. Beni. A validity measure for fuzzy clustering, 1991.

[102] J. Yamaguchi. Global viewpoints: More on insight. *Automotive Engineering International*, Jan. 2000.

[103] E. Zitzler. *Evolutionary Algorithms for Multiobjective Optimisation: Methods and Applications*. PhD
thesis, Eidgenöissiche Technische Hochschule Zürich, Nov. 1999.

[104] E. Zitzler, K. Deb, and L. Thiele. Comparison of multiobjective evolutionary algorithms: Empirical results. *Evolutionary Computation*, 8(2):173–195, 2000. URL `http://www.tik.ee.ethz.ch/~zitzler/`.

Appendix A

Appendix

A.1 District Heating Problem

A.1.1 Total Cost and Pollution Cost with Multi-Dimensional Scaling
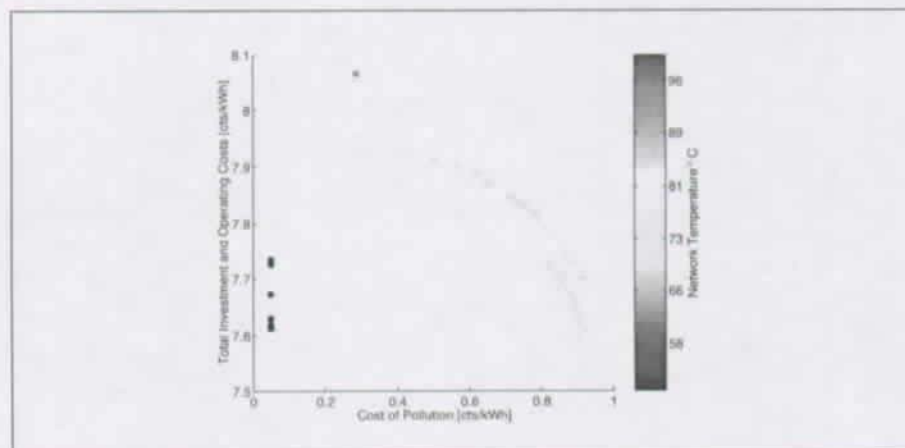
Fig. A.1: Pareto Surface showing Cost vs Cost of Pollution with clustering on reduced dimensions
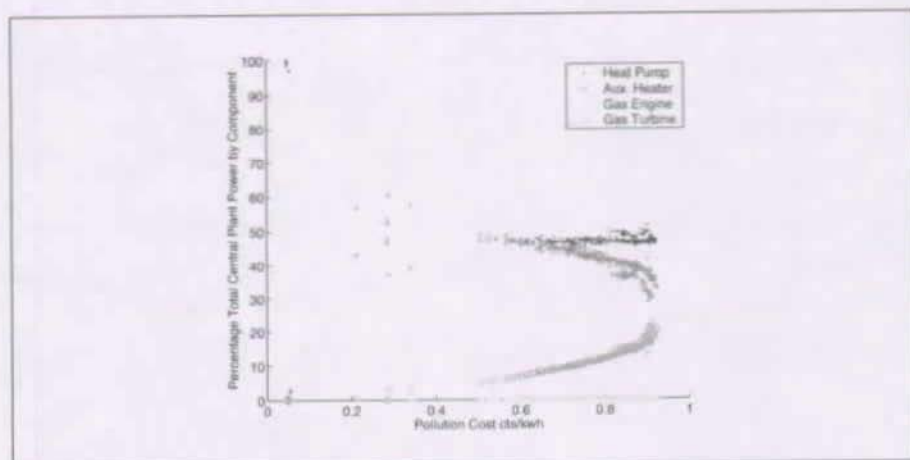
A.1.2 Cost and Pollution with European Electricity Mix

Fig. A.2: Multi-objective optimisation of Total Cost and Cost of pollution. Central Plant composition as a percentage of total heat produced vs pollution cost.
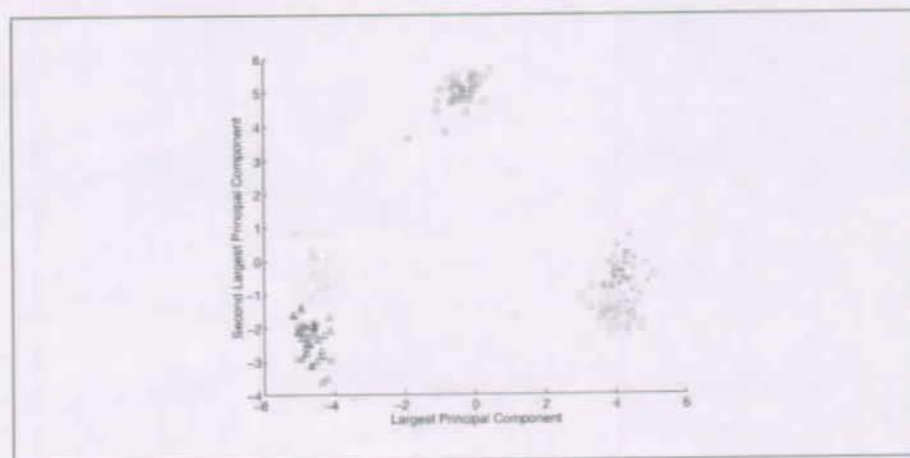
Fig. A.3: Total Cost and pollution costs. Clusters shown reduced principal component space
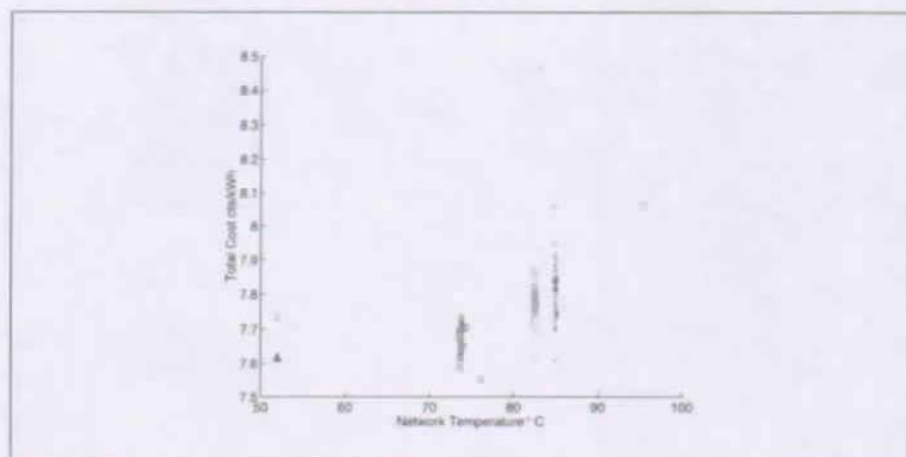
Fig. A.4: Total Cost and pollution costs. Clusters shown in Temperature and total cost space

A.2   Vehicle Drivetrain

A.2.1   IC Engine Cartographies
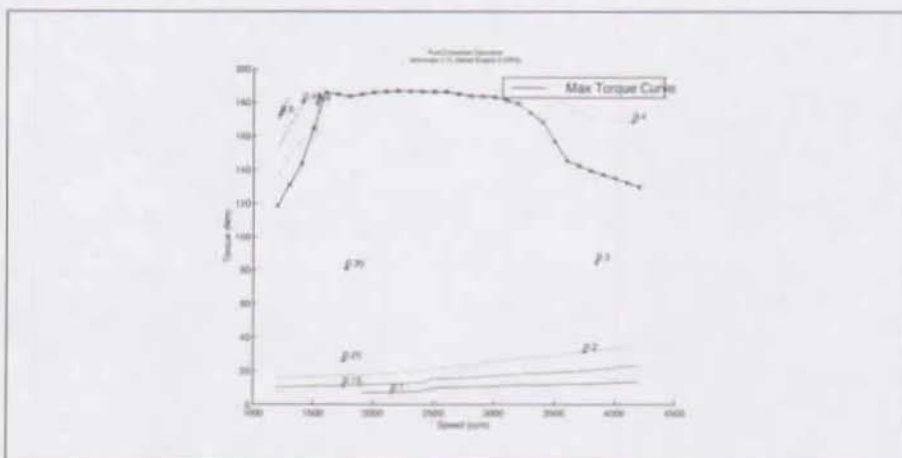
Fig. A.5: ICE Efficiency versus Speed and Load for 60kW CI engine
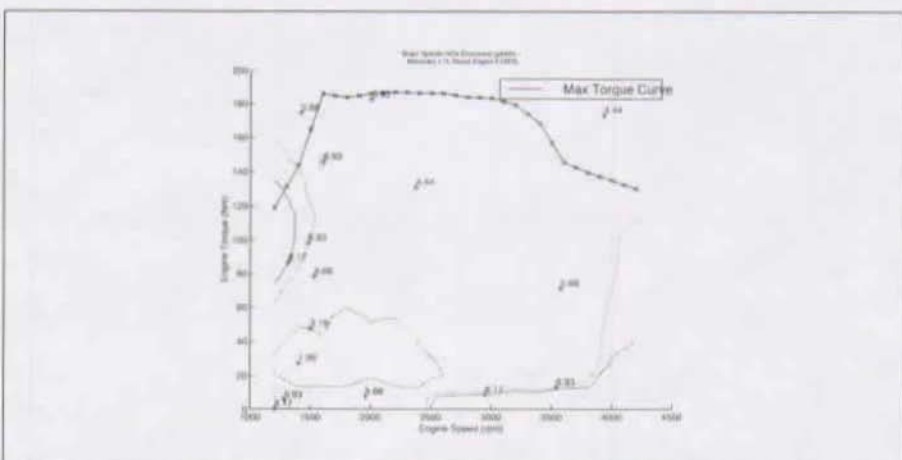
Fig. A.6: ICE NO_x versus Speed and Load for 60kW CI engine
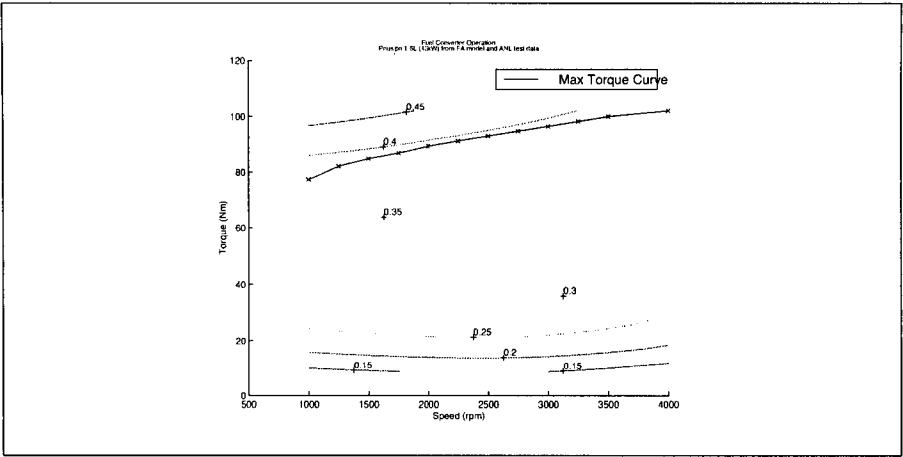
Fig. A.7: ICE Efficiency versus Speed and Load for Prius engine
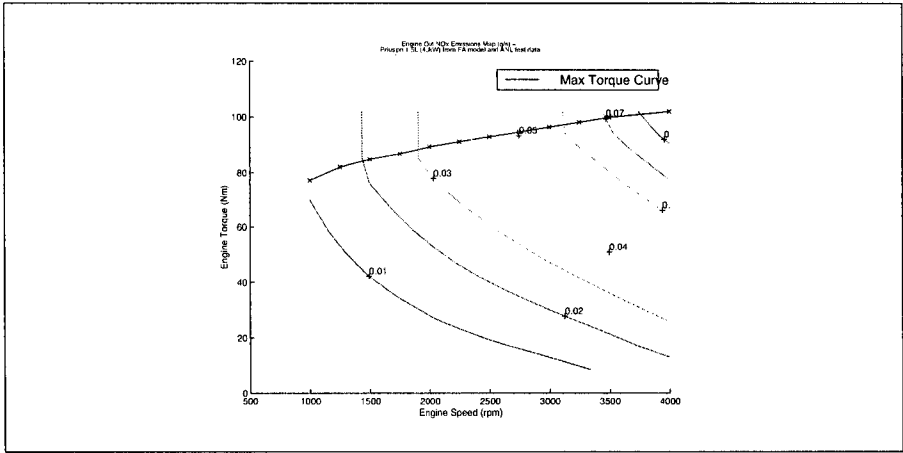
Fig. A.8: ICE NO_x versus Speed and Load for Prius engine
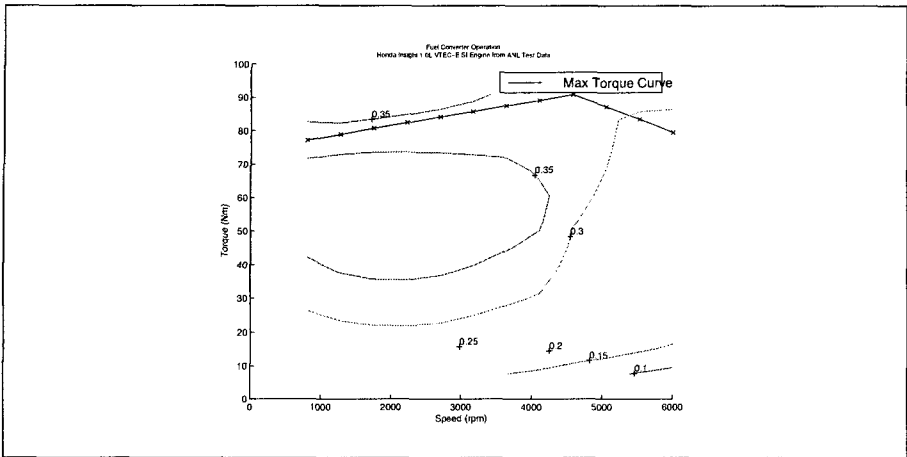
Fig. A.9: ICE Efficiency versus Speed and Load for Insight engine

A.2.2 Fuel Economy vs NO_x

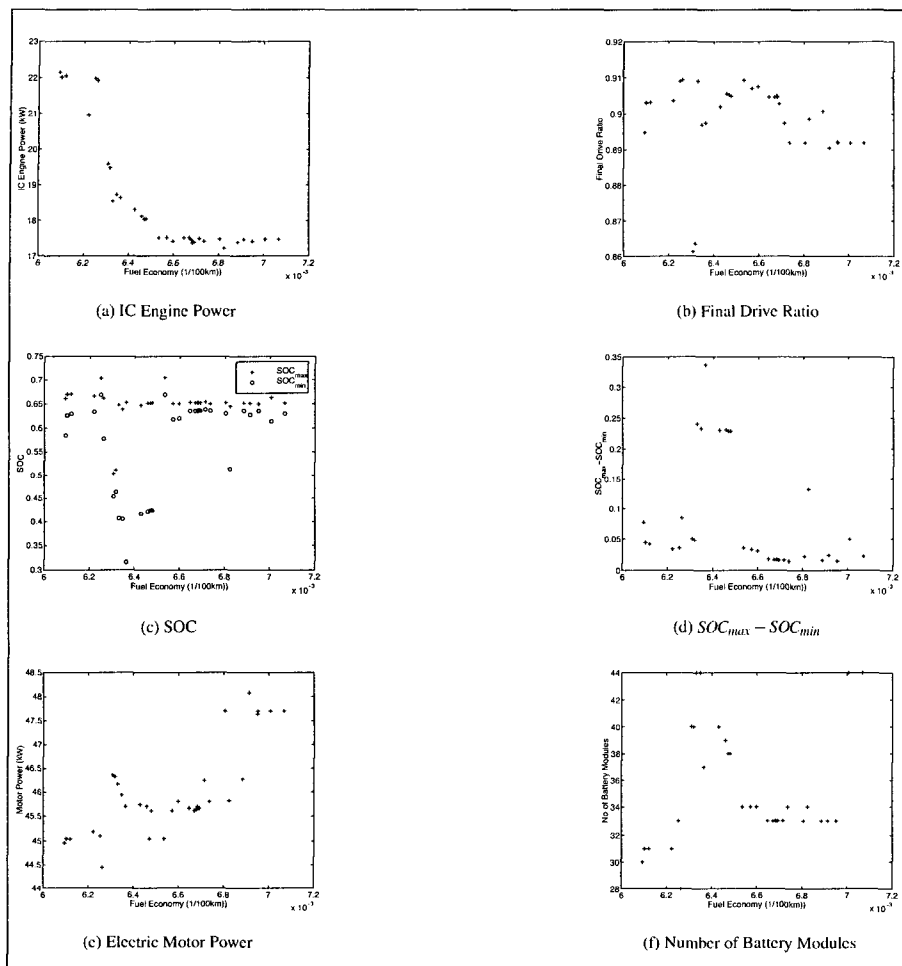Full results for the fuel economy vs NO_x optimisation
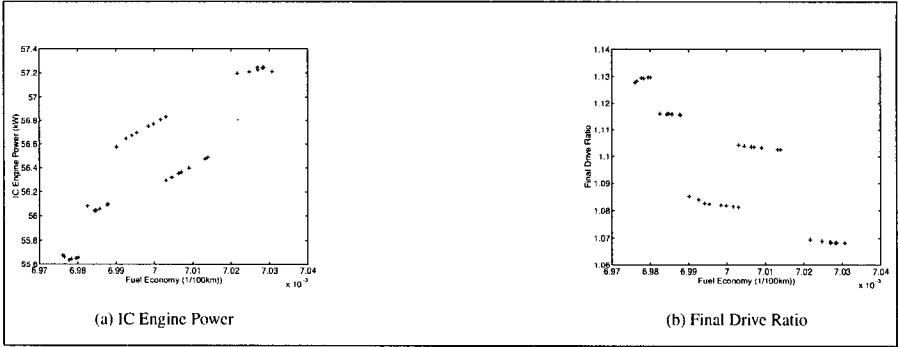
Fig. A.10: Independent variable values for the conventional SI vehicle from the optimisation of economy vs nox.

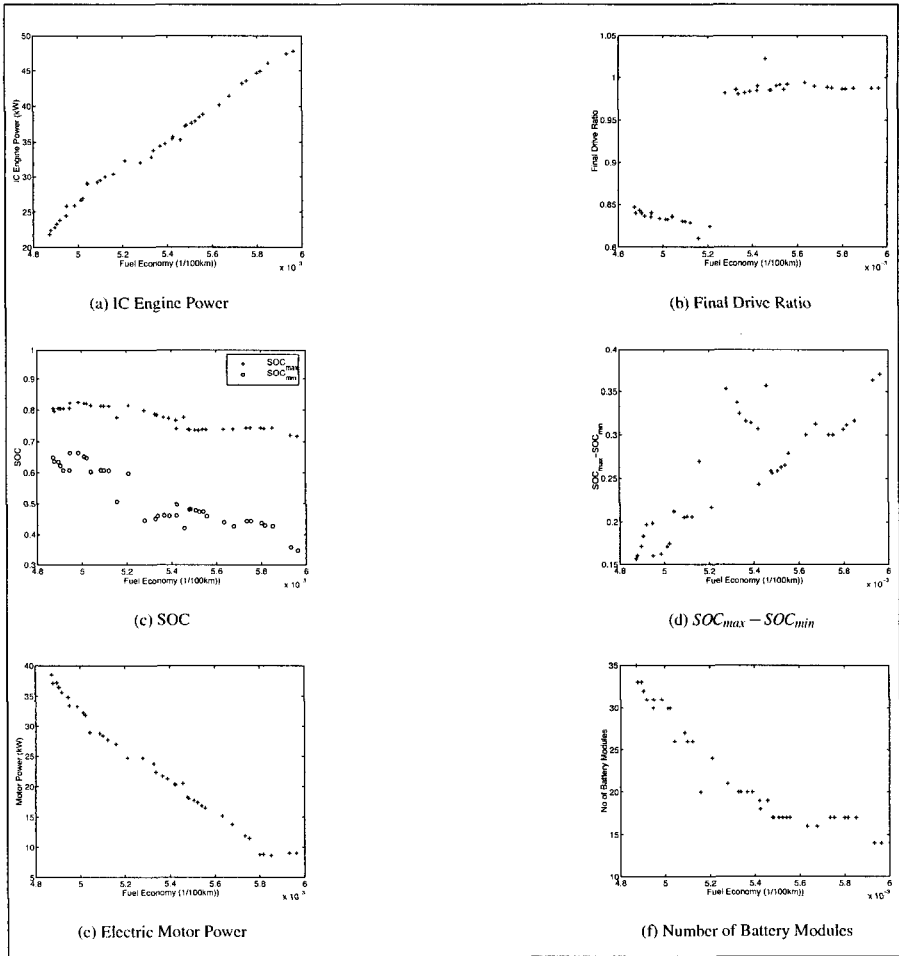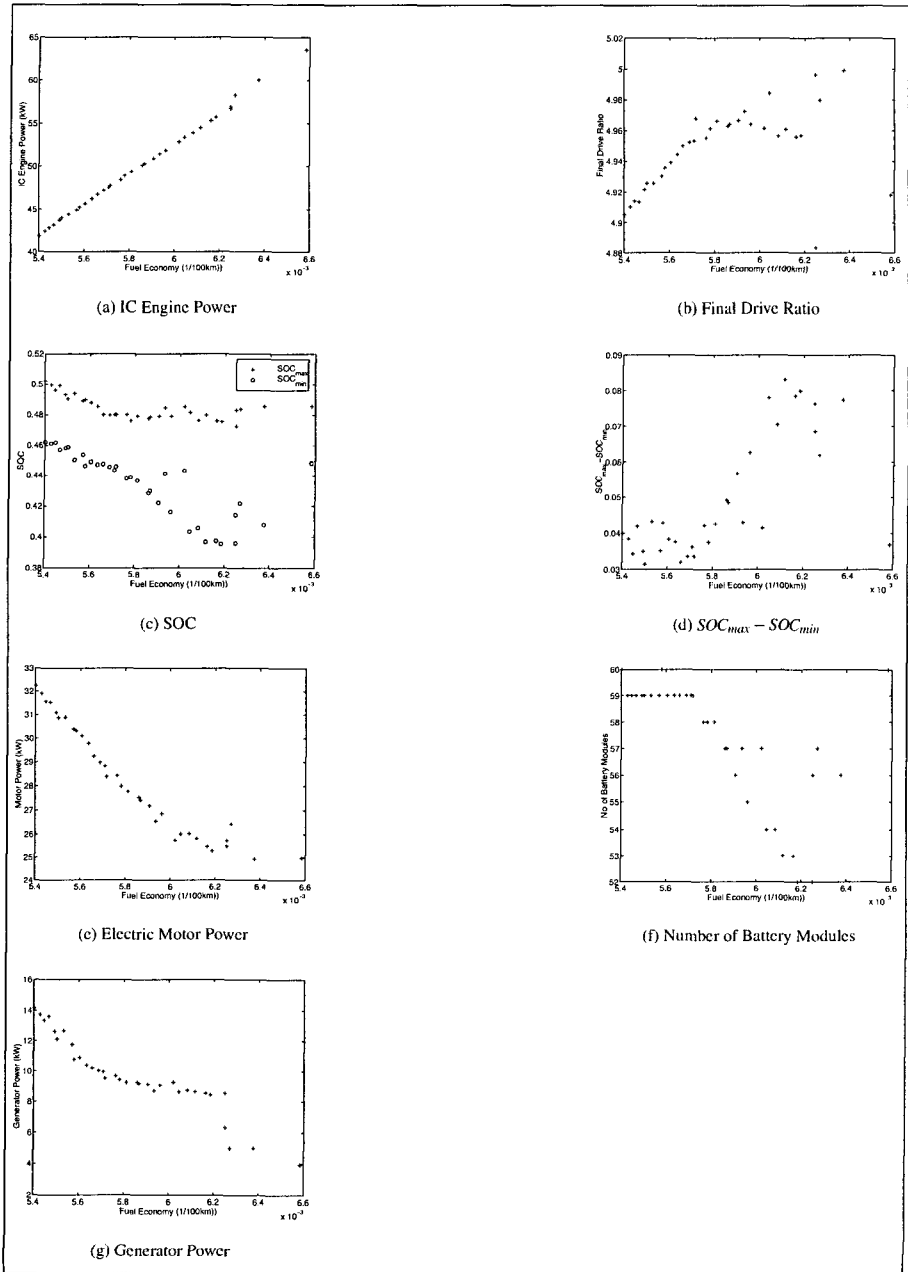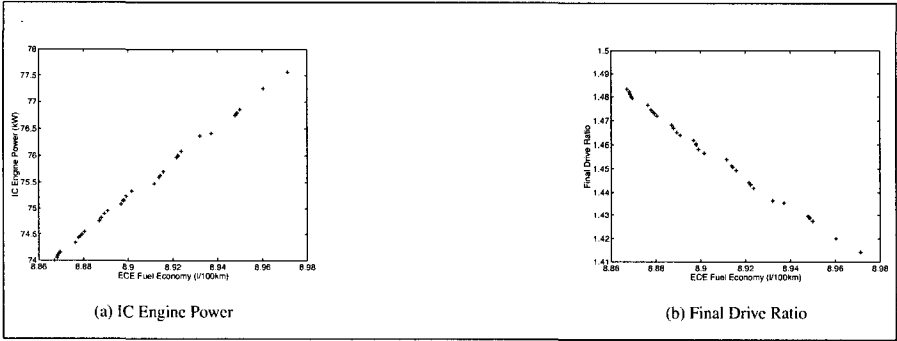Fig. A.11: Independent variable values for the conventional CI vehicle from the optimisation of economy vs nox.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.12: Independent variable values for the SI series hybrid vehicle from the optimisation of economy vs nox.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.13: Independent variable values for the CI series hybrid vehicle from the optimisation of economy vs nox.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

Fig. A.14: Independent variable values for the parallel hybrid SI vehicle from the optimisation of economy vs nox.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

Fig. A.15: Independent variable values for the parallel hybrid CI vehicle from the optimisation of economy vs nox.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power
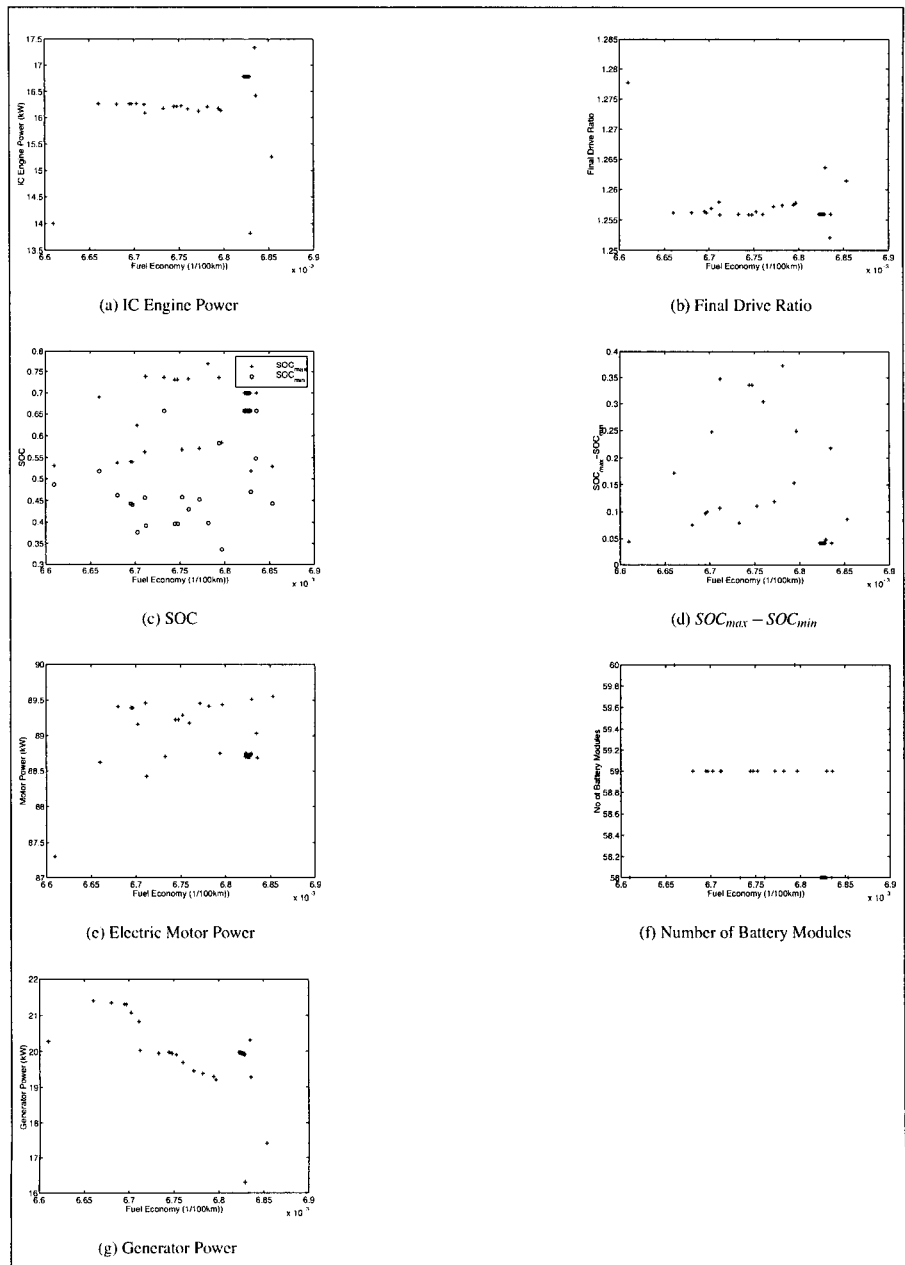
Fig. A.16: Independent variable values for the modified Prius from the optimisation of investment vs operating cost.

A.2.3 Multi Cycle

ECE-EUDC vs US06HWY

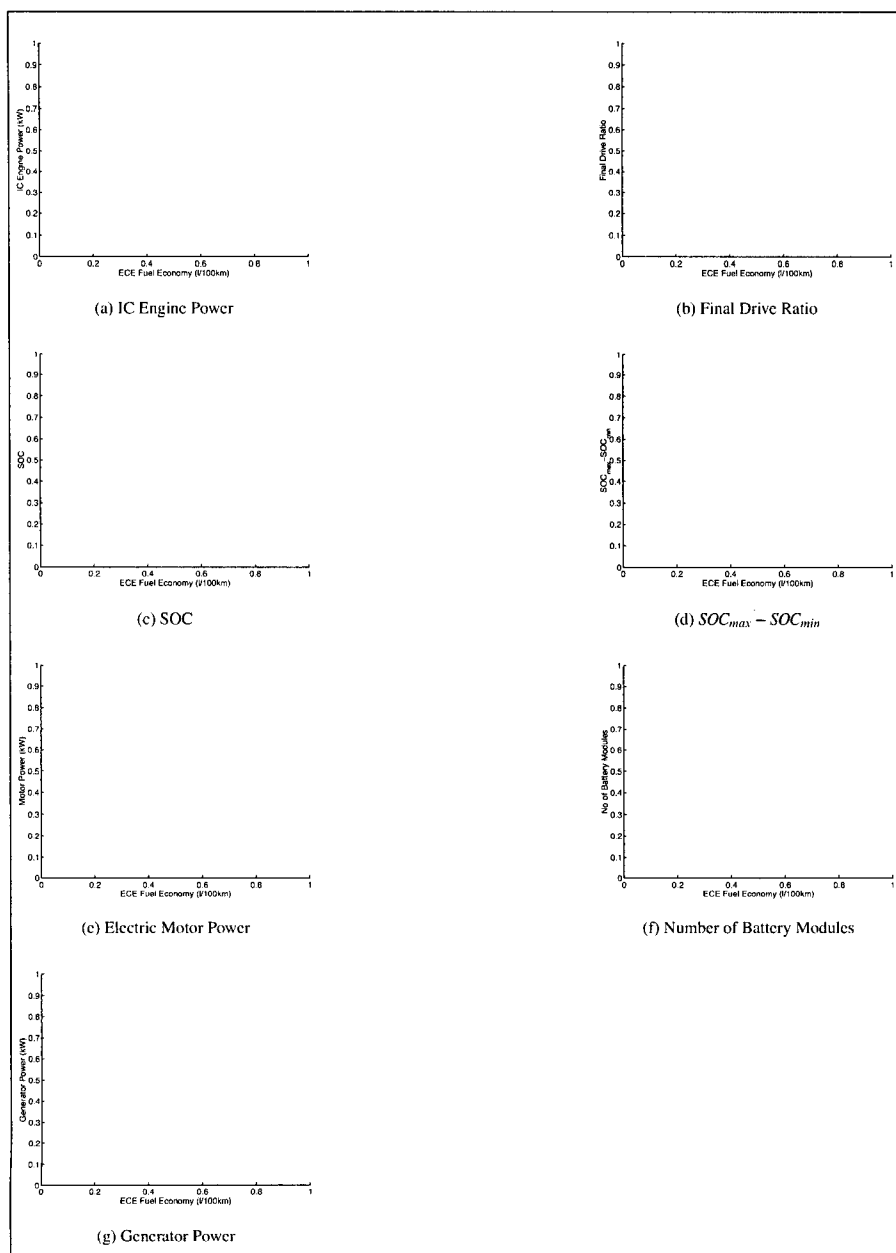Full results for the ECE-EUDC vs US06HWY fuel economy optimisation.

Fig. A.17: Independent variable values for the conventional SI vehicle from the optimisation of fuel economy over US06HWY vs ECE-EUDC cycles.

Fig. A.18: Independent variable values for the conventional CI vehicle from the optimisation of fuel economy over US06HWY vs ECE-EUDC cycles.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.19: Independent variable values for the SI series hybrid vehicle from the optimisation of fuel economy over US06HWY vs ECE-EUDC cycles.
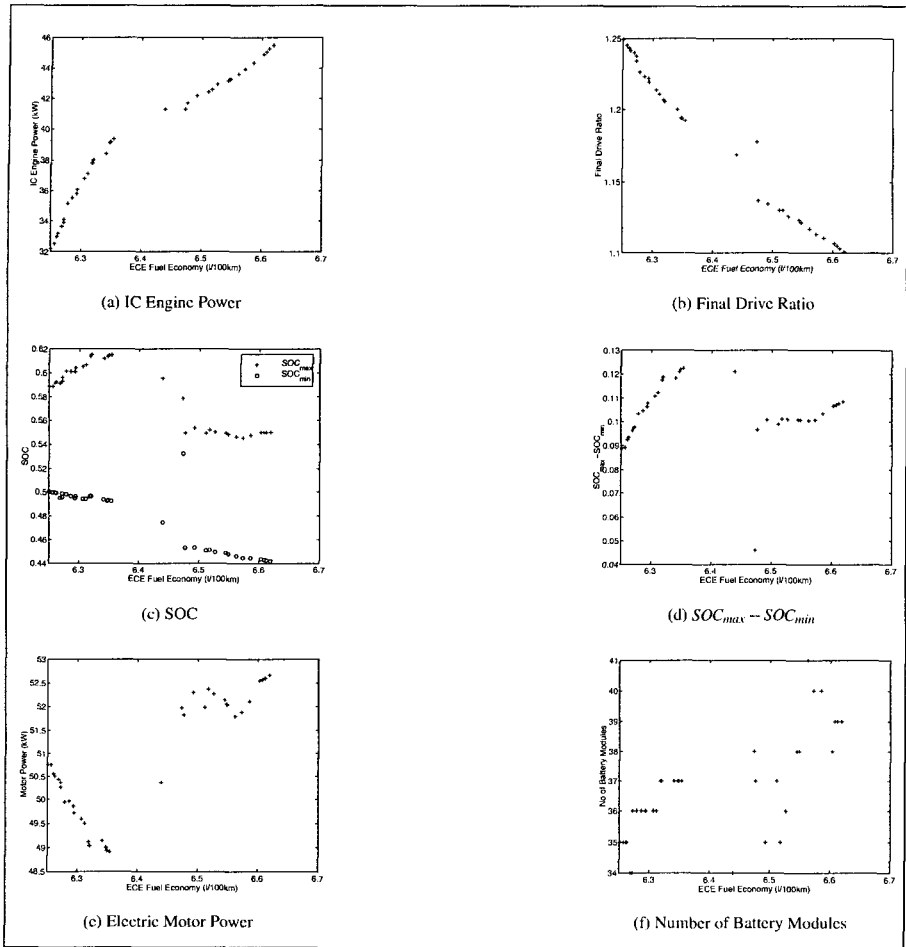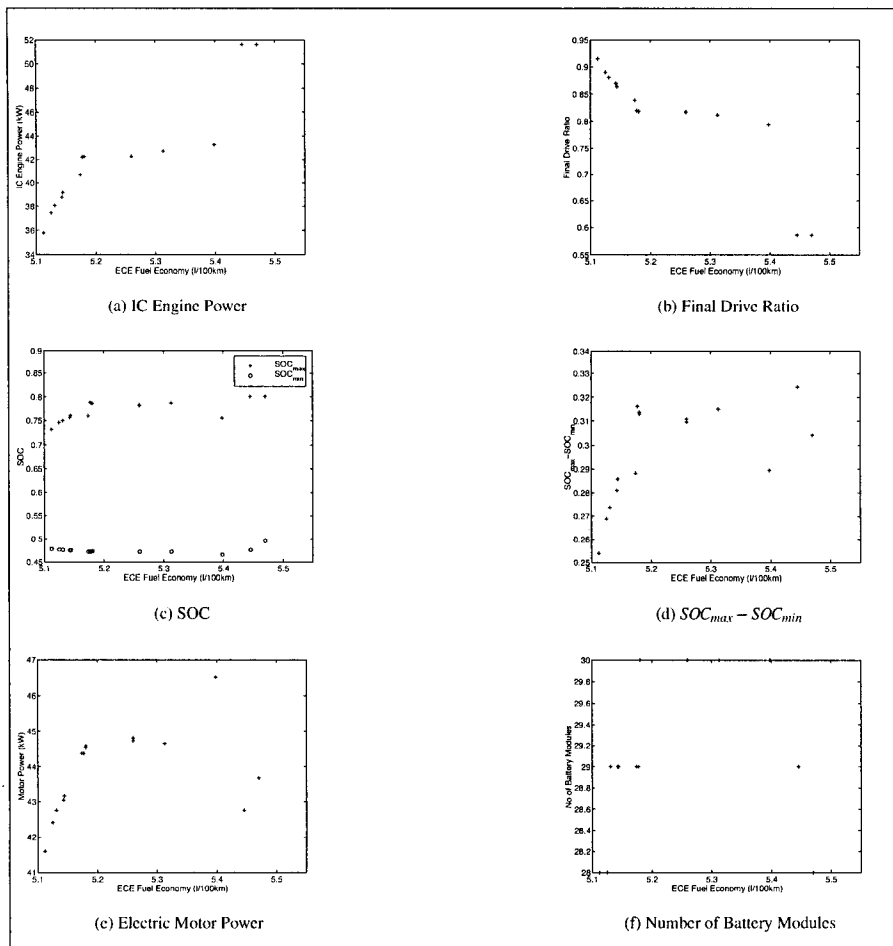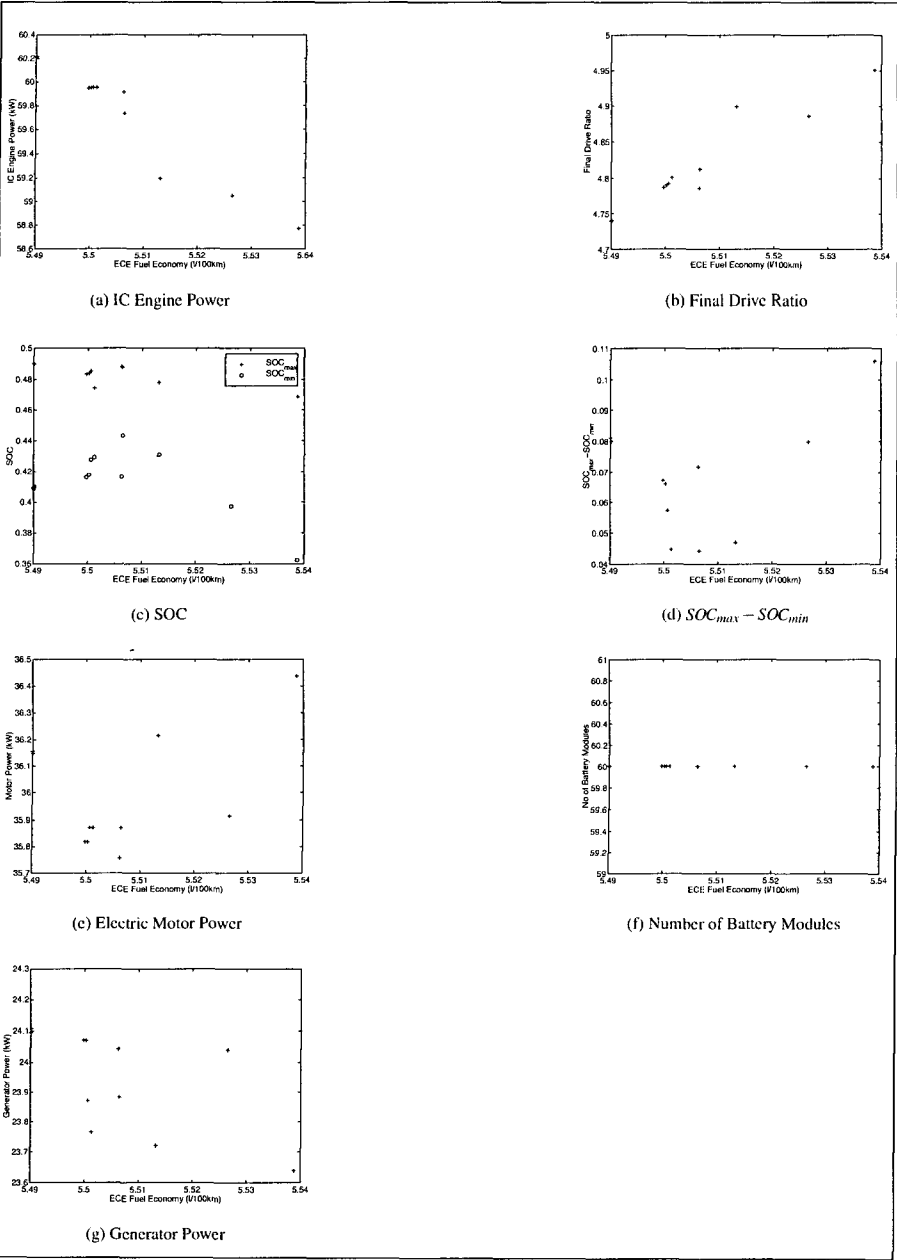
(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.20: Independent variable values for the CI series hybrid vehicle from the optimisation of fuel economy over US06HWY vs ECE-EUDC cycles.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power
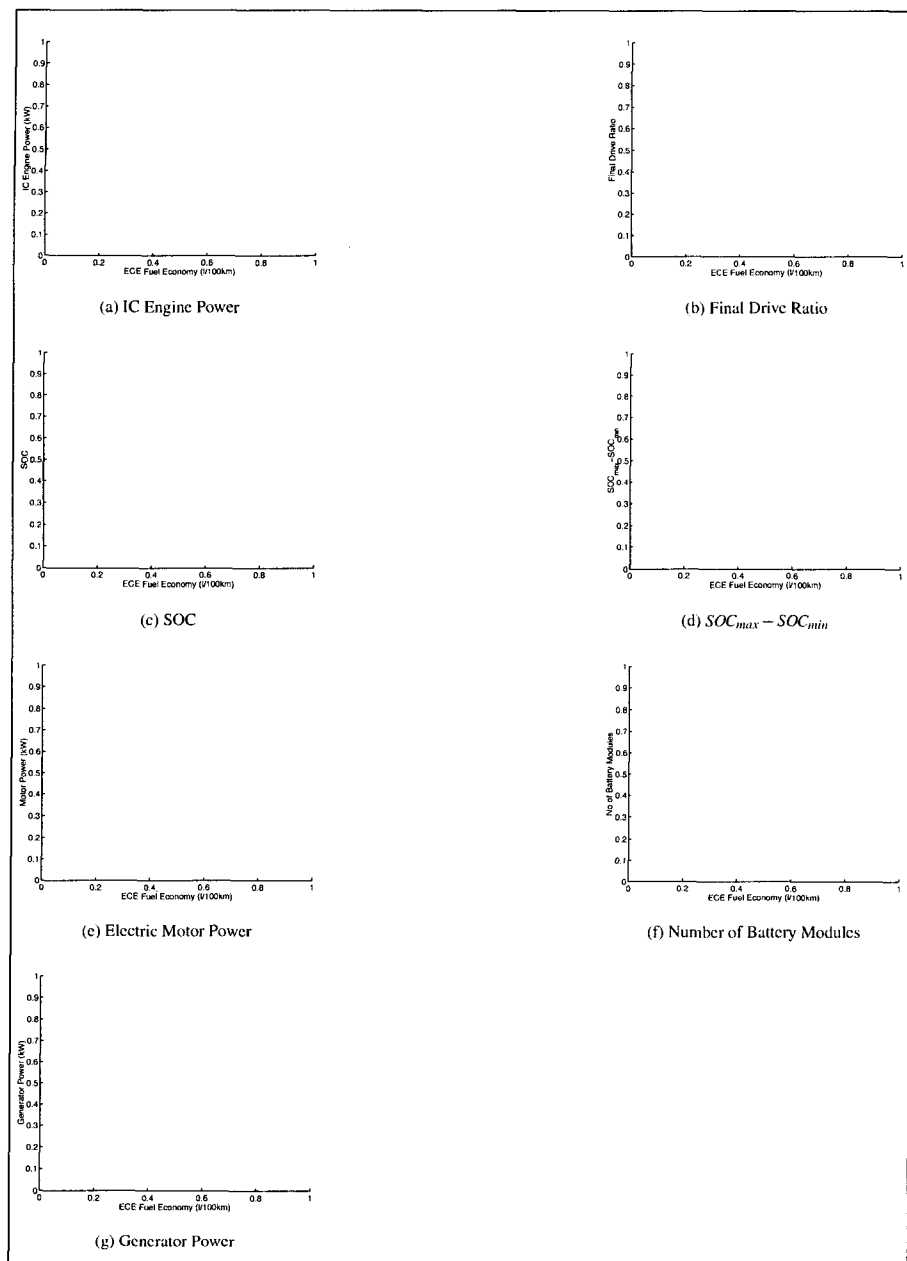
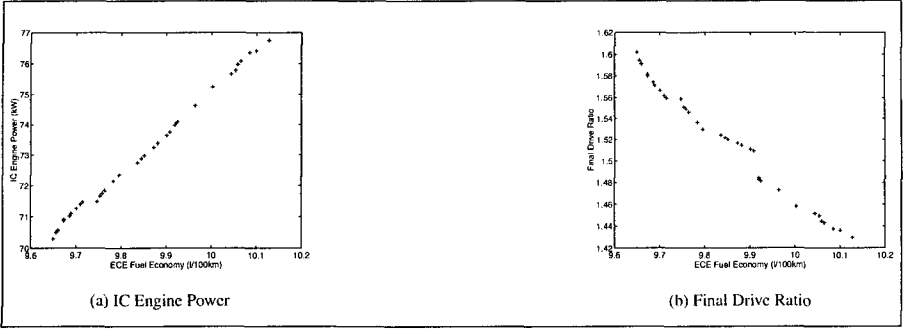(f) Number of Battery Modules

Fig. A.21: Independent variable values for the parallel hybrid SI vehicle from the optimisation of fuel economy over US06HWY vs ECE-EUDC cycles.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$
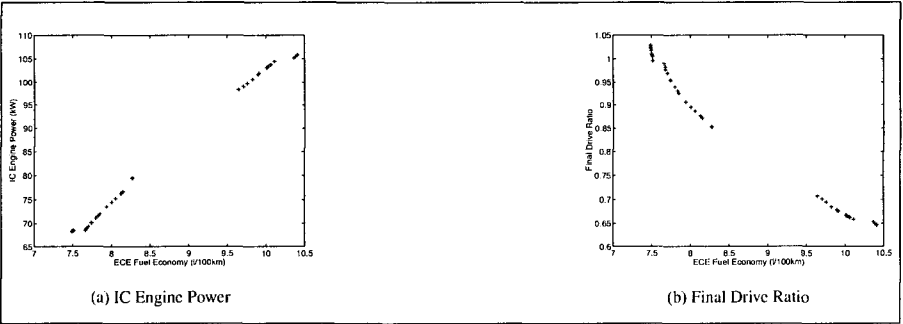
(e) Electric Motor Power

(f) Number of Battery Modules

Fig. A.22: Independent variable values for the parallel hybrid CI vehicle from the optimisation of fuel economy over US06HWY vs ECE-EUDC cycles.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.23: Independent variable values for the modified Prius from the optimisation of fuel economy over US06HWY vs ECE-EUDC cycles.
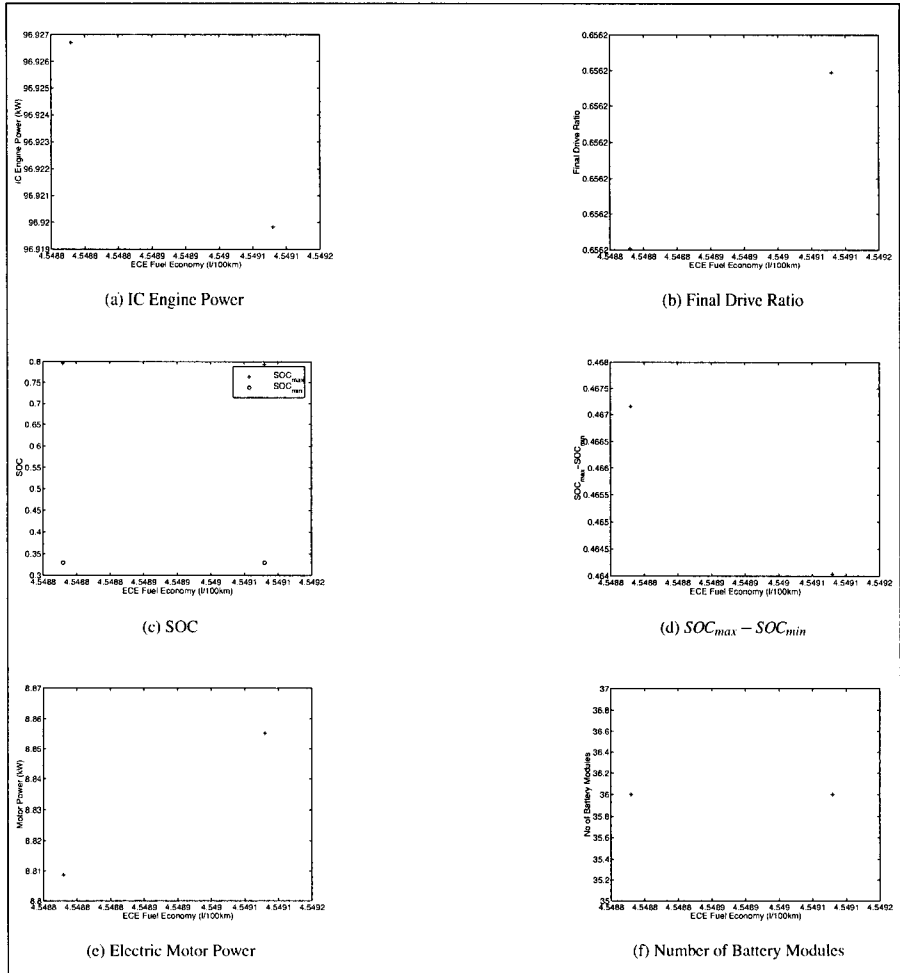
Fig. A.24: Independent variable values for the modified Insight from the optimisation of fuel economy over US06HWY vs ECE-EUDC cycles.

ECE vs US06HWY

Full results for the ECE vs US06HWY fuel economy optimisation.

(a) IC Engine Power   (b) Final Drive Ratio

Fig. A.25: Independent variable values for the conventional SI vehicle from the optimisation of fuel economy over US06HWY vs ECE cycles.
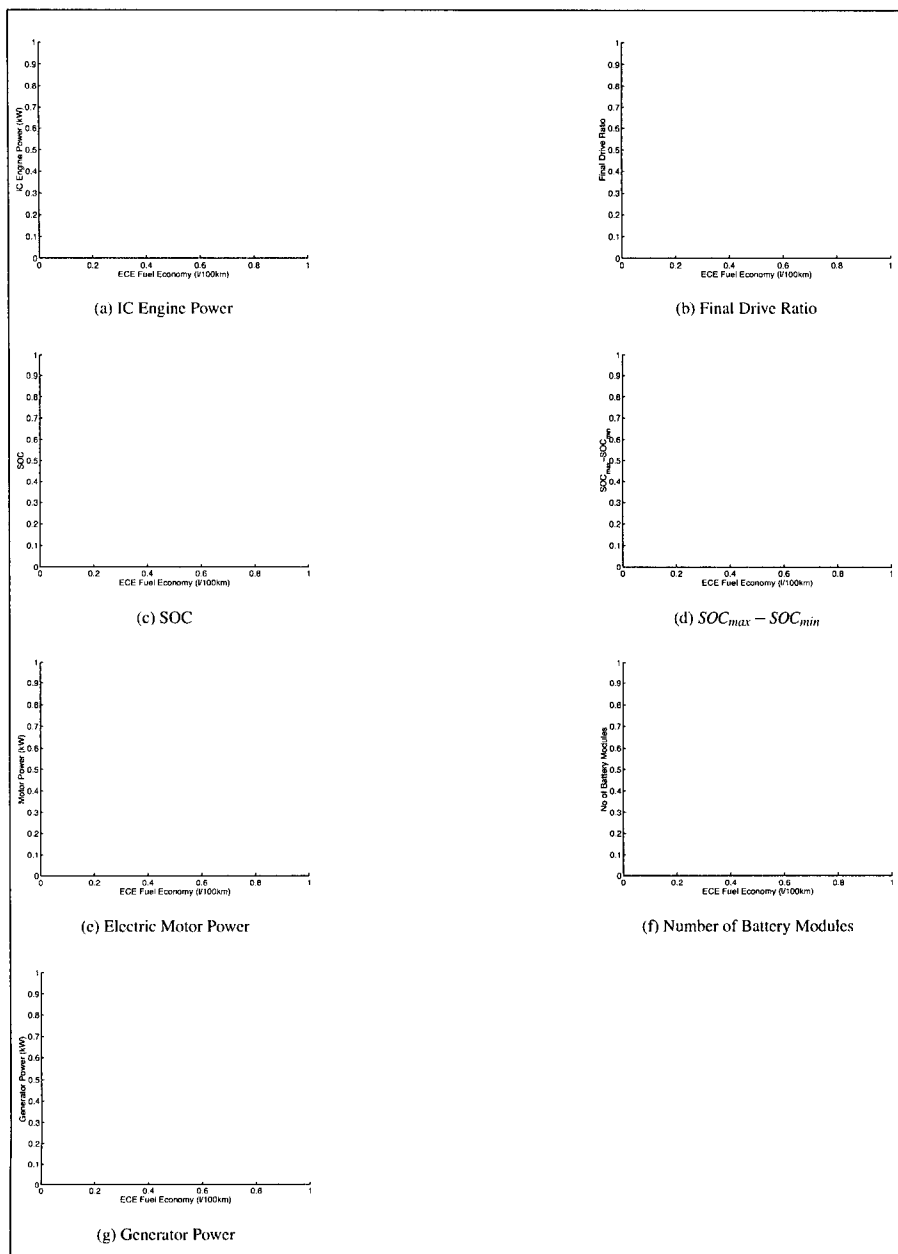

(a) IC Engine Power   (b) Final Drive Ratio

Fig. A.26: Independent variable values for the conventional CI vehicle from the optimisation of fuel economy over US06HWY vs ECE cycles.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.27: Independent variable values for the SI series hybrid vehicle from the optimisation of fuel economy over US06HWY vs ECE cycles.

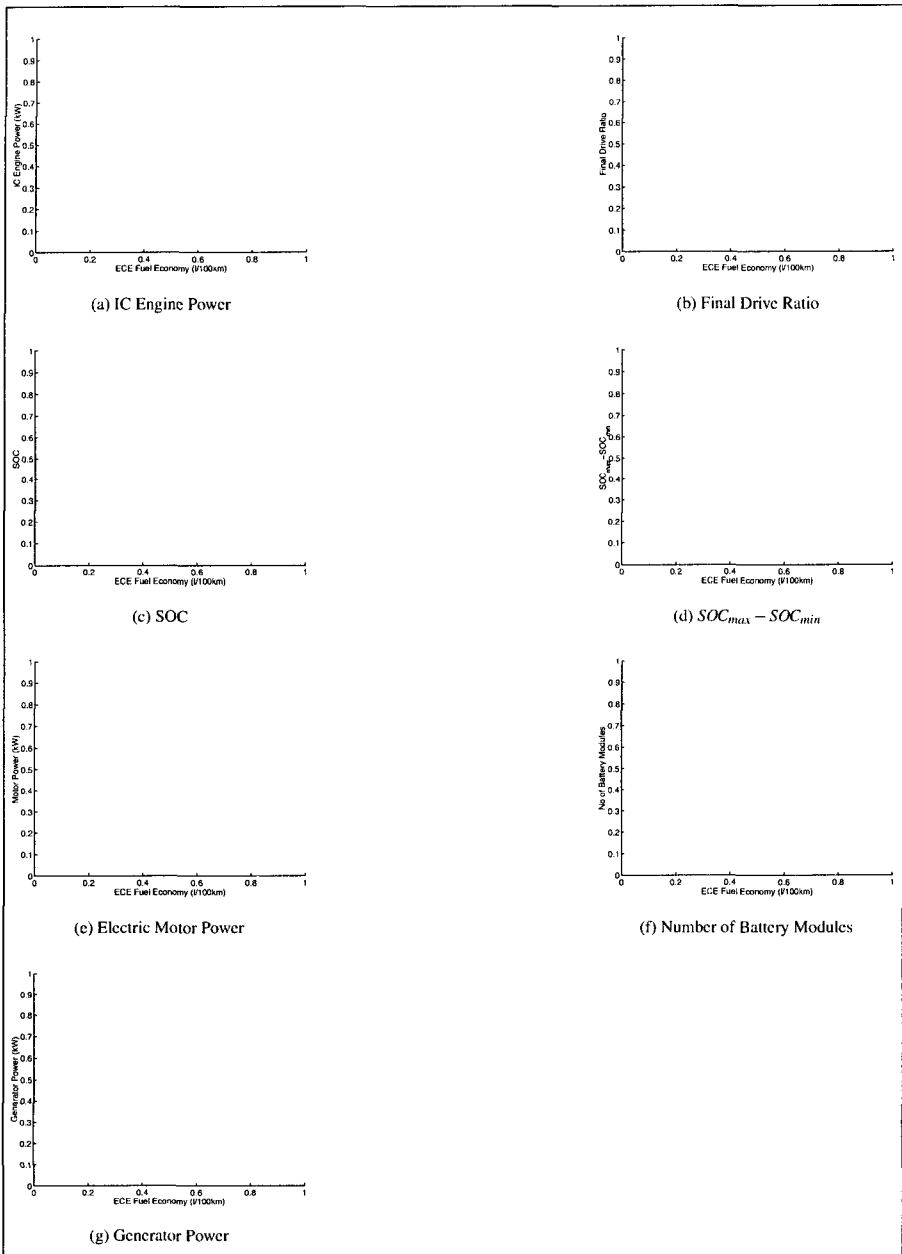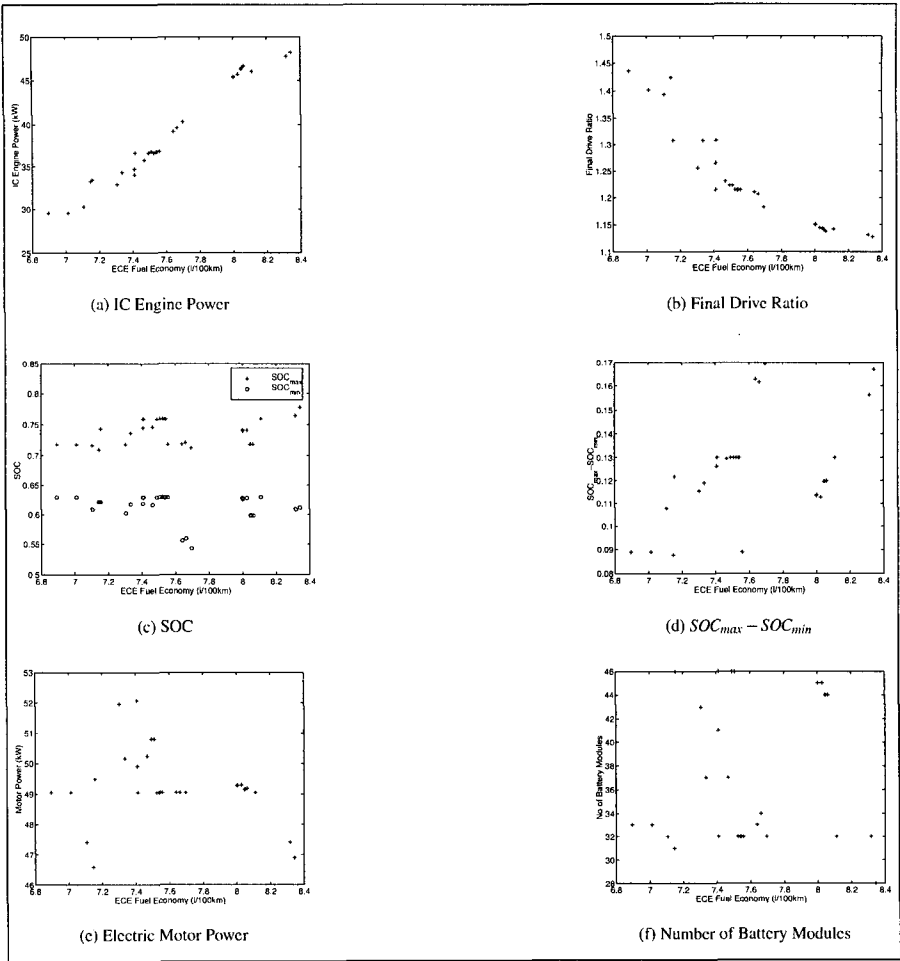(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.28: Independent variable values for the CI series hybrid vehicle from the optimisation of fuel economy over US06HWY vs ECE cycles.

Fig. A.29: Independent variable values for the parallel hybrid SI vehicle from the optimisation of fuel economy over US06HWY vs ECE cycles.
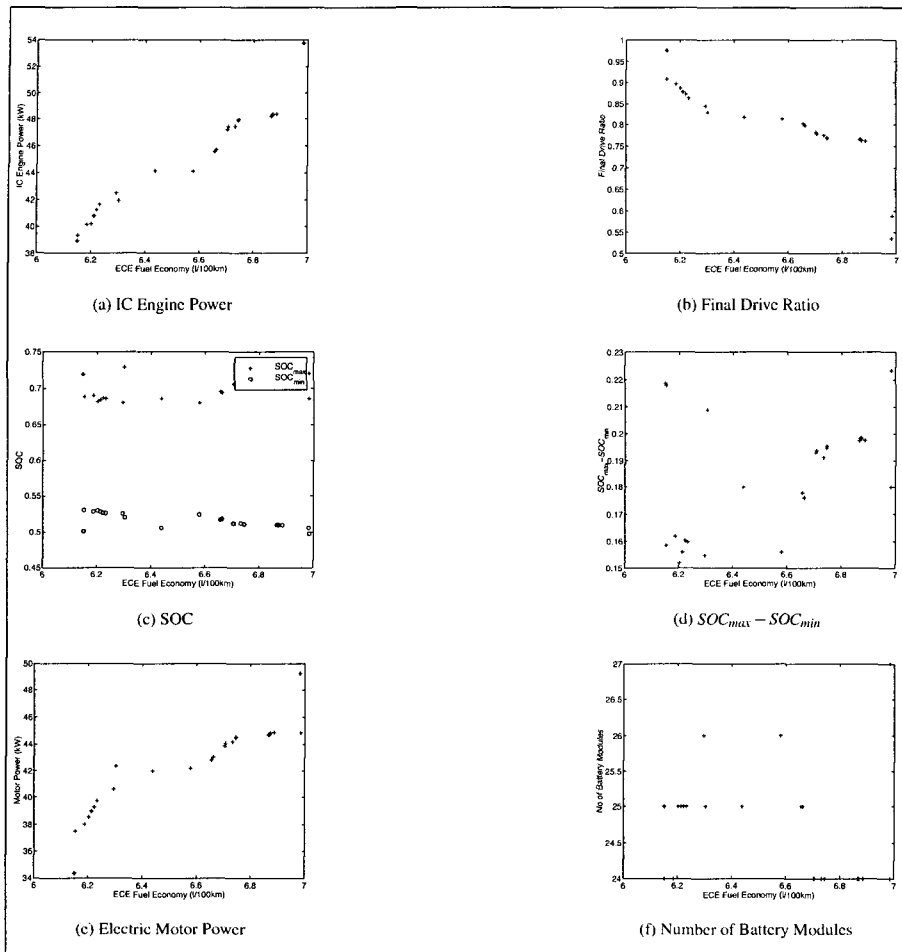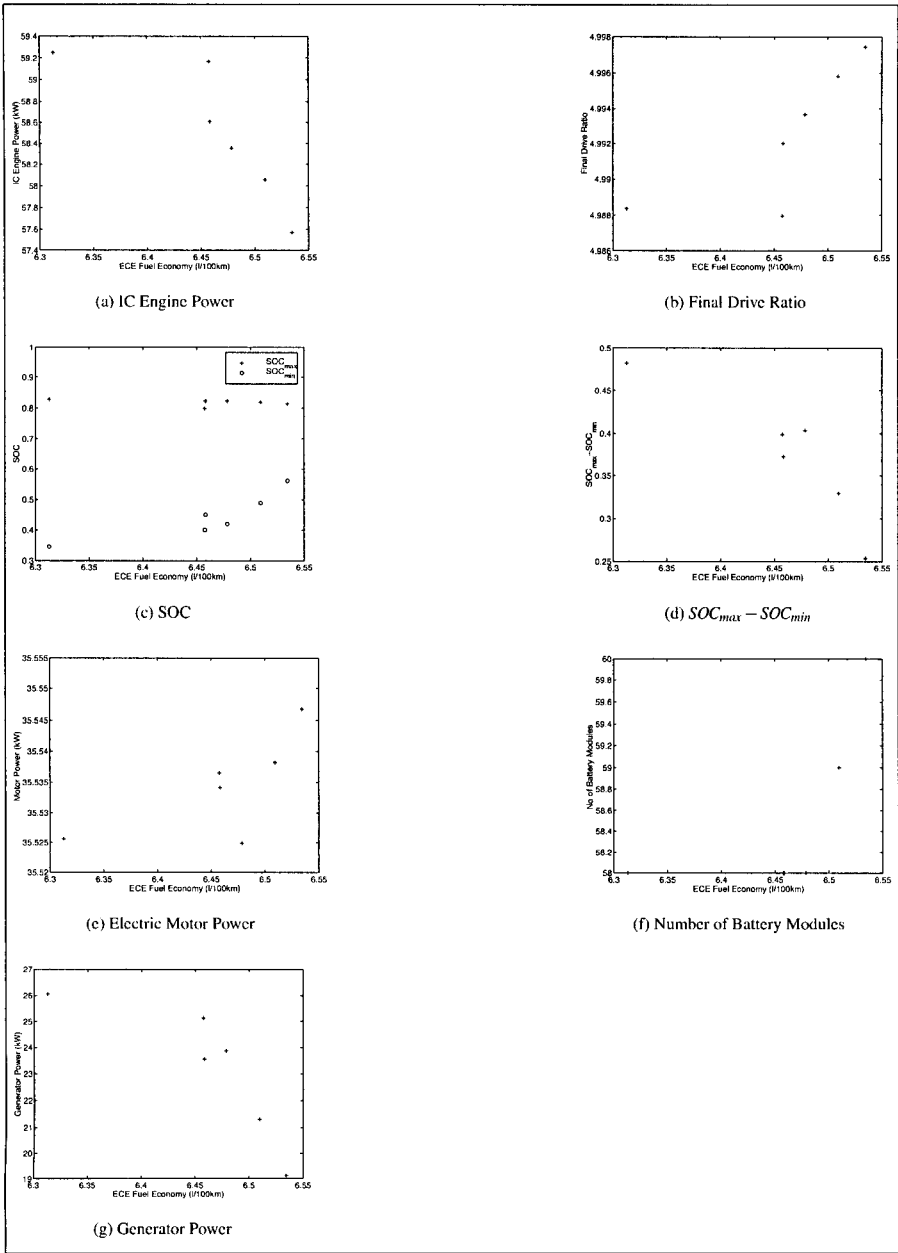
(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

Fig. A.30: Independent variable values for the parallel hybrid CI vehicle from the optimisation of fuel economy over US06HWY vs ECE cycles.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.31: Independent variable values for the modified Prius from the optimisation of fuel economy over US06HWY vs ECE cycles.
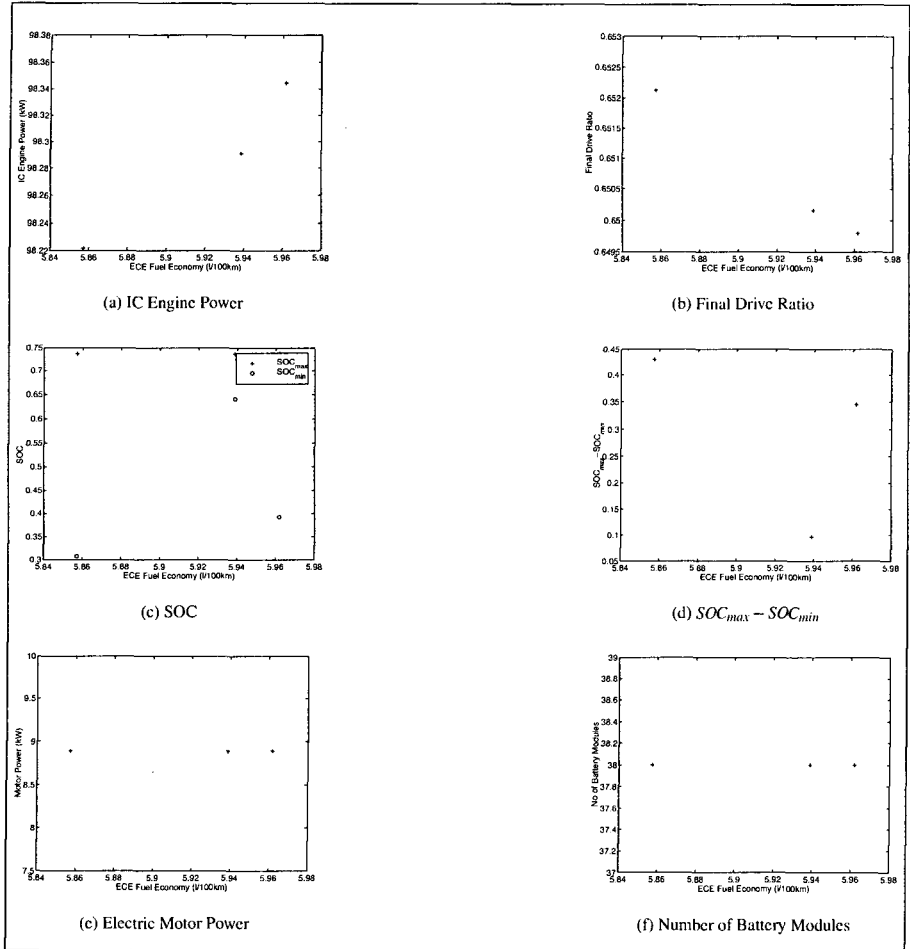
Fig. A.32: Independent variable values for the modified Insight from the optimisation of fuel economy over US06HWY vs ECE cycles.

A.2.4 Investment Cost vs Operating Cost

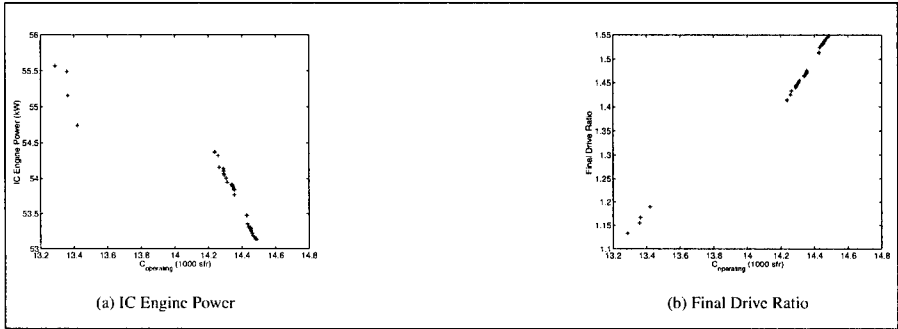Full results for the investment vs operating cost optimistaion.

(a) IC Engine Power

(b) Final Drive Ratio

Fig. A.33: Independent variable values for the conventional SI vehicle from the optimisation of investment vs operating cost.
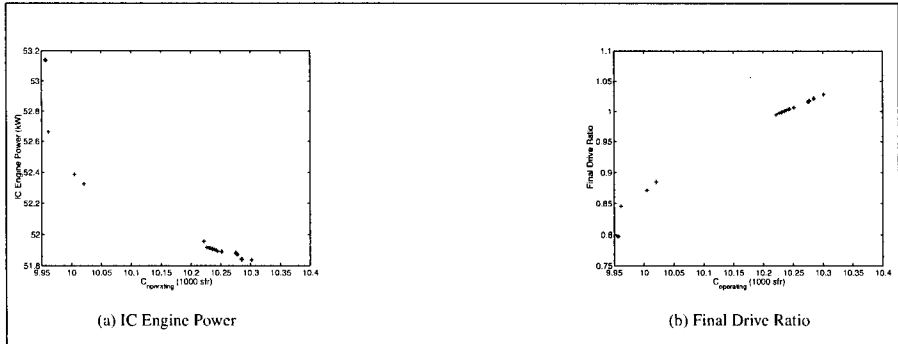

(a) IC Engine Power

(b) Final Drive Ratio

Fig. A.34: Independent variable values for the conventional CI vehicle from the optimisation of investment vs operating cost.

(a) IC Engine Power
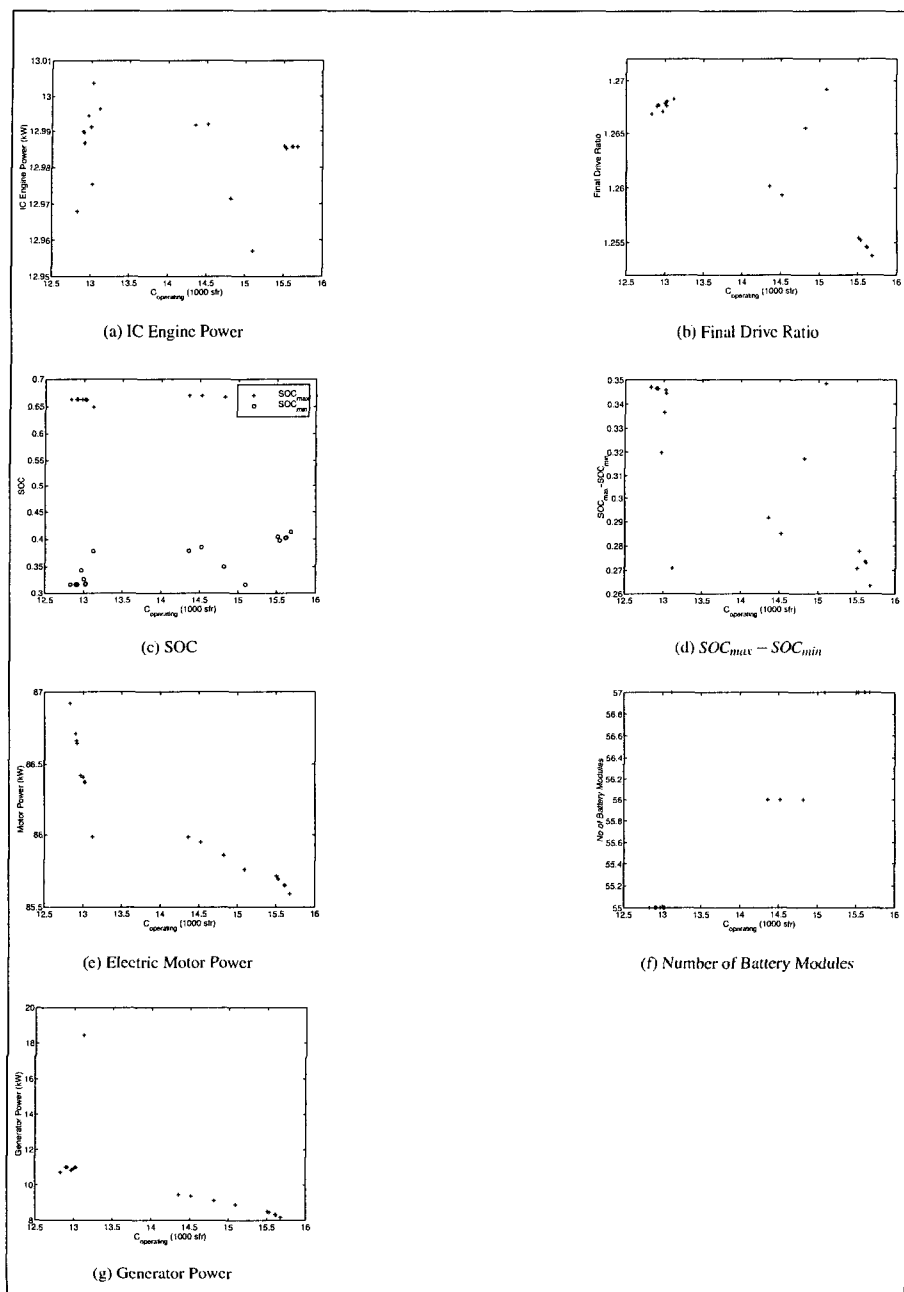
(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.35: Independent variable values for the SI series hybrid vehicle from the optimisation of investment vs operating cost.
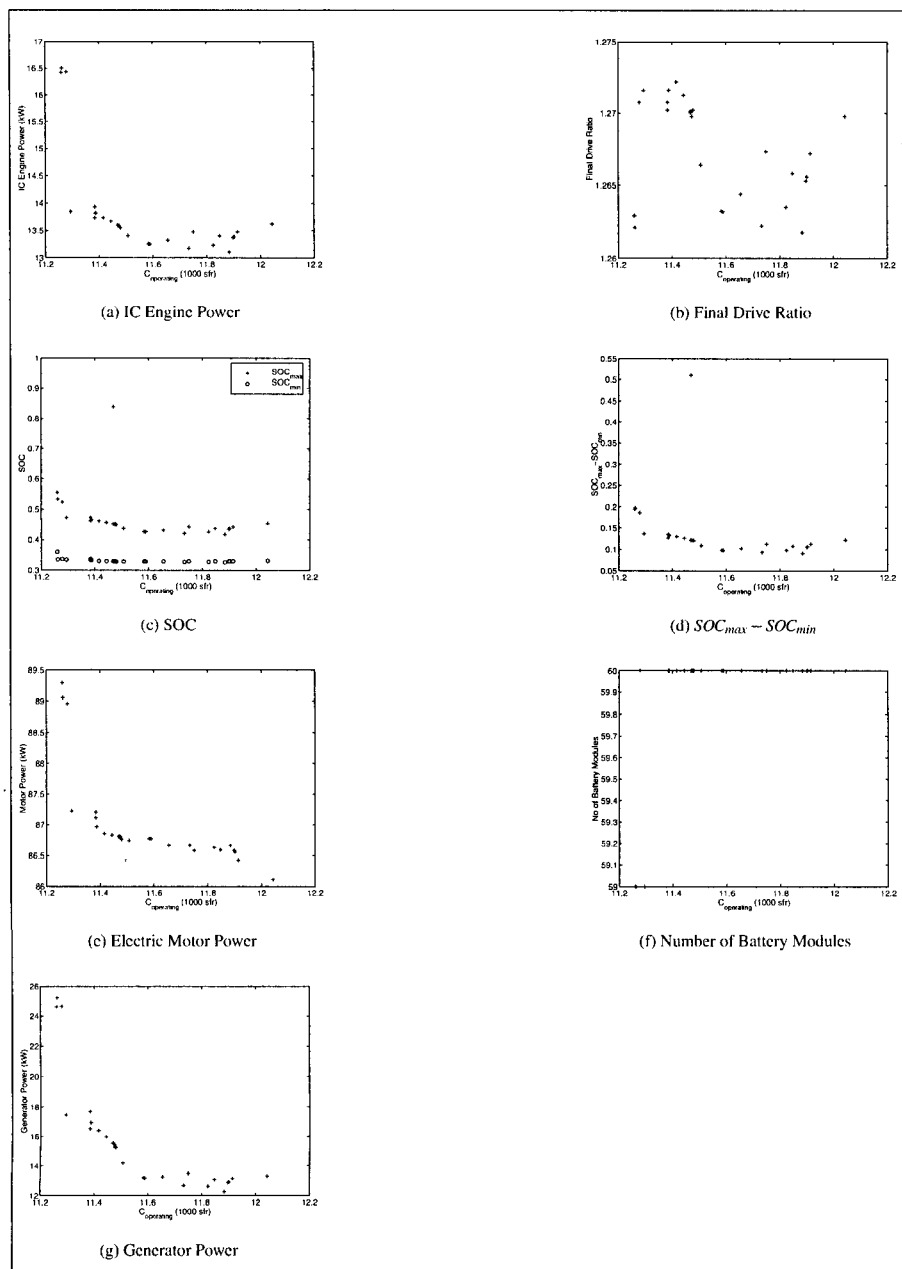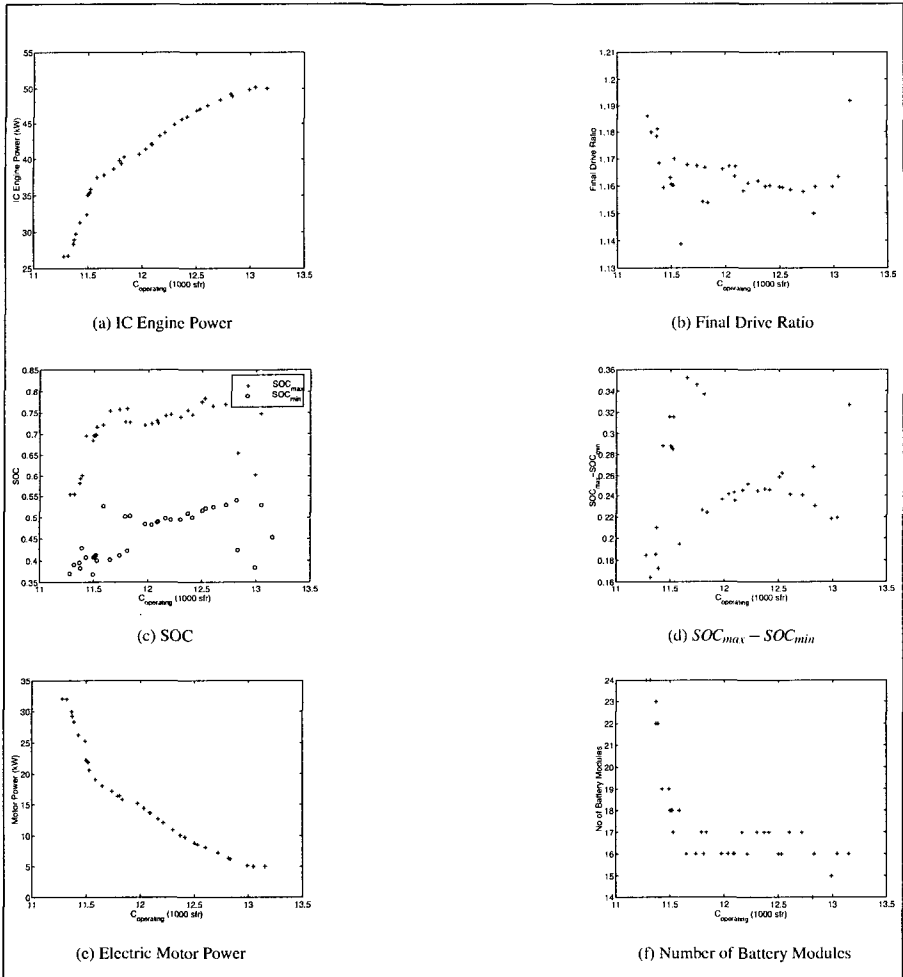
(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.36: Independent variable values for the CI series hybrid vehicle from the optimisation of investment vs operating cost.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

Fig. A.37: Independent variable values for the parallel hybrid SI vehicle from the optimisation of investment vs operating cost.

Fig. A.38: Independent variable values for the parallel hybrid CI vehicle from the optimisation of investment vs operating cost.
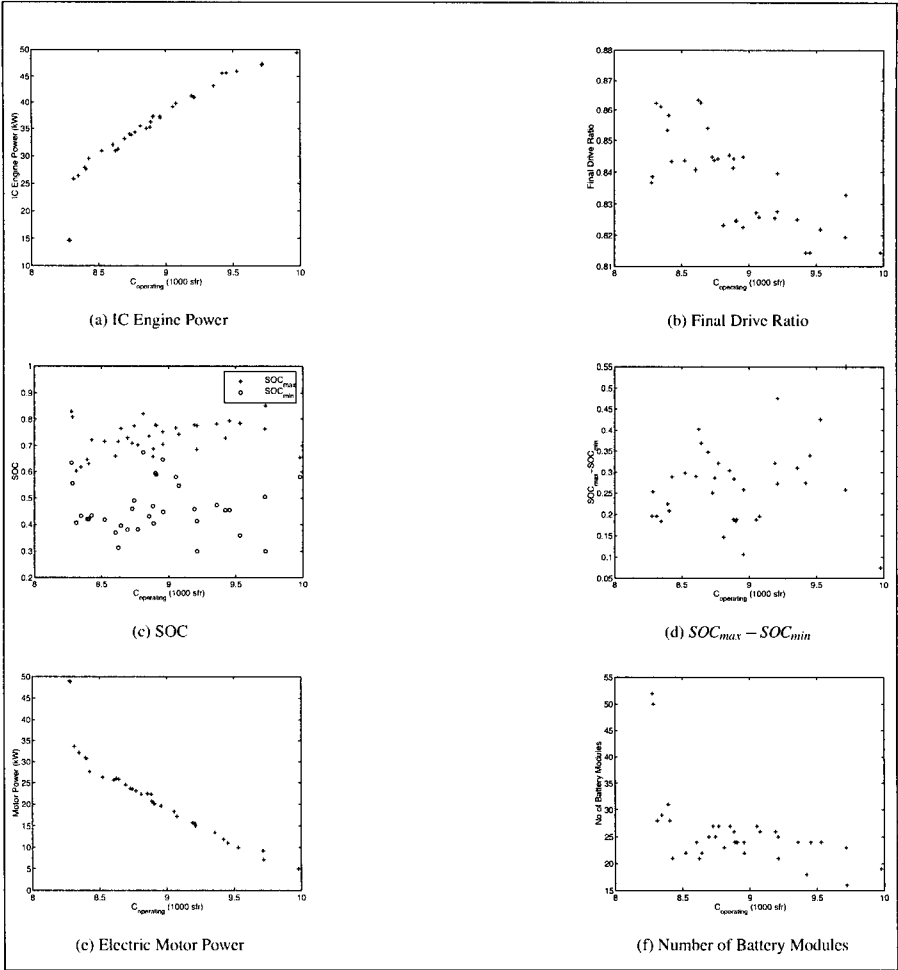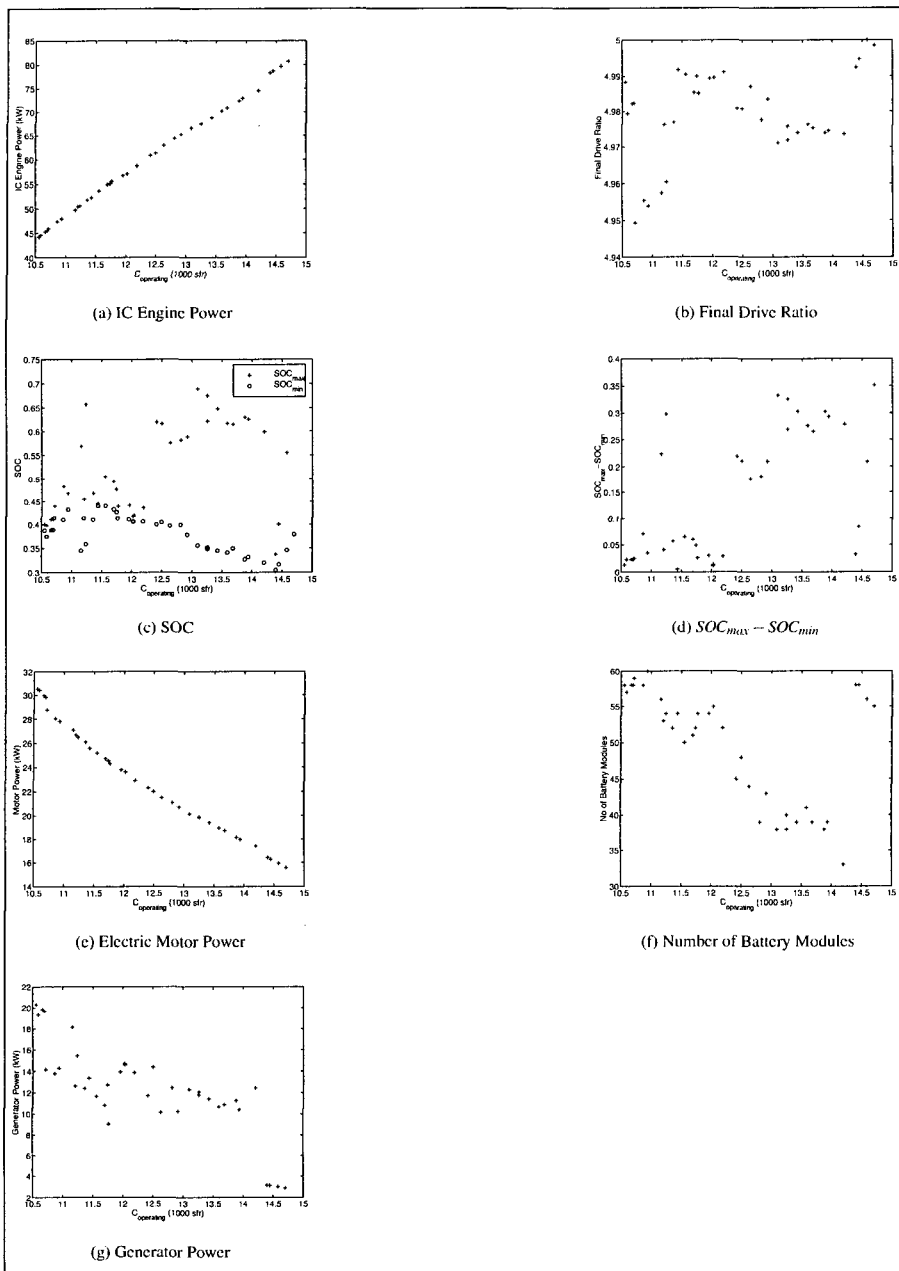
(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.39: Independent variable values for the modified Prius from the optimisation of investment vs operating cost.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power
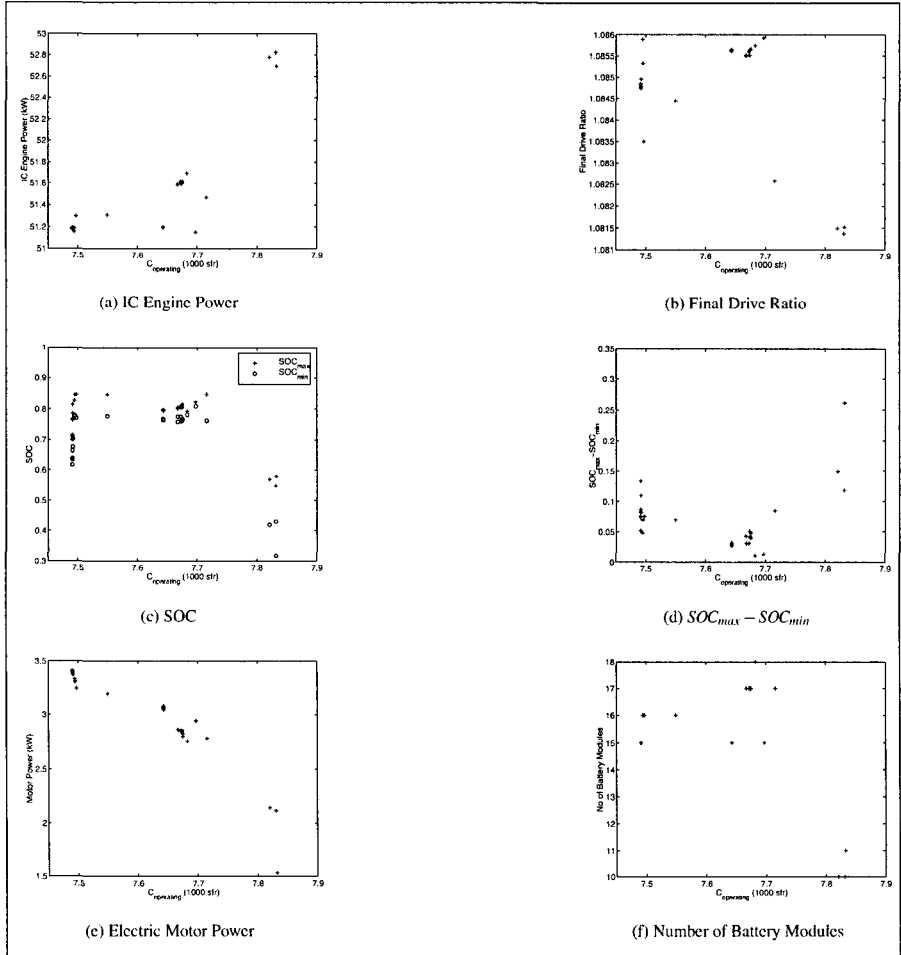
(f) Number of Battery Modules

Fig. A.40: Independent variable values for the modified Insight from the optimisation of investment vs operating cost.

A.2.5 Total Cost vs Pollution Cost

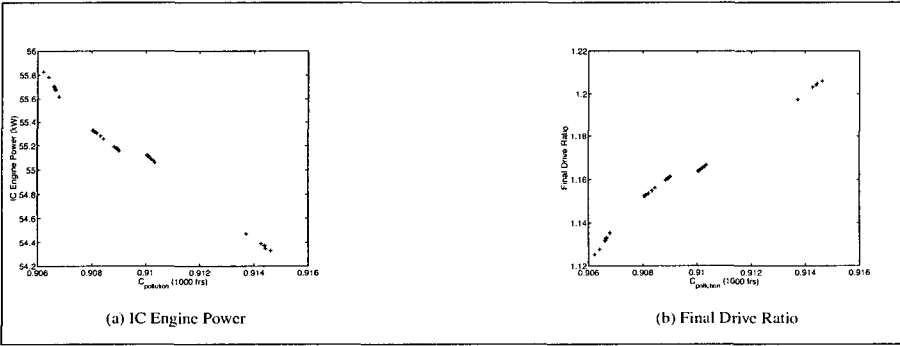Full results for the total cost vs NO$_x$ optimisation.

Full C_{elec}

(a) IC Engine Power

(b) Final Drive Ratio

Fig. A.41: Independent variable values for the conventional SI vehicle from the optimisation of total cost vs pollution cost.
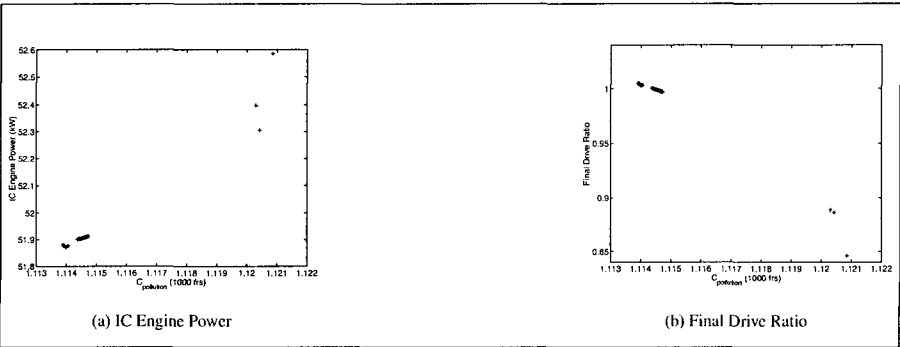

(a) IC Engine Power

(b) Final Drive Ratio

Fig. A.42: Independent variable values for the conventional CI vehicle from the optimisation of total cost vs pollution cost.

(a) IC Engine Power
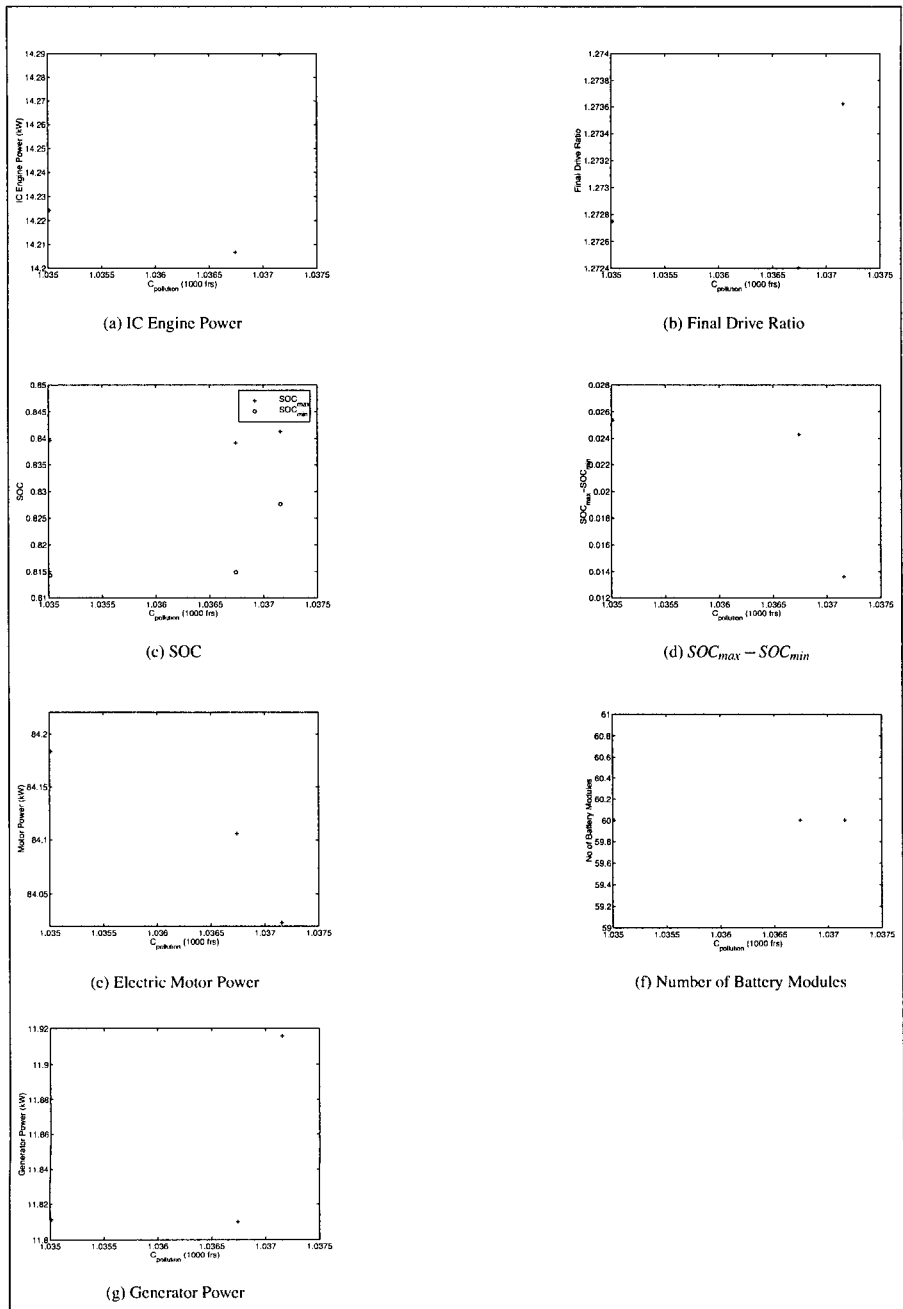
(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.43: Independent variable values for the SI series hybrid vehicle from the optimisation of total cost vs pollution cost.
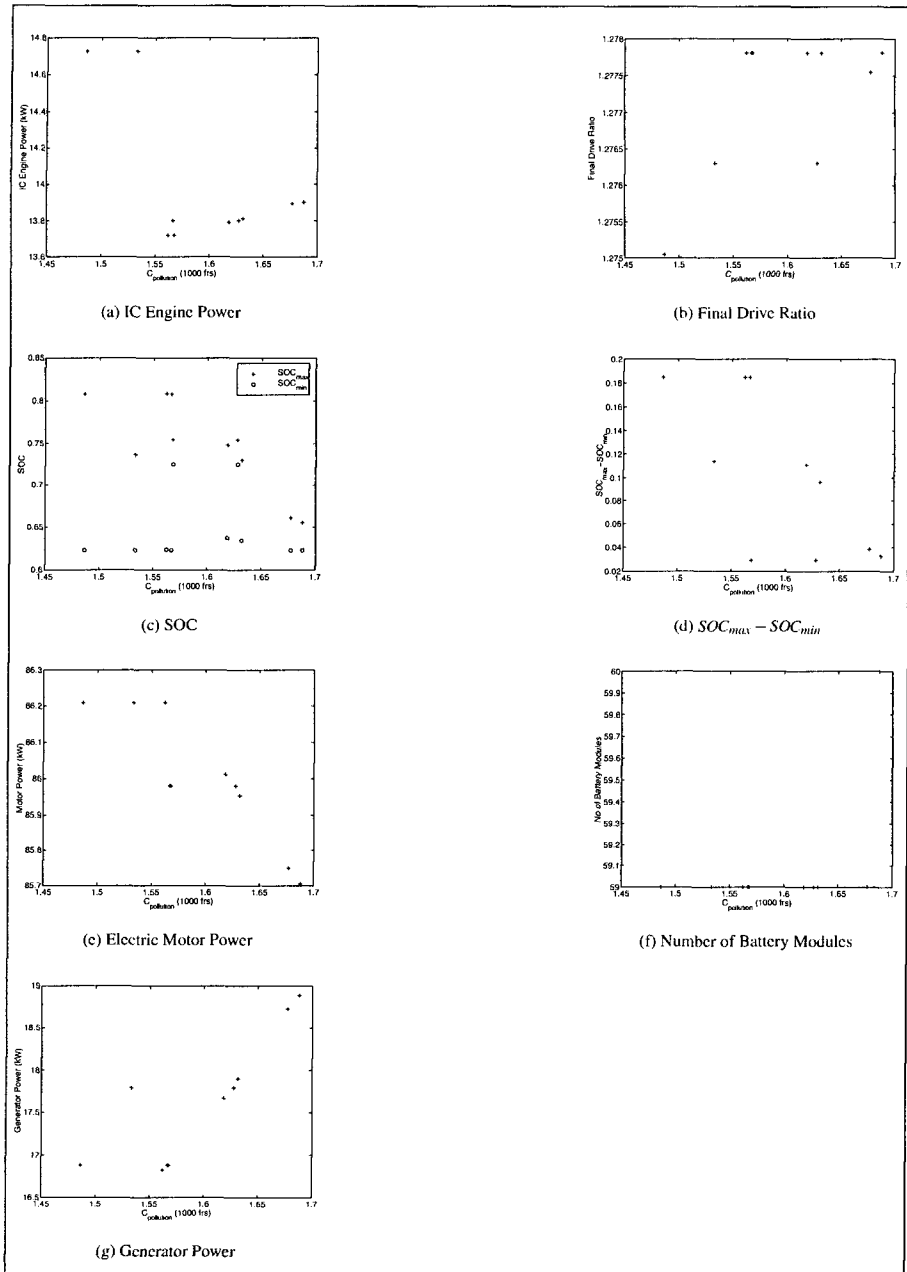
(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.44: Independent variable values for the CI series hybrid vehicle from the optimisation of total cost vs pollution cost.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$
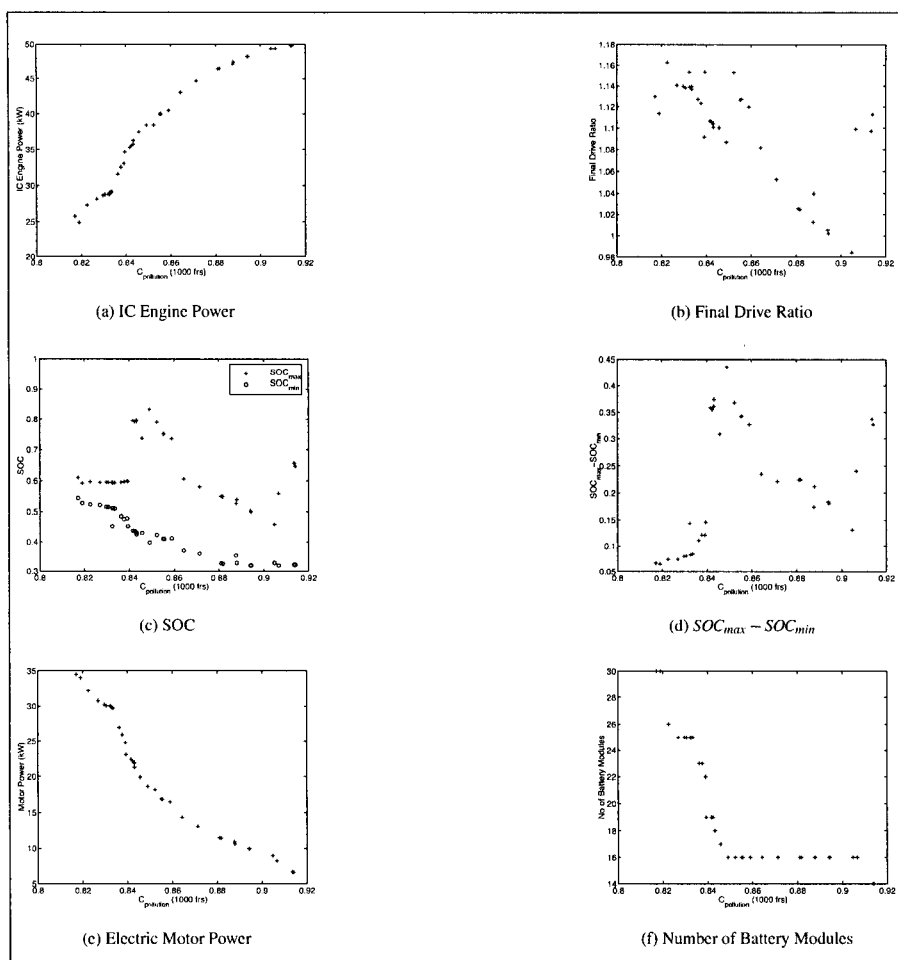
(e) Electric Motor Power

(f) Number of Battery Modules

Fig. A.45: Independent variable values for the parallel hybrid SI vehicle from the optimisation of total cost vs pollution cost.

(a) IC Engine Power
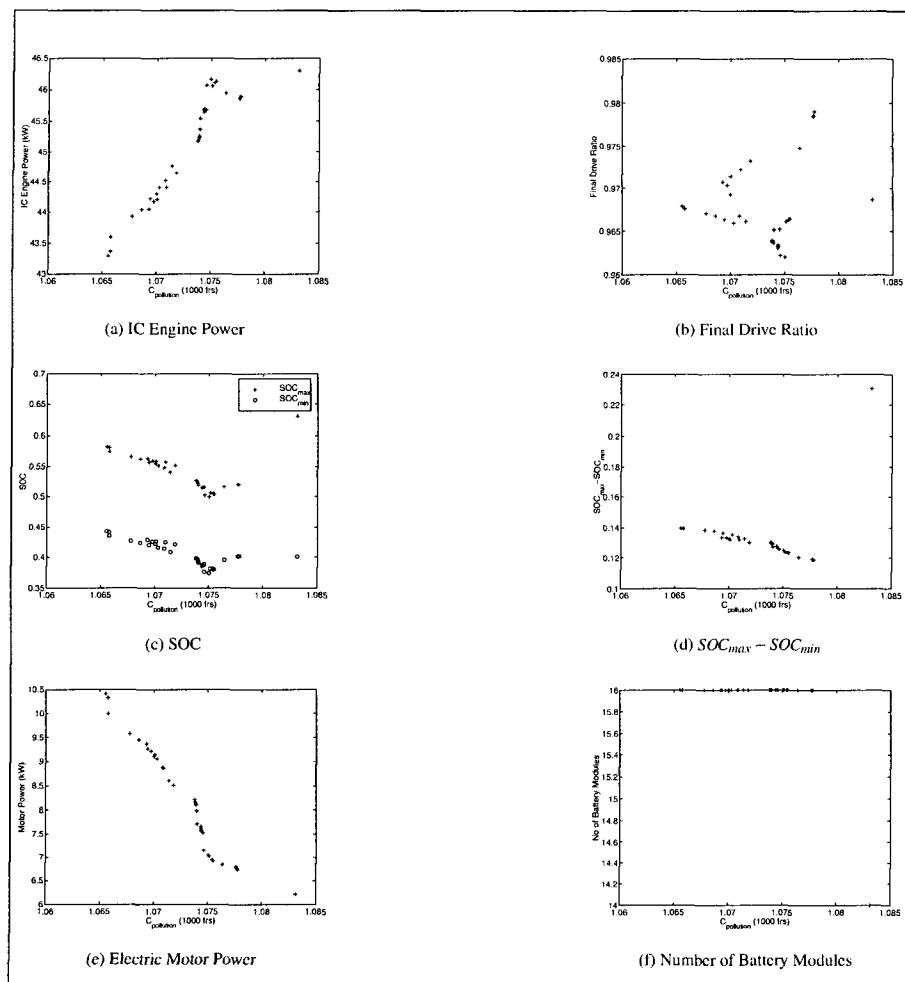
(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

Fig. A.46: Independent variable values for the parallel hybrid CI vehicle from the optimisation of total cost vs pollution cost.

(a) IC Engine Power
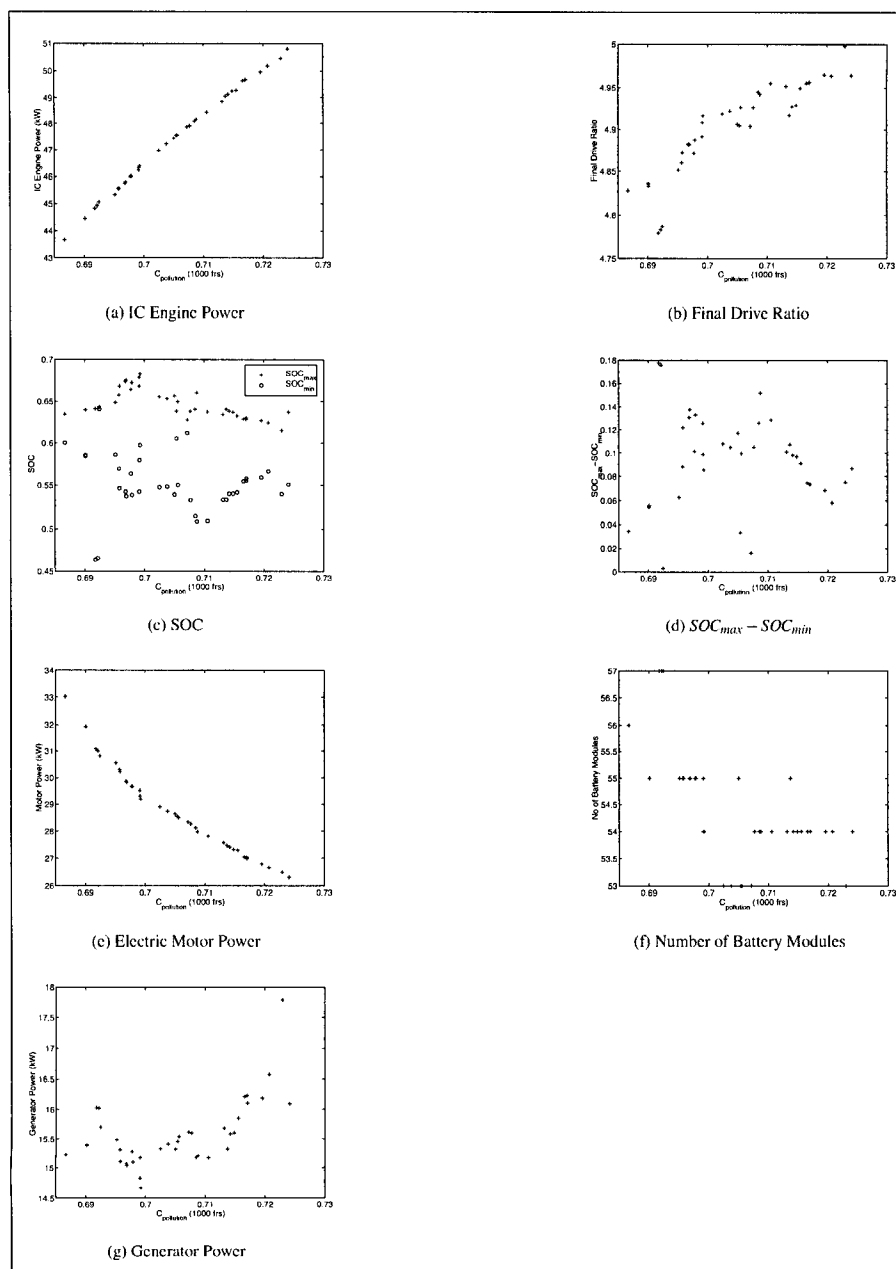
(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.47: Independent variable values for the modified Prius from the optimisation of total cost vs pollution cost.

Reduced C_{elec}

(a) IC Engine Power   (b) Final Drive Ratio
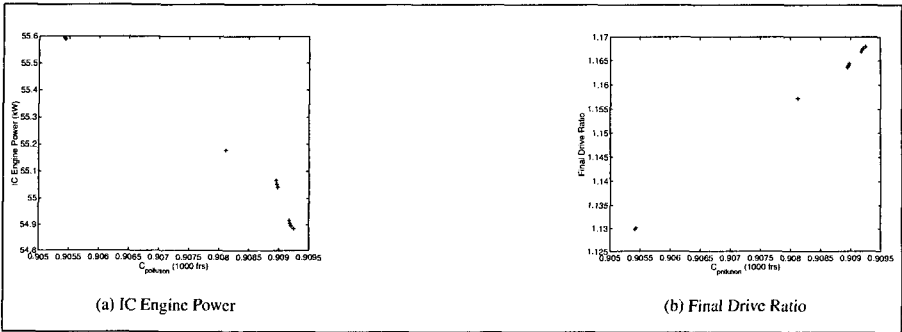
Fig. A.48: Independent variable values for the conventional SI vehicle from the optimisation of total cost vs pollution cost with c_{elec}=198.

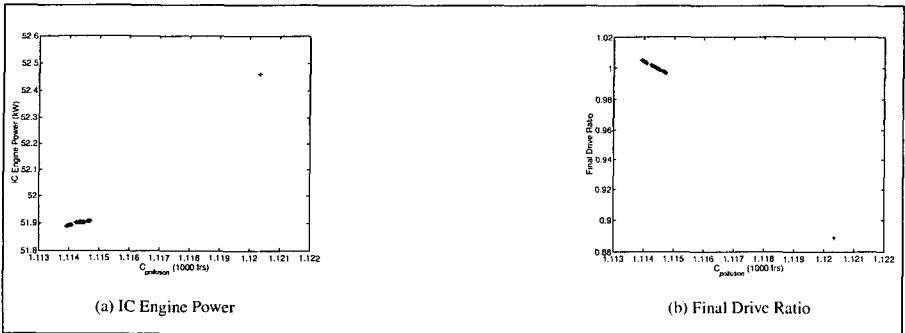(a) IC Engine Power   (b) Final Drive Ratio

Fig. A.49: Independent variable values for the conventional CI vehicle from the optimisation of total cost vs pollution cost with c_{elec}=198.

(a) IC Engine Power

(b) Final Drive Ratio

(c)
SOC

(d)
$SOC_{max} - SOC_{min}$

(e) Electric Motor Power
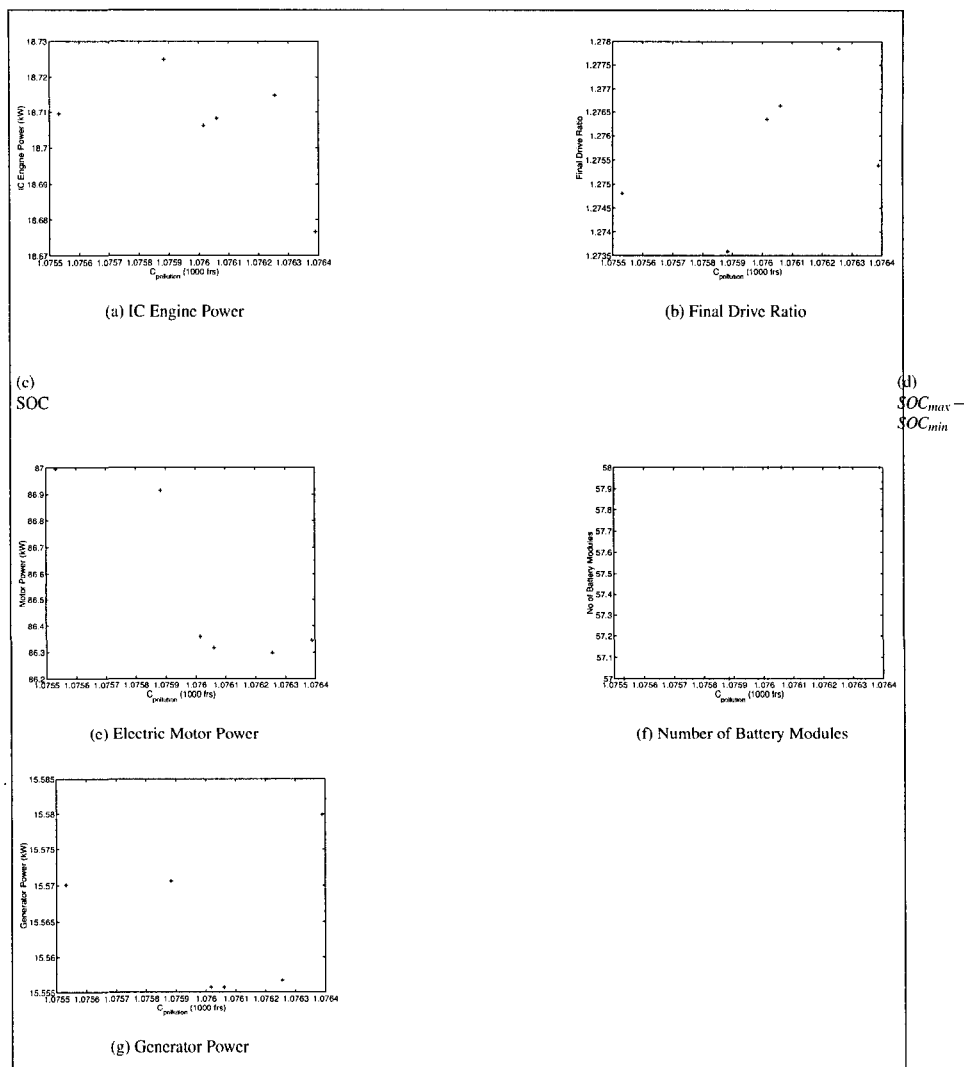
(f) Number of Battery Modules

(g) Generator Power

Fig. A.50: Independent variable values for the SI series hybrid vehicle from the optimisation of total cost vs pollution cost with $c_{elec}=198$.

(a) IC Engine Power

(b) Final Drive Ratio

(c)
SOC

(d)
$SOC_{max} - SOC_{min}$

(c) Electric Motor Power
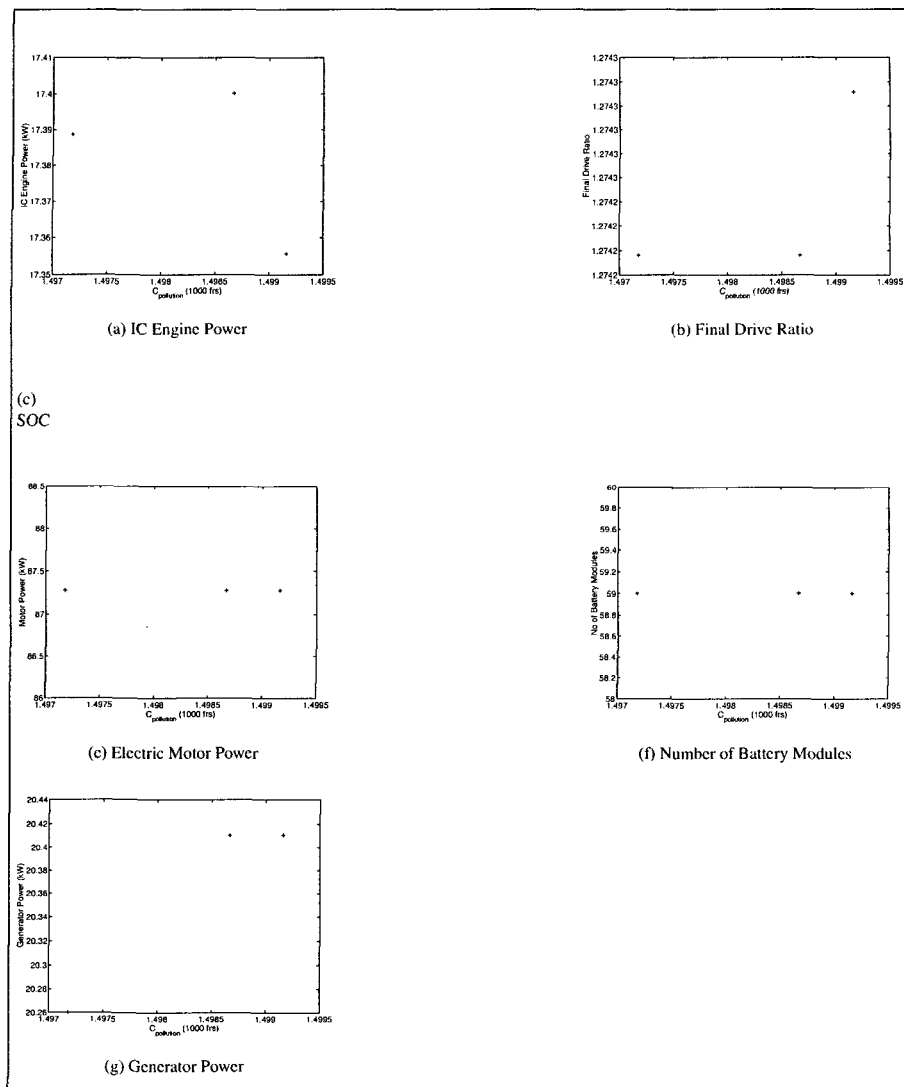
(f) Number of Battery Modules

(g) Generator Power

Fig. A.51: Independent variable values for the CI series hybrid vehicle from the optimisation of total cost vs pollution cost with c_{elec}=198.

(a) IC Engine Power

(b) Final Drive Ratio

(c)
SOC

(d)
$SOC_{max} - SOC_{min}$

(e) Electric Motor Power
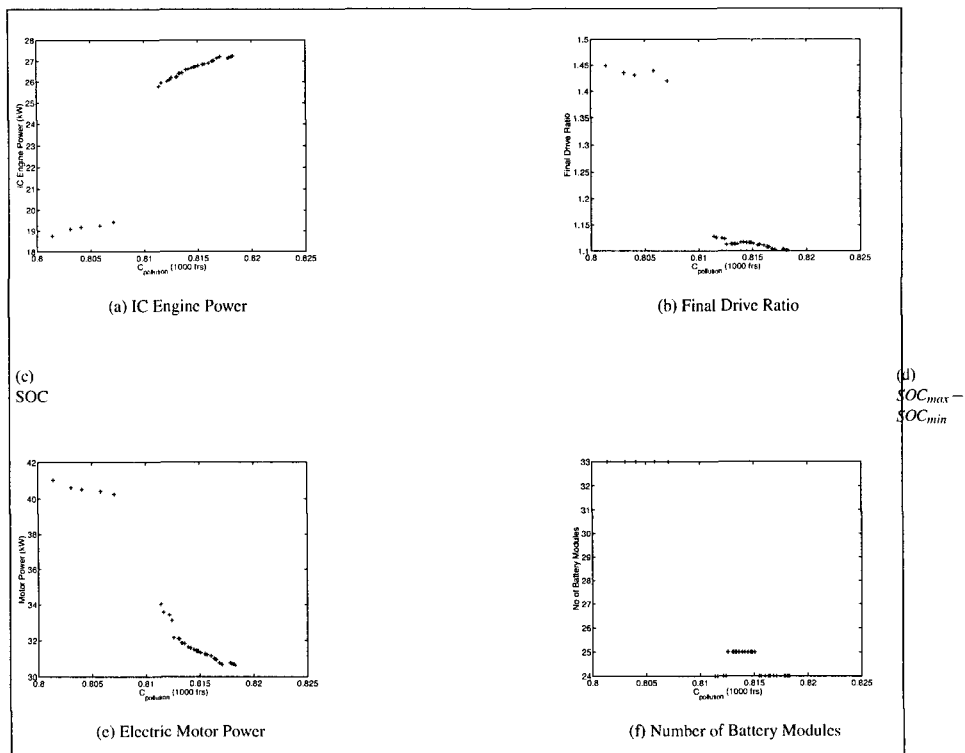
(f) Number of Battery Modules

Fig. A.52: Independent variable values for the parallel hybrid SI vehicle from the optimisation of total cost vs pollution cost with c_{elec}=198.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

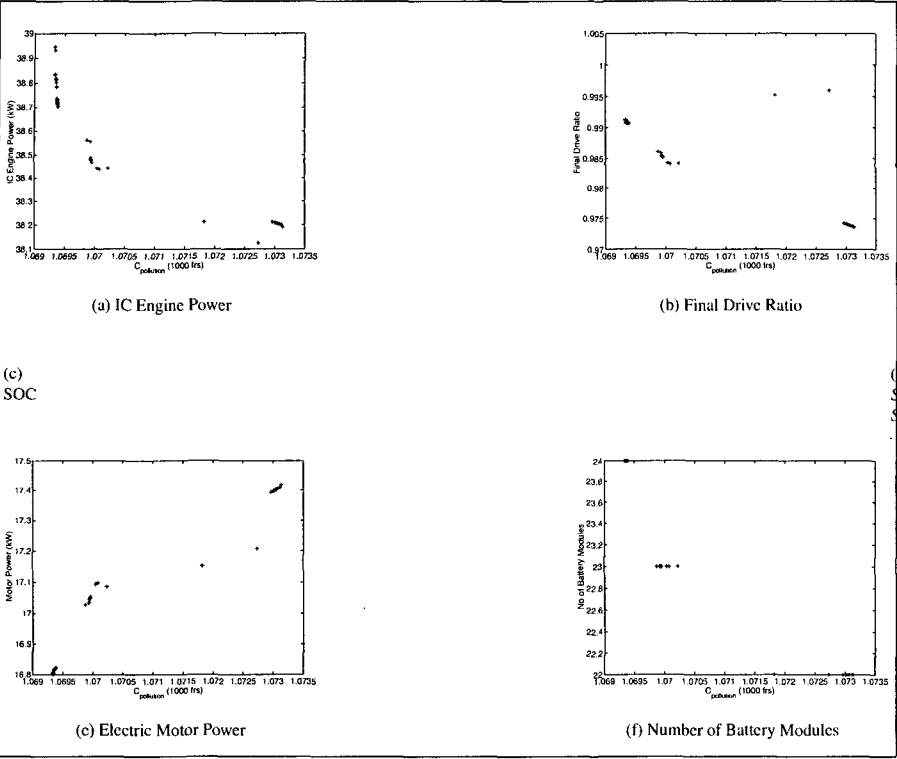(e) Electric Motor Power

(f) Number of Battery Modules

Fig. A.53: Independent variable values for the parallel hybrid CI vehicle from the optimisation of total cost vs pollution cost with c_{elec}=198.
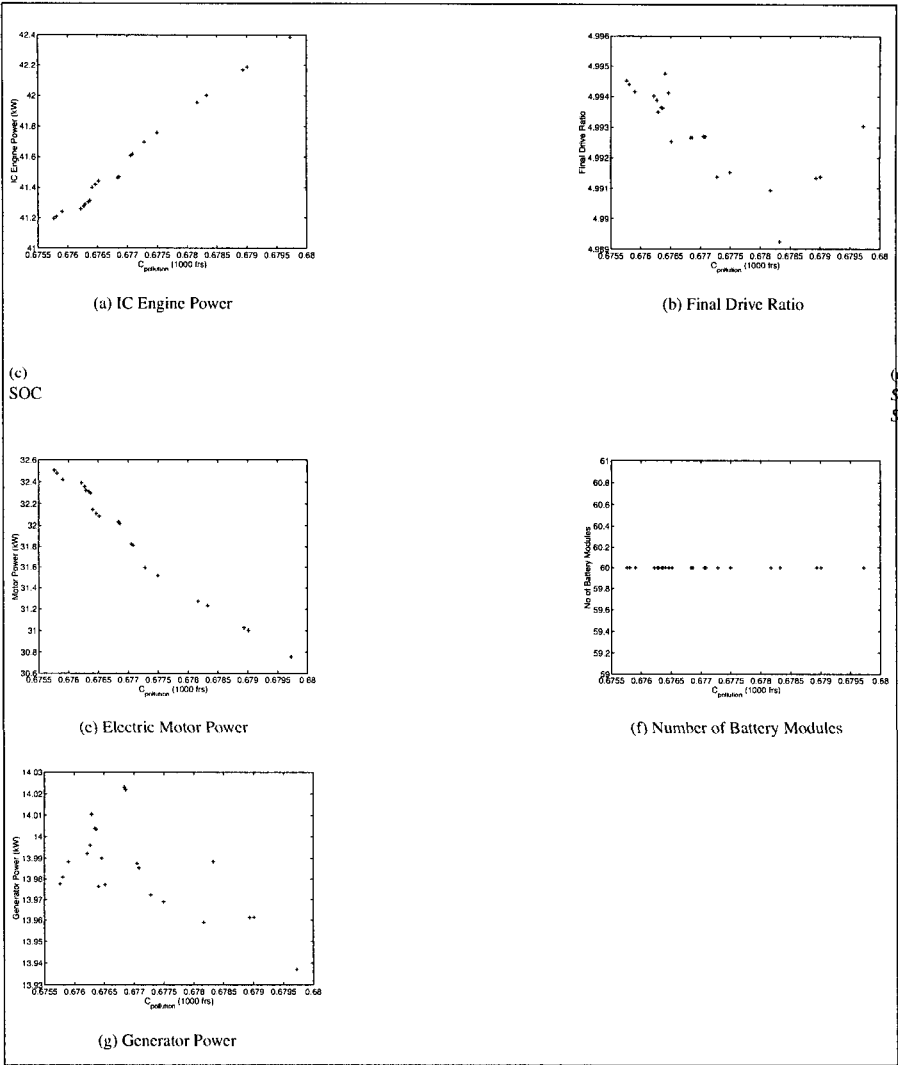
(a) IC Engine Power

(b) Final Drive Ratio

(c)
SOC

(d)
$SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.54: Independent variable values for the modified Prius from the optimisation of total cost vs pollution cost with $c_{elec}=198$.

A.2.6   Total Cost vs NO_x

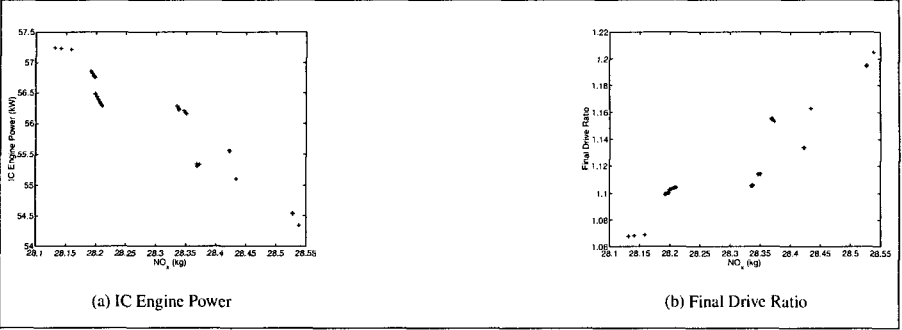Full results for the total cost vs NO_x optimistaion.

Fig. A.55: Independent variable values for the conventional SI vehicle from the optimisation of total cost vs NO_x.
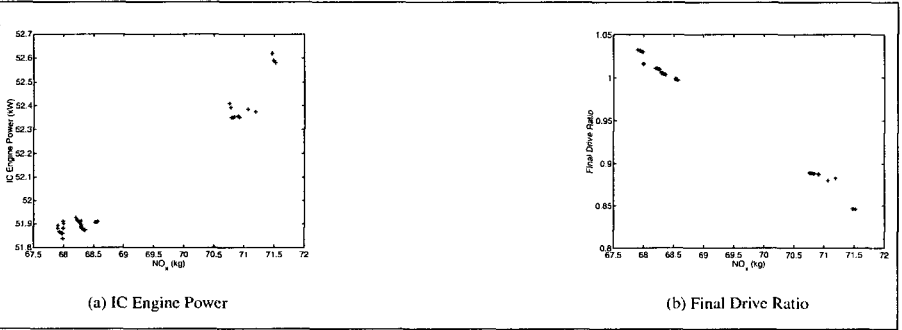

Fig. A.56: Independent variable values for the conventional CI vehicle from the optimisation of total cost vs NO_x.
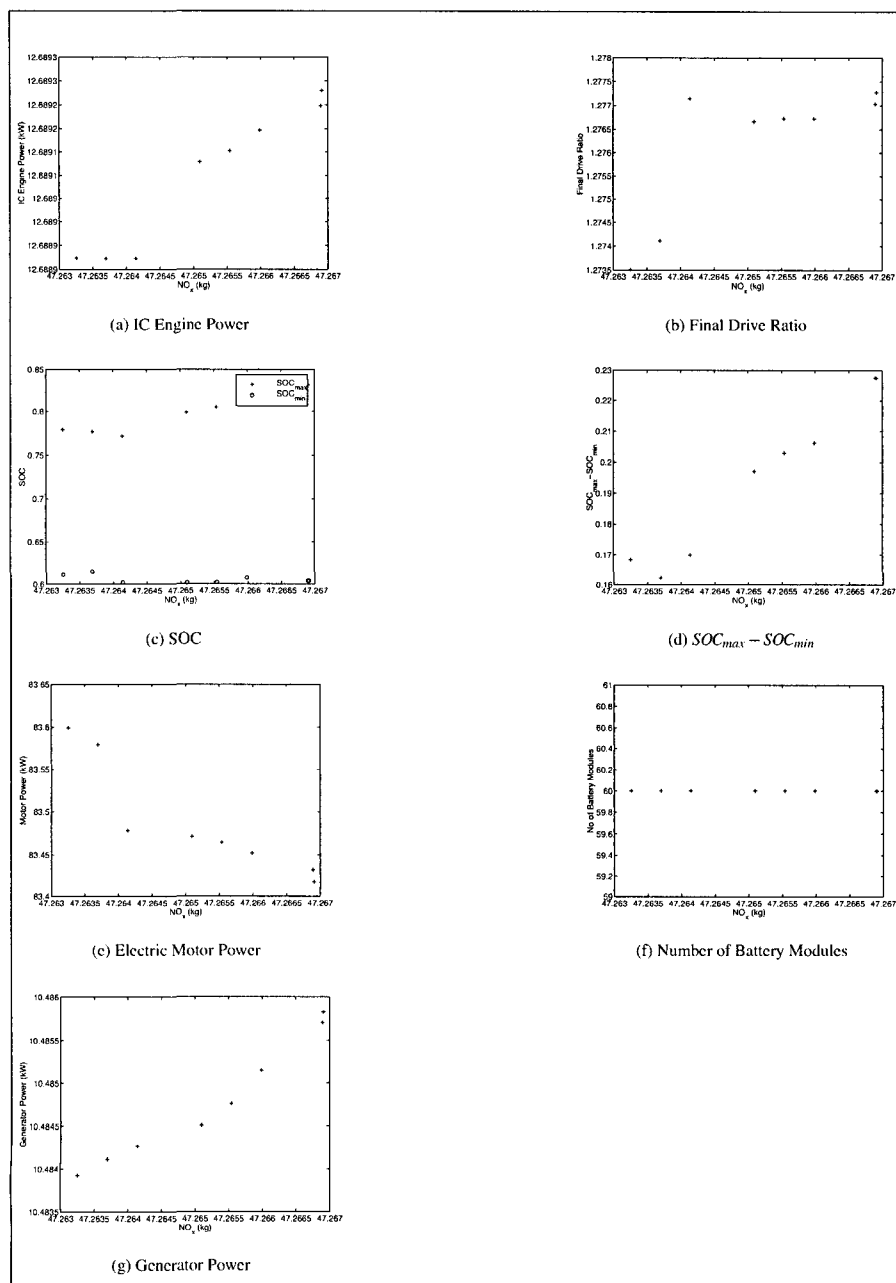
(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.57: Independent variable values for the SI series hybrid vehicle from the optimisation of investment vs operating cost.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

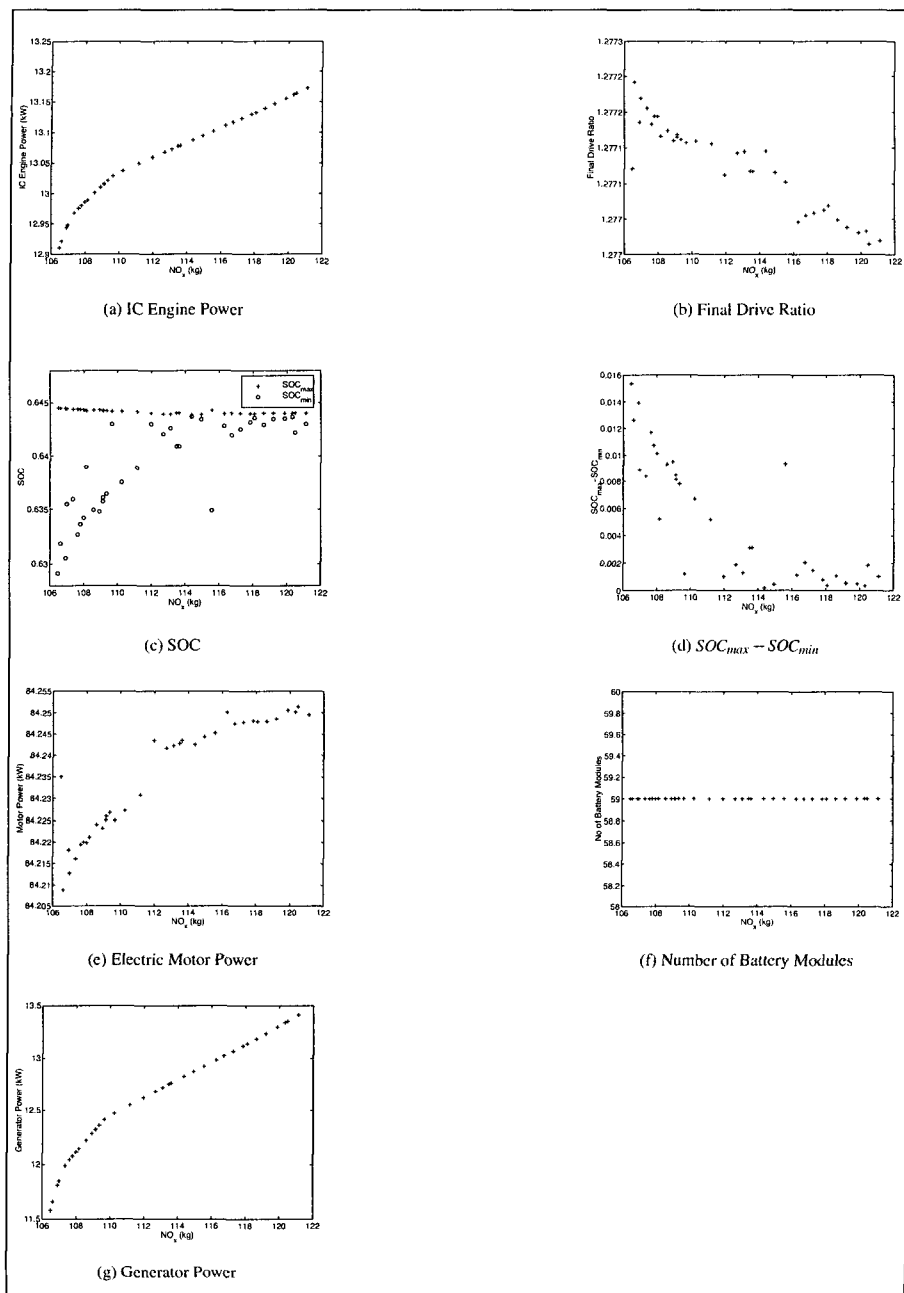(e) Electric Motor Power

(f) Number of Battery Modules

(g) Generator Power

Fig. A.58: Independent variable values for the CI series hybrid vehicle from the optimisation of investment vs operating cost.

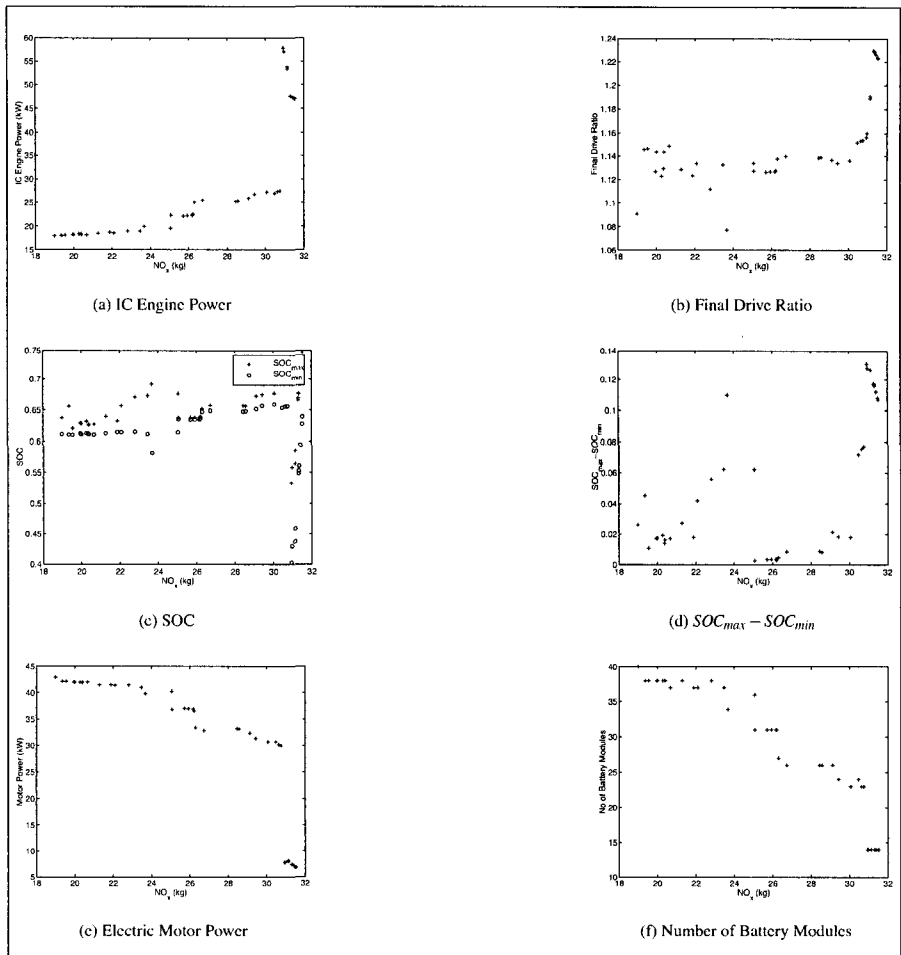(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules

Fig. A.59: Independent variable values for the parallel hybrid SI vehicle from the optimisation of total cost vs NO_x.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

(f) Number of Battery Modules
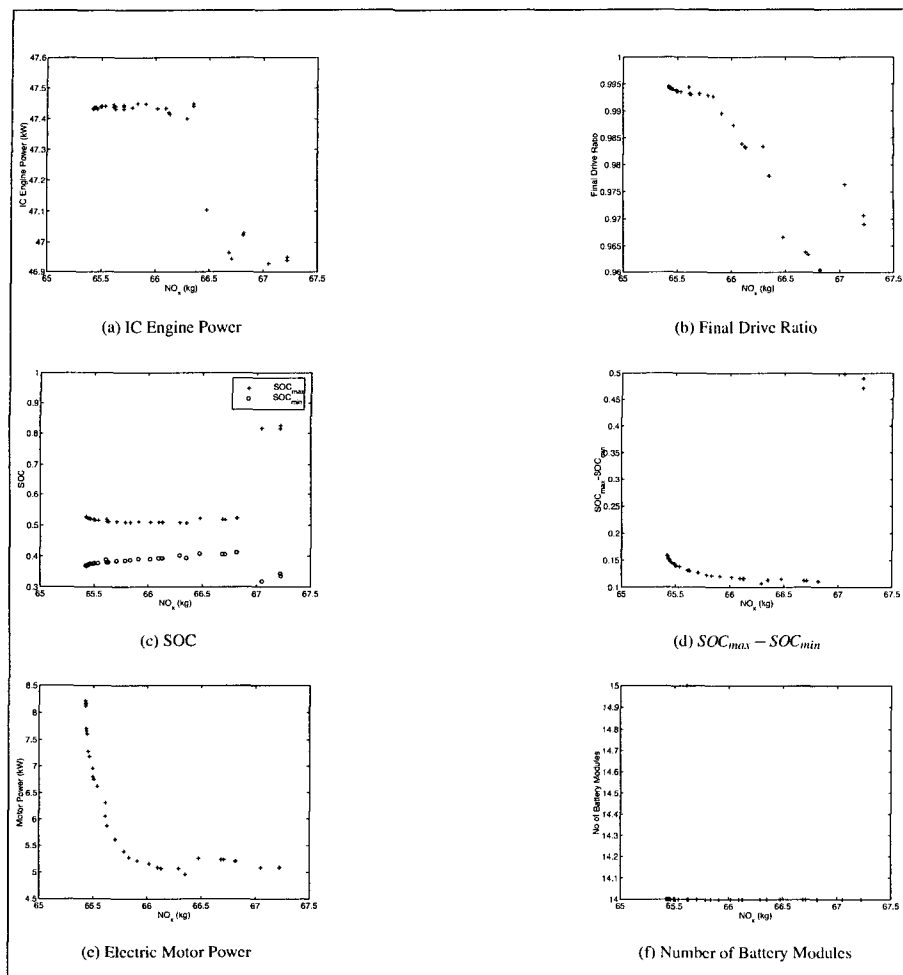
Fig. A.60: Independent variable values for the parallel hybrid CI vehicle from the optimisation of total cost vs NO_x.

(a) IC Engine Power

(b) Final Drive Ratio

(c) SOC

(d) $SOC_{max} - SOC_{min}$

(e) Electric Motor Power

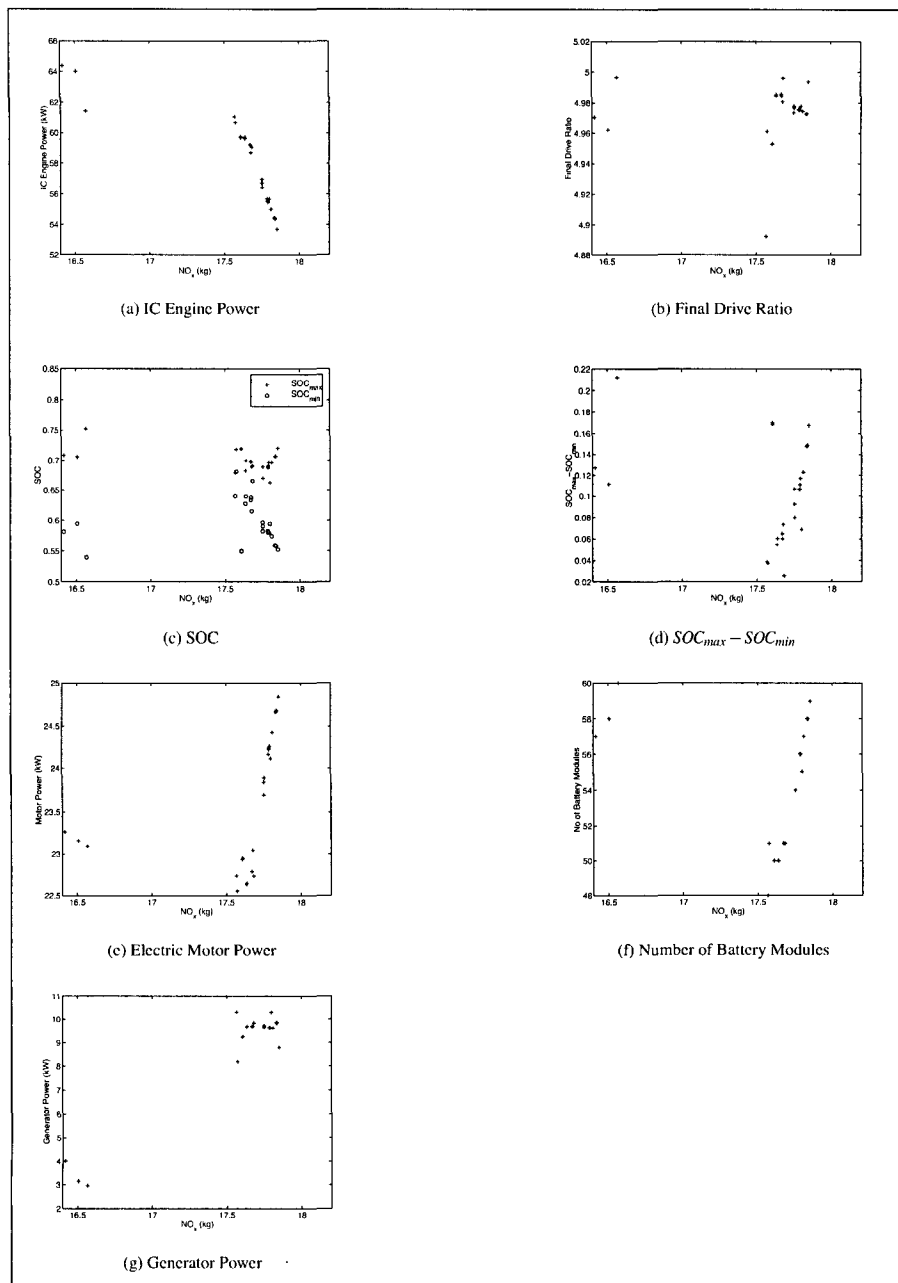(f) Number of Battery Modules

(g) Generator Power

Fig. A.61: Independent variable values for the modified Prius from the optimisation of investment vs operating cost.

Curriculum Vitae

Adam Molyneaux

born on the 7th of September, 1966 in Bath, England.

British Citizen

Education

1990	MSc Centre for Composite Materials, Imperial College of Science, Technology and Medecine.
1989	BSc Mechanical Engineering, Imperial College of Science, Technology and Medecine.

Professional Experience

1996-2001	Research assistant at the Laboratoire d'Energétique Industrielle (LENI), EPFL. Development of high-performance multi-objective evolutionary algorithms, simulation of district heating system and hybrid electric vehicle system performance.
1994-1996	*Independent engineer working for Peter Heppel Associates. Development of a high-speed structural/fluid dynamic simulation,* **RelaxII** *framework for battened fabric structures such as sails. Clients included several America's Cup teams.*
1990-1994	Research Officer at the Centre for Composite Materials, Imperical College. Included lecturing the MSc, and research into failure growth modeling of CFRP plates.

Publications

2001   A K Molyneaux, G. Leyland and D. Favrat *A New, Clustering Evolutionary Multi-Objective Optimisation Technique*, International Symposium on Adaptive Systems, Havana, Cuba, March 2001.

2001   G. Leyland, A K Molyneaux and D. Favrat *A New Multi-Objective Optimisation Technique Applied to a Vehicle Drive Train Simulation*, ECOS—Efficiency, Costs, Optimization, Simulations and Environmental Impact of Energy Systems, Istanbul, Turkey, July 2001.

1991   A K Molyneaux and A R Curtis *'Buckling and Postbuckling of Carbon Fibre PEEK Composite Panels'*, Sixth International Conference on Composite Structures (ICCS/6), Paisley College, Sept 1991, publ. ElsevierApplied Science, ed. I H Marshall.