ON A FLEXIBLE MODEL FOR NEW ZEALAND'S HYDRO-THERMAL ELECTRICITY GENERATION SYSTEM

A THESIS PRESENTED IN PARTIAL FULFILMENT
OF THE REQUIREMENTS FOR
THE DEGREE OF PH.D. IN
OPERATIONS RESEARCH AT
MASSEY UNIVERSITY

Shane Dye

1994

Abstract

This thesis investigates the modelling of the New Zealand hydro-thermal electricity generation system in order to determine an optimal strategy for generation, in terms of minimizing fuel costs. The model currently used by ECNZ (Electricity Corporation of New Zealand) uses an SDP (Stochastic Dynamic Programming) method for solution; this allows little detail of the physical system, and models *two* explicit hydro reservoirs. The model developed in this thesis is flexible, in order to allow the balance between ensuring stochastically stable solutions and the detail of the physical system, to be altered, whilst ensuring computational tractibility. The areas of the system which are important to be modelled accurately are isolated, as are those which may lead to computational intractibility if they are modelled in too much detail. The flexibility in the model also allows the effects of the approximations used on solutions to be explored in a wider framework.

The time horizon of the model is one to two years, with time steps of the order of a week. The time horizon describes the level to which many aspects of the system are to be modelled. Transmission is modelled explicitly so as to include information on the geographical locations of power stations and power users; this takes the form of a network structure underlying the model. The load at each geographic location is represented by a Load Duration Curve (which is more robust, in terms of forecasting, than a direct representation of load with respect to time). Hydro river chains are modelled as single power stations with a single reservoir and connect the model temporally; we model six explicit hydro river chains. Thermal stations are modelled individually, and the generation from run-of-river and geothermal stations is removed from load before solution begins.

The initial approach considers a model which, upon further investigation, is unacceptable. However, examination of the issues highlighted by this approach provide insight into the system. The resultant re-modelling of the problem leads to a linear model which does not explicitly model the uncertainty in the generating

capacity of stations due to forced outages. This accentuates the reason why the usual approach to explicitly modelling the uncertainty of supply (within a week) cannot be used in the case where the geographic distribution of generation has been explicitly modelled. The deterministic model may then be formulated as a Generalized Network with side constraints.

The deterministic model developed can be extended stochastically in many ways. The stochastic extension investigated uses Rockafellar and Wets' *Progressive Hedging Algorithm*. This takes a scenario approach, in which the stochastic variables are approximated by a number of scenarios of observed values. A policy is required which minimizes the expected cost of generation over these scenarios, ensuring that information on the observed values of the stochastic variables is not used before it would be available in practice.

Results and implementation issues are discussed for both the deterministic and stochastic models. Consideration is given to the implementation of a finished product, as well as implementation for the purposes of investigating the feasibility and examining the computational effectiveness of approximations made in the model.

Acknowledgements

I would like to thank ECNZ for their funding of this project and for sending me to the 15th Symposium on Mathematical Programming which gelled many of the ideas in this thesis. My chief supervisor Dr. John Giffin deserves a lot of praise for his help, support, extensive knowledge of the English Language, and, "in some ways," for telling me when to stop. I would also like to thank my second supervisor Dr. Jonathan Lermit for his help, ideas (it's a shame about the cumulants), valiantly trying to print the colour figures within this thesis, for coming up every month or so to ensure I was actually doing some work, and for his nice comments about markets (which I have studiously tried to avoid). I would also like to thank Philippa, my wife, for her love and support, for her proof reading (even though she complained about not understanding most of it), and especially for her interest and enthusiasm in my thesis.

There are many other people I would like to thank but don't have the room for, although a special note of thanks goes to Richard "Vancouver Coast Guard" Rayner for getting Saruman and Matlab back up again (well, for calling the repairman) when I had only two weeks to finish this think, and for being away the six days before I sent it in to be bound, while I was trying to get the colour pages printed.

Special thanks goes to Philippa (my wife) and Mark Johnston for helping me to "colour in" the coloured pictures, *by-hand*, the night before binding. Thankyou.

Addendum

Page 25, line 11; "*optimallity*" should read "*optimality*".

Page 109, §6.2, sentences two and three should become:

"The general stochastic program can be written as a multi-stage stochastic program with recourse. The two-stage stochastic program with recourse can be written as follows:

$$\text{Min} \quad f_1(x) + \underset{\tilde{\xi}}{E}\left[f_2(x,\tilde{\xi})\right]$$

subject to: (6.0a)

$$Ax = b$$

where x is the decision variable representing a decision that must be implemented prior to the realization of the random variable $\tilde{\xi}$, $f_1(x)$ represents the cost of decision x, and $f_2(x,\xi)$ (where ξ is a single observation of $\tilde{\xi}$) is defined as:

$$f_2(x,\xi) = \text{Min} \quad g(y)$$

subject to: (6.0b)

$$Wy = \xi - Tx,$$

$$y \geq 0$$

where this involves the determination of the optimal recourse variable y given the initail decision x. Extension to the multi-stage case invloves defining (6.0b) in a similar manner to (6.0a). In our case the recourse variables are the releases of the subsequent weeks."

Page 113, paragraph 3; should be appended with the sentence:

"While there are many other stochastic techniques which could be considered, since the focus here is to show that it is *feasible* to extend the deterministic model developed to a full stochastic model and we cannot cover every method here, the following are a *selection* of approaches which have been used in the past to model such a system."

Page 115, §6.5, line 2; "...as it offers the greatest flexibility in the extent ..." should read "...as it appears to offer the greatest flexibility, of any of the many possible stochastic approaches which could be used, in the extent ..."

Contents

List of Figures

List of Tables

Chapter 1

Problem Definition

Before proving a Mathematical theorem one needs to have a clear definition of the premise. So it is with Operations Research; before modelling, one needs a clear definition of the problem to be modelled—the features that it is important to model well, and those for which a coarse approximation suffices. This becomes very important in the case of a large, complex, problem such as the New Zealand hydro-thermal electricity scheduling problem; designing a model which precisely models the entire system, but is practically insolvable, is of little real use.

The New Zealand hydro-thermal electricity scheduling problem is a "large-scale" problem with additional highly variable stochastic elements, making it extremely difficult to model in an accurate, consistent fashion. This means that there needs to be a clear definition of the aspects of the problem which should be emphasized.

The perspective taken here is to develop a model which gives more physical detail than the model currently used by ECNZ (Electricity Corporation of New Zealand). In particular, the goal is to include the explicit modelling of six separate hydro reservoirs. Consequently, it may not be possible to model the stochastic elements to a level of detail similar to that of the current model. This investigation is not intended to produce a finished product for ECNZ, but to develop a model, investigate whether it is viable, and determine whether it is useful enough to develop further. In particular, the focus of this thesis is directed to the *development* of a useful model, rather than rigorous testing of the stability, robustness and quality of solutions produced; this is seen as an appropriate second stage of model development—important if the model is to be used, but beyond the scope of this thesis.

1.1 Description of the System

The system being modelled is New Zealand's hydro-thermal electricity generation system. Electricity is generated by various power stations and distributed via the National Grid (a transmission network of power lines) to meet the current load. Stations are powered by various "fuels". Thermal stations generate power using heat created, in general, by burning various hydrocarbon based fuels, mainly gas by-products and coal. Hydro stations generate power using water flow in rivers, which is partially controlled by hydro dams further up the river. Power from other sources include geothermal stations; there are *no* nuclear power plants in New Zealand.

The stations tend to be partitioned into three groups, characterized by their generation constraints: thermal stations have a cost applied to the fuel, but little constraint on the usage of this fuel; hydro stations have no direct cost attached to water use, but they have limited storage and uncertain replenishment of this water; and, auxiliary stations which, unlike the first two groups, have no useful control over their level of generation—they tend to be small capacity stations and are run continuously. We also include as auxiliary stations, those which are quite complicated to model but have a very minor effect on the overall total generation. These include some minor hydro stations, which are not part of other hydro systems, and are isolated from them; such stations can be thought of as acting like "free" or base-loaded thermal stations.

Some of the statistics of the major stations are given below. This is to provide some background on the New Zealand system, as well as to provide an idea of the *scale* of the system, and to illustrate the systems capabilities and constraints. New Zealand has six major hydro systems. The Waikato system (in the North Island) has Lake Taupo (New Zealand's largest lake) as its reservoir, and eight hydro stations in the river system. Lake Taupo can hold in the order of 600 Gigawatt hours (GWh) of potential energy, and the hydro system has a generation capacity in the order of 900 Megawatts (MW). The Tongariro system (also in the North Island) has three hydro stations and can only hold up to 1 GWh of potential energy; the generation capacity of this system is 400 MW. The last hydro system in the North Island is the Waikaremoana system. This system has three hydro stations, a combined generating capacity of 100 MW and its reservoir can store in the order of 50 GWh of potential generation. The Waikaremoana system is the only New Zealand system with significant storage depletion over time.

Of the South Island hydro systems, the Waitaki system is the most extensive. It has three major reservoirs and 10 hydro stations (two minor separate hydro systems are also included in this system). The Waitaki system can store up to 2000 GWh of potential generation, and has a combined generation capacity of 1600MW. The Clutha system of the South Island has two hydro stations with a combined capacity of 700 MW and can store up to 300 GWh of potential generation. The final South Island hydro system, the Manapouri system, has the most extreme inflows and hence spills most often. It is a simple hydro system consisting of a single reservoir and station. The station's generation capacity is about 600 MW and the reservoir can hold up to 400 GWh of potential generation.

All of the thermal stations are in the North Island. The northernmost is the Marsden station which runs on residual fuel oil (from the Marsden gas to gasoline plant), and has a generation capacity of 100 MW. The Otahuhu station (in Auckland) runs on distillate oil and has a generation capacity of 100 MW. The Huntly station can run on a mixture of both coal and natural gas, and has a generation capacity of 1000 MW. The New Plymouth and Stratford stations both run on natural gas and have generation capacities of 600 MW and 200 MW, respectively. Finally, the Whirinaki thermal station runs on distillate oil and has a generation capacity of 200 MW. The cheapest fuel is Maui Gas which supplies the New Plymouth, Stratford, and Huntly thermal stations via a single pipeline. The next cheapest fuel is coal, then residual fuel oil, with the most expensive being distillate oil.

Auxiliary stations are used to model run-of-the-river hydro stations, small isolated hydro stations with little major effect on the system, as well as the geothermal stations near Taupo. Waihapa Gas, burned at the Stratford thermal station, may also be modelled in this way, as it is otherwise flared. There is also some fixed coal generation at the Huntly thermal plant, which is assumed to act as an auxiliary station.

The National Grid is a large complex system running AC lines of various voltage levels between major locations, and a high voltage DC link connecting the two Islands. As most of the *hydro generating capacity* is in the South Island and most of the *power use* occurs in the North Island, the DC link is crucial to the operation of the system.

The major differences between the New Zealand system and other electricity generation systems in other parts of the world are the high reliance on hydro generation (with a total hydro storage capacity of only about three months' generation), and the great unpredictability of the inflows into this hydro system. These differences imply the need to design our own specialized model, or heavily tailor another model to our needs, this is expanded upon in Chapter 2, together with discussion on models of other systems.

As well as the major components of the New Zealand system already mentioned there are additional constraints and properties of the system; for instance, the available Maui Gas supply, used by three of the thermal stations, has an upper limit on usage over any week. Such additions are dealt with separately from the overall model, in Chapter 7, so that they do not obfuscate the construction of the overall model. This is because they are of relatively lesser importance, and not central to the model being developed.

In this thesis we seek to model the entire system to a level of detail allowable by a time horizon of the order of one year. It is important at this stage to settle on the order of the time horizon of the model as, under differing time horizons, different areas of the system become more important and particular stochastic elements have different levels of effect. This is, at least in part, due to the computational complexity involved in implementing a model of the system as a whole. For example, if we are solving a model with a time horizon of a day, the lag time of water travelling in various river chains from one station to the next becomes important, whereas, if we have a time horizon of about 30 years, the actual entire river flows for each week are of lesser import.

1.2 Objectives of the Model

In attempting to model the New Zealand system in this way, we seek to satisfy three conflicting objectives. The first is to provide enough detail of the physical system (we would *like* to model every station individually and the transmission network exactly); the second is to provide a good account of the stochastic elements, so as to perform well in an uncertain environment; the third is the ability to obtain numerical solutions in reasonable time. It seems unlikely that there is one way of satisfying all three of these objectives to a high degree—an informal *conservation of effort* law seems to apply. One could draw a diagram of the relative positions of models with respect to these three aspects, and determine how close each of the models is to the mythical centre of this diagram (which simultaneously satisfies all three objectives to a high degree). Unfortunately the exact placement of a model on this diagram is dependent on the relative importance one puts on the three aspects, which is not only a subjective decision, but also dependent on the aims of the model's user.

The system is currently modelled by ECNZ in a framework appropriate for the use of Stochastic Dynamic Programming (SDP) as a solution technique. This approach tends to emphasize the stochastic elements since, due to the aptly named *curse of dimensionality*, it can cope with only a very limited model of the physical system. The motivation behind this thesis is to "attack the problem from the other side" and design a model which provides considerable detail for the physical system and, consequently, has less detail for the stochastic elements.

Some aspects of the system are more difficult to usefully model to a particular level of detail than their potential effect on system operation would suggest is worthwhile; usually this is due to the computational complexity added by such an approach. Instead we must choose an approximation which embodies the essential character of the particular aspect whilst being implementably achievable. Isolating such aspects is often a difficult task and one must be guided by previous experience and the intuition of those who actually run the system.

1.3 Uncertain Supply

It is necessary, in practice, for maintenance and repairs to be carried out on thermal and hydro stations. Some of this maintenance can be scheduled beforehand and taken into account by removing that station from the model for a corresponding period of time. However, maintenance may take place over a time period which is too short to be modelled effectively, or at a future time which is not known exactly when the model is run. We must also take into consideration breakdowns and other outages which cannot be scheduled, and any uncertainty in the fuel supply.

Unscheduled down-time can be modelled as random outages, sometimes called forced outages. For each station, we specify a probability distribution signifying the probable maximum generation by that station. These probability distributions will be only discrete in nature since a generation turbine can be either up (and so able to generate at full capacity) or down (and so not able to generate at all).

1.4 Hydro Detail

The detail sought by ECNZ is mainly in the form of information about individually modelled hydro reservoirs, since it is the presence of these hydro reservoirs, and the uncertainty involved in their water supply, which makes it so important to have a model with the time horizon considered in our model. The current (SDP) model incorporates only two hydro reservoirs, one for each island; a paper investigating the possible approaches of a new model, Lermit *et al.* [9], recommended six separate hydro reservoirs as being a desired level of approximation for the New Zealand system. It is the intention of the model developed in this thesis to model six spatially distributed hydro reservoirs.

When moving from two hydro reservoirs to six, one must also make appropriate changes to the coarseness with which the rest of the system is modelled—modelling different aspects of the system at vastly differing levels of accuracy can cause unwanted additional structure within the solutions which is an artifact of the model rather than the system, and which could be avoided by using a more evenly approximated model. For this reason we seek to distribute *all* stations spatially and hence model transmission of power from the stations to the power users.

In seeking to model six hydro reservoirs individually we need to determine an appropriate level of detail for them. The scheduling of the hydro stations is the most difficult, and most important, job of the model; without hydro stations we would only need to optimize the scheduling of the system over a day or week, as no other part of the model has as much effect on the decisions made later in the year.

One stipulation by ECNZ for the model was that it did *not* model river chains using a "spill past" model, as such a model would add complexity to the system which would swamp the rest of the model and hence affect the level of detail obtained. The time delay of the water travelling from the hydro reservoir, and between stations, means that, when a spill past model is used, the generation at each hydro station must reflect this delay and include the possibility of stations further up the river spilling so as to allow generation by lower (possibly higher capacity) stations at some future, higher load, time. These hydro systems are complicated enough to model, to such a level of detail, in isolation without the added complexity of requiring that they, together with the thermal stations, meet the load requirements of the system.

In modelling hydro stations individually, with or without a spill past model, one encounters many more difficulties. Hydro reservoirs control the flow of water down river chains, each of which may contain many hydro stations. The situation is compounded by having multiple reservoirs and river chains interconnected by controlled canals. Also, many hydro stations have "forbidden generation zones", which define minimum generation levels for each turbine of a station if that turbine is to be used. These also occur in thermal stations but are compounded in hydro stations by the fact that the water flow through many hydro stations is controlled by the same reservoir.

In an effort to avoid these difficulties, we make a quite heavy-handed approximation of the hydro systems: each river chain is represented by a single hydro reservoir and station. This level of approximation may seem unreasonably coarse; however, this model is expected to provide generation information for more detailed models of the individual hydro systems, which in turn return "local" solution information with which we can fine-tune the hydro system approximation in the longer time horizon model (see Section 1.6).

Water enters the hydro systems via inflows, i.e. streams and runoff either entering the reservoir, or entering the river chain downstream from the reservoir. Inflows entering the reservoir can be stored or released for generation and so are known as controlled inflows; the other inflows are known as tributary or uncontrolled inflows.

In our single station approximation of the hydro system we assume the controlled inflows enter the reservoir and the uncontrolled inflows enter the system between the reservoir and the station (see Figure 1.1).

The capacity of the new amalgamated station is taken to be the sum of the capacities of each station in the chain. Each station in the chain is assigned a factor to represent the fraction of water released from the reservoir which can be used by that station; another factor gives the fraction of the total tributary inflow which flows through that station. There is also a conversion factor (or function) for each station which gives its efficiency in converting water to electricity. These conversion factors are combined, in terms of the fraction of water passing through each station, to yield an overall generation efficiency for the amalgamated station.

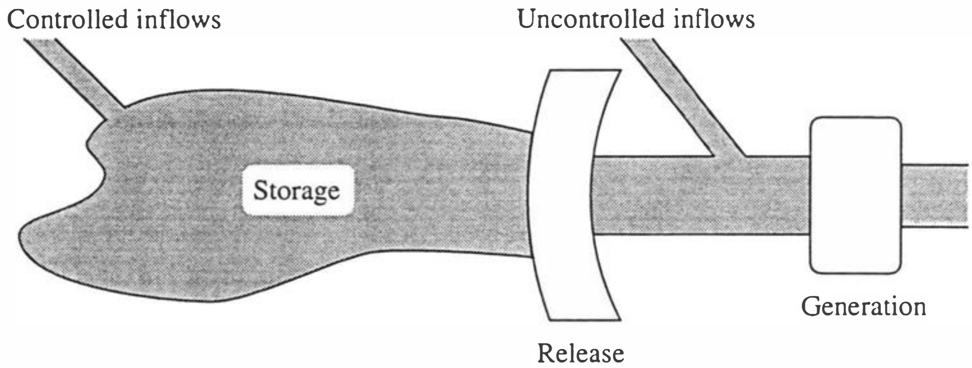


Figure 1.1: Amalgamated hydro system and representative inflows

1.5 Transmission Detail

As mentioned earlier, the transmission network (National Grid) is extremely complex: transmission characteristics depend on the load and generation at various points in the system; the network is mainly AC, of differing voltages, with a high voltage DC link used to get power across Cook Strait; the National Grid has 13 000 km of transmission cable and 180 substations and switching yards; the line capacities and power loss within the grid vary with the load and generation; also, the use of AC power means there is power loss from both resistance and reactance.

A detailed model of the transmission network is not required by our model, as we are interested in the generation schedule (more specifically hydro station generation) rather than an accurate generation-transmission schedule. For our purposes the transmission network needs to be represented by only the major lines between power stations and power users, and the losses and capacities need be modelled only simply. It is intended that, as with the hydro station generation, our model will interact with a more detailed model of the transmission network so as to use "local" solution information about the transmission network to update our capacity and loss models.

1.6 Putting it in Perspective

In actuality there is an hierarchical structure linking all of the models used. The model developed here (which shall be referred to as "our model" to avoid confusion) interacts with other models within this structure, gaining local information from highly detailed short-term models and value or volume information past its planning horizon from long-term models. Our model also provides similar information for the other models.

The reason for the hierarchical structure is that different models provide different information for different purposes; this structure also embodies an informal decomposition of the system into manageable pieces which portray, in some way, the manner in which we intend to operate the system. Each model finds a solution which is locally optimal in terms of the constraints *implied* by the information given to it from other models, as well as those *explicitly* defined.

Our model sets generation or storage targets for hydro stations, taking into account the long term needs for water. Short-term models determine the actual schedule for a river chain for the coming hour (and indeed a real-time model can be used to determine the actual generation at the time). A long-term model is concerned with developing future resources and predicting the needs of the system as a whole into the future.

The presence of the hierarchical structure defines the environment in which our model exists. In light of this we must consider the ways in which our model interacts with other models within this structure. This is important in terms of the information exchanged, which depends on our model and on the other models within this structure; this is discussed further in Chapter 6.

1.7 An Implementable Model

In developing a model, implementation and formulation are inextricably linked. A model which models every aspect of the system to a high and desirable level is all but useless if it can be solved only by brute-force methods which do not converge in a reasonable time, or even at all! Similarly, a model which has been designed specifically for solution via an elegant solution method, but which does not adequately represent the system, is also effectively useless.

Therefore, in developing a model, careful thought needs to be given to the solution method as well as ensuring that the formulation represents the system well. This does not mean, however, that one must decide on a fixed solution strategy; it is better to consider many options in both the formulation and implementation phases which can later be tested to decide upon the best combination for the required task.

Generally a method will require tailoring to solve the model developed smoothly and efficiently.

One of the most significant partitions in developing formulations is that of linear and non-linear models. The advantages of a linear model are that there are many solution methodologies available for them and that one can solve larger linear models than non-linear ones. Another significant partition is that of convex and non-convex models. Results abound on the properties of solutions to convex problems, however in the non-convex case one cannot be sure whether the solution obtained is indeed globally optimal or not.

It is important that linearity and convexity be achieved, where possible, by transformation, rather than approximation, of the formulation. However, if the non-linearity or non-convexity is slight (in some sense) it would probably not change the model much by making appropriate approximations. In such cases one needs to consider the consequences of such action carefully, although often such approximations are all but necessary to allow tractable solution.

Chapter 2

Past Solutions

In this Chapter we do not present an exhaustive survey of all related literature. We only highlight those papers which are important in model development or which illustrate an important modelling or solution technique. The reason for this is that we are modelling a specific system and so few reported models are of a direct relevance. We seek to highlight literature which is relevant to the modelling of the New Zealand system, or which contrasts with features of the New Zealand system, leading to important observations about the structure of the model to be developed. This is most effectively achieved without the unnecessary clutter an exhaustive survey would inevitably create.

The New Zealand system has unique attributes, including: the mix of hydro and thermal generation, the lack of any import or export of electricity, the unpredictability and high variance of the inflows, and the relatively small total storage capacity. The effect of this is that none of the models presented in the literature for other systems can be *directly* used in New Zealand. The difficulties inherent in the New Zealand system mean that, even when converted to allow for the New Zealand conditions, the models and algorithms presented in the literature are challenged, computationally, by the New Zealand system.

This provided the initial motivation for our approach. We seek to develop a model which accurately portrays the New Zealand system, highlighting those aspects which are seen as important to the system, in terms of the structure imposed upon solutions. Because of this we do not intend to *adapt* an existing model to the New Zealand system (such an investigation is currently being carried out at ECNZ), but to create one from scratch. In doing so it is *important* to investigate other approaches, both to solving the New Zealand system and other systems, which have been taken.

When modelling such a complex system one must inevitably make simplifications and approximations. Often these depend on the system being modelled, whereby simplifications which are reasonable for one system may be unreasonable for another. For this reason it was decided that the model being developed should "aim for the stars", i.e. we want to describe the system in as much generality as practicable in the model, and *then* make any simplifications necessary to allow the model to be computationally feasible to solve. This should have a two-fold effect; firstly, it will go a long way towards defining exactly *what* simplifications are made, and, secondly, it will make it easier to expand the model in light of advances in modelling and computer solution techniques. It is realized that not all simplifications and approximations can be left to be applied when the model is completed; some are fundamental to the view of the system (such as those made in Chapter 1), and others are needed to define the structure within which we model other aspects. For this reason, in investigating other approaches to similar problems, we highlight the inherent simplifications as well as the novel modelling techniques.

Many of the papers to be discussed present algorithms as well as formulations. As noted in Section 1.7, in the development of a model both the formulation and method of solution must be addressed. Formulations are often developed in a particular way so that special structure may be exploited in the solution method, or to provide an illustrative example of a solution technique. We do not intend to examine this trade-off here, since the models developed are for many different systems, each with different attributes and different aspects of importance. A discussion considering the trade-off only in terms of the New Zealand system would obviously be highly biased towards models developed specifically for the New Zealand setting.

None of the approaches that will be examined here take explicit account of the geographic location of power stations and load; it appears the main reason for this involves the difficulties that such structure evokes.

2.1 Maximizing Generation

Various models have been developed for systems in which hydro generation is not crucial, or is a beneficial side effect of scheduling hydro releases for other purposes, such as irrigation. In these situations the emphasis is not on the cost of power generation but on maximizing the amount of hydro generation that can be coaxed from the system. This differs from the situation where the emphasis is on the *cost* of alternative generation or in explicitly attempting to meet the load requirements of the system.

Both Ikura, Gross, and Hall [8] and Soliman and Christensen [21] describe systems where the emphasis lies entirely on the accurate modelling of the hydro systems involved. For the New Zealand system, such accurate hydro system models are within the realms of the short term modelling (with a planning horizon of about a week). Due to the time scales used in both [8] and [21], however, there are modelling techniques which can be exploited in our model.

The lack of explicit modelling of thermal station generation in these systems makes most of the modelling techniques irrelevant to the New Zealand situation, since incorporation of these aspects would tend to make any model so developed computationally infeasible. Also, in the New Zealand system, the load levels are very important to the running of the system, as it is the load to be met which determines the cost of generation. The maximization of hydro generation is not an adequate substitute for this. The major modelling technique that appears to be most useful for incorporation in a model of the New Zealand system is the representation of hydro systems as networks, using bounds on the arcs to represent constraints on river flows, generation and storage. Storage is represented by temporal arcs which represent the volume of water carried from one time step to the next.

2.2 Purely Deterministic

For many generation systems the hydro component is either relatively minor, overly constrained by external limitations, or reasonably predictable in its inflows. In such situations the use of a purely deterministic model, with the possibility of considering a few inflow scenarios, is adequate. This allows for a very detailed model of the physical system and is especially useful when the interactions within the system have more of an effect on solution structure than any stochastic elements. It also allows explicit expression of the effect of reservoir levels and turbine flow rate on the generating efficiency of hydro stations and other non-linear dependencies of the system.

Due to large inflows which occur in Spring for the system modelled in [8], it is more concerned with minimizing the adverse effects of too much water, which are not always quantifiable. For this reason they use a deterministic model and evaluate the effects of various scenarios manually. In New Zealand it tends to be the long term *lack* of water that is of most concern.

Lyra and Tavares [11] and Rosenthal [20] both use deterministic approaches in which the cost of thermal generation is an explicit, and fixed, function of the load not met by hydro generation; this appears to be a very "Engineering" type of approach. The advantage here is that one can model quite complex functions of the efficiency of various thermal stations, and so it appears that, in these situations, this thermal efficiency has more of an effect on the structure of solution than any stochastic effects. In this case many stochastic aspects could adequately be evaluated by comparing the solutions for a few important scenarios.

This approach has also been taken in Nabona [13], where the uncertainty of inflows *have* been taken into account. The implicit assumption that inflows are totally correlated in time, and the small number of different inflow sequences investigated for a few river systems, mean this approach has most of the advantages and disadvantages of a purely deterministic approach.

Boshier and Lermit [1] use a deterministic approach to the New Zealand system. The hydro reservoirs are amalgamated into catchment areas, and hydro generation is assumed to depend linearly on the volume of water released. Similar assumptions about totally reliable thermal generation transform the formulation into a linear Network, which is especially useful. A single transmission line (the DC link between the North and South Islands) is modelled explicitly, but *no* transmission losses are applied. Again the stochastic aspects could be taken into account by moving to a scenario type approach.

Unfortunately the unpredictability, high variance, and lack of spatial correlation in New Zealand's hydro reservoir inflows make an approach which ignores the stochastic aspects, or evaluates the effects of such manually, unacceptable. It is, also, important to have an accurate model of the thermal generation as, in our model, we seek to include information on the geographic distribution of stations and load; this would completely eliminate the usefulness of modelling the cost of thermal generation explicitly in terms of load not met by hydro generation.

2.3 Stochastic Aspects

There are various methods for stochastic modelling and solution methodologies which account for stochastic elements explicitly. In the New Zealand system, the uncertain aspects with the most effect on system operation are the inflows into various reservoirs. For this reason we separate our discussion on stochastic aspects into two parts: those concerned with hydro reservoir inflows, and all other aspects of the problem. The discussion on the exact method of accounting for inflows is left until Chapter 6. For now we accept the necessity of modelling these stochastic elements, but make no judgments on how such aspects will be modelled. Instead we concentrate on the physical system being modelled, explicitly including the stochasticity inherent in load and thermal supply.

There are many ways in which to incorporate the uncertainty in future loads; one way is to use the *average* load for each week. The advantage of this approach is that the average load can often be forecast with reasonable certainty and it also simplifies the model with respect to thermal generation. When one takes into account the fact that releases and inflows into the hydro system are often only specified as totals over a week, this is not such an oversimplification.

Li, Yan and Zhou [10] and [11] use such an approximation for load. However, [10] does try to take account of peak loads (and forced outages) by derating station capacities. This approximation does not allow for the fact that peak load is apparent for only part of the week, and only particular stations (which may not be known in advance) are available to meet peak load. It also takes only *above* average load into account, not below average load; this may mean that certain stations do not generate at peak efficiency, and is liable to ruin any advantage gained by explicitly using generation efficiency.

2.4 Load Duration Curves

The use of average load comes about because of the difficulty in approximating and forecasting load. Most models divide the time horizon into discrete time steps whose duration is a day, week, or month. Load is often characterized by two major peaks in each day, as is the case in New Zealand. When forecasting load the uncertainty is not only in the height of the peaks but also in their times of occurrence. The degree of certainty with which load is forecast can be increased a little by moving to a slightly different representation—rather than considering load as a direct function of time (as in a Load Curve representation), one can use a curve which gives the fraction of time each load level occurs, i.e. a Load Duration Curve (LDC) (Electric Power Research Institute [6]). Figure 2.1 gives an example of a Load Curve over a day, and its corresponding LDC. In using an LDC one removes the uncertainty of exactly when peaks occur as well as information on the difference in height of the two peaks, so LDC's can be forecast with more accuracy than Load Curves.

Most approaches use a single LDC to represent all of the load, e.g. Boshier, Manning and Read [2], Dembo *et al.* [4] and [20]. In this case, the generation pattern can be determined by finding the actual load, at a particular time, on the curve and reading off the corresponding generation. Of course with such an approximation one cannot specify start-up costs, as it is unclear how often the station will be turned on and off. This is actually not such a burden as start-up costs usually require integer variables to model them and so either become part of the station's efficiency curve or are left out of models of this scale entirely (and are instead modelled in shorter time horizon models).

Another feature of the system which lessens the need to model start-up costs accurately, is that not all small stations are modelled explicitly. There are various stations which are modelled as auxilary stations, for convience and computational tractibility, but which have more control over their generation than this would imply; these stations can be used to smooth over start-up periods, and handle discrepancies between the forecast and actual load.

These problems can be partially side-stepped by also specifying an approximate "unsorting" of the LDC; this is done in Pereira and Pinto [14] (also [18]), where discrete Load Curves are used. However, the time dimension is partitioned into, possibly, unconnected regions over which the load is constant, which effectively models the load as a discrete LDC with a specified "unsorting" to allow some intra-week constraints to be applied.

In explicitly modelling the geographic distribution of load, there is another dilemma to be faced. If all load, for one time step, is treated as a set of LDC's, once the generation schedule has been determined it is difficult to determine the actual generation of each plant, since we have an implicit assumption that all load is coherent (peaks occur at the same time); if a peak occurs at one location before it occurs elsewhere, the exact generation schedule to use is unclear. In reality, there are shorter time-horizon models which are used to determine the *actual* generation schedule used. The model being developed is used only to determine the mix of hydro and thermal generation that should be used during the week, taking into account the need for hydro over the year. The load is included explicitly to ensure feasibility of the generation schedule, and *not* to provide explicit generation timing.

Given a Load Curve (forecast or from past data), the corresponding Load Duration Curve can be calculated by sorting the Load Curve from highest to lowest load. In theory one could determine the exact generation schedule by unsorting the load and generation given by the model; however, there will be the same uncertainty in the generation as for the forecast Load Curve. Furthermore, generating an LDC from a forecast Load Curve is not as robust a method as forecasting the LDC directly, in which case there is no *specific* underlying Load Curve.

One of the advantages of using LDC's, which is exploited to include the uncertainty in supply, is that the inverse of an LDC is a probability distribution function (see Figure 2.2). This function gives the probability that the *forecast* load is above each power level. Because of this we can include uncertainty in the forecast LDC by changing the probability function used to be the *total* probability that the load is above each power level, i.e. explicitly including the uncertainty of the forecast. The LDC used is then the inverse of this probability distribution function, for which there is *no underlying Load Curve.*

2.5 Filling an LDC using Thermal Stations

Given an LDC, we can determine the generation schedule of a given set of power stations by determining the load each station must meet; this is known as *filling* the Load Duration Curve. For a totally reliable, purely thermal system, this is achieved by scheduling stations in increasing order of cost ([6]). Each station is scheduled to generate at its peak capacity or at the remaining load level, whichever is lowest; this generation is then removed from the load and the next station is similarly scheduled. This leads to generation schedules similar to that shown in Figure 2.3.

When scheduling thermal stations one would like to take account of possible forced outages (see Section 1.3). This can be done by re-solving the model, for each state of each station, to determine the expected cost of generation. It should be

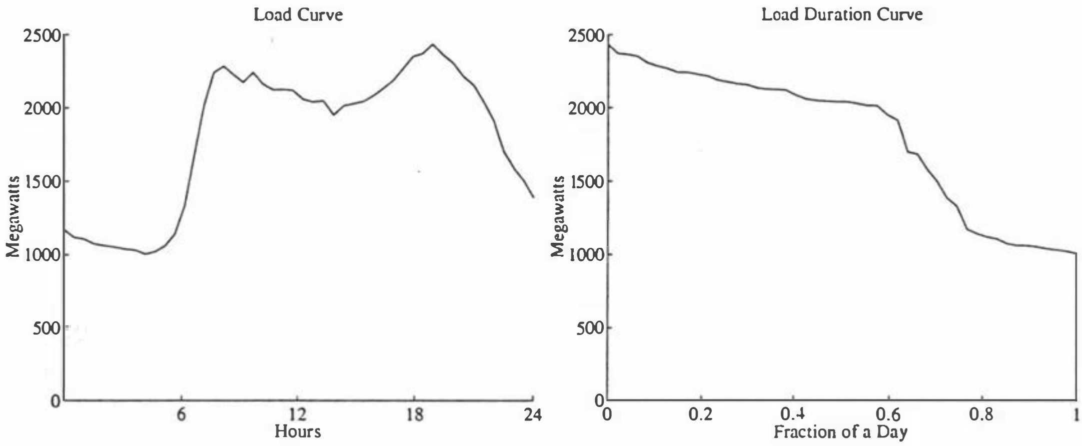


Figure 2.1: A Load Curve and corresponding Load Duration Curve

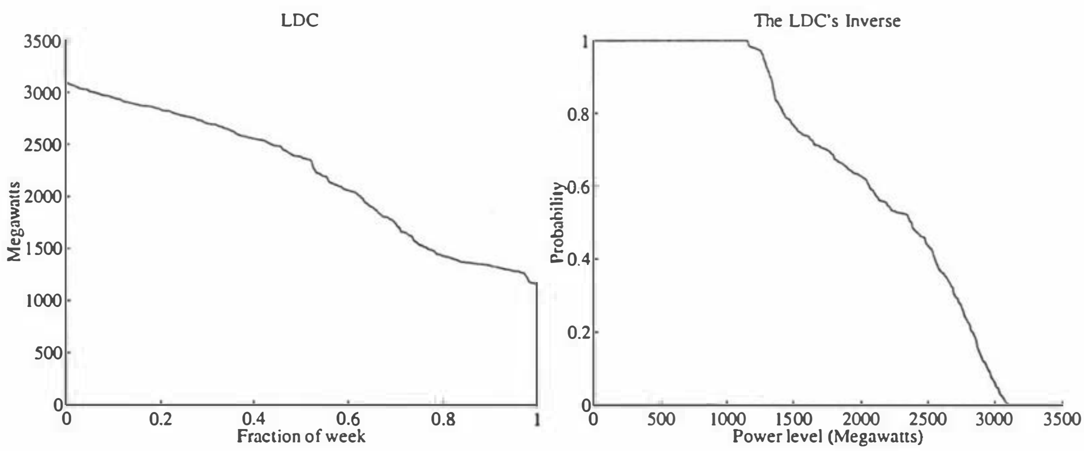


Figure 2.2: An LDC and its inverse, a probability distribution function

obvious that such a task becomes extremely time-consuming for even a moderate number of stations, as the total number of states is exponential in the number of stations. Luckily, when using an LDC to represent load this can be achieved with much less effort. The convolution of the probability distribution function representing the probability that the station can meet various load levels with the probability distribution function associated with the LDC gives a probability distribution function representing the probability that the remaining load is above various levels. Subsequent stations fill the remaining load in a similar manner (see [6]).

An example of an LDC filled by such a method is used is given in Figure 2.4. It must be remembered, however, that this schedule is derived *only* for the purpose of determining the expected cost of generation. For the *actual* schedule of generation, one must schedule all currently operational stations, at their current capacity, *as if they were totally reliable.*

When one attempts to include information on the geographical distribution of load and stations, it becomes necessary to calculate the expected cost by explicitly re-solving the model for every state of the stations, since a change in the geographic distribution of power available may also change the optimal distribution of generation.

2.6 Hydro Stations Filling LDC

For hydro stations one schedules their generation in one of two dual ways: given the dual cost associated with the hydro station, it can be scheduled as if it were a thermal station with this dual cost as the cost of generation; alternatively, given the volume of water released (in terms of average potential generation), the station can split the LDC, removing a section with height equal to its generating capacity and area equal to the potential generation of the release (this is shown in Figure 2.5). Station generation efficiencies can be modelled as functions of average potential generation, and the effect on the generating efficiency of the reservoir level can be modelled as a (non-linear) function of both average potential generation and storage.

The difficulty involved in specifying the dual price of a hydro station is that often this price is exactly the same as the fuel cost of one of the thermal stations,

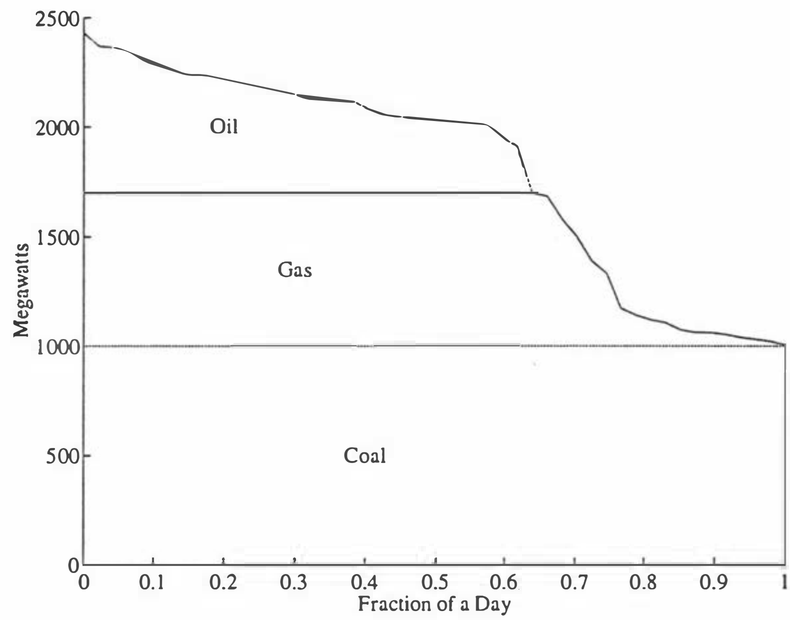


Figure 2.3: Totally reliable stations filling an LDC

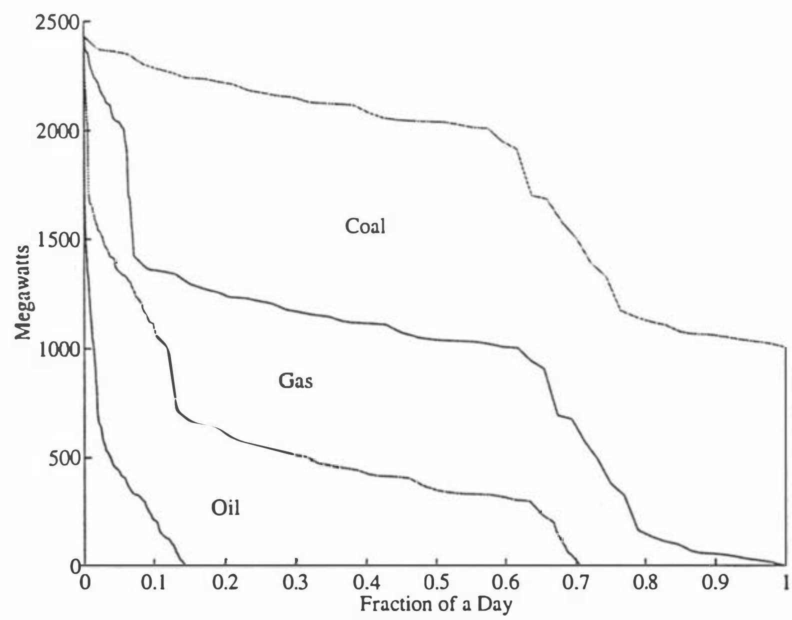


Figure 2.4: Unreliable stations filling an LDC

namely the station whose generation is effectively split (see Figure 2.6). In this situation, it is unclear as to which station should be scheduled first. Often a rule of thumb that the thermal is scheduled first is used, as, since both stations cost the same and hydro inflows are uncertain, it is better to act on the side of caution. In actuality it is probably the case that scheduling either of the stations first does *not* yield the required generation. The problem in this case is that, if the hydro is scheduled first, more water will be used than desired and so later, the order will be reversed (and *vice versa*), often causing an oscillation in which the hydro station and thermal station alternate in being run at full capacity and turned off.

The difficulty in scheduling hydro stations in a primal manner (given the release) is that one needs to determine exactly *where* the station splits the curve—that is the δ shown in Figure 2.5. As with thermal stations one can include hydro station uncertainty here; in the case of the primal method, the probability distribution function corresponding to load is shifted by δ before convolution, and in the case of the dual method, one must *also* schedule all of the stations as if they were totally reliable in order to determine the actual release.

2.7 Geographical Distribution of Power

None of the models in the literature deal explicitly with the geographic distribution of load and stations. Some of the models do include some transmission constraints; in general, these are in the form of capacities (and possibly losses) between the station and the "pooled" load (see [18] and [10]). Models of the New Zealand system explicitly include only the North-South DC link, due to its importance to the system ([1]): this effectively partitions stations into two sets, those in the North Island and those in the South Island.

The reasons for these omissions seem to be the difficulties in implementing such an approach, and, at least in New Zealand, the actual freedom within the transmission system in terms of capacity of lines. However, it is not just the capacity of lines that has an effect; there is power loss in the transmission system and, most importantly, the actual *distribution* of power. When considering the distribution of power generation over multiple hydro reservoirs, it becomes important to also consider the distribution of this power geographically. This geographic distribution of power will be dealt with explicitly in the model developed.

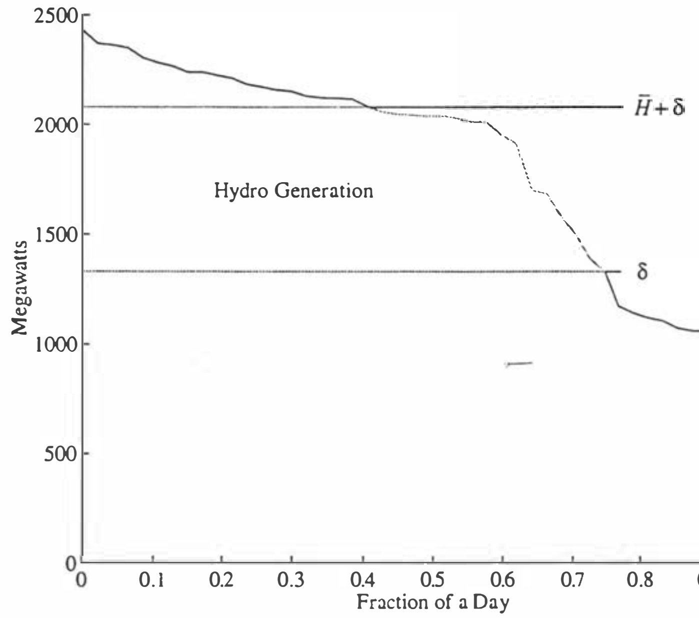


Figure 2.5: Hydro station splitting an LDC

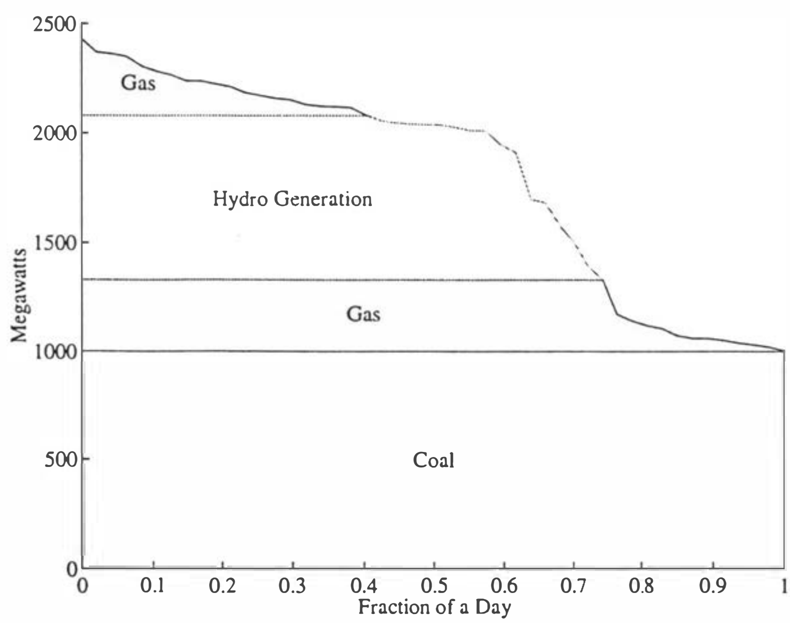


Figure 2.6: Hydro station splitting a thermal station's generation

2.8 How Many Hydro Reservoirs?

The number of hydro reservoirs modelled by the various methods varies wildly. For instance, there are 37 reservoirs modelled in Pereira and Pinto [15] but only two separate reservoirs modelled in [2]. The reason behind this variation involves the effects of the stochastic elements on the system—in a deterministic model there is effectively no limit on the number of reservoirs which can be modelled, but, for a stochastic model, this limit is very dependent on the correlations, predictability and variance in inflows between reservoirs and over time.

The New Zealand system has six or seven important separate catchment areas; however, to date, the system has only been able to be modelled effectively using two separate catchment areas, because of the unpredictable inflow patterns, the small storage capacity (in terms of total yearly generation), and the fraction of hydro generation in the total generation ([2]). All of these factors mean that the stochastic aspects can not be smoothed out of the system operation in any way, i.e. the system must be run in response to the stochastic aspects as they occur.

Other systems with a larger volume of hydro storage can effectively smooth out the variance in inflows with this storage, and the average predicted inflows are useful in scheduling the system. In New Zealand it is the actual inflows which are of prime importance, making the task of scheduling generation quite difficult.

Our model attempts to model the New Zealand hydro system using six separate reservoirs. It is hoped that the information gained about the running of the geographically distributed physical system goes some way towards balancing the inevitable loss of stochastic information required to make a computationally feasible model.

Chapter 3

Desirable Features of the Model

This chapter describes the aspects of the problem seen as important to be well modelled, and features which the model should possess. This includes both features which are modelled well by other models (as described in Chapter 2), and features important to the New Zealand system which have not yet been modelled well elsewhere. The goals are to give all of the important aspects a consistent level of detail, to seek to isolate these aspects, and to provide approximations which give an appropriate level of detail.

The intention is to propose elegant approximations and approaches to be incorporated into the model. However it is not always possible to find an elegant approach for every aspect of the model and therefore in this case we settle for a *good* approximation. Of course all approximations, both elegant and otherwise, need to be modelled *well* in terms of implementability (including solution time) and *closeness* to reality.

The first part of this chapter focuses on a "first step" model in which the details of the approximation of the load and generation are not fixed; the actual approximation to be used is discussed in the second part. The reason for this is that the choice of our approximation depends on the interaction of load and generation with, and within, the system.

3.1 A Flexible Model

As mentioned in Section 1.2, our model attempts to satisfy three conflicting objectives: detail of the physical system, effective planning for an uncertain future, and computational efficiency. However, we do not seek to explicitly define the best way to achieve this, nor do we intend to rank the three; such decisions lie ultimately with the end-users of the model and it should be able to survive a *re-prioritorization* of the goals of these users. To accommodate these goals, the model should be made as flexible as practicable. This does raise another difficulty, however, as now there is need to investigate the effects of making allowable changes to the model on the quality and structure of the solutions given.

Flexibility also allows the advantage that the same base model can be used for different purposes; for instance, it may be desirable on occasion to make computationally expensive runs incorporating greater detail than usual to estimate the deviation from *optimallity*, and the changes in structure, of the solutions usually obtained.

The flexibility needs to be easy to control, in that the overall structure of the model needs to remain constant. We desire flexibility that allows us to change the level of detail with which we model one aspect of the model, whilst leaving the rest of the model alone. As an example, we do not want to change the model from a linearly constrained problem to one with non-linear constraints, even though this may provide a finer approximation for load. Of course such problems in allowing flexibility may be unavoidable, and the aspects of the problem for which they occur need to be identified.

A related aspect of the flexibility is that we seek flexibility in terms of changes to the model which only affect the (local) structure of the model, whilst leaving the *character* of the solutions unchanged. We seek the model to be robust in terms of the changes to the approximations we may make. Explicitly, if we move to a coarser approximation which could have been thought of as valid under the previous approximation, then the solution to the coarse approximation needs to be a *feasible* solution to the finer approximation, and *no better solutions to the finer approximation should be feasible for the coarse approximation.* That is, we want coarser approximations to be sub-sets of the finer approximations.

3.2 Internal Consistency

In seeking flexibility and elegance we need to be careful that we do not end up with a model which is a collection of different, elegant, ideas tied together only by the fact that they all model different parts of the same problem. Such a model will tend to enforce structure on the solutions not present in reality—structure which reinforces the differences in the modelling of various parts of the system, and which may, in effect, represent differing management styles, or policies, for these parts of the system, which do not occur in practice. For instance, if at some point during the year we change from inflows of a stochastic nature to deterministic ones (in an effort to cut down on the solution time, say), we are implying that after that point in time our knowledge of the future becomes exact, and so we could leave lake levels in a significantly worse condition then than we could possibly have allowed before that time; in terms of management policy, after this point in time, management of lakes becomes less conservative, and the solutions given by the model will reflect this change.

We require the model to be as internally consistent as practicable. It may appear that this is the same thing as requiring the detail of every element of the model to be at a consistent level, but, to allow flexibility in the model, we seek to allow similar parts of the model to be modelled at differing levels of detail and yet with an overall consistent approach. This consistency requires that aspects of the model which are the same in scope be modelled via similar methods. For example, we do not mind if hydro releases are modelled as volumes of water over the entire week while generation is modelled as a function of time, but we do not want thermal generation to be modelled as power output over time while hydro generation is modelled as just total energy output for the week.

3.3 Geographic Distribution

The most important difference between our approach and the SDP approach currently being used by ECNZ, is that our aim is to allow stations and power users to be distributed over different geographic locations. This is best done by introducing some sort of network structure into the model.

We leave describing the actual form the network *structure* will take until Section 3.5. It may very well turn out to be a *Linear Network* in the strict Operational Research sense of the term—however, we do not mean to imply this is the only structure it can have. *Network structure*, in our interpretation, will be taken to mean that we have an underlying structure which can be thought of as an (un)directed graph in which "arcs" which represent possible paths of flow of some commodity and "nodes" represent points at which this flow interacts. In essence we are more interested in the pictorial nature of the network than the strict mathematical structure.

Facilities deemed to be coincident will be at the same node, and transmission lines between these locations will act as arcs. As it is the transmission network which represents the arcs, and thus relative displacement of nodes, it is a misnomer calling this the geographic network; since we want to distinguish between this network and the actual transmission network it is an approximation of, and since different places in terms of the transmission network will be at different geographic locations, we will continue to label it as such.

The desire for a flexible model means we would like the geographic network used to be altered with little difficulty, in terms of the appearance rather than the structure. This means that, when designing the model, we should not fix the network to be used, but think in terms of an arbitrary geographic network. However, for the purposes of implementation and to allow for the creation of a working model to work with, we define the network of Figure 3.1 to be the representative geographic network of the model.

3.4 Time

As described in Section 1.1, the time horizon this model should cater for is of the order of a year, but this needs to be allowed to be flexible. Different time horizons can mean that different aspects of the problem become more, or less, important so we need to keep in mind the intended time horizon for the model.

In choosing a fixed time horizon one needs to take into account the structure in the various forms of data used in the model, especially any periodicity in this data. Load curves have recurring patterns each day and week (as well as seasonal effects), while lake and inflow levels express seasonal patterns; these patterns mean that our model will need to have a time horizon which is an integral number of weeks, and, in fact, to account for the seasonal effects, an integral number of years. The problem of a calendar year not being an integral number of weeks is considered later in this section.

It would be nice to be able to represent the time dimension of this model

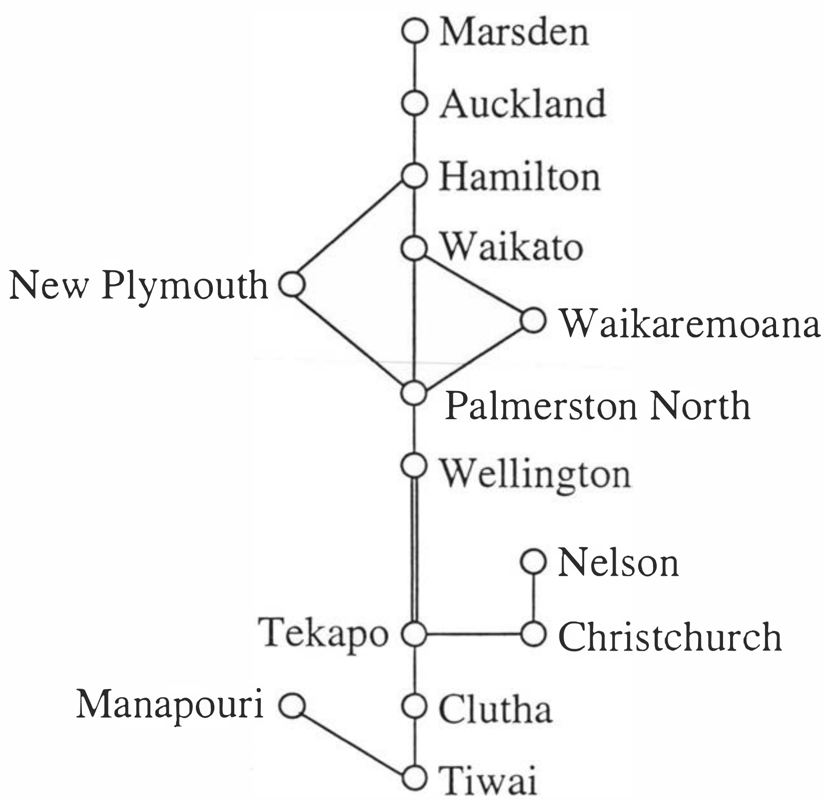


Figure 3.1: The representative geographic network

continuously. Comparing the two graphs of Figure 3.2, which show the entire load for the North Island for one year and for a single day, it seems that trying this type of approach will inevitably lose a lot of important information *and* require horrendously large approximations (there are more than 720 local maxima for the curve showing the entire year). On the other hand, splitting time up into smaller portions seems to be trying to impose structure on the model which is not there in practice. However, there does not seem to be any reasonable alternative.

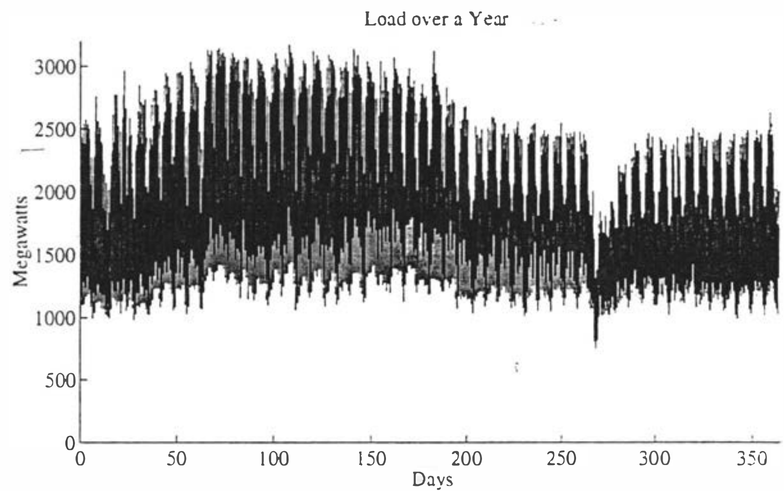
Therefore, we choose to split the time horizon up into smaller time steps. As with the time horizon itself, we need to base the length of these time steps on the periodicity of the various problem data. Flexibility requires that we do not fix the length of the time steps outright; as we require a fixed case for our working model, we choose a weekly time step for this. This choice is reinforced slightly by the fact that, for the current model used by ECNZ, Taupo, the major North Island hydro, has a storage cycle length of about a month.

The problem that months, seasons and years are not integral numbers of weeks can, in general, be ignored, as the effects of changing the length of these longer periods to become an integral number of weeks should be minimal, and, due to the periodicity of the load, this seems more appropriate than to change the size of a week so as to fit an integral number into a year. Of course one must be careful when using *data* which is taken over periods which are not an integral number of weeks long; as an example of this the inflow data used for this thesis was given as 52 evenly spaced inflow levels for each year.

Since the length of a time step is not fixed we will assume that all time steps are normalized, i.e. have a "length" of one; this effectively sets our unit of time. The time horizon is taken to be Y time steps long, where Y is a positive integer. Irrespective of the length of the time steps, we will henceforth refer to them generically as *weeks* and the to time horizon generically as a *year*.

3.5 Transmission

Before talking about the transmission of power around the system it would seem that we should decide on the form of the power representation. However, this form depends on the interactions we require our power representations to have, which in turn depends on how we are going to model transmission and generation



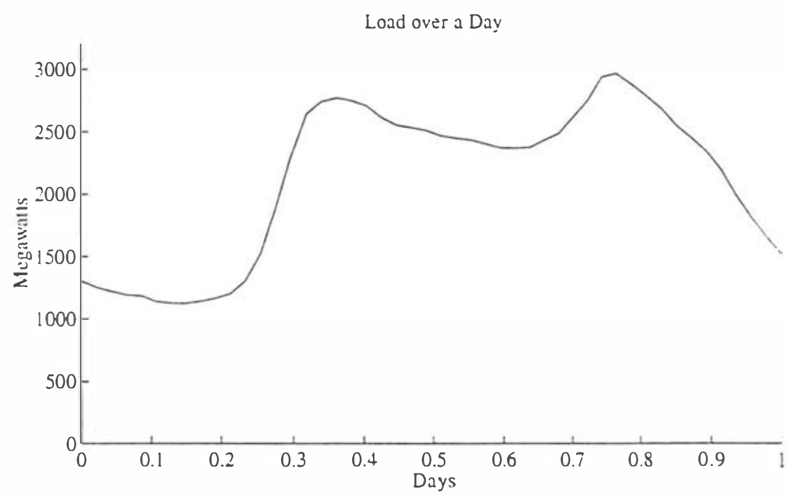


Figure 3.2: The North Island load for a whole year and a single day

and how we approximate load. This Section describes which features we wish the transmission to possess.

As mentioned in Section 1.5, we do not seek to model the transmission network exactly. Also, to deal with the need for consistency, we should not model the characteristics of the transmission lines beyond the level at which we model generation and load. Each arc of the geographic network represents transmission over a part of the National Grid; for each arc we specify a capacity and loss function, representing similar characteristics displayed by *that* part of the National Grid.

If the loss function is chosen to be non-linear (quadratic is a good approximation of loss and stems from well-grounded theory) the model will then have non-linear constraints, be they modelled explicitly as constraints or as a penalty function in the objective. To allow flexibility, we shall consider the possibility of either linear and non-linear losses in the model. However, for the working model, we settle for linear line losses.

The DC link is very important, in terms of the system operation, and it is deemed important, by ECNZ, that the loss structure on this arc is modelled with, possibly, greater accuracy than that of the other lines. Our desire for consistency in approach would seem to require that all lines be modelled in the same way, however the DC link does have a different structure to other lines; it is DC and so has no reactance loss, it is the *only* connection between the South and North Islands, and it is also important to the system in terms of the reliability of supply. For these reasons, and that of flexibility, we allow the loss of the DC link to be modelled non-linearly even when other line losses are being modelled linearly.

If the transmission lines are to have an inherent power loss, one must be careful about the direction of transmission. One can think of negative transmission as being transmission in the opposite direction, but when we apply the loss function to this negative power, the (negative) power seen at the other end of the line needs to be sufficient to ensure that when it enters the line (as positive power transmitted in the opposite direction) the power seen at this end of the line is the original power transmitted negatively in the other direction. Mathematically this means that, for any loss function, f, we require $f(-f(-x)) = x$. However this is not the case for a linear loss, since, if $f(x) = ax$ for some $a \in (0, 1]$, then;

$$f(-f(-x)) = f(-a(-x)) = f(ax) = a^2x \neq x, \quad \text{unless } a = 1$$

Figure 3.3 shows this pictorially.

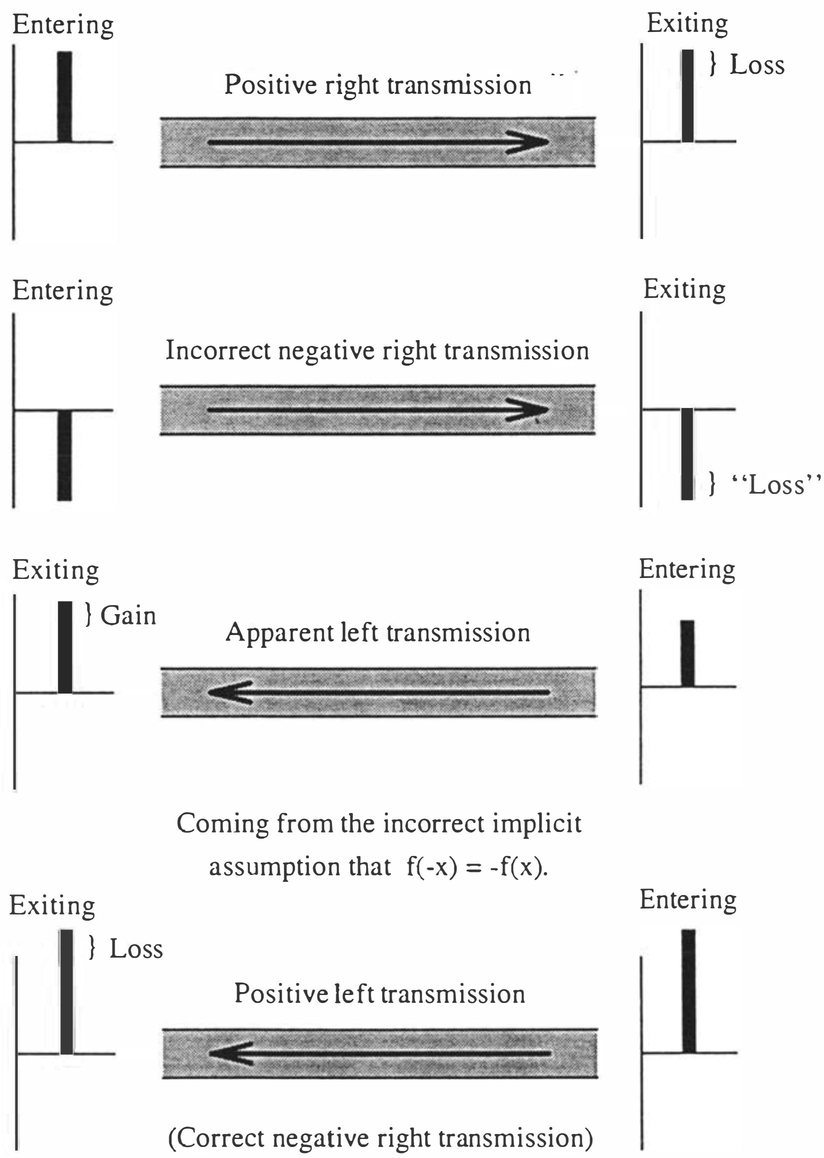


Figure 3.3: The effect of allowing negative power transmission with a linear loss

Hence, for linear loss, we need to split the arcs into two oppositely directed arcs and allow only positive power to be transmitted in each direction. We may also have to apply this arc-splitting for non-linear losses since, in requiring the above condition, the loss function may be non-differentiable at zero. This can be seen from the linear loss function example.

3.6 Load

Load is modelled as occuring at the nodes of the geographic network, represented as a Load Duration Curve over a week, as discussed in Section 2.4. Although this representation removes some local information about the load, having a separate LDC for each week means we do retain some of this information, with the advantage that it is in a form that is more easily approximated well. The exact approximation we use for LDC's and other electricity curves is dealt with in Section 3.10.

By including an underlying network one must address the issue of the interaction between Load Duration Curves, be this direct interaction in which *load* moves around the network or indirect interaction where generation from the same station is used to meet the load at two nodes. In investigating this interaction it should be recalled that the Load Duration Curves are forecasts of load and so inherently include some uncertainty. There are two fundamental forms of interaction which can be considered: interaction in which the Load Duration Curves are considered to be coherent, or independent.

For coherent interaction we are assuming the LDC's can all be specified from one parameter, which is some combination of time and the forecast uncertainty. Therefore, if we know the value of one LDC for a particular parameter value, we can infer *exactly* the value of all other LDC's at the same value. This form of interaction is achieved by adding the appropriate curves.

For independent interaction the assumption is that there in *no* correspondence between the values of one LDC and the values of any other. Knowing the value of one LDC at a particular time gives no information whatsoever about the other LDC's values at that time. This interaction is achieved via a convolution of the directly interacting curves.

In actuality, the LDC's of a particular week are highly correlated and so combining them as if they are coherent is a reasonable approximation. It may appear that the correlation can be increased by using Load Curves rather than Load Duration Curves, but this is not necessarily the case. One of the reasons for using LDC's is the reduced uncertainty in forecasting them relative to forecasting Load Curves (refer to Section 2.4). The coherence which is lost in using Load Duration Curves instead of Load Curves is made up for in the greater certainty with which they can be forecast. So, in terms of uncertainty there is little difference in using Load Curves or LDC's. The approximation advantages therefore make using LDC's the preferred approximation.

For the artificial case where the actual Load Curves are known, Figure 3.4 shows the real (addition of the Load Curves), coherent approximation (addition of the LDC's), and independent approximation (convolution of the LDC's) of the interaction between two LDC's.

It should be noted here that if one wishes to use Load Curves instead of LDC's, the model should allow this by using a finer approximation for the load; since the decreasing nature of each LDC is exploited later, this would require major revision to the model, and so is not investigated directly. However, the tools used in other areas of the model should be sufficient to allow a reasonable approximation to be made to this end.

Having decided on the type of interaction between load at different nodes means that we have settled on the structure underlying the network. The interaction at a node, j say, *without* any stations, is simply

$$\sum_{i \in \text{IN}(j)} f_i(X_i(t)) - \sum_{i \in \text{OUT}(j)} X_i(t) = L_j(t) \quad \forall t \in \text{Week}_w \tag{3.1}$$

where, $\text{IN}(j)$ and $\text{OUT}(j)$ are the set of arcs entering and leaving node j respectively, X_i is transmission as a function of time over week w, and f_i is the loss function for arc i, and L_j is the load as a function of time over week w at node j. In all cases the *time* dimension is in terms of the LDC parameter (which could be thought of as "sorted" time), and *w.l.o.g.* we assume the domain of all LDC's (and hence all related curves) is the interval $[0, 1]$; if this is not the case, an affine transformation can be used to map $\text{Week}_w \mapsto [0, 1]$, mapping the beginning of the week to 0, and the end of the week to 1. Equation 3.1 requires the transmission in and out of the node to exactly meet the load there; for a node with a station, a similar equation will apply but the station generation needs to be included (see below). For a fixed t and linear loss on each arc, Equation 3.1 characterizes the network constraint of

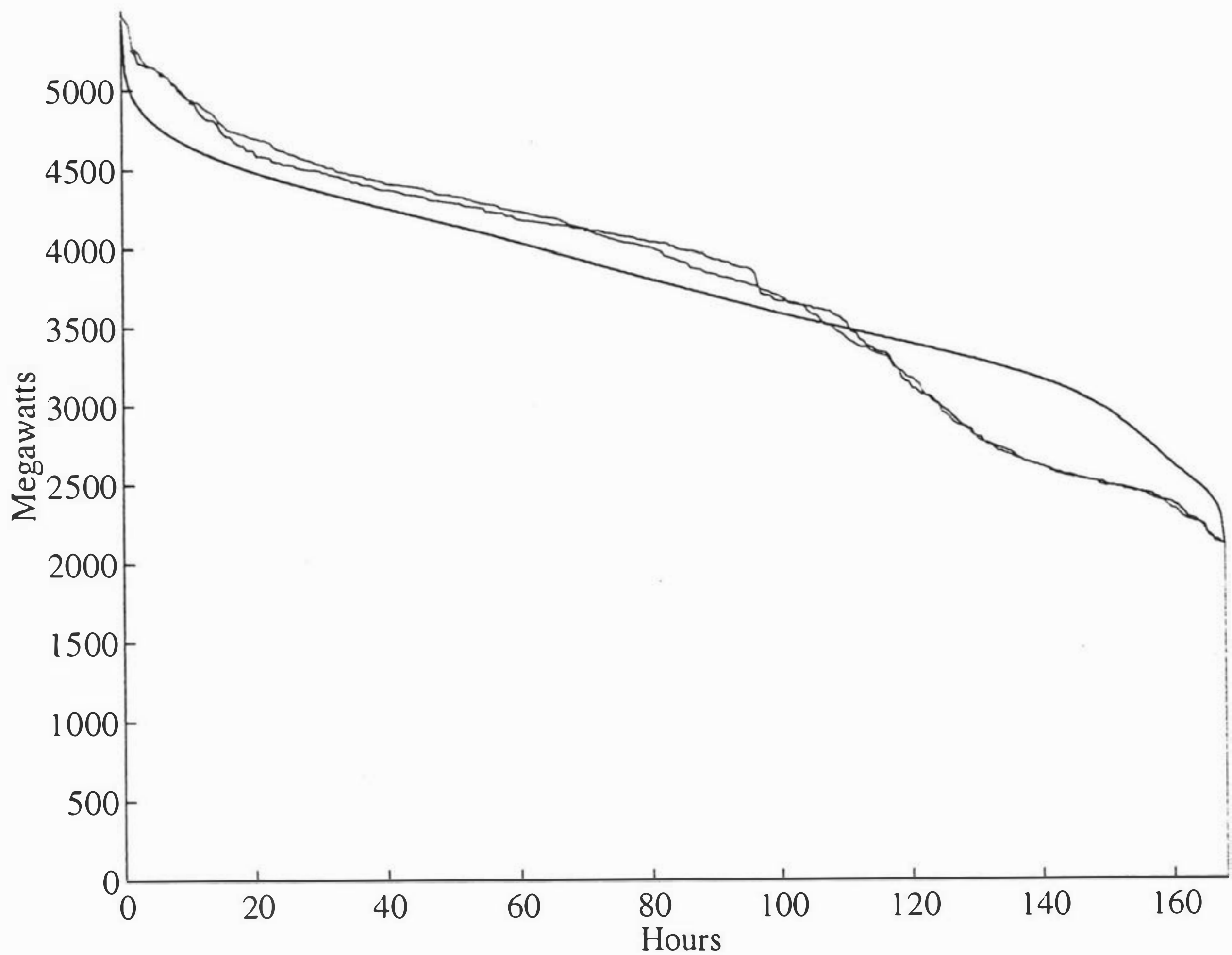


Figure 3.4: Actual (green), correlated (red) and independent (blue) interaction between two Load Duration Curves.

a Generalized Network[1].

3.7 Stations

Each station needs to generate power so as to meet the load at its node and the nodes to which its power is transmitted. Constraints on station generation include a generating capacity, possible uncertainty in supply, a fuel cost (for thermal stations) and a limited supply of water (for hydro stations).

To incorporate an uncertainty of supply in generation meeting load at *other* nodes, we need to identify, for all load, the stations which were used to meet that load. This requires a vast amount of information, as every station would need to be represented at every node. We could, instead, re-solve the network for every "state" of the stations, but this would require a lot of computational effort (see Section 2.5). Alternatively, we could model this uncertainty so as to try to make it independent of the actual meeting of the load. This independence could be obtained by separating the generation of power from the supplying of load. One way to achieve this is to require that stations at the same node present a "Contract Curve" (C.C.) representing the load they have chosen to meet. The C.C.'s are required to meet all load when distributed via the transmission network, and the stations at each node generate power so as to fill the C.C. at that node. Unfortunately changing the state of a station could also change the optimal distribution of the Contract Curves. However, it *may* be a reasonable approximation to assume that it does not. This assumption means power stations can compensate for uncertain generation of other stations *at the same node* (but not at other nodes), since C.C.'s are filled, by the stations at that node, in the same way LDC's are filled by other modelling methods (*c.f.* Section 2.5).

The inclusion of the C.C. means that, at station node, j, Equation 3.1 becomes

$$G_j(t) + \sum_{i \in \mathrm{IN}(j)} f_i(X_i(t)) - \sum_{i \in \mathrm{OUT}(j)} X_i(t) = L_j(t) \tag{3.2}$$

where G_j is the C.C. for node j.

The C.C.'s are obviously non-negative, since allowing them to become negative means we get paid for *creating* load at a station. We assume the C.C.'s are decreasing. The reason for this is that, since all of the LDC's are decreasing, the

[1]Here the load represents a sink of the network (in standard network terminology)

marginal *cost* of meeting any load, by any station, is also decreasing, so it will not be of benefit to generate more at a time of lower load than when the load is higher. Also, since the load is in terms of LDC's, if we do have generation at a station which is not decreasing then, when we come to implement this solution, it is not clear how that station's generation should be resolved. By imposing this constraint we are also implicitly imposing some sort of regularity constraint on a station's generation. In light of this there seems little point in allowing load to be given in terms of Load Curves, as proposed at the end of Section 3.6.

Further, we can impose such structure on the C.C. and not affect the optimal solution of our model. The advantage of this is that the non-negativity condition and any capacity constraint on a C.C. simply become

$$G(1) \geq 0, \quad \text{and,} \quad G(0) \leq \overline{G} \tag{3.3}$$

where $G : [0,1] \rightarrow \Re$ is the C.C. and $\overline{G}$ is its capacity.

Since the thermal and hydro stations fill the C.C. in different ways, we consider the two separately, so as to investigate the impact these two methods of scheduling have on both the way in which the problem is modelled, and the form of the solutions given.

3.7.1 Thermal Stations

Thermal stations each have an associated cost of generation, so when filling a C.C. we schedule the cheapest stations first. As part of this scheduling we incorporate the uncertainty of supply of stations (see above discussion) by convolving a probabilistic generation profile from the C.C.; as was discussed in Section 2.5.

To perform this convolution one requires the inverse of the C.C., which is the probability distribution function form of the C.C., since the C.C. is over the interval $[0,1]$ (see Section 3.6). To calculate this inverse in practice can be difficult and computationally intensive. It is, however, possible to *approximate* the inverse instead of calculating it exactly. Discussion of how the convolution is to be achieved is described later in the context of the model developed in Chapter 4.

To calculate the cost of generation for a thermal station we merely need to calculate the difference between the amount of required generation before and after that station is scheduled. All this requires is to be able to find the area under the resulting C.C., or, in terms of the inverse, the area under the associated probability distribution function over the interval $[0, \infty)$.

3.7.2 Hydro Stations

Hydro stations have the added complication of having to use water, a limited resource, to generate their power. Limited storage capacity and uncertain inflows into the hydro system make hydro stations difficult to model. The added difficulty of having many stations on a river chain is removed by amalgamating river chains so that each hydro system is represented by a single reservoir and station (see Section 1.4). This choice is reinforced by the fact that, over the weekly time step, the lag time between stations becomes reasonably insignificant, and so, the implicit assumption that all stations on a river chain generate in phase is reasonable.

The data for the controlled and uncontrolled inflows (see Section 1.4) is given as the average inflow for each $\frac{1}{52}$ of a year, and the inflow sequences for the last 60 years are used to predict the future inflows. Since we have effectively sorted the time dimension during each time step by using LDC's, we can not use any finer time scale information than the total inflow during each time step. Therefore we must make a decision about how to model the inflows with respect to the generation. The assumption is made that the controlled (and some of the uncontrolled) inflow arrives in such a manner as to accommodate any generation sequence required for the station; the rest of the uncontrolled inflow is then assumed to arrive at a constant rate throughout the time step. A factor is assigned to each hydro station indicating the fraction of the uncontrolled inflow which arrives at a constant rate. These assumptions are not unreasonable when one takes into account the river chain structure incorporated within the amalgamated hydro station and the fact that many hydro stations have some control over the "local" water flow. In terms of the model, the whole uncontrolled inflow provides a minimum generation amount for the whole time step, and the uncontrolled inflow arriving at a constant rate provides a minimum generation level for each hydro station's generation curve (assuming no spill).

Each station has a conversion factor for converting the gravitational potential energy of the released water into electricity. This conversion allows us to consider the water in terms of its potential generation, rather than its volume. As well as the usual constraints on all stations, the river systems themselves may have minimum and maximum reservoir release levels and minimum flows from the river mouth specified; these allow for environmental and recreational concerns. However, due to the uncertain nature of the inflows and the amalgamation of river chains, we need not model these constraints too precisely if it would be difficult to do so, as we do not want them to overly constrain the solutions when the coarseness of our approximation of the hydro stations may make such constraints unnecessary. If it transpires that the solution requires a particular station to fall irretrievably outside these release bounds, then the problem could be re-solved with a more precise approximation of these conditions, allowing greater flexibility through the balancing of solution time against the incorporation of all relevant constraints. We therefore need to consider how to model these constraints in both a precise and imprecise fashion.

Hydro stations fill C.C.'s via the (dual) methods described in Chapter 2 for hydro stations filling an LDC, i.e. we need to either fix a water cost for the hydro station, or fix the reservoir release. Since load is distributed over the geographic network, load at different nodes will "see" different costs for power from the same station. Hence, it becomes difficult (and computationally expensive) to specify the water costs for a hydro station. Also, ensuring that the water released from the reservoir each week allows enough storage for subsequent weeks requires the actual release anyway. For these reasons we use the *release* from the reservoir for each week to specify the generation with respect to the local C.C..

Storage in a reservoir is limited in terms of the time horizon and, since the inflows include seasonal patterns, this means that we need to include storage information for the whole year. This is most easily done by attaching a *waterflow network* to each hydro station, i.e. an inter-temporal linking of the hydro station from week to week. The waterflow network is the obvious time discretized network discussed in Chapter 2, containing arcs for releases, inflows (the uncontrolled inflows can either be separated from the controlled inflows or be included with them and appear as a minimum release as in Boshier and Lermit [1]) and storage from one week to the next for the reservoir; see Figure 3.5. In determining the best release and storage for each week, we need to determine a generation schedule for each week of the year requiring a copy of the geographic network to be solved for each week.

The release needs to be independent of any uncertainty in supply, as otherwise this uncertainty would need to be carried through the whole year. Therefore the

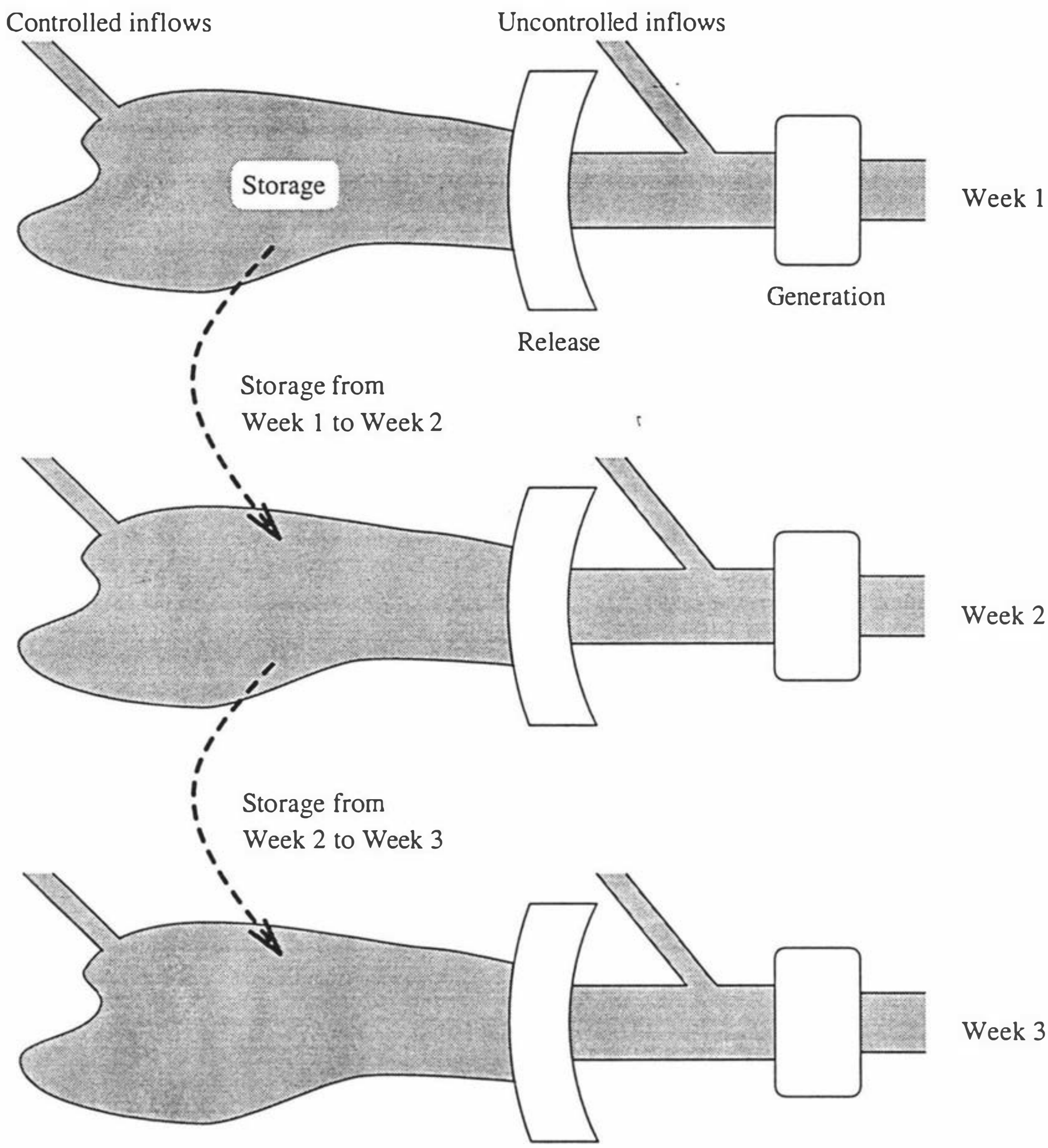


Figure 3.5: An example of a waterflow network

release may be no less than the maximum total generation under any possible future. Including uncertainty when filling a C.C., will give the *expected* generation and so the maximal generation would need to be calculated separately. If we used the expected release as an approximation to the maximum release, when the station does not breakdown (which is more often than not) we will end up with *less* water in the lake than planned for. In a dry year this could have disastrous results.

In the working model, we disallow the possibility of an uncertain supply for a hydro station. For flexibility we should allow this in general, although it would require more work as we would need to determine both the expected and maximum generation.

Having determined a release, H (total volume of water released during the week, in Megawatt hours), from the waterflow network, the hydro station generation then *splits* the C.C. so as to generate at peak capacity for the longest period, and generate exactly H MWh if possible (see Section 2.6). If the station cannot generate H MWh, the remaining release is spilled.

3.7.3 Auxiliary Stations

Auxiliary stations (see Section 1.1) are assumed to run continuously at a constant level during each time step; they are given no cost, limit or control on generation. These stations are modelled by removing their proposed (constant) generation from the load before solution begins, which may involve allowing some nodes to have negative load. To allow for the fact that system constraints may not permit all of this negative load to be used, we allow the "resulting load" at this node to be negative, so long as it is above the initial negative load.

3.8 Non-Supply

The term *non-supply* is usually used to describe load which cannot be met by the generation system. In allowing uncertain generation by some of the stations, we need to cater for such a possibility, as there is always a (tiny) probability that the load can not be completely satisfied due to breakdowns. The amount of non-supply needs to be minimized by some method, since we do not want to allow the optimal solution to generate no power at all since this involves no cost! This is done by introducing a *cost of non-supply*, which is attached to any load which is not met by some station.

Allowing non-supply in the model ensures that there is *always* a feasible generation schedule, i.e. generate nothing. Another reason for allowing non-supply is that there is then no longer any reason to enforce all the stations at a node to *completely* fill their C.C., as any contracted generation which can not be filled becomes non-supply and effectively "filled" by an imaginary non-supply station. As a result of this, filling a C.C. involves *only* determining the cost to fill it, and thus takes place during evaluation of the objective function. This implies that, if determining the generation of each station requires the use of some non-linear equations, we are only introducing non-linearities into the objective function and *not* into the constraint set.

Allowing for non-supply also means that we are only interested in the curve resulting from the scheduling of some of the stations in order to schedule the remaining stations—if we can discover some information about a curve which will allow *scheduling* of a station and the computation of the area under the resulting curve easily, we would prefer to utilize this information rather than having to determine the *actual* curve at each stage of the scheduling process. Once an optimal solution has been determined, we can use computationally expensive methods to determine the actual filled curves (if this is required).

In terms of filling C.C., we allow stations at the same node to explicitly allow for others' uncertain generation. The residual unmet contracted load is then penalized at the cost of non-supply. However, this load *could* have been met by a station at another node, so the cost that should actually be applied is the *cost of re-supply* by stations at other nodes. In general, this cost could be estimated for each node and could either be incorporated in a different cost of non-supply at every node, or explicitly split up into the cost of non-supply and a cost of re-supply for each node.

The cost of non-supply, for our working model, is assigned a constant value, but for the sake of flexibility we allow it to take different values at every node, if required.

To prevent non-supply being allowed only at station nodes (and hence requiring "non-supplied generation" to be transmitted via the geographic network), we introduce a non-supply curve at each node. Such non-supply curves can be used to identify badly modelled or highly constrained areas of the geographic network and be used to ensure feasibility.

3.9 Stochastic Elements

The two major elements of the model being developed are that it should provide a high level of detail about the physical system, and that it should account well for the stochastic elements of the system. To incorporate these two elements at a consistent level of detail, and to allow for flexibility, in that each element may change its level of approximation relatively easily without affecting the modelling of the other element, they need to be separable in our model. It is, in fact, the difficulty in providing an adequate balance between these two aspects that initially prompted a flexible approach.

To make these two elements separate in some implementational sense, whilst not enforcing too much of this separation into the structure of the solutions, is difficult. We chose to initially develop a deterministic model which would then be extended to a stochastic model, consistent with the model's objectives, as described in Section 1.2. The actual stochastic extension to the deterministic base model need not be fixed. This will allow new (possibly better) methods to be used, as well as allowing different stochastic extensions to be tested under similar conditions. Flexibility in the deterministic model could be used to enhance particular stochastic extensions.

It may be that not all stochastic extensions could be realistically applied to the base model developed, and so we need to be careful, in the development of the deterministic model, that there are at least *some* stochastic extensions that can be used. To try to develop a deterministic base model which would allow *any* stochastic extension is beyond the scope of this thesis, and not necessarily useful in terms of its approximation to reality—a deterministic base model, specifically tailored for the particular stochastic extension, would do better in this regard. In light of this it may seem that we should have started with a stochastic extension and tailored our deterministic approach to that; however, using this approach, the stochastic extension used tends to set limits on the amount of detail in the deterministic approach used, but in this thesis we were seeking to make the level of detail in the deterministic base problem drive the stochastic extension (see Chapter 1).

We do not *blindly* hope that there will be a reasonable stochastic extension available at the end of the deterministic model development; we plan on using a scenario aggregation approach, as well as investigating other possible extensions. Scenario aggregation methods allow for stochastic detail in the number and variety of scenarios used, while deterministic detail is dependent on the underlying deterministic subproblems; both of these can be independently varied. The discussion of stochastic extensions is left for Chapter 6, so they may be investigated in terms of the deterministic model developed.

3.10 The Electricity Curve Approximation

We use the term *Electricity Curve* to refer to any curve which, represents the generation, transmission or use of electricity, e.g. an LDC or a C.C..

There is no point in using an LDC approach if one is going to keep the approximation of the LDC (and hence other Electricity Curves) in the form given by the data, namely the average load level for each half-hour. We require an approximation which can store much of this information using only a few significant values, since power station generation (a *variable* of our model) needs to be approximated in the same way. In looking for such an approximation we need to consider the impact on the entire model, and to allow for flexibility.

In light of the points made in Section 3.6 about the interaction of LDC's, it seems appropriate that the information stored about the Electricity Curves should be in the form of coefficients of some fixed basis used to approximate the curves, i.e. we would like to write each approximated Electricity Curve, $G(t)$, as

$$G(t) = g_1 B_1(t) + \cdots + g_N B_N(t) \tag{3.4}$$

where $\{B_1, \ldots, B_N\}$ is the chosen basis.

Due to the nature of the LDC, we need to ensure the approximations we choose for them are decreasing, even if there is a *closer* approximation (in terms of least squares, say) for them which is *not* decreasing. As an example of this, Figure 3.6 shows where the least squares 3-piecewise quadratic approximation to an LDC is not decreasing.

The basis used needs to not only be a good approximation to the LDC, but also should allow the types of structures, required by system, for the other Electricity Curves (this is reinforced by the findings of Chapter 8). We want to keep the number of basis elements in this representation low, as it affects the number of variables in our final model. Stations and transmission arcs have capacity constraints and require non-negative power, so the approximation needs to be able to be constrained within a fixed range. The existence of this fixed range will also mean that some of the curves required will *want* to be at capacity or zero for part of the week and this may cause non-smooth points on the curve (see Figure 3.7). Smooth bases trying to approximate non-smooth behaviour and constant function values over positive length intervals often oscillate or admit superfluous optima and do not approximate such phenomena well (see Figure 3.8). This suggests the use of a piecewise basis of some kind.

The fixed range implies the use of a piecewise linear approximation in an effort to keep the constraining equations linear. Unfortunately, piecewise linearity is not a very good approximation for an LDC without a fine partition. Piecewise linearity may seem reasonable in light of the uncertainty in the Load Duration Curves, but, we seek to incorporate the particular structure of the solutions as much as possible, as it is the *structure* of the solutions which is of most interest. Also, in allowing a flexible approach, we seek to find the limits of the approximations used. By enforcing C.C.'s to be decreasing, as explained in Section 3.7, the non-negativity and capacity constraints of Equation 3.3 simply become

$$g_1B_1(1) + \cdots + g_NB_N(1) \geq 0, \text{ and, } g_1B_1(0) + \cdots + g_NB_N(0) \leq \overline{G}$$

in terms of the basis approximation.

To enforce a C.C. to be always decreasing via linear constraints indicates the use of a quadratic basis. As a piecewise quadratic basis provides a good approximation to an LDC, moving to a cubic approximation would achieve little more in the way of accuracy to this approximation; furthermore since there is no motivation to use any type of basis other than a piecewise polynomial, we stay with the piecewise quadratic. Figure 3.9 shows a 3-piece, 4-piece, and 5-piece piecewise quadratic and an 8-piece piecewise linear approximation to an LDC with evenly spaced breakpoints.

Empirical experimentation shows a 4-piece piecewise quadratic (NB: piecewise quadratic and *not* quadratic spline) to be a *good* approximation to LDC. We choose this for our central model, with partition {0, 0.1, 0.4, 0.7, 1.0}; Figure 3.10 shows such a piecewise quadratic approximation. For flexibility we allow for an n-piecewise quadratic (or linear or discrete) approximation. We still have the open problem of constraining the transmission, but this is left to be dealt with by the

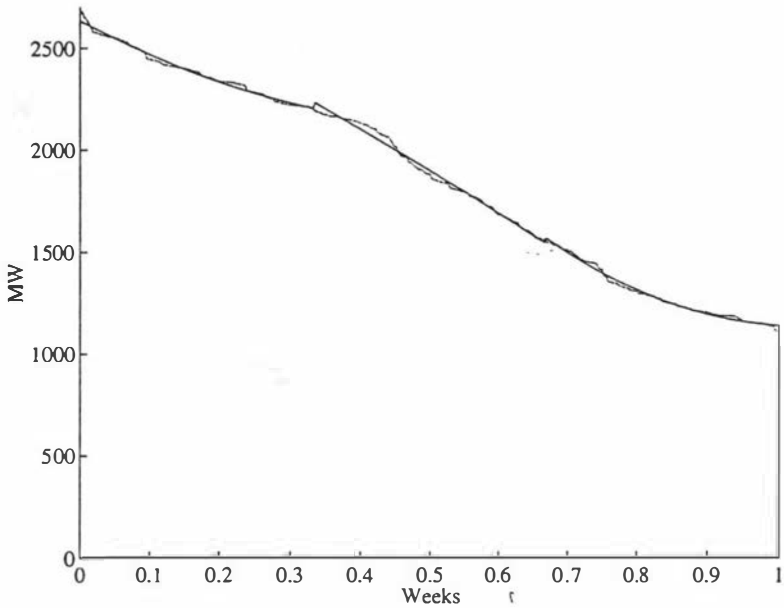


Figure 3.6: The best least squares 3-piecewise quadratic approximation to an LDC

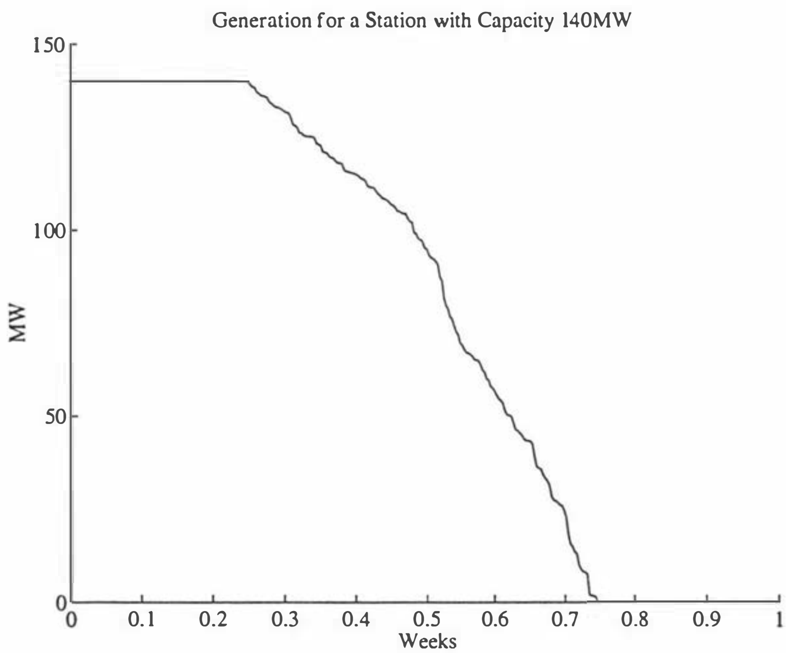


Figure 3.7: Intended generation of some station

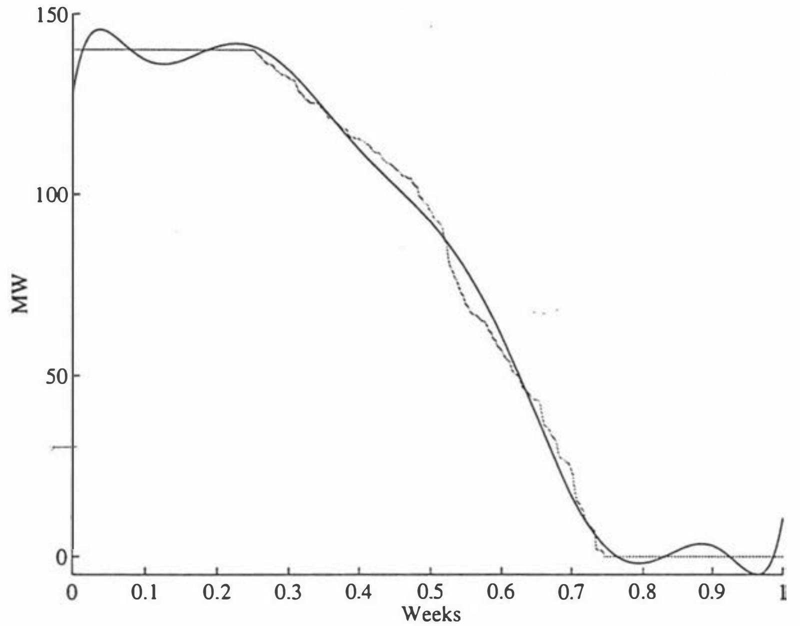


Figure 3.8: Best fit 12-nomial approximating a non-smooth curve

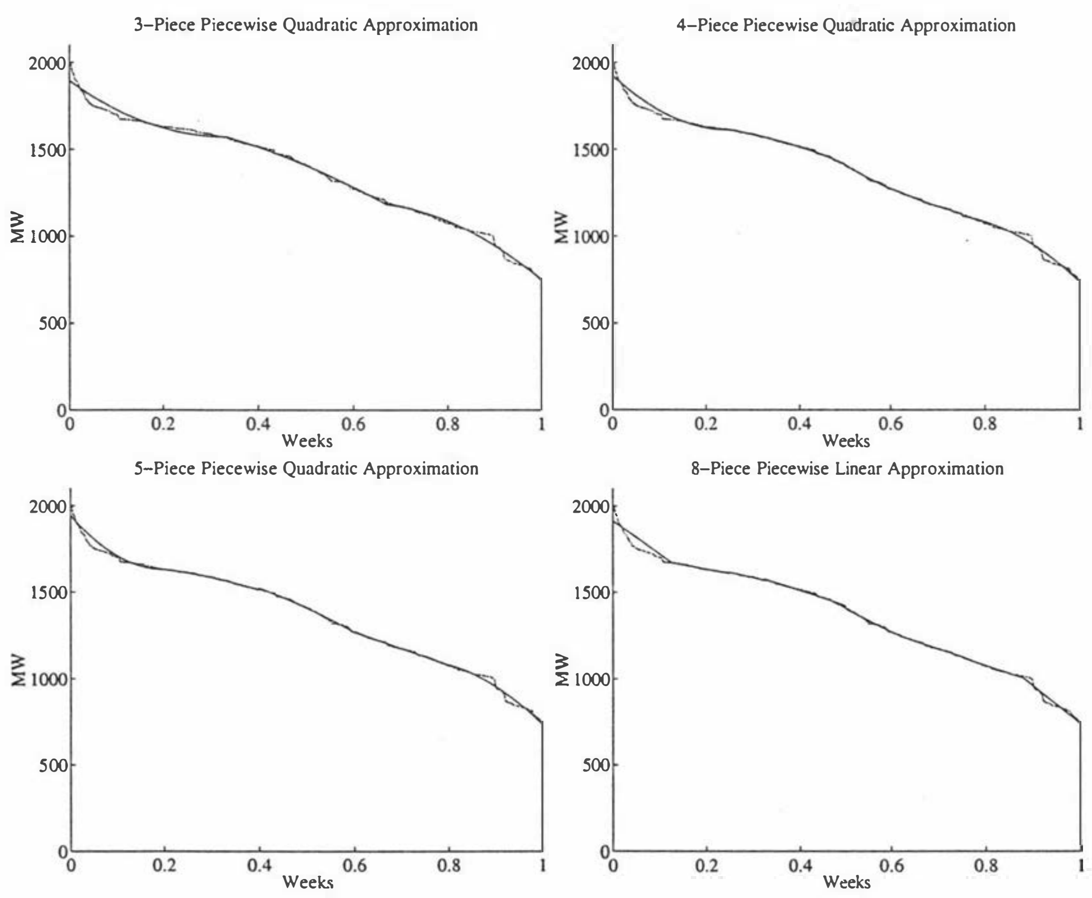


Figure 3.9: Various piecewise approximations to an LDC

model.

For reference, in connection with the filling of C.C., the inverse of a decreasing piecewise quadratic is reasonably easy to determine; unfortunately it is a rational polynomial with (possibly) more pieces than the original, and initially unknown break points.

3.11 Discussion

An underlying geographic network describes the distribution of load and power stations. The electricity flow in the network is approximated for each week, over the year, by piecewise quadratics. Each power station node defines a Contract Curve which represents the load to be met by the station(s) at that node. These stations fill the C.C. in the cheapest possible manner, given that unmet contracted generation is penalized at a cost of non-supply. The C.C.'s must satisfy all load in the system, represented by LDC's, as ensured by the transmission network constraints.

Generation by each hydro station is restricted by its generation capacity and the release from the reservoir for that week. The hydro stations have an attached waterflow network which describes the lake levels and releases for each week; these networks tie the weeks together.

Every node has a non-supply curve to allow feasibility. This power is again penalized at the cost of non-supply.

Such a representation creates a natural division in which the (contracted) supply of load takes place via a constraint set (the geographic network) and the generation of power (filling the Contract Curves) takes place in the objective function.

The next Chapter describes various methods considered for modelling the system. In order that there is no need to cover the fundamental elements of the system for each of the various methods, most of the basic structures defined in this Chapter are taken to hold for each of the methods described, with modifications and refinements as are necessary to develop the particular model.

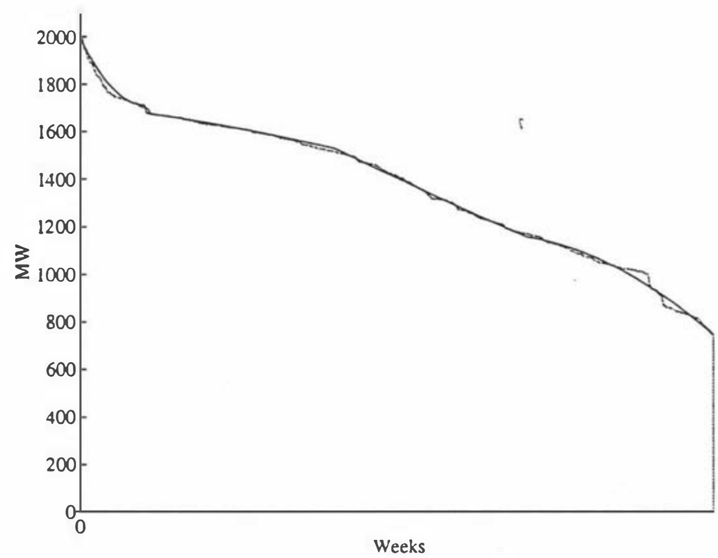


Figure 3.10: A 4-piecewise quadratic approximating an LDC

Chapter 4

Inappropriate Approximations

Description of the various wrong turns, blind alleys, and computationally unworkable approaches taken on the model before a final, workable, model was produced, is given in this Chapter. The reason for including such a chapter is partly because a considerable amount of the development work for the model went into such areas, partly in an effort to express why the problem was, or was not, modelled in particular ways, and also, partly as a warning to later modellers of such systems: "don't try this at home!" In our opinion, the two major reasons for *not* including an approximation in the model appear to be either, that it created significant undesirable structure in the solutions which was an artifact of the approximation rather than the problem, or that it was unimplementable in terms of solution time or convergence. In the latter case it must be remembered that the deterministic model being developed is only a part of a stochastic extension, and therefore needs to be robust.

It may seem that in any approximation one includes unwanted structure, if only in the sense that solutions can only be in the form of the approximation. It is not this type of unwanted structure in the problem that is of concern; that structure is obvious, and is clearly taken as being acceptable when the approximation is made. The extra structure of concern is the insidious structure *implicitly* imposed by the approximation. This structure is almost never obvious from the outset, and can be formed by a combination of seemingly unrelated approximations. Such structure is mainly noticed in the *form* of the actual solutions obtained; as a symptom, it is often difficult to trace. The effects of such structure are, unfortunately, often unnoticed as they tend to be in the form of different management styles; these can range from over-conservation of water, to full on/full off policies for some type of station. Of course it is sometimes impossible to avoid this extra structure, but it always pays to be aware of its causes and effects. So it is with care, and an ultimately pragmatic view, that *any* approximations are made.

In the previous chapter the approximation details of most of the model were left undeveloped, the intention being to create a skeletal model made up of elegant, and necessary, parts which could be enriched in the development of an implementable model. The skeletal model is not meant to be totally inflexible, since to create a *good* implementable model may require some re-working of the model's basics; it is supposed to embody the essence of the model, to emphasize the underlying direction of model development. This model also allows this Chapter to concentrate on specific areas of the model from within an encompassing structure which describes the interaction of the model's components, without the added complication of needing to explain an entire model.

In this Chapter, the aspects modelled will be discussed in the generality required for the desired flexibility and consistency. Where moving the level of approximation past some point would require making changes to other aspects of the model, these levels are explained and the reasons for, or against, moving beyond such a limit are discussed.

4.1 Filling Contract Curves

Probably the most obvious omission of Chapter 3 is the actual filling of the Contract Curves. We need to determine a method of scheduling thermal stations in a (reasonably) fast, automatic, manner which can be achieved without *prior* knowledge of the exact C.C.. The resulting curve is given by:

$$F(x) = \max\left\{(G^{-1} * F_Q(x))^{-1}, 0\right\} \tag{4.1}$$

where G is the C.C., F_Q is the probability distribution function associated with the station's probable maximum generation, and the convolution is given by:

$$G^{-1} * F_Q(x) = \int_{-\infty}^{\infty} G^{-1}(x-z)\ \mathrm{d}F_Q(z)$$

When the information about the uncertain supply of the station is in the form of a *single* probability of complete breakdown this becomes:

$$F = \max\left\{\left(pG^{-1}(x) + (1-p)G^{-1}(x+\overline{Q})\right)^{-1}, 0\right\}$$

where p is the probability of breakdown, and $\overline{Q}$ is the capacity of the thermal station. For a decreasing piecewise quadratic we can calculate the inverse exactly; however, F will be a piecewise function with more pieces than G. Keeping track of the new pieces can be extremely difficult, *viz* the following case, with $\overline{Q} = 1$, $p = 0.9$, and G given by;

$$G(t) = \begin{cases} 1.2 - t & t \in [0, 0.5] \\ 2(t-1)^2 & t \in (0.5, 1] \end{cases}$$

so G^{-1} becomes;

$$G^{-1}(x) = \begin{cases} 1 - \sqrt{\frac{x}{2}} & x \in [0, 0.5) \\ 0.5 & x \in [0.5, 0.7) \\ 1.2 - x & x \in [0.7, 1.2] \\ 0 & x > 1.2 \end{cases}$$

giving

$$F(x) = \begin{cases} 1.2 - 10t & t \in [0, 0.05) \\ 2(1-10t)^2 & t \in [0.05, 0.0684) \\ 0.314 - 1.111t - 0.1\sqrt{0.193 - 0.686t} & t \in [0.0684, 0.28] \\ 0 & t \in (0.28, 1.0] \end{cases}$$

Taking $\overline{Q} = 0.4$ gives a *very* messy F with 6 pieces; one can see how difficult such an approach would become when multiple stations try to fill a C.C.. However, we can extract one advantage: as we are only interested in the *amount* of generation for any station, when determining the cost of a solution, there is no need to re-invert F. We can simply use F^{-1}, since:

$$\int_0^1 \max\{F(t), 0\}\, \mathrm{d}t = \int_0^\infty F^{-1}(x)\, \mathrm{d}x \tag{4.2}$$

Equation 4.2 possesses the advantages that we do not need to calculate the, analytically difficult, left-hand-side, and we need never know F^{-1} for negative x.

For the moment we assume that either of the above methods, or some other approach, is used, and concentrate on some other aspects of the model. Later in this Chapter we will re-visit the filling of C.C., but for now it is sufficient to assume that we have a function $E(G, F_Q)$ which gives the total remaining contracted load from the C.C. G after scheduling a station with probable maximum generation described by the probability distribution function F_Q. To cope with multiple stations we also define the related function $F(G, F_Q)$ which gives the curve *resulting* from scheduling the station. Note that we have the following relationship:

$$E(G, F_Q) = \int_0^1 \max\{F(G, F_Q)(t), 0\} \, dt$$

The amount of fuel used by a second station, Q_2, is given by

$$E(G, F_{Q_1}) - E(F(G, F_{Q_1}), F_{Q_2})$$

We also define the following, in anticipation of approximations to come, and to clean up the notation somewhat. For C.C. G, recalling that a C.C. is non-negative, we write

$$m(G) = \int_0^1 G(t) \, dt \tag{4.3}$$

When there is no confusion over the C.C. involved we will often write m for $m(G)$.

Notice that when F_Q represents a single probability, p, of complete station failure, then $E(G, F_Q)$ is exactly given by

$$E(G, F_Q) = (1-p)E(G, \chi_{[\overline{Q},\infty)}) + pm(G)$$

where $\chi_{[\overline{Q},\infty)}$ is the characteristic function of the set $[\overline{Q}, \infty)$ having value one on the set and zero elsewhere.

4.2 Hydro Stations Filling Contract Curves

As mentioned in Section 3.7, hydro station generation splits the C.C.. Since this takes place *prior* to the scheduling of thermal stations, it requires determination of the resulting curve, which in turn requires determining exactly *where* the station splits the curve. For a C.C., G, this requires determining $\delta \geq 0$ such that:

$$\int_0^1 \max\left\{\min\left\{G(t), \overline{H} + \delta\right\} - \delta, 0\right\} \, dt = H$$

where $\overline{H}$ is the capacity of the hydro station. This is easier to calculate when put in terms of G^{-1}, as the above equation then becomes

$$\int_{\delta}^{\overline{H}+\delta} G^{-1}(x) \, dx = H$$

If no such δ exists, then the generation is only

$$\int_0^1 \min\{G(t), \overline{H}\} \, dt < H$$

with the remaining release being spill. Figure 4.1 shows the two cases when, firstly, a δ exists and, secondly, does not exist, for the same H and $\overline{H}$ values.

Once δ has been found, it is easier to specify the inverse of the resulting curve than the curve itself; this inverse is given by $G^{-1}(x)$ for $x \in [0, \delta]$, and $G^{-1}(x + \overline{H})$ for $x > \delta$. This reiterates that, for the purposes of evaluating the cost of a C.C., it is only the inverse of the C.C. that interests us.

To avoid the difficulty of needing to determine δ, we can specify that each hydro station must have its *own* Contract Curve. In this case if such a δ exists, the total generation is given by H, and if δ does not exist, the total generation is given by $m(G) - E(G, \overline{H})$; here we slightly abuse notation by allowing the second argument of E be a single value, representing the probability distribution function which is equal to zero below this value and equal to one above it. This means the amount of non-supplied contracted load for a hydro station's C.C. is given by

$$\max\left\{m(G) - H, E(G, \overline{H})\right\} \tag{4.4}$$

Immediately one can see a potential problem with this: any combination of release, H, and C.C. G, with

$$m(G) - H = E(G, \overline{H}) \tag{4.5}$$

will be a non-differentiable point of the objective function. It may seem that we could eliminate this by requiring no spill; apart from the fact that this creates a non-linear constraint (due to the non-linearity of E), it may also cause some solutions to be infeasible (as there are times when spill is unavoidable, especially when trying to cater for the uncertain environment).

A possible remedy might be to "smooth-off" these edges via an approximation to Equation 4.4. Discussion of this is left until an exact representation of E is given, since such an approximation will depend on the nature of this representation.

In the case of two hydro stations filling the same C.C., the order of scheduling is unimportant. To see this, consider the scheduling of two hydro stations, with releases H_1 and H_2, and capacities $\overline{H}_1$ and $\overline{H}_2$ respectively. First schedule them individually into the C.C.; if there is no overlap on the energy they wish to meet, then, regardless of the order in which they are scheduled, they will still split the curve in the same place (see Figure 4.2). If there is overlap, call the height of the overlap η (see Figure 4.3). Notice that if we schedule the two stations as one large station, with release $H_1 + H_2$ and capacity $\overline{H}_1 + \overline{H}_2$, the split of the C.C.

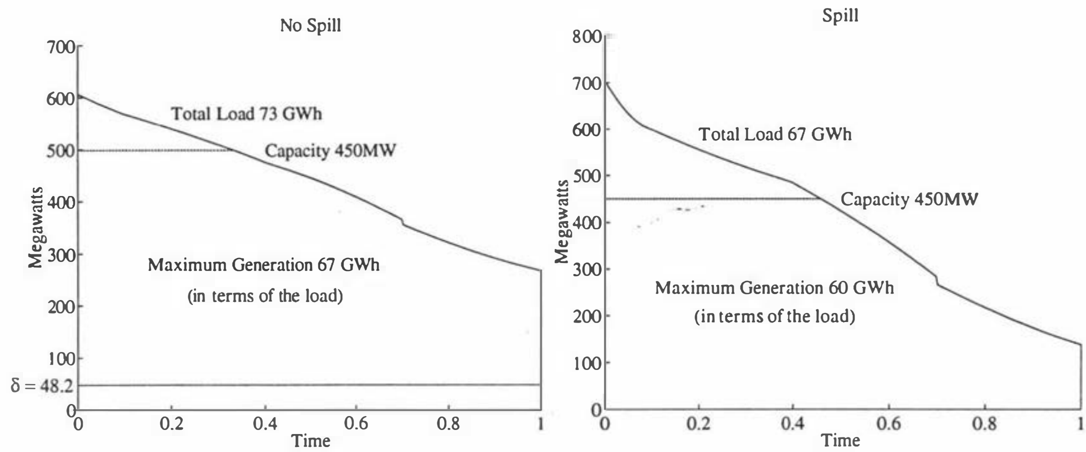


Figure 4.1: The two cases for a hydro station filling a C.C.

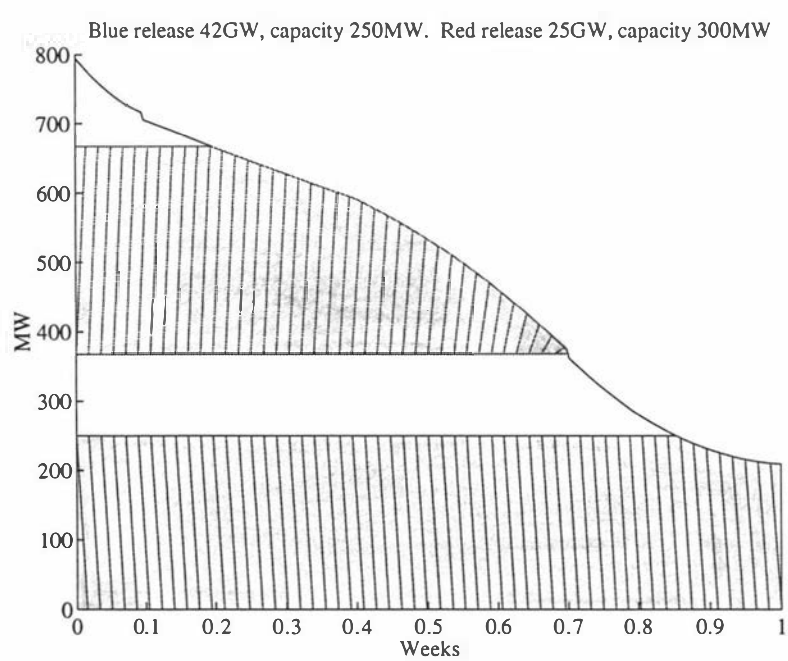


Figure 4.2: Two hydro stations filling a C.C. with no overlap

would contain both of the individual splits, since the extra area must exactly match the overlapping area. Consider the portion of the curve that would be scheduled to the combined station: if we remove the generation from one of the stations, the resulting curve has height equal to the capacity of the other station, and area equal to the release of that station—this is exactly where this station would split the resulting curve. The stations may therefore generate differently, but the resulting curve is independent of the order of scheduling them. This readily generalizes to more than two hydro stations, by considering them in overlapping pairs.

In light of this, it may seem that, instead of requiring each hydro station to have an individual C.C., we merely need to require that *all* of the hydro stations, at a single node, have a collective C.C.. However, this allows underutilized stations to generate using the *spill* from other stations. Consider the situation where the generation of two stations does not overlap (when scheduled individually into a C.C.), with some *positive* gap between the two schedules (as in Figure 4.3), and the lower station (in terms of splitting the C.C.) is spilling. Because of the positive gap, if the stations were scheduled using the same release (as they would be if scheduled as a combined station), the non-spilling station could generate using some of the spill of the other station by decreasing this gap.

4.3 Transmission Capacity

In Chapter 3 the modelling of transmission line losses was mentioned. Here, we choose to use a linear line loss as, otherwise, we are incorporating a non-linear *equality* constraint for every *coefficient* of every arc of every week (i.e. in the order of 10 000 such constraints!). Even just including a non-linear line loss for the North-South DC link would incorporate in the order of 600 non-linear equality constraints. Also, due to the basis representation, the non-linear constraints would necessarily be an *approximation* in any case.

For transmission capacity, the fact that transmission curves are not constrained to be decreasing implies that the constraints required to keep transmission within a fixed range must be non-linear. To see this, consider a quadratic, $at^2 + bt + c$, over $[0, 1]$ which we seek to constrain in the range $[A, B]$. This is exactly the same as requiring the polynomial $at^2 + bt$ to be constrained in the region $[A - c, B - c]$,

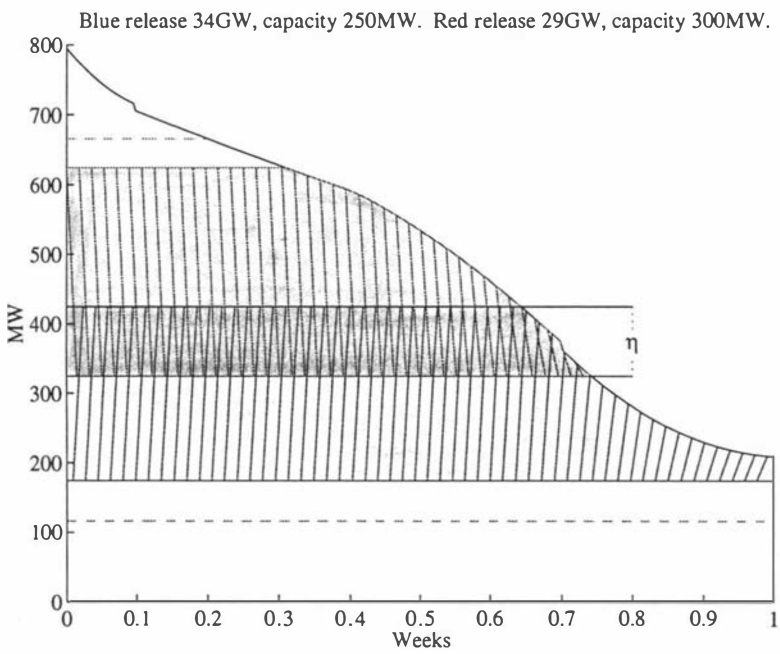


Figure 4.3: Two hydro stations filling a C.C. with overlap

with $A \leq c \leq B$; hence we require the following conditions:

$$A \leq a + b + c \leq B$$
$$A \leq -\frac{b^2}{4a} + c \leq B \text{ when } -\frac{b}{2a} \in [0,1]$$

The region of allowable coefficients is shown in Figure 4.4.

If we require these constraints on every transmission line, we would have in the order of $15 \times 2 \times 52 \times 4 = 6250$ non-linear constraints, i.e. two for every arc in the geographic network (one for each direction), for every week in the year, for each quadratic piece of the transmission curve. Including these constraints explicitly would adversely affect solution time. Instead we recognize that the maximum transmission capacities are not truly hard constraints, so that they may be better modelled as penalty functions in the objective.

To this end, suppose we have a function E which gives the amount of load remaining after scheduling a station. If we could treat transmission as we treated the scheduling of a station, and penalize untransmitted power, this would act like a penalty function, penalizing over-capacity transmission. When the transmission curve is $X(t)$ and the line's capacity is $\overline{X}$, the untransmitted power, for that line, is exactly $E(X, \overline{X})$, and the negative transmission is $E(-X, 0)$ (where we use the same slight abuse of notation here as we did in Section 4.2). These two values are penalized in the objective function; since negative transmission incurs a power *gain* in transmission, this will probably be penalized more heavily than the over capacity transmission. An advantage of this approach is that it keeps the transmission and station generation in a consistent form.

It may seem that one could perform a similar trick to model the unreliability of thermal stations when "scheduling" the electricity through unreliable transmission lines, however, to do so is not valid[1]. This would give the expected *transmission*, the effect of which would be to derate the capacity of the line. What we seek is the expected *cost* to the system, given that the line is unreliable, and might, therefore, not be used. The other problem with such an approach is that not every transmission line in the system is modelled; the geographic network arcs actually represent many lines, so that allowing an arc to be either "off" or "on" (or even

[1] It may seem that this is exactly the approximation used for thermal generation, however there are two important differences; the first is that there is no *cost* directly applicable to the uncertain transmission to apply (to use the penalty is clearly artificial); the second is that we do not *just* use the cost of the expected generation to approximate the expected cost of generation, in the case of thermal generation.

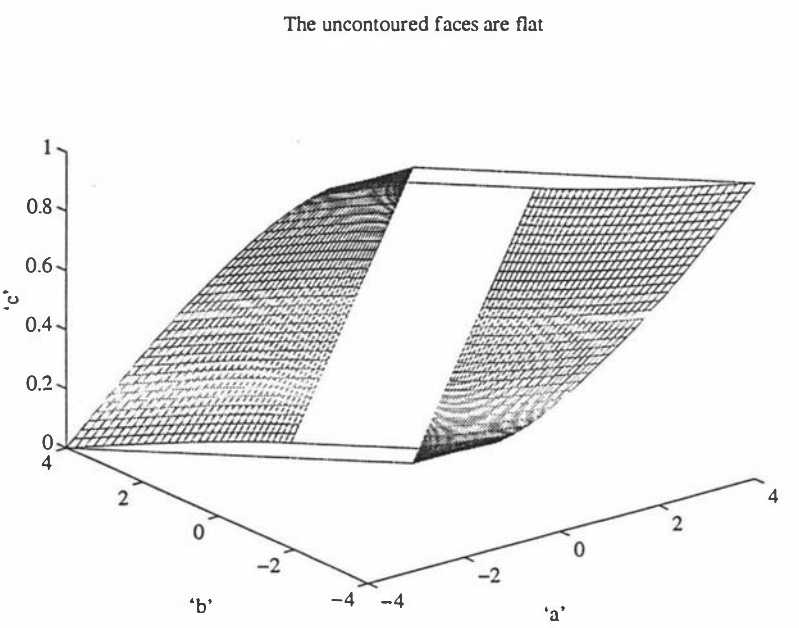


Figure 4.4: Values of $[a\ b\ c]$ for which $at^2 + bt + c \in [0, 1]\ \forall t \in [0, 1]$

assigning an arc a probability of being at a lower capacity) does not effectively model what does, or even could, happen in practice.

The one exception is for the DC link: there is only one transmission link between the North and South Islands, comprising twelve (parallel) lines. If one of the lines breaks down, the transmission from the South Island to the North Island (the predominant direction of transmission) is restricted to a lower capacity, and stations in the North Island need to generate more to cope with the lost potential energy. In this case the penalty factor could act, appropriately, as a cost of re-supply. As the model is affected little by whether or not this aspect is included, flexibility dictates that the specific modelling of the DC link be left unresolved.

4.4 The Objective Function and Convexity

Letting c_{NS} be the cost of non-supply, c_Q the cost of fuel at thermal station Q, and P the penalty cost associated with over-capacity and negative transmission. We can then write the objective function as:

$$\begin{aligned} Z = & \sum_{j \in \text{NODE}} c_{NS} m(G_j) + \sum_{j \in \text{HYDRO}} c_{NS} \max\{m(G_j) - H_j, E(G_j, \overline{H})\} \\ & + \sum_{j \in \text{THERMAL}} \sum_{k=1}^{K_j} c_{Q_{jk}} E(G_{(k)}(G_j), F_{Q_{jk}}) + \sum_{i \in \text{ARC}} P(E(X_i, \overline{X}) + E(-X_i, 0)) \end{aligned}$$

where $G_{(k)}(G)$ is the resulting C.C., after the first $k-1$ stations have been scheduled into G. In particular $G_{(1)}(G) = G$. NODE is the set of all nodes, HYDRO and THERMAL are the sets of C.C. for nodes with hydro stations and nodes with thermal stations, respectively (the *same* node may be in both sets), and ARC is the set of all arcs.

Notice, in particular, that the objective function is convex if E and m are convex, since the pointwise maximum of two convex functions is convex, as is the sum of convex functions. Also, if E admits multiple optima, then so will the objective function. From Equations 3.4 and 4.3 we see that $m(G)$ is linear in terms of the basis coefficients, since

$$m(G) = \int_0^1 G(t)\ \mathrm{d}t = g_1 \int_0^1 B_1(t)\ \mathrm{d}t + \cdots + g_N \int_0^1 B_N(t)\ \mathrm{d}t$$

As the convexity depends only on E, it would, therefore, be beneficial if E were convex. We actually have more reason to desire E convex than just making our objective convex; it turns out, that more generally, E is convex over the set of integrable functions. Theorem 4.1 proves the case where the supply is certain, and Theorem 4.2 extends this to the discrete uncertain supply case.

Theorem 4.1 *Let $\mathcal{G}$ be a convex set of Lesbegue integrable functions, and define $E \mid \mathcal{G} \to \Re$, by*

$$E(G) = \int_0^1 \max\{G(t) - \overline{X}, 0\}\ dt$$

for some fixed $\overline{X}$. Then E is convex over $\mathcal{G}$.

Proof We can write E as

$$E(G) = \int_{A_G} G(t)\ \mathrm{d}t - \overline{X}\mathrm{m}(A_G)$$

where $A_G = \{t \in [0,1] \mid G(t) \geq \overline{X}\}$, and m is the Lesbegue measure. Let $P, Q \in \mathcal{G}$, and $\lambda \in (0,1)$. Put $R = \lambda P + (1-\lambda)Q$. Let A_P, A_Q, and A_R be defined in a like manner to A_G. Then,

$$\begin{aligned}
E(R) &= \int_{A_R} R(t)\ \mathrm{d}t - \overline{X}\mathrm{m}(A_R) \\
&= \lambda\left(\int_{A_R} P\ \mathrm{d}t - \overline{X}\mathrm{m}(A_R)\right) + (1-\lambda)\left(\int_{A_R} Q\ \mathrm{d}t - \overline{X}\mathrm{m}(A_R)\right) \\
&= \lambda\left(\int_{A_P} P\ \mathrm{d}t - \overline{X}\mathrm{m}(A_P)\right) + (1-\lambda)\left(\int_{A_Q} Q\ \mathrm{d}t - \overline{X}\mathrm{m}(A_Q)\right) \\
&\quad + \lambda\left(\int_{A_R\setminus A_P} P\ \mathrm{d}t - \overline{X}\mathrm{m}(A_R\setminus A_P)\right) \\
&\quad - \lambda\left(\int_{A_P\setminus A_R} P\ \mathrm{d}t - \overline{X}\mathrm{m}(A_P\setminus A_R)\right) \\
&\quad + (1-\lambda)\left(\int_{A_R\setminus A_Q} Q\ \mathrm{d}t - \overline{X}\mathrm{m}(A_R\setminus A_Q)\right) \\
&\quad - (1-\lambda)\left(\int_{A_Q\setminus A_R} Q\ \mathrm{d}t - \overline{X}\mathrm{m}(A_Q\setminus A_R)\right)
\end{aligned}$$

Now if $t \in A_R \setminus A_P$ then $P(t) < \overline{X}$ so that

$$\int_{A_R\setminus A_P} P\ \mathrm{d}t \leq \overline{X}\mathrm{m}(A_R\setminus A_P)$$

and if $t \in A_P \setminus A_R$ then $P(t) \geq \overline{X}$ so that

$$\int_{A_P\setminus A_R} P\ \mathrm{d}t \geq \overline{X}\mathrm{m}(A_P\setminus A_R)$$

The same is true if we replace P with Q so that

$$\begin{aligned}
E(R) &\leq \lambda\left(\int_{A_P} P\ \mathrm{d}t - \overline{X}\mathrm{m}(A_P)\right) + (1-\lambda)\left(\int_{A_Q} Q\ \mathrm{d}t - \overline{X}\mathrm{m}(A_Q)\right) \\
&= \lambda E(P) + (1-\lambda)E(Q)\ \blacksquare
\end{aligned}$$

Theorem 4.2 *Let $\mathcal{G}$ be a convex set of Lesbegue integrable functions over $[0,1]$, F_Q be a discrete probability distribution function, and m be the Lesbegue measure on $\Re$. For any $G \in \mathcal{G}$ define the function $F(G)$ by:*

$$F(G)(x) = \mathrm{m}(\{t \in [0,1] \mid G(t) \geq x\})$$

Then $\tilde{E} \mid \mathcal{G} \rightarrow \Re$, given by

$$\tilde{E}(G) = \int_0^\infty F(G) * F_Q(x)\ dx$$

is convex over $\mathcal{G}$.

Proof Note, in particular, that $F(G)$ is a probability distribution function with

$$\lim_{x\rightarrow\infty} F(G)(x) = 0 \ \ \textit{and} \ \ \lim_{x\rightarrow-\infty} F(G)(x) = 1$$

Since F_Q is a monotonic step function, label the steps, in increasing order, as $\{Q_1, \ldots, Q_N\}$. Let the step height at Q_n be p_n. Then:

$$\begin{aligned} F(G) * F_Q(x) &= \int_{-\infty}^{\infty} F(G)(x-z)\ \mathrm{d}F_Q(z) \\ &= \sum_{n=1}^{N} p_n F(G)(x - Q_N) \end{aligned}$$

Therefore

$$\begin{aligned} \tilde{E}(G) &= \int_0^\infty \sum_{n=1}^{N} p_n F(G)(x-Q_N)\ \mathrm{d}x \\ &= \sum_{n=1}^{N} p_n \int_0^\infty F(G)(x-Q_N)\ \mathrm{d}x \\ &= \sum_{n=1}^{N} p_n \int_{-Q_N}^{\infty} F(G)(x)\ \mathrm{d}x \\ &= \sum_{n=1}^{N} p_n \int_0^1 \max\{G(t)-Q_N, 0\}\ \mathrm{d}t \end{aligned}$$

from the definition of $F(G)$, so $\tilde{E}$ is convex from Theorem 4.1 and the facts that a finite sum of convex functions is convex, and all of the p_n are non-negative. ■

Theorems 4.1 and 4.2 show that any non-convexity, and multiple optima, bought into the objective function by $E(G, F_Q)$ are *only* artifacts of the approximation of E.

4.5 A Cumulant Approximation

This Section examines one possibility for the form of the functions E and F. We take into account the previous comments regarding the use of the inverse of the Contract Curve rather than the curve itself.

We require a representation, or an approximation, of the inverse of a partially filled C.C., which gives enough information to schedule another station and find the area under the curve. Recall that the inverse of a C.C. can be thought of as a probability distribution function, and the act of scheduling a station can be thought of as a convolution (see Equation 4.1). In Probability Theory, cumulants are used to calculate convolutions since the cumulants of the resulting curve are the sum, or difference, of the cumulants of the initial curves; these cumulants could be used, in a truncated Gram-Charlier Type A expansion (Cramér [3]), to approximate the curve (such an expansion is suggested by Electric Power Research Institute [6]).

Cumulants depend polynomially on the moments of a probability distribution function [3]. The first moment (and cumulant) is the mean of the distribution; the second cumulant is the variance (the square of the standard deviation). The k'th *moment* of a standard (increasing) probability distribution function, P, is given by:

$$\alpha_k = \int_{-\infty}^{\infty} x^k \, \mathrm{d}P(x) \tag{4.6}$$

In terms of a C.C., $G(t)$, since G^{-1} is decreasing, this means that

$$\begin{aligned} \alpha_k &= \int_{-\infty}^{\infty} x^k \, \mathrm{d}(1 - G^{-1})(x) \\ &= -\int_0^1 G(1-t)^k \, \mathrm{d}t \\ &= \int_0^1 G(t)^k \, \mathrm{d}t \end{aligned} \tag{4.7}$$

using the variable substitution $x = G(1-t)$. In particular the *mean* (the first cumulant) for any C.C., G, is exactly $m(G)$ from Equation 4.3. The variance (the second cumulant), σ^2, is given by;

$$\begin{aligned} \sigma^2 &= \int_0^1 (G(t) - m(G))^2 \, \mathrm{d}t \\ &= \int_0^1 G(t)^2 \, \mathrm{d}t - m(G)^2 \\ &= \alpha_2 - {\alpha_1}^2 \end{aligned} \tag{4.8}$$

In terms of the basis approximation, the moments are polynomial in the basis coefficients, since:

$$\begin{aligned}\alpha_k &= \int_0^1 (g_1B_1 + \cdots + g_NB_N)^k \, \mathrm{d}t \\ &= \int_0^1 \left(g_1^kB_1^k + kg_1^{k-1}g_2B_1^{k-1}B_2 + \cdots + g_N^kB_N^k\right) \, \mathrm{d}t \\ &= g_1^k \int_0^1 B_1^k \, \mathrm{d}t + kg_1^{k-1}g_2 \int_0^1 B_1^{k-1}B_2 \, \mathrm{d}t + \cdots + g_N^k \int_0^1 B_N^k \, \mathrm{d}t\end{aligned}$$

This, and the fact the cumulants depend polynomially on the moments, implies that the cumulants depend polynomially on the basis coefficients. These polynomial equations are not given here as they are tedious and not necessary for the current development.

The cumulants of the probability distribution function associated with the uncertain supply can be calculated *before* solution, since this function is fixed, and different, for each station (and possibly, for each week). In the case where the uncertain supply is given by a fixed probability of breakdown, the cumulants are quite easy to calculate, since the associated probability distribution function is given by:

$$F_Q(x) = \begin{cases} 0 & x \leq 0 \\ p & x \in (0, \overline{Q}] \\ 1 & x > \overline{Q} \end{cases} \tag{4.9}$$

where p is the probability of breakdown. Section 1.3 notes that all outage distributions will be discrete, so they will all have a similar form to Equation 4.9, but with more discrete pieces. This actually represents the probability that there is unserved power when the power level needed is x. The moments of this function are given by:

$$\alpha_{Q,k} = (1-p)\overline{Q}^k$$

The first four cumulants are given by:

$$\begin{aligned}\chi_{Q,1} &= (1-p)\overline{Q} \\ \chi_{Q,2} &= p(1-p)\overline{Q}^2 \\ \chi_{Q,3} &= p(1-p)(2p-1)\overline{Q}^3 \\ \chi_{Q,4} &= p(1-p)(6p^2-6p+1)\overline{Q}^4\end{aligned}$$

In filling a C.C. a polynomial calculation is required to derive the cumulants of the C.C.. Subsequently everything can be left in terms of cumulants, and only sums and differences are required. To calculate the load not yet satisfied in a partially filled C.C. one can approximate the associated probability distribution function, P, via a truncated Gram-Charlier expansion of Type A, namely:

$$P(x) = \Phi\left(\frac{m-x}{\sigma}\right) + \sum_{k=3}^{M} \frac{a_k}{k!}\Phi^{(k)}\left(\frac{m-x}{\sigma}\right) \tag{4.10}$$

where Φ is the *normal distribution function*, m and σ^2 are the mean and variance (first and second cumulants) of P, and the a_k are constants depending on P. Notice the use of $\frac{m-x}{\sigma}$, negative normalized variables, to convert to mean zero and standard deviation one, and to allow for P being decreasing, while the standard normal distribution function is increasing. Given

$$\phi^{(k)}(x) = (-1)^k H_k(x)\phi(x) \tag{4.11}$$

where $\phi = \Phi' = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$ is the *normal frequency function* and H_k is the k'th Hermite polynomial. Using the properties that Φ is an odd function, ϕ is even, and H_k is odd for odd values of k and even for even values of k, we obtain:

$$\begin{aligned} P(x) &= \int_{-\infty}^{\frac{m-x}{\sigma}} \phi(y)\ \mathrm{d}y + \sum_{k=3}^{M}(-1)^{k-1}\frac{a_k}{k!}H_{k-1}\left(\frac{m-x}{\sigma}\right)\phi\left(\frac{m-x}{\sigma}\right) \\ &= \int_{\frac{x-m}{\sigma}}^{\infty} \phi(y)\ \mathrm{d}y + \sum_{k=3}^{M} c_k H_{k-1}\left(\frac{x-m}{\sigma}\right)\phi\left(\frac{x-m}{\sigma}\right) \end{aligned} \tag{4.12}$$

where the c_k are given by Equation 4.13 and can be expressed as rational polynomials of the cumulants (from Equation 4.6 and the fact that cumulants depend polynomially on the moments), *viz*:

$$c_k = \frac{1}{k!}\int_{-\infty}^{\infty} P(x)H_k(x)\ \mathrm{d}x \tag{4.13}$$

The M, of Equation 4.12, is chosen to ensure a succinct approximation. Since we are only interested in this approximation so as to calculate unserved contracted load, the approximation needs to be *good* in terms of calculating the area under the resulting curve.

4.5.1 Contract Curve Filling Functions

The function $F(G, F_Q)$ described in Section 4.1, is given by:

$$F(G, F_Q)(x) = \int_{\frac{x-m}{\sigma}}^{\infty} \phi(y)\ \mathrm{d}y + \sum_{k=3}^{M} \tilde{c}_k H_{k-1}\left(\frac{x-m}{\sigma}\right)\phi\left(\frac{x-m}{\sigma}\right) \tag{4.14}$$

where the $\tilde{c}_k$ are given, in terms of the cumulants, by the same formula as for the c_k, with the cumulants given instead by

$$\tilde{\chi}_k = \chi_k + (-1)^k \chi_{Q,k}$$

We can think of $P(x)$, given in Equation 4.12, as being $F(G, 0)$, using the same abuse of notation, for the second argument, as used in Section 4.2 for E.

It may appear that the mean gives the area under the curve (see Equation 4.7); however, this includes *fictitious* negative generation. In terms of the approximation given in Equation 4.14, the unmet contracted load is given by:

$$\begin{aligned} E(G, F_Q) &= \int_0^\infty F(G, F_Q)(x)\ \mathrm{d}x \\ &= \underbrace{\int_0^\infty \int_{\frac{x-m}{\sigma}}^\infty \phi(y)\ \mathrm{d}y\ \mathrm{d}x}_{A} + \underbrace{\sum_{k=3}^{M} \tilde{c}_k \int_0^\infty H_{k-1}\left(\frac{x-m}{\sigma}\right)\phi\left(\frac{x-m}{\sigma}\right)\ \mathrm{d}x}_{B} \end{aligned}$$

Changing the order of integration for A, and using $\lim_{x\to\infty}\phi(x) = 0$, and Equation 4.11, gives

$$\begin{aligned} A &= \int_{-\frac{m}{\sigma}}^\infty (\sigma y + m)\phi(y)\ \mathrm{d}y \\ &= \sigma\left[-\phi(y)\right]_{-\frac{m}{\sigma}}^\infty + m\int_{-\frac{m}{\sigma}}^\infty \phi(y)\ \mathrm{d}y \\ &= \sigma\left(\phi(x_0) - x_0\int_{x_0}^\infty \phi(y)\ \mathrm{d}y\right) \end{aligned}$$

where $x_0 = -\frac{m}{\sigma}$ represents the zero prior to normalization. To simplify B we use $H_k(x)\phi(x) = -\frac{\mathrm{d}}{\mathrm{d}x}(H_{k-1}(x)\phi(x))$, obtainable from Equation 4.11, and the fact that $\lim_{x\to\infty} x^k\phi(x) = 0$ for any positive integer k, giving:

$$\begin{aligned} B &= \sigma\sum_{k=3}^{M} -c_k\left[H_{k-2}\left(\frac{x-m}{\sigma}\right)\phi\left(\frac{x-m}{\sigma}\right)\right]_0^\infty \\ &= \sigma\sum_{k=3}^{M} c_k H_{k-2}(x_0)\phi(x_0) \end{aligned}$$

It should be noted here that for this form the differential of E can be calculated explicitly, if somewhat tediously. This is extremely useful in terms of optimizing an objective which contains such terms, since, as noted in Section 3.11, filling the C.C. takes place in the objective function.

4.5.2 Transmission Revisited

At this point it may seem that the use of cumulants means that we cannot apply the penalty function to the transmission as explained in Section 4.3, since the transmission curves are not necessarily monotonic. However, Equation 4.7 indicates that the moments are defined for *any* integrable function. For a transmission curve these moments are exactly those for the curve when it has been "sorted" from highest transmission to lowest. Regardless of *how* the transmission function is sorted the area above, or below, any fixed level is the same. So it turns out that the moments, and hence cumulants, of a curve always exist and are meaningful.

4.5.3 Dispense with Piecewise Quadratics?

We have moved from an almost hopeless situation of explicitly calculating inverses of convolutions of inverses of piecewise quadratics, to evaluating polynomials multiplied by normal functions. It may even be that we could dispense with the need to evaluate a polynomial expression by maintaining the Contract Curves in the form of cumulants throughout. However, when discussing the use of cumulants for the transmission capacity constraints above, it was noted that, regardless of how a function is sorted, its cumulants are the same.

Consider: the two functions $G(t)$ and $G(1-t)$, over the interval $[0,1]$, have the *same* cumulants. Given a node at which $G(t)$ is arriving from a lossless arc, and $G(t)$ is leaving down another arc, and a second node at which $G(t)$ is arriving from a lossless arc, but $G(1-t)$ is leaving down another arc. A *correlated* interaction at the first node results in no load left at that node, whereas a correlated interaction at the second node will not, unless G is even about 0.5. Any interaction of cumulants will result in the same load left in both situations, regardless of the form of G.

Of course, if we allow an independent interaction for transmission, cumulants create no difficulties, but, as noted in Section 3.6, the interaction between electricity curves needs to be modelled as being correlated.

4.6 Why a Normal Approximation?

The next obvious question is, perhaps, "how many cumulants are needed to provide a reasonable approximation?" Actually this is *not* the correct question to ask, partly because of flexibility issues. The main reason for this is illustrated in Figure 4.5, which shows a truncated Gram-Charlier Type A approximation of an LDC, to 7 terms. Notice the oscillations at the extreme points: these disappear when only the first two cumulants are used, or, when the full Gram-Charlier Type A series is used (if it converges!). The oscillations make this approach unacceptable, since, if we do nothing about them, they create extraneous local optima in the objective function. It may seem that since these oscillations are shallow, and, since we are interested in the area under curves, they will not have much effect; but, note that an optimal solution will minimize the amount of non-supply (i.e. the contracted load remaining after every station at the node has been scheduled) and therefore, near the optimal solution, the area under the remaining curve will be small. Since we calculate this area from an approximation of G^{-1} (of Equation 4.2), when this approximation is negative the amount of non-supply can be negative, which would act like being paid the cost of non-supply, rather than paying this cost; in terms of the objective function this is worthwhile, and so the solution will "stick" at points where there is negative non-supply—an artifact of our approximation. Also, having alternative optima will lead to uncertainty in whether or not we are at a truly global optimal solution; this could have potentially disastrous results when one considers that the deterministic problem is only a small part of a stochastically extended model.

It may seem that it would be useful to retain the cumulant approach, but approximate the resulting curves where they begin to oscillate, in a monotonic fashion. However, it is at these end pieces of the curves that we calculate the non-supply for each thermal node, and, as this has already been approximated *twice*, a *third* approximation, which may be a reasonable approximation of the current approximation, will not necessarily be a good approximation of the actual scheduled C.C.. The *real* difficulty, however, lies in the fact that finding where the approximation should begin, is difficult. This is because it involves finding where the cumulant approximation of the curve is equal to some fixed value, ϵ, for *the first time*. If the presence of multiple solutions is not enough to deter proceeding with this approach, consider that the equation we seek to solve is to find the smallest x for which:

$$\int_{\frac{x-m}{\sigma}}^{\infty} \phi(y)\, \mathrm{d}y + \sum_{k=3}^{N} c_k H_{k-1}\left(\frac{x-m}{\sigma}\right) \phi\left(\frac{x-m}{\sigma}\right) = \epsilon$$

where m, σ^2, and the c_k are all dependent on variables from the model, H_k is the k'th Hermite polynomial, and $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ is the normal frequency function. If this was achieved numerically one could not then find the differential of the objective function—the existence of this differential is effectively necessary for a problem of this size to be solved in reasonable time in order that the stochastic extension then have a reasonable solution time.

The only option remaining seems to be to only use the first *two* cumulants, i.e. approximating the inverse by a normal distribution with like mean and standard deviation. As we are interested only in the area under part of the curve, this may not be too bad an approximation—it certainly seems to be a reasonable approximation for the LDC, as shown in Figure 4.6.

In this situation, $F(G, F_Q)$ and $E(G, F_Q)$ are given by

$$F(G, F_Q)(z) = \int_{x(z)}^{\infty} \phi(y)\ \mathrm{d}y$$

and

$$E(G, F_Q) = \sqrt{\sigma(G)^2 + \sigma(F_Q)^2}\left(\phi(x(0)) - x(0)\int_{x(0)}^{\infty} \phi(y)\ \mathrm{d}y\right) \tag{4.15}$$

where x is given by

$$x(z) = \frac{z - (m(G) - m(F_Q))}{\sqrt{\sigma(G)^2 + \sigma(F_Q)^2}}$$

To try to improve the approximation, we could use some sort of skew factor, a role played by the third cumulant in the cumulant approximation. This could be achieved by having effectively *two* standard deviations, one associated with the resultant curve over $[0, 0.5]$, and the other associated with the same curve over the interval $[0.5, 1]$. Such splitting could become more generalized, having multiple splits, each with its own mean and standard deviation. Unfortunately, in this case the method of performing the convolution (required for scheduling thermal stations) becomes difficult. It was the ease of performing this convolution in terms of the cumulants that lead to a cumulant approach initially.

4.6.1 Contract Curve Corners

In the meantime we will retain the normal approximation of the inverse of a Contract Curve. This also means (as we only use the mean and standard deviation) that we are using a normal approximation to the breakdown probability, so it would probably be better to use a *direct* normal approximation of these rather than using a normal approximation of the current discrete approximation. The normal approximation, of course, breaks down when the standard deviation is zero, as is the case when the C.C. is constant. This implies that E and the objective function, have corners; Figure 4.7 shows an example of this.

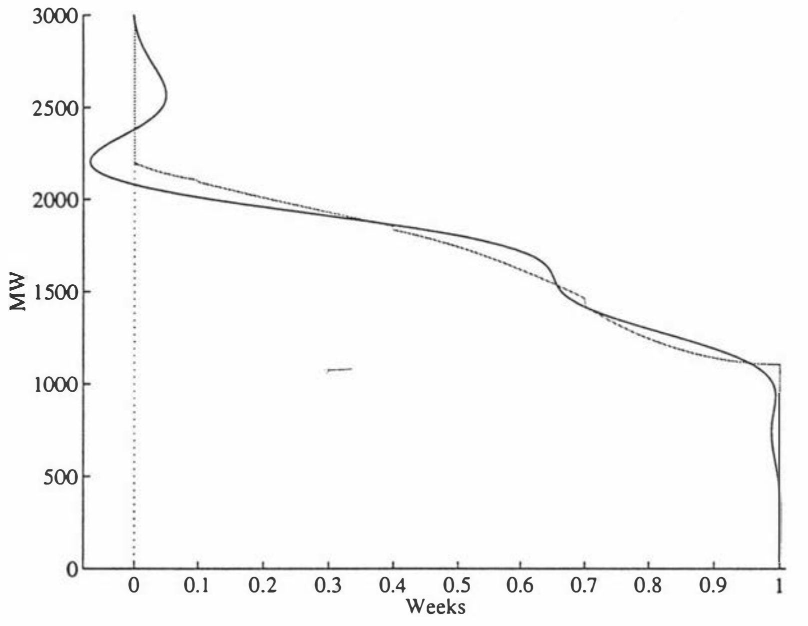


Figure 4.5: A truncated Gram-Charlier Type A expansion of an LDC

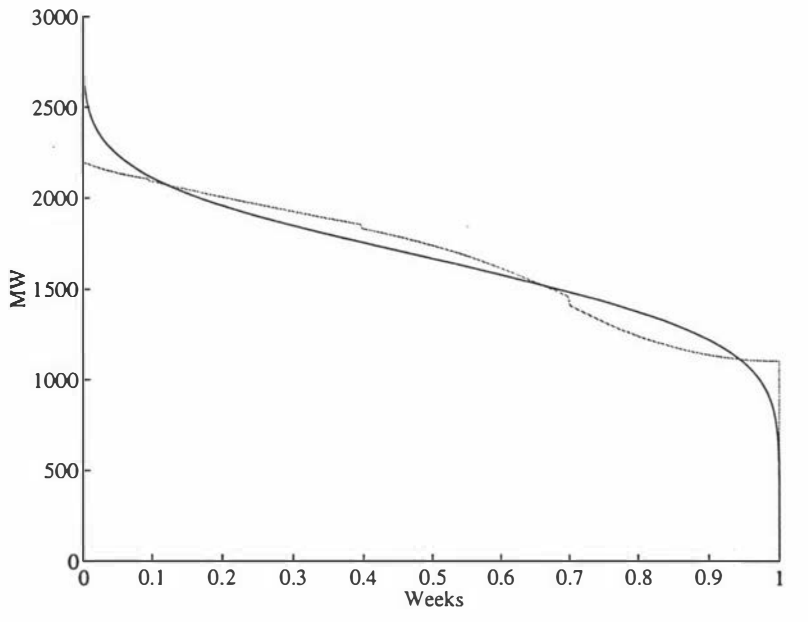


Figure 4.6: A normal approximation of an LDC

If these corners create a difficulty, we can bound solutions away from them by ensuring that one (or more) of the coefficients of the basis elements are non-zero for any Electricity Curve represented in the objective function. The problem with this is that, since we represent over-capacity transmission in arcs as a penalty in the objective function in this way, we will have some non-zero transmission in *every* arc, which, due to arc loss, will increase the load.

4.6.2 Convexity of the New E

Recall from Section 4.4 that the convexity of the objective function depends on the convexity of E. The reason for coveting a convex objective function is that it will mean that all locally optimal solutions are globally optimal, and most Mathematical Program solvers are robust on convex problems; we desire robustness in our deterministic model as it is to be the basis of an extended stochastic model, and may need to be "solved" many times to obtain a solution to the stochastic model.

It turns out the current E, given by Equation 4.15, is convex, as is shown in the following results.

Lemma 4.3 *Let* $\{B_1, \ldots, B_N\}$ *be a basis of integrable functions over* $[0,1]$, *and* $G = [g_1, \ldots, g_N] \in \Re^N$. *Put*

$$L(t) = g_1 B_1(t) + \cdots + g_N B_N(t)$$

and define σ *by:*

$$\sigma(G)^2 = \int_0^1 L(t)^2 \, dt - \left(\int_0^1 L(t) \, dt \right)^2$$

Then σ *is non-negative and convex over* $\mathcal{G}$, *and the kernel of* σ *is contained in one-dimensional subset of* G.

Proof Notice σ^2 is just the variance of G (as given by Equation 4.8), and we can write

$$\sigma(G)^2 = \int_0^1 \left(L(t) - \int_0^1 L(t) \, \mathrm{d}t \right)^2 \, \mathrm{d}t$$

so that σ is just a translated norm and, hence, both convex and nonnegative. Also $\sigma(G) = 0$ if, and only if

$$L(t) = \int_0^1 L(t) \, \mathrm{d}t \quad \forall t \in [0,1] \tag{4.16}$$

i.e. when $L(t)$ is constant. Since $\{B_1, \ldots, B_N\}$ is a basis of functions then either Equation 4.16 holds nowhere, or describes a one-dimensional subspace of $\text{span}\{B_1, \ldots, B_N\}$. ∎

Theorem 4.4 *Let functions m and σ be defined over some open, convex, N-dimensional set, $\mathcal{G}$, with $N > 2$, where m is an affine function and σ is a non-negative convex function, the kernel of which is closed relative to $\mathcal{G}$ and contained within some one-dimensional set. For any $G \in \mathcal{G}$ define*

$$E(m, \sigma) = \int_0^\infty \int_{\frac{x-m}{\sigma}}^\infty \phi(y)\ dy\ dx \tag{4.17}$$

when $\sigma \neq 0$ and

$$E(m, 0) = \max\{m, 0\} \tag{4.18}$$

Then $E(m(G), \sigma(G))$ is convex, with respect to G.

Proof: Put $G = [g_1 \ldots g_N]$, where $N > 2$. First consider the case where $\sigma(G) \neq 0$. Using the change of variable, $x \leftarrow \frac{x-m(G)}{\sigma(G)}$, in Equation 4.17 gives

$$E(m(G), \sigma(G)) = \sigma(G) \int_{\left(-\frac{m(G)}{\sigma(G)}\right)}^\infty \int_x^\infty \phi(y)\ \mathrm{d}y\ \mathrm{d}x$$

Putting $x_0(G) = -\frac{m(G)}{\sigma(G)}$, and writing $E = E(m(G), \sigma(G))$, we get

$$\begin{aligned}\frac{\partial^2 E}{\partial g_i \partial g_j} = &\frac{\partial^2 \sigma(G)}{\partial g_i \partial g_j} \int_{x_0(G)}^\infty \int_x^\infty \phi(y)\ \mathrm{d}y\ \mathrm{d}x + \sigma(G) \frac{\partial x_0(G)}{\partial g_i} \frac{\partial x_0(G)}{\partial g_j} \phi(x_0(G)) \\ &- \left(\frac{\partial \sigma(G)}{\partial g_i} \frac{\partial x_0(G)}{\partial g_j} + \frac{\partial \sigma(G)}{\partial g_j} \frac{\partial x_0(G)}{\partial g_i} + \sigma(G) \frac{\partial^2 x_0(G)}{\partial g_i \partial g_j} \right) \int_{x_0(G)}^\infty \phi(x)\ \mathrm{d}x\end{aligned}$$

Using the definition of $x_0(G)$ and changing the order of integration in the double integral gives

$$\frac{\partial^2 E}{\partial g_i \partial g_j} = \left(\frac{\partial^2 \sigma(G)}{\partial g_i \partial g_j} + \sigma(G) \frac{\partial x_0(G)}{\partial g_i} \frac{\partial x_0(G)}{\partial g_j} \right) \phi(x_0(G)) \ ,$$

so the Hessian of $E(m(G), \sigma(G))$, $H(E)$, can be written as $H(E) = A + v^T v$, since both σ and ϕ are positive functions. By the convexity of σ, A is positive semi-definite. Now

$$\begin{aligned} xH(E)x^T &= xAx^T + x(v^T v)x^T \\ &= xAx^T + (vx^T)^T (vx^T) \\ &= xAx^T + \|vx^T\|^2 \geq 0 + 0 = 0 \end{aligned}$$

and so $H(E)$ is positive semi-definite. Hence $E(m(G), \sigma(G))$ is convex over any open set where $\sigma(G) \neq 0$.

For the case where σ is allowed to be zero, notice, from Equations 4.17 and 4.18, that $\lim_{\sigma \to 0} E(m, \sigma) = E(m, 0)$, so that, as m is linear and σ is continuous, $E(m(G), \sigma(G))$ is also continuous.

For each $G, H \in \mathcal{G}$ define

$$I(G,H) = \{\lambda G + (1-\lambda)H \mid \lambda \in [0,1]\}$$

Calling $\mathcal{K}$ the kernel of σ for every $G, H \in \mathcal{G}$ one of the following holds

Case (a): $I(G,H) \cap \mathcal{K} = \emptyset$,

Case (b): $I(G,H) \subset \mathcal{K}$,

Case (c): $I(G,H) \cap \mathcal{K} = \{D\}$ $\exists D \in \mathcal{K}$.

If Case (a) holds then $\mathcal{K}$ and $I(G,H)$ are mutually exclusive closed convex sets of $\mathcal{G}$, so there exists an open convex set $C \subset \mathcal{G}$, such that $I(G,H) \subset C$, and $C \cap \mathcal{K} = \emptyset$. In this case we have already shown that if

$$\lambda \in (0,1) \quad \text{and} \quad D_\lambda = \lambda G + (1-\lambda)H$$

then

$$E(m(D_\lambda), \sigma(D_\lambda)) \leq \lambda E(m(G), \sigma(G)) + (1-\lambda)E(m(H), \sigma(H)) \tag{4.19}$$

If Case (b) holds then we see that Equation 4.19 holds from the definition of E over $\mathcal{K}$, and the fact that the pointwise supremum of two convex functions is a convex function.

For Case (c), if $D \in \{G, H\}$, *w.l.o.g.* $D = H$. So there is $\{D_n\} \subset I(G,H) \setminus \{H\}$ such that $D_n \to H$, and Equation 4.19 holds where H is replaced by D_n for each n. Hence, by the continuity of $E(m(G), \sigma(G))$, Equation 4.19 holds for G and H. If $D \notin \{G, H\}$, there is $\{G_n\} \subset \mathcal{G} \setminus \mathcal{K}$ such that $G_n \to G$, and $I(G_n, H) \cap \mathcal{K} = \emptyset$, since $\mathcal{K}$ is contained in a subset of dimension 1 of $\mathcal{G}$, which is of dimension at least 3. The convexity of E therefore follows from the continuity of $E(m(G), \sigma(G))$ and the case where σ is nonzero. ∎

Changing the order of integration in Equation 4.17 transforms it to the form of Equation 4.15. By replacing the function m by $m(G) - m(F_Q)$, and σ by $\sqrt{(\sigma^2 + \sigma(F_Q)^2)}$, in Theorem 4.4, (recalling F_Q is fixed) we obtain the convexity of $E(G, F_Q)$.

4.6.3 Hydro Creases

Recall from Section 4.2 that the non-supply from a hydro station's C.C. is given by Equation 4.4, which implies that the objective function may be non-differentiable where Equation 4.5 holds. We will call this phenomena a "crease" in the objective function. Having now a fixed formula for E we can investigate these creases further. Specifically the approximation of G^{-1} is given by

$$G^{-1}(x) \approx \int_{\left(\frac{m(G)-x}{\sigma(G)}\right)}^{\infty} \phi(z)\ \mathrm{d}z$$

so $E(G, \overline{H})$ is given by

$$\begin{aligned} E(G,\overline{H}) &= \int_{\overline{H}}^{\infty} \int_{\left(\frac{m(G)-x}{\sigma(G)}\right)}^{\infty} \phi(z)\ \mathrm{d}z \\ &= \sigma(G)\phi\left(\frac{m(G)-\overline{H}}{\sigma(G)}\right) + \left(\overline{H} - m(G)\right) \int_{\left(\frac{m(G)-\overline{H}}{\sigma(G)}\right)}^{\infty} \phi(z)\ \mathrm{d}z \end{aligned}$$

Since the term $E(G, \overline{H})$ has no dependence on the variable H, there will be a discontinuity in the differential of the objective here (if only in the direction of the variable H). The crease created will be sharp—the differential with respect to H will be one, on one side, and zero, on the other. There will also be discontinuities in the gradient in the direction of the other variables, as m is linear but E is not; the values of the variables of G where Equation 4.5 holds are dependent on the value of H, so the gradient of E will vary for different G, but the gradient of m will not. Figure 4.8 gives an example of this crease over two dimensions.

We would like to smooth off these creases whilst maintaining the convexity of the objective function. This means either determining exactly where the crease occurs, or applying some approximation based only on the relative values of $m(G) - H$ and $E(G, \overline{H})$. To achieve the former we need to be able to solve *analytically* equations of the form:

$$E(G,\overline{H}) + H - m(G) = \delta \tag{4.20}$$

for G and H. The form of E makes this impractical.

An approximation based *only* on the *relative* values of $m(G) - H$ and $E(G, \overline{H})$, which is required to be *convex* everywhere, can be written as;

$$f(G,H) = \begin{cases} m(G) - H & \text{when } m(G) - H \geq E(G,\overline{H}) + \delta_1 \\ E(G,\overline{H}) & \text{when } E(G,\overline{H}) \geq m(G) - H + \delta_2 \\ A(G,H) & \text{otherwise} \end{cases}$$

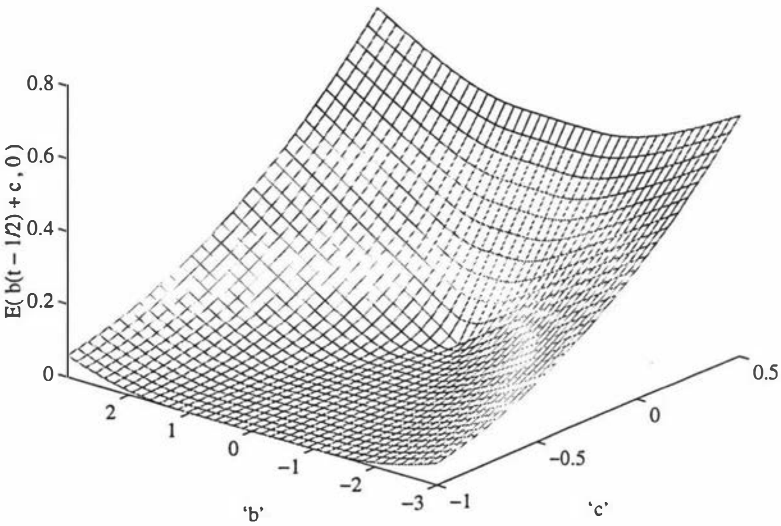


Figure 4.7: Area above zero for the normal approximation of $b(t-\frac{1}{2})+c$ on $[0,1]$

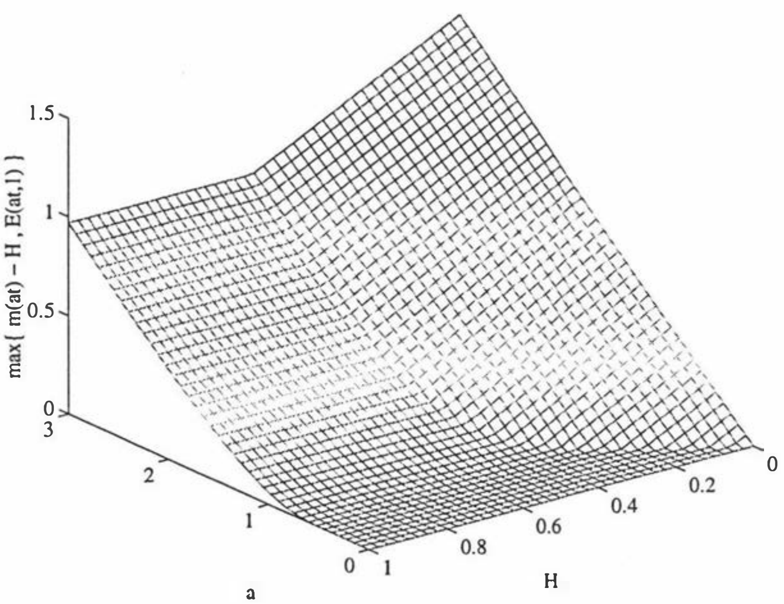


Figure 4.8: Crease in hydro station non-supply over variables H and $G(t)=at$

where the function A must pass through the points where Equation 4.20 holds for δ being, variously, δ_1, 0, and δ_2. Consider any subspace over which $m(G) - H$, and $E(G, \overline{H})$ are both linear; the subspace $\{(G, H) \mid G = 0\}$ is one such subspace. On this subspace, the only possible convex approximation which passes through the points where Equation 4.20 holds for $\delta = 0$ in this subspace, is the function $\max\{m(G) - H, E(G, \overline{H})\}$. Consequently, we do not change the function on this subspace, so a crease remains. Therefore we can not eliminate creases, in this way, from the objective function in the current form and keep it convex.

4.7 The *really* bad news!

It was claimed in Section 4.6 that the normal distribution function was a reasonable approximation for the LDC's. Unfortunately this is not the only Electricity Curve requiring approximation. When approximating Contract Curves, the normal distribution function is used to determine the amount of power which can not be used—in particular, the amount of power *above* the total capacity of the stations present. Since we are using a normal approximation to the inverse of the C.C., there will *always* be some non-supply, as the normal distribution curve is positive everywhere. We are interested in how large this amount of non-supply can become.

Consider a situation where a Contract Curve is being filled by only one totally reliable station. If the station is required by the model to generate at capacity for only part of the week in order to minimize non-supply, the C.C. at this node will want to mimic this required generation pattern of the station. The situation where the station generates at full capacity for 75% of the week, and has no generation for the rest of the week, is shown in Figure 4.9.

This shows a substantial amount of fictitious non-supply from a Contract Curve which should produce none. To illustrate the ramifications of this problem, Figure 4.10 shows the percentage of actual generation wrongly non-supplied for a C.C. having a fixed non-zero level on the interval $[0, r]$ and a zero level for the rest of the period (black); and for a similar C.C. with the generation linearly decreasing to zero at the end of the period (green). Since the cost of non-supply is in the order of ten times the cost of the most expensive thermal station, this fictitious non-supply *is* significant. Therefore, stations are penalized for having constant generation rates for only part of the week, especially if the generating range is quite large. In

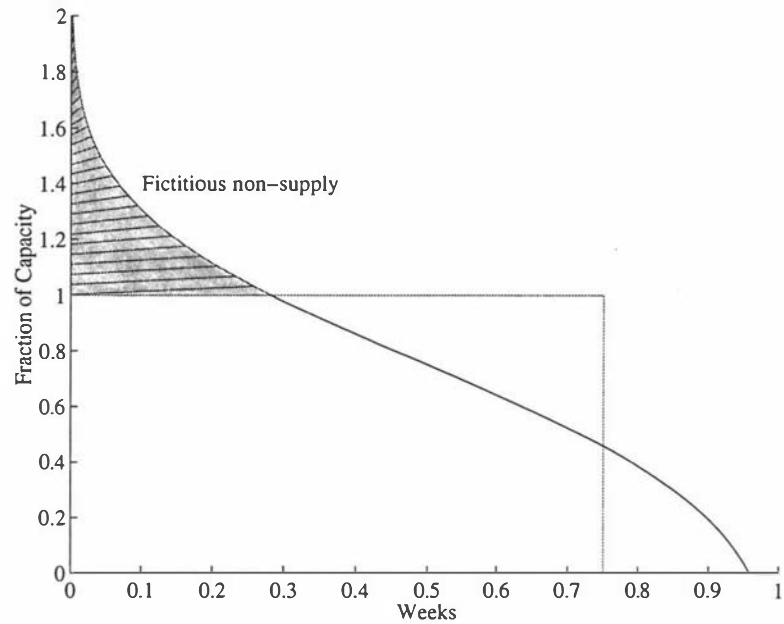


Figure 4.9: A C.C. and its corresponding normal approximation

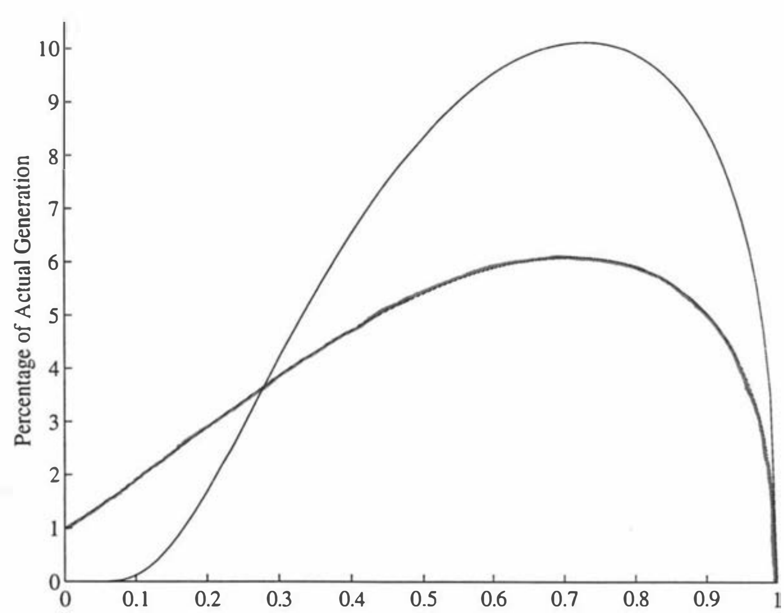


Figure 4.10: Percentage of actual generation wrongly non-supplied

general, the generation over the week is decoupled, in that the generation at one time during the week should not be affected by the generation at any other time, so the added structure is artificial.

Notice also that the worst point on Figure 4.10 is near the value of $r = .7$, which happens to be one of our breakpoints (we would need one near this anyway), so it is not unlikely for a station to want to generate at peak capacity (or some other fixed level) on the interval $[0, 0.7]$ and then drop off to zero (say) on the interval $[0.7, 1]$. Unfortunately this model would discourage such a solution.

When one considers the similar graphs for transmission, the situation is worse, since in transmission the negative transmission is also penalized. Figure 4.11 shows this for a transmission function at full capacity for a proportion r of the week, and at zero for the remainder of the week. Such a model will try to (incorrectly) encourage flat generation and transmission at all stations and along all arcs.

4.8 Discussion

It appears the model we have developed is riddled with superfluous unwanted structure and aspects which make it difficult to model. Such a model is unacceptable as the deterministic base of a large-stochastic extension. It also appears that under the current skeleton model for the system there are no "quick fixes". It may be that the model could be linearized, via a coarse linear approximation of the various elements, from its current form, with the hope that this would remove much of the extraneous and difficult structure currently present. Unfortunately, such a brutal linearization would bear little resemblance to the system we seek to model, being a coarse approximation of an already bad approximation.

The only options appear to be to completely re-model the system from scratch, or to change some of the desired structure in the skeletal model, so as to remove the difficulties which arise from there. It seems that the majority of the difficulties arise from the need to find the inverse of the Contract Curves, which, in turn, result from the filling of the C.C..

Actually the situation is not as grim as it may appear. We have only investigated one information structure for calculating E, namely cumulants, and from this information structure, only two approximations. It may be that there is a better information structure than cumulants to use, or that there are cumulant expansions without the problems associated with the Gram-Charlier Type A expansion. Such an investigation should possibly attract more attention than is given here; however the rewards of such an investigation will probably be quite limited, in terms of the model being developed, as the problem of the creases brought into the objective function via the hydro station generation will still be present. It is the investigation of the elimination of these creases from the objective function (as discussed in the next Chapter) which leads, naturally, to a solution to many of the other dilemmas discussed in this Chapter.

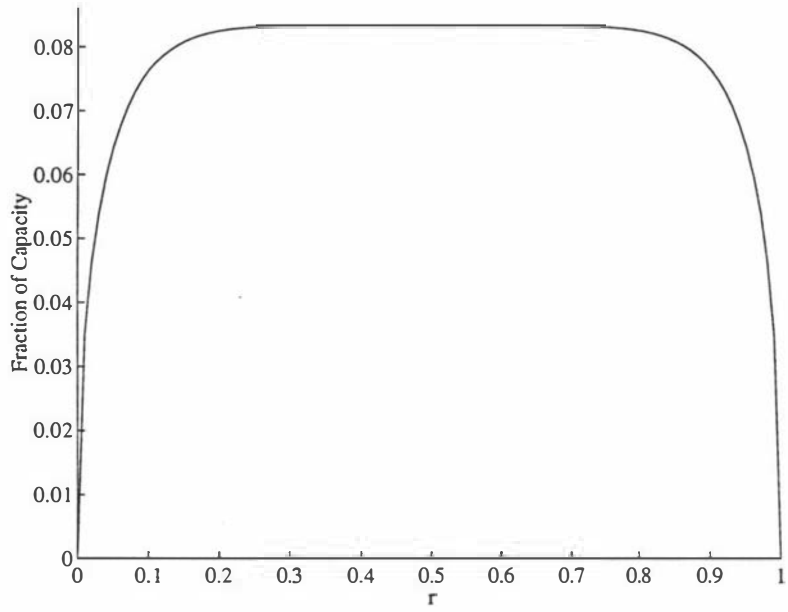


Figure 4.11: The fraction of actual transmission wrongly penalized

Chapter 5

The Model

Various approximations examined in the previous Chapter each appeared to impose unwanted structure to the problem, in such a way as to make the resultant model unsuitable for its desired purpose. This Chapter examines one way to suppress the unwanted structure and hence develop a more suitable model.

In seeking a method for removing such difficulties, it is *better* to re-model the system in such a way that the difficulty does not arise, rather than to *approximate* the difficulty out of the model. Of course the former is not always possible, without creating more difficulties, and approximations applied to the model can often motivate better methods for modelling the system.

5.1 Removing the Hydro Crease

Initial motivation for the *actual* change made came from examining ways of eliminating the "crease" induced in the objective function by the hydro station generation; this is a natural way to introduce it.

Recall, from Section 4.2, that the non-supply remaining after a single hydro station fills is own Contract Curve, G, is given by:

$$\max\{m(G) - H, E(G, \overline{H})\}$$

where H is the release from the hydro reservoir and $\overline{H}$ is the capacity of the station. This feature is not *just* an artifact of the approximations used, and becomes difficult to handle when optimizing the model. Given that the deterministic model being developed is to be used as a basis for a full stochastic model, we would like to eliminate as many difficulties in the computation of a solution as possible.

In Section 4.6 we examined the possibility of "smoothing" the corners; however, this proved elusive, and it also appears to have little justification in practice. By examining what the crease represents, we can provide a better opportunity to eliminate its effects.

The crease is the *only* link between the geographic networks and the hydro waterflow networks; since the waterflow networks link the geographic networks of each week, this crease is extremely important. If we approximated the crease by *assuming* we are always on one side of it, we effectively de-couple the geographic and waterflow networks (and hence the weeks), and the problem becomes very simple to solve, but of little use. It is therefore important that we model this connection well.

One side of the crease ($E(G, \overline{H})$) represents the situation where there is some *spill* at the hydro station, while the other side ($m(G) - H$) occurs when there is none. It may be that if we include the spill explicitly we can remove the crease. Using W to represent the spill (the total *generation* is, therefore, given by $H - W$) would require the inclusion of a constraint to ensure that we do not spill more than we release, i.e. $H \geq W$, and another to ensure that all of $H - W$ can be generated given the hydro stations C.C., i.e.

$$H - W \leq m(G) - E(G, \overline{H}) \tag{5.1}$$

which is a non-linear constraint.

The right-hand side of Equation 5.1 represents the maximum generation allowed for by the station's C.C. and its capacity. The non-linear term, $E(G, \overline{H})$, represents the area under the Contract Curve above the capacity of the station. Since each hydro station has its own C.C., this part of the curve could be transferred to the non-supply curve at this node, retaining the same solution (within the limits of the piecewise quadratic approximation). The only difference in solution is that, instead of having $E(G, \overline{H})$ of non-supply in the hydro station's C.C., it would be in the non-supply curve; we are, afterall, merely interested in the *amount* of non-supply.

Removing the area above the station's capacity is exactly the same as requiring that the station's C.C., G, is constrained below the station's capacity, $\overline{H}$, and, since the C.C. is decreasing, this can be achieved by requiring

$$G(0) \leq \overline{H}$$

which is a linear constraint. If this is included, Equation 5.1 becomes:

$$H - W \leq m(G) \tag{5.2}$$

a linear constraint. The non-supply remaining after scheduling the station is now given by $m(G) - H + W$.

It was noted in Section 4.2 that, if more than one station fill the same C.C. (as a combined station), some of the stations may use the spill at other stations for generation. Since we model river chains as a single amalgamated station, this allows explicit expression of the assumed approximations in this amalgamation.

Consider a simple hydro chain with just one reservoir, and some hydro stations in series down the river (see Figure 5.1). The assumption is made that the uncon-

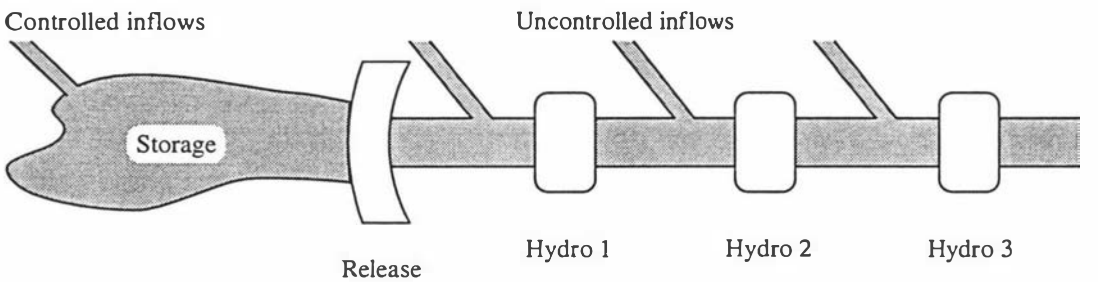


Figure 5.1: A simple hydro river chain

trolled inflow arrives in fixed proportions above each station. The hydro stations are scheduled as a single station with capacity equal to the sum of the capacities of all of the individual stations, and release equal to the sum of the individual flows (which is just the release from the reservoir). In other words, the approximation used is equivalent to approximating all river chains as simple chains, and to assuming complete correlation in the uncontrolled inflows within this chain, together with the added assumption that stations with spare capacity can generate using the *spill* of other stations. There is also the implicit assumption that the storage for individual stations is not used for week-to-week storage.

We can, in theory at least, use more sophisticated networks to approximate hydro river chains. This would be most useful for modelling controlled canals between river chains, linked reservoirs and other controllable phenomena, although by including too much detail here we would most likely make the model computationally intractable. Another reason for inclusion of such detail is to better model conservation constraints of the minimum and maximum flows along various parts of the river network.

For a simple chain it is possible to model conservation constraints such as minimum and maximum flow constraints. The Contract Curves now act as generation curves. Since the time dimension is effectively sorted we can apply constraints only on the total release, or, as minimum and maximum generation levels. The problem with using only minimum and maximum generation levels to approximate the corresponding release levels is that they *ignore* spill. For maximum release level, $\overline{F}$, we can include the constraint in two forms, namely

$$G(0) \leq \overline{F} \text{ and, } \quad H \leq \overline{F} \tag{5.3}$$

to effectively limit the spill to be less than the slack in the "total release" form (see Figure 5.2). For a minimum release level, $\underline{F}$, we require a constraint of the form

$$G(1) \geq \underline{F} - W \tag{5.4}$$

where the " $-W$" in this equation allows the use of constant spill to augment some of the minimum level release. Equations 5.3 and 5.4 actually refer to flows from the mouth of the river. For constraints on the flow at an arbitrary point in the *simple* river chain, we need to determine the proportion, α, of the uncontrolled inflows which arrive *below* this point, and instead use the following constraints:

$$G(0) \leq \overline{F} + \alpha\tilde{U} \text{ and, } \quad H \leq \overline{F} + \alpha\tilde{U}$$

(which correspond to *also* requiring some of the $\alpha\tilde{U}$ flow be at a constant level) and:

$$G(1) \geq \underline{F} - W + \alpha\gamma\tilde{U} \text{ and, } \quad H \geq \underline{F} + \alpha(1-\gamma)\tilde{U}$$

where γ is the fraction of the uncontrolled inflow which is deemed to arrive at a constant rate. These constraints correspond to ensuring that there is enough flow above $\underline{F}$ at the bottom of the river chain to allow for the extra $\alpha\tilde{U}$ which arrives below the point in the river chain where this constraint is being applied. Most of these types of constraints have $\alpha \in \{0, 1\}$, i.e. are at either end of the river.

After removing *some* of the non-supply from the hydro station's C.C. it may seem that we could (and probably should) remove *all* of the non-supply from the C.C.. This could easily be done by converting Equation 5.2 into an equality constraint. In this case we would be explicitly treating all of the non-supply consistently, by requiring that it is present *only* in the non-supply curve. When we use Equation 5.2 as it stands, however, we are allowing a little more flexibility at

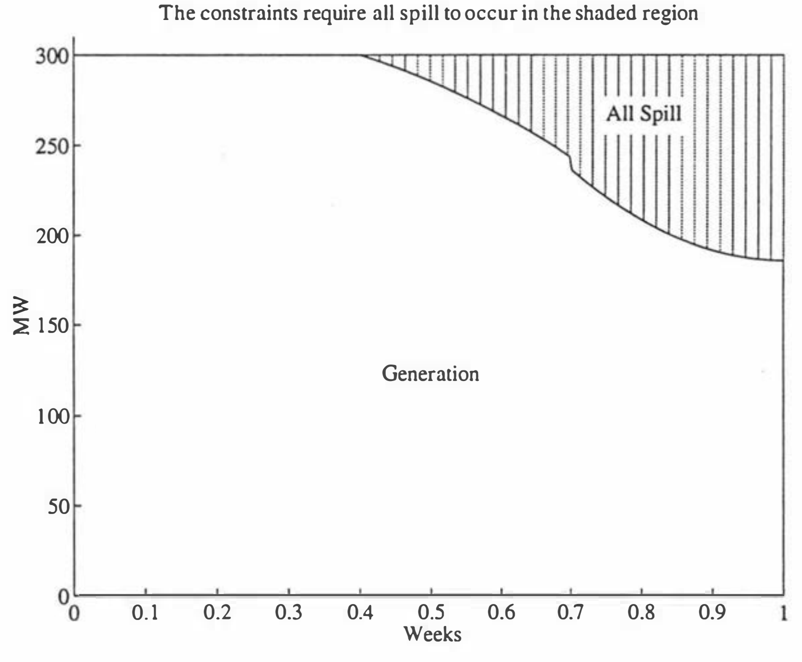


Figure 5.2: A maximum level release constraint applied in both forms

Table 5.1: Connections between the release and total generation

Constraint	Non-supply	Spill	Comments
$H - W = m(G)$	none	W	Less flexibility in the structure of non-supply.
$H - W \leq m(G)$	$m(G) - H$	W	Non-supply is split between the C.C. and non-supply curve, leading to possible alternative solutions.
$H \geq m(G)$	none	$H - m(G)$	No explicit spill variable, and less flexibility in the structure of non-supply.

these nodes, as it is no longer required that all non-supply must be in terms of the approximation. Figure 5.3 shows an example of this in which the first graph shows the desired non-supply form (the shaded area), and the second graph shows the closest approximation allowable in the non-supply curve due to the basis used for all Electricity Curves.

At this point it should be noted that if, instead, we require the constraint $H \geq m(G)$, then we do not need to include a spill variable. This means we can include the connection between the release H, and the total contracted load, $m(G)$, in three equivalent ways, each of which endows the model with different properties; these are listed in Table 5.1. For flexibility, we leave open the exact modelling of this constraint. Notice that, in all three cases, the equation for determining non-supply, and hence the hydro station's contribution to the objective function, is linear. For the working model, we employ the constraint $H - W \leq m(G)$.

In light of the fact that there is no need to *explicitly* define spill to remove the hydro station crease from the objective, it is apparent that the benefits derive from the removal of the above-capacity non-supply from the hydro station's C.C.. It may be that similar benefits could be gained by applying this type of transformation to the thermal station C.C.'s; we now investigate the effects of such an application.

5.2 Thermal Contract Curves

Consider a thermal station with its own Contract Curve, G, fuel cost c_Q, and probability distribution associated with failure, F_Q. The cost of generation and non-supply for this station is:

$$c_Q m(G) + (c_{NS} - c_Q)E(G, F_Q)$$

where c_{NS} is the cost of non-supply (or re-supply, see Section 5.2.1). If F_Q represents only a single probability, p, of total plant failure, the term $E(G, F_Q)$ is exactly given by:

$$E(G, F_Q) = (1-p)E(G, \overline{Q}) + pm(G)$$

where $\overline{Q}$ is the capacity of the station (see Section 4.1). This means that if we perform the same transformation as in Section 5.1, and transfer the above capacity non-supply to the non-supply curve at this node, the cost of generation and non-supply for this C.C. becomes

$$c_Q m(G) + p(c_{NS} - c_Q)m(G) = ((1-p)c_Q + pc_{NS})m(G)$$

which is linear. It also has the added advantage that we no longer need an approximation for calculating E in this case.

Of course, a similar trick applied to a C.C. for more than one station will not work, since, when scheduling the *first* station, there will be contracted load above its capacity. However, it does open up the possibility of allowing each station to have a C.C. of its own; if we do this, we transform those parts of the objective function dealing with filling Contract Curves into being linear. This then implies that the only part of the objective function still with non-linearities would be the penalty on over-capacity transmission.

The advantage of taking this approach is enormous. We move from a situation where we have no *realistic* approximation for an apparently non-linear energy function, to one where this function can be calculated linearly. The disadvantage is that we can no longer use stations at the same node to cover stations which may break down; this means that we are effectively using the cost of completely reliable generation as an approximation to the expected cost of unreliable generation. The cost of reliable generation provides a lower bound for the expected cost of unreliable generation. Obviously the advantages outweigh the disadvantages.

Recall from Section 2.5 that the non-linear method of filling is used to calculate the expected value of generation to meet the load. This is not quite what it is used for here, as we calculate the expected value of meeting the contracted load using only stations *at one node*! However, in the geographic network chosen for the working model there is only *one* node which has more than one thermal station at it, and it has only two! This means that we were implicitly using an approximation to the expected cost of unreliable generation, where load which was to be met by a plant which fails, is just not met at all. This, therefore, may not provide a very good approximation to the expected cost of generation, but it does provide an upper bound for the expected cost of generation.

The non-linear method of filling an LDC also appears to be inconsistent in the way it treats various stations. Stations at the same node can be used in case another station at that node breaks down; however, for single stations, a breakdown means the load is non-supplied, and so is penalized at the cost of non-supply. This would seem to suggest that stations at nodes with more than one station are *more* reliable than those at nodes with only one, which is extraneous structure we should seek to deter.

Hence, if we *are* willing to accept one of the approximations, then the linear version is better; if we are *not* willing to accept an approximation, then neither will suffice. Of course, we could always consider solving the network for every "state" of each station, in which case the linear approximation is fine. This is, however, computationally infeasible (with one possible exception, mentioned below). For our model we choose to accept the linear approximation.

5.2.1 Approximations to Handle Breakdowns

Unfortunately, there is no elegant representation (as opposed to approximation) of station breakdowns when the transmission network is included. The reason for this (as stated in Chapter 3) is that changing the state of a station could also change the optimal distribution of the Contract Curves, which requires solving a Generalized Network.

One approximation, already mentioned, is to use a cost of re-supply, rather than the cost of non-supply, when filling Contract Curves. To obtain this cost of re-supply we could use the cost of the next-most-expensive thermal station, modified by the power loss incurred in transmission to this node and, possibly, a factor to allow for the probability that this station is already being used. This would effectively mean that we are defining a fixed cost for covering the load met by this station, if it were to break down. Such an approximation may seem more reasonable when it is taken into account that, at the time of failure, stations are not usually scheduled so as to *optimally* meet the load, in terms of the global situation.

We could also use past data to estimate the *cost* to the system of a station breaking down, and use this for the cost of re-supply. It would appear that this would be more effective; however, this cost is dependent on the current load and the amount of hydro station generation in the system for that week, and so a fixed figure would be difficult to obtain, and often be inapplicable.

Another possibility is to consider re-solving *only* the first week for every "state" of each station since, after the first week, the amount of stored water (and hence the hydro station generation) is uncertain, so that using a "cost of re-supply" approximation is reasonable. Also, in this first week, we need not consider *every* possible state of the generation system. For instance, the situation where every thermal station fails is unlikely in the extreme, and, if it *did* happen, emergency steps (which cannot be modelled) would be taken to minimize the effects. We could therefore presume that such situations need not be catered for explicitly, unless this is easily done.

We could also take a scenario approach to station failures, in which the "scenarios" chosen would define which stations had failed. It would seem reasonable to include those scenarios which are most likely to happen, along with a few which are seen as including important events to consider.

For the working model we use a cost of re-supply equal to the cost of non-supply. However, for flexibility, we consider having differing re-supply costs, as well as using a scenario approach in the first week.

5.3 Transmission

As mentioned above, the only non-linearities left in the objective function are those associated with the penalty applied to over-capacity transmission. If we enforced the transmission line's capacity, this would not occur. However, it was the difficulty in enforcing this capacity constraint, in light of the fact that the transmission curves are not necessarily decreasing, that led us to the use of objective penalties.

It would be better to neither force the transmission to be within capacity, nor apply a penalty to over-capacity transmission, but to use a non-linear increasing line loss which could either explicitly or implicitly enforce the capacity; Figure 5.4 shows an example of an implicit and explicit capacity through line loss. Of course, the problem with this approach is that it introduces many (about 10 000) non-linear *equality* constraints, which are extremely difficult to handle computationally.

The only realistic option, therefore, appears to be to bound the transmission via linear approximations of the non-linear capacity constraints. To do this we need to examine the constraints themselves. Recall from Section 4.3 that the constraints needed to keep the quadratic $at^2 + bt + c$ within the range $[A, B]$ over the interval $[0, 1]$ are:

$$\begin{aligned} A \leq c \leq B \\ A \leq a + b + c \leq B \\ A \leq -\frac{b^2}{4a} + c \leq B \text{ when } -\frac{b}{2a} \in [0,1] \end{aligned} \tag{5.5}$$

Consider the situation where c is at one of the bounds; *w.l.o.g.* let this bound be A, corresponding to points on the bottom face of the region shown in Figure 4.4. The region into which a and b are allowed to fall is shown in Figure 5.5; notice, in particular, that the *edge* between this face and one of the non-linear faces is linear. A tight linear approximation which only allows *feasible* solutions should, therefore, pass through this edge. Consider the other edge which also passes through the point $(a, b, c) = (0, 0, A)$, and is defined by the intersection between the two faces described by

$$\begin{aligned} a + b + c = A \\ -\frac{b^2}{4a} + c = A. \end{aligned}$$

It is described by the curve

$$\{(a, b, c) = (t - A, -2(t - A), t) \mid t \in [A, B]\}$$

Notice that, since this is linear, the two edges describe a linear face; this face is the linear approximation used, and is given by:

$$b + 2c \geq 2A \tag{5.6}$$

Similarly, the linear approximation used for the other non-linear face is obtained by replacing A with B in Equation 5.6. Therefore, the constraints to ensure that the transmission curves are within $[0, \overline{X}]$ are:

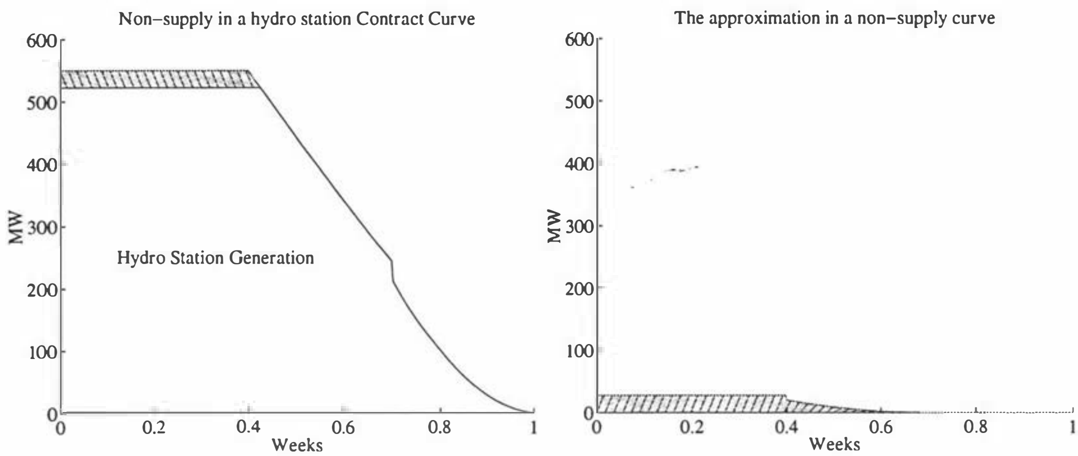


Figure 5.3: Non-supply in a hydro station's G.C., and its "approximation"

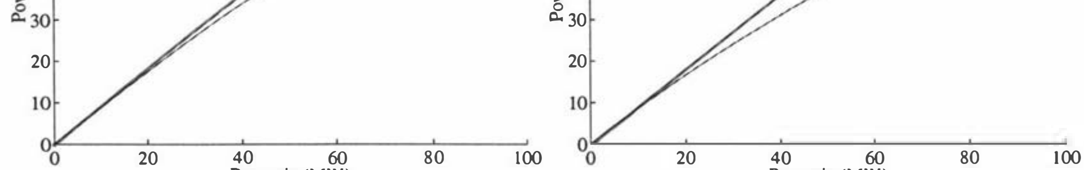


Figure 5.4: Capacity constraints solely through line loss; green curves show 10% linear loss

$$0 \le x_{2k}t_{k-1}^2 + x_{1k}t_{k-1} + x_{0k} \le \overline{X} \quad (5.7)$$
$$0 \le x_{2k}t_k^2 + x_{1k}t_k + x_{0k} \le \overline{X} \quad (5.8)$$
$$0 \le \tfrac{1}{2}x_{1k} + x_{0k} \le \overline{X} \quad (5.9)$$

where the transmission curve is given by $x_{2k}t^2 + x_{1k}t + x_{0k}$ over the subinterval $[t_{k-1}, t_k]$ (for $k = 1, \ldots, 4$ for the working model).

It is worthy to note that, if we used a piecewise linear approximation (or even piecewise discrete) for Electricity curves instead of piecewise quadratic, we could easily bound the transmission to be within capacity using linear constraints *without* approximation. The reason we do not *restrict* ourselves to a piecewise linear approximations is that, in seeking flexibility, we need to explore the limits of the model being developed.

Another point to note is that our capacity approximation does allow all feasible piecewise *monotonic* transmission. Also, in computational testing of these constraints, when we instead impose a linear approximation which *allows* all feasible transmission (for the capacity bound only), the optimal solution remains unchanged.

Having moved to this approach allows a strategy for approximating convex line losses. A piecewise linear approximation of the line loss could be achieved by splitting the arc into many smaller sub-arcs, each with its own capacity (called sub-capacities) and line loss, representing separate pieces of the piecewise linear approximation. One advantage to this approach over a more usual piecewise linear approximation is the effective smoothing off of the corners for particular transmission shapes. If the solution prefers to transmit a shape which spans two sub-capacities within a single partition interval (of the Electricity Curve approximation), this shape is split into two similar shapes. Each shape is sent through one of the sub-arcs, attracting appropriate linear line losses. The combined line loss of the whole shape is thus somewhere between the line losses of the sub-arcs. Figure 5.6 shows an example of this.

5.4 A Better Basis

When considering computational implementation of the model developed, it quickly becomes apparent that we have few explicit bounds on variables. Instead, we have

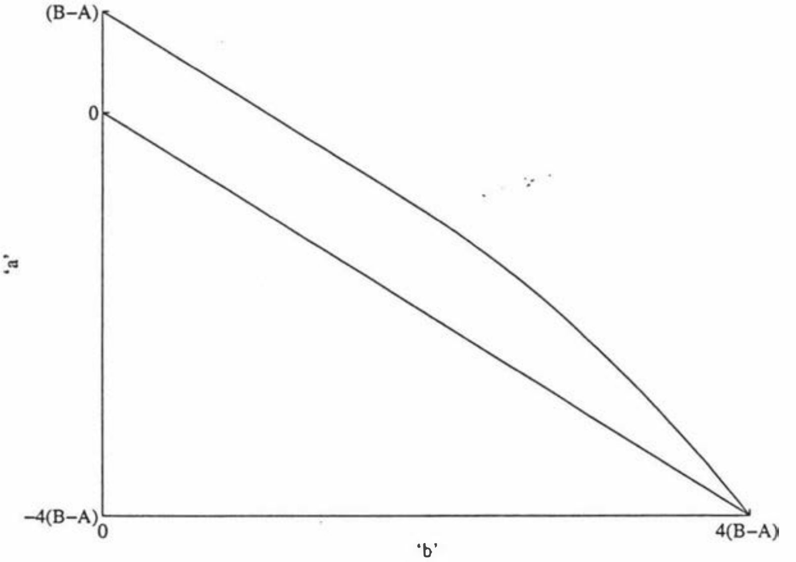


Figure 5.5: Values of a and b for which $at^2 + bt + A \in [A, B]\ \forall t \in [0, 1]$

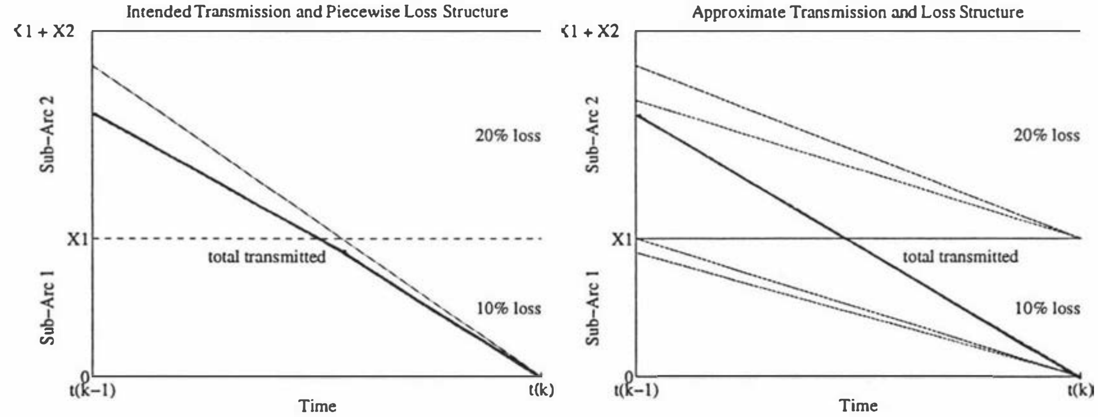


Figure 5.6: Effective smoothing of piecewise linear line losses

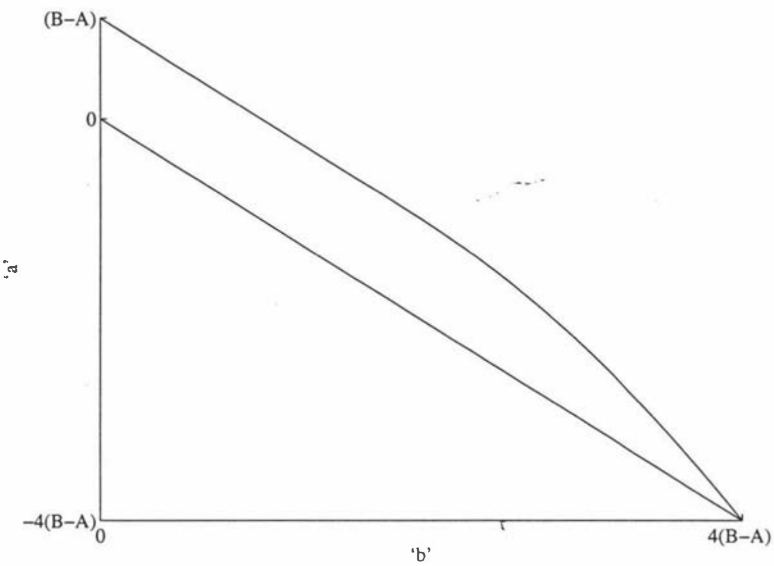


Figure 5.5: Values of a and b for which $at^2 + bt + A \in [A, B]\ \forall t \in [0, 1]$

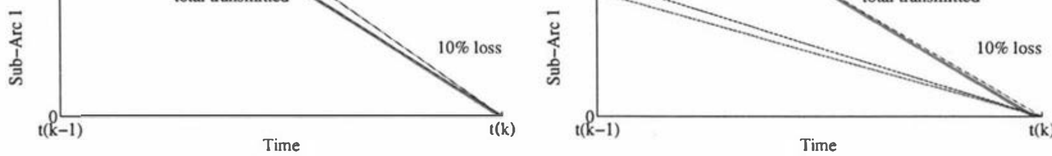


Figure 5.6: Effective smoothing of piecewise linear line losses

constraints which implicitly impose bounds. When solving a problem numerically, it becomes important to have bounds on variables; any constraints which can be imposed as bounds are dealt with intrinsically.

Rather than explicitly determining implicit bounds, it would be better if we could transform some of the constraints already imposed into bounds, or non-negativity conditions, by a change of basis. Here, a change of basis in the variables of the formulation amounts to changing the basis of the piecewise quadratics in the model.

There are two candidate sets of equations for such a transformation: either the capacity bound constraints for transmission, or the decreasing constraints for generation. The capacity constraints include an approximation, which we may wish to change, and only two other constraints for each quadratic piece, so we would be left to choose another arbitrary constraint for each quadratic piece. The decreasing constraints are enough to specify all but one of the new basis elements; we do indeed use these constraints as well as the related non-negative generation constraint.

The constraints we wish to convert to non-negativity conditions are therefore,

$$2g_{k2}t_{k-1} + g_{k1} \leq 0 \quad \forall k = 1 \cdots K \tag{5.10}$$

$$2g_{k2}t_k + g_{k1} \leq 0 \quad \forall k = 1 \cdots K \tag{5.11}$$

$$g_{k2}t_k^2 + g_{k1}t_k + g_{k0} - g_{k+1,2}t_k^2 - g_{k+1,1}t_k - g_{k+1,0} \geq 0 \quad \forall k = 1 \cdots K-1 \tag{5.12}$$

$$g_{K2} + g_{K1} + g_{K0} \geq 0 \tag{5.13}$$

where $\{0 = t_0, t_1, \ldots, t_K = 1\}$ is the partition used in the approximation, and the Electricity Curve constrained is given by $G = g_{k2}t^2 + g_{k1}t + g_{k0}$ over the subinterval $[t_{k-1}, t_k]$. Equations 5.10 and 5.11 ensure the quadratic pieces are decreasing, Equation 5.12 ensures the step discontinuities are decreasing, and Equation 5.13 ensures the curve is non-negative.

We need to ensure that the left-hand-sides of Equations 5.10–5.13 indeed define a basis. To show this we merely need to show that we can write the natural basis in terms of the left-hand-sides of these equations. The natural basis can be written as

$$\mathcal{B} = \left\{ \chi_{[0,t_1]}, \chi_{[0,t_1]}t, \chi_{[0,t_1]}t^2, \ldots, \chi_{(t_{K-1},1]}, \chi_{(t_{K-1},1]}t, \chi_{(t_{K-1},1]}t^2 \right\}$$

where $\chi_{\mathcal{I}}$ is the characteristic function of the set $\mathcal{I}$. We define the new coefficients, f_{kj}, in terms of the natural basis coefficients, g_{kj}, as

$$\begin{aligned} f_{k0} &= g_{k2}{t_k}^2 - g_{k+1,2}{t_k}^2 + g_{k1}t_k \\ &\qquad - g_{k+1,1}t_k + g_{k0} - g_{k+1,0} \quad \forall k = 1 \cdots K-1 && (5.14) \\ f_{K0} &= g_{K2} + g_{K1} + g_{K0} && (5.15) \\ f_{k1} &= -2g_{k2}t_{k-1} - g_{k1} \qquad \forall k = 1 \cdots K && (5.16) \\ f_{k2} &= -2g_{k2}t_k - g_{k1} \qquad \forall k = 1 \cdots K && (5.17) \end{aligned}$$

If we call the new basis elements $\{B_{10}, \ldots, B_{42}\}$, to write an Electricity Curve in terms of the new basis, we substitute in the natural coefficients, g_{kj}, using Equations 5.14–5.17 and collect terms, and obtain the natural basis elements in terms of the new basis:

$$\begin{aligned} \chi_{[0,t_1]} &= B_{10} \\ t\chi_{[0,t_1]} &= t_1 B_{10} - B_{11} - B_{12} \\ t^2\chi_{[0,t_1]} &= {t_1}^2 B_{10} - 2t_1 B_{12} \\ &\vdots \\ \chi_{[t_{K-1},1]} &= B_{K0} - B_{K-1,0} \\ t\chi_{[t_{K-1},1]} &= B_{K0} - B_{K1} - B_{K2} - t_{K-1}B_{K-1,0} \\ t^2\chi_{[t_{K-1},1]} &= B_{K0} - 2t_{K-1}B_{K1} - 2t_{K-1}B_{K2} - {t_{K-1}}^2 B_{K-1,0} \end{aligned}$$

showing the set of functions $\{B_{10}, \ldots, B_{42}\}$, indeed form a basis. It is interesting to examine the form of these new basis elements; Figure 5.7 shows the new basis elements for 2-piecewise quadratics with partition $\{0, .4, 1\}$. It can be seen from Equations 5.14–5.17 that a similar transformation can be used in the piecewise linear case.

Having obtained a new basis representation, we need to convert all of our constraints into this new form. However, there is no need to explicitly do so here. Also, there will be no effect on constraints which are valid for any basis, namely the conservation of power constraints for nodes in the geographic network.

5.5 Exploiting Flexibility

There are some parts of the model for which we can use flexibility in ways as yet unconsidered. We currently have the implicit assumption that all time steps are of the same, fixed, length, e.g. all weeks or months; however, the model developed has the flexibility to allow us to use time steps of differing length. We could, for

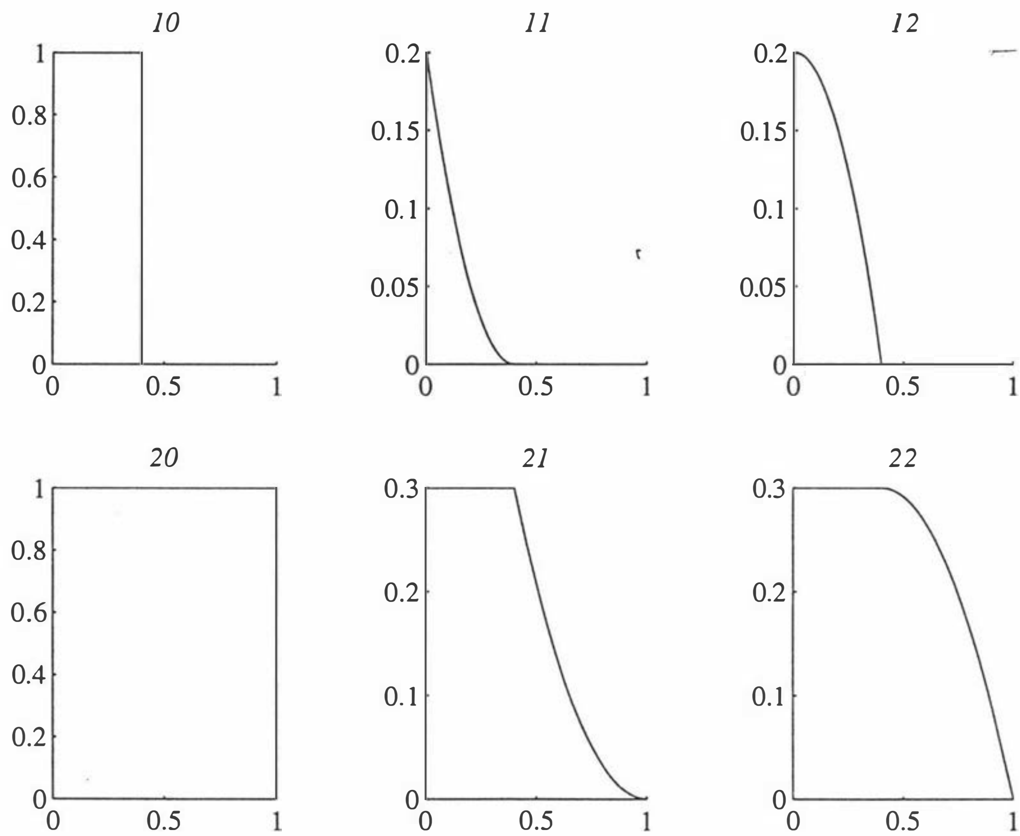


Figure 5.7: Structure of the new basis elements

instance, have time steps of a week, for the first six weeks, then use time steps of four weeks, and finish the year with two time steps of thirteen weeks each. This would reduce the problem to about a quarter of its original size.

The model developed allows infinite variability in the time steps' lengths, and the lengths of time steps to be used are dependent on the trade-off between computability and accuracy of the approximation. As time passes in the model the decisions made become less certain, so it seems that increasing the length of some of the later time steps should not have too much effect on a solution's accuracy. In increasing the length of time steps in later periods, we are effectively relaxing the maximum and minimum storage levels. The effect of this could be to use more reckless policies on long time steps, as there is the ability to "push" our minimum storage level during that time, effectively borrowing water during that time step. This is not to say that this type of relaxation does not occur for constant length time steps, but in this case we are consistent in the amount by which such bounds can be pushed. In having different length time steps we are changing this amount, from one time step to the next, and may find that solutions tend to take more risks in earlier periods, as later periods have more chance of correcting any bad consequences.

The further into the future we look, the more uncertain we are regarding the exact state of the system, and so the effects of such changes in the length of time steps may become swamped by our uncertainty. Also, since it is the intention to re-solve this system each week, if the change in time step lengths occurs far enough in the future, the effect on this week's decision will be minimal.

Differing time step lengths could be used to implicitly model *real* changes in the quality of our future knowledge. For instance, we could increase the length of our time steps for periods in which the uncertainty inherent in the forecasts exceeds a certain level. This would, hopefully, have the effect of including some information on the accuracy of forecasts used, in that each time step would have similar uncertainty in the values being used.

Another unconsidered use of flexibility is to have different geographic networks for different weeks. The reasons for using flexibility in such a way are similar to those for having differing length time steps. In fact it is probably useful to use both of these techniques together. The concerns raised over differing length time steps also apply to having differing geographic networks; however, it is the extra flexibility in transmission which is at issue in this case. This flexibility in the geographic network used also extends to the approximation used for Electricity Curves—it makes sense that, as the uncertainty of the forecasted load increases, our accuracy in approximating it should decrease, as this would lead to less precision in the costs associated with generation. However, the exact effect of such a loss in accuracy is difficult to predict, as it is unclear whether the cost would tend to be consistently lower or higher than the more accurately produced cost. Intuitively, it would appear that on average these costs would be about the same, since in each case we use a best fit approximation of the load, and it is just the coarseness of this approximation that we are changing.

We also consider the aggregation of some hydro or thermal stations at future time steps of the model. This would most likely occur in conjunction with the use of differing geographic networks. Aggregation of thermal stations is quite easy to apply, as the aggregated stations have no direct connection, in terms of the formulation, with the same stations of the previous weeks.

Future aggregation of hydro stations is a little more difficult as it requires integration of their respective waterflow networks. In actuality, this allows multiple storage arcs to enter the *same* reservoir (the aggregated one) for a particular week. There is no realistic way of splitting aggregated reservoirs at later time steps, as it is difficult to decide how much water to assign to each of the new reservoirs; this should not cause a problem, as there is no reason to approximate a group of reservoirs *more* accurately further into the future. Such an aggregation means that we need to include possible multiple storage arcs arriving from the previous week in the conservation of water constraints for the waterflow network. If we define the set $\text{LAST}(h, w)$ to be the set of hydro reservoirs, present (in week $w - 1$) which are to be aggregated into hydro reservoir h in week w, these constraints now become:

$$S_{hw} + H_{hw} - \sum_{j \in \text{LAST}(h,w)} S_{j,w-1} = \tilde{I}_{hw} + \tilde{U}_{hw}$$

where, for hydro reservoir h during week w, S_{hw} and H_{hw} are the storage and release, respectively, and, $\tilde{I}_{hw}$ and $\tilde{U}_{hw}$ are the controlled and uncontrolled inflows, respectively. Note that the inflows are actually stochastic in nature but, for the deterministic model, they are assumed to take on fixed values.

To realistically make a decision on the use of any of these techniques would take rigorous computational testing and simulation of the system, which is beyond the scope of this study. For the moment, the working model will assume that all of the features discussed above are constant over each time step and that there is no further aggregation of hydro stations. However, for flexibility, we allow for the *modelling* techniques described above to be included.

5.6 End Effects and Discount Factors

Thus far there has been no mention of the storage levels at the end of the year. If no constraints are used to handle these, then it is most likely that *all* of the lakes will be empty at the end of the year in an optimal solution, as this makes fullest use of the "free" resource, water. We are not taking into account the fact that we will need water beyond the end of the planning horizon, but we do not have information now on how much we will need; therefore we need to define some terminal conditions.

We do not intend to adhere to more of the solution produced than that for the first week, since, in each future week, we will be using the new information on hand to refine the decisions made. This implies that the terminal conditions do not need to be too precise, as the effect of these conditions on the the first week's solution should be minimal. However, as we use the model over time, this small effect could propagate through the solutions from one week's run to the next, until it begins to have a major effect on the types of solutions generated. For instance, it may be that if we allow the lakes to finish empty, use of this model over time may slowly lower the average level of the lakes, since it appears that there is more water available than is actually the case.

One solution is to estimate the future benefit of the final period's lake levels, for each lake, and include this in the objective function. A drawback with this is that of *estimating* this future benefit—it is not a simple task, as, including too much of a benefit will mean we leave the lakes full at the end of the year under the impression that we can make better use of the water next year, and, too small a benefit means we will use all of the water this year. Another problem with using a future benefit is that, although this is only a coarse estimate, it has the potential to swamp other more certainly known details in the objective function, due to factors such as machine precision.

Another reasonable solution is to fix the lake levels at the end of the year. We could require these to be at some previously decided "best" level for the time of year, or to be at the same level as they currently are. Another option is to fix the volume of water over a particular set of lakes, and let the final solution decide the exact distribution of this water. This will mean that, while the average lake level remains constant, the relative levels of the lakes will change. This is the option used in the working model, although it does tend to empty the most usefully located and reliable lakes in preference to those which have a high variability of inflow or are geographically challenged[1].

Of course, these options could be used in combination, with bounds also added to final levels; however, they have little effect on the *form* of the model developed, so we can assume the model can have any combination of the above terminal conditions. There is little point in becoming too elaborate here, since the effect on the first week's solution is limited.

Another point to briefly mention is that of using discount factors. These are used to discount the value of the objective function for later weeks, so as to to place more emphasis on earlier weeks than later weeks since, when we *actually* come to schedule later weeks, we will have more information on the conditions of the system. This means that, if a the value of using water this week is the same as using it next week, we will, in preference, use the water this week. For flexibility, we allow any discount factor; a discount factor of 1 relates to the situation where we do not discount at all.

5.7 The Deterministic Model

Table 5.2 shows a slightly generalized version of the full working model developed so far. For simplicity and compactness, as many of the constraints as possible are written in terms of the Electricity Curves themselves, rather than the coefficients of the basis used to approximate these curves. To expand these constraints, the Electricity Curve need only be replaced by its basis representation. This working model uses a K-piecewise quadratic approximation to the Electricity Curves.

The notation of the formulation in Table 5.2 is explained in Table 5.3. A w subscripted on a set represents only those elements in the set which are from week w. Also, all elements of sets are assumed to be associated with a fixed week, e.g. a

[1]isolated

Table 5.2: The Full Deterministic Working Model

$$\text{Min} \sum_{w=1}^{Y} \tau_w \left(\sum_{j \in \mathcal{N}_w} c_{NS} m(F_j) + \sum_{h \in \text{HYDRO}_w} c_{NS}(m(G_h) - H_h) + \sum_{q \in \text{THERMAL}_w} c'_q m(G_q) \right)$$

subject to:

$$\sum_{q \in \text{THERMAL(j)}} G_q + \sum_{h \in \text{HYDRO(j)}} G_h + F_j + \sum_{i \in \text{IN}(j)} (1-\beta_i) X_i - \sum_{i \in \text{OUT}(j)} X_i = L_j \quad \forall j \in \mathcal{N} \tag{5.18}$$

$$S_h + H_h - \sum_{k \in \text{LAST}(h)} S_k = \tilde{I}_h + \tilde{U}_h \quad \forall h \in \text{HYDRO} \tag{5.19}$$

$$H_h - W_h \leq m(G_h) \quad \forall h \in \text{HYDRO} \tag{5.20}$$

$$0 \leq W_h \leq H_h \quad \forall h \in \text{HYDRO} \tag{5.21}$$

$$\left.\begin{array}{r} \underline{S}_h \leq S_h \leq \overline{S}_h \\ \max\{\underline{F}_h, \tilde{U}_h\} \leq H_h \leq \overline{F}_h \\ \underline{R}_h + (1-\gamma_h)\tilde{U}_h \leq H_h \leq \overline{R}_h + \tilde{U}_h \end{array}\right\} \quad \forall h \in \text{HYDRO} \tag{5.22}$$

$$\left.\begin{array}{r} G_h(0) \leq \min\{\overline{F}_h, \overline{R}_h + \tilde{U}_h\} \\ W_h + G_h(1) \geq \max\{\underline{F}_h, \underline{R}_h + \gamma_h \tilde{U}_h\} \end{array}\right\} \quad \forall h \in \text{HYDRO} \tag{5.23}$$

$$X_i \in \mathcal{C}_{\overline{X}_i} \quad \forall i \in \mathcal{A} \tag{5.24}$$

$$G_s(0) \leq \overline{X}_s \quad \forall s \in \text{POWER} \tag{5.25}$$

$$g_{i,s} \geq 0 \;\; \forall s \in \text{POWER}, \quad f_{i,j} \geq 0 \;\; \forall j \in \mathcal{N} \quad \forall i \in \{10, \ldots, K2\} \tag{5.26}$$

Table 5.3: Notation used in Table 5.2

$\mathcal{N}$, $\mathcal{A}$	All nodes and arcs, respectively, in geographic networks.
TYPE	All *type* stations (POWER = HYDRO ∪ THERMAL); a dependence on j represents *only* those present at node j.
IN(j), OUT(j)	All arcs entering and exiting, respectively, node j.
LAST(h)	All hydro reservoirs, from the previous week, aggregated into hydro reservoir h.
$\mathcal{C}_{\overline{X}}$	Approximated set of all curves with range in $[0, \overline{X}]$.
$\{10, \ldots, K2\}$	Subscripts of the coefficients of the basis elements.
$m(G)$	Area under the curve defined by G.
G_s, $g_{i,s}$	C.C. for station s and i'th coefficient of this, respectively.
F_j, $f_{i,j}$	Node j non-supply curve and its i'th coefficient, respectively.
X_i	Transmission curve for arc i.
$\underline{R}_h$, H_h, $\overline{R}_h$	Minimum, actual and maximum release, respectively, from hydro reservoir h (including spill).
$\underline{S}_h$, S_h, $\overline{S}_h$	Minimum, actual and maximum storage levels, respectively, for hydro reservoir h at the end of the current week.
W_h	Spill from hydro reservoir h.
L_j	Load Duration Curve for node j.
$\overline{X}_j$	Capacity of station (or transmission line) j.
$\underline{F}_h$, $\overline{F}_h$	Minimum and maximum flow, respectively, from the river mouth of hydro chain h.
τ_w	Factor for week w including discount factor and week length.
c_{NS}	Cost of non-supply.
c'_q	$(1-p)c_q + pc_{NS}$ where c_q is the fuel cost for thermal station q.
β_i	Fraction of power loss for transmission arc i.
γ_h	Fraction of uncontrolled inflow, for hydro station h, which arrives at a constant rate.

hydro station at week w is a separate element of the same station at week $w + 1$. The set $\mathcal{C}_{\overline{X}}$ refers to all Electricity Curves whose coefficients satisfy Equations 5.7–5.9. The formulation shown in Table 5.2 assumes that each hydro station has its own C.C..

To obtain an idea of the size of the formulation in Table 5.2, we count the number of constraints (the Equations in this Table each describe many constraints). The assumption is made that the same geographic network, number of stations, and approximation of the Electricity Curves are used for each week. Equation 5.18 (conservation of power at geographic network nodes) represents one constraint for the coefficients of each basis element for each node of a geographic network, giving

$3KYN$ constraints for a K-piece piecewise quadratic approximation of the electricity curves, where Y is the number of weeks in a year, and N is the number of nodes in the geographic network for each week. Equation 5.19 (conservation of water at waterflow network nodes) represents YH constraints, where H is the number of hydro stations. Equations 5.20 and 5.21 (ensuring W_h is the total spill) also each represent YH constraints. Equations 5.22 just represent bounds on storage and release, and are generally dealt with separately, and so do not explicitly add to the number of constraints; the same is true of the non-negativity conditions (Equation 5.26) which ensure some of the Electricity Curves are decreasing. Equations 5.23 and 5.25 (minimum and maximum generation levels) represent $Y(Q+2H)$ constraints, where Q is the number of thermal stations. Finally, Equation 5.24 represents $6K$ constraints for every arc (this can be halved by enforcing bounds on the slack of each constraint, giving a total of $3KY(2A)$ where A is the number of transmission lines in the geographic network, being *half* the number of arcs since these are split (see Section 3.5)).

These give a total of $Y(3K(N+2A)+5H+Q)$ constraints. The number of variables is $Y(3K(N+2A)+3H)$, so the number of variables is approximately the same as the number of constraints. For the working model as described in this and preceding Chapters there are 29 328 variables and 28 392 constraints.

5.8 Generalized Network with Side Constraints

The model, as formulated in Table 5.2, can be re-formulated as a Generalized Network with side constraints. This is most helpful in determining a solution procedure as the Generalized Network structure is easily exploited to allow faster solution times. The re-formulation is *mostly* a change in the way the problem is interpreted. To show that a formulation is indeed a Generalized Network with side constraints, we merely need to demonstrate that the formulation exhibits a Generalized Network substructure which includes *every variable.*

Each basis element (used in the Electricity Curve approximation) has an associated Generalized Network for each week, corresponding to the geographic network. The variables of this network, for week w, are exactly the coefficients of the associated basis element corresponding to each Electricity Curve of week w. For each basis element's network, Equation 5.18 corresponds to the "conservation of mass" constraint at a node, with the transmission variables corresponding to arcs (with losses) connecting these nodes. The generation variables and non-supply variables each correspond to an arc from the node, at which they appear, to a fictitious "power supply" node (a super source of the network). The load coefficients correspond to the sinks of this Generalized Network.

The waterflow networks are obviously (Pure) Networks (and therefore Generalized Networks), however we need to incorporate the spill as a network variable, or eliminate it from the formulation. It may be eliminated by replacing Equation 5.20 with the constraint

$$H_h \geq m(G_h) \quad \forall h \in \text{HYDRO}$$

and, consequently, removing Equation 5.21. This slightly alters the *model*. The *same* model can be maintained by the introduction of a new variable, V_h, for each hydro station h (with a *particular* week corresponding to each h). V_h exactly represents the part of the release which flows through the turbines of hydro station h. This new variable is given by

$$V_h = H_h - W_h \tag{5.27}$$

and *may* be used to eliminate H_h from the formulation. However, since the removal of H_h converts the bounds on H_h into explicit constraints, it may be more useful to retain this variable.

The waterflow network, including the spill variable and the new variable V_h, is now a (Pure) Network with each variable corresponding to an arc. Equation 5.19 describes a storage node, and Equation 5.27 describes a "river mouth" node (the arc corresponding to H_h may be retained or removed). Equation 5.21 implicitly holds (by Equation 5.27 and the non-negativity of V_h and W_h), so it can be removed from the formulation. Figure 5.8 shows the part of a waterflow network, for a single week, under this formulation.

Since every variable of the re-formulation may be considered to be part of a Generalized Network, the re-formulation is a Generalized Network with side constraints. Constraints corresponding to Equations 5.20 and 5.23–5.25 are the side constraints, and the remainder of the constraints correspond to either nodes of the Generalized Network, or to bounds on the arc flows.

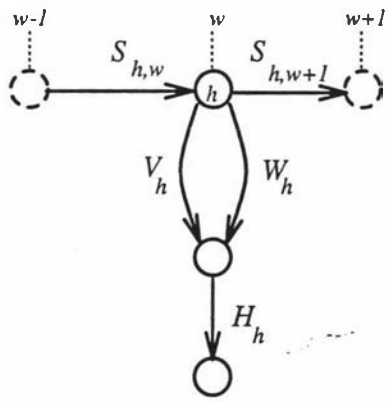


Figure 5.8: Part of the waterflow network with the addition of V

5.9 Discussion

This Chapter described a deterministic model for the New Zealand hydro-thermal electricity generation system. Although it includes some stochastic features (for instance estimates on expected generation costs in terms of station failures), it treats the major stochastic elements of the system as deterministic. These elements are the inflows into the hydro reservoirs.

Due to the high variability and unpredictability of these inflows, the solution to the deterministic model with particular inflows is not very robust as an implementable solution, since it relies on the fact that the inflows are fixed. This does *not* imply that the solutions of the deterministic problem are of no use, merely that they should not be implemented *as they stand*. They are useful in determining the effect of uncertainty in the system, and for producing a lower bound on costs for particular scenarios of inflows.

The major advantage of a purely deterministic model is the ability to allow vast amounts of detail which it is not computationally reasonable to have in a stochastic model. The next Chapter will describe how the deterministic model developed can be usefully extended to encompass the stochastic elements.

Chapter 6

Modelling Stochastic Inflows

A linear model of the physical system discretized over the time horizon was developed in Chapter 5. It included methods for coping with the stochastic elements brought in by supply uncertainty and load forecasts, but it did not explicitly deal with the stochasticity of hydro reservoir inflows. Any model developed needs to take adequate account of this stochasticity to be effective.

A reason for leaving the discussion of stochastic aspects until this Chapter is that they are very difficult to take accurate account of, in the sense of developing a computationally tractible model. In the New Zealand system, the hydro inflows have a high variance; this is illustrated in Figure 6.1, which shows 10 years of inflows into South Island lake, Te Anau (the reservoir for the Manapouri hydro station). Note that the highest inflows into lake Te Anau are able to almost fill the lake from empty. Also, there are differing correlations *between* reservoirs (even those in the same Island); Figure 6.2 illustrates this by showing two scatter plots of the inflows given in Figure 6.1 against those into South Island lakes Pukaki and Hawea for corresponding weeks.

Every reservoir can be modelled by a separate random variable (or possibly a combination, if partial correlations are included). Including temporal independence increases the number of random variables involved to the order of 300. Given that each random variable adds another dimension to the problem (in terms of right-hand-sides and coefficients), taking all of these factors explicitly into account can increase computation time to an unreasonable level if a brute-force search is used.

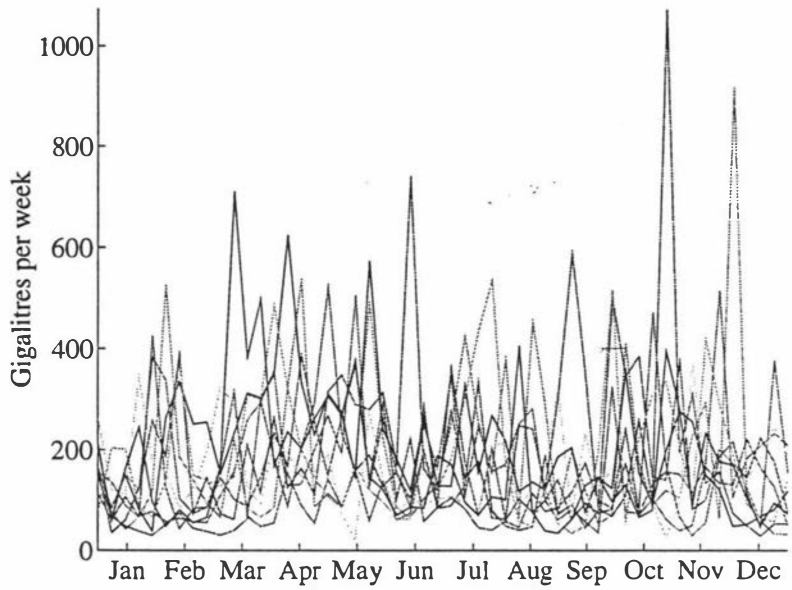


Figure 6.1: Ten years of inflows into lake Te Anau

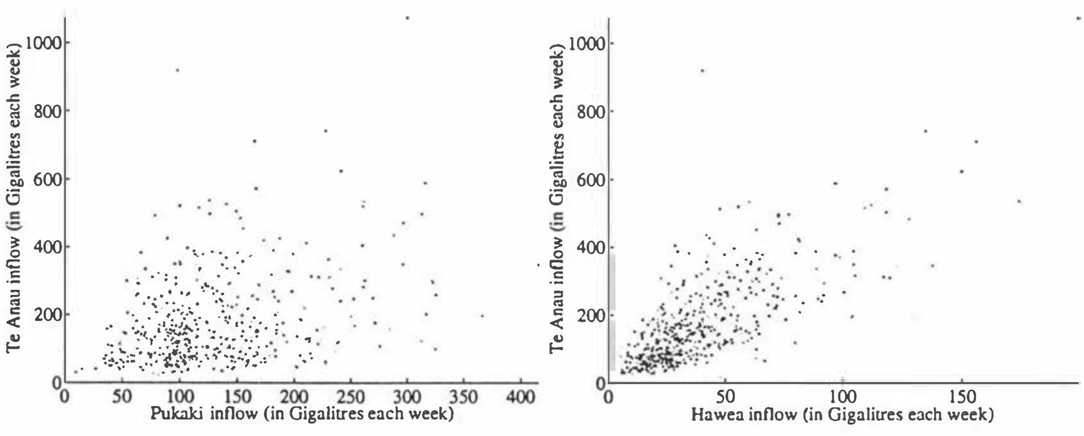


Figure 6.2: Te Anau inflows against those into Pukaki and Hawea

6.1 Optimality in a Stochastic Setting

Extending the constraint set to include stochastic elements is a reasonably simple exercise. We merely need to define which values are stochastic and allow some of the variables to depend on the observed value of various stochastic variables. We can model correlations and dependencies by allowing some of the stochastic variables to be given by a weighted sum of various random variables. For instance, suppose there is a correlation between the controlled and uncontrolled inflows of a particular river chain. If we let ξ, ζ and ψ be random variables from appropriate distributions, then the correlation can be modelled by setting $\tilde{I} = \xi + \lambda_I \zeta$ and $\tilde{U} = \psi + \lambda_U \zeta$, where the λ's represent correlation factors. In general each inflow would be a function of a random variable corresponding uniquely to that inflow, and various other random variables (of which other inflows are also functions) corresponding to correlations bought in by various environmental effects (local weather patterns, for instance).

A difficulty which arises when specifying the objective function is that of determining exactly what we are trying to optimize—the answer is not at all obvious. The usual objective used for Stochastic Programming is that of minimizing the expected cost (or maximizing the expected benefit); the actual objective depends on the intention of the model and what is reasonably achievable.

The "tails" of random variable distributions present a difficulty when using expected cost, since these tails are often not well approximated due to a lack of information about this area of the distribution (consider approximating a statistically 1 in 100 year drought using only 50 years of past data). However, these tails may actually *drive* the solution, since the costs associated with such tails are often large and could swamp the data which would otherwise lead to more reasonable solutions. Another difficulty with these extreme values is that, generally, approximations and constraints of the model are based on near-average values of the stochastic variables and such approximations and constraints may break down, or become unreasonable, in the extreme. Also, some constraints are not hard, but are more easily modelled in this way, especially in the face of extreme conditions; given a serious drought, noone can expect minimum river levels to be maintained. Therefore, the difficulty may be the way the problem is modelled. However, in this situation, it may *not* be desirable to re-model such constraints, as this re-modelling may cause the model to become computationally intractible.

In general, the approximation used for random variable distributions truncates the tails of the underlying distribution (e.g. a discrete approximation). Such a situation can be thought of as optimizing over some set of "reasonable" values, and treating extreme situations as "acts-of-god", for which special actions will be taken which are not (or can not be) modelled explicitly. By specifying where the truncations are made, we are defining such unreasonable situations.

Minimizing the expected cost is not the only objective that could be used; it assumes that the intention is to do well in the long-term. To do better in the short-term, one could include criteria for taking risks on the forecasts, or, risk trying to do better in an average year by foresaking security in an extreme year. Given the unpredictability of inflows into hydro reservoirs in New Zealand and the lack of imported power, such risk taking is most likely untenable. Because of this, and the difficulties inherent in specifying other objectives, we use the minimization of expected cost, or an approximation thereof, as the objective for this model. By changing the random variable distributions used, or incorporating a weighting function into the objective function, we can change the importance given to various probable futures and hence include some flexibility in the definition of the objective.

6.2 The General Problem

Recall (from Table 5.2) that the stochastic variables in the model are present only as right-hand sides of some constraints and bounds. Hence, the general stochastic program can be written as the multi-period stochastic program

$$\text{Min } \underset{\xi|y}{E}\left[Z(x,y)\right]$$

$$A\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} b \\ \xi_1 \end{bmatrix}$$

$$\xi_2 \leq \begin{bmatrix} x \\ y \end{bmatrix} \leq \xi_3$$

where Z is the cost of generation, $\xi = [\xi_1\ \xi_2\ \xi_3]^T$ is the set of random variables (with possible correlations), y is some type of state or history information from previous time steps upon which our decisions (and some random variables) are contingent, and x represents the decision variables. Here the expectation is calculated as an integral over all possible random variable values which form some multi-dimensional set. It is the calculation of this integral that creates the difficulty, since, as pointed out by Wets [22], while there are adequate ways of numerically computing integrals over one (and possibly two) dimensions, there are no reasonable ways for doing so over three or more dimensions.

It is well known that when the probability distributions underlying the stochastic variables are discrete, the formulation can be written as an equivalent large-scale deterministic program (with a stochastic interpretation). Furthermore, when the model of the physical system underlying the stochastic problem is linear (as it is in this case), the equivalent large-scale deterministic problem obtained is an LP. Many methods discretize the probability distributions underlying the stochastic variables (either before or during solution) to take advantage of this property, however the large-scale nature of the equivalent deterministic problem means that even then one must limit the size of "local" searches (possibly by limiting the size of the problem investigated) so as to make the problem tractable.

In Section 6.3 we consider using a continuous approximation, via a fixed basis, to the distributions underlying the stochastic variables, in a similar manner to the approximation of Electricity Curves. The benefit of using a fixed basis is that it eliminates the difficulty of calculating expected values, because one can obtain a fixed polynomial expression.

A difficulty which arises when discretizing time is that of deciding exactly when the random variable is observed (and when this information can be used). Such knowledge at the beginning of the week assumes perfect foresight over the week, while allowing this knowledge only at the end of the week assumes that we cannot react to knowledge gained during the week. For the latter, however, some state variables (in our case, storage at the end of the week) must depend upon the actual value of the random variables. To approximate a *limited* use of the knowledge before the end of the week, we could assume, say, that the Contract Curve of the station needs to be fixed at the beginning of the week; however, the release, storage and spill can all depend on the value of the random variable, allowing hydro generation to be replaced by non-supply (or re-supply) depending on the observed value of the random variable, simulating an ability to react to information gained over the week.

For the working model we assume, for simplicity, that all decisions must be made at the beginning of the week. This allows the possible first week's inflows to realistically model inflow forecasts for the coming week, rather than needing to forecast only *one* inflow for the first week. All decisions are extendable, with appropriate modification, to any of the cases discussed above. Often we will use the simplest case for a particular situation, so as to avoid obfuscatory complications.

6.3 A New Method

Consider for the moment the situation for which the reservoirs have independent inflows which, individually, are completely correlated with respect to time. The case for two reservoirs is easily generalizable to the multi-reservoir situation and so, for simplicity, we use this as an example. We also make the assumption that there are no uncontrolled inflows, although this can be extended to the assumption that either the uncontrolled inflows are fixed, or totally correlated with the controlled inflows. For the purpose of this discussion, we will also assume that the decisions can be made with perfect foresight of the week ahead; this avoids the need to distinguish between structure admitted by the method described and that created by this lack of foresight.

Due to temporal correlation and spatial independence, inflow into a reservoir for each week can be determined using a single random variable. This can be represented by having, for each reservoir j, a parameter $x_j \in [0,1]$ and, for each week, w, a function of this parameter, f_{jw}, which gives the actual inflow for that week. The conservation of water constraint, and the constraint linking the hydro station release with its generation (for hydro reservoir j during week w), are respectively:

$$S_{jw}(x_1,x_2) + H_{jw}(x_1,x_2) = S_{j,w-1}(x_1,x_2) + f_{jw}(x_j) \tag{6.1}$$

$$H_{jw}(x_1,x_2) - W_{jw}(x_1,x_2) \leq m(G_{jw}(x_1,x_2)) \tag{6.2}$$

where each variable represents a two-dimensional function of a fixed basis so that these constraints can be put in terms of their coefficients (which would be the variables of the Mathematical Program). The variables are similar in design to those given in Table 5.3.

Unfortunately, the assumption of temporal correlation is not a good approximation for the New Zealand system. However, if one introduces temporal independence, then the decision variables for each week become functions of *all* of the inflows for *all* of the preceding weeks! One way to circumvent this is to assume that our decision can be based only on fixed horizon hindsight. Suppose, for instance, that we can base our release decisions only on the previous two weeks' inflows. A natural extension to this is to allow the probability distribution of this week's inflows to depend on last week's inflows as well (introducing a lag-1 temporal correlation). Equations 6.1 and 6.2 therefore become

$$\begin{aligned} S_{jw}(x_{1w}, x_{2w}, x_{1,w-1}, x_{2,w-1}) + H_{jw}(x_{1w}, x_{2w}, x_{1,w-1}, x_{2,w-1}, x_{1,w-2}, x_{2,w-2}) \\ = S_{j,w-1}(x_{1,w-1}, x_{2,w-1}, x_{1,w-2}, x_{2,w-2}) + f_{jw}(x_{jw}, x_{j,w-1}) \\ H_{jw}(x_{1w}, x_{2w}, x_{1,w-1}, x_{2,w-1}) - W_{jw}(x_{1w}, x_{2w}, x_{1,w-1}, x_{2,w-1}) \\ \leq m(G_{jw}(x_{1w}, x_{2w}, x_{1,w-1}, x_{2,w-1})) \end{aligned} \tag{6.3}$$

The conceptual difficulty here is that we are assuming that our release decision for this week cannot depend (explicitly) on the storage level at the beginning of the week.

What we would *like* to do is to introduce a "forgetting" function, g, which approximates the distribution of storage levels for the beginning of the week (which is currently a decision function of many parameters) by a single probability distribution (we could also include finite horizon hindsight if desired). If we have no hindsight (except for the lake level at the beginning of the week), the conservation of water constraint becomes:

$$S_{jw}(x_{jw}, y_{jw}) + H_{jw}(y_{jw}) = g[S_{j,w-1}](y_{jw}) + f_{jw}(x_{jw})$$

In addition to the difficulty of specifying g, such an approach suffers from the large number of variables needed. If we consider having six independent hydro reservoirs, the function H_{jw} of Equation 6.3 is a function of 18 parameters, so that if, for each parameter, we have a basis of n functions, then to specify each function H_{jw} would take in the order of n^{18} variables! An in-depth investigation of such an approach, even for a small number of reservoirs, would take a lot of theoretical research to ensure that it is of a robust nature. Such an investigation is beyond the scope of this thesis, and this approach is therefore not taken any further.

6.4 Stochastic Approaches

There are many approaches to solving Stochastic Programs in the literature. To fully examine these alternatives as feasible extensions to the deterministic model would not only require adjustment of this model to fit the extension, but also require rigorous testing of the viability and robustness of the resulting stochastic model. Such an investigation is also well beyond the scope of this thesis, since the intention here is to develop a detailed model of the physical system which can be extended to a full stochastic model. To this end we examine only how *one* such extension may be achieved and perform very minor testing so as to address some of the implementation issues involved.

We do, however, identify the need to carry out testing and simulations on various stochastic models so as to identify those which best meet the needs of New Zealand's power scheduling system. Such a study, purely for the New Zealand system, has not yet been initiated. In performing such a study, there would be considerable benefit gained from the use of consistent models of the physical system because this would tend to remove discrepancies which are based on differences in the way in which physical system is actually modelled in different stochastic models. It would mean that all approaches could be developed and coded together, removing some of the arbitrariness in separately developed (and programmed) models. Also, the solutions from such models could then be realistically compared, as the structure of solutions obtained would tend to be consistent.

For the moment we give a brief account of some of the stochastic approaches presented in the literature and discuss their possible use as extensions to the model developed here.

6.4.1 Stochastic Dynamic Programming

Stochastic Dynamic Programming (SDP) uses the concept of state variables, which are variables that react to the value of the decision variables and to the random variables' observations. In solution, SDP discretizes the random variables and the state variables upon which the current decision is being made, so as to determine the best current decision based on each of these values, and calculates the cost of making such a decision. It then interpolates the costs, and uses this as an approximation of the future cost of decisions for the decisions of the previous time period (in a backwards recursion).

The undesirable feature associated with this approach is the aptly named "curse of dimensionality". Because of this, SDP can use few (two or three) random variables at each stage whilst remaining computationally feasible. Such an approach is taken by Boshier, Manning and Read [2] for the New Zealand system.

An effort to circumvent the curse of dimensionality resulted in the use of an Aggregation-Decomposition approach (Durán, Puech, Díaz, and Sánchez [5]). The principle behind this approach is to aggregate all but one of the reservoirs and solve the resulting system using an SDP approach, repeating this for each reservoir. The "solution" is then taken to be the combination of the individual solutions for each reservoir. If such an approach were to be used in New Zealand, one would need two aggregated reservoirs, one for the North Island and one for the South Island, since the physical barrier between the two needs to be well represented.

Another attempt to side step the difficulties of SDP has recently been developed. Stochastic Dual Dynamic Programming (Pereira and Pinto [16]) discretizes the state variables near points of interest (locations where a solution is likely to venture) for the SDP backward recursion and then performs a forward simulation to determine new "interesting" values of the state variables. The algorithm calculates lower bounds and estimates upper bounds, giving an idea of the convergence of the objective value. It appears that such an approach could work well as a stochastic extension here.

Lagrangian relaxation has also been used in a number of decomposition approaches. One method was to decompose the system into separate stations by fixing the Lagrange multipliers corresponding to meeting demand, Li, Yan, and Zhou [10]. The hydro systems are then individually solved using SDP and the thermal station's generation is directly determined. The Lagrange multipliers can then be updated so as to ensure global convergence. Due to the use of an underlying network in this model, such a method would be difficult to implement as it stands, although for this model we could, instead, relax the Lagrange multipliers of the constraints linking hydro reservoir release and hydro station generation. Lagrangian relaxations are also used in conjunction with a scenario-aggregation approach.

6.4.2 Scenario-Aggregation

A scenario-aggregation approach approximates the random variable distributions with a number of fixed, appropriately weighted "scenarios". The problem can then be formulated as an equivalent large-scale deterministic problem in which the deterministic base model is replicated for the various scenarios and so-called non-anticipavity conditions are introduced to ensure that no decision made uses foreknowledge which would not be available at the time. The large-scale nature of this problem all but forbids direct solution, and so the problem is often decomposed by relaxing these non-anticipavity constraints.

In situations where the effects of the stochastic variables are slight, often only the relaxation is solved and the solutions are combined by hand; see, for instance, Dembo *et al.* [4]. Such a method is not appropriate for the New Zealand system because of the high variability of the inflows and the significant influence this has on deterministic solutions.

The structure of the equivalent large-scale deterministic problem lends itself to the use of Bender's Decomposition. Unfortunately, to be effective in reducing the size of the problem to manageable portions would require many successive applications of the decomposition, and, in practice, the number of Bender's Cuts necessary to obtain a solution makes this approach intractable.

Several methods have been developed which apply a Lagrangian relaxation to the non-anticipavity conditions. One in particular, the Progressive Hedging Algorithm of Rockafellar and Wets [19], uses an augmented Lagrangian technique to successively tighten the non-anticipavity condition relaxation. This method is described in more detail in the next Section.

6.5 Applying Progressive Hedging

A scenario aggregation approach was chosen to be used as the stochastic extension, as it offers the greatest flexibility in the extent to which the modelling of stochastic elements dominates the solution procedure. Furthermore, it allows correlations and future forecasts to be included (if only implicitly). In taking such an approach we are, in essence, approximating a multi-dimensional solution space by a few selected snap-shots of this space. It is hoped that the physical detail given by solutions to the thus created full stochastic model goes some way to making up for the lack of stochastic detail. It is intended that this approach should be used in tandem with another approach which is more thorough in dealing with the stochastic elements (and hence has a less well defined physical system). Indeed, the flexibility of this approach allows it to fulfill both roles.

In reality, however, the users of the model will probably want to use a single model which gives consistently sensible solutions. Therefore, such a model needs to be sufficiently flexible so as to be capable of producing solutions which can be well stochastically hedged or well detailed (since, in practice, achieving both is not computationally tractable). The use of a scenario aggregation method, with the deterministic model developed in Chapters 3–5, creates such a model.

The particular solution method chosen is Rockafellar and Wets' *Progressive Hedging Algorithm* [19]. The benefits of this method are that it has well-grounded theory and may be solved on parallel processors (which would greatly enhance solution time). It also has the advantage that, at each iteration, it produces a solution which obeys all of the non-anticipavity conditions (although these may not be feasible in terms of the formulation); this solution can then be used as an approximation to the optimal solution. In terms of the solution process, the Progressive Hedging Algorithm amounts to re-solving the deterministic model under various scenarios, with a quadratic augmentation to the objective function, many times.

6.5.1 A Brief Description of Progressive Hedging

The deterministic model for each scenario of inflows is known as a scenario subproblem. Each subproblem, i, is given a positive weight, p_i, which can be thought of as its probability of occurrence. The optimal solution to each of these subproblems is found, and these are used as the initial subproblem solutions, x_i^*. We begin with an initial estimate of the value of the Lagrange multipliers associated with the non-anticipavity conditions, W (usually $W = 0$ is used).

Using the current subproblem solutions, a *policy*, $\hat{X}$, is determined for each scenario. The value of each component of the policy for any scenario is equal to the corresponding component of the current subproblem solution, or, if this component is required to obey a non-anticipavity condition, it is equal to the average value (weighted by the p_i's) of the components (of the current scenario subproblems) which must also satisfy *the same* non-anticipavity condition. That is

$$\hat{X}_{ij} = \frac{\sum\limits_{k \in \mathcal{H}(i,j)} p_k x_{kj}^*}{\sum\limits_{k \in \mathcal{H}(i,j)} p_k}$$

where $\mathcal{H}(i,j)$ is the set of all subproblems for which the j'th variable is required to be equal to the j'th variable of the i'th subproblem by the non-anticipavity conditions. If x_{ij} is not constrained by any non-anticipavity condition, we have $\mathcal{H}(i,j) = \{i\}$.

The current Lagrange multiplier estimates are updated via

$$W_i \leftarrow W_i + r\left(x_i^* - \hat{X}_i\right)$$

where $r > 0$ is a fixed penalty parameter. The objective functions of the scenario subproblems are augmented so as to include the Lagrangian term for the non-anticipavity conditions, $\sum_j p_i W_{ij} x_{ij}$ (often written as $\langle W_i, x_i \rangle$), and a quadratic term to limit the step length taken, $\frac{r}{2}\|x_i^* - \hat{X}_i\|^2$ (where the norm corresponds to the inner product used in the Lagrangian term). Each of these new scenario subproblems is then solved to obtain new current subproblem solutions and the process repeated until adequate convergence is obtained. Figure 6.3 describes this process diagrammatically.

The implementation of this algorithm is reasonably straightforward, however there are one or two implementation issues to be considered. These, and the reported experiences of others who have also used this algorithm, are discussed in Chapter 9.

6.6 Choosing Scenarios

Developing the exact method for choosing the scenarios to be used demands considerable attention. It requires computational testing and simulation to convey an appreciation for the effects of this choice in practice. When developing a model to deal with stochastic aspects of a problem, there is no alternative to empirical testing, as it is the quality of solutions produced *for the system involved* that is of real interest. Such an exhaustive study is beyond the scope of this thesis.

In choosing a stochastic extension to the deterministic model developed we do, however, have some expectations of the model to be implemented. A scenario approach was chosen specifically because of the freedom allowed in the choice of scenarios. Since the number of scenarios needed to take reasonable account of the stochastic elements may be intractably large, it *could* be, nonetheless, that this freedom is a false security in this case. The intention is to run the model under a very few *representative* scenarios to gain insight into the running of the system as a whole, and more importantly, for this model, to give a great deal of detail about the physical implementation of such a solution.

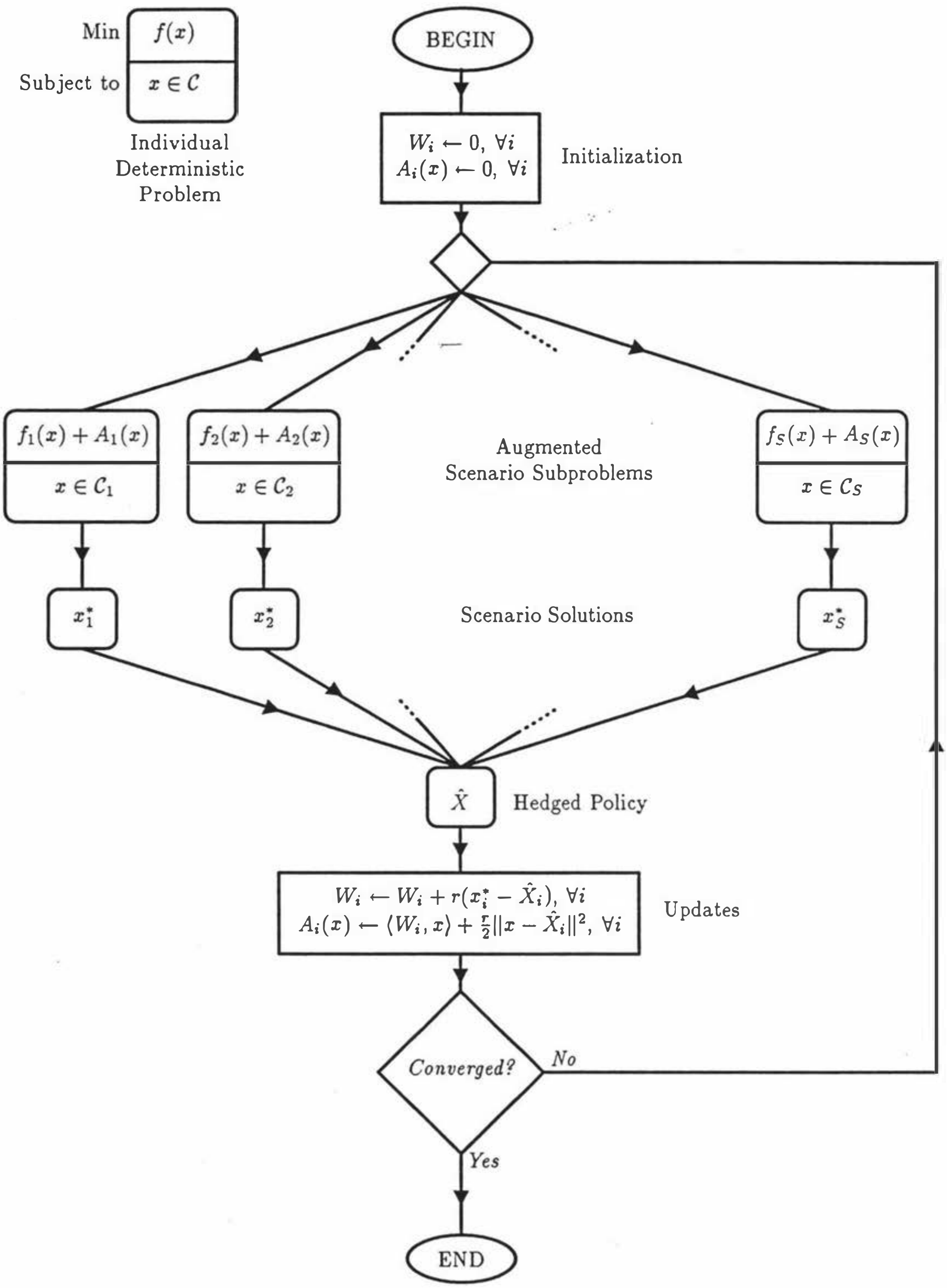


Figure 6.3: Rockafellar and Wets' *Progressive Hedging Algorithm*

One original intention was to solve the deterministic model under a number of different scenarios (compiled from historic data), only ensuring that the first week's releases were all the same. Currently, the option that appears most attractive (based purely on intuition) is to start with several (five, say) important *scenes* (yearly inflow sequences) based on the volume of inflow. From these scenes, then construct a scenario tree based on possible transitions from one scene to another as the year progresses, where some transitions are expressly forbidden (e.g. transition from a very wet scene to a very dry scene and then back again). The scenario tree so obtained would be unmanageably large (for five scenes we would obtain in the order of 5^Y scenarios). However, beyond some indeterminate horizon, the prediction of inflows is no longer very precise, and, the effect of wrong predictions on the first week's decisions is slight; we call this horizon the *short horizon*. It seems to be reasonable to approximate decisions made beyond this horizon by, say, deterministic solutions; other possibilities beyond this horizon are discussed in Section 6.8. Taking this short horizon to be four weeks gives in the order of 600 scenarios; from this a sub-tree can be chosen which has at most three arcs splitting from any node, and consists of 3 to 20 scenarios. This sub-tree defines the scenarios which are to be used.

6.7 Reducing Effort when Progressive Hedging

When using a scenario approach one creates a scenario tree, as in Figure 6.4, showing the various interactions between scenarios. Paths from the root node to a leaf node represent the individual scenarios (see Figure 6.5).

The Progressive Hedging Algorithm solves the scenarios individually as (augmented) deterministic subproblems. Notice that we require that some *nodes* in the scenario tree to be "*solved*" many times for each hedging iteration, to ensure a feasible solution in each deterministic subproblem. In Figure 6.5 the dashed ellipses represent sets of nodes which are solved more than once; notice that their *history* (path to the root node) must be the same in order that they can be considered as being solved multiple times.

In the setting where the nodes of the scenario tree actually represent quite large physical systems (such as the current problem), such an overlap would represent a large overhead in solution time. To avoid this we should therefore decompose

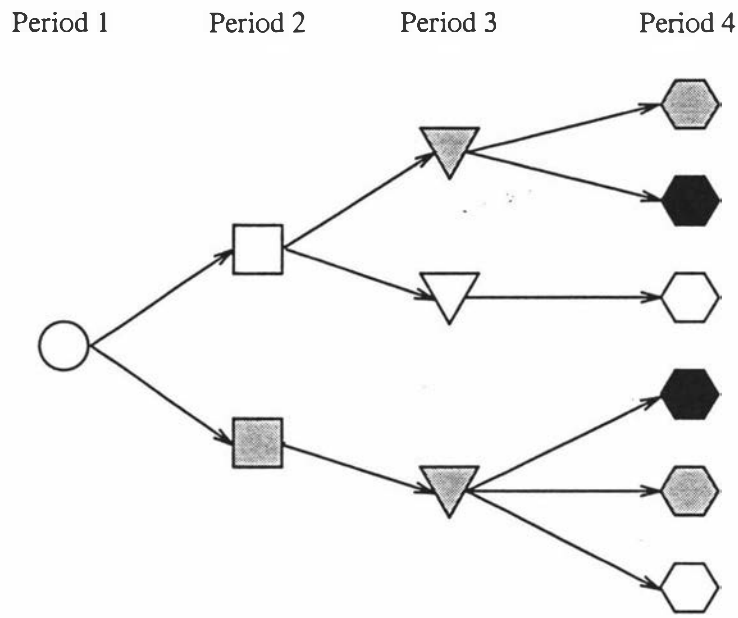


Figure 6.4: A scenario tree

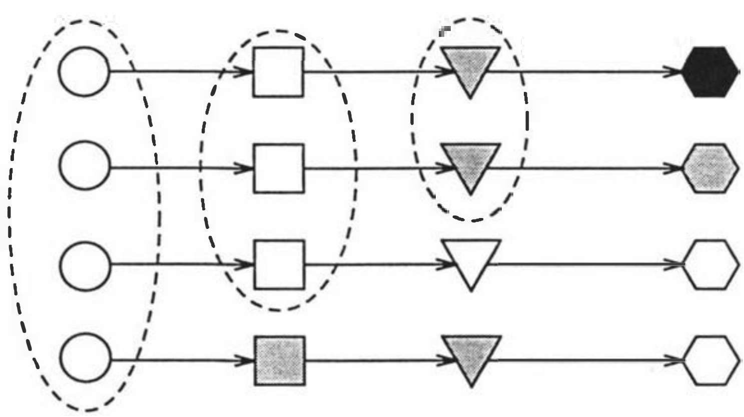


Figure 6.5: Various scenarios from the scenario tree

the system in another way. The reasons for choosing the natural decomposition (of Figure 6.5) are that it is easy to implement, easy to understand, and, that it exhibits similar structure in each subproblem (which may be exploited to obtain faster solutions or a more compact representation).

The Progressive Hedging Algorithm decomposes the problem by relaxing non-anticipavity conditions and then sequentially tightening them. Since the Progressive Hedging Algorithm makes *no* assumptions about the scenario structures used, one answer is to break up the scenario tree into subtrees rather than paths. For the model developed there are few (very important) inter-period links, namely the storage variables for each hydro station. It may, therefore, be better to split these variables "breadth-wise" rather than "height-wise", taking the sub-trees to be as in Figure 6.6. In doing so, we are actually hedging on a "non-anticipation variable", i.e. an artificial variable equal to the value of the non-anticipavity constraint. To give an example of the resulting difference in the decomposition used, consider the following: let

$$A_t \begin{bmatrix} u \\ x \\ w \end{bmatrix} = b_t \tag{6.4}$$

be the partial constraint set for period t, and

$$A^1_{t+1} \begin{bmatrix} w \\ y \end{bmatrix} = b^1_{t+1}$$

and

$$A^2_{t+1} \begin{bmatrix} w \\ y \end{bmatrix} = b^2_{t+1}$$

to be two scenarios of partial constraint sets for period $t+1$, where u and w are the variables linking periods $t-1$ and t, and, t and $t+1$, respectively, and x and y are the decision variables for periods t and $t+1$, respectively. Usually we would have the following partial constraint sets:

$$\begin{bmatrix} A_t & \vdots & 0 \\ 0 & \vdots & A^1_{t+1} \end{bmatrix} \begin{bmatrix} u^1 \\ x^1 \\ w^1 \\ \cdots \\ w^1 \\ y^1 \end{bmatrix} = \begin{bmatrix} b_t \\ \cdots \\ b^1_{t+1} \end{bmatrix}$$

$$\begin{bmatrix} A_t & \vdots & 0 \\ 0 & \vdots & A_{t+1}^2 \end{bmatrix} \begin{bmatrix} u^2 \\ x^2 \\ w^2 \\ \dots \\ w^2 \\ y^2 \end{bmatrix} = \begin{bmatrix} b_t \\ \dots \\ b_{t+1}^2 \end{bmatrix}$$

with non-anticipavity condition

$$w^1 = w^2 (= W)$$

(Note that we effectively solve Equation 6.4 twice). Decomposing as proposed above would split W and have the following partial constraint sets:

$$A_t \begin{bmatrix} u \\ x \\ W^a \end{bmatrix} = b_t$$

$$\begin{bmatrix} A_{t+1}^1 & \vdots & 0 \\ 0 & \vdots & A_{t+1}^2 \end{bmatrix} \begin{bmatrix} W^b \\ y^1 \\ \dots \\ W^b \\ y^2 \end{bmatrix} = \begin{bmatrix} b_{t+1}^1 \\ \dots \\ b_{t+1}^2 \end{bmatrix}$$

with non-anticipavity condition

$$W^a = W^b$$

Such a decomposition breaks the problem up into small pieces, whose size is controlled only by the number of arcs leaving a scenario tree node and the detail in the physical system of each period. However, we further propose to "glue" some of these sub-trees together so as to make *large* blocks of about the same size (e.g. Figure 6.7).

One advantage to this approach is that it implicitly incorporates the fact that the first period (the root node) is affected more by decisions made in close periods than by distant periods. There may be an asynchronous parallel solution method to solving these sub-trees (possibly an adapted version of the Progressive Hedging Algorithm) which takes advantage of this by having smaller trees at the "narrow" end of the scenario tree which are updated more often (but we do not pursue this possibility further).

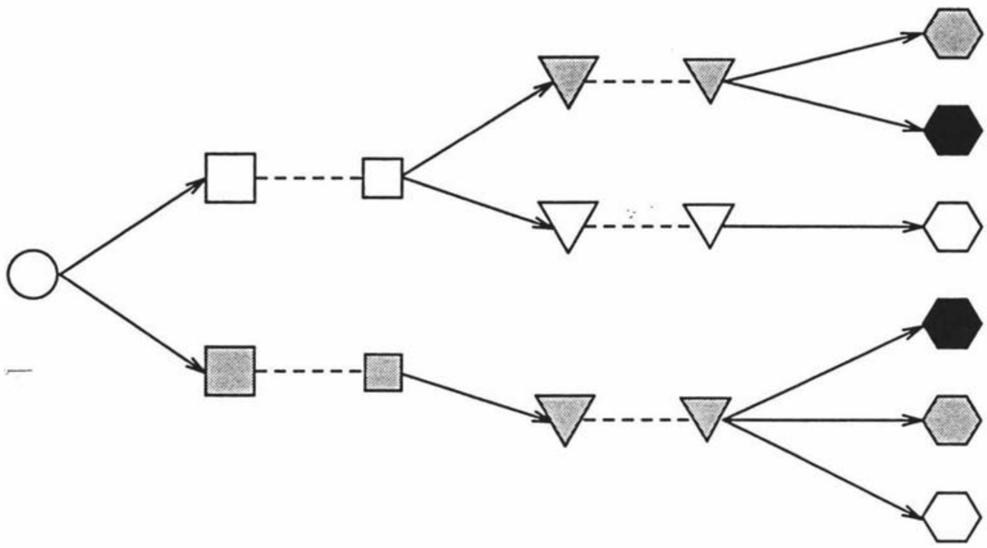


Figure 6.6: Split sub-trees

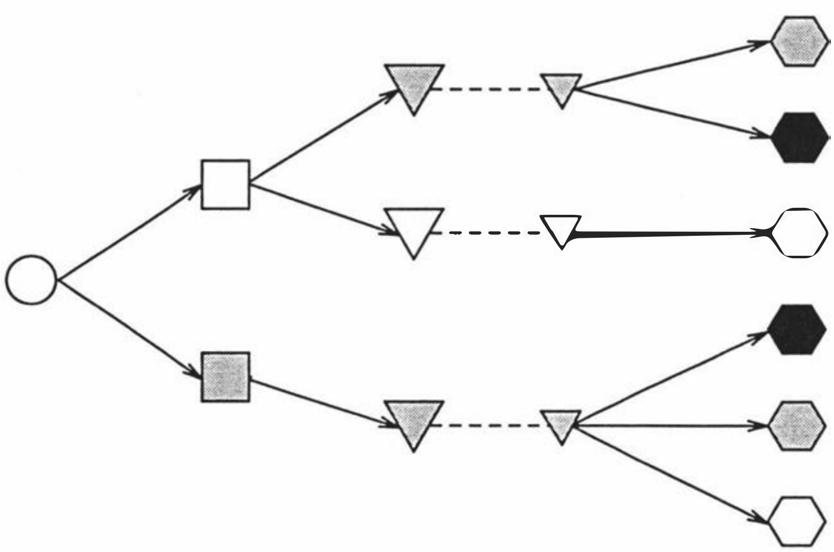


Figure 6.7: Joined sub-trees

The construction of these larger blocks is completely dependent on the base model and the scenario tree used. In solving the problem using the natural decomposition, the subproblems solved each use Y nodes of the scenario tree (where Y is the number of weeks in a year); it is therefore reasonable to break the problem up into subtrees, each with a depth of $\log_s Y$, where s is the average number of arcs splitting from a scenario tree node.

For scenario trees similar to those described in Section 6.6, a natural construction of the large blocks is evident: the first block consisting of the scenario tree with the deterministic "tails" removed (i.e. the part of the tree before the short horizon), and the other blocks representing the deterministic tails. Figure 6.8 shows an example of this. While this procedure removes only the same order of computation as the removal of *one* scenario (for each progressive hedging iteration), the number of relaxed non-anticipavity variables is decreased from 54 to 18; this should considerably reduce the number of iterations required. Most importantly, the new structure will increase the convergence rate of the first week's solution.

6.8 Non-anticipavity

In the scenario aggregation setting, the issue of when information on the observed values of random variables becomes available begs the question as to which variables should be the non-anticipation variables. Consistency of the model requires that the same variables are used as non-anticipation variables each week. The choice of non-anticipation variable is also dependent on the information to be sent to shorter time models.

Where the assumption about the observation of the random variables is either, that of perfect foresight over the week, or, the case where we can not react until the end of the week, both the hydro station's C.C. and release need to be based only upon the information available at the beginning of the week. This can be achieved by imposing the non-anticipavity condition only on the release, if desired, since, when the releases are fixed for a week, the decisions for that week become decoupled from the rest of the model. The case where we approximate partial foresight over the week (as in Section 6.2), amounts to requiring the hydro station's C.C. to be non-anticipated but the release to be anticipated, hence the release may react to the observed values of the random variables for that week.

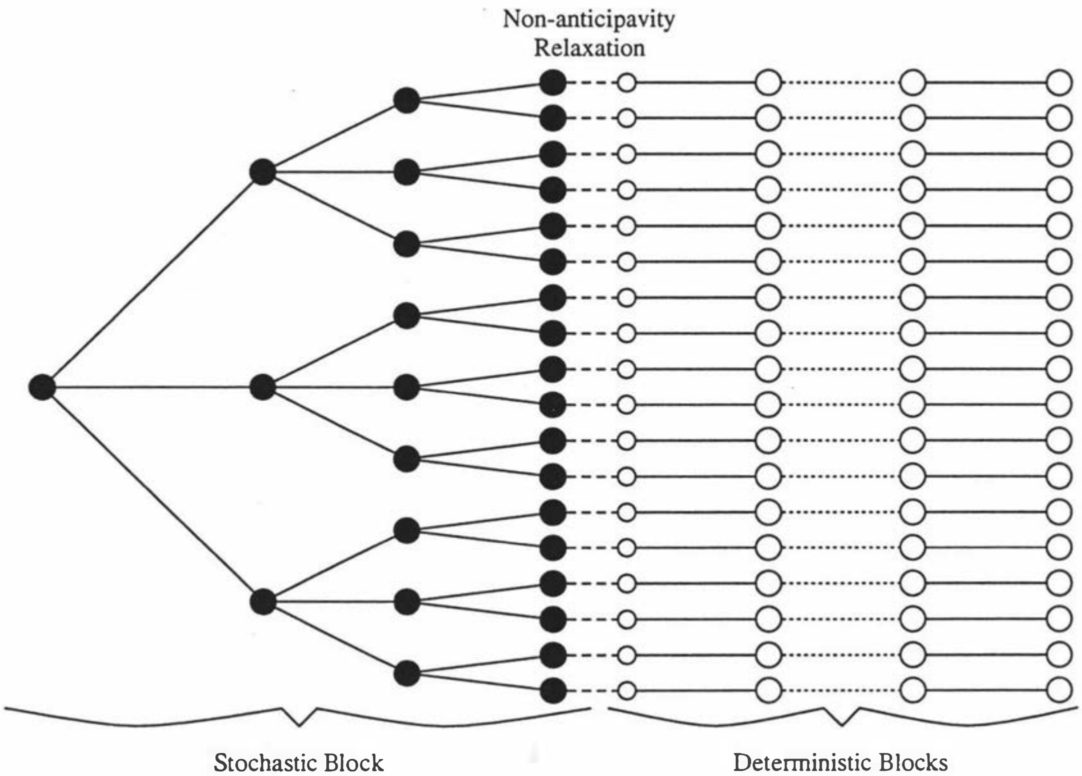


Figure 6.8: Suggested break-up of a particular scenario tree

Another method of approximating partial foresight is to use storage as the non-anticipated variable, and allow the release and generation to react to the observed values of the random variables—this amounts to setting storage targets. Unfortunately, in the situation where there is a large range of possible inflows for a week (which is common for later weeks), the target storage is limited by the possible storage given that the lowest inflow occurs, requiring that, for high inflows, the solution should *spill* rather than not meet the storage target!

A slight generalization may, however, work well in tandem with the short horizon proposed in Section 6.6. The idea is to approximate the effects of stochasticity beyond the short horizon, by applying the non-anticipavity conditions to the target storage level for *every* scenario (regardless of its history) for every week beyond the short horizon. Applying a penalty for under-achieving the storage target simulates stochastic decision-making for these weeks. Of course, the solution is dependent on the penalty actually used, but the effect on the first week's decision may not be overly great.

Such an approximation provides a setting for investigating the effects of different approximations beyond the short horizon. Having a low (or zero) penalty parameter is equivalent to assuming a deterministic solution beyond this point, whereas using a very high penalty assumes that no benefit can be obtained from considering either the past or possible futures of a decision regarding storage levels.

This approximation also comes with a warning. Both it, and the use of a short horizon, are artificial structures which are included *only* to reduce the amount of work involved in taking some account of the stochastic elements involved. They are not approximations of real phenomena and will induce unwarranted structure in solutions. The only way to fully evaluate their effect on solutions is to run and compare exhaustive empirical studies and simulations.

There are various possibilities regarding the choice of information which is given to the shorter time horizon models. The exact choice depends on the short term models and information exchange structures used. Communicating the generation schedule requires a minimum release from the short term model. Release information, or equivalently a target storage level, sent to the short term model requires maximum generation from this model. Often price information, in the form of water values, is given, in an attempt to provide local information in the face of uncertainty about load and inflow information. Unfortunately, this often leads to an oscillating off/on situation when the water value is near the fuel cost for some thermal station.

Incorporating information about the geographic location of load and power stations means that price information is in terms of the dual variables of a network, so that each power station "sees" a different value for its generation with respect to different load locations. If a single cost were required, we could identify some fixed load which a particular station is deemed to meet, and use this single value (or obtain an average value over all load met by the station). If partial (or no) foresight is allowed over the week, the price information with also depend on the inflow.

It appears that it is best to send both release information (possibly in the form of a target storage level) and some price information, on the value of stored water. This provides the short term model with a method for evaluating the effect of approximations used in the longer term model (in terms of the local solution at least) and may provide the basis for a decision on whether local price information is necessary in the longer term model. Unfortunately this would also cause confusion over which piece of information to use (the release or price) in scheduling the system. Price information is more robust under the assumption that it does not vary over the week.

We see the inclusion of another model as being a more robust option—one which considers the whole system for just one week (and which can be used to give the local stations and river chains more detailed daily information (even if this information is only given once a week).

6.9 Advantages of a Scenario Method

In choosing a scenario approach we take advantage of some hidden benefits. By discretizing the hydro reservoir inflows we can include constraints with a non-linear dependence on the hydro inflows. This is most useful for constraints can be relaxed in extreme inflow situations. For example consider constraints specifying minimum and maximum flow somewhere on a hydro river chain—when inflows are below (or above) certain specified levels, often the minimum (or maximum) flow level is relaxed to take account of this. This can be modelled by removing the constraint in a given week for a scenario with an extremely low (or high) inflow for that week.

6.10 Discussion

This Chapter considered many of the issues involved in moving from a deterministic model to a stochastic one, specifically for the deterministic model developed in Chapters 3–5. Various stochastic extensions which have been used to model similar systems were reviewed. It was decided that a scenario approach should be taken as this offered the most flexibility in the extent to which the stochastic extension is modelled.

Specifically, Rockafellar and Wets' *Progressive Hedging Algorithm* was choosen. Many advantages in taking a scenario approach and particularly in using the Progressive Hedging Algorithm have been highlighted. The implementation of both the deterministic and stochastic approaches (for testing purposes) are discussed in Chapter 9, however the enhancements proposed in this Chapter have not been implemented since any effective comparison would require a thorough investigation, which is beyond the scope of this thesis.

Chapter 7 extends the model to include some local constraints, which cannot be modelled via techniques already discussed. Chapter 8 describes a theoretical investigation of the effects of a particular approximation.

Chapter 7

House Rules

Up until this point the model has been developed for a general setting, ignoring locally applied constraints which are not central to the structure of the model. These local features need to be addressed if the model is to be used in practice. Their previous exclusion was to facilitate model development by avoiding the complications they create.

The features addressed in this Chapter are those which do not immediately fall within the framework of the model developed so far. There are many local idiosyncrasies which may be adequately dealt with within the current framework, e.g. the North-South DC link may be modelled with a greater precision than the other transmission lines merely requiring the use of two different approximations which have already been developed.

7.1 Huntly and Stratford

Huntly thermal station can be fuelled by any mixture of Maui Gas and coal. The amount of Maui Gas which can be used is constrained (see Section 7.2) and coal is used from a stockpile. The stockpile can be modelled in a similar way to the hydro waterflow networks with the "inflows" becoming decision variables. Since there is enough flexibility in the system to take adequate account of the coal stockpile by examining possible future needs for coal, this is not included in the working model. It may be useful, however, to include a stockpile model so as to model the cost of coal as the price paid when the coal is purchased, rather than as a cost applied when it is used.

To model the dual fuel aspect of generation at Huntly, the Contract Curve for Huntly, G_H, is split into two parts; G_g for generation by Maui Gas and G_c for generation by coal. The generation capacity constraint for Huntly becomes

$$G_g(0) + G_c(0) \leq \overline{Q}_H$$

where $\overline{Q}_H$ is Huntly's generating capacity. The amounts of Maui Gas and coal used over the week are given by $m(G_g)$ and $m(G_c)$, respectively. Huntly's contribution to the objective function is

$$c_g m(G_g) + c_c m(G_c)$$

where c_g and c_c are the cost of Maui Gas and coal generation at Huntly, respectively.

Stratford thermal station is also fuelled by two fuels, Maui Gas and Waihapa Gas. Waihapa Gas is otherwise flared and so it is base-loaded at the Stratford plant. This is modelled by treating the Waihapa Gas generation as an auxiliary station, and derating Stratford's capacity accordingly. However, if required, the dual fuel nature of Stratford may be modelled in the same manner as the Huntly dual fuel.

7.2 Gas Deliverability

Maui Gas is extracted off the Taranaki coast. It provides fuel for three stations, Huntly, New Plymouth and Stratford, via a single pipeline. This pipeline imposes various constraints on the amount of gas which can be used. The nature of the pipeline provides a buffer to changes in the rate of gas usage.

The C.C.'s of Huntly Maui Gas generation, G_g, the New Plymouth thermal station, G_{NP}, and the Stratford thermal station, G_S, give the *generation* from Maui Gas at each of these thermal stations. To obtain the *actual* usage of Maui Gas, we use the reciprocal of the generation efficiency, called e_j for station j. A maximum level constraint would have the form

$$\frac{1}{e_g} G_g(0) + \frac{1}{e_{NP}} G_{NP}(0) + \frac{1}{e_S} G_S(0) \leq \overline{\text{Gas}}_L$$

where $\overline{\text{Gas}}_L$ is the maximum instantaneous usage of gas. Due to an ability to use the gas at a faster rate than it is pumped into the pipeline, for short periods, the maximum level is a very soft constraint and so it is not used in the working model. Maximum usage constraints (over an integral number of weeks) are given by constraints of the form

$$\sum_{w \in \mathcal{W}} \left(\frac{1}{e_g} m(G_{g,w}) + \frac{1}{e_{NP}} m(G_{NP,w}) + \frac{1}{e_S} m(G_{S,w}) \right) \leq \overline{\mathrm{Gas}}_{\mathcal{W}}$$

where $\mathcal{W}$ is the set of weeks that the constraint is taken over, and, $\overline{\mathrm{Gas}}_{\mathcal{W}}$ is the maximum total Maui Gas usage over this period.

The working model includes a constraint for only the maximum Maui Gas usage over each week. This is generally the tightest of these types of constraints.

7.3 Security of Supply

Security of supply constraints are conditions imposed upon the system so as to help to maintain a level of security of supply in the face of forced outages and transmission failures, including running some hydro station turbines so that the turbine is spinning but not producing any output; this is to allow a quick reaction to failures in the system, and is known as a Spinning Reserve. The other major security of supply constraint is that of ensuring that no station, nor the North-South DC link, may supply more than a fixed fraction of the total generation for a particular Island (including the contribution from the North-South DC link), at any time.

To accurately model the optimal scheduling of Spinning Reserve within our model would require integer variables, since the generating characteristics of the amalgamated station are different if a turbine in one of the stations on the river chain is being used as Spinning Reserve. A better method is to treat this constraint as part of the generating characteristics of the river chain, effectively choosing which stations will provide Spinning Reserve prior to solution. Interaction with shorter time horizon models may be used to fine-tune such constraints for the first week. The effect is to derate the capacity (and also possibly the generation efficiency), and impose minimum generation, release or spill bounds on the (amalgamated) hydro station involved.

For the constraint on the maximum generation of a particular station we consider only northwards transmission on the North-South DC link; this is easily generalized to be any power station or transmission direction on this power line.

The constraint required can be formulated as

$$X_{S\text{-}N}(t) - \frac{\alpha}{1-\alpha} \sum_{j \in \text{NORTH}} G_j(t) \leq 0 \tag{7.1}$$

where NORTH is the set of all power stations in the North Island for this week, and $X_{S\text{-}N}$ is this week's northwards transmission on the North-South DC link. Since the piecewise quadratic on the left-hand side of Equation 7.1 is not necessarily decreasing, we need to apply similar constraints as those applied to transmission lines in order to enforce the transmission capacity constraints (as in Chapter 5).

This constraint is probably better achieved by using another approximation which involves fewer variables, and so should create less overhead in solution. We assume that the total generation for the North Island can be approximated by the total North Island load increased by a factor β (approximating average line loss). The constraint can now be written as

$$X_{S\text{-}N}(t) - \alpha(1+\beta) \sum_{j \in \text{NI}} L_j(t) \leq 0 \tag{7.2}$$

where NI is the set of all North Island nodes. The constraints used to enforce transmission capacity are used here to enforce Equation 7.2.

No security of supply constraints are used in the working model.

7.4 Implementation of these Constraints

A number of the constraints described in this Section are rarely active at an optimal solution. This means that including them will often increase solution time, for little or no gain in the quality of solution obtained. This is particularly important when considering that the solution method used requires deterministic subproblems to be solved many times.

In tuning the model to provide reduced solution time, constraints which are rarely active (such as some of those given in this Chapter) should therefore be omitted from the model for an initial solution. If any are violated by the solution, the model can be re-run (from this solution) with the appropriate constraints, which were initially not present, included (possibly all of them) to obtain an optimal solution which is now feasible in terms of both those constraints included and not included (hopefully).

Given the stochastic nature of the problem being solved, it may be allowable for some of the constraints which have been left out, to be violated slightly in a later week of some scenario. This is because the remedial action may not greatly affect the first week's solution, which is the solution of most interest, and the constraint is just an *approximation*, so that, although it may be violated in the model, the *actual* constraint may not be violated by the solution implemented.

Chapter 8

Function Formulation

Modelling a complex system inevitably requires approximations. These approximations have an indeterminate effect on the solutions given by the model. It is hoped that enough of the *essence* of the system is endowed in the model so as to provide good solutions in practice. One way of investigating the effects of approximations made, *as a whole*, is by empirical testing and simulation. However, this only highlights the symptoms of the approximations made, and isolating the approximations from which superfluous structure arises, is often difficult.

We seek to better understand, qualitatively, some of the approximations made in modelling the system. The approximation under investigation in this Chapter is that used for the Load Duration Curve. To investigate possible effects of this approximation, we generalize the way the LDC is modelled so as to bring it "closer" to reality.

There is an (unknown) point at which the Mathematical tools available are not sufficient to study the differences between the approximation and reality. The difficulty, here, is characterizing the reality rather than isolating some of the essential ingredients of that reality (which is the main tool of Mathematics). Our attempt to get closer to reality may admit unrealistic solutions. It may also be that reality does not exhibit worthwhile, exploitable, properties which can be used to facilitate such an investigation.

There is, however, still value in such an investigation, with the benefit of empirical testing and simulation to complement it. Where empirical testing highlights symptoms, qualitative investigations, such as this one, indicate causes and possible remedial actions.

8.1 The Generalization

In investigating effects of the approximation used, it is not enough to only examine the difference between the actual LDC and the approximation. We are more interested in the effect on the model (and hence on the optimal solution) of this approximation. The intention is not to pursue such an investigation to its bitter end, but to provide a framework within which such an investigation may be carried out.

Consider allowing all Electricity Curves to be any function from some set of implementable schedules, in terms of the system; call this set $\mathcal{I}$. With no loss of generality we can assume that $\mathcal{I}$ is a subspace of the space of all Riemann Integrable functions. We are interested in investigating how the optimal solution to the generalized model differs from that of the original (where the Electricity Curves are piecewise quadratics).

Consider the Mathematical Program given in Table 8.1 defined over $\mathcal{I}$, which is a subspace of the space of Riemann Integrable functions on the set S. Note the specific form of the Objective and Constraints. The reason for this form is that these constraints become linear under the transformation used to make the formulation tractable. Further, these constraints lend themselves well to formulating a system with an underlying Network, or Generalized Network, structure, in which the arcs transmit functions rather than values. It also generalizes most of the functional constraints given in the deterministic formulation of Table 5.2. The definition of each variable is given in Table 8.2.

Table 8.1: Function formulation

$$\text{Min} \sum_{i\in\mathcal{V}} c_i \int_S F_i(t)\ \mathrm{d}t$$

subject to:

$$\sum_{i\in\mathcal{V}} A_{ij}F_i(t) = Y_j(t) \quad \forall t \in S\ \forall j \in \mathcal{E} \tag{8.1}$$

$$\sum_{i\in\mathcal{V}} a_{ik} \int_S F_i(t)\ \mathrm{d}t \geq y_k \quad \forall k \in \mathcal{C} \tag{8.2}$$

$$F_i \in \mathcal{I} \quad \forall i \in \mathcal{V}$$

In the form given, the Mathematical Program is unsolvable, as $\mathcal{I}$ may be of

Table 8.2: Variables used in Table 8.1

$\mathcal{V}$	Variable index set.
$\mathcal{E}$	Functional equality constraint index set.
$\mathcal{C}$	Integral constraint index set.
F_i	Decision function. (A decision variable in $\mathcal{I}$.)
Y_j	RHS function, of the jth (functional) constraint. (Element of $\mathcal{I}$.)
y_k	Right hand side of the kth integral constraint. (Real)
c_i	Cost associated with variable F_i. (Real)
A_{ij}	Coefficient of F_i in the jth constraint. (Real)
a_{ik}	Coefficient of $\int_S F_i(t)\, dt$ in the kth constraint. (Real)

infinite dimension. The following definition facilitates the transformation of this Mathematical Program to an LP.

Definition 8.1 *Let $\mathcal{V}$ be a (possibly infinite dimensional) vector space, and $\mathcal{B}$ a finite dimensional subspace of $\mathcal{V}$. If every element of $\mathcal{V}$ can be approximated "well" by some element of $\mathcal{B}$, then $\mathcal{B}$ is said to* **approximate** $\mathcal{V}$ *"well".*

In the above definition the term "*well*" is purposefully left vague, as what constitutes a function being approximated well is often dependent on the circumstances of the approximation.

If there is some subspace, $\mathcal{B}$, of $\mathcal{I}$ with basis $\{B_1, \ldots, B_N\}$, which approximates $\mathcal{I}$ well, then we can approximate the Mathematical Program above by a Linear Program. Write the approximation of each Y_j, and F_i, as:

$$Y_j(t) = \sum_{n=1}^{N} y_{nj} B_n(t) \quad \forall t \in S$$
$$F_i(t) = \sum_{n=1}^{N} f_{ni} B_n(t) \quad \forall t \in S$$

The formulation in Table 8.1 can be approximated as shown in Table 8.3. The form of the formulation given in Table 8.1 is overly restrictive, partially due to constraints of the form $f(t) \geq g(t)\, \forall t$ being all but impossible to implement without knowledge of the structure of $f(t)$ and $g(t)$ *a priori*. If the matrix of coefficients in Equation 8.1, $[A_{ij}]$, has a Network or Generalized Network structure, this structure is preserved by the transformation. In this case capacity constraints of the type given in Equation 8.2 transform to side constraints (rather than capacity constraints as might be anticipated).

Table 8.3: Approximation of the formulation of Table 8.1

$$\text{Min} \sum_{i\in\mathcal{V}} \sum_{n=1}^{N} \left(c_i \int_S B_n(t)\, \mathrm{d}t \right) f_{ni}$$

subject to;

$$\sum_{i\in\mathcal{V}} A_{ij} f_{ni} = y_{nj} \quad \forall n \in \{1,\ldots,N\}\ \forall j \in \mathcal{E}$$

$$\sum_{i\in\mathcal{V}} \sum_{n=1}^{N} \left(a_{ik} \int_S B_n(t)\, \mathrm{d}t \right) f_{ni} \geq y_k \quad \forall k \in \mathcal{C}$$

$$f_{ni} \in \Re \quad \forall n \in \{1,\ldots,N\}\ \forall i \in \mathcal{V}$$

Equation 8.2 and the Objective Function of the Table 8.1 formulation can be generalized whilst retaining the same approximation. Assume we have M measures defined over S, $\{\mu_1,\ldots,\mu_M\}$, such that every $F \in \mathcal{I}$ is μ_m-measureable. The Objective Function can be generalized to

$$\sum_{m=1}^{M} \sum_{i\in\mathcal{V}} c_{im} \int_S F_i \, \mathrm{d}\mu_m$$

and Equation 8.2 generalized to

$$\sum_{m=1}^{M} \sum_{i\in\mathcal{V}} a_{ikm} \int_S F_i \, \mathrm{d}\mu_m \geq y_{km} \ \forall k \in \mathcal{C}$$

Note that this form includes constraints and costs pertaining to both integration over some subset of S and the value of F at particular points in S.

We would like to show that if $\mathcal{B}$ is close to $\mathcal{I}$ then the formulation given in Table 8.1 is also close to that given in Table 8.3, in the sense that their optimal solutions are close.

8.2 Unsettling Results

It appears that the linear approximation in Table 8.3 should be *close* to the original formulation of Table 8.1, especially as we take better approximations by increasing the size of $\mathcal{B}$. However, we can have the case where the original has a unique optimal solution, but some or all linear approximations have no feasible solution. Consider the following:

Let $\{\mathcal{B}_n\}$ be a sequence of subspaces of $\mathcal{I}$ with the following properties: each $\mathcal{B}_n$ has $\{B_1, \ldots, B_n\}$ as a basis, making $\mathcal{B}_{n+1}$ a refinement of $\mathcal{B}_n$; for each $F \in \mathcal{I}$ there exists a unique approximation in $\mathcal{B}_n$ which shall be denoted $F_n = \sum_{k=1}^{n} f_k B_k$; $\mathcal{I}$ is approximated well by each $\mathcal{B}_n$, with $\|F - F_n\| \rightarrow 0$, where $\|F\| = \int_S |F(t)|\, \mathrm{d}t$.

We write $\mathcal{B} = \bigcup_{n=1}^{\infty} \mathcal{B}_n$, and refer to the sequence $\{\mathcal{B}_n\}$ by referring to $\mathcal{B}$ or to $\{B_1, B_2, \ldots\}$, the basis of $\mathcal{B}$. The LP obtained by using $\mathcal{B}_n$ as an approximation of $\mathcal{I}$, shall be denoted $\mathcal{F}_n$, with the original formulation being $\mathcal{F}$.

Consider the formulation of Table 8.4, in which $\mathcal{I}$ is the space of all Riemann Integrable functions. It has a single feasible point, and hence unique optimal solution, at $F = e^{-t}$. Consider approximations based on the polynomials, $\{1, t, t^2/2, \ldots, t^k/k!, \ldots\}$; $\mathcal{F}_n^{(1)}$ (corresponding to $\mathcal{F}^{(1)}$) is given in Table 8.5. This formulation only has feasible solutions for odd n.

Table 8.4: $\mathcal{F}^{(1)}$

$$\text{Min } c \int_0^1 F(t)\, \mathrm{d}\, t$$

subject to:

$$\begin{aligned} F(t) &= e^{-t} \\ \int_0^1 F(t)\ \mathrm{d}t &\leq 1 - e^{-1} \\ F &\in \mathcal{I} \end{aligned}$$

Table 8.5: $\mathcal{F}_n^{(1)}$

$$\text{Min } \sum_{k=1}^{n} \frac{c}{(k+1)!} f_k$$

subject to:

$$\begin{aligned} f_k &= (-1)^k && \forall k = 1, \ldots, n \\ \sum_{k=1}^{n} \frac{f_k}{(k+1)!} &\leq 1 - e^{-1} \\ f_k &\in \Re && \forall k = 1, \ldots, n \end{aligned}$$

When the formulation of Table 8.6, $\mathcal{F}^{(2)}$ is approximated by the same basis, *none* of $\mathcal{F}_n^{(2)}$ has a feasible solution, while the original formulation still has a unique optimal solution.

Table 8.6: $\mathcal{F}^{(2)}$

$$\text{Min } c\int_0^1 F(t)\ dt$$

subject to;

$$F(t) = e^t$$
$$\int_0^1 F(t)\ dt \geq e-1$$
$$F \in \mathcal{I}$$

However consider the basis

$$\{e-1, (e-2)(2t-1), (e-5/2)(3t^2-2t), (e-8/3)(4t^3-3t^2), \ldots\}$$

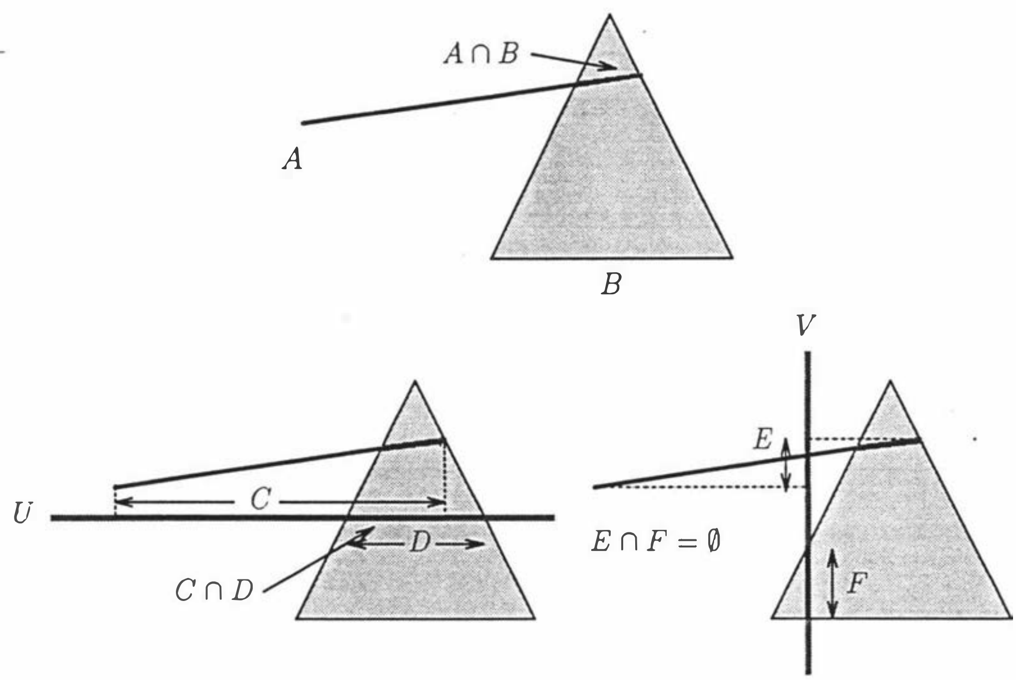
where the k'th basis element is a polynomial of degree k such that the sum of the first k basis elements is $1+t+t^2/2+\cdots+t^{k-1}/(k-1)!+\alpha t^k$, with α chosen so the integral over $[0,1]$ of this sum is $e-1$. Using this basis, each $\mathcal{F}_n^{(2)}$ has a unique optimal solution, and these optimal solutions have the optimal solution of $\mathcal{F}^{(2)}$ as their limit.

The reason for this phenomenon is that the equality constraints are *projected* onto the subspace while the integral constraints are *restricted* within the subspace. This means that the intersection of the two transformed regions is neither the projection nor the restriction of the intersection of the original two, and so depends heavily on the subspace used. Figure 8.1 shows how the choice of subspace can adversely change the intersection of the two resulting sets.

8.3 Discussion

Notice that the bad cases presented are very specific and rely on a restrictive feasible region. In most applications the system being modelled has greater freedom; it is difficult to express this freedom in Mathematical terms.

This investigation does, however, accentuate some positive aspects. In considering the approximation to use, we must not only take into account the approximations of the right-hand side functions, but also the structure imposed upon these



Set A is a line segment and set B is a triangle. Using the line U as the subspace, C is the *projection* of A onto U, and D is the *restriction* of B to U; $C \cap D \neq \emptyset$. Using the line V as the subspace has E the *projection* of A and F the *restriction* of B; here $E \cap F = \emptyset$.

Figure 8.1: The projection/restriction of two sets onto two subspaces

functions by the rest of the model. This was done in our model when investigating the approximation for the Electricity Curves in Chapter 3.

The need for investigation into the approximation of potential solutions as well as right-hand side functions is further reinforced by the experiences of Chapter 4. In that Chapter, an approximation which appeared reasonable in approximating LDC's (the normal approximation) led to badly structured solutions due to an inability to model potential solutions well (those with flat generation for part of the week).

The investigation of this Chapter could be taken much further. We could investigate the limit of the optimal solutions to each $\mathcal{F}_n$ in the case where each optimal solution exists. The difference between a particular $\mathcal{F}_n$ and $\mathcal{F}$ could also be examined. Further constraints with a linear interpretation could be imposed, or we could explicitly limit $\mathcal{B}$ to having particular structure so as to allow the application of a more diverse range of constraints. However, such an investigation would prove lengthy and, most likely, provide little help with the actual modelling of the system considered. Such an investigation is beyond the scope of this thesis and is not pursued further here.

Chapter 9

Implementation and Results

One of the most time-consuming aspects in the development of the model was the computer implementation. Often a seemingly minor change to an approximation within the model can lead to a major revision of the computer implementation. Ensuring that the computer implementation corresponds to the model is a long and arduous task, but is necessary to eliminate many unworkable approximations considered.

This Chapter discusses two phases of implementation: implementation of the working model for the purposes of testing and exploring the consequences of various approximations; and, some of the issues involved in massaging the final implementation into a finished product. We emphasize that the intention is *not* to create a finished product; insights gained from model development may prove useful in the further development of the model and in solution strategies to be used in a finished product. The initial implementation was developed in tandem with the model, and it was through testing of the model at various stages that many refinements were made and unsuitable approximations were highlighted.

For both phases of implementation there are many issues involved in addition to the computational expression of the model and the coding of an algorithm. For the different phases these issues need to be addressed differently: e.g. in final implementation, the input of data needs to be integrated smoothly into the finished product; however, in testing the working model, we will often require data in different forms and so the data manipulation should be more flexible. Flexibility in the development phase is of paramount importance; it is impossible to anticipate later needs for the model. In comparison, the final implementation needs to professional and efficient.

The results given here are derived from testing, and highlight features of solutions. Since these results are from a developmental model, they can not be directly compared with those from other methods; instead they are intended to provide guidelines on the solution structure and computational feasibility.

In this Chapter we do not address the question of how well the algorithm performs in an uncertain environment. This *important* question requires much further work to be addressed with any authority, and requires comparison with other methods. Any lesser investigation cannot adequately investigate the method's ability to react to the uncertain future, or compare it with contemporary methods.

In this Chapter we also present suggestions on the form of the model which should provide a basis for a final implementation.

9.1 Input Data

For any final implementation the input data needs to be provided by an integrated system which provides direct access to the necessary data, and direct conversion to necessary data formats. Care should be taken to ensure reasonable flexibility, but efficiency and robustness are more important issues. Data used needs to be accessible and its rôle within the system needs to be clear. The format and application of this is entirely dependent on the system within which the model is integrated.

The form of data input into the working model is not so well defined. It needs to be flexible enough to allow for format changes or for further manipulation. It is also useful to be able to manually change some of the data for experimentation, so the formats used need to allow for this. In this Section we examine the integration of data from five major areas: load data, transmission network data, thermal station descriptions, hydro system physical descriptions, and hydro inflow data.

To allow flexibility (so as to make debugging the process easier) the information needed for a deterministic solution was collated in a collection of files (called `WeekRef` files), one for each week, which contained all of the information about the entire system model for that week. A master file, `Master`, gave information and parameters which were relevant to the complete problem. These weekly files were then combined into a format which could be used as input to the solution platform.

Appendix A gives input files for a single example. The descriptive file names given both here and in the Appendix are to allow easier reference.

9.1.1 Load

The raw data on load is past data, given as the average load over each half hour for the North and South Islands, with various constant loads removed. The load at any node is represented by a portion of the total load for the Island it originated from, combined with any previously removed constant load deemed to originate at that node. The generation by auxiliary stations is removed from the load at this point.

To achieve this, a file, `NodeRef`, is created for each geographic network used. This file contained information on the distribution of the each Island's load, the constant loads, and auxiliary generation over the geographic network. In this file every node was named and the node *order* specified the node numbers. These node names were used by other procedures to specify station placement and transmission line end points, so as to facilitate changes to the geographic network.

A routine was written in Matlab to convert the data into a weekly 4-piece piecewise quadratic LDC for each node of the geographic network. The year-long load curves were converted into weekly LDC's, from which the best least-squares, 4-piece piecewise quadratic approximation was found and heuristically converted to be decreasing. Routines for converting from the natural basis to the basis described in Section 5.4 (and *vice-versa*) were also used.

To allow scaling (so as to facilitate solution robustness), a routine was included which scaled the coefficients so that each had an average (over all coefficients of the same basis element) of order 1. Allowance for the use of different geographic networks and different sized weeks, for the same system model, was made through the use of different input files.

The information obtained was transferred into the `WeekRef` files in the form of an LDC for each node. Scaling information was appended to the master file.

9.1.2 Transmission

Raw transmission data was given in the form of line characteristics of the higher voltage lines which make up major components of the National Grid. The make-up of arcs (in terms of these major lines connected in series and parallel) were also given, so that characteristics of the *arcs* could be produced. For each arc of the geographic network a representative capacity was estimated. In practice the capacity of components of the National Grid depend on the actual load and generation over the whole of the New Zealand system; such complex dependencies were deemed as unsuitable to be modelled here.

From this data, two types of file are created. The first, `LineRef`, gives representative loss characteristics (at 100MW) and the voltage level of each power line used in a geographic network arc. This file also specifies the file names of the second type of file, `ArcRef`, which represent the geographic network for each week. These secondary files give the power line representation of the transmission arcs, so that the arc loss can be calculated. The capacity of the transmission arcs and the nodes to which each arc is connected are also given. The power loss on the North-South DC link is specified as its resistance.

The arcs specified in the `WeekRef` files are directional arcs. The information in the `WeekRef` files pertaining to each arc is: a linear approximation of power loss, a representative capacity, and the entering and exiting nodes of the arc.

9.1.3 Thermal Stations

The raw data for thermal stations was included in an input file intended for the current program used at ECNZ to schedule the system. This format was unacceptable for use as an input file, partially due to its size, to difficulties in interpreting all data contain therein, and to difficulties in changing the data when required. The data was, therefore, manually compiled into a single file, `ThermalRef`, giving data on each fuel type, together with constraints applied to some of the fuels (see Chapter 7) for each week.

The data given for each thermal station is: its forced outage rate (which is assumed to be constant over the year), its capacity over the year (including scheduled outages), the fuel type, and a conversion factor (the "heat rate") for converting fuel into electricity. The node at which the station appears for each week is also specified. Each station is given a name so as to distinguish stations at the same node, and allow the appropriate stations to have fuel constraints applied.

For fuels, the cost and calorific value of the fuel are given, together with constraints on the usage of that fuel. The calorific value is used only in the specification of fuel constraints, and to allow fuel consumption information in the output.

The information from this file is included in the `WeekRef` files in the following form: for each station its associated node, generation cost, capacity, and forced outage rates are included; the coefficients of the various stations in the Maui Gas deliverability constraint are also given.

9.1.4 Hydro Station Data

The raw data for the physical attributes of the hydro systems were also included in the input file intended for the current program used at ECNZ. This was manually transferred into three files, `HydroRef`, `StationRef`, and `InflowRef`.

`HydroRef` specifies the attributes of each hydro river chain, including the node at which it is based, its initial storage level and the stations and inflows which form this river chain. It also specifies the directory which holds all of the inflow data.

`StationRef` describes the attributes of each hydro station, giving the controlled and uncontrolled inflows which feed that station. This file also gives the generation capacity, generation efficiency and fraction of uncontrolled inflow which passes through each station. A fraction of uncontrolled inflow which represents the amount of this inflow which arrives at the hydro station but which cannot be stored *during* the week, is also included.

`InflowRef` specifies data on the inflows, and reservoirs of the river chains. For inflows into reservoirs, the maximum level of the reservoir is specified. For both controlled and uncontrolled inflows, bounds on the flow level at the beginning and end of the river chains are specified. The name of the file containing the inflow data, the starting date of that file (the information in all files ends on the same date) and the start date to be used for this particular scenario are also included.

The specifications for hydro stations in the weekly `WeekRef` files are complex. Each hydro station in this file corresponds to a hydro system specified by the previous files. The node at which each hydro station is present is given, along with a "hydro number" which specifies which hydro stations have been aggregated into the *current* hydro station (this is to allow correct specification of the waterflow networks). The hydro number for hydro stations of week one are unique powers of two; those for aggregated hydro stations (in later weeks) are the sum of the hydro numbers of the stations combined. For each hydro station, the amount of controlled and uncontrolled inflow, the fraction of uncontrolled inflow which cannot be stored *during* the week, and the generation and storage capacity are specified. There are also specifications on the maximum and minimum release rates from the reservoir, and flow rates from the river mouth (which is the flow rate leaving the final station on the river chain).

9.1.5 Inflow Data

Inflow data is given in a file for each inflow, which specifies the level of inflow, in average litres per second over the week. This data is used for both the construction of the `WeekRef` files and for directly converting the solution platform input files, to represent a new scenario (as to be described in Section 9.2).

9.1.6 Weekly System Input Data

Information from the above files is compiled into the `WeekRef` files. The form of these files is easy to read and change, and much more compact than the MPS format—this relative compactness is mostly due to the specialized nature of MPS format (for a case in which *all* of the `WeekRef` files required half a Megabyte, the corresponding MPS file was over 10 Megabytes!). A program written to change inflow scenarios directly using the MPS file meant that the `WeekRef` files can be seen to embody the *structure* of the particular model version, and so are used to store "fleshed out" models. The layout of the files made it particularly easy to construct simple examples by hand for debugging purposes.

Each section of these files was created separately to allow for the different forms of raw data given. This also allowed changes made to the individual sections to be self-contained, thereby allowing easier experimentation in the modelling of various aspects of the system.

9.2 Solution Platform

The solution platform used was MINOS 5.4. The platform needed to be flexible enough to withstand changes in the model—indeed, the initial model being solved was non-linear. MINOS also proved to be effective in the solution of the stochastic case, since it can be used as a subroutine to a master program. The input format for specification of the linear part of a formulation into the stand-alone version of MINOS is MPS format. A routine was written to convert the `WeekRef` files into MPS format. This routine also produces some of the parameters necessary for MINOS, which are formulation specific. For easy interpretation of the output data, another file (known as the "info-file") is produced by this routine which contains information, relevant to every constraint and variable, which is not directly obtainable from the MPS file or MINOS output files.

Since the volume of inflow affects only some right-hand side and bound values, a program was written to convert an MPS file from describing one inflow sequence to describing another. This removes the need to reconstruct all of the `WeekRef` files when the only change is to the inflows. This program is most useful in conjunction with the stochastic model.

For solution of the non-linear version of the model (described in Chapter 4), C functions were written to compute the value of the objective and its differential, as required during solution by MINOS. For testing of the objective function and model, various small test problems were created which focussed on particular areas of the model. However, for the reasons noted in Chapter 4, this model was dropped as being unsatisfactory.

9.2.1 Viewing Solutions

The model is large. A 52 week example, which uses 17 nodes in each geographic network, was specified by 42 000 variables, 40 000 constraints, 260 000 non-zero matrix elements and 19 000 non-zero objective coefficients. The size of the model together with difficulties inherent in interpreting solution values of various basis coefficients, made it impossible to investigate solutions manually. The sheer size of the solution output made producing a full hard copy of the solution (even for just one week) infeasible, so Matlab was used as a platform through which to view solutions graphically.

To facilitate solution viewing, a program was written to convert the information output from MINOS and the information contained in the "info-file" to a form more readily usable by Matlab. The GUI (graphical user interface) features of Matlab 4 where used to allow easy movement through the information, and compilation of some overall statistics. An example of the output given is shown in Appendix B for the input files given in Appendix A.

The solution times were considerable for the large problems. Using only the "crash" start option of MINOS (i.e., where no initial solution is specified) solutions take in the order of 30 hours, for the model version specified above. Starting from the optimal solution to another "scenario", for the same model version, takes in the order of 12 hours, but there a large variation in this solution time depending on how "close" the scenarios are. This suggests very long solution times for only moderately sized stochastic extensions.

The solution times indicated above need to be taken in context; they are for an "untuned" solution platform, with no "reasonable" choice of starting solution. They are also more detailed than would be appropriate for a final implementation; the physical system of the first week is as intricately detailed as that for the last week, whereas in practice this would seem to add unnecessary difficulties for negligible improvement in accuracy. Finally, the solution platform takes no account of the structure of the problem, which could be exploited in a tailored solution platform.

Investigation of the fully detailed model is useful for the purpose of highlighting implementation issues for the entire model. The necessity for the investigation of these issues, in the context of a more highly detailed model than would be used in practice, is to allow insight into the interactions of approximations, and the feasibility of the model as a whole. It is *not* intended that this full model version should be used in a final implementation. The final implementated version of the model needs to be easily expandable, and flexibility in the entire model allows this. Also, describing the universal model, of which the final model is a part, allows interpretation of some of the approximations made in the final model and investigation of the accuracy of the implemented model through comparison with more detailed models, which are not appropriate for generating weekly generation schedules.

Probably the most expensive approximation used, in terms of problem size and solution time, is the Electricity Curve approximation. Forcing the Electricity Curves to be continuous removes about a quarter of the variables, and a number of constraints, for little loss of approximation. Moving to piecewise linear Electricity curves reduces the number of variables by a third, and the number of constraints is also reduced.

An investigation of the benefits and the concomitant loss of information is not appropriate for this thesis. Comparing objective values only considers the solutions in terms of themselves and not in the setting of the system they model (which is the only important comparison). Comparing the solutions produced by various approximations can only be authoritatively achieved by simulation of the system (which would need to be carried out in a deterministic manner, to be consistent with the expectations of the models). Performing the exhaustive simulations necessary are also beyond the scope of this thesis.

Comparing only problem size and solution times gives an idea of the potential savings in computer overheads. The piecewise linear model equivalent to the 52 week 17 node example above is specified by 28 000 variables, 37 000 constraints, 186 000 non-zero matrix entries and 13 000 non-zero objective coefficients. The solution time for this example is of the order of 12 hours from a crash start and 6 hours from a previous scenario's solution. For a piecewise quadratic continuous model, equivalent to the same example, the specifications are: 32 000 variables, 38 000 constraints, 199 000 non-zero matrix elements, and 15 000 objective coefficients. The solution time, here, was in the order of 18 hours from a crash start and 6 hours from a previous scenario's solution.

For the final implementation, rather than using a generic solution platform, one should be tailored to the model to allow fast, robust, solution. The platform should exploit the structure of the model, especially the self-similarity of the weekly systems. In particular, exploiting the fact that the problem may be formulated as a Generalized Network with side constraints would decrease solution time.

To further reduce solution time, the final implementation should include a start-up procedure which finds a "reasonable" solution. This would considerably save on solution time, especially if the initial solution is feasible. Such a procedure was *not* constructed for the initial implementation because, with a continually changing model, such a procedure would also require continual change. This would hinder model development by making it difficult to implement changes.

A fast, feasible initial solution *could* be constructed by decomposing the problem. A reasonable schedule of storage and release could be constructed for each waterflow network by estimating the *value* of water released for each week based purely on the average load for that week, and then determining an *optimal* water schedule which maximizes the total value of water released. Such a schedule could be found extremely quickly using a Pure Network solver. Given the releases from this schedule, a feasible generation schedule could be found. All hydro stations would first meet as much of the load at their nodes as possible, and then the stations with remaining capacity and release closest to the node with the highest remaining load would be scheduled to meet as much of this remaining load as possible. Thermal stations would meet the load remaining after this in a similar manner, scheduling the cheapest stations first. Load which cannot be met by this method would be non-supplied. If fast solutions were available for single weeks, it may be useful to use the optimal solution for each week (with fixed releases), starting from an initial solution for that week determined as described.

9.3 Stochastic Implementation

Initial implementation of the stochastic model was difficult because it required the solution of many similar subproblems. MINOS allows iterative solution of inter-related problems; however, the size of the necessary changes to the constraint matrix for each subproblem rendered this method unsuitable. Instead, MINOS was used as a Fortran subroutine to a master C program.

We were unable to use the MPS reader incorporated into MINOS since the constraint matrix needed to be transformed to include extra linear terms in the objective row for the Lagrangian term. This meant that the master program needed to include a matrix reader. It was decided that an MPS format file reader was more appropriate for the initial implementation than direct construction of the matrix from the `WeekRef` files—it was easier to debug and it was easier to incorporate multiple scenarios in an MPS file, since this merely required (in the case of our model) the specification of multiple right-hand sides and bounds.

To speed up the time spent on the input phase, allowance was made to save and load the model's important information directly. This allowed the MPS file to be read on the first Progressive Hedging run, with subsequent runs being started with less overhead.

Direct implementation of the Progressive Hedging Algorithm (using the natural decomposition of scenarios) is straight-forward, and so was used for the working implementation. The non-linear part of the objective function is a simple quadratic and easily programmed. The linear part needs to be specified differently for each subproblem and for each iteration; this was easily done by updating a current objective and inserting the appropriate row into the constraint matrix (where MINOS stores the objective function coefficients). This meant that the constraint matrix of each subproblem was of the same dimensions as the deterministic case and so we could directly use solutions to the deterministic cases as initial solutions for the Progressive Hedging Algorithm.

Difficulties were experienced running MINOS in a linear manner (to obtain the initial solutions) and *then* in a non-linear manner, as required by the Progressive

Hedging Algorithm. For this reason, initial solutions to the subproblems were obtained purely through MINOS, and stored as start-up files. This proved to be a very satisfactory method as it separated the time taken to determine an initial solution from that for stochastic solution, and allowed many runs using the same scenarios under different parameter values to be performed efficiently.

In a final implementation, the solution obtained last week should provide a good initial solution for this week, (hopefully) even under a change in the scenarios used. This would be most effective when used in tandem with an heuristic procedure for dealing with the increment in weeks, i.e. a procedure which takes the solution obtained last week, removes the first week and adds a feasible solution for the last week (probably similar to the solution for the, now, second-to-last week). Changes to inflow and load forecasts would be harder to deal with, however the use of the optimal policy and price variables from the solution found last week should be reasonable.

9.4 Progressive Hedging Convergence

Convergence for the Progressive Hedging Algorithm (on a convex example such as this) is guaranteed to be linear for the situation where every subproblem is solved to *optimality*. This result holds even for the situation where the subproblems of successive Progressive Hedging iterations are progressively solved to tighter tolerances under an explicit scheme where the tolerance (on the magnitude of the gradient) is given by

$$\mu(1-\epsilon)^{n+1}\min\left\{1, \|x-\hat{X}_i\|\right\} \tag{9.1}$$

where x is the solution point being considered, $\hat{X}$ is the current policy and μ and $\epsilon < 1$ are positive constants.

Unfortunately, the large-scale nature of the model is prohibitive in the *numerical* procurement of a truly optimal solution, i.e. we must be content with a solution which is within some fixed tolerance of the optimum. In actuality, the tolerance of Equation 9.1 *will* apply for a number of initial iterations (for some choice of μ and ϵ), but beyond some unspecified point convergence will be no longer guaranteed, because of machine precision and the possibility that $\hat{X}_i$ is intolerably close to the optimal solution of scenario i.

Convergence of the Progressive Hedging Algorithm can be measured explicitly via the the convergence of

$$\delta^n = \left(\|\hat{X}^n - \hat{X}^{n-1}\|^2 + \sum_{i \in \{1,\ldots,S\}} p_i \|x_i^{*,n} - \hat{X}_i^n\|^2 \right)^2 \tag{9.2}$$

to zero, where n represents the Progressive Hedging iteration. In practice, convergence is initially monotonic and fast. Subsequently, however, the convergence rate slows, but, while it is no longer monotonic, progress continues to be made. Figure 9.1 shows an example of this. To ensure convergence of the dual variables it was found to be necessary to use an inner product (see Section 6.5) which was weighted according to the relative importance of each variable (in terms of the amount of electricity it represented). The difference in convergence is illustrated in Figure 9.2.

However, solutions exhibited a lack of convergence after some indeterminant point, and so it was necessary to devise a more useful stopping procedure. Stopping when δ was within some tolerance of zero was *not* useful, since the tolerance may never be achieved or, if too loose, may not provide a solution adequately close to optimum. In practice δ seldom dropped below 10^{-3}. The choice of an adequate stopping criterion was made more difficult by the presence of spikes and non-monotonic behaviour in the value of δ as the algorithm progressed. These combined to make the best stopping point an almost purely subjective decision; the real test of adequacy of an optimal solution is in terms of its application to the system, which is not addressed here.

We focussed on the the value of δ, one of the few measures of convergence that was available to us (as observed by Helgason and Wallace [7]). When the value of δ is no longer decreasing over time, the solution was halted so that a manual investigation of convergence could be conducted (by viewing the rate of change of various parameters and values) and a decision could be made to restart the solution, or not.

Creating an *automatic* procedure to decide when δ is no longer decreasing was not simple. Trials which considered the gradient of the best linear least-squares fit of the past ten δ values, were very susceptible to spikes (sudden jumps in the value of δ which did not often indicate a lack of convergence). To alleviate the effect of spikes we considered the last twenty such *gradients*, and a consensus of them indicating an increasing δ was taken to be the stopping criterion. However, this also continued to be affected by spikes (as the height of the spikes was often

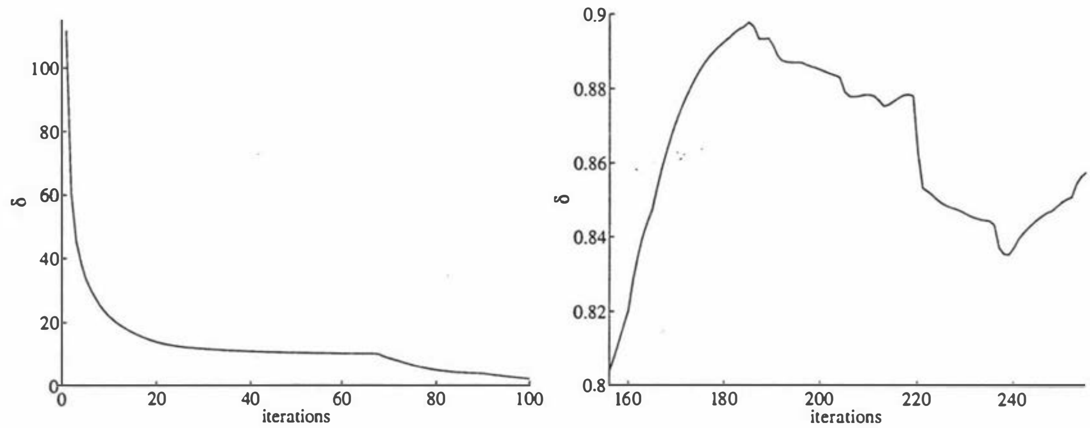


Figure 9.1: Convergence using sub-optimal subproblem solutions

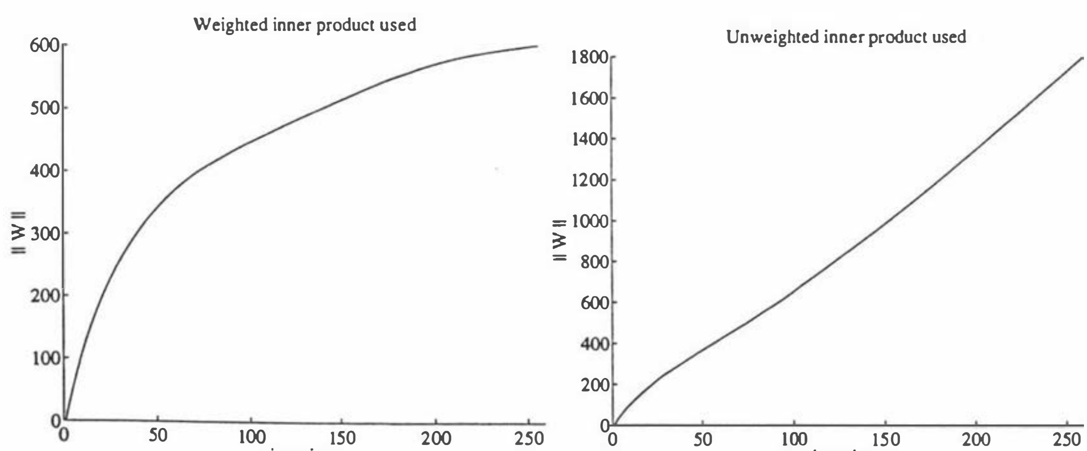


Figure 9.2: Convergence of W with and without a weighted inner product

large relative to the size of δ). An effective stopping criterion was developed by considering the most negative gradient obtained by successively ignoring one of the previous ten δ values. This was used in conjunction with the consideration of the consensus of the previous twenty gradients so chosen.

To incorporate some extra robustness into the stopping criteria, consideration was given to the convergence (to zero) of another parameter, γ, proposed by Mulvey and Vladimirou [12], where

$$\gamma^n = \left| \sum_{i \in \{1,\dots,S\}} p_i \left\langle W_i^n, x_i^{*,n} \right\rangle \right| + \sum_{i \in \{1,\dots,S\}} p_i \| x_i^{*,n} - \hat{X}_{i-1}^n \|^2$$

While γ could not be expected to *monotonically* decrease to zero, in practice it seemed to pass through continuos periods of decreasing or increasing. The γ values had the same stopping criterion applied as used for the δ values, and solution was stopped when *both* showed a lack of further decrease. Often this extra consideration of γ was enough to carry δ to a point where it had begun to decrease again.

It must be stressed here that these stopping criteria are *not* in terms of acceptable convergence, but indicate a point where convergence had ceased. This means the solution times (as given in this Chapter) cannot be thought of (*at all*) as the time taken to reach an optimal solution; they are the time taken to *first* be stopped by the stopping criteria described here. A large amount of convergence has taken place at this point, but the solutions may remain a long way from an *optimal* solution. Since the same stopping criterion is applied to all problems, it is hoped this will give an indication of the *relative* solution times to a solution which is acceptably close to optimallity.

Due to the fact that we do not investigate to optimality of solutions in terms of the system they model, we needed to consider many measures of convergence. This was also useful in ensuring that some progress was being made, and for debugging. Among the values considered were: the norm (in terms of the inner product used) of the dual variables, the norm of the difference between the current optimal solutions to the subproblems and the current policy, the norm of the change in policies from one iteration to the next, the Lagrangian part of the objective function, the objective function value, the part of the objective function deriving solely from the non-anticipative variables, the norm of the non-anticipative variables, and the norm of the policy-dual variable pair [19]. Figure 9.3 shows the progress of some of these measures for an eight scenario example. It is difficult to comment on the efficiency

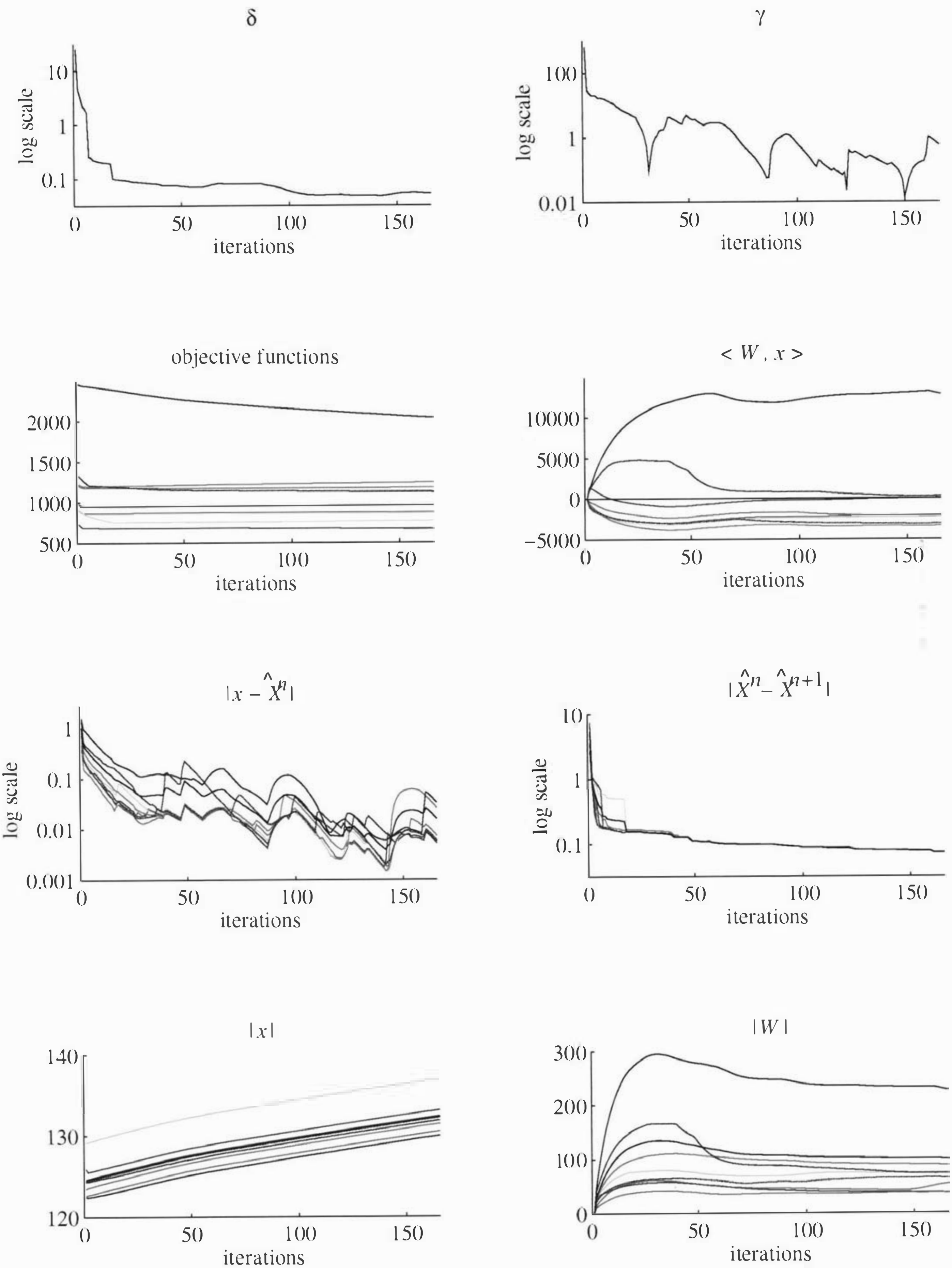


Figure 9.3: Various measures of convergence for an eight scenario problem

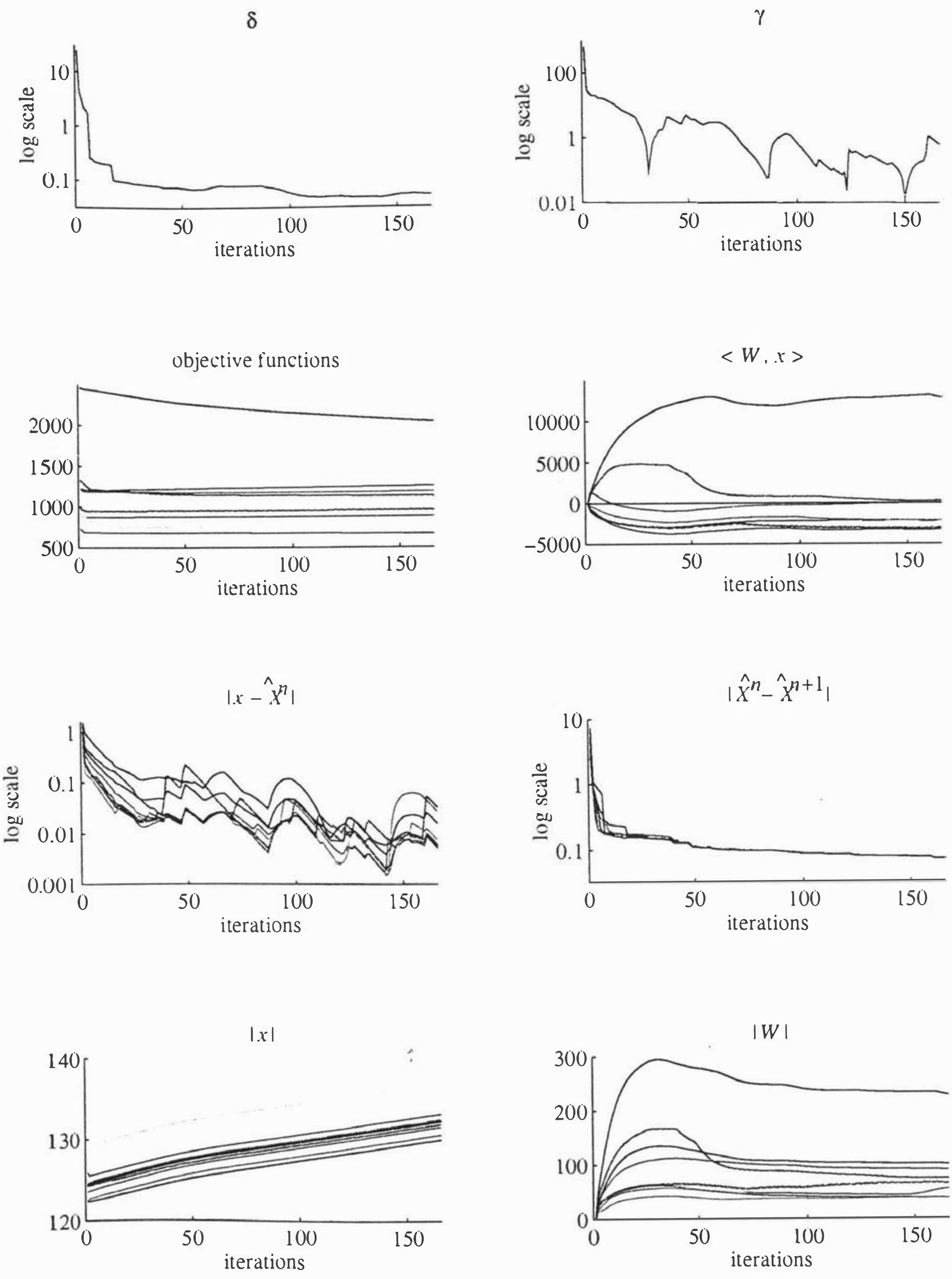


Figure 9.3: Various measures of convergence for an eight scenario problem

of such measures when no rigorous concept of convergence has been obtained. However, Figure 9.4 shows some of these parameters for a run which was continued until there was no longer any apparent change in the value of the non-anticipative variables. As can be seen from Figures 9.3 and 9.4, many of the measures shown mimic the behaviour of others (e.g. the magnitude of the subproblem solutions and the magnitude of the policies) while some appear to give little information on convergence (e.g. the objective function values and the inner product).

9.5 Fine Tuning—A Discussion

There are few, as yet, case studies in the literature on the application of the Progressive Hedging Algorithm. In this Section we discuss the recommendations made in the literature and our experience with the algorithm. Many of the recommendations given in the literature pertain more to a final implementation than to experimental implementation.

For most of the stochastic testing a four week version of the model was used. The main reason for this is that the solution times for the full 52 week version of the model were too long to allow adequate experimentation. A 52 week implementation with only *two* scenarios which allowed deterministic solutions *after the first week* took 70 hours of computer time in solution(!), with each scenario subproblem taking an average of 15 minutes to solve. This would suggest that the fully detailed model is too restrictive to be used as a basis for the stochastic extension (confirming the comments made in Chapter 3). If solution was performed in a parallel manner, the solution time would be in the order of 35 hours, which is still unacceptable.

Choice of the penalty parameter r has a major effect on the convergence rate. Mulvey and Vladimirou [12] suggested a process of dynamically changing r during solution; the effects of this on solution time are not investigated here, as the particular values and the regime for changing of r is heavily dependent on the complete and final formulation. It did not appear appropriate to invest the vast amount of time necessary to carry out such an investigation on a trial model, with no way to investigate the quality of solutions produced and so no true method of determining adequate convergence.

To obtain convergence in reasonable time, different values of r were investigated. The results of [12], [7] and Philpott and Leyland [17], suggested that low values

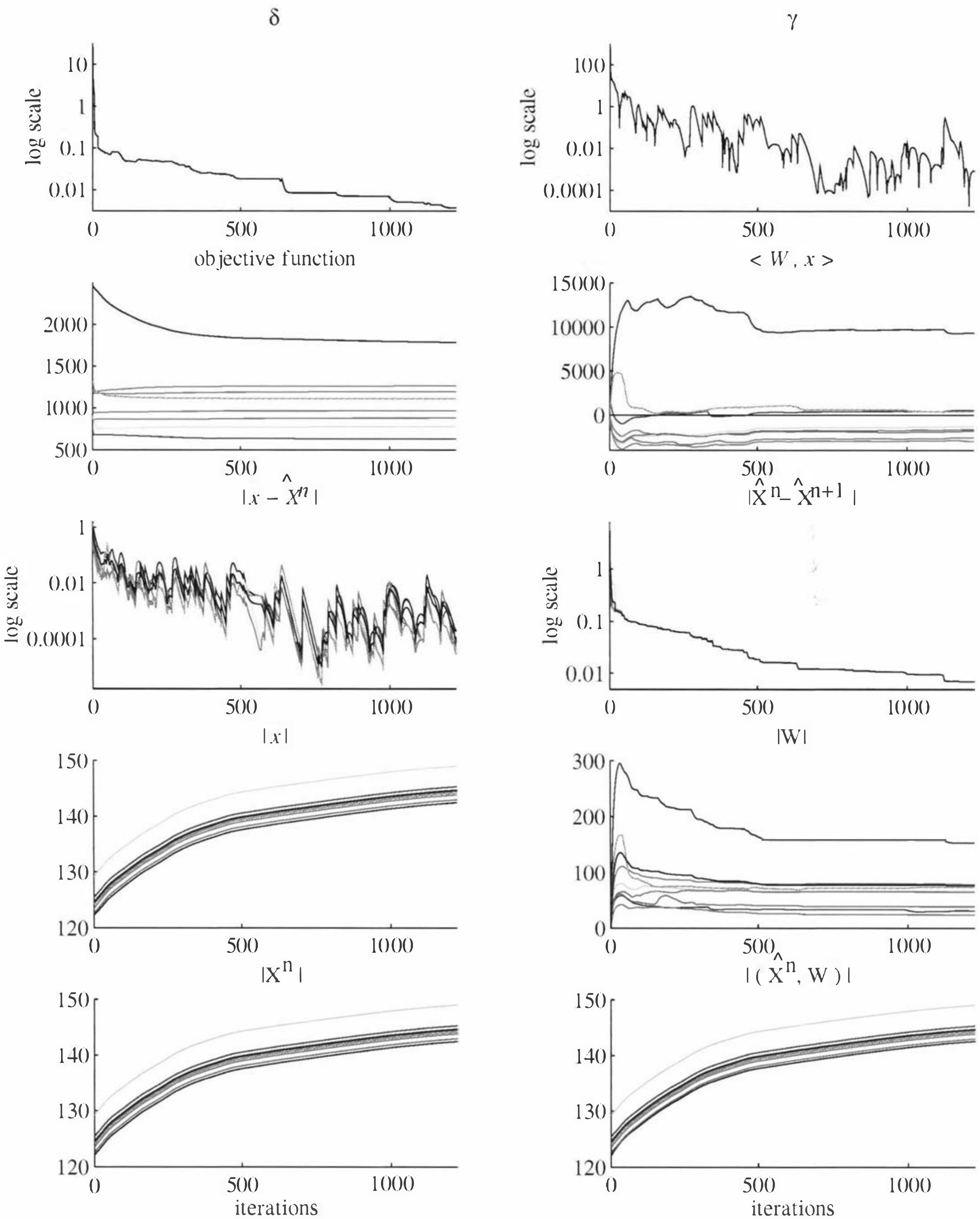


Figure 9.4: Convergence measures from an exhaustive run using eight scenarios

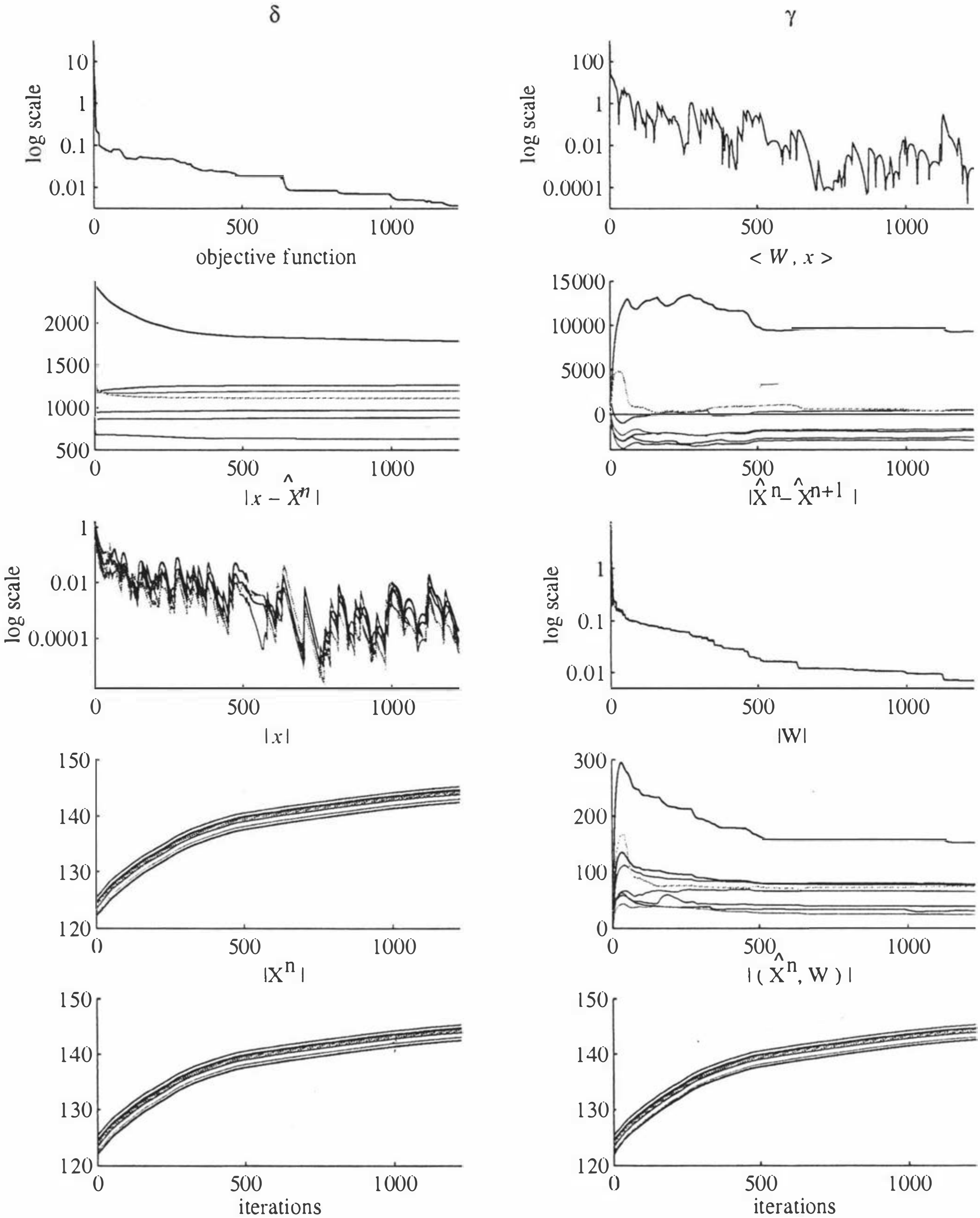


Figure 9.4: Convergence measures from an exhaustive run using eight scenarios

of r produced slow convergence of the primal solution (very low values produced no convergence at all), whereas high values of r quickly produced solutions which adhered to the non-anticipavity conditions, but, from there, convergence to an optimal solution was very slow. Our results, however, were inconclusive, as we have no measure of optimality beyond the convergence measures used. The *value* of many of these convergence measures is dependent on the r value used—lower values of r did tend to produce more erratic behaviour in the value of δ. Figure 9.5 shows an example of the convergence of some parameters for differing values of r. Table 9.1 compares the solution times for various r values.

Table 9.1: Solution times (in minutes) for various r values, the number of iterations is given in brackets.

r	Solution Time			
	Unweighted inner product		Weighted inner product	
10	80	(200)	300	(1000)
30	105	(200)	400	(1000)
100	200	(200)	680	(1000)
300	360	(200)	940	(1000)

The nature of the model allows some additions which should have a beneficial effect on the solution time. Since the objective function of the subproblems is quadratic, the unconstrained minimum can be calculated explicitly. If this point is feasible, this is exactly the optimal solution. In this case it is quite simple to find the unconstrained minimum explicitly; for the i'th subproblem it is exactly given by

$$\hat{X}_i - r^{-1}(f + W_i)$$

where the objective function of the deterministic problem is $f^T x$, and no weighting function is used in the inner product. Solution time may be saved by checking the feasibility of this point.

Another solution point of interest is the *policy*. In the situation where this point is feasible it would, most likely, be a better initial solution than the optimal solution found during the previous Progressive Hedging iteration (which has the advantage of *always* being feasible). Needless to say, it is better to start from a feasible solution than an infeasible one, as there is no *nice* way of discovering a feasible solution which is close to an infeasible one (in terms of both Euclidean distance and objective value).

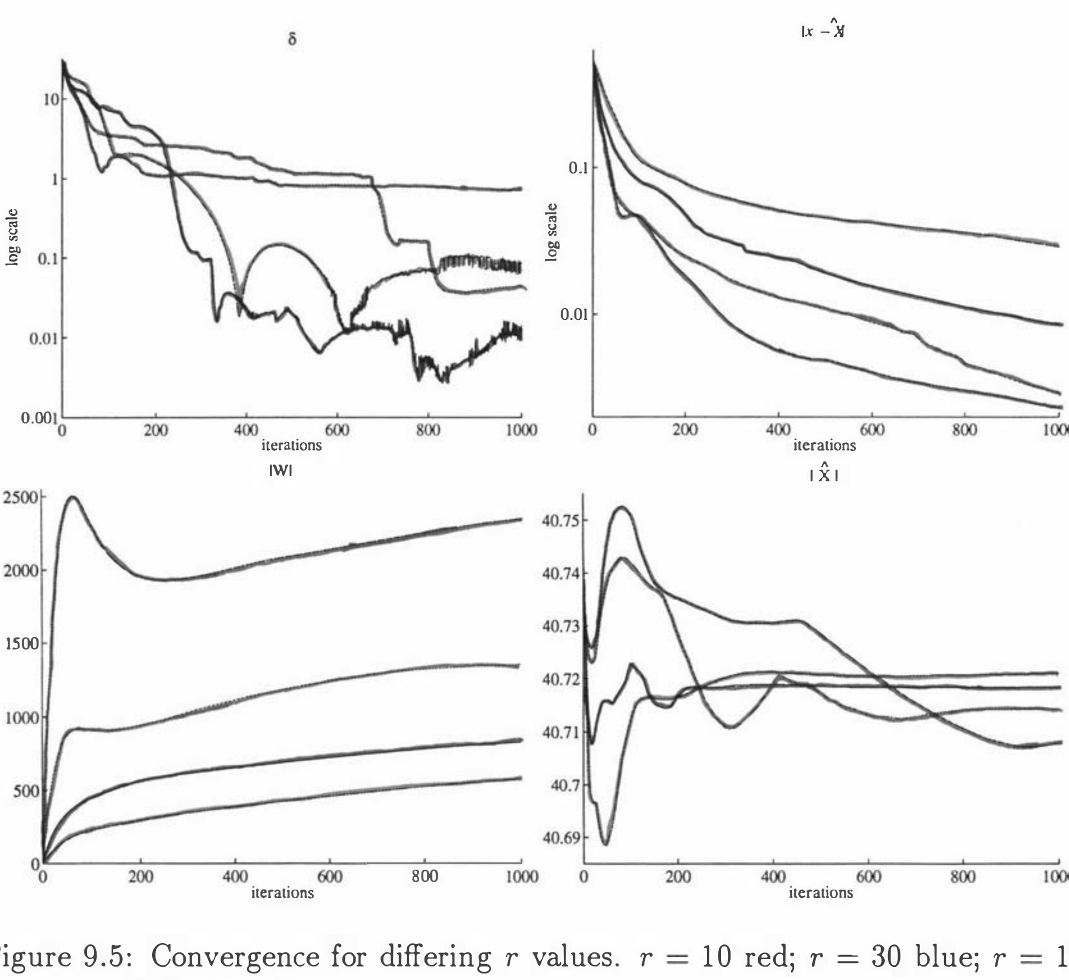


Figure 9.5: Convergence for differing r values. $r = 10$ red; $r = 30$ blue; $r = 100$ green; $r = 300$ magenta.

The convergence under different non-anticipative variables was also investigated. Generation, releases and storage were variously used as the non-anticapative variables. The convergence rates seemed to reflect the "freedom" allowed by each of the non-anticipavity variables; generation was the most constraining, release the next and, finally, storage seemed to allow the most freedom, especially when used in conjunction with a penalty for not meeting the target level. Since having different non-anticipated variables changes the formulation, the best value of r for each such variable will be different. This, combined with the stopping criterion used, means there would be little use in comparing the solution times using various non-anticipation variables.

Changing the number of scenarios had a major effect on the time until solution. To illustrate this, Table 9.2 shows the solution times for the same formulation, and the same value of r, while using various numbers of scenarios. It also shows an

Table 9.2: Solution times under differing numbers of scenarios

Scenarios	Serial Time	Estimated Parallel Time	Iterations	Final δ
2	310	160	260	0.00324
4	830	207	255	0.858
8	2450	306	202	1.81
16	7115	445	337	0.372

estimate of the solution time had solution been carried out on parallel processors which is determined by dividing the total solution time by the number of scenarios used. This estimate is reasonable given the fact that the solution time spent *between* subproblems was negligible when compared to the total solution time.

9.6 Discussion

This Chapter opens up many possibilities for future research. The model has been brought to a point where the possibilities for implementation appear almost boundless. Such an investigation would require the implementation of a simulation environment and definition of a test model, whose scope is beyond the focus of this thesis.

We feel that this juncture provides a very good starting point from which an investigation on the usefulness, in practical terms, and implementation issues. Any investigation from this point must consider many possibilities, comparing and contrasting their effects. Any lesser investigation would surely provide a very biased view thereby giving little credit to the wealth of possibilities which were *not* investigated. Thereby making this stage a very appropriate, and natural, point at which to finish the development of the model.

Having looked ahead, in an attempt to anticipate some of the implementation issues of the model, we have provided an initial basis for any encompassing research in the stochastic and implementation issues of the model. The development of the model has constructed a justifiable and useful framework within which to investigate the effects of various approximations, and trade-offs which may arise from these.

Chapter 10

Conclusions

Broadly speaking, the initial intention of this thesis was to develop a model of the New Zealand hydro-thermal electricity generation scheduling system with a time horizon of one to two years, for subsequent use by ECNZ. The model was required to provide detail of the physical system, including the explicit incorporation of six separate hydro reservoirs. The model was also required to account for the stochastic aspects of the inflows into these reservoirs. The current model used by ECNZ (for this specific scheduling instance) is based on an SDP approach, which includes two hydro reservoirs with little detail of the physical system; it does account well for the stochasticity in inflows.

The most significant contribution of this thesis is the development of a model which provides a flexible level of detail for *both* the physical system *and* the stochastic elements. Flexibility in the modelling of the physical system provides a *framework* within which the effects of various approximations used may be investigated. Determination of the limitations to this flexibility provided boundaries on the approximations which could be used, as well as insights into the efficiacy of various modelling techniques.

The *specifics* of the intentions for the thesis metamorphosed as model development progressed and we became more familiar with the limitations of the framework used and the fundamental characteristics of the system. The emphasis moved from the development of a *specific model* to the framework of a *general model* which provided flexibility in many aspects of the system. This was in order that that the level of approximation of the *important* aspects of the system did not preempt further development, especially in the stochastic extension used. Broadening this emphasis reduced specificity in the model which, therefore, meant that deep exploration of a stochastic extension became less meaningful to consider, in the context of practicability and comparison of solutions.

The framework developed here will provide a good *basis* for a final implementation which is usable by ECNZ. It also provides the platform for a thorough investigation into the best design of a *specific* model, i.e. determination of the *actual* level of approximation for each aspect of the system, so as to provide an adequate representation of the system, reasonable solution time, and useful solutions.

The framework developed allows the balance between computational tractibility, detail of the physical system, and representation of the stochastic elements. This is to allow the model to be "tuned" so as to provide well balanced solutions. The flexibility also allows the use of sensitivity analysis for investigation into the effects of approximations used, in terms of the wider framework provided.

10.1 The Deterministic Model—A Summary

The model's development naturally separates into two parts: the deterministic framework (in which the hydro inflows are treated as fixed), and the stochastic extension (which allows future inflows to be uncertain). We now summarise the achievements derived from the development of the deterministic framework.

With the emphasis of the model on its flexibility, the deterministic framework was required to provide physical system detail and to allow a stochastic extension. For this reason, the deterministic framework defines the structure underpinning both the thesis and any model so derived.

A survey of literature highlighted the differences between the New Zealand system and other systems, as well as the need for a unique approach to the modelling of the New Zealand system. The features which are important to be modelled well, and those for which it is reasonable (or even desirable) to be more coarsely modelled, were discussed. Many of these features were *defined* by the fact that the model was required to incorporate six explicit hydro reservoirs as well as a detailed physical system.

The need for consistency within the model demanded the inclusion of information on the geographical distribution of generation and power use. This feature was incorporated through the use of a geographic network connecting locations of interest. Load was specified at the nodes of this network structure, as were the significant power stations. To ensure ease of approximation, and to make the forecasting more robust, Load Duration Curves were used to represent this load. To facilitate this, the time horizon was split into time steps (of the order of a week) so as to allow the use of time-dependent hydro system information and decision making.

Hydro and thermal stations may be present, in any number, at the nodes of the network structure. Thermal stations have a fuel cost associated with generation, and hydro stations have use of a limited supply of water which may be stored over multiple time steps; it is the scheduling of the stored water which links the generation schedule over the time horizon. The arcs of the network structure represent transmission over part of the National Grid. To model this, the arcs take on representative characteristics in the form of capacity and line loss information.

Load Duration Curves, transmission and generation curves are all approximated by piecewise quadratics. A non-supply curve (together with a cost of non-supply) is introduced to ensure feasibility in the meeting of demand.

An initial approach considered incorporating information on the uncertainty of the generating capacity of stations (due to forced outages) by using a cumulant approximation of the Electricity Curve inverse. Unfortunately, the approximation used artificially induced non-convexity in the objective function and admitted multiple locally optimal solutions. The objective function also exhibited non-differentiable "corners", some of which were an artifact of the way in which the problem was modelled.

Other approximations of the Electricity Curve inverse were proposed, but, these were not investigated further, for several reasons. One of the major reasons was that the way in which the effects of the uncertainty in station generating capacity were being modelled, was *not* a good approximation. Moreover, this approximation induced many of the poor features apparent in the objective function. The other major reason was that the re-modelling of the system, so as to eliminate the "corner" arising from the hydro stations' contribution to the objective function, also removed the *need* to use the inverse of the Electricity Curves. This re-modelling included other beneficial side-effects, such as making the problem linear. The explicit use of Electricity Curve inverses does, however, allow for a better account of forced outages in the case where the physical system has many thermal stations at the same node. This becomes important when the deterministic framework provided is used to define a simplistic physical model so as to afford more detail for the stochastic elements; this would be useful in conjunction with the exploration of stochastic extensions proposed in Chapter 6, and to allow more flexibility in the deterministic framework. Since the intention of this thesis is to provide a detailed physical system (in the New Zealand context) the existence and construction of approximations to Electricity Curve inverses which give a convex objective function were not fully investigated. Such an investigation provides a worthwhile direction for future research.

The re-modelling alluded to above was designed to facilitate modelling the proposed generation of each station explicitly rather than as part of a collective contract curve for each node. This re-modelling is reasonable, in the New Zealand context, in light of the fact that there are at most two thermal stations at any node in the *representative* geographic network, and that the hydro stations were already individually scheduled. Modelling each station's proposed generation individually meant that the capacity of each station could be *directly* applied to the proposed generation, rather than needing to be implicitly enforced via a non-supply cost for over-capacity generation. It was the calculation of this non-supplied over-capacity generation which induced the numerically difficult features in the objective function.

To fully remove the need for approximate Electricity Curve inverses, the transmission capacity constraints (which were being modelled as penalties for over-capacity transmission) also required re-modelling. Since the transmission curves are piecewise quadratics, the constraints to explicitly enforce transmission capacities are non-linear. To circumvent this, a linear approximation of these constraints was used which ensured that *only* feasible (below capacity) transmission was allowed.

The model was expanded to include some features of the New Zealand system which were not seen as central to the model, but which would need to be addressed in a final implementation, for ECNZ. More specifically, the features included were: the use of two fuels by some thermal stations, limitations on a fuel supplying three of the thermal stations, and the modelling of consideration of the security of supply. Constraints were designed which were well suited to the framework developed.

10.2 Investigating the Deterministic Framework

The deterministic model developed could be formulated as a Generalized Network with side constraints. This structure could be usefully exploited to ensure fast solution times for a final implementation. Solution time could also be reduced by the use of a less detailed model for later time steps, which would obviously result in a concomitant loss of detail in solutions. The use of differing length time steps, coarser Electricity Curve approximations and differing amounts of detail in the physical system, were all discussed. None of these modelling techniques were actually implemented, as fully evaluating the effects of such approximations would necessitate rigorous testing and comparison of solutions, in order to determine how the loss detail from solutions might affect system operation. Such testing is beyond the scope of this thesis, as it requires a more *specific* model than the framework developed here. We do identify the need for this testing to explore the effects that such approximations have upon solutions. This will provide valuable information on the amount of detail required from later weeks so as to continue to provide good first week solutions. This is extremely important in the context of developing a fast, efficient, final implementation for ECNZ.

A more theoretical exploration was initiated on the effects of the Electricity Curve approximation. The approach taken was to investigate the behaviour of structured approximated formulations derived from sequentially finer approximations of the Electricity Curves. These approximated formulations tend to an "unapproximated" formulation (in which the Electricity Curves are allowed to be *any* functions which are implementable in terms of the model) in the sense that the right-hand-side functions of the approximated formulations converge to the appropriate right-hand-side functions of the unapproximated formulation. The point of interest is whether the sequence of optimal solutions of the approximated formulations converge to the optimal solution of the unapproximated formulation. Examples showed that either some (or all) of the approximated formulations may have no feasible region, but such occurences appeared to be dependent on the approximations used. This indicated that the approximations used needed to take the *types* of solutions produced by the model into account, as well as the right-hand-sides.

A full investigation would be lengthy and may not produce results of direct importance to the model being developed. This is because the convergence (or non-convergence) of a sequence gives no indication of how well a single element of that sequence approximates the sequence limit. For these reasons the investigation was *not* taken to its conclusion. In practice, investigation into the effects of the Electricity Curve approximation may be better served by an empirical study into the quality of solutions obtained.

This is not to say that a *full* theoretical investigation of this nature is not important and worthwhile; it just does not fit into the framework of this thesis. We offer the full investigation into the limit of optimal solutions (when they all exist) of the approximate formulations as a worthwhile direction for future research, as well as an investigation into the difference between a single approximated formulation and its corresponding unapproximated formulation.

10.3 Deterministic Implementation

Implementation of the deterministic model was discussed. An formative implementation was used to isolate many of the difficult approximations, and some of the issues involved in a final implementation were also explored. The formative implementation used a working model (as described in Chapters 3–5). This working model was designed to test the limits of the framework where such limits were seen as important (e.g. the use of piecewise quadratics), to be internally consistent (e.g. all weeks were specified to the same level of detail), and to provide a level of detail which was *at least* a level desirable to ECNZ (e.g. the representative geographic network). Such a model would be detailed enough to highlight inconsistencies in the framework.

The formative implementation demanded procedures to allow specification of the model being solved (from within the framework provided), and specification of the input data that were required. The continual change of the model and the deterministic framework meant that the solution input needed to be flexible. This made the input structures used unsuitable for use in a final implementation.

Due the the enormity of evaluating solutions directly from the output of MINOS 5.4 (the solution procedure used), procedures were written to allow solutions to be viewed using the GUI (graphical user interface) features of Matlab 4. This proved to be an effective method for interrogating solutions.

The need for a procedure to determine a feasible (and reasonable) initial solution, so as to speed up solution times, was highlighted. The inclusion of such a procedure in a formative implementation merely increases the difficulties associated with making amendments to the model, and so was seen as inappropriate during this development phase. Development of such a procedure, therefore, provides a direction of future research; however, the development of a corresponding procedure for the stochastic case may supercede this.

10.4 Stochastic Extension

Allowing future inflows to be uncertain, and future decisions to depend on previous inflows (once they are known), increases the difficulty of the problem. There are many ways in which the deterministic model developed can be extended to include such uncertainty; several of these were discussed. The necessity of comparing the effectiveness of these methods under similar conditions was also identified. An authoritative comparison would necessarily be extensive, requiring simulation of the system to evaluate various solutions in terms their benefit to the system. Such rigorous testing is well beyond the scope of this thesis, and provides an important direction for future research in development of a full working implementation for use by ECNZ.

For a *stochastic* model it is much more important to investigate the robustness and effectiveness of solutions produced. This is often done though simulation of the system, and comparison with current policies and those produced by other methods. Since the focus of this thesis was on the development of a deterministic framework for use as the *basis* to a full stochastic model (and *not* on the development of a *specific* full stochastic model itself), there are many issues and modelling aspects of a stochastic model which cannot be meaningfully explored here. Instead we provided a brief examination of the feasibility of extending a deterministic model (developed from the framework provided) stochastically, and explored some of the issues which arose in order to "set the scene" for the exhaustive testing and analysis of stochastic extensions, which is seen as an important next phase. This meant that any examination undertaken here could not directly involve investigation of the *quality* of solutions produced, making any comparison of solutions obtained, during testing performed here, effectively meaningless in this context.

A scenario approach, using Rockafellar and Wets' *Progressive Hedging Algorithm*, was used to *illustrate* one stochastic extension, and to allow preemptive investigation of some of the implementation issues in an effort to provide guidelines for the future development of a stochastic extension. This particular stochastic extension was used because it provides flexibility in the amount of stochastic information which can used; this flexibility is bounded only by the solution time of the consequent model (this is, of course, a very *significant* bound). It also does not limit the formulation of the underlying deterministic model of the system in any, explicit, way.

The modelling issues, generated through the use of the Progressive Hedging Algorithm, included ideas on the choice of scenarios, the choice of the non-anticipative variables, and a possible method for reducing the solution time through the use of a different decomposition. These issues were not fully addressed computationally, as the benefits they provide need to be evaluated within the context of the solutions they produce.

The convergence of the Progressive Hedging Algorithm is guaranteed only when the subproblems are solved to successively tighter tolerances each iteration under a strict regime. Unfortunately, the large-scale nature of the deterministic model, induced by the detail required in the physical system, means that convergence beyond some fixed tolerance is impossible. Therefore, in theory, convergence was not guaranteed beyond some indeterminant tolerance. In practice, the algorithm did converge. Due to the distance of the solutions to the deterministic subproblems from their respective optimal solutions, scaling was an important consideration. Better convergence was achieved through the use of a scaled inner product (for the Lagrangian term and the quadratic augmentation), where the variables were scaled relative to their importance to solutions (i.e. in terms of the amount of generation they represent). This produced a most satisfactory result, and its inclusion should make implementation of a final model more robust.

A brief examination of the effect of the choice of the non-anticipative variable on various convergence measures was made. This showed that the use of storage in this context appeared to allow the most freedom, with the use of generation allowing the least freedom, and the use of release giving slightly more freedom than that given by the use of generation. A variety of convergence measures were investigated, including the measure proposed by Rockafellar and Wets' [19] as the definitive convergence measure, another proposed by Mulvey and Vladimirou [12], and the convergence of various primal and dual variables. Many of these produced similar patterns of convergence.

Due to extremely slow convergence experienced beyond an unpredictable point, a stopping criterion was needed which indicated where convergence appeared to have stopped, since there was no longer a guarantee of convergence to within a pre-specified tolerance. This proved satisfactory and allowed the user to determine whether further convergence was possible, or whether the convergence measure used was merely exhibiting temporary non-monotonic behaviour. With no way to examine the difference in quality between solutions obtained at various points during use of the Progressive Hedging Algorithm, there is little use in taking such an examination further. This does highlight another direction for future research—that of investigating the correspondence between the values of various convergence measures and the quality of solutions obtained; this would provide valuable information on the convergence requirements, as well as possibilities for limiting solution time, in a final implementation.

An examination of the convergence for various values of the Progressive Hedging penalty parameter was carried out, in terms of the convergence measures mentioned above. The results obtained were inconclusive, since the stopping criterion used gave no indication of the *quality* of solutions (or how close to optimallity these were), and also because the values of many of the convergence measures depended on the value of the penalty parameter used. Such an investigation would be more appropriate in the context of a full investigation of this particular stochastic extension.

10.5 Future Directions

We see the main directions for future research as being able to be encompassed within an investigation which furthers the development of a *specific* model into a form which is directly usable by ECNZ. There are three directions that such an investigation could take. The development of a full model for ECNZ would be best served by pursuing these three directions simultaneously, enabling the results of these investigations to be compared and contrasted.

The three directions are: the development of a *specific* deterministic model, and an examination of the sensitivity in the quality of solutions to changes in the various approximations used; an investigation comparing the various stochastic extensions used in conjunction with appropriately approximated deterministic models arising from the deterministic framework developed here; and, an investigation into the implementation issues arising from the use of the Progressive Hedging Algorithm as a particular stochastic extension.

The development of a specific deterministic model requires close consultation with ECNZ. The investigation should be concerned with the *sensitivity* of the first week's solution to the use of various approximations. The effects on solution time and solution quality of these approximations also needs to be addressed. The *specific* deterministic model used will depend, not only on the results of this investigation, but also on the stochastic extension chosen and the physical detail allowed by this extension to ensure computationally tractable models.

An investigation into the various stochastic extensions would need to follow two paths. The first would be to compare solutions and solution times under the *same* deterministic base model. The second would be a comparison of the quality of solutions when the physical systems were tailored so that the solution times were all within some pre-specified bound. The latter investigation would be more useful in terms of the development of a full model, usable by ECNZ; however, it would be difficult to compare solutions obtained via different methods since the level of detail in both the stochastic elements and the physical systems would be different. This means the solutions will need to be compared in terms of how well they perform on simulations of the system.

The investigation into the implementation of the Progressive Hedging Algorithm, as an extension to a deterministic model constructed from the deterministic framework developed here, may seem preemptive *vis-à-vis* the outcome of the investigation into all of the stochastic extensions. This need not be the case. The investigation into possible stochastic extensions will, due to constraints of time, not be able to "fine-tune" each stochastic extension so as to allow fastest, most efficient, solution time. To ensure that each method is treated with fairness, this means that *none* of the methods should be tuned to any greater extent than others. An extensive investigation into a *particular* stochastic extension will provide information on the speed-up which could be expected from tuning of the stochastic extensions. Also, a scenario approach (and more specifically the Progressive Hedging Algorithm) provides the most flexibility for the model as a whole, making such an investigation worthwhile for its own sake.

The other future research directions, which have been outlined in this Chapter, while being *worthy* of further investigation, are not directly relevant to the further development of a model for use by ECNZ, and so are seen, in terms of the aims of this thesis, as being of secondary importance.

10.6 Discussion

The framework developed in this thesis allows flexibility in all aspects of the modelling of New Zealand's hydro-thermal electricity generation system. This will allow the developers of a full model access to information on the cost (in terms of the loss of information) of approximations made within this framework. The deterministic framework also allows for many different stochastic extensions, so as to allow investigation into the one which *best* serves the needs of the user.

Development of the framework has been taken to a stage which allows future developers a platform upon which to base their investigations. Some of the conclusions about the system and possible modelling extensions provide useful insight for future modellers which will help to direct their investigations towards fruitful areas.

The *framework* is fully developed at this point. An investigation into a *specific* representation of the physical system will be dependent on the stochastic extension to be used, and the quality of solutions produced by the consequent full stochastic model—whereas a stochastic extension requires a specific representation of the physical system on which to base itself.

Future investigations will require a re-prioritization of aims and intentions from this point, while the development of the framework, for both designing a specific model and determining the effects of approximations used within this *specific* model, has reached a point of natural conclusion, making this an appropriate point at which to conclude this thesis.

Appendix A

Sample Input Files

This Appendix gives a selection of the input files used to specify the system. The example chosen has a time horizon of two weeks. For illustrative purposes, we use the same geographic network for both weeks, consisting of four nodes (three nodes in the North Island and one in the South Island) with two hydro stations (one in each Island) and two thermal stations (both in the North Island). Files which specify multiple weeks contain some fields which specify their values for each week separately.

The MPS and information files which are created from the `WeekRef` files are not given—the layout of the information file is superfluous and the MPS structure is standard. Furthermore, these files contain transformed data which would be tedious to explain and which is discussed thoroughly in the description of the model, so that no benefit is obtained by the inclusion of these files.

The problem described herein is specified by 420 variables, 360 constraints, 2 500 non-zero elements in the constraint matrix and 220 non-zero objective coefficients. The problem took under 2 seconds to solve.

A.1 Master File

The file `Master` specifies most constants of the system. It also contains information for later use as it is used throughout the entire solution procedure.

```
Name:               Master

InputFiles:
```

```
 ThermalFile:     ThermalRef
 MainArcFile:     LineRef
 HydroFiles:
  Reservoirs:     HydroRef
  Inflows:        InflowRef
  Stations:       StationRef
 MPSFiles:
  MPSOutput:      example.mps
  SPECSOutput:    example.spc
  Information:    example.Info

TransitFiles:
 ToMPS:           WeekRef.%w   % %w is replaced by appropriate week

Maximums:
 NumberOfWeeks:   2
 MaxNodes:        4
 MaxArcs:         8
 MaxHydros:       2
 MaxThermals:     2

Detail:
 WeekSizes:       1 1 –
 DataDeckName:    EXAMPLE
 DiscountRate:    0.07   % per year
 NonSupplyCost:   300
 01Partition:     0.0 0.1 0.4 0.7 1.0

Units:
 TimeInHours:     168.0    % scaling factors for the
 PowerInMW:       1000.0   % objective function

WeeklyNodeFiles:  NodeRef NodeRef –
Unscaling;   % from scaling for basis coefficients
```

```
Unscale:Row1:     1.0e1 1.0e3 1.0e2
Unscale:Row2:     1.0e1 1.0e2 1.0e2
Unscale:Row3:     1.0e1 1.0e2 1.0e2
Unscale:Row4:     1.0e2 1.0e2 1.0e2
```

A.2 Load Input File

The file NodeRef specifies the make-up of load at each node of a geographic network for a single week. The amount of generation from auxiliary stations is specified by Matlab variables, as is each Island's yearly load curve.

```
Name:           NodeRef
Week:           1
SizeOfWeek:     1   % in weeks

Start:
Name:           Auckland
Island(N/S):    N
FlatLoad:       890000   % in GWh
FlatGenerator:  HuntlyCoal
LoadFraction:   0.52

Name:           Taupo
Island(N/S):    N
FlatLoad:       1320000
FlatGenerator:  Wairaki&Ohaaki
LoadFraction:   0.39

Name:           NewPlymouth
Island(N/S):    N
FlatLoad:       0.0
FlatGenerator:  None
LoadFraction:   0.09

Name:           Christchurch
```

```
Island(N/S):    S
FlatLoad:       4340000
FlatGenerator:  Arnold
LoadFraction:   1.00
```

A.3 Transmission Input Files

The file `LineRef` gives specification of each power line, and the network file for each week.

```
Name:           LineRef
NetworkFiles:   ArcRef ArcRef -

Lines:
Name            RL      XL      Volt
ALB-HEN-3       0.00188 0.01537 220  % proportional losses @100MW
ARA-WRK         0.00058 0.00342 220  % RL = real loss
ATI-OHK         0.00119 0.00564 220  % XL = imaginary loss
AVI-WTK         0.00163 0.00803 220  % Volt = line voltage
BEN-AVI-1       0.00325 0.01509 220
BEN-AVI-2       0.00325 0.01509 220
BPE-HAY-1       0.02198 0.10411 220
   ⋮               ⋮       ⋮     ⋮
WRK-WHI         0.00823 0.06765 220
WTK-LIV         0.00629 0.02983 220
```

The network file, `ArcRef`, specifies the make-up and distribution of arcs in the geographic network. Node names correspond to those given in `NodeRef`.

```
Name:          ArcRef

Network:
NodesFromTo:   Auckland Taupo
Capacity:      1300  % Megawatts
```

```
LineMakeUp:    ((OTA-WKM-1 // OTA-WKM-2) // OTA-WKM-3) // (OTA-HLY
               + TAK-HLY + GLN-HLY + HLY-HAM + HAM-WKM + WPA-MTI +
               (MTI-WKM-1 // MTI-WKM-2) + (WKM-TKU-1 // WKM-TKU-2)
               + (OKI-WRK-1 // OKI-WRK-2) + ARA-WRK + (TRK-ATI-1 //
               TRK-ATI-2) + (TRK-EDG-1 // TRK-EDG-2) + EDG-KAW +
               KAW-OHK + WRK-RPO + ((WKM-ATI + ATI-OHK + OHK-WRK)
               // WKM-WRK)) ! % // in parallel, + in series

NodesFromTo:   Auckland NewPlymouth
Capacity:      360
LineMakeUp:    HLY-SFD // (HLY-TMN + TMN-SFD) !

NodesFromTo:   NewPlymouth Taupo
Capacity:      600
LineMakeUp:    (NPL-SFD-1 // NPL-SFD-2) + ((SFD-BRK-1 // SFD-BRK-2)
               // SFD-BRK-3) !

NodesFromTo:   Taupo Christchurch
Capacity:      1240
LineMakeUp:    22.8 !    % Ohms, DC resistance
```

A.4 Thermal Station Input File

The file ThermalRef gives information on the fuels and fuel constraints, as well as data on each thermal station.

```
Name:             ThermalRef


Fuels:
FuelName:         Coal
FuelCost:         2.4  % $ per Gigajoule
CalorificValue:   22  % GJ/unit
Constraint:       None None –

FuelName:         Maui
```

```
FuelCost:        2
CalorificValue:  1000
Constraint:      Maui Maui –

FuelConstraints:
MaxWeeklyMaui:  2300 2300 –   % PJ (units)

Thermals:
Node:            Auckland Auckland –
Thermal:         Huntly
ForcedOutage:    0.033
Capacity:        960 960 –   % MW
FuelUsed:        Maui Coal –
HeatRate:        10  GJ/MWh

Node:            NewPlymouth NewPlymouth –
Thermal:         NewPlymouth
ForcedOutage:    0.073
Capacity:        580 580 –
FuelUsed:        Maui –
HeatRate:        10.5
```

A.5 Hydro Station Input Files

The file HydroRef gives information on the characteristics of each river chain which is to be used as a single hydro station in the model. Information on the inflows and stations of each river chain are given in other files.

```
ReservoirsName:  HydroRef
DataDirectory:   $HOME/InflowData

HydroReservoirs:
Name:            Taupo
NodeAtEachWeek:  Taupo Taupo –
Island:          N
```

```
Initial:          240000  % MWh
INames:           Taupo —  % controlled inflows
UNames:           Waikato —  % uncontrolled inflows
Stations:         Aratiatia Arapuni Atiamuri Karapiro Maraetai
                  Ohakuri Whakamaru Waipapa —

Name:             Waitaki
NodeAtEachWeek:   Christchurch Christchurch —
Island:           S
Initial:          1240000
INames:           Cobb Coleridge Pukaki —
UNames:           Benmore Ohau Tekapo —
Stations:         Aviemore Benmore Cobb Coleridge OhauA OhauB
                  OhauC TekapoA TekapoB Waitaki —
```

The file StationRef gives information on every hydro station. Hydro stations have specified controlled and uncontrolled inflows (which are respectively combined for the single station representation).

```
StationsName:   StationRef

Name:           Aratiatia
IInflow:        Taupo  % controlled inflow
UInflow:        Waikato  % uncontrolled inflow
MaxGeneration:  84 84 —  % MW
Cumec/MW:       3.6  % generation efficiency
PartOfUThru:    0.07  % fraction of uncontrolled flowing through
FractionFlat:   0.17  % immediately generated uncontrolled flow

Name:           Arapuni
IInflow:        Taupo
UInflow:        Waikato
MaxGeneration:  140 140 —
Cumec/MW:       2.23
PartOfUThru:    1
FractionFlat:   0.12
```

```
⋮                 ⋮
Name:             TekapoB
IInflow:          –  % none
UInflow:          Tekapo
MaxGeneration:    160 160 –
Cumec/MW:         0.78
PartOfUThru:      1
FractionFlat:     1.0

Name:             Waitaki
IInflow:          Pukaki
UInflow:          Benmore
MaxGeneration:    100 90 –
Cumec/MW:         6.2
PartOfUThru:      1.1
FractionFlat:     0.5
```

The file `InflowRef` gives information on the reservoirs and inflows (both controlled and uncontrolled), including the file in which the past data on each inflow is given.

```
InflowsName:      InflowRef

Controlled:
Name:             Taupo
FileName:         taupo.dat  % containing inflow data
FileStartDate:    1/4/31
InflowStart:      16/2/66
MaxLevel:         9900 9900 –  % m^3s^-1 days
MaxRel:           220 220 –  % m^3s^-1
MinRel:           35 35 –  % m^3s^-1
MinFlow:          160 160 –  % m^3s^-1

Name:             Cobb
FileName:         cobb.dat
```

```
FileStartDate:  1/4/31
InflowStart:    16/2/66
MaxLevel:       280 280 —
MaxRel:         inf inf —
MinRel:         0 0 —
MinFlow:        0 0 —

Name:           Coleridge
FileName:       coleridg.dat
FileStartDate:  1/4/31
InflowStart:    16/2/66
MaxLevel:       1600 1600 —
MaxRel:         inf inf —
MinRel:         0 0 —
MinFlow:        0 0 —

Name:           Pukaki
FileName:       pukaki.dat
FileStartDate:  1/4/31
InflowStart:    16/2/66
MaxLevel:       38000 38000 —
MaxRel:         440 440 —
MinRel:         0 0 —
MinFlow:        120 120 —

Uncontrolled:
Name:           Waikato
FileName:       waikato.dat
FileStartDate:  1/4/31
InflowStart:    16/2/66
MaxRel:         inf inf —
MinRel:         0 0 —
MinFlow:        0 0 —
```

```
Name:           Benmore
FileName:       benmore.dat
FileStartDate:  1/4/31
InflowStart:    16/2/66
MaxRel:         inf inf -
MinRel:         0 0 -
MinFlow:        0 0 -

Name:           Ohau
FileName:       ohau.dat
FileStartDate:  1/4/31
InflowStart:    16/2/66
MaxRel:         inf inf -
MinRel:         0 0 -
MinFlow:        0 0 -

Name:           Tekapo
FileName:       tekapo.dat
FileStartDate:  1/4/31
InflowStart:    16/2/66
MaxRel:         inf inf -
MinRel:         0 0 -
MinFlow:        0 0 -
```

A.6 Weekly System File

The WeekRef files specify the whole system for a single week. They were constructed from the previous files. The format used was simple to change, allowing the construction of examples by hand for debugging purposes.

```
Name:                WeekRef.1

NODES:
GeneralNodeData:
NumberOfNodes:       4
```

```
SpecificNodeData:
Node:               1
MatrixG:Row1:       7.908e-01 2.128e+00 0.000e+00
MatrixG:Row2:       0.000e+00 6.432e+00 2.436e+00
MatrixG:Row3:       2.484e+00 1.016e+00 1.226e+00
MatrixG:Row4:       6.362e+00 1.044e+00 3.467e+00
Node:               2
MatrixG:Row1:       5.931e-01 1.596e+00 0.000e+00
MatrixG:Row2:       0.000e+00 4.824e+00 1.827e+00
MatrixG:Row3:       1.863e+00 7.623e-01 9.191e-01
MatrixG:Row4:       1.929e+00 7.829e-01 2.600e+00
Node:  3
MatrixG:Row1:       1.369e-01 3.684e-01 0.000e+00
MatrixG:Row2:       0.000e+00 1.113e+00 4.215e-01
MatrixG:Row3:       4.299e-01 1.759e-01 2.121e-01
MatrixG:Row4:       9.248e-01 1.807e-01 6.000e-01
Node:               4
MatrixG:Row1:       1.197e+00 1.941e+00 2.169e+00
MatrixG:Row2:       0.000e+00 4.053e+00 4.681e+00
MatrixG:Row3:       6.220e-01 4.012e-01 9.119e-01
MatrixG:Row4:       8.920e+00 9.663e-01 3.009e+00

THERMALS:
GeneralThermalData:
ThermalNodes:       1 3 -
ThermalNames:       Huntly NewPlymouth
MauiMax:            2300
MauiWeightings:
Thermal             Node      Fuel      Weight
Huntly              1         1         0.01
NewPlymouth         2         1         0.0105

SpecificThermalData:
ThermalNode:        1
```

```
ForcedOutage:        0.033
GenerationCost:      20 24 -
Capacity:            960
ThermalNode:         3
ForcedOutage:        0.073
GenerationCost:      21 -
Capacity:            580

HYDROS:
GeneralHydroData:
HydroNodes:          2 4 -
HydroNumbers:        1 2
HydroNames:          Taupo Waitaki
SpecificHydroData:
HydroNode2:I:        2.81e+05  % controlled
          :U:        0.00e+00  % uncontrolled
FractionOfUFlat:     0.34
MaximumGeneration:   930
StorageInterval:     0.00e+00 5.82e+05
TopInterval:         8.57e+01 5.39e+02  % release
BottomInterval:      3.92e+02 inf  % flow from river mouth
HydroNode4:I:        1.36e+06
          :U:        6.81e+04
FractionOfUFlat:     0.56
MaximumGeneration:   1600
StorageInterval:     0.00e+00 2.12e+06
TopInterval:         0.00e+00 1.17e+03
BottomInterval:      3.18e+02 inf

ARCS:
GeneralArcData:
NumberOfArcs:        8
SpecificArcData:
Arc0:FromTo:         1 2
```

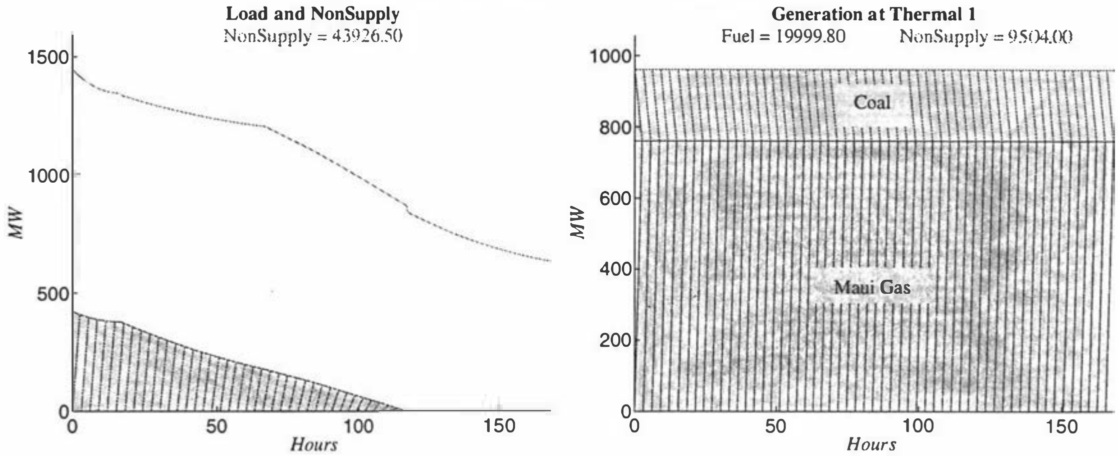
```
PowerLoss:          0.075
Capacity:           1300
Arc1:FromTo:        2 1
PowerLoss:          0.075
Capacity:           1300
Arc2:FromTo:        1 3
PowerLoss:          0.053
Capacity:           360
Arc3:FromTo:        3 1
PowerLoss:          0.053
Capacity:           360
Arc4:FromTo:        3 2
PowerLoss:          0.026
Capacity:           600
Arc5:FromTo:        2 3
PowerLoss:          0.026
Capacity:           600
Arc6:FromTo:        2 4
PowerLoss:          0.23
Capacity:           1240
Arc7:FromTo:        4 2
PowerLoss:          0.23
Capacity:           1240
```

Appendix B

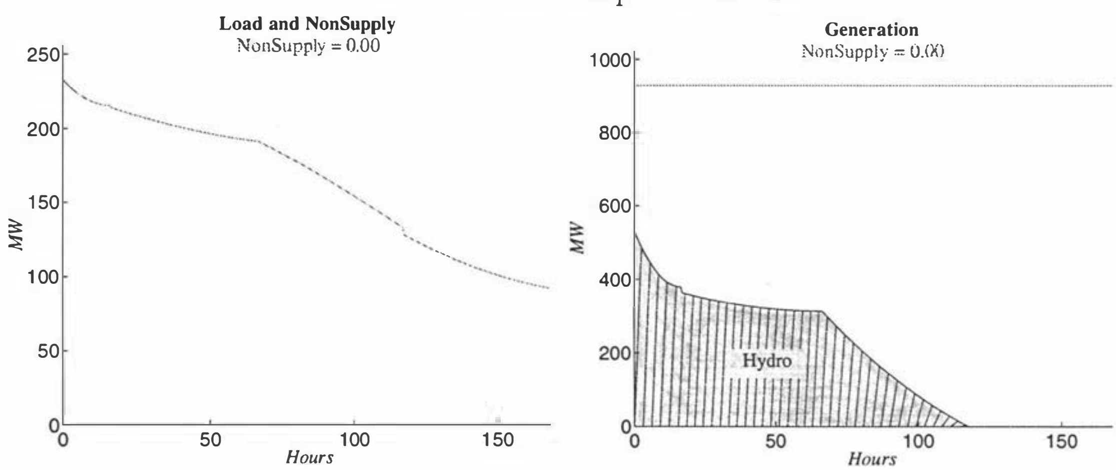
Model Output

Here we illustrate the first week's solution for the small example for which some of the input files are given in Appeddix A. This small example is not very realistic as requires that only *four* power stations are used to meet *all* of New Zealand's load for two weeks. It does, however, illustrate many features of the model. The LDC's show the non-supply curves shaded in red. Both the release and storage levels for both weeks of solution are given. In these plots, the black curve is the minimum level, the red curve gives the maximum level and the blue curve gives actual release or storage. For the transmission arcs, the blue curve is the power entering the line, and the cyan curve is the power exiting the line. The final plot shows a break-down of the schedule for both weeks. The green line shows the total load, the cyan line shows the total generation, and the black line shows the total generation plus non-supply; the difference between the total load and the total generation plus non-supply is due to line losses. The red curve gives total non-supply, the blue curve gives total hydro generation, and the magenta curve gives total thermal generation.

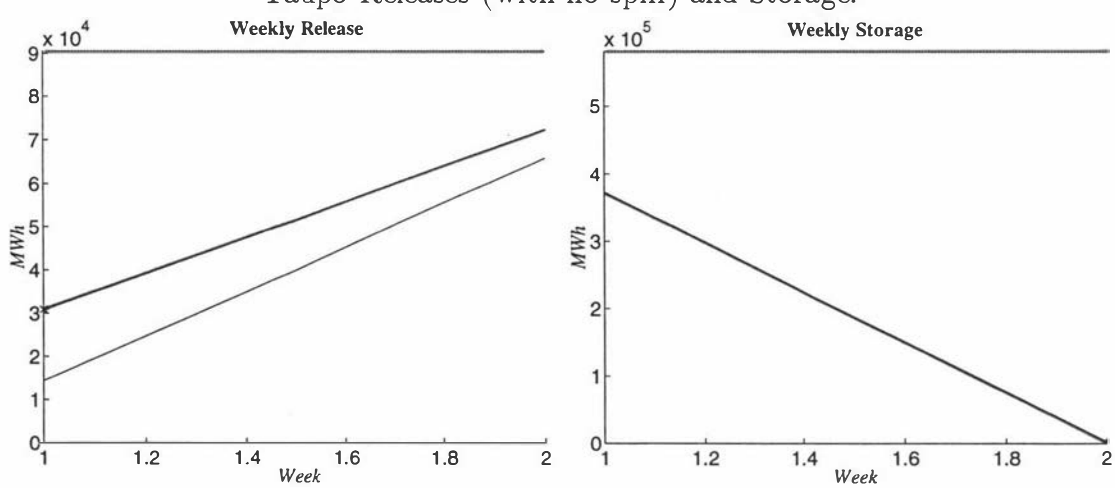
Node 1. Load and Huntly Generation.



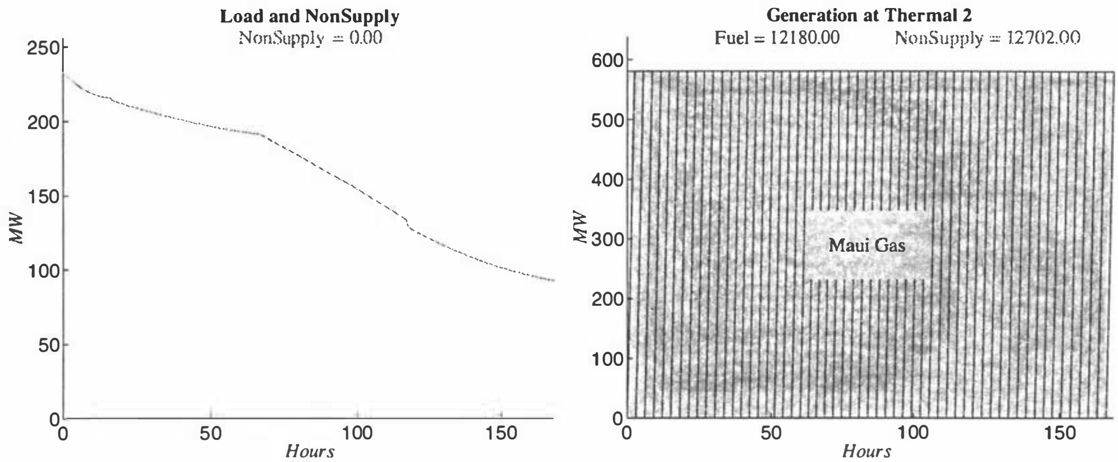
Node 2. Load and Taupo Generation.



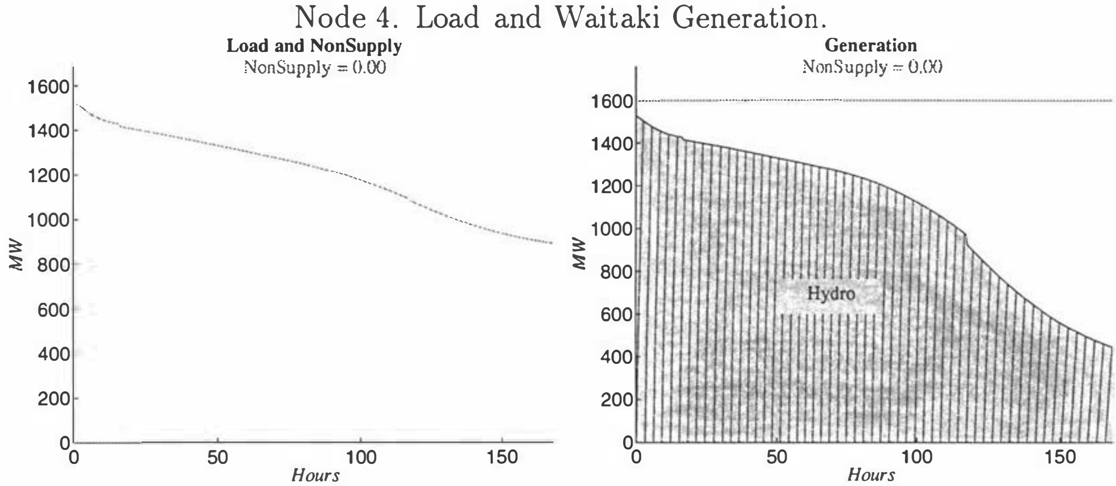
Taupo Releases (with no spill) and Storage.



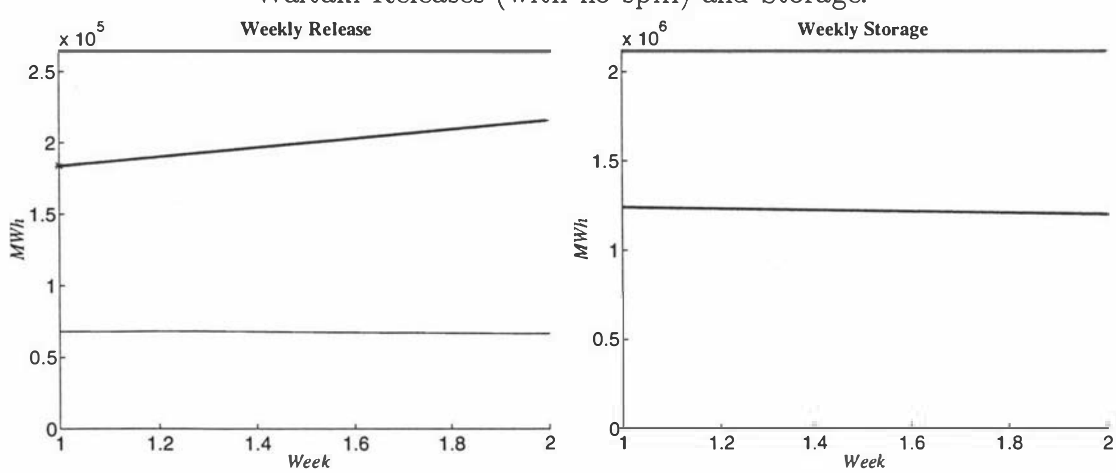
Node 3. Load and New Plymouth Generation.



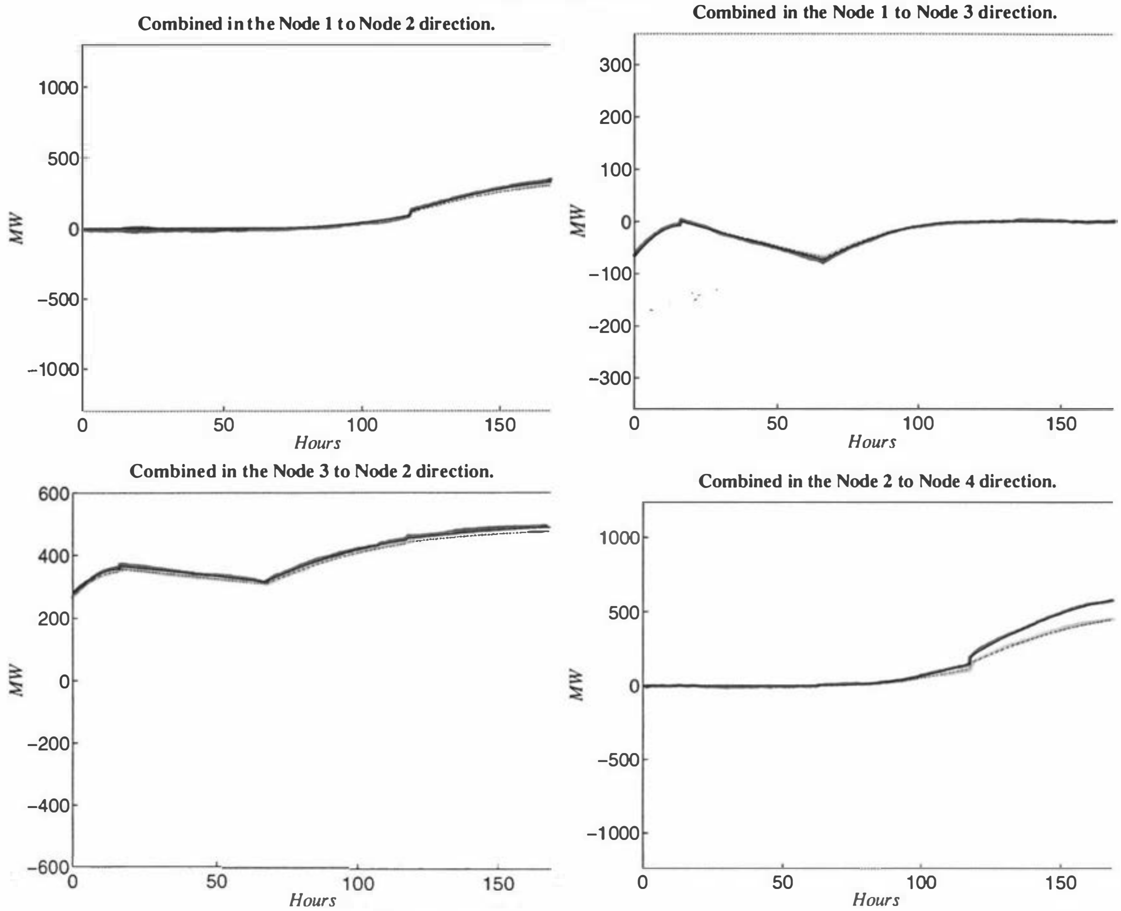
Node 4. Load and Waitaki Generation.

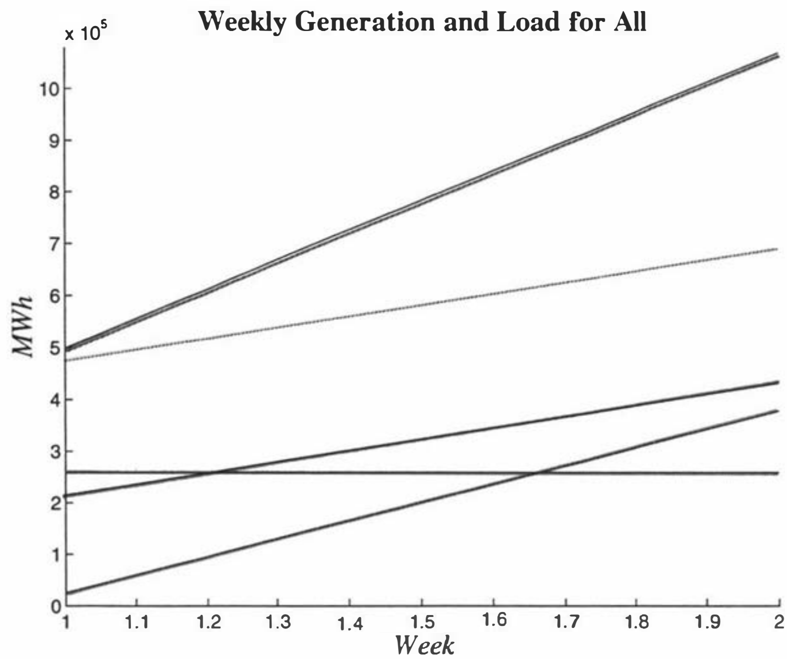


Waitaki Releases (with no spill) and Storage.



Transmission.





Bibliography

[1] John F. Boshier and R. Jonathan Lermit. A network flow formulation for optimum reservoir management of the New Zealand power generating system. *New Zealand Operational Research*, 5(2):85–99, 1977.

[2] John F. Boshier, G. B. Manning, and E. Grant Read. Scheduling releases from New Zealand's hydro reservoirs. *Institute of Professional Engineers in New Zealand, Transactions*, 10(2):33–41, 1983.

[3] Harald Cramér. *Mathematical Methods of Statistics*, volume 9 of *Princeton Mathematical Series*. Princeton University Press, 8th edition, 1958.

[4] Ron S. Dembo, Angel Chiarri, Jesus Gomez Martin, and Luis Paradinas. Managing Hidroeléctrica Española's hydroelectic power system. *Interfaces*, 20(1):115–135, 1990.

[5] H. Durán, C. Puech, J. Díaz, and G. Sánchez. Optimal operation of multireservoir systems using an aggregation-decomposition approach. *IEEE Transactions on Power Apparatus and Systems*, 104(8):2086–2092, 1985.

[6] Electric Power Research Institute. Electric Generation Expansion Analysis System, 1982. EPRI EL-2561.

[7] Thorkell Helgason and Stein W. Wallace. Approximate scenario solutions in the Progressive Hedging Algorithm. *Annals of Operations Research*, 31:425–444, 1991.

[8] Yoshiro Ikura, George Gross, and Gene Sand Hall. PGandE's state-of-the-art scheduling tool for hydro systems. *Interfaces*, 16(1):65–82, 1986.

[9] J. Lermit, H. Barr, K. Garner, D. Rhoades, and G. Read. A review of system characteristics and modelling techniques for multi-reservoir hydro-thermal electricity systems. Internal ECNZ Report, 1990.

[10] Chao-an Li, Rui Yan, and Jing-yan Zhou. Stochastic optimization of interconnected multireservoir power systems. *IEEE Transactions on Power Systems*, 5(4):1487–1496, 1990.

[11] C. Lyra and H. Tavares. A contribution to the midterm scheduling of large scale hydrothermal power systems. *IEEE Transactions on Power Systems*, 3(3):852–857, 1988.

[12] J. M. Mulvey and H. Vladimirou. Applying the Progressive Hedging Algorithm to stochastic generalized networks. *Annals of Operations Research*, 31:399–424, 1991.

[13] N Nabona. Multicommodity network flow model for long-term hydrogeneration optimization. *IEEE Transactions on Power Systems*, 8(2):395–404, 1993.

[14] M.V.F. Pereira and L. M. V. G. Pinto. Application of decomposition techniques to the mid - and short - term scheduling of hydrothermal systems. *IEEE Transactions on Power Apparatus and Systems*, 102(11):3611–3618, 1983.

[15] M.V.F. Pereira and L. M. V. G. Pinto. Stochastic optimization of a multireservoir hydroelectric system: A decomposition approach. *Water Resources Research*, 21(6):779–792, 1985.

[16] M.V.F. Pereira and L. M. V. G. Pinto. Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52:359–375, 1991.

[17] Andrew B. Philpott and Geoffrey B. Leyland. Scenario aggregation for weather-dependent race planning. In *Proceedings 29th Annual Conference*, pages 250–258, Auckland, 1993. Operational Research Society of New Zealand (Inc.).

[18] Power Systems Research Inc. *Stochastic Operations Scheduling of Multi-reservoir Hydrothermal Systems with Transmission Constraints—the SDDP Model.*, 1993. Version 1.2.

[19] R. T. Rockafellar and Roger J.-B. Wets. Scenarios and policy aggregation in optimization under uncertainty. *Mathematics of Operations Research*, 16(1):119–147, 1991.

[20] Richard E. Rosenthal. A nonlinear flow algorithm for maximization of benefits in a hydroelectric power system. *Operations Research*, 29(4):763–786, 1981.

[21] S. A. Soliman and G. S. Christensen. Application of functional analysis to optimization of a variable head multireservoir power system for long-term regulation. *Water Resources Research*, 22(6):852–858, 1986.

[22] Roger J-B Wets. "...the real problem" (George B. Dantzig). Lecture giving Special Honours to George B. Dantzig at the 15th International Symposium on Mathematical Programming, Ann Arbor, Michigan, 1994.