KTH Electrical Engineering

Efficient Operation and Planning of Power Systems

Lennart Söder *Mikael Amelin*

Eleventh edition

Royal Institute of Technology
Electric Power Systems
Stockholm 2011

Cover illustrations:
Virginia wind power plant, Näsudden, Gotland, Sweden
Kisiizi Hospital hydro power plant, Kisiizi, Rukungiri, Uganda
Bristaverket combined heat and power plant, Märsta, Uppland, Sweden

PREFACE

This compendium provides a general overview of how an electricity market works, as well as descriptions of various models and methods that can be used for calculations concerning efficient balance keeping between production and consumption in a power system. The texts presented in the compendium have evolved during several years and have been used in the education at KTH and the Mittunversitetet.

This eleventh edition of the compendium is more or less identical to the previous editions; some misprints have been corrected, and a few examples and exercises have been added. An important difference compared to the seventh edition concerns the notation used in the section on hydro power planning. In order to be more in line with the standards used by the Swedish power companies, this compendium uses the symbol Q for discharge and V for inflow. Unfortunately, the seventh edition used exactly the opposite notation.

We would like to give a warm thank you to the graduate and undergraduate students who have pointed out errors and contributed with suggestions for improvements of the compendium.

Stockholm *Lennart Söder*
January 2011 *Mikael Amelin*

CONTENTS

ABBREVIATIONS AND NOTATION

Since this compendium covers quite a wide area, it is unfortunately impossible to avoid that the same symbol may have different meanings in different contexts. An overview of the most important abbreviations and symbols used in the compendium is listed here in order to at least simplify somewhat for the reader.

Abbreviations

AB	aktiebolag (Swedish for limited company)
AC	alternating current
DC	direct current
HE	hour equivalent
HVDC	high-voltage direct current
LP	linear programming
MILP	mixed integer linear programming
MSEK	million Swedish Kronor
SEK	Swedish Kronor

Modelling of Electricity Markets and Power Systems

Indices

$\lozenge_g$	(thermal) power plant
$\lozenge_i$	hydro power plant
$\lozenge_j$	segment in piecewise linear function
$\lozenge_n$	area
$\lozenge_m$	area
$\lozenge_t$	time period
$\hat{\lozenge}$	installed capacity of $\lozenge$
$\bar{\lozenge}$	maximal value of $\lozenge$
$\underline{\lozenge}$	minimal value of $\lozenge$

Sets

$\mathcal{G}$	(thermal) power plants
$\mathcal{K}$	hydro power plants directly upstream
$\mathcal{M}$	all downstream hydro power plants

$\mathcal{N}$	areas
$\mathcal{P}$	transmission lines

Functions

$B_{\Diamond}(\Diamond)$	benefit function of $\Diamond$
$C_{\Diamond}(\Diamond)$	cost function of $\Diamond$
$L_{\Diamond}(\Diamond)$	loss function of $\Diamond$
$\alpha_{\Diamond}$	constant term in the function $\Diamond$
$\beta_{\Diamond}$	coefficient of linear term in the function $\Diamond$
$\gamma_{\Diamond}$	coefficient of quadratic term in the function $\Diamond$

Variables and parameters

D	load
E	equivalent load
ENS	unserved energy
f	frequency
G	(thermal) generation
H	hydro generation
h	heat contents of fuel
$LOLO$	power deficit
M	contents of reservoir
$MTTF$	mean time to failure
$MTTR$	mean time to repair
O	outage in power plant
P	transmission
p	purchase from power exchange
p	availability
Q	discharge in hydro power plant
q	unavailability
R	gain
r	sales to power exchange
S	spillage
s^*	start of thermal power plant
s^+	start of thermal power plant
s^-	stop of thermal power plant
T	duration of time period
t	time deviation
t^+	up time
t^-	down time
TOC	total operation cost
U	unserved power
$\bar{U}_W$	maximal unused capacity in non-dispatchable power plants
$\bar{U}_{WG}$	maximal unused capacity in all power plants
u	unit commitment of thermal power plant
V	inflow
W	non-dispatchable generation
z	active segment in piecewise linear function
γ	production equivalent
η	(relative) efficiency
λ	electricity price

λ	failure rate
μ	marginal production equivalent
μ	repair rate
τ	water delay time between two hydro power plants
τ	time constant of thermal power plants
ϕ	fuel price

System indices

EENS	expected energy not served
EG	expected generation
ETOC	expected total operation cost
LOLP	risk of power deficit

Random Variables and Monte Carlo Simulation

$a_{\Diamond}$	coefficient of variation for ◊
$Cov[\Diamond_1, \Diamond_2]$	covariance between $\Diamond_1$ and $\Diamond_2$
$E[\Diamond]$	expectation value of ◊
$F_{\Diamond}(x)$	distribution function of ◊
$\tilde{F}_{\Diamond}(x)$	duration curve of ◊
$f_{\Diamond}(x)$	density function of ◊
$g(\Diamond)$	mathematical model
$\tilde{g}(\Diamond)$	simplified mathematical model
$m_{\Diamond}$	estimated expectation value of ◊
$m_{\Diamond h}$	estimated expectation value of ◊ in stratum *h*
$N(\mu, \sigma)$	normal distribution with mean μ and standard deviation σ
n	number of samples
$P(\Diamond)$	probability of ◊
$s_{\Diamond}$	estimated standard deviation of ◊
$s_{\Diamond h}$	estimated standard deviation of ◊ in stratum *h*
U	pseudorandom number
$U(a, b)$	uniform distribution between *a* and *b*
$Var[\Diamond]$	variance of ◊
X	result variables
Y	scenario parameters
Z	control variates
$\Phi(x)$	distribution function of the standardised normal distribution
$\phi(x)$	density function of the standardised normal distribution
$\mu_{\Diamond}$	true expectation value of ◊
$\mu_{\Diamond h}$	true expectation value of ◊ in stratum *h*
$\rho_{\Diamond_1, \Diamond_2}$	correlation coefficient of $\Diamond_1$ and $\Diamond_2$
$\sigma_{\Diamond}$	true standard deviation of ◊
$\sigma_{\Diamond h}$	true standard deviation of ◊ in stratum *h*
ω_h	stratum weight of *h*
$\Diamond^*$	complementary random number of ◊

Chapter 1
INTRODUCTION

This compendium describes methods and models for operation planning of electric power systems. The compendium includes both a general overview of the function of electric power systems and electricity trading, as well as a more technical part. The general part is opens with this chapter, where the subject system planning is introduced and a brief summary of the history of electricity and the structure of a power system. In chapter two it is explained which players can be found in an electricity market and how they can trade with each other. Then, in chapter three provides an outline of the factors which influence the price in an electricity market.

The more technical part of the compendium is started in chapter four, which describes the control systems which are necessary to maintain the instantaneous balance between production and consumption in a power system. Chapter five considers short-term planning, i.e., methods used by the players of the electricity market to plan their actions for the closest future (i.e, the coming hours or days). Then, chapter six describes different simulation methods which can be used to analyse the expected behaviour of an electricity market. The compendium is then closed by a few appendices which primarily describes the mathematics used in the compendium, but which also discusses software which can be used in this context.

1.1 SYSTEM PLANNING

System planning is about using the resources of a system in the best possible way. This planning has to consider the technical prerequisites of the system as well as the economy; one desires the best possible performance to the least possible price. This of course means that there is a trade-off between technology and economy. This trade-off depends to a great extent on the time period of the planning. Some examples are given in table 1.1.

In a very short time perspective, seconds and minutes, the technology and safety of the system is the most important. In an aircraft there are for example a number of control systems, which function is to at any cost avoid a disaster. In an office each of the co-workers are devoted to less dramatic—but for the office system equally important—tasks like writing letters, communication with customers and other duties. In a power system it is necessary to keep a stable frequency, since otherwise it might be necessary to shut down parts of the system.

In a slightly longer time perspective, maybe hours or days, maintaining the function of the system is still very important, but it is also necessary to include economical considerations. If an aircraft needs to make an emergency landing then cost is not a matter, but during normal conditions it is more important to keep the timetable and to use the fuel in a reasonable way. In an office the

Table 1.1 Comparison of the planning for different systems

System	Regulation	Operation	Planning	Investments
Aviation	technical function of aircraft and ground control	emergency plans, keeping schedules, economical flying	air routes, maintenance, pricing	new aircraft
Papermill	machine function	economical operation, varying raw material, optimal quality	submitting offers, pricing, maintenance	new machines
Office	the individual task of each co-worker	coordination, deadlines	submitting offers, estimating costs, education	new staff, new equipment, new premises
Power system	primary control, secondary control	coordination of power plants, economical operation	pricing, maintenance	new power plants, grid expansion, long-term contracts

seconds → years
technology → economy
security → uncertainty

short-term planning is about coordinating the duties of the co-workers so that all problem can be solved in time, but economical questions (e.g. "should we postpone the deadline of this project or should we hire a consultant to finish it in time?") also need to be considered. A power company must for example plan which power plants should be used, and if the company should trade at the power exchange.

The economical decisions become more dominating when the time perspective is weeks or month. The most important considerations are pricing and planning how to use the available resources; an aviation company must decide which routes they should operate; a papermill must decide which kind of paper they should have in their assortment; in an office it must be decided which kind of projects the company should undertake, etc. A power company must among other things plan how to use the available hydro energy over a longer time period (this is in practice equivalent to assigning a price to the water) and decide whether or not the company should buy an insurance against too high or too low electricity prices.

Maintenance planning of machines, education staff, etc., also has a time perspective of weeks or months. It is necessary to plan these activities so that the impact on the daily operation is minimal, while considering some technical limitations: How many aircraft do we have? Which kind of paper products can we manufacture in the existing machines? Can we close these power plants for maintenance without having a significant risk of power deficit? It is also necessary to include a certain level of uncertainty in the planning: Will the equipment continue to work properly if we postpone the maintenance another month? Will we be able to solve the new project without further education of the staff?

In a really long time perspective, i.e., one or more years into the future, the technical limitations are few, as it is always possible to make investments in new equipment or hiring new staff members. The investment decisions are only about economy: Is it profitable to purchase new aircraft? Will we have enough new projects for the office in order to keep new staff members busy? Is it profitable to build a new power plant or a new transmission line? These questions are complicated by the need to include a considerable amount of uncertainty of future events. Therefore, it is necessary to have a good understanding of the underlying technology in order to decide whether or not an investment is profitable.

In this compendium we will consider operation and planning of electric power systems, but the models and methods which are used show many similarities to corresponding problems in other systems. There are also many similarities between planning problems with different time perspectives, but it is always necessary to adjust the models to the particular system and time perspective that is studied.

1.2 BRIEF HISTORY

The first documented reference to an electric phenomenon is from Ancient Greece, where people had noted that a strange force was created when amber was rubbed with a piece of fur; the amber could then attract light objects such as hair. The strange force was what we today refer to as static electricity. It is possible that electricity also came to practical usage already during the antiquity. In a village close to Baghdad, archeologists have found clay jars which possibly could have been used as simple batteries. The findings are hard to date, but might be from about 250 BC. The German Wilhelm König suggested the theory that the "batteries" could have been used for electroplating of metallic objects. The Canadian Paul Keyser presented an alternative theory, in which the batteries would have been used for medical or religious purposes. The skeptics have a more commonplace explanation: the clay jars could have been used as storage vessels for sacred scrolls.

Regardless of whether the findings from Baghdad were used as batteries or not, it would take another two millennia before electricity could be explained and utilised for practical applications. In 1600 the Englishman William Gilbert published a paper on magnetism, *De magnete, magneticisique corporibus*, where he coined the word "electricity" from "elektron", which is the Greek word for amber. However, during the 17th and 18th centuries, only minor progress was made in the exploration of electricity. In 1660 the German Otto van Guericke invented a machine to generate static electricity, and using this device the phenomenon could be studied closer. In 1747 the Englishman William Watson could show that a spark of static electricity was the same as an electric current. Using a kite during a thunderstorm, the American Benjamin Franklin showed in a famous (and extremely dangerous) experiment from 1752 that lightning also constituted an electric current.

The development took a leap during the 19th century. Inspired by the discoveries of the Italian physician and physicist Luigi Galvani during the 1790s, his compatriot Alessandro Volta could develop the first battery in 1800. During the following decades, the battery was further refined by several inventors, for example the Frenchman Gaston Planté who designed a rechargeable battery. The next major step was the generator, which made it possible to transform mechanical energy into electric current. The principles of the direct current generator was developed by the Englishman Michael Faraday 1831-32.

When it was possible to generate a continuous current and not just bursts of static electricity, it also became possible to design devices powered by electricity. The first electric lamp was designed by the Englishman Humphry Davy in the beginning of the 19th century, when he connected a battery to a thin piece of carbon, which began to glow. The first direct current motor was created by Faraday in the 1820s.[1]

Up to this time, electricity had only been of interest to scientist or possibly as a fascinating phenomenon to show off at private parties. The first major application of electric energy was the electric telegraph. In 1820 the Dane Hans Christian Ørstedt had discovered that a electric current could affect a compass needle. This discovery was followed in 1825 by the first electromagnet, designed by the Englishman William Sturgeon. Thanks to these advances it became possible to transmit messages using electricity. Patents of electric telegraph systems were issued in 1837 to the Englishmen William Dothergill Cooke and Charles Wheatstone, as well as the American Samuel Morse.

1. A more modern direct current motor was invented by coincidence at an exhibition in Vienna 1873 when the Belgian Zénobe Gramme connected a rotating generator to a similar generator, which then started to work as a motor.

In the 1870s the incandescent lamp was developed in parallel by the Englishman Joseph Swan and the American Thomas Edison. Electric lighting was the main usage of electricity during the 1880s, when the first public power systems were built. The first enterprise was in Godalming, England, 1881 and was followed in 1882 by Holborn Viaduct, London, and Pearl Street, New York. These enterprises quickly spread to the rest of Europe and North America (for example, the first Swedish public power system was opened in the small city of Härnösand in 1885).

The first power systems comprised limited geographical areas, because the electric losses were too large on long power lines. To reduce the losses it was necessary to increase the voltage of the power lines and this was not possible until the alternating current systems replaces direct current. During the 1880s, several inventors and engineers contributed to the development of the alternating current generator and the transformer. With that it became possible to transform low voltage to higher voltage levels, thus making it possible to build larger power systems.

During the first half of the 20th century, electricity slowly but surely gained ground, and the isolated systems of the 19th century could be interconnected to national and international power systems. The electricity was as before mainly generated in hydro power plants and thermal power plants fuelled by coal and oil. Although the principles of the fuel sell had been described already in 1839 by the Swiss Christian Friedrich Schönbein (and already the same year the Welshman William Grove built a working prototype), but the fuel cell is still today too expensive to compete with other electricity generation other than for some specific purposes. The same goes for photovoltaics, which generate electricity using the photoelectric effect.[2] Instead it was the nuclear power which became the first addition of non-traditional electricity generation. The first nuclear reactor used for electricity generation was tested in USA in 1951 and the first nuclear power plant used for large-scale electricity generation was opened in the Soviet Union in 1954.

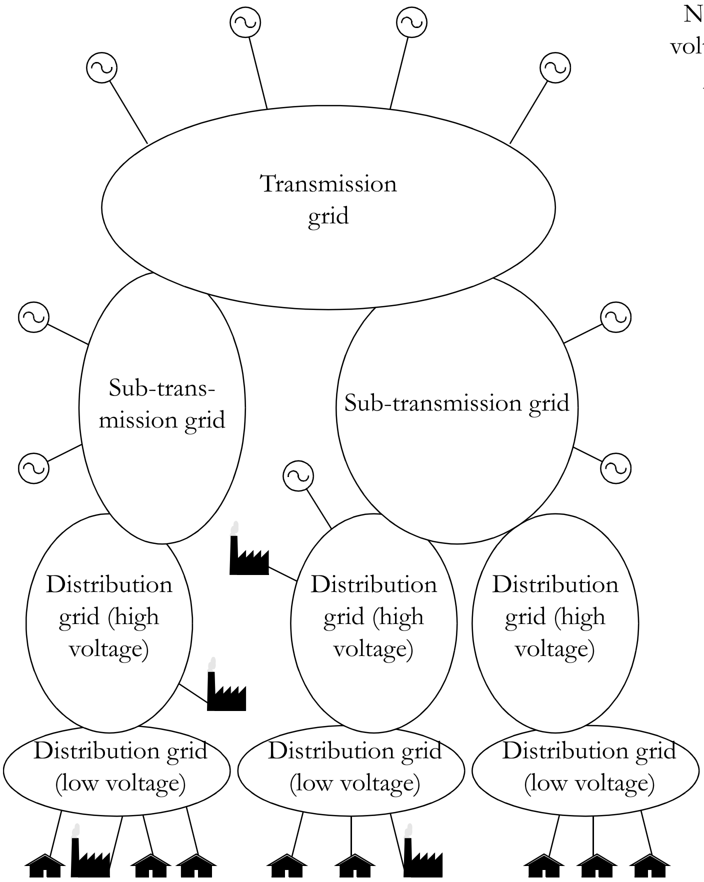
The technology development within electronics was huge during the second half of the 20th century, causing an increase in electricity demand, but the results were used within the power industry. Using new components as for example thyristors increased the possibilities to control and efficient operation of large power systems. One of the applications of the new power electronics was HVDC (High Voltage Direct Current). However, the first HVDC links—for example the Soviet Union link Moscow-Kashira[3] from 1951 and the Swedish link Västervik-Gotland from 1954—used mercury arc valves (invented by the American Peter Hewitt Cooper in 1902) instead of thyristors. Another usage for power electronics are so-called FACTS components,[4] i.e., components which increase the controllability of AC lines, making it possible to utilise them in a more efficient manner.

As a result of the oil crisis in the beginning of the 1970s and an increased environmental awareness during the 1980s, more and more interest has been given to renewable alternatives for electricity generation. Most of the renewable electricity generation comes from hydro power (which has been used since the beginning of electricity). Since the 1980s, wind power has been subject to a rapid technology development and the installed wind power capacity has increased by 35% annually the last decade. Also the increase of biofuel in thermal power plants has increased.

1.3 ELECTRIC POWER SYSTEMS

By an electric power system we refer to all kinds of system including one or more generators, which supplies one or more electric load via an electric grid. This is a rather wide definition, which

2. By the way, it was his explanation of the photoelectric effect that earned Albert Einstein the Nobel Prize in physics 1905.
3. This link was by the way based on a never finalised German experiment, which was conquered by the Red Army during the World War II.
4. Short for *Flexible Alternating Current Transmission System.*



Nominal voltage[kV]	Notion	
1 000		} ultra high voltage
800	800	extra high voltage
400	400	
220	200	
132	130	high voltage
66	70	
45	50	
33	30	
22	20	
11	10	
6.6	6	industrial grids only
3.3	3	
0.4		} low voltage

Figure 1.1 The structure of a larger power system.

includes for example simple household appliances (such as an electric torch) and various vessels (cars, aircraft, ships, satellites). However, in this compendium we will focus on larger power systems, which supplies a large number of consumers with varying demand for electricity, i.e., the type of power system which constitutes such an important part of the infrastructure of a modern society.

Such power systems have a hierarchical structure, as shown in figure 1.1. The purpose of the grid is to transfer electric energy from the generation sources to the final consumers. The transmission grid constitutes the top level of the grid. The transmission grid of a country is often directly connected to the transmission grid of the neighbouring countries (cf. figure 1.2). The purpose of the transmission grid is to transfer large amounts of power over long distances. Earlier it was necessary to use AV to achieve this, but today HVDC is used too (see section 1.2). Consumers are rarely connected directly to the transmission grid, but many large power plants are connected at this level.

Below the transmission grid we find the sub-transmission grid (sometimes referred to as regional grids), which serves as a link between the transmission grid and the distribution grid. Generally, the transferred power is smaller and the distances shorter in the sub-transmission grids compared to the transmission grid; hence, somewhat lower voltage levels are used. AC is most common, but HVDC can also be used. Power plants might be connected directly to the sub-transmission grid, but it is still uncommon with consumers at these voltages.

The final link in the public electricity grid is the distribution grids, which extends all the way to the final consumers. It is common to differentiate between distribution grids for high voltage, to which larger consumers (for example industries) might be connected, and distribution grids for

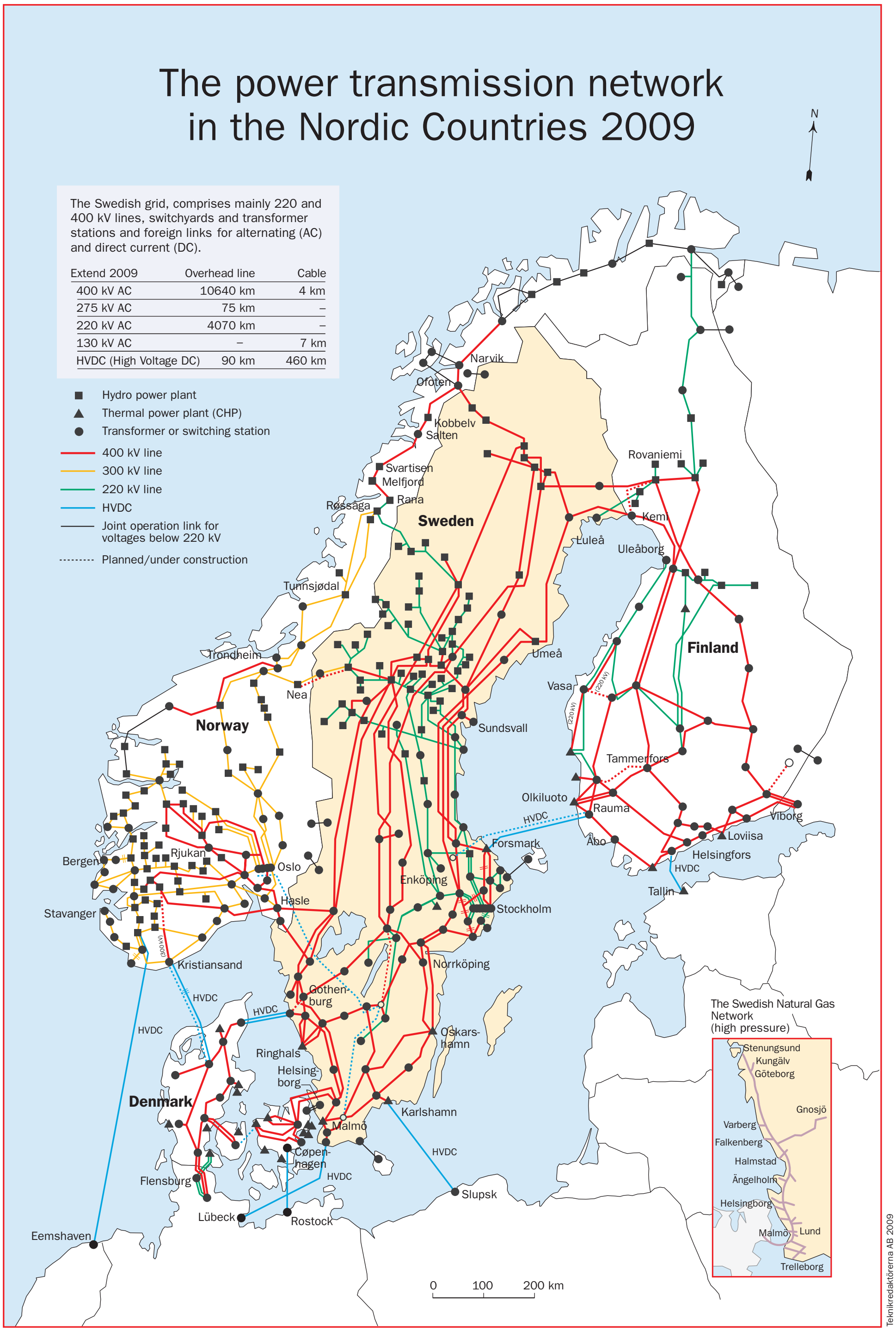


Figure 1.2 The transmission grid in Sweden and its neighbouring countries.

low voltage, which use the voltage levels found in ordinary wall sockets. Most of the consumers are connected to the distribution grids. There are also small-scale power plants which are connected to the distribution grid; this is referred to as distributed generation.

It is common that the consumers which are connected to the public grid have internal grids, which transfers the power to the electric devices the consumer wishes to use. These internal grids are today almost exclusively AC grids, but considering that electronic devices constitute a large part of the load nowadays, it is possible that DC will be used to a larger extent in the future.

Chapter 2

THE STRUCTURE OF AN ELECTRICITY MARKET

Once a market was simply a place people where people would meet to buy, sell and trade goods. Today the word is used in a more figurative meaning to describe an arrangement for exchanging goods. Hence, an electricity market is an arrangement to transfer electric energy from producers to consumers. Electric energy is a somewhat special commodity, which has an impact on the structure of an electricity market. Roughly speaking, the electricity market can be divided into three parts (see figure 2.1).

To transfer electric energy it is necessary to have a special infrastructure in the form of wires between producers and consumers. It would of course possible that each consumer had an exclusive direct wire to the producer from which the consumer is buying electricity, but such a solution would be very expensive. The only reasonable way for a large group of producers and consumers to trade electric energy is through a common *power system*. Electric energy cannot be stored; therefore, the only way to maintain the balance between generation and load is to use automatic control system (these systems will be further discussed in chapter 4). It is also necessary to have technical systems which monitors that individual lines are not overloaded and knocked out, as this could lead to the collapse of the entire or large parts of the power system.

Hence, the power system secures that the consumers receive power when they need it. Naturally,

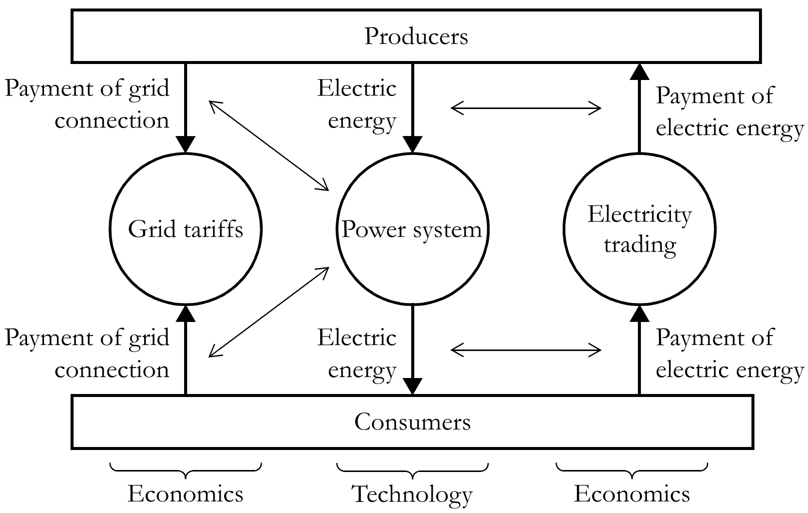


Figure 2.1 Overview of an electricity market.

it costs to build and operate the grid on which the power is transferred; therefore, it has to be arranged that the players of the electricity market pays the investment and operation costs of the grid. This is achieved by *grid tariffs*, which can be designed in several ways. In this presentation, we will however not discuss this issue any closer.

The players who are using electricity should of course also pay for the generation of the energy. To administrate the payments of electric energy it is necessary to establish a system for *electricity trading*. The electricity trading can also be organised in different ways, and in this chapter we will try to provide an overview of the basic concepts of various solutions. The players that might exist in an electricity market is introduces in the first part of the chapter, and the second part describes different kinds of trading that can be found in an electricity market.

2.1 PLAYERS IN AN ELECTRICITY MARKET

This section describes the various functions that the players may have in an electricity market. It is rare that each player in an electricity market has a single function; usually each company or authority can appear in several roles. Exactly which combinations are possible varies from electricity market to electricity market, but some examples that can be mentioned are large power companies that often have both electricity generation and a trading division, some process industries which may both be large consumers of electricity but which also have some electricity generation of their own, and public agencies which appear both as owners of the transmission grid and as system operator.

Producers and Consumers

Quite naturally, the producers are the players that own and operate the power plants in the electricity market. Equally obvious is that the consumers are the players that are the final consumers of the electric energy transferred through the grid. In figure 2.1 we find producers and consumers in all three sectors; they are of course connected to the power system, both producers and consumers have to pay for the benefit of being connected to the grid and the consumers pay the producers for their electricity generation through the electricity trading.

It might be noted that electricity generation is an activity with major economics of scale, which in several electricity markets have resulted in a few, large electricity producers. The number of consumers on the other hand is considerably higher and their consumption can vary significantly concerning peak power and annual energy consumption.

Retailers

Many consumers in electricity markets are to small to purchase directly from producers or a power exchange. These consumes can then turn to a retailer, who will purchase electricity on their behalf. Thus, the business idea of the retailers—which sometimes are referred to as traders or independent traders (if they lack generation resources of their own)—is to buy electricity directly from producers or the power exchange and resell it to consumers. Hence, the retailers are only involved in the electricity trading sector in figure 2.1; they serve as a link between producers and consumers.

It might appear as if these retailers were just unnecessary, price increasing middlemen—and in the worst case this might actually be true—but they can also supply important functions to the electricity market. Primarily the existence of retailers means larger freedom of choice for the consumers, resulting in increased competition compared to if retailing was run by producers only. The increased competition may not just apply to the electricity price, but it is also possible that enterprising retailers offer better service (for example more employees answering the telephones at the customer service) or special electricity products (such as for example power produced in environmentally benign power plants). The retailers may also take over part of the risks (both towards producers and consumers) by offering stable prices during longer periods than one trading period.

System Operator

It would be very difficult for an electricity market to be organised by its own; there is a need for a player which maintains safe operation of the power system and which administrates the electricity trading. This player is referred to as system operator (commonly abbreviated ISO[1] or TSO[2]). The system operator may have many tasks. One of the most important is to be responsible for the technical operation of the power system, which among other things means to be responsible for frequency control (which is described in chapter 4). It is also usually the system operator who manages the post trading (see section 2.2.3).

It is self-evident that the system operator has a large influence on the electricity market. If the system operator also was a producer or retailer, it would be possible for the system operator to favour itself in an appropriate manner; therefore, the system responsibility is normally given to an independent organisation. It is however common that the system operator also acts as a grid owner, which means that the system operator may have to be a consumer in the electricity market when buying electricity to cover for losses.

It should be noted that a single electricity market may have several system operators. For example, Denmark, Finland, Norway and Sweden have a common electricity market, where each country has its own system operator. The practical division of the system responsibility is regulated by a contract (see the literature at the end of this chapter).

Balance Responsibility

As neither producers nor consumers can know beforehand exactly how much they will produce and consume respectively, it is unavoidable that the energy which is actually transferred by the grid deviates more or less from the plans of the players. These deviations are compensated in the physical power system by automatic control systems and the actions of the system operator. However, they deviations must also be accounted for in the electricity trading, to assure that the players are paid for all the energy that they have supplied to the grid and are paying for all the energy which has been extracted. The players which are responsible for this financial adjustment are referred to as *balance responsible* players. In figure 2.1 the balance responsible players are depicted as the arrows between the power system and the electricity trading.

All players who participate in the electricity trading do not have to be balance responsible, but it is possible to transfer the responsibility to another player. It would for example be difficult for the average residential consumer to manage their own balance responsibility; therefore, it is common that the retailer take over their balance responsibility. As there is a cost for being balance responsible (see section 2.2.3), the player who takes over the responsibility will of course require a payment for this service.

Grid Owners

Unlike electricity generation it is not appropriate to have a competitive market for transmission and distribution of electricity, as they constitute so-called natural monopolies. This means that the costs of the investments necessary to enter the market are so high that it is not beneficial to the so-

1. Independent System Operator.
2. Transmission System Operator.

ciety to have several competing firms. In other words, the cost of building parallel grids would be significantly higher than the possible price press that would be the result of competing grids. Rather than letting the grid owners compete it is often decided to give some companies or municipal authorities the exclusive right of power distribution within a limited area. Moreover, there is generally a company which owns the transmission network.

The main task of the grid owner is to operate and maintain the grid, and to provide an adequate power quality. The grid owner is also responsible for measuring generation and consumption for the producers and consumers who are connected to the grid. The grid owner may also have to buy electricity to cover the electric losses of the grid. To cover the costs of these tasks, the grid owner charges grid tariffs of the users. The grid tariffs are usually divided in a power part and an energy part. The power part corresponds to the fixed costs for building and maintaining the grid—these costs are primarily depending on the peak power of the grid user. The energy part corresponds to the variable costs, i.e., primarily the electric losses.

To prevent the grid owners from exploiting their monopoly, it is common to regulate the grid tariffs. The regulation may either be cost-based or performance-based. In short, the cost-based regulation means that the grid owner may charge tariffs which cover the actual costs of the grid and a certain profit. In performance-based regulation the grid tariffs depend on how well the grid works; the better the performance, the higher tariffs can be charged.

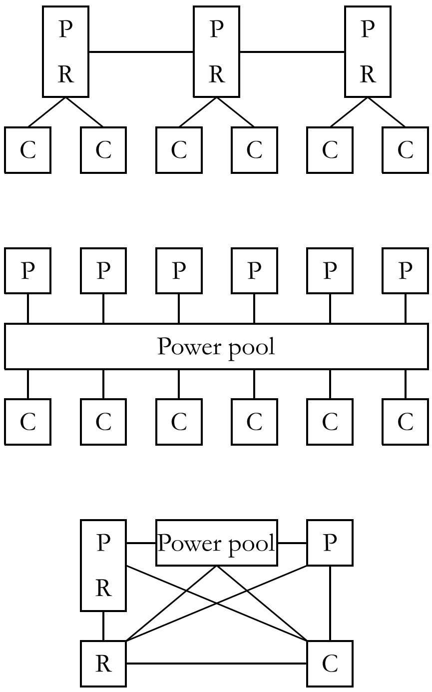
2.2 ELECTRICITY TRADING

In the introduction of this chapter it was established that the objective of the electricity trading is to ensure that the producers are paid for the energy they have generated and that the consumers pay for the energy they have consumed. However, as the power system is partly operated by automatic control system, the payment cannot be performed in real-time. The solution is to introduce trading periods, the duration of which can be chosen arbitrarily. The most common choice is to trade per hour, but some parts of the world (for example Great Britain) have chosen to trade in half-hour periods.

There are many ways to organise the electricity trading. To overview the different solutions, we may divide them in three major categories (see figure 2.2). The oldest form is the vertically integrated electricity market, where each power company combines the roles of producer, retailer, grid owner and system operator. (The designation vertically integrated actually refers to that the companies manage all steps of the power delivery.) Each company has a franchise for one or more geographical areas, where they have a monopoly of retailing; the consumers of a vertically integrated electricity market has no possibility to choose supplier, but must buy from the local power company. To prevent the power companies from taking advantage of their monopoly, the activities are regulated; hence, it is stated which electricity prices they may charge and which other responsibilities they have.

As the power companies of the vertically integrated electricity market do not compete, they can decide themselves how to solve the technical issues of safe system operation in the best manner. To reduce the operation costs they can also get involved in trading with each other in those cases when a power company has unused production capacity which is cheaper than the power plants of the other companies. It is even possible for the power companies to jointly plan the operation of all power plants in the system. However, sometimes there are difficulties also in the trading between vertically integrated companies. An example is when two companies trade, but have no direct electric connection to each other's grids; the trading will take place via the grid of a third company instead, which causes losses (and possibly other inconveniences) for the third company, which therefore wants to be compensated.

The advantage of a vertically integrated electricity market is that simpler technical solutions can



a) Vertically integrated electricity market. The consumers are forced to buy from the local power company. The power companies may trade freely.

b) Centralised electricity market. All producers must sell to the centralised power pool and all consumers must buy from the pool.

c) Bilateral electricity market. All players may trade freely.

P - Producer C - Consumer R - Retailer

Figure 2.2 Three ways of structuring the electricity trading.

be used when one company manages all parts of the power system, and the investment in generation, transmission and distribution may be coordinated. The obvious disadvantage is that there will not be the same pressure to improve the performance as when several companies compete for the favours of the consumers. Hoping to increase the efficiency of electricity markets, a global restructuring process was initiated during the late 20th century. In the restructured electricity markets the vertically integrated companies are divided, so that competitive activities (production and possibly retailing) are separated from monopolies (system operator and grid owners).

Among the restructured electricity markets we can distinguish centralised and bilateral electricity markets. Characteristic of a centralised electricity market is that producers and consumers may not trade directly. The producers have to submit their sales bids to a central power pool, which is managed by the system operator. In some cases the consumers (maybe represented by a retailer) also submit purchase bids to the power pool, whereas in other cases the system operator forecasts the load during the trading period and buys the same amount from the power pool; thus, the system operator serves as retailer for all consumers. For each trading period the pool either determines an electricity price for the whole market or a number of different electricity prices, which each apply to a part of the market.

In a bilateral electricity market the system operator has a more supervising role. The players do not have to trade through a power pool, but may sell and purchase freely (all transactions must however be reported to the system operator, so that after the trading period it is possible to control that the players have fulfilled their undertakings). There is no official electricity price, but generally there is a power pool also in bilateral electricity markets and the price of the pool serves as a guideline to those players trading bilaterally.

The electricity trading is divided in several steps, as the players of the electricity market do not know exactly how much electricity will be traded during a certain trading period. In the first step—the ahead trading—the players buy and sell as much as they think they need. During the actual trading period there is a real-time trading, the objective of which is to enable the system operator to reschedule generation and consumption to maintain safe operation of the power sys-

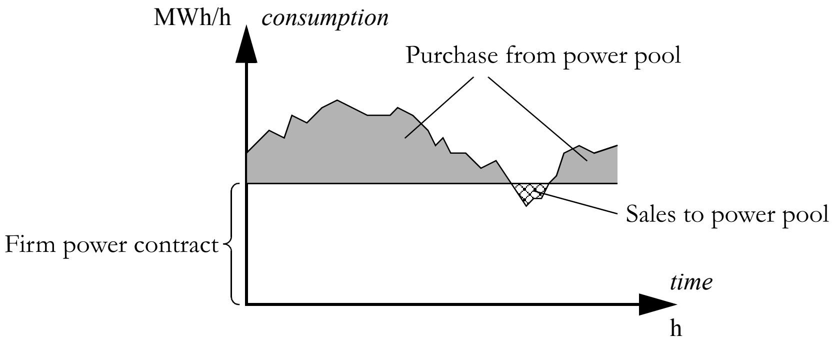


Figure 2.3 Example of the electricity purchase of a consumer. This consumer has a long-term contract to buy a certain number of MWh/h. During those times the consumer plans to use more the long-term contract is complemented by purchase from the power pool (i.e., a short-term contract)., whereas if the consumption is lesser the remainder is sold to the power exchange.

tem. Finally, there is a post trading, where the deviations between planned generation and consumption (i.e., the ahead trading and to some extent the real-time trading) and the actual generation and consumption. Closer descriptions of these three steps follow below.

2.2.1 The Ahead Trading

By the ahead trading we refer to all trading which occurs before the actual trading period. The players are assumed to plan so that they buy or produce themselves as much as they sell or consume themselves. The players may make a number of different agreements, from long-term contracts which are valid for several trading periods, to short-term contracts which only apply to a single trading period and which typically are made the day before the actual trading period. By combining various contracts the players can obtain favourable prices for most of their turnover by long-term contracts, while they can adept to the circumstances at hand using shorter contracts (cf. figure 2.3). Below we will briefly describe the most common types of ahead trading.

Power Pools

Players who want to trade in a power pool submit purchase and sell bids respectively. These bids can be designed in different ways. The most simple form of a purchase bid is to specify the highest price that the player is willing to pay for a certain amount of electric energy during a certain trading period. In a similar manner, a simple sell bid specifies the least price for which the player is willing to sell a certain amount of electric energy during a certain trading period. Hence, the common purchase and sell bids are valid for one particular trading period. It is also possible to allow bids that are valid for several trading periods. Such bids are called *block bids* and must be accepted as a whole. If a player submits a purchase block for 1 000 MWh/h during 5 hours if the price is at most 20 €/MWh, then this bid will only be accepted if the average electricity price during all five hours does not exceed 20 €/MWh.

Also the principle for how the electricity price is calculated may differ between different power pools. The most common arrangement is a power pool with a price cross. In such power pools the sell bids are arranged to a supply curve, by sorting the sell bids in ascending order according to the requested least price. The purchase bids are arranged to a demand curve in a similar manner, by sorting the purchase bids according to descending willingness to pay. The electricity price is then

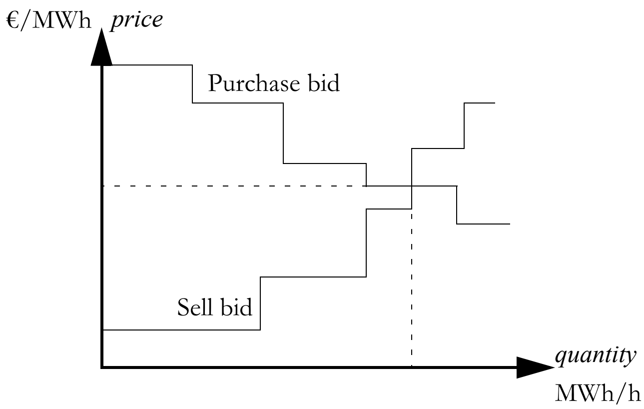


Figure 2.4 Principle of a power pool with a price cross.

Table 2.1 Some examples of power pools.

Power Pool	Area	Trading period	Pricing	Bid have to be submitted
European Power Exchange	Germany, Austria	1 hour	Price cross	No later than 12 a.m. the day before.
Nord Pool Elspot	Denmark, Finland, Norway, Sweden	1 hour	Price cross	No later than 12 a.m. the day before.
Nord Pool Elbas	Denmark, Finland, Sweden	1 hour	As in bid	Not earlier than 2 p.m. the day before and no later than one hour before the trading period.

determined by the price cross, i.e., the intersections between the two curves (see figure 2.4). All bids to the left of the price cross are accepted and they all receive the same electricity price. Thus, all purchase bids where the highest price is higher than the electricity price given for the trading period are accepted, as well as all sell bids where the least price is lesser than the electricity price. Notice that most players receive a price which are more beneficial than the price they were willing to pay and sell for respectively.

Another variant is to have a separate price for each transaction. The power pool lists all submitted bids and if a player finds the price appealing they can accept the bid.

Some power pools (for example the Nordic power exchange Nord Pool) have a limitation to how much may be traded between different parts of the grid. In this case the grid is divided in *price areas*. When a player submits a bid to the power pool, they have to specify the price area in which energy will be injected or extracted as well as the usual quantity, price and trading period. If the trading between two areas should exceed the stipulated limitation, then the market is split in multiple parts with individual electricity prices for each part.

Bilateral Trading

By bilateral trading we refer to all agreements which are made directly by two players. The bilateral trading must be reported to the system operator, as it is to be accounted for in the balances of the post trading market (see section 2.2.3). Bilateral trading is not allowed in centralised electricity markets, where all transactions have to pass the power pool.

Several kinds of contracts are used for bilateral trading. The two most common types of contracts are firm power contracts and take-and-pay contracts. Firm power means that a given constant power is traded during a given time period. This sort of contract is usually used to cover the base load. Take-and-pay contracts mean that the buyer can consume as much as he or she wants, up to a certain limit. This type of contract is for example used by regular residential consumers; in this case the power limit is set by the size of the main fuse.

A bilateral contract also states the electricity price paid by the buyer. The price can either be the same for the entire duration of the contract (fixed price) or changing over time (variable price). A variable price can for example be determined by the electricity price at a power exchange plus a certain uplift.

It can be noted that with the exception of a firm power contract with fixed price, a bilateral contract will result in some uncertainty. A consumer who is buying electricity at a variable price will not know the price in advance and a producer who is selling electricity in a take-and-pay contract will not know the exact sales until after the end of the trading period.

Financial Trading

The financial trading is a trading of various sorts of price insurances. A financial instrument is a bilateral agreement between two players. The financial trading *is not* reported to the system operator and *is not* accounted for in the balances of the post trading market. Therefore, financial trading is allowed also in centralised electricity market, although they actually constitute bilateral contracts.

To simplify the financial trading it is possible to introduce different standard contracts. These contracts can then be traded at special markets at a power exchange. Some examples of financial instruments are options and futures:

- **Options.** Assume that company B wants to buy a given amount of power during a specified time period, but the company is not willing to pay more than a certain maximum price. Company B can make an agreement with company A, in which company A promises to pay the difference if the price on the power exchange is higher than the desired maximum price. Company B can then buy power from the power exchange. Company B pays company A for the agreement. Company A does not have to be a power company.
- **Futures.** Company B wants to buy a given amount of power during a specified period to a specified price. Company B agrees with company A for the specified price. Company A then takes risk. Company A obtains the benefit or the loss depending on the difference between the agreed price and the exchange price. Company A does not have to be a power company.

2.2.2 Real-time Trading

The real-time trading refers to the trading which occurs during a trading period. The most important player in the real-time market is the system operator, who has the responsibility to maintain safe operation of the power system. Therefore, the system operator sometimes has to change the production and consumption respectively, and this task is solved by real-time trading. As usual, there are several ways to design the real-time trading. In this compendium we will distinguish between to main types of real-time trading: a real-time balancing market and central dispatch.

Real-time Balancing Market

In a real-time balancing market, the normal condition is that the players themselves decide how

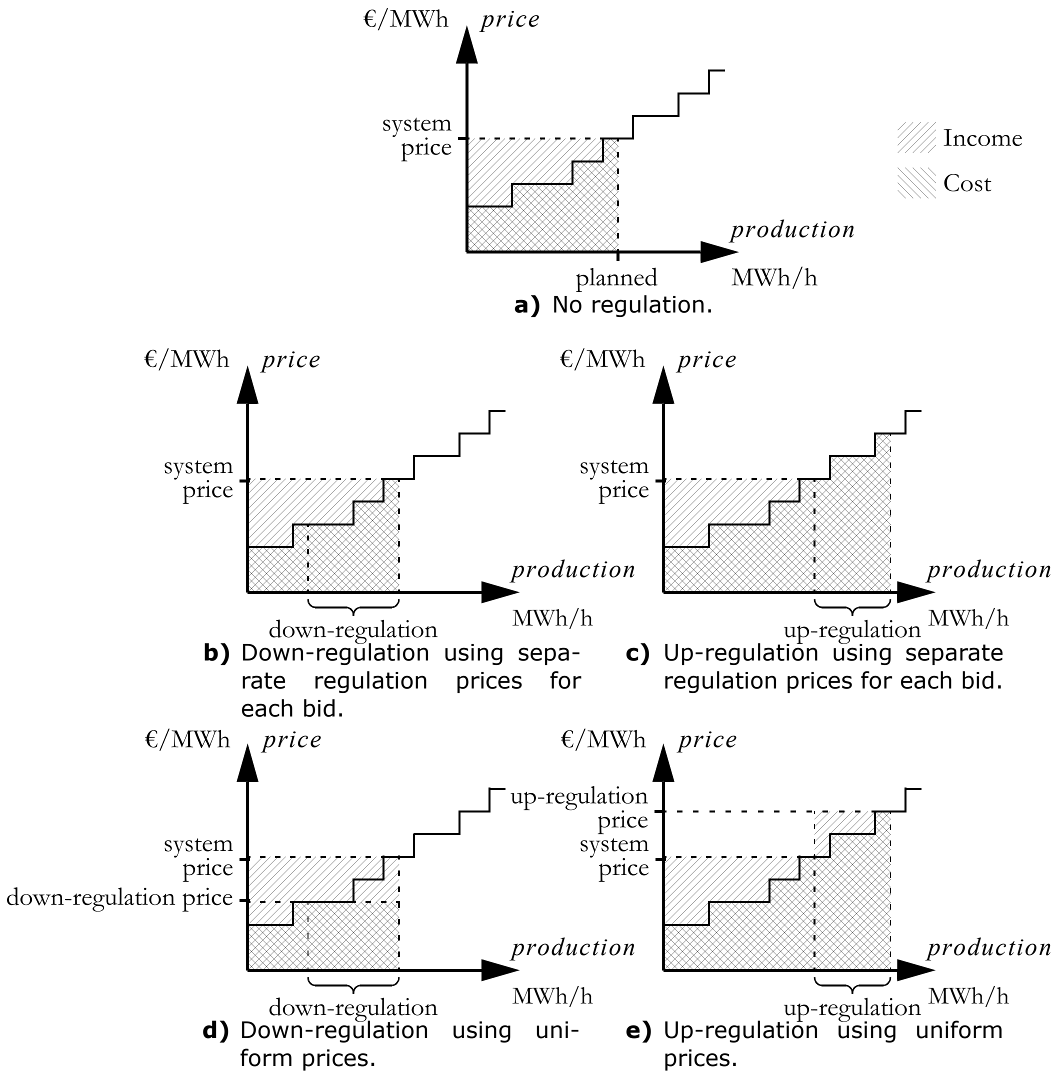


Figure 2.5 Principles of pricing in a real-time balancing market. In this example there are seven power plants with increasing variable production costs. The planned production according to panel a has been sold in the ahead market, resulting in the indicated system price.

Each producer now submits bids to the real-time balancing market. The power plants which can down-regulate are willing to pay a price not exceeding their variable production cost (otherwise buying regulating power is a loss). In a similar way, the power plants which can up-regulate are only willing to do so if they get paid at least as much as their variable production costs.

In panels b and c regulating power is traded using separate prices (corresponding to the upper and lower limits respectively of the submitted bids). In panel d and e uniform down- and up-regulating prices are used instead. The first activated bids now receive a more favourable price than the variable production cost.

much to produce and consume respectively, but if necessary the system operator can persuade a player to change their production or consumption by activating bids which have been submitted to the real-time balancing market. Two kinds of bids can be submitted to the real-time balancing market. When down-regulating the system is supplied with less energy than agreed upon in the ahead market (thus, a producer carries out down-regulation by reducing the production, whereas a consumer carries out down-regulation by increasing the consumption). Down-regulation means that the player buys regulating power from the system operator and a down-regulation bid must therefore state how much the player can down-regulate (in MW) and the maximal price (in €/MWh) which the player is willing to pay for the regulating power. Similarly, up-regulation means that the player sells regulating power to the system operator, i.e., an up-regulation bid states how much the player can up-regulate and the minimum price for which the player is willing to sell regulating power. Unlike the bids to the ahead market, bids to the real-time balancing market are not just financial but physical undertakings—the system operator measures generation and load and may control that activated bids really have been carried out within time.

The pricing of the real-time balancing market can either be different for each activated bid or there may be uniform prices for up- and down-regulation respectively. Separate pricing means that those players buying regulating power pay exactly the price they stated in their down-regulation bid and the players selling regulating power get paid as much as they stated in their up-regulation bids. Using this pricing scheme the players will—provided that the competition is good—not be able to make any profits from the real-time trading; for example, a producer will buy regulating power at the same price as it would have cost to produce it (cf. figure 2.5b) and will when selling receive just as much to cover the production cost (cf. figure 2.5c). Such arrangements are not very attractive neither for producers nor consumers; it implies that they submits bids out of public duty or that they are simply forced to submit bids whenever possible. To make a real-time balancing market more appealing it is possible to use marginal pricing instead, which means that a down-regulation price is defined, which is equal to the lowest price among the activated down-regulation bids, and an up-regulation price, which is equal to the highest price among the activated up-regulation bids. All activated bids will then obtain these regulating prices. Thus, a producer may for example buy regulating power to a price which is less than what it would have cost to produce the same amount in an own power plant (cf. figure 2.5d) or may sell regulating power to a price which is higher than the production cost cf. figure 2.5e).[3]

Central Dispatch

When central dispatch is applied, the system operator decides how much the other players should produce or consume. The dispatch is based on the bids submitted by the other players during the ahead trading. To give the players the possibility of correcting bids based on mistaken forecasts, the players may be allowed to adjust their bids before the real-time trading (to reduce the risk of market manipulation it is possible to require that the players only may change their bids if they provide a reasonable motivation). Each trading period is divided in phases (in for example Australia, five minute intervals are used) and for each phase the system operator performs an economic dispatch, which means that the system operator solves a short-term planning problem based on the bids of the players. The result of the dispatch states for example how much is to be produced in each power plant. Separate prices are obtained for each phase (or varying prices for different parts of the grid if the economic dispatch accounts for losses and/or transmission limitations); these prices are referred to as real-time prices.

2.2.3 Post Trading

As noted earlier, the objective of the electricity trading is to assure that somebody is paying for the

3. It is actually possible to combine the two pricing schemes. An example of this is the Nordic electricity market, where the regulating bids activated to maintain balance between production and consumption are paid uniform up- and down-regulation prices, whereas those bids activated to prevent a part of the grid to be overloaded (so-called counter trading) receive separate prices.

energy transferred in the power system. When a trading period is ended the system operator can compile how much the balance responsible and his or her clients have actually produced and consumed, as well as how much they have bought or sold in the ahead and real-time markets. This will almost certainly result in a deviation between supplied and extracted energy. The purpose of the post trading is to settle these deviations.

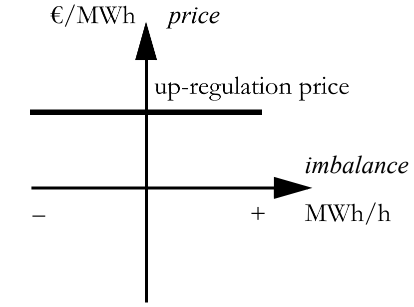
All balance responsible players having an imbalance have to trade in the post market in order to achieve balance. Players having positive imbalances (i.e., they have supplied more energy than they have extracted) sell imbalance power to the system operator. If there is a negative imbalance instead then the player has to buy imbalance power from the system operator.

The price of the imbalance power is generally related to the prices used during the real-time trading and as usual there are several variants. The question is partly whether average pricing is used (as for example in Denmark, England-Wales and Spain) or marginal pricing (as for example in Australia, Finland, Norway and Sweden), and partly whether or not separate prices are used for buying and selling imbalance power. A one-price system means that all imbalance power is bought and sold for the same price (this is the case in for example Australia, Norway, Spain and Germany), whereas a two-price system means that there are separate prices for negative and positive imbalances respectively (this is the case in for example Denmark, Finland, Sweden and England-Wales).

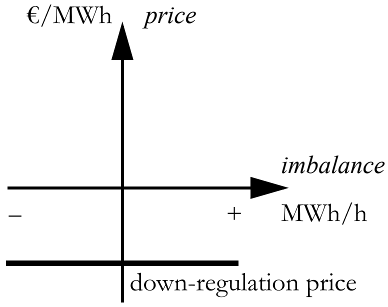
An overview of the principles for one-price and two-price systems is given in figure 2.6. If the real-time trading has required both up- and down-regulations then it must first be decided the direction of the net regulation during the trading period; if the system operator has bought more regulating power than they have sold then the trading period counts as up-regulation and vice versa. In a one-price system the up-regulation price is used for all imbalance power during up-regulation periods (a) and the down-regulation price is used during down-regulation periods (figure 2.6b). If the real-time market uses central dispatch rather than a real-time balancing market, the real-time price is used instead of up- and down-regulation prices.

The two-price system means that a less favourable price is given to those players who are assumed to be the cause of the activation of regulating bids. During an up-regulation period those players who have not supplied enough energy, i.e., which have negative imbalances, must pay the up-regulation price (which is higher than the price in the ahead market), while the players having positive balance receive the same price as in the ahead market (figure 2.6c). During down-regulation on the other hand, those players having positive imbalance are assumed to have caused the need for regulation and therefore are getting paid the down-regulation price (which is less than the ahead market price), while the other players receive the same price as in the ahead market (figure 2.6d). This approach can also be applied to real-time prices originating from a central dispatch; if the real-time price is higher than the ahead market price then there has been an up-regulation period and vice versa.

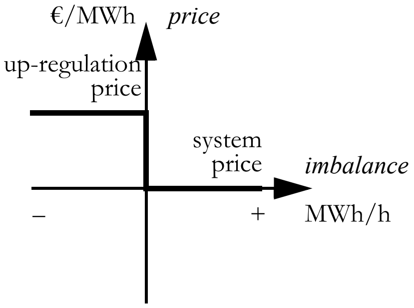
Using a one-price system a balance responsible having an imbalance in “the right direction” (i.e., positive imbalance when the system has a net up-regulation and negative imbalance when the system has a net down-regulation) may buy or sell imbalance power to a price more favourable than in the ahead market, which means that there in some cases will be profitable to have an imbalance. Thus, the one-price system introduces a possibility to intentionally obtain an imbalance, but it is hard to see how a player systematically could take advantage of this possibility. However, the one-price system means that the costs decrease for those being balance responsible of unpredictable production or consumption, because the cost of the occasions when the player has an imbalance in “the wrong direction” is to some extent compensated by the income of the occasions when the player has an imbalances in “the right direction”. In a two-price system it is never possible to make any profits from an imbalance, which results in higher costs for the balance responsible players, who accordingly can be assumed to feel more pressure to keep their own balance in every trading period.



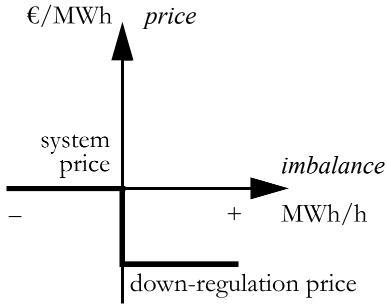
a) Single-price system during an up-regulation period.



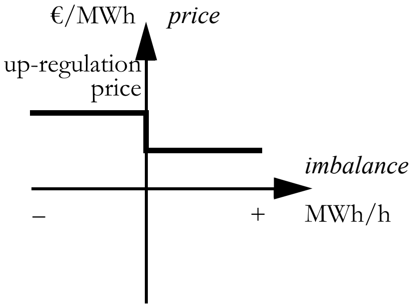
b) Single-price system during a down-regulation period.



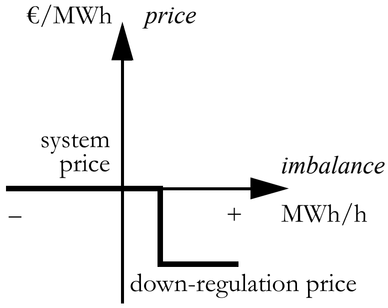
c) Two-price system during an up-regulation period.



d) Two-price system during a down-regulation period.



e) Mixed price system during an up-regulation period. The cost of positive imbalance is the mean value of the system price and the up-regulation price.



f) Two-price system with dead-band during a down-regulation period.

Figure 2.6 Different pricing schemes for the post market. Players have positive imbalance when supplying more energy during the trading period than they have extracted. The system price refers to the electricity price in the ahead market, whereas up- and down-regulation prices refer to the prices of the real-time balancing market (corresponding to the real-time price if central dispatch is used).

Example 2.1 (trading of imbalance power): Consider two balance responsibility players in an electricity market. Player A is both acting as a producer and as a retailer and must plan generation and trading at the power pool to cover the expected load of the consumers during the hour. Player B is a pure producer. Table 2.2 shows the plans of the two players for the trading period between 10 am and 11 am.

At 10:15 a power plant generating 200 MW belonging to player A fails. To at least partly compensate the lost generation A increases the generation in their hydro power plants by 180 MW at 10:30.

Player B submitted a sell bid to the power pool for a power plant having an operation cost of 24 €/MWh, but the price of the power pool became just 20 €/MWh; hence, the bid was not accepted. As player B has free capacity in a power plant that can be started quickly, they have submitted an up-regulation bid of at most 300 MW for 24 €/MWh. At 10:20 the system operator activates 240 MW of this bid and the power plant is started at 10:35.

When the system operator collects meter values of the trading period in question it turns out that the consumers of player A (for which the company is balance responsible) have consumed more than what the company had expected, more precisely 1 240 MWh instead of 1 200 MWh. Concerning player B, which owns a few wind farms, it turns out that they have produced 20 MWh more than planned.

No other up-regulation bids have been activated during the trading period; thus, the up-regulation price became 24 €/MWh.

a) What will the price of imbalance power become in a one-price system? Will the players earn or loose on their imbalances?

b) What will the price of imbalance power become in a two-price system? Will the players earn or loose on their imbalances?

Table 2.2 Planned generation and trading in example 2.1.

	Player A	Player B
Own generation	1 000	500
Purchase from power pool	200	
Sales to power pool		500
Sales to final consumers	1 200	

Solution: We start by calculating the imbalances of the players. Player A had planned to produce 1 000 MWh, but due to the failure the generation was reduced by 45/60 h · 200 MW = 150 MWh. This was slightly compensated by generating 180 MW more than planned for 30 minutes, i.e., an increase by 90 MWh. They have also purchased 200 MWh from the power pool, which gives a total input of 1 140 MWh. The companies load is however 1 240 MWh, which yields a negative imbalance of 100 MWh.

Player B should have produced 500 MWh according to the plan, but the wind power gave an extra contribution of 20 MWh. Moreover, they increased the generation by 25/60 h ·240 MW = 100 MWh. In total they have produced 620 MW, while they have sold 500 MWh to the power pool and 100 MWh up-regulation to the system operator. This results in a positive imbalance of 20 MWh.

a) In a one-price system all players trade imbalance power to the same price (in this case the up-regulation price 24 €/MWh). Player A has to buy 100 MWh, which costs 2 400 €. If they had known this in advance they could have bought 100 MWh at the power pool for 2 000 € (assuming that the quantity trades at the power pool is so large that an additional purchase would not affect the electricity price). Thus, the imbalance costs player A 400 €.

Player B has to sell 20 MWh, which results in an income of 480 €. If they had known the increased wind power generation in advance and sold it to the power pool they would however only gain 400 €. Thus, player B has earned 80 € from the imbalance.

b) In a two-price system the players buy imbalance power for the up-regulation price 24 €/MWh and sell for the electricity price of the power pool, i.e., 20 €/MWh. The cost of player A is therefore the same as in a one-price system, i.e., player A loses

400 € on the imbalance. For player B the income of the sold imbalance power is decreased to 400 €, which means that player B neither gains nor loses from the imbalance.

Finally, it can be mentioned that there are some other variants of pricing imbalance power. It might for example be desirable to compromise between a wish to motivate the balance responsible players to keep their balance and a wish to not disadvantage players having difficulties predicting or regulating production and consumption. Such a compromise is to use a two-price system, where those players having imbalance in "the right direction" will not receive a price as favourable as the corresponding regulating price, but still better than the price of the ahead market (for example a mixed price system as in figure 2.6e). It is also possible to refrain from punishing lesser imbalances by an unfavourable price (so-called dead-band; seefigure 2.6f).

EXERCISES

2.1 Which are the responsibilities of the system operator?

2.2 What does it mean to be balance responsible?

2.3 What are the responsibilities of a grid owner?

2.4 How is a vertically integrated electricity market organised?

2.5 How is a centralised electricity market organised?

2.6 How is a bilateral electricity market organised?

2.7 What does ahead trading refer to?

2.8 How is electricity traded in a power exchange?

2.9 What is a block bid?

2.10 What is a firm power contract?

2.11 What is a take-and-pay contract?

2.12 What does real-time trading refer to?

2.13 What does an up-regulation bid mean?

2.14 What does a down-regulation bid mean?

2.15 What does post trading refer to?

FURTHER READING

- S. Stoft, *Power System Economics*, IEEE Press, 2002. — *Basic textbook on electricity markets.*
- I. Wrangensteen, *Power System Economics – the Nordic Electricity Market*, Tapir Academic Press, Trondheim 2007. — *Basic textbook with focus on the Nordic electricity market.*
- "Balance Responsible Agreement", Svenska kraftnät. [Available at http://www.svk.se] — *The regulation for balance responsible players in Sweden.*
- "EPEX Spot Market Rules and Regulations", European Power Exchange, November 2011. [Available at http://www.epexspot.com] — *The regulation for the European Power Exchange.*
- "System Operation Agreement 2006-06-13", Nordel, 2006. [Available at https://www.entsoe.eu] — *The contract which regulates the cooperation between the Nordic system operators.*
- "Trade at NPS", Nord Pool Spot. [Available at http://www.npspot.com/trading/] — *Description of the Nordic power exchange Nord Pool.*
- "Trade at Nord Pool ASA's Financial Market", Nord Pool, March 2010. [Available at http://www.nasdaqomxcommodities.com/trading/] — *Description of the Nordic power exchange Nord Pool.*

Chapter 3
ELECTRICITY PRICING

This chapter describes a simple model for electricity pricing. The objective of the model is to provide a basic understanding on the factors which influence the electricity price and the relative importance of individual factors.

3.1 PRICING

As for any good which is traded in a free market environment, the price of electricity will be set by supply and demand. Assume that there is a number of producers[1] and it is known how much they can produce of a particular good. Given a number of producers, the price for which they are willing to sell a particular good and which production costs they have. It is then possible to arrange a supply curve, which will show the total production of the good for a given price. In the same way it is possible to create a demand curve, which shows the total consumption for a given price. Figure 3.1 shows examples of such curves.

The market price is set by the intersection of the supply and demand curves. It can be shown that this trading maximises the total surplus, i.e., this is the most efficient solution from the society's point of view. It should be noted that this is stable equilibrium point, because no player can gain anything from trying to deviate from the market price. Consider for example a producer which can not sell his or her production, because the producer is requiring a higher price than the market price. No consumer will be interested in buying from this producer, as they can buy the same good from some other producer to the market price instead. Admittedly, there are also some consumers who are not supplied at all in the current equilibrium, but they are on the other hand not even willing to pay the market price. The only thing that our example producer can do to get more customers is to reduce the price. This is only profitable if the marginal production cost, i.e., the variable manufacturing cost, is less than the market price.

If we instead consider a producer who is selling the good, this producer could increase his or her by demanding a higher price. However, this is only possible if the wanted price not is higher than the market price. If the producer is trying to exceed this limit then the consumers instead will turn to the least expensive producer having free production capacity. This does of course presuppose that there are a lot of producers in the market and that the marginal cost is not showing too large differences; when there will always be a producer having unused production capacity available wo a marginal cost which is only slightly higher than the market price.

1. We assume that all producers manufacture equivalent products.

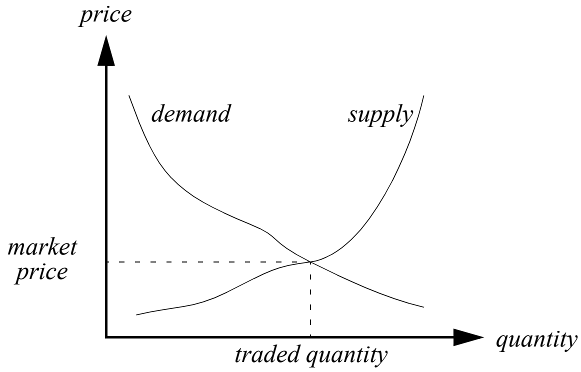


Figure 3.1 Supply and demand in a market.

Similar reasoning as for the producers can be applied to the consumers. From this we can conclude that if competition in the market is strong then all producers are offering their good to marginal costs, which means that the supply curve corresponds to a marginal cost curve. In the same way, the demand curve corresponds to a marginal value curve. We can also conclude that all producers and consumers will trade to the market price, even though some producers will have a marginal cost which is less than the market price.

3.2 A SIMPLE PRICE MODEL

A closer analysis of the price variations in an electricity market would require a more detailed simulation models, similar to those described in chapter 6. In this section we will present a more simplified model, which is appropriate for rough estimations or to obtain an idea of the influence different factors will have on the electricity price.

Assumptions

The idea behind the simple price model is to neglect the load variations during a day and only estimate the electricity price by studying supply and demand *on an annual basis*. To achieve this, we need to make a few simplifying assumptions. (Later in this section we will study what happens if some of the assumptions are skipped.)

- **Perfect competition.** In order to have perfect competition in a market, a number of constraints have to be fulfilled; the players must for example be rational, they must be free to trade with each other and there should be no market power. If these requirements are fulfilled, the price will be decided according to the principles described in section 3.1.
- **Perfect information.** When we refer to perfect information we mean that all players have exact knowledge of all parameters of importance for their decisions.
- **No capacity limitations.** This means that those power plants which are needed to supply enough energy per year also have enough installed capacity to cover the load during every instant of the year.
- **No transmission limitations.** It is not sufficient that there is enough capacity available in a system; it must also be possible to distribute the power to the final consumers. The assumption of no transmission limitations means that this always will be possible.
- **No reservoir limitations.** It must be possible to freely use the available hydro

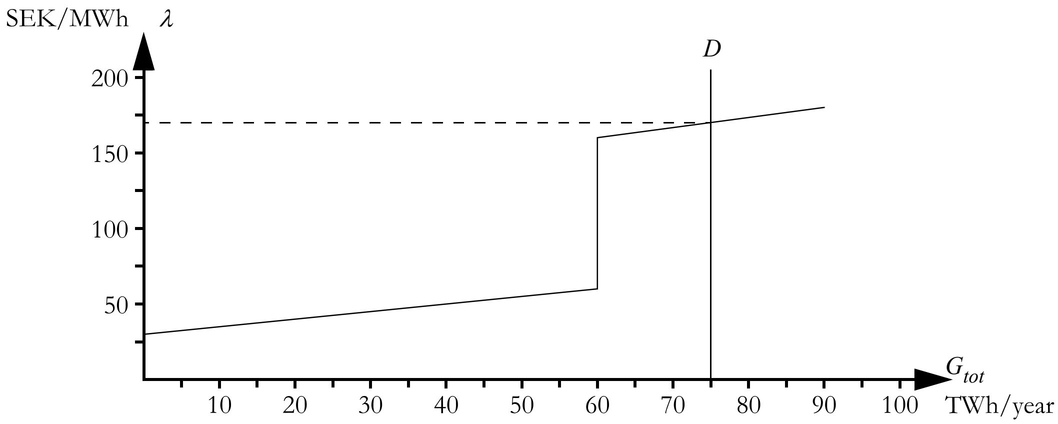


Figure 3.2 Supply and demand of the electricity market in example 3.1.

power energy during the year. To make this possible there must not be any reservoir limitations, which prevents us from storing enough hydro energy from one part of the year to another.

- **The load is not price sensitive.** If the load is not price elastic then the demand curve is corresponding to a vertical line at the quantity corresponding to the annual load. On most markets it is reasonable to simplify the demand curve in this way, because most consumers have take-and-pay contracts[2] where they pay a fixed price per kWh. These consumers have no reason to change their electricity consumption even if the electricity price is increasing.

The supply curve can be compiled by combining the variable operation cost of the power plants and the potential energy generation in each power source. The average electricity price during a year (or any other longer time period) can then be estimated by determining the intersection of the supply and demand, according to the pricing theory described above.

Example 3.1 (price in a simple electricity market): Assume that we have a power system supplied by hydro power and coal condensing. The potential hydro power generation is limited by the inflow, which is 60 TWh/year. The potential of the coal condensing units is 30 TWh/year. The load is 75 TWh/year. The variable operation costs are assumed to be 30-60 SEK/MWh for hydro power and 160-180 SEK/MWh for coal condensing. The operation costs are assumed to be linear within the stated intervals, i.e., if the generation is equal to zero then the price is equal to the lower value, and for maximal generation the price is equal to the higher value. Assume perfect competition, perfect information and that there are neither any capacity, transmission nor reservoir limitations. Which electricity price will we have in this system? Also calculate the difference between variable operation cost and total income of the power producers.

Solution: The supply curve of this system is shown in figure 3.2. If the annual consumption would have been less than 60 TWh then the system could have been supplied solely by hydro power. The assumption of not having any capacity, transmission or reservoir limits makes it possible to distribute the available hydro energy during the year so that the whole load can be covered. If the price would be different during different parts of the year, the hydro power producers could earn money by saving water in the reservoirs from low price periods to high price periods. This would result in

2.See section 2.2.1.

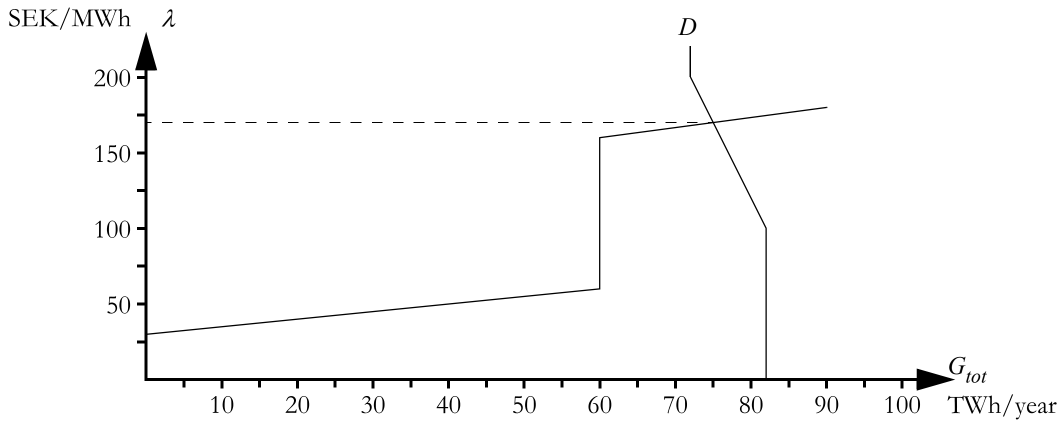


Figure 3.3 Supply and demand of the electricity market in example 3.2.

increasing prices during the lower price periods and decreasing prices in the high price periods, until the price has been levelled out between the different parts of the year.

For a load which is larger than 60 TWh some of the coal condensing units will be needed. When the load is 75 TWh all coal condensing up to the marginal cost 170 SEK/MWh. The assumption of perfect information means that the hydro power owners are aware of this and will demand the price170 SEK/MWh for their electricity generation too. The assumption of perfect competition means that nobody will be able to sell power to a price higher than 170 SEK/MWh, because then they will loose their sales to somebody who offers to sell for 170 SEK/MWh.

The total variable operation cost in this example is

$$PC = 60 \cdot \frac{(30+60)}{2} + 15 \cdot \frac{(160+170)}{2} = 5\ 175 \text{ MSEK},$$

while the income is

$$PV = 75 \cdot 170 = 12\ 750 \text{ MSEK},$$

i.e., the total producers' surplus is 12 750 – 5 175 = 7 575 MSEK. It should however be noted that this surplus is not pure profit, but should also cover fixed costs as capital cost and labour costs.

Price Sensitive Load

The simple price model does not change much if we release the assumption that the load is not price sensitive; we will still look for the intersection of the supply and demand curves.

Example 3.2 (price sensitive load). Consider the same electricity market as in example 3.1, but assume that part of the load is not price sensitive and a part of it that is price sensitive. The price insensitive load is 72 TWh/year, while the price sensitive load is 10 TWh if the price does not exceed 100 SEK/MWh and will then decrease linearly until the price reaches 200 SEK/MWh; at this price level the elastic consumption has decreased to zero. Assume perfect competition, perfect information and that there are neither any capacity, transmission nor reservoir limitations. Which electricity price will we have in this system? How large will the total consumption become?

Solution: The supply and demand curves of this system are shown in figure 3.3. The hydro power is not sufficient to cover the firm load; hence, a certain amount of coal condensing must be used. As this production is more expensive than 100 SEK/MWh some price sensitive consumers will decrease their consumption. Assume that the price is λ. The contribution from the coal condensing can be written as

$$30 \cdot \frac{\lambda - 160}{180 - 160}.$$

When the price exceeds 100 SEK/MWh the price sensitive load decreases by 0,1 TWh for each SEK/MWh of price increase. The total load is therefore $82 - 0{,}1(\lambda - 100)$. If production equals consumption then we must have

$$60 + 30 \cdot \frac{\lambda - 160}{180 - 160} = 82 - 0{,}1(\lambda - 100).$$

The solution to this equation yields the electricity price $\lambda = 170$. At this price the price sensitive consumption is reduced by 7 TWh, which means that the total consumption is 75 TWh.

Impact of Reservoir Limitations

In reality the inflow to the hydro reservoirs is not evenly distributed over the year. In the Nordic countries most of the inflow is during the spring flow (April to June) whereas the electricity consumption has its peak during the winter. It is therefore necessary to store large amounts of water from the spring and summer to the winter. If the inflow is too large compared to the storage capacity of the reservoirs it becomes necessary to increase the hydro power generation to avoid spillage. This might result in capacity limitations during those periods when the hydro power must increase its generation or that not enough water is stored in the reservoir when the peak load period begins.

In the price model the impact of reservoir limitations can be studied by dividing the year in separate periods which are treated independently:

Example 3.3 (reservoir limitation). Assume that the reservoirs in the electricity market from example 3.1 have a storage capacity of 18 TWh. During the first half of the year the load is 35 TWh and the inflow is 50 TWh, which means that the load and inflow during the second half is 40 TWh and 10 TWh respectively. Moreover, assume that the reservoirs are empty at the beginning of the year and completely filled

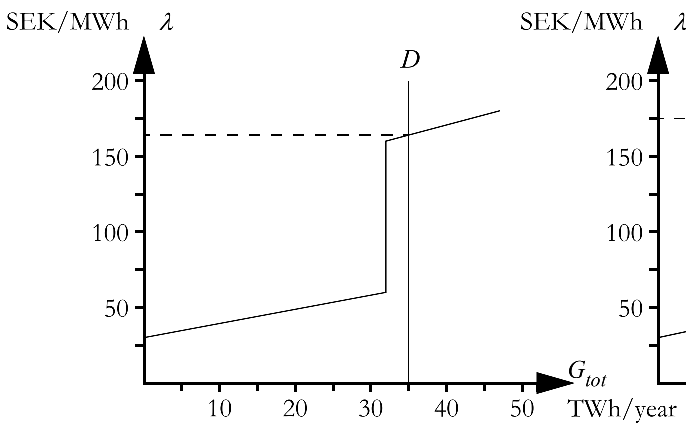


Figure 3.4 Supply and demand during the first six months.

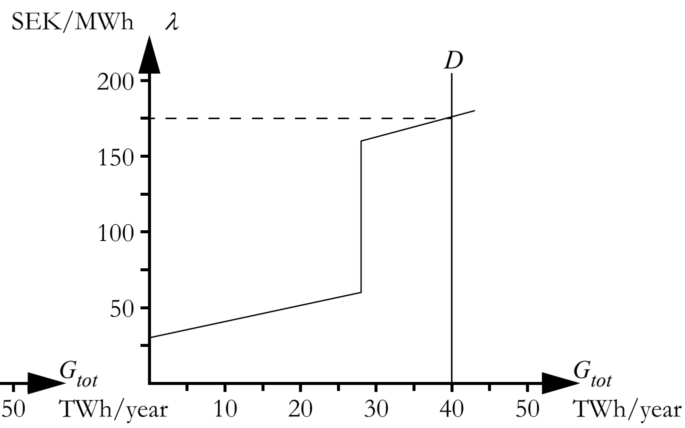


Figure 3.5 Supply and demand during the last six months.

after the first half of the year. Determine the electricity price during the year. The coal condensing potential can be assumed to be evenly distributed over the year.

Solution: During the first half of the year it is possible to generate 32 TWh of hydro power, i.e., the inflow during the period minus the water which is stored in the reservoir after six months. Further, it is possible to use half of the coal condensing potential. The resulting marginal cost curve is shown in figure 3.4. The electricity price becomes 164 SEK/MWh.

During the second half of the year only 18 + 10 = 28 TWh hydro energy is available. Together with half of the coal condensing potential we get the marginal cost curve shown in figure 3.5. The electricity price during the second half of the year is 176 SEK/MWh.

Notice that the electricity price was higher in the period after the reservoirs were filled. If the price would have been lower during the second half of the year then it would have been profitable to store less water for that part of the year and use it to lower the price during the first six months instead. This way it would have been possible to "move" water from the second half of the year to the first until the price difference had been levelled out.

Impact of Forecast Uncertainties

In reality it is impossible to have perfect information about all factors affecting the operation of the power system. This is primarily a problem in power system with a large share of dispatchable hydro power, because long-term hydro power planning require knowledge about inflow, availability in thermal power plants and load in order to utilise the hydro energy in an optimal manner. All these quantities must however be estimated by forecasts.

One of the most uncertain forecasts is the annual inflow. If it is believed that the inflow will be high then the need for thermal power will decrease and the marginal cost will be relatively low. If it on the other hand is believed that the inflow will be low then higher marginal costs will be expected. This principle is illustrated in the following example:

Example 3.4 (uncertain inflow forecasts): Assume that an electricity market has a potential of 30 TWh/year in the coal condensing units. There is also dispatchable hydro power in the system, but the information about the inflow for the coming twelve months are changing according to figure 3.6, i.e., at the beginning of the year the inflow is expected to be 60 TWh, but due to increased precipitation during the

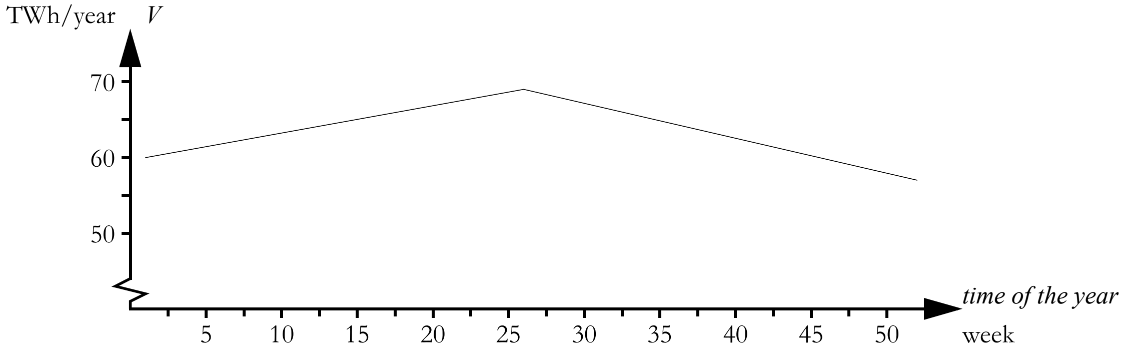


Figure 3.6 Forecasted hydro power potential for the next twelve months in example 3.4.

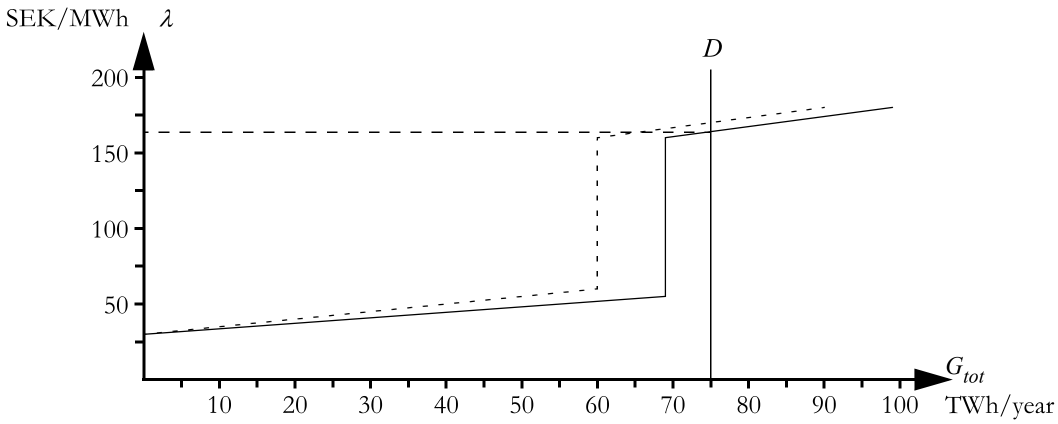


Figure 3.7 Supply and demand after six months in example 3.4.

spring the expectations are increased to 69 TWh after six months. In the autumn, the inflow is however far less than expected, which causes the expectations to decrease continuously to 57 TWh at the end of the year. The load is 75 TWh/year. The operation costs are assumed to be linear within the stated intervals, i.e., if the generation is equal to zero then the price is equal to the lower value, and for maximal generation the price is equal to the higher value. Assume perfect competition, perfect information about everything except the inflow and that there are neither any capacity, transmission nor reservoir limitations. How will the electricity price develop during the studied year?

Solution: The conditions in the first week of the year are exactly the same as in example 3.1; hence, the electricity price is the same, i.e., 170 SEK/MWh. It can be noticed that this conclusion is based on the assumption that all players have the same forecast about the inflow.

After six months the players of the market believe that the inflow will be 69 TWh the next twelve months. The supply curve in this case is shown in figure 3.7. The result from example 3.1is also shown in the figure by a dotted line. The difference is that the increased amount of hydro energy has shifted the coal condensing part of the supply curve rightwards. The same reasoning as in example 3.1 yields that the electricity price is 164 SEK/MWh.

As the marginal cost of the coal condensing units is a linear function the electricity price λ will be a linear function of the expected hydro power generation the next twelve months:

$$\lambda = 160 + \frac{180 - 160}{30}(75 - V) = 210 - \frac{2}{3}V \text{ SEK/MWh.}$$

This function is of course only valid when the electricity consumption is 75 TWh/year and the marginal cost is set by the coal condensing units, i.e., when $45 \leq V \leq 75$. Using this relation and the given the expected hydro power potential according to figure 3.6 the electricity price must vary as shown in figure 3.8.

Market Power

Market power arises when one or more players have such a large market share that they can affect

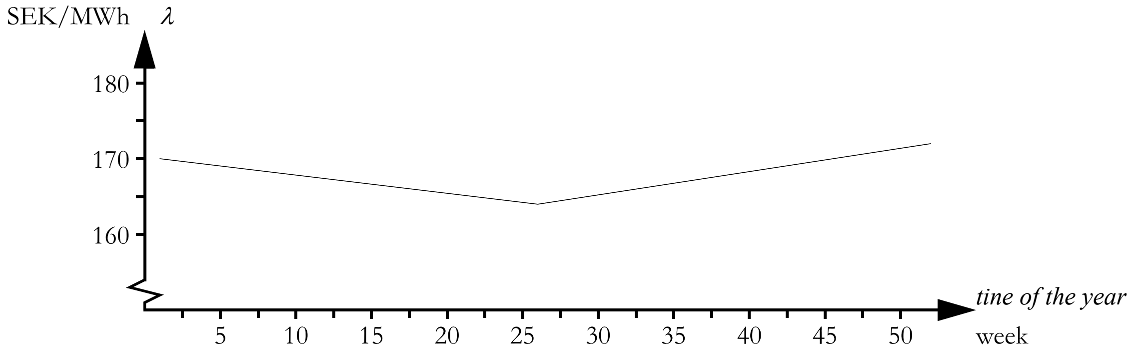
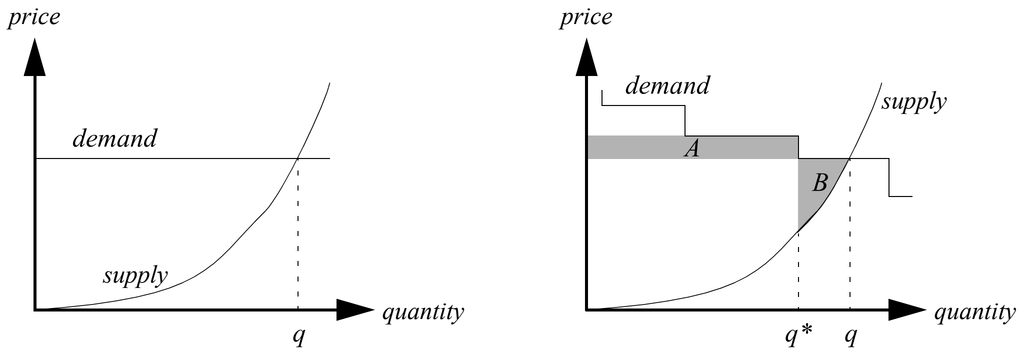


Figure 3.8 Development of the electricity price during the year in example 3.4.

the market price. A producer having market power can increase his or her profit by decreasing the electricity generation. They can also force competitors to leave the market, by selling power to less than marginal cost and by that means lower the electricity price to a level where the competitor is no longer profitable. In other words, the producer is willing to accept a limited loss of profits in order to get rid of a competitor and increase the profits in a long-term perspective.

A player with market power is often referred to as a *price setter*, unlike other players which are *price takers*. The difference between a price setter and a price taker is illustrated in figure 3.9. A price taker has such a small production capacity compared to the total turn-over of the market that the actions of the producer do not affect the demand, as shown in figure 3.9a. This producer will maximise his or her profit by producing as much us possible as long as the marginal cost (i.e., the supply curve) is less than the willingness to pay of the consumers (i.e., the demand curve). If all producers as producing so much that their marginal costs are equal to the marginal benefit, then the market has reached the equilibrium point which maximised the total surplus.

However, if the producer is a price setter the situation is different. A price setter has such a large share of the market that his or her actions affect the price. In figure 3.9b we can see that if the producer decreases the production from q to q^* then the profits are decreased by an amount corresponding to the shaded area B, but on the other hand the price is increasing, which means that the income of the remaining sales increase by an amount equal to the shaded area A. If A is larger than B then it would be profitable to decrease the production, even though the producer's marginal cost is less than the market price. It must be noticed that the reduced production only is profitable



a) Company specific supply and demand of a price taker.

b) Company specific supply and demand of a price setter.

Figure 3.9 Illustration of the difference between a price taker and a price setter.

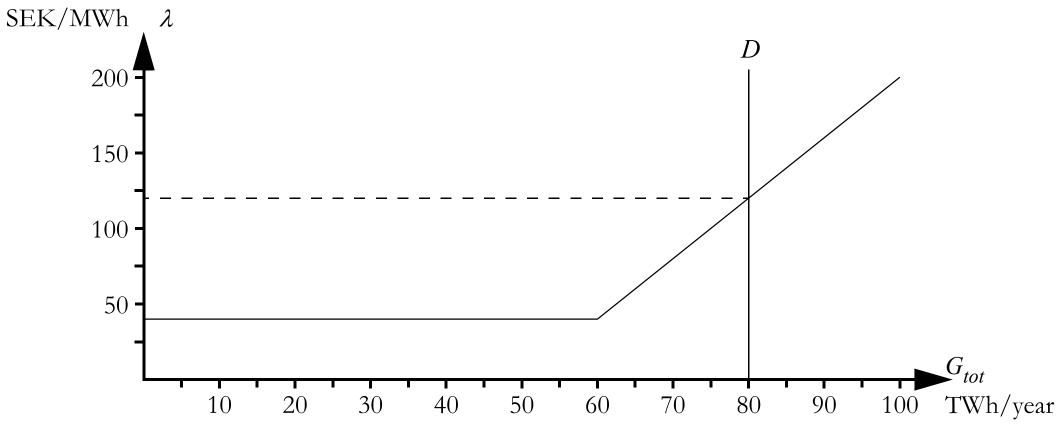


Figure 3.10 Supply and demand of the electricity market in example 3.5.

for the price setter, but not for the society as a whole!

Since market power is not good for the total surplus, the authorities usually try to limit market power by legislation. This is however in many cases difficult, as it can be hard to prove that a company is misusing its market power.

> ***Example 3.5 (one player with market power):*** Assume that there is an electricity market with one large producer, AB Kraftjätten, which can generate up to 60 TWh/year at the cost 40 SEK/MWh. The remaining producers all have very small capacity compared to the turn-over of the market. Their marginal costs vary between 40 SEK/MWh and 200 SEK/MWh, as shown in figure 3.10.
>
> During a certain year the demand is 80 TWh and we can assume that the load is not price sensitive. Assume perfect information and that there are neither any capacity, transmission nor reservoir limitations. Which electricity price would we have if there is perfect competition? What would the electricity price be if AB Kraftjätten chooses to utilise its position as a price setter?
>
> ***Solution:*** In a perfect market the price would be set by the intersection of supply and demand, i.e., 120 SEK/MWh. At this point Kraftjätten AB would generate its full potential. By reducing the generation the company can however increase the electric-

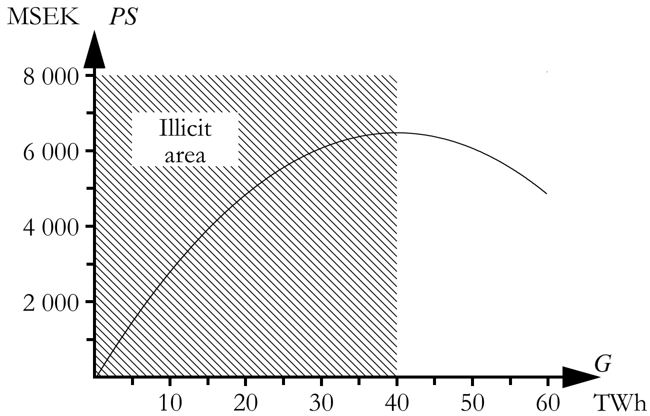


Figure 3.11 The surplus of AB Kraftjätten as a function of chosen annual generation.

ity price. For each TWh of reduced generation the electricity price will increase by 4 SEK/MWh, i.e., we can express the electricity price as

$$\lambda(G) = 120 - 4(G - 60) = 360 - 4G,$$

where G is the annual generation of AB Kraftjätten. This relation is only valid when $40 \leq G \leq 60$; if the generation is lower than 40 TWh then there would be an energy deficit, which hardly would be tolerated by the system operator. In the allowed interval the surplus of AB Kraftjätten (i.e., income minus costs) by

$$PS(G) = G \cdot \lambda(G) - 40G = 320G - 4G^2.$$

This function is shown in figure 3.11. As can be seen, the company is maximising its surplus by decreasing the annual generation to 40 TWh. Such a large reduction may however draw attention from the authorities, which for example could force the company to be split in several smaller companies to increase competition. In the long run it might therefore be profitable for AB Kraftjätten to just reduce the generation a little, which will not give the same profits, but will allow the company to keep its position as a price setter.

EXERCISES

3.1 Determine the electricity price in a small electricity market, where the annual consumption is 72.5 TWh/year. The power is generated in nuclear power plants having a variable operation cost between 100 and 120 ¤/MWh (annual potential 40 TWh) and combined heat and power plants with a variable operation cost between 80 and 150 ¤/MWh (annual potential 35 TWh). The operation costs are assumed to be linear within the stated intervals, i.e., if the generation is equal to zero then the price is equal to the lower value, and for maximal generation the price is equal to the higher value. Assume perfect competition, perfect information and that there are neither any capacity nor transmission limitations.

3.2 What will happen to the electricity price of the market in the previous exercise if 10 TWh of the nuclear generation is replaced by 10 TWh wind power?

3.3 Figure 3.12 shows supply and demand for a certain electricity market.

a) What will the electricity price become in this electricity market if we assume perfect competition, perfect information and that there are neither transmission-, reservoir nor capacity limitations?

b) Assume that a certain company owns all hydro power plants in this electricity market and have fixed costs of 100 M¤/year. How large is the profit of the company?

3.4 Consider a simplified model of the electricity market in Land. Data for the power plants are given in table 3.1. The operation costs are assumed to be linear within the stated intervals, i.e., if the generation is equal to zero then the price is equal to the lower value, and for maximal generation the price is equal to the higher value.

a) Assume that the electricity market in Land has perfect competition, all players have perfect information, and there are neither transmission nor capacity limitations. How large is the electricity consumption in Land if the electricity price during a certain year is 220 ¤/MWh?

b) AB Vattenkraft owns hydro power plants with a combined production capacity of 10 TWh/year. The fixed costs of the company are 650 M¤/year. How large are the

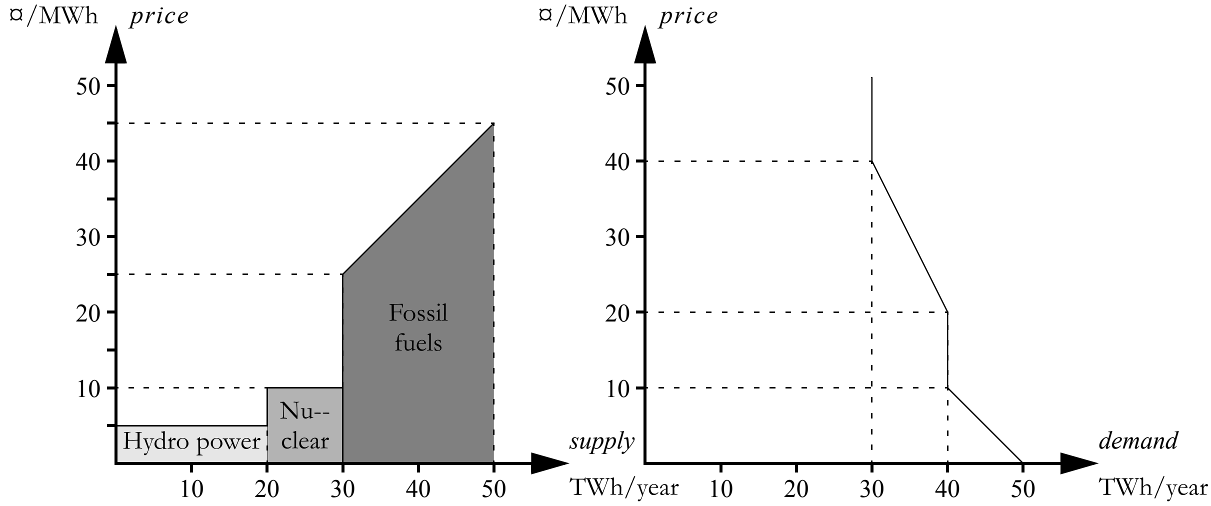


Figure 3.12 Supply and demand in exercise 3.3.

Table 3.1 Data for the power plants in Land.

Power source	Production capability [TWh/year]	Variable costs [¤/MWh]
Hydro power	55	5
Nuclear power	50	100–120
Bio fuel	16	100–260
Fossil fuel	20	190–390

profits of the company if the electricity price during a certain year is 220 ¤/MWh?

3.5 AB Skog is usually a large consumer of electric energy. If the electricity price is 200 SEK/MWh they consume 2 TWh/year. If the electricity price is higher than that, the electricity consumption of the company decreases, because the cost of their forest products increase (hence, they sell less) and it also becomes profitable to utilise rest products from the industrial processes of the company to generate electricity. If the electricity price is 400 SEK/MWh then the net consumption of AB Skog equals zero, i.e., the internal generation of the company equals the consumption. The relation between electricity price and the net consumption of AB Skog is assumed to be linear in the interval 200–400 SEK/MWh. If the electricity price increases above 400 SEK/MWh then the internal generation of AB Skogs will exceed the consumption. At the electricity price 500 SEK/TWh the company will generate a net surplus of 1 TWh/year. The relation between electricity price and net surplus is assumed to be linear in the interval 400-500 SEK/MWh.

The other consumers in the electricity market consume 100 TWh/year and are not price sensitive. The other producers in the electricity market can generate 50 TWh hydro power for 10 SEK/MWh, 40 TWh nuclear power for the price 100 SEK/MWh, 8 TWh from combined heat and power plants for the price 240 SEK/MWh, 3 TWh from fossil gas fuelled power plants for the price 320 SEK/MWh, 3 TWh from oil-fired power plants for the price 420 SEK/MWh and 1 TWh in gas turbines for the price 600 SEK/MWh.

What will the electricity price become in this electricity market if we assume perfect competition, perfect information and that there are neither transmission-, reservoir nor capacity limitations?

3.6 Table 3.2 shows the available generation resources, marginal generation costs and annual consumption in the Nordic countries during year 2000. The operation costs are assumed to be linear within the stated intervals, i.e., if the generation is equal to zero then the price is equal to the lower value, and for maximal generation the price is equal to the higher value. Assume perfect competition, perfect information and that there are neither any capacity, transmission nor reservoir limitations. Finland imports 4 TWh from Russia, Denmark exports 5 TWh to Germany, Sweden exports 0.5 TWh to Germany as well as 0.5 TWh to Poland.

Table 3.2 Available generation resources and annual consumption.

Power source	Production capability [TWh/year]				Cost [SEK/MWh]
	Sweden	Norway	Finland	Denmark	
Hydro power	78	142	14	-	40
Wind power	0.5	0	0	4	20
Nuclear power	55	-	21.5	-	50–75
Industrial backpressure	5	-	13	2	40–100
Comb. heat and power	5	-	13	24	60–140
Coal condensing	-	-	13	24	120–140
Consumption	146	124	79	35	

a) Estimate the electricity price.

b) Estimate the trading between the Nordic countries.

c) How would the electricity price be affected if there is a reservoir limitation so that the reservoirs are filled on July 31? Everything else is assumed to be the same as above. No numerical values have to be calculated; just describe the trend.

3.7 Consider the Nordic electricity market model in exercise 3.6. Assume that all players—except one—are small enough to be considered as price-takers. Investigate whether or not it is profitable to exercise market power for the large company in the following two cases:

a) Assume that one company, Voimajättiäs OY., owns all Finnish hydro power plants and nuclear power plants.

b) Assume that one company, Kraftjätten AB, owns all Swedish hydro power plants and nuclear power plants.

3.8 An external cost is a cost which is due to the production or consumption of a good, which is neither paid by the producer nor the consumer, but paid by a third party instead. Emission of dangerous pollutants from combustion of fossil fuels is an example of an external cost in an electricity market.

The maximal annual production of the power plants in Land are shown in table 3.3 below. The variable production costs are assumed to be linear in the intervals, i.e., the production is zero if the price is on the lower price level and the production is maximal at the higher price level. The annual consumption in Land is 100 TWh.

a) Assume perfect competition, perfect information and that there are neither capacity, transmission nor reservoir limitations. Moreover, assume that the cost of the emissions is an external cost. Estimate the electricity price in Land.

Table 3.3 Potential for electricity production in Land.

Power source	Production capacity [TWh/year]	Variable costs [¤/MWh]	Harmful pollutants [ton/MWh]
Wind power	10	1	0
Hydro power	60	10	0
Coal condensing	30	200–320	1
Fossil gas	20	230–390	0.4

b) The benefit to the society of the electricity generation can be measured by the total surplus, which is obtained by summation of

- the profit of the consumers, i.e., the value of the consumption (assumed to be 500 ¤/MWh) minus the purchase cost
- the profit of the producers, i.e., the value of sold electricity minus the production cost
- the damage to the environment, which is assumed to be 50 ¤/ton pollutants.

Determine the total surplus of the electricity market in Land.

c) Suppose that Land introduces emission rights trading and that the price of the emission rights becomes 40 ¤/ton, i.e., for each ton harmful pollutants emitted, the producer must pay 40 ¤ for the emission right. Which electricity price will they get in Land? How large is the total surplus going to be? Assume that the income Land receives from selling emission rights is used to counteract the damage to the environment caused by the emissions (i.e., the net damage of each ton pollutants is 10 ¤).

FURTHER READING

- M. L. Katz & H. S. Rosen, *Microeconomics*, third edition, Irwin/McGraw-Hill, 1998. — *Good basic text book in microeconomics.*
- L. Söder, "Analysis of Electricity Markets", compendium, Electric Power Systems Lab, KTH, Stockholm 2007. — *Deeper analysis of the models presented in this chapter.*

Chapter 4
FREQUENCY CONTROL

A very important property of electric energy is that it can not be stored. A battery for example stores chemical energy (which readily can be converted to electric energy). Due to this property all electric power systems must at all moments have balance between generation of electric energy and consumption. To keep this balance it is necessary to have automatic control systems which can respond swiftly (within seconds). This function is called primary control. Those familiar with control theory recognises the primary control as the proportional part of a PI-regulator. Furthermore, it is necessary to have additional controls which restores the balance in a slightly longer time perspective (minutes); this corresponds to the integrating part of an PI-regulator. This control is referred to as secondary control and can either be operated automatically or manually.

4.1 PRIMARY CONTROL

The primary control considers the capability of the system to meet sudden load and/or generation changes. The primary control is separate for each *synchronous* power system A synchronous power system is a power system where all producers and consumers are connected to each other through transformers and AC transmission and distribution lines. Anything from a diesel generator set supplying a single load to a multi-national grid as the Nordel system (which connect Norway, Sweden, Finland and the eastern part of Denmark) can constitute a synchronous grid.

An AC line has to have the same electric frequency at both ends of the line. If there were different frequencies at the ends then the voltage angle shift would increase until it reaches 180°, resulting in unacceptable large currents on the line. The same is valid for transformers. The conclusion is that in a synchronous grid the average electric frequency must be the same. For short periods the frequency can be slightly different, but that will cause large power oscillations in the grid, which can result in serious disturbances.

The Function of the Primary Control

In a synchronous grid there has to be some kind of primary control. A short description of this type of control is given here:

- Assume that a power system is in balance, i.e., electricity generation and consumption are equal.
- At a certain time the consumption increases, while the generation remains the same.
- There is rotational energy stored in all synchronous machines connected to the sys-

tem, more exactly in the rotors and the connected turbine shafts. Since the generation and consumption always have to be in balance the load increase is compensated using the rotational energy, which causes the rotor speed to decrease in the synchronous machines.

- In synchronous machines there is a strong connection between rotor speed and electric frequency. This means that all synchronous machines in a synchronous grid in principle rotates at the same speed.[1] The reduced rotor speed therefore results in a frequency decrease in the grid.
- In some power plants there is frequency sensitive equipment, which forces the generation to increase when the frequency decreases. These power plants are referred to as primary controlled units.
- As long as the frequency keeps falling the primary controlled units will increase their generation. Finally, the balance between generation and consumption will be restored and the frequency will be stable again. It should be noted that the new stable frequency will be less than the original frequency!

In the description above we assumed that it was a load increase that triggered the primary control. A corresponding course of events will come up if there is a failure in a generating unit—if a generating unit is stopped and the load is constant then the remaining units in the system will have to increase their generation. A load decrease and/or a generation increase also causes similar courses of events. The only difference is that the frequency is increasing, which results in generation decreases in some power plants.

Permissible Frequency Range

In each power system there is a nominal frequency— North and South America usually 60 Hz, whereas most of the remaining world uses 50 Hz—and it is not possible to allow too large deviations from the nominal frequency, because there then is a risk that important components in the power system are damaged. If for example the electric frequency would coincide with the harmonic oscillation frequency of turbine blades or shafts, then these power plants have to be shut down to avoid expensive repairs. Another problem is that generators and transformers are designed for the nominal frequency and a deviating frequency can cause heating of the windings, resulting in damages. Finally, there might be certain kinds of loads that do not work satisfactory if the frequency deviates too much from the nominal.

The exact requirements on the frequency varies from power system to power system. As an example we may consider the Swedish regulation (see table 4.1). If the frequency deviation is at most 0.1 Hz then the frequency control is maintained by the primary control. For larger deviations the primary control reserves are exhausted and other countermeasures must be taken. In the first place international export on HVDC links is reduced and electric boilers and heat pumps are turned off.[2] If these actions are not sufficient then it is necessary to disconnect load. Everyone who is connected to the main grid of Svenska Kraftnät south of latitude 61° N are forced to install equipment which can automatically disconnect a part of the load (excluding electric boilers and heat pumps). This disconnection is activated in five steps, approximately equal in size. When all steps have been activated at least 30% of the load should be disconnected.As a last resort Svenska

1. It can be added that almost all larger power plants (larger than 3 MW) use synchronous machines. For small-scale generation (less than 1 MW) induction machines are often used, and such do not have that same strong connection between frequency and rotor speed.
2. I similar action in other power systems would be to stop any pumping in hydro power plants that can be run in "reverse" (such that water is pumped to the reservoirs during periods of low electricity prices and then later used for electricity generation when the electricity price is higher). This kind of hydro power is however not used in Sweden.

Kraftnät can order manual load shedding. The manual load shedding should be carried out within 15 minutes and include about 50% of the consumption, divided into five approximately equally large steps. Manual load shedding is often performed as so-called rotating load curtailment, which means that one or more distribution network is disconnected for a while. The distribution networks are then connected again at the same time as one or more other distribution networks are disconnected. The objective of this procedure is to shed the load as fair as possible. However, it should be noted that manual load shedding has never been used in Sweden so far. On the other hand, load shedding might be very common in countries with a rapid load growth, as it might not be possible to build new power plants fast enough.

Table 4.1 Frequency control in the Swedish power system.

Frequency interval [Hz]	Actions
49.9 - 50.1	Primary control.
49.0 - 49.8	Control changes on HVDC-links.
49.0 - 49.4	Automatic disconnection of electric boilers and heat pumps.
48.0 - 48.8	Automatic load shedding.
< 48.0	Manual load shedding (rotating load curtailment).

Gain

As explained above there are some power plants which automatically adjust their generation whenever the frequency of the system changes. The gain is a measure of how large the change in generation is for a certain change in frequency

Definition 4.1. The gain, R, indicates how the generation in a power plant is changed when the frequency changes. It is measured in MW/Hz. If G_0 is the generation at nominal frequency, f_0 = 50 Hz, it holds that $G = G_0 - R(f - f_0)$.

The notion of gain is illustrated in figure 4.1.

In practice it is often inconvenient to use the formula given in definition 4.1. It is easier to just calculate how the changes in generation and frequency relates to each other:

$$\Delta G = R \cdot \Delta f. \tag{4.1}$$

Then it is simple to conclude whether a certain change corresponds to an increase or a decrease; there must be a shortage of generation if the frequency is falling and therefore the primary control must result in an increase of generation, etc.

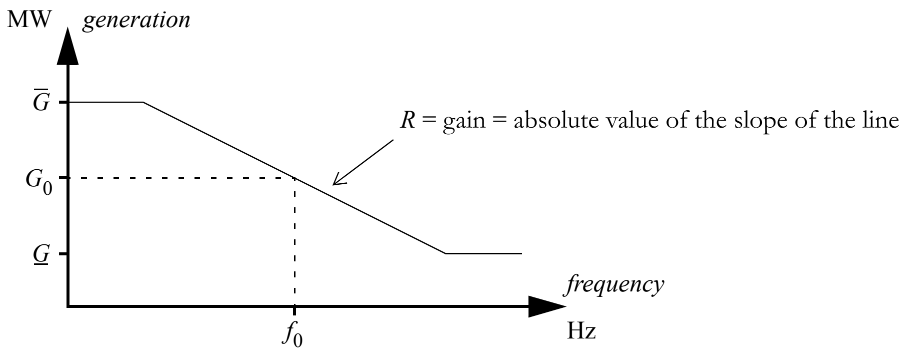


Figure 4.1 The relation between frequency and generation in a power plant participating in the primary control.

When calculating frequency changes it is necessary to also consider the reserves available in the primary controlled power plants; it is impossible to increase the generation of a power plant operating at maximal capacity regardless of the gain.

Example 4.1 (limitations in primary control): There are two primary controlled power plants in a certain power system. The installed capacities of the two units are 500 MW and 100 MW respectively. The gain is set to 300 MW/Hz in both units. The power plants are producing 420 MW and 75 MW respectively at an occasion when the frequency is exactly 50 Hz. A fault suddenly occurs in one of the other power plants in the system, which means that 85 MW of generation is lost. Which frequency will there be in the system when the primary control has restored the balance between generation and consumption?

Solution: Since both units have the same gain, a certain frequency change will result in equally large changes of generation in the two units. The frequency change caused by a loss of 85 MW of generation would thus make both units increase their generation by the same amount, i.e., by 42,5 MW each. However, this is not possible, because the smaller unit can only increase its generation by 25 MW. The remaining 60 MW must be generated in the larger unit. This requires that the frequency changes by

$$\Delta f = \frac{\Delta G}{R} = \frac{60}{300} = 0.2 \text{ Hz}.$$

Hence, the frequency must decrease to 49.8 Hz before the increased primary controlled generation matches the lost generation due to the outage.

Impact on the Transmission System

A consequence of the function of the primary control is that the generation change in a certain power plant only depends on the frequency change in the system; therefore, it is in principle independent of where the cause of the frequency deviation is located. If for example a nuclear power plant in southern Sweden must be stopped then the frequency in the whole Nordic power system will decrease. This results in generation increases in the primary controlled units in not only Sweden but also in Norway, Finland and Denmark. Consequently, the power flows on the transmission lines will be changed as generation is transferred from the nuclear power plant to other power plants in Sweden, Norway, Finland and Denmark. We can conclude that it is necessary to distribute the gain so that the transmission network can manage to transfer the power flows induced by the primary control.

The following example shows how bad things can turn out if the gain is not properly distributed:

Example 4.2 (inappropriate distribution of the gain). The three areas A, B, C are operated synchronously via the transmission lines AB, BC and AC. The total gain of the three systems R_A = 900 MW/Hz, R_B = 50 MW/Hz and R_C = 50 MW/Hz respectively. The transmission lines are equipped with protection systems which after a short time delay disconnects the lines if the power flow exceeds the capacity of each lines, which is $\bar{P}_{AB}$ = 300 MW, $\bar{P}_{BC}$ = 120 MW and $\bar{P}_{AC}$ = 150 MW respectively.

At a certain occasion the frequency is exactly 50 Hz and the flows on the transmission lines are as shown in figure 4.2.A fault occurs in a 240 MW power plant in area B and the generation is immediately stopped. Calculate the resulting frequencies and power flows when the system is stabilised again.

Solution: When 240 MW generation is lost the generation in the primary controlled

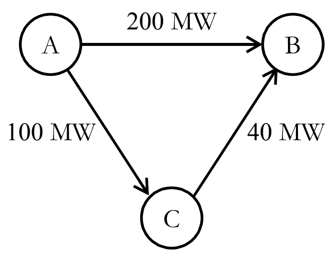


Figure 4.2 Transmission before the generation is reduced in area B.

power plants must increase by that amount. The increase is distributed between the areas according to their share of the total gain, R_{ABC}, which equals 1 000 MW/Hz:

$$\Delta G_A = \frac{R_A}{R_{ABC}} \cdot 240 \text{ MW} = 216 \text{ MW},$$

$$\Delta G_B = \frac{R_B}{R_{ABC}} \cdot 240 \text{ MW} = 12 \text{ MW},$$

$$\Delta G_C = \frac{R_C}{R_{ABC}} \cdot 240 \text{ MW} = 12 \text{ MW}.$$

Since the load in area A remains unchanged the whole generation increase must be exported. There is however only 150 MW free capacity on the connections AB and AC. One of the lines and then the other will be overloaded and disconnected regardless of how the generation increase in area A is divided between the two lines. Thus, there will be two separate systems: A and B + C respectively.

In system A the load and the generation not involved in the primary control is the same as before the fault in area B. The earlier export of 300 MW has however been lost, which implies that the primary controlled generation must decrease correspondingly. Thus, we will have a frequency increase $\Delta f_A = 300/900$ Hz = 0.33 Hz, which yields $f_A = 50.33$ Hz.

In the system B + C there is a deficit of 240 MW due to the failure. Moreover, the import from area A has been lost, which gives us a total deficit of 540 MW. As both areas have equally large gain, the generation will increase by 270 MW in both areas. In area C, we have an generation increase of 270 MW and a reduced import of 100 MW. There are no changes in load and the generation in those power plants which are not participating in the primary control; hence, the export to area B must be increased by 170 MW. Thus, that transmission line will also be overloaded.

Compared to the situation before the failure, area C has lost an import of 100 MW and an export of 40 MW. This leaves us with a deficit of 60 MW, which gives us the frequency decrease $\Delta f_C = 60/50$ Hz = 1.2 Hz. The frequency in area C must then be $f_C = 48.8$ Hz.

Area B is now left to its own gain. In total, the area has lost 240 MW generation and 240 MW import from areas A and C. The frequency change required to compensate for the lost generation is $\Delta f_B = 480/50$ Hz = 9.6 Hz! Such a large frequency decrease is however not acceptable. Probably the system operator will order rotating load curtailments in area B to prevent the system from a complete blackout. If the gain in the example had been more evenly distributed it would have been possible to maintain the operation of the system without disconnection of transmission lines and rotating load curtailments (see exercise 4.8).

4.2 SECONDARY CONTROL

When an imbalance occurs between generation and consumption it will be eliminated by the primary control, as described above. However, when the balance is restored the frequency will deviate from the nominal value. Moreover, the system reserves will partly be used up, which means that it can be difficult to respond to new imbalances.

The secondary control is used to solve these problems. Its function is thus the following:

- Restore the frequency to its nominal value.
- Release used primary control reserves so that there is enough margins to respond to new load and/or generation changes.
- Take measures to prevent the time deviation from becoming to large.

The time deviation is the difference between synchronous time and normal time. By synchronous time we mean the time shown by a clock triggered by the grid frequency. If the grid frequency constantly was equal to its nominal value, synchronous time would correspond to the normal time. However, if the grid frequency is less than the nominal value then the synchronous time loose time compared to the normal time, and then the grid frequency is higher than the nominal value, it would gain time.

In a 50 Hz system one second corresponds to fifty cycles. If the grid driven clock sees an extra cycle per second (i.e., if the grid frequency is 51 Hz) then this correspond to 0.02 seconds. The total time deviation can hence be written as

$$t_i = \int_0^t \frac{(f(\tau) - 50)}{50} d\tau, \tag{4.2}$$

where

t_i = time deviation,
t = continuous time.

In for example the Nordic electricity market it has been agreed that the system operators in Sweden and Norway together are responsible for preventing the time deviation to become larger than ±10 s.

Example 4.3 (time deviation): At a certain occasion the frequency in the system is 49.8 Hz and the time deviation is +1 s. How large is the time deviation after five minutes?

Solution: During each second the grid driven clock will loose (49.8 – 50)/50 = 0.004 s. In 300 s the clock will therefore loose 1.2 s. The time deviation after five minutes is consequently 1 – 1.2 = –0.2 s.

The operation of the secondary control is simple. Assume that a large generator is lost. The primary control will replace this generator by increasing the generation in the primary controlled units until a new, stable and lower frequency is reached. This means that parts of the reserves have been used. If a new power plant (which does not participate in the primary control) is started then the frequency will increase, which is balanced by decreased generation in the primary controlled units. Hence, in that way it is possible to restore the frequency and regain the primary control reserves. The time deviation is reduced by starting up “too much” secondary power if the frequency has been too low for too long time.

Secondary control is a necessary function in all power systems, but it can be implemented in several ways. In many power systems the so-called AGC (Automatic Generation Control) is used, which means that the control systems in the power plants in addition to the gain also have an integrating part, which eliminates the frequency deviation. AGC is not used in the Nordic synchronous grid; instead, all secondary control is managed manually form the control rooms of the system operators.

It should be noted that if generation is lost in a certain part of the system then it might be necessary to start up new generation in the same part of the system in order to prevent the transmission from changing too much.

4.3 COST OF FREQUENCY CONTROL

The frequency control is, as said above, necessary if the power system should be able to function at all; hence, it can be seen as an example of a technical function which is far more important than economy (cf. section 1.1); the costs caused by a total collapse of the power system is by far larger than the costs of maintaining the frequency. Still, it can be interesting to have an idea of what is causing the costs of frequency control.

All power plants have certain operation points where the efficiency is as high as possible. When performing short-term planning of power plants it is of course desirable to use these best efficiency points as much as possible. The primary controlled units can however not be planned to operate at maximum efficiency, as the generation is partly determined by the gain and the system frequency. The primary control therefore means that some primary controlled units will be operated at a lower efficiency than the maximal. In most cases the decrease in efficiency is small and the costs are consequently also rather small.

Another cost for the primary control is that it costs to keep margins in the primary controlled units. Sometimes it might occur that a more expensive power plant, which is not part of the primary control, is used even though there is unused capacity in a less expensive primary controlled unit. The unused capacity of the primary controlled unit can however not be utilised, since the power plant then would be unable to increase its generation if the frequency decreases. It is also possible that situations occur when a more expensive primary controlled unit is used, even though there is unused capacity in a less expensive power plant which is not participating in the primary control. In this case the primary controlled unit is operated at part capacity in order to be capable of reducing its generation if the frequency should be increased.

The secondary control includes rescheduling the operation of hydro power plants and—in those system where AGC is used—thermal power plants so that better efficiency points can be used, i.e., a cost decrease. Extra costs can however occur during short time periods when the primary control reserve is restored, for example if an expensive, but fast-started, power plant (for instance a gas turbine) is started while waiting for less expensive reserves to be ready for operation.

EXERCISES

4.1 A power plant generates 100 MW when the frequency is exactly 50 Hz. When the frequency decreases to 49.95 Hz the power plant increases its generation to 110 MW. What is the gain of this power plant?

4.2 The frequency in a power system is exactly 50 Hz. The total gain in the system is 2 000 MW/Hz. Calculate the new frequency when the primary control has restored balance after the following events:

a) The load increases by 100 MW.

b) The load decreases by 80 MW.

c) A power plant, which is not participating in the primary control, increases its generation by 80 MW.

d) A power plant with the gain 200 MW/Hz, which is currently generating 400 MW, is stopped due to a serious fault.

4.3 The total gain in a certain power system is 5 000 MW/Hz. The largest power plant in the system is a nuclear power plant of 800 MW; this power plant does not participate in the primary control. The second largest power plant is a hydro power plant of 650 MW, which has the gain 50 MW/Hz. In which of the following cases will it be possible to maintain a frequency within the interval 49.9 to 50.1 Hz?

a) Generation and consumption are in balance, and the frequency is 50.05 Hz when the nuclear power plant — which is operated at maximal generation — must be stopped immediately.

b) Generation and consumption are in balance, and the frequency is 49.98 Hz when the hydro power plant must be disconnected due to a short-cut in a transformer. The hydro power plant was generating 400 MW before the error.

c) Generation and consumption are in balance, and the frequency is 49.99 Hz. During the half-time pause of a football game, 200 000 TV viewers start making coffee, which causes a load increase of 250 MW within just a few minutes.

d) The generation exceeds the consumption by 150 MW, and the frequency is 49.93 Hz. A stroke of lightning causes a major black-out in a city, which results in a load decrease of 650 MW.

e) Generation and consumption are in balance and the frequency is exactly 50 Hz. The time deviation is –16 seconds and it is desired to reduce the deviation to –10 seconds within 10 minutes. No other generation or load changes occur during these 10 minutes.

4.4 The hydro power plant Fors has the installed capacity 100 MW. It is not possible to generate less than 40 MW in the power plant, because then there would be a risk of damaging the turbines. The gain of the power plant is set to 200 MW/Hz and the generation is 70 MW when the frequency is exactly 50 Hz.

a) How much is Fors generating when the system frequency is 49.82 Hz?

b) How much is Fors generating when the system frequency is 49.94 Hz?

c) How much is Fors generating when the system frequency is 50.06 Hz?

d) How much is Fors generating when the system frequency is 50.18 Hz?

4.5 The island Ön has two power plants participating in the primary control. The gain in each power plant is 50 MW/Hz. One of the power plants has an installed capacity of 50 MW and generates 44 MW at 8:00 AM. The other power plants has an installed capacity of 100 MW and generates 62 MW at 8:00 AM.

Ön is connected to Land via an AC transmission line. This line has a maximal transmission capacity of 400 MW and is equipped with a protection system which after a short time delay disconnects the line if the maximal capacity is exceeded. At 8:00 AM the transmission on the line is 390 MW from Land to Ön.

The power plants located in Land which are participating in the primary control has a total gain of 4 950 MW/Hz available in the frequency range 50±0.4 Hz.

At 8:00 AM the frequency is 49.98 Hz. Just after that a nuclear power plant is stopped in Land. The nuclear power plant has an installed capacity of 1 200 MW and was generating 1 006 MW when the failure occurred.

a) How much does the electricity generation in Ön increase due to the failure?

b) Will the transmission line to Land be able to transfer the generation increase without being disconnected?

c) Which frequency will be obtained in Land and Ön? Answer with three decimals!

4.6 Assume that a power system is supplied by a thermal power plant with the maximum capacity 220 MW and three hydro power plants with the capacities 200 MW, 400 MW and 600 MW respectively. At a certain occasion the frequency of the system is exactly 50 Hz, the thermal power plant is generating 190 MW and the hydro power plants are operating at 80% of their installed capacity. The two smaller hydro power plants have a gain of 250 MW/Hz and the largest hydro power plant has the gain 500 MW/Hz. At this occasion the thermal power plant is suddenly stopped. Calculate the system frequency and the generation in the hydro power plants when the balance is restored.

4.7 A power system is divided in two areas, A and B. The maximal transmission capacity between the two areas is 500 MW. The connections between A and B are equipped with protection systems which after short time delay disconnects overloaded transmission lines. In average, area A exports 200 MW to area B. The dimensioning fault (i.e., the largest production loss which can be caused by a failure in a power plant) is 800 MW in area A and 400 MW in area B. The dimensioning fault *does not* affect power plants participating in the primary control.

a) Assume that it has been decided that the frequency may not drop more than 0.2 Hz if a dimensioning fault occurs in either of the areas. How large must the total gain be then?

b) Moreover, assume that it has been decided that the system should be able to manage a dimensioning fault in either of the areas, which occurs when the power flow between the areas correspond to the average, without any transmission lines becoming overloaded. How must the total gain be distributed between area A and B?

4.8 Consider the same system as in example 4.2, but assume that the gain is 300 MW/Hz in each area. Check that the system can manage the transmission changes induced by the primary control and calculate the resulting system frequency.

4.9 Consider the same system as in example 4.2, but assume that the gain in the three systems are $R_A = 650$ MW/Hz, $R_B = 500$ MW/Hz and $R_C = 450$ MW/Hz respectively. At a certain occasion the frequency is exactly 50 Hz and the power flows of the transmission lines are as shown in figure 4.3. A short circuit, caused by a stroke of lightning, causes the protection system to disconnect line AC. Calculate the resulting frequencies and the transmission on the tie lines when the system is stabilised again.

4.10 Table 4.2 shows how the total gain is divided among the countries in the synchronous Nordel system at a certain occasion. The table also shows the transmission between the countries. At this occasion the frequency is exactly 50 Hz and the time deviation is +1 s. At this occasion a 200 MW coal condensing unit is stopped in Sjælland. Five minutes later the load in Norway increases by 100 MW and five minutes after that a 300 MW thermal power plant is started in Finland. Calculate the following quantities 10 minutes after the start of the Finnish power plant: the frequency, the time deviation and the transmission between the countries.

All necessary data are not given for this problem, so one assumption has to be

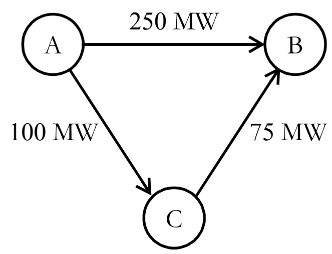


Figure 4.3 Transmission before the fault on the line AC.

made! Describe this assumption.

Table 4.2 Data for the Nordel system in exercise 4.10.

Country	Gain [MW/Hz]	Connection	Transmission [MW]
Sweden	2 500	Norway → Sweden	2 000
Norway	2 000	Finland → Norway	100
Finland	1 250	Sweden → Finland	200
Denmark (Sjælland)	250	Sweden → Denmark (Sjælland)	500

4.11 Assume that the total gain for the Nordel countries and the transmission between them are as recorded in table 4.3. At this occasion the frequency of the system is exactly 50 Hz and the time deviation is –1 s. Suddenly there is a problem on the HVDC-link between Jutland (which is not a part of the synchronous Nordel system) and Norway. The earlier 500 MW import to Norway changes to an export of 500 MW. (This has happened in reality in USA when static electricity in a control circuit cased a sudden change in direction of the power flow). Also assume that the load and generation in the power plants not participating in the primary control remain constant. Calculate the following quantities 10 minutes after the error: the frequency, the time deviation and the transmission between the countries.

Table 4.3 Data for the Nordel system in exercise 4.11.

Country	Gain [MW/Hz]	Connection	Transmission [MW]
Sweden	2 500	Denmark (Jutland) → Norway	500
Norway	2 000	Norway → Sweden	600
Finland	1 250	Finland → Norway	100
Denmark (Sjælland)	250	Sweden → Finland	200
		Sweden → Denmark (Sjælland)	500

FURTHER READING

- A. J. Wood & B. F. Wollenberg, *Power Generation Operation and Control*, second edition, John Wiley & Sons, 1996. — *General text book which among other things describe the control theory behind frequency control.*

Chapter 5
SHORT-TERM PLANNING

There is a large variety of models and calculations methods that can be used for operation planning of power plants in a system. Although the various models have common characteristics, it is always necessary at least to some extent to adjust a planning problem to the circumstances at hand. In this chapter we will present a selection of basic models for short-term planning of hydro power plants and thermal power plants respectively.

The models described in this chapter are linear, and readers who are not familiar with linear programming are advised to study appendix A in order to better profit from the chapter.

5.1 OBJECTIVE AND CONDITIONS

The objective of a short-term planning is to determine detailed plans how to act for the closest future. Normally a short-term plan would comprise a time span between a day and a week, which is subdivided into a number of periods. The length of the periods may vary, but the natural choice is to let each period correspond to a trading period in the electricity market—in this description we will assume that each period corresponds to an hour. Thus, the result of a short-term plan is a plan, which for example states how much each power plant should generate during each hour of the planning period. It is also possible to decide how much should be purchased or sold to other players (in particular the power pool).

The Difference Between Short-term Planning and Season Planning

It should be noted that when we in this compendium refer to short-term planning, we mean a problem where *all important factors influencing the operation schedule are known* or that *there is a reliable forecast*. In reality this condition will be hard to fulfil. If we for example consider a player who intend to trade at the power exchange Nord Pool, then this player must submit bids of a particular 24-hour period no later than at noon the day before. Then all submitted bids are compiled and it is notified which bid were accepted and which electricity prices will apply during the 24 hours. Hence, the player must decide how to act before the prices are set and the prices can then be affected by the decision made by the player.

When planning the operation it is possible to account for several plausible chains of events; such planning problems are in this compendium referred to as season planning problems, as they usually include a longer time perspective than the short-term planning problem. Season planning problems are much larger than short-term planning problems, and are more difficult to manage from a mathematical point of view.

A General Short-term Planning Problem

Short-term planning is about trying to utilise the resources as efficiently as possible, while considering physical and legal limitations.[1] Utilising the resources as efficiently as possible is usually equivalent to maximising the profits. A planning problem is therefore a natural application of optimisation theory. In short, the optimisation theory is a branch of mathematics concerned with maximisation or minimisation of the value of a function defined on a set of feasible solutions (see appendix A for further details). In this chapter we will just formulate the optimisation problems; we assume that the problems then can be solved using standard software (see appendix B for some examples of appropriate software).

For a player in an electricity market, we can provide this very general formulation of the short-term planning problem:

maximise	*the income during the planning period + future income* *– the costs during the planning period – future costs*,	(5.1)
subject to	*physical limitations*,	(5.1a)
	economic/legal limitations.	(5.1b)

The objective function (5.1) is to maximise the profit of the player. Notice that it sometimes is necessary to consider that a decision made during the planning period can have consequences after the end of the planning period. To find an optimal solution it is then necessary to include possible future income and costs.

The constraints consist of physical limitations and economic/legal limitations. The physical limitations constitute physical laws and limitations that are part of the power system and which cannot be exceeded (for example, a power plant cannot generate more than its installed capacity and a hydro reservoir cannot hold any more water when it is full). These constraints depend on how we choose to model the power plants and the remainder of the power system; several examples will be given in the following sections. The economic and legal constraints includes a variety of rules that the player have to adjust to, for example limitations in how much water may be diverted from a river and how large emissions of carbon dioxide are permissible for a power plant.

5.2 HYDRO POWER

Hydro power plants are characterised by very low variable operation costs[2] and in this presentation we will assume that the variable operation cost is negligible. Characteristic for hydro power is that there is a limited amount of energy (in the form of water in the reservoirs) and this energy should be used in an optimal manner. Therefore, it is desirable to use as good efficiency as possible and the generation should preferably be scheduled for those time periods when the electricity prices are high. At the same time, it must be considered that power plants in the same river system cannot be operated independently from each other.

1. This description would by the way fit most planning problems.
2. By and large, the operation cost constitutes of some wear and tear of the electrical and mechanical equipment of the power plant, as well as a small usage of lubricants.

5.2.1 General Description of Hydro Power Plants

A hydro power plant generates electric power by utilising the difference in potential energy between an upper and a lower water level. The potential energy is converted to kinetic energy when the water is discharged from the upper to the lower water level via a turbine. The turbine in its turn is driving a generator, where the kinetic energy is transformed to electric energy.

The most important parts of a hydro power plant are shown in figure 5.1. The water is often collected in one or more hydro reservoirs, but there are also so-called run-of-the-river hydro power plants which do not have any possibility to store water. Through the intake the water is led via an headrace tunnel (or a headrace channel) and the penstock to the turbine itself. The exact layout varies from plant to plant and depends on several factors, as for example the bedrock, the head and the distance between reservoir and turbine. Figure 5.1a shows the typical layout of a hydro power plant with low head, where the water is led directly from the intake through a very short penstock to the turbine. After passing the turbine, the water is released directly out into the natural riverbed, which may have been dredged in order to increase the head. It is also possible to spill water past the power station, by opening gates in the dam.

In figure 5.1b a power plant with higher head is shown. In this case the water is led via a headrace tunnel through the rock before it reaches the penstock. Also the powerhouse and the tailrace tunnel are excavated in the rock. The natural riverbed is used if there is a need to spill water past the power plant; hence, it is normally drained or has a very reduced water flow.

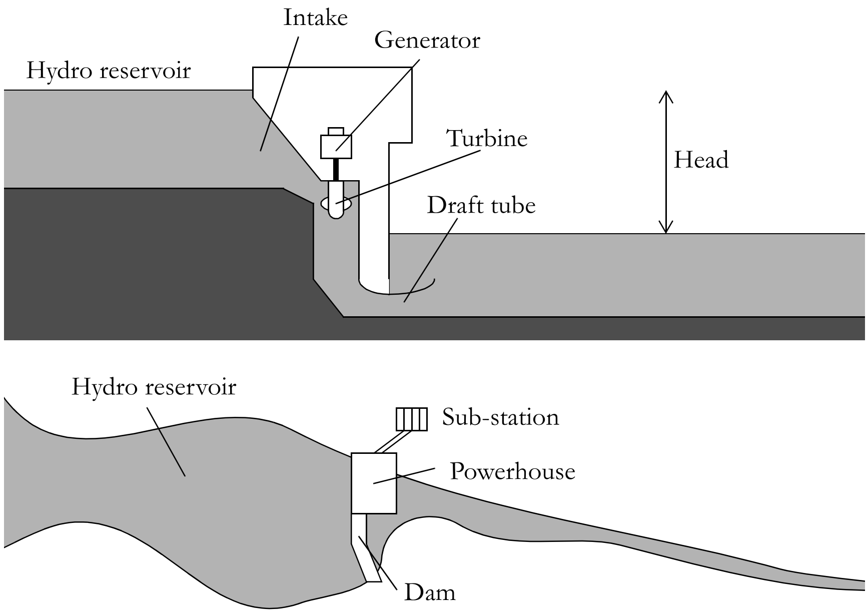
Discharge, spillage and reservoir contents are all measured in hour equivalents (HE), which corresponds to the water flow 1 m^3/s during 1 hour. Observe that the unit HE sometimes is interpreted as a volume and sometimes as water flow! It should however always be clear from the context what is meant.

5.2.2 Discharge and Efficiency

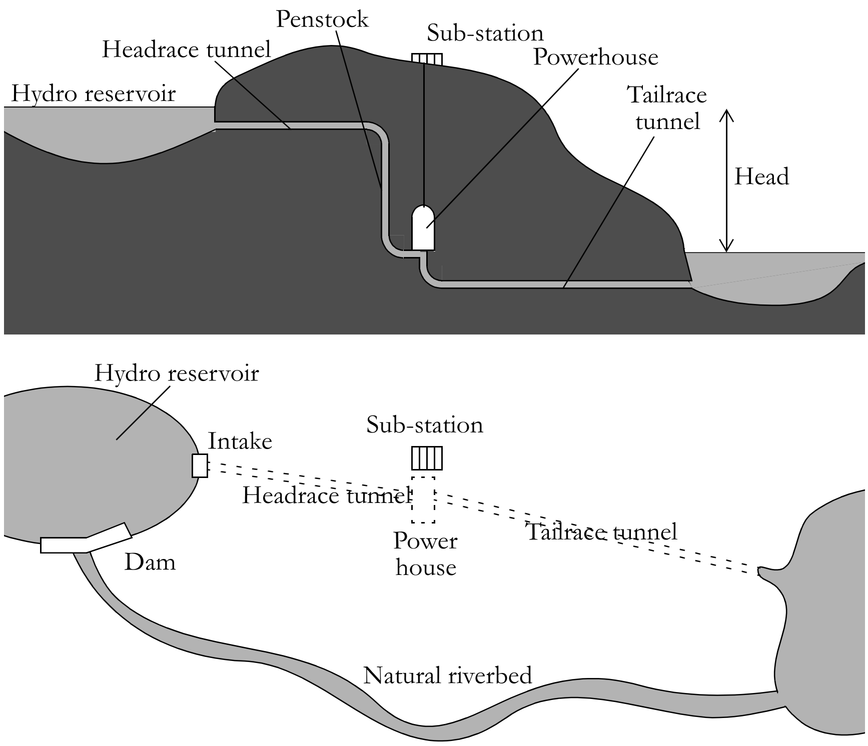
The efficiency of a hydro power plant is depending on the head (i.e., the height difference between the water level at the intake and the water level at the end of the tailrace) and the discharge through the turbine. The relation between power generation, discharge and head is a non-linear function, which somehow must be approximated by a linear or piecewise linear function in order to be included in an LP model. The impact of the head is rather small and we will therefore neglect this factor. The problem is then to find a linear approximation of the power generation as a function of the discharge, $H(Q)$.

When calculating the power generation in a hydro power plant it is important to understand the difference between the following three notions:

> ***Definition 5.1.*** The production equivalent is the quota between the energy generation and the discharge through the turbines. The production equivalent is different for different discharges. It is denoted by γ and measured in MWh/HE. Given the generation and the discharge, the production equivalent is calculated by
>
> $$\gamma(Q) = \frac{H(Q)}{Q}.$$
>
> ***Definition 5.2.*** The marginal production equivalent is a measure of how much the power generation will increase for a small change of the discharge, i.e., $dH(Q)/dQ$. The marginal production equivalent is denoted by μ and is also measured in MWh/HE.
>
> ***Definition 5.3.*** The relative efficiency normally refers to the production equivalent for some discharge compared to the maximal production equivalent of the power plant. The relative efficiency is denoted η and measured in per cent. Given the pro-



a) Hydro power plant with low head.



b) Hydro power plant with high head.

Figure 5.1 Overview of the layout of typical Swedish hydro power plants.

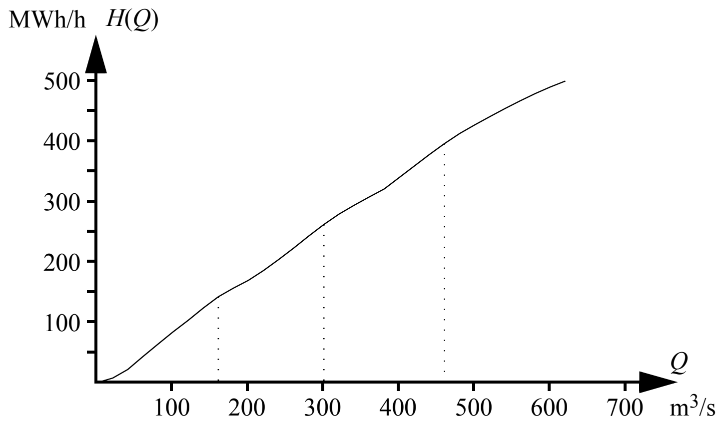


Figure 5.2 The electricity generation of a particular hydro power plant.

duction equivalent as a function of the discharge the relative production equivalent is calculated by

$$\eta(Q) = \frac{\gamma(Q)}{\gamma_{max}},$$

where

$$\gamma_{max} = \max_{Q} \gamma(Q).$$

The relative efficiency shows how much energy can be extracted from each m^3 water compared to the maximal possible. If we only had to consider the energy generation then we would always operate hydro power plants at the maximal efficiency. There are however other constraints which need to be considered, as we will see later in this chapter.

At low discharges the efficiency is poor, which is illustrated in the following example:

Example 5.1 (relative efficiency in a hydro power plant): In a certain hydro power plant the electricity generation is—slightly simplified—a function of the discharge according to table 5.1; this relation is also shown in figure 5.2. The curve is the result of a station optimisation, i.e., for each level of the total discharge, the water flow is distributed between the turbines so that as much power as possible is generated.[3] The points where there are local maxima in the efficiency are indicated by dotted lines.

Determine the relative efficiency curve for this hydro power plant.

Solution: In this example the best efficiency is obtained for the discharge 160 HE, then the power plant generates 126.56 MW. Hence, we get

$$\gamma_{max} = \frac{126.56}{160} = 0.791 \text{ MWh/HE}.$$

The relative efficiency as function of the discharge is now according to the definitions above given by

3. In this power plant one turbine is used for discharges up to about 210 m^3/s At this discharge another turbine is engaged. When the discharge is larger than approximately 380 m^3/s a third turbine is also used. This explains the appearance of the curve.

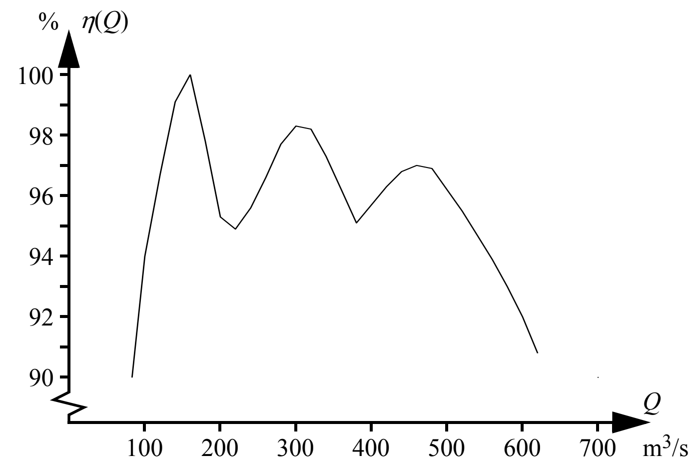


Figure 5.3 Relative efficiency curve of the hydro power plant in example 5.1.

Table 5.1 The electricity generation of a particular hydro power plant.

Q [m³/s]	H [MW]	Q [m³/s]	H [MW]	Q [m³/s]	H [MW]	Q [m³/s]	H [MW]
0	0	160	126.5600	320	248.5638	480	367.9099
20	6.3280	180	139.2476	340	261.6786	500	380.4710
40	18.9840	200	150.7646	360	273.9391	520	392.8106
60	37.9680	220	165.1450	380	285.8516	540	404.5016
80	56.5090	240	181.4870	400	302.7948	560	415.9394
100	74.3540	260	198.6676	420	319.9279	580	426.6654
120	91.7876	280	216.3860	440	336.9027	600	436.6320
140	109.7433	300	233.2659	460	352.9442	620	445.3014

$$\eta(Q) = \frac{\gamma(Q)}{0.791} = \frac{H(Q)}{0.791Q},$$

which results in the relative efficiency curve shown in figure 5.3.

A piecewise linear model of the power generation in a hydro power plant means that the generation as function of the discharge is divided into one or more segments. The breakpoints between the segments are preferably located to those discharges where there are local maxima in the efficiency, because these discharges are then more likely to appear in the solution of the planning problem.[4] In each segment the marginal production equivalent is approximated by a constant, as shown in figure 5.4. This interpretation means that the total discharge through power plant i must be divided in one variable per segment:

$$Q_{i,\,t} = \sum_{j=1}^{n_i} Q_{i,j,\,t}, \tag{5.2}$$

where

$Q_{i,\,t}$ = discharge through power plant i during hour t,
$Q_{i,\,j,\,t}$ = discharge in power plant i, segment j, during hour t,
n_i = number of segments in power plant i.

4. To understand why it is like this it would be necessary to further study feasible solutions to LP problems, which is beyond the scope of this compendium.

The total power generation in the power plant is calculated by

$$H_{i,t} = \sum_{j=1}^{n_i} \mu_{i,j} Q_{i,j,t}, \tag{5.3}$$

where

$H_{i,t}$ = power generation in power plant *i* during hour *t*,
$\mu_{i,j}$ = marginal production equivalent for power plant *i*, segment *j*.

In appendix A (examples A.12 and A.13) it is shown that integer variables might be required to represent a piecewise linear function in a correct manner. In this case we have a certain total discharge, which is to be divided between the different segments.[5] There is however just one division which is corresponding to the physical interpretation of the model, i.e., the division where first segments are fully utilised before any discharge is made from the next segment. To avoid integer variables, the piecewise linear function must be such that it is more profitable to use the first segment than the second, and the second segment must be preferable to the third, etc. This means that the marginal production equivalents have to be *decreasing* in the linear model (as it is more profitable to use a segment having a high marginal production equivalent, since more electric energy is obtained per HE of water), i.e.,

$$\mu_{i,j} > \mu_{i,k} \text{ if } j < k. \tag{5.4}$$

Let us now study an example of a model of a hydro power plant:

Example 5.2 (linear model of a hydro power plant): Determine a piecewise linear model of the hydro power plant in example 5.1.

Solution: As discussed above it is important that the model is convex, i.e., the more water that is discharged the lesser is the efficiency. Such a model can be obtained by using the local best efficiency points as breakpoints for the piecewise linear function. It can be seen in figure 5.2 and figure 5.3 that the efficiency has a local maximum for the discharges 60 m^3/s, 300 m^3/s and 460 m^3/s respectively. The electricity generation for these discharges are given in table 5.1.

The piecewise linear function is shown in figure 5.4. The marginal production equivalents correspond to the slope of the linear segments of the piecewise linear curve:

$\mu_1 = (126.5600 - 0)/(160 - 0) \approx 0.79$ MWh/HE,

$\mu_2 = (233.2659 - 126.5600)/(300 - 160) \approx 0.76$ MWh/HE,

$\mu_3 = (352.9442 - 233.2659)/(460 - 300) \approx 0.75$ MWh/HE,

$\mu_4 = (445.3014 - 352.9442)/(620 - 460) \approx 0.58$ MWh/HE.

Hence, the model for this hydro power plant can be formulated mathematically as

$$H = \sum_{j=1}^{4} \mu_j Q_j,$$

$0 \le Q_1 \le 160$,

$0 \le Q_2 \le 140$ (= 300 – 160),

$0 \le Q_3 \le 160$ (= 460 – 300),

5. For example, $Q_1 = 20$ and $Q_2 = 0$ provides the same total discharge according to (5.2) as $Q_1 = 0$ and $Q_2 = 20$.

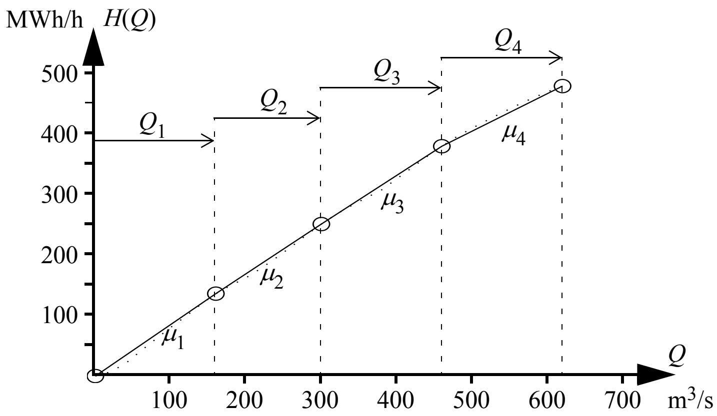


Figure 5.4 Piecewise linear model of the electricity generation of the hydro power plant in examples 5.1 and 5.2. The real electricity generation as a function of the discharge is shown by a dotted line.

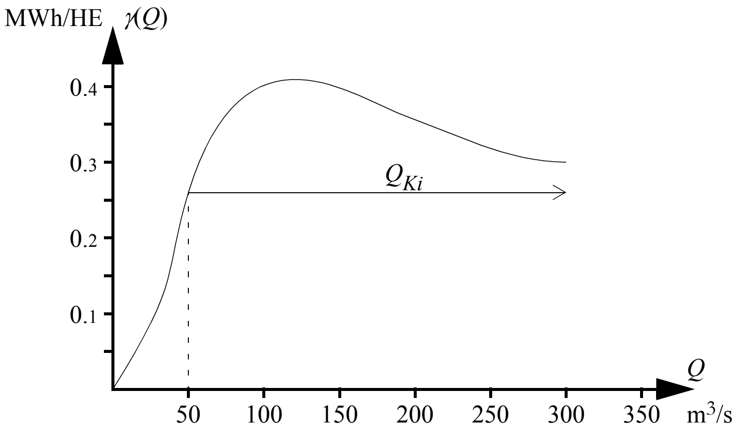


Figure 5.5 Example of modelling the electricity generation in a hydro power plant using a mixed integer linear model.

$0 \le Q_4 \le 160$ (= 620 – 460).

In figure 5.4 we can se that the resemblance between the real electricity generation and the linear model is quite good (in particular close to the break points of the piecewise linear function). However, it is also possible to see that for low discharges the linear model is significantly overestimating the electricity generation.[6] A straightforward method to avoid this problem is to quite simply forbid certain discharges. The principle is illustrated in the following example:

Example 5.3 (forbidden discharge domain): Suppose that the production equivalent as a function of the discharge in a certain power plant *i* is as shown in figure 5.5. As can be seen in the figure, the efficiency is relatively low for discharges below 50 m^3/s (and consequently so is the production equivalent); therefore, the plant should be scheduled so that either more than 50 HE is discharged or nothing at all. Create a linear model of the power generation as a function of the discharge.

6. If we for example consider the electricity generation at the discharge 20 m^3/s then the linear model states that the generation is $20\mu_1 \approx 15.8$ MW, whereas it in reality is about 6.3 MW. Thus, the linear model overestimates the generation by 150%!

Solution: For simplicity we choose to divide the generation as function of the discharge in only two segments. The first segment includes only one point, namely the origin. The other segment includes all discharges between 50 HE and maximal discharge (300 HE). We assume that the marginal production equivalent in the second segment is a constant, μ_{Ki}, which for example could be chosen as the average of $\gamma(Q)$ in the interval between 50 and 300 HE.

Now we introduce a binary integer variable which represent the active segment during a certain hour:

$$z_{i,\,t} = \begin{cases} 0 \text{ if the discharge is equal to 0 HE,} \\ 1 \text{ if the discharge is larger than or equal to 50 HE.} \end{cases}$$

The discharge in the continuous segment is denoted by $Q_{Ki,\,t}$. The total discharge in the power plant can thus be written as

$$Q_{i,\,t} = 50z_{i,\,t} + Q_{Ki,\,t},$$

which means that the power generation is

$$H_{i,\,t} = 50\mu_{Ki}z_{i,\,t} + \mu_{Ki}Q_{Ki,\,t}.$$

The equations above will only provide correct results if we introduce a constraint which forces the discharge of the second segment to be equal to zero when $z_{i,\,t} = 0$. This is achieved by the following inequality constraint:

$$Q_{Ki,\,t} \le \overline{Q}_{Ki}\, z_{i,\,t}.$$

Finally, we have to introduce variable limits for $Q_{Ki,\,t}$ and $z_{i,\,t}$:

$$0 \le Q_{Ki,\,t},$$

$$z_{i,\,t} \in \{0, 1\}.$$

It can be noted that avoiding running the power plant at a poor efficiency level is not the only reason to forbid certain domains of the discharge. It can also be possible that certain discharges causes vibrations in the turbines which may result in expensive repairs. The method described in the example can be generalised to a model where the accepted levels of discharge are no charge at all, a number of local best efficiency points and a continuous interval up to maximal discharge.

5.2.3 Hydrological Coupling Between Hydro Power Plants

In order to use the available water as efficiently as possible it is necessary to coordinate the operation of hydro power plants located in the same river system. The coordination partly has to try to achieve the best possible efficiency in all power plants and partly has to avoid situations where spillage is necessary (for example if a power plant is operated at maximal discharge even though the downstream reservoir can not store the water). The modelling of the efficiency was described in the previous section. To avoid spillage it is necessary to consider the hydrological coupling between the power plants, which means that each power plant has to fulfil the following balance equation:

$$\begin{aligned} &\textit{new reserovir contents} = \textit{old reservoir contents} + \\ &\qquad\qquad \textit{water inflow} - \textit{water outflow.} \end{aligned} \tag{5.5}$$

This is a physical necessity and has to be valid for each hour (or any other specified time period).

Mathematically (5.5) is expressed as

$$M_{i,t} = M_{i,t-1} - Q_{i,t} - S_{i,t} + \sum_{j \in \mathcal{K}_i} Q_{j,t-\tau_{j,i}} + \sum_{j \in \mathcal{K}_i} S_{j,t-\tau_{j,i}} + V_{i,t}, \tag{5.6}$$

where

$M_{i,t}$ = contents of reservoir i at the end of hour t [HE],
$Q_{i,t}$ = discharge in power plant i during hour t [HE],
$S_{i,t}$ = spillage past power plant i during hour t [HE],
$\mathcal{K}_i$ = the set of indices for power plants directly upstream of power plant i,
$\tau_{j,i}$ = the delay time for the water between power plant j and the closest downstream power plant i [h],
$V_{i,t}$ = local inflow to reservoir i during hour t [HE].

It must be noted that the delay time τ_{ji} between two power plants is a relatively complex function of season,[7] the water flow and the reservoir levels. For simplicity we assume that the delay time is constant, i.e., water discharged or spilled from a certain reservoir will reach the next reservoir after a given number of minutes. Assume that the delay time between j and i is h_j hours and m_j minutes, independent of the discharge, etc. The following expression for the discharge can then be used:

$$Q_{j,t-\tau_{j,i}} = \frac{m_j}{60} Q_{j,t-h_j-1} + \frac{60-m_j}{60} Q_{j,t-h_j}, \tag{5.7}$$

i.e., a weighted average of the discharge h_j and $h_j + 1$ hours earlier. A similar expression can be formulated for the spillage.

Juridical restrictions and physical laws imply that reservoir levels as well as discharges are limited. For the last hour in the planning period it is common that there is an upper and a lower limit for the reservoir contents, which is obtained from long-term plans. These limitations can be formulated as

$$\underline{Q}_i \le Q_{i,t} \le \overline{Q}_i, \tag{5.8}$$

$$\underline{M}_i \le M_{i,t} \le \overline{M}_i, \tag{5.9}$$

$$\underline{M}_{i,T} \le M_{i,T} \le \overline{M}_{i,T}, \tag{5.10}$$

where

$\underline{Q}_i$ = lower limit for discharge in power plant i during hour t,
$\overline{Q}_i$ = upper limit for discharge in power plant i during hour t,
$\underline{M}_i$ = lower limit for the contents of reservoir i,
$\overline{M}_i$ = upper limit for the contents of reservoir i,
$\underline{M}_{i,T}$ = lower limit for the contents of reservoir i at the end of hour T, i.e., at the end of the planning period,
$\overline{M}_{i,T}$ = lower limit for the contents of reservoir i at the end of hour T.

Generally the allowed reservoir levels are given in metres above the sea level. These limits can however easily be translated into limits for the reservoir contents.

5.2.4 Value of Stored Water

When planning the operation of hydro power plants it is necessary to consider that the electricity

7. The delay time is for example affected by whether or not the river is icy.

price can vary. At occasions when the electricity price is low it might very well be preferable to generate nothing at all and rather save the water of the hydro reservoirs for a later occasion, when the electricity price is more favourable. Consequently, it is not sufficient to maximise the income during the planning period, but it is also necessary to consider the future income that can be obtained from the stored water at the end of the planning period.

There are two methods to include the value of stored water. The most simple is to decide in advance how much water that should be stored at the end of the planning period. This preset value could for example originate from the company long-term plans. A more advanced method is to determine a function to value the water. The value of the water depends partly on how much electric energy can be generated from the water and partly on which electricity prices can be expected when the power is sold, i.e.,

$$B_i(M_i) = \lambda_e M_{i,\,T} \sum_{j \in \mathcal{M}_i} \gamma_j, \tag{5.11}$$

where

$B_i(M_i)$ = value of the water stored in reservoir *i*,
λ_e = expected electricity price,
$M_{i,\,T}$ = contents of reservoir *i* after the end of the planning period,
γ_j = expected future production equivalent in power plant *j*,
$\mathcal{M}_i$ = the set of indices for all power plants downstream of reservoir *i* (including power plant *i* itself).

It should be noted that we have to consider that the water in a certain reservoir eventually will reach the reservoirs downstream and therefore will be used for power generation in these power plants too!

5.2.5 Some Short-term Hydro Power Planning Problems

In the previous sections we have shown how to model the electricity generation in hydro power plants and the hydrological coupling between power plants in the same river system as well as how to calculate the value of stored water. With that, it is time to give a few examples how these models can be applied when formulating short-term planning problems for hydro power plants:

Example 5.4 (planning of sales to a power pool): Assume that you are the owner of the hydro power plants Degerforsen and Edensforsen in River Ångermanälven. (Degerforsen is located upstream of Edensforsen.) Data for the two power plants are given in table 5.2. The electricity generation in the two power plants are sold to a power exchange, where the electricity price the next six hours is forecasted to 165, 168, 194, 202, 187 and 183 SEK/MWh respectively. Assume that water which is stored after the sixth hour can be discharged later and that the electricity which then is generated will be sold for 185 SEK/MWh. The inflow during the period is equal to the mean annual flow and the reservoirs are half-filled at the beginning of the planning period. Assume constant efficiency for the power plants, i.e., constant production equivalent, and that the installed capacity is reached at maximum discharge. Neglect the head losses and the delay time between the power plants. Suggest a discharge plan for the two power plants during the next six hours.

Solution: Since the operation cost in a hydro power plant is negligible, the planning problem can be formulated in words as

maximise	*income of sold energy + value of stored water*,
subject to	*hydrological balance in the hydro power plants*.

Table 5.2 Data for Degerforsen and Edensforsen.

	Degerforsen	Edensforsen
Year of completion	1966	1956
Type of power plant	Above ground	Below ground
Gross head [m]	24	28
Mean annual flow [m^3/s]	163	164
Design flow[m^3/s]	300	270
Number and type of turbines	2 × Kaplan	2 × Kaplan
Installed capacity[MW]	62	63
Average annual energy production [GWh]	295	320
Type of dam	Rockfill/earthfill	Earthfill
Height of dam[m]	18	19
Length of dam crest[m]	2 150	700
Volume contents of the dam[m^3]	235 000	100 000
Active storage of the reservoir [m^3]	5 000 000	4 000 000
Maximum discharge capacity of spillways[m^3/s]	1 365	1 350

Before we formulate this problem as an LP problem, we make sure that we have defined all variables and parameters involved. We start by defining Degerforsen as power plant number 1 and Edensforsen as number 2. Then we define the optimisation variables (i.e., the variables controlled by the decisions of the company):

The following parameters are given directly in the problem text:

$$\overline{H}_i = \text{installed capacity in power plant } i = \begin{cases} 62 & i = 1, \\ 63 & i = 2, \end{cases}$$

$$\overline{Q}_i = \text{maximal discharge}^8 \text{ in power plant } i = \begin{cases} 300 & i = 1, \\ 270 & i = 2, \end{cases}$$

$$\lambda_f = \text{expected future electricity price} = 185,$$

$$\lambda_t = \text{expected electricity price during hour } t = \begin{cases} 165 & t = 1, \\ 168 & t = 2, \\ 194 & t = 3, \\ 202 & t = 4, \\ 187 & t = 5, \\ 183 & t = 6. \end{cases}$$

The remaining parameters are obtained through simple calculations. The maximal reservoir contents[9] must be converted from m^3 to HE:

$$\overline{M}_i = \text{maximal contents of reservoir } i = \begin{cases} 5\,000\,000/3\,600 \approx 1\,389 & i = 1, \\ 4\,000\,000/3\,600 \approx 1\,111 & i = 2. \end{cases}$$

The start contents is 50% of the maximal, i.e.,

8. Found under the heading “design flow” in table 5.2.

9. Found under the heading “active storage of reservoir” in table 5.2. The heading ”volume contents of the dam” refers to the volume of the dam construction itself.

$$M_{i,\,0} = \text{start content of reservoir } i = 0.5\overline{M}_i \approx \begin{cases} 694 & i = 1, \\ 556 & i = 2. \end{cases}$$

The local inflow is assumed to be equal to the mean annual flow at Degerforsen. For Edensforsen the local inflow is equal to the difference between the mean annual flow in the two power plants:

$$V_i = \text{local inflow to reservoir } i = \begin{cases} 163 & i = 1, \\ 164 - 163 = 1 & i = 2. \end{cases}$$

Finally, we have to calculate the constant production equivalent according to the data in the assignment:

$$\gamma_i = \text{production equivalent in power plant } i = \frac{\overline{H}_i}{\overline{Q}_i} \approx \begin{cases} 0.207 & i = 1, \\ 0.233 & i = 2, \end{cases}$$

Then we define the optimisation variables (i.e., the variables which are depending on the decisions of the company):

$Q_{i,\,t}$ = discharge in power plant i during hour t, $i = 1, 2$, $t = 1, \dots, 6$,
$S_{i,\,t}$ = spillage from reservoir i during hour t, i = 1, 2, $t = 1, \dots, 6$,
$M_{i,\,t}$ = contents of reservoir i at the end of hour t, $i = 1, 2$, $t = 1, \dots, 6$.

The next step is to formulate the objective function:

$$\text{maximise} \quad \sum_{t=1}^{6} \lambda_t \sum_{i=1}^{2} \gamma_i Q_{i,\,t} + \lambda_f((\gamma_1 + \gamma_2)M_1(6) + \gamma_2 M_2(6)).$$

The constraints are the hydrological balance for Degerforsen and Edensforsen respectively:

$$M_{1,\,t} - M_{1,\,t-1} + Q_{1,\,t} + S_{1,\,t} = V_1, \qquad t = 1, \dots 6,$$

$$M_{2,\,t} - M_{2,\,t-1} + Q_{2,\,t} + S_{2,\,t} - Q_{1,\,t} - S_{1,\,t} = V_2, \qquad t = 1, \dots 6.$$

Finally we have to consider the limits for the optimisation variables:

$$0 \le Q_{i,\,t} \le \overline{Q}_i, \qquad i = 1, 2, t = 1, \dots, 6,$$

$$0 \le S_{i,\,t}, \qquad i = 1, 2, t = 1, \dots, 6,$$

$$0 \le M_{i,\,t} \le \overline{M}_i, \qquad i = 1, 2, t = 1, \dots, 6.$$

Solving the optimisation problem above yield the discharge plan displayed in table 5.3.

Table 5.3 Discharge plan in example 5.4.

Hour, t	0	1	2	3	4	5	6
M_1 [HE]	694.4	857.4	1 020.4	883.4	746.4	609.4	772.4
M_2 [HE]	555.6	556.6	557.6	588.6	619.6	650.6	651.6
Q_1 [HE]		0	0	300	300	300	0
Q_2 [HE]		0	0	270	270	270	0
H_1 [MWh]		0	0	62	62	62	0
H_2 [MWh]		0	0	63	63	63	0
S_1 [HE]		0	0	0	0	0	0
S_2 [HE]		0	0	0	0	0	0

In the example above there was no need for an extra variable representing sales to the power exchange, since the whole generation was sold to the exchange. By introducing separate variables for purchase and sales at the exchange, it becomes possible to manage situations where there is a contracted load which must be supplied, as well as a possibility to trade at the exchange (cf. example 5.10).

It may be noticed that we did not state any upper limit for the spillage, although table 5.2 stated that there is a limitation in the discharge capacity through the spillways. This limitation does however refer to *controlled* spill. If the reservoir is full and the capacity of turbines and spillways is not large enough, the dam will be flooded, which in its own way is a form of spill. It is of course a very undesirable form of spill, because it would cause large damages, but it on the other hand in practice no risk that these situations should occur. Spillage always means lost income, so the solution to the planning problem will always minimise the spill. If a discharge plan nevertheless includes spill then it is simple to test check that the discharge capacity of the spillways is sufficient. If this is not the case (which should be very rare) then the spillage must be distributed over a longer time period.

Example 5.5 (the situation during the spring flood): Consider the same power plants as in example 5.4, but assume that the reservoirs are filled to 80% at the beginning of the period and that the inflow is 150% higher than the annual mean. Suggest a discharge plan for the period.

Solution: We use the same optimisation problem as in the previous example, but change the following parameters:

$$M_{i,0} = \text{start content of reservoir } i = 0.8\overline{M}_i \approx \begin{cases} 1\,111 & i = 1, \\ 889 & i = 2, \end{cases}$$

$$V_i = \text{local inflow to reservoir } i = \begin{cases} 2.5 \cdot 163 = 407.5 & i = 1, \\ 2.5(164 - 163) = 2.5 & i = 2. \end{cases}$$

The solution to the new problem is shown in table 5.4.

Table 5.4 Discharge plan in example 5.5.

Hour, t	0	1	2	3	4	5	6
M_1 [HE]	1 111.1	1 218.6	1 326.1	1 066.4	1 173.9	1 281.4	1 388.9
M_2 [HE]	888.9	581.4	613.9	1 013.6	1 046.6	1 078.6	1 111.6
Q_1 [HE]		300	300	300	300	300	300
Q_2 [HE]		270	270	270	270	270	270
H_1 [MWh]		62	62	62	62	62	62
H_2 [MWh]		63	63	63	63	63	63
S_1 [HE]		0	0	367.2	0	0	0
S_2 [HE]		340	0	0	0	0	0

In the example above, the reservoirs are filled at the end of the period, even though both power plants are operated at maximum capacity. Spillage is necessary, since the reservoirs are not large enough to store the inflow. In this solution the spillage occurs in hour 3 at Degerforsen and hour 1 at Edensforsen, but there are several other alternatives which are equally good. It is for example possible to spill the same amount of water in each hour, or to change the time for the spillage to some other hours. Thus, this problem has a so-called degenerated solution, which means that there are several other solutions which will provide the same value of the objective function.[10]

Example 5.6 (planning to serve a given load): Once again, consider the same power plants as in example 5.4, but assume that the generation has to follow a certain load instead of adjusting the generation after the price. The forecasted load for the next six hours is 90, 98, 104, 112, 100 and 80 MWh/h respectively. Suggest a discharge plan which covers the forecasted load.

In this example the efficiency should be depending on the discharge. Assume that both power plants have their best efficiency at 75% of maximal discharge. Furthermore, assume that for discharges above 75% of the maximal, the marginal production equivalent is 5% less than at best efficiency. Installed capacity is still obtained at maximal discharge. It can be assumed that stored water can be used for electricity generation at best efficiency sometime in the future.

Solution: Since the load is determined in advance, so is the income from the power sales; hence, the income can be omitted from the objective function. In words the planning problem becomes

maximise	*value of stored water*,
subject to	*hydrological balance in the hydro power plants*,
	delivery of the contracted load.

Most parameters are the same as in the previous example, but we also have to introduce some new ones. The contracted load is given in the problem text:

$$D_t = \text{contracted load for hour } t = \begin{cases} 90 & t = 1, \\ 98 & t = 2, \\ 104 & t = 3, \\ 112 & t = 4, \\ 100 & t = 5, \\ 80 & t = 6. \end{cases}$$

We must also calculate a linear model of the electricity generation. We choose to divide the linear model in two segments, and the break point between them is located to the discharge for best efficiency. The maximal discharge in each segment is calculated using data in the problem text:

$$\overline{Q}_{i,1} = \text{maximal discharge in power plant } i\text{, segment 1} = 0.75\overline{Q}_i =$$

$$= \begin{cases} 225 & i = 1, \\ 202.5 & i = 2, \end{cases}$$

$$\overline{Q}_{i,2} = \text{maximal discharge in power plant } i\text{, segment 2} = \overline{Q}_i - \overline{Q}_{i,1} =$$

$$= \begin{cases} 75 & i = 1, \\ 67.5 & i = 2. \end{cases}$$

The marginal production equivalents can be determined from the fact that we know the generation at maximal discharge and that the marginal production equivalent in segment 2 is 95% lower than the one for segment 1:

$$\overline{H}_i = \mu_{i,1}\overline{Q}_{i,1} + \mu_{i,2}\overline{Q}_{i,2},$$

10. Cf. appendix A, example A.6.

Table 5.5 Discharge plan in example 5.6.

Hour, t	0	1	2	3	4	5	6
M_1 [HE]	694.4	656.0	594.0	532.0	460.4	398.4	407.8
M_2 [HE]	555.6	555.5	565.3	548.4	514.0	515.0	467.1
$Q_{1,1}$ [HE]		201.4	225	225	225	225	153.6
$Q_{1,2}$ [HE]		0	0	0	9.6	0	0
$Q_{2,1}$ [HE]		202.5	202.5	202.5	202.5	202.5	202.5
$Q_{2,2}$ [HE]		0	13.6	40.4	67.5	22.6	0
Q_1 [HE]		201.4	225	225	234.6	225	153.6
Q_2 [HE]		202.5	216.1	242.9	270	225.1	202.5
H_1 [MWh]		42.2	47.1	47.1	49	47.1	32.2
H_2 [MWh]		47.8	50.9	56.9	63	52.9	47.8
S_1 [HE]		0	0	0	0	0	0
S_2 [HE]		0	0	0	0	0	0

$\mu_{i,2} = 0.95\mu_{i,1}$.

Solving these systems of equations, we get

$\mu_{i,j}$ = marginal production equivalent of power plant i, segment j ≈

$$\approx \begin{cases} 0.209 & i = 1, j = 1, \\ 0.199 & i = 1, j = 2, \\ 0.236 & i = 2, j = 1, \\ 0.224 & i = 2, j = 2. \end{cases}$$

The optimisation variables are also more or less the same as in example 5.4 except that we now need two discharge variables per power plant and hour:

$Q_{i,j,t}$ = discharge in power plant i, segment j,during hour t, $i = 1, 2, j = 1, 2,$ $t = 1, \ldots, 6.$

The objective function in this problem is just the value of stored water:

maximise $\lambda_f((\mu_{1,1} + \mu_{2,1})M_1(6) + \mu_{2,1}M_2(6)).$

The hydrological constraints are almost the same as in example 5.4, but we have some more variables for the discharge:

$$M_{1,t} - M_{1,t-1} + Q_{1,1,t} + Q_{1,2,t} + S_{1,t} = V_1, \qquad t = 1, \ldots 6,$$

$$M_{2,t} - M_{2,t-1} + Q_{2,1,t} + Q_{2,2,t} + S_{2,t} - Q_{1,1,t} - Q_{1,2,t} - S_{1,t} = V_2, \qquad t = 1, \ldots 6.$$

The requirement to deliver contracted load is included in the following load balance constraints, which in this case states that the total hydro power generation should equal the contracted load:

$$\sum_{i=1}^{2} \sum_{j=1}^{2} \mu_{i,j} Q_{i,j,t} = D_t, \qquad t = 1, \ldots 6.$$

Finally, we have to consider the following limits:

$$0 \le Q_{i,j,t} \le \overline{Q}_{i,j}, \qquad i = 1, 2, j = 1, 2, t = 1, \ldots, 6,$$

$0 \le S_{i,\,t}$, $\quad i = 1, 2, t = 1, \ldots, 6,$

$0 \le M_{i,\,t} \le \overline{M}_i$, $\quad i = 1, 2, t = 1, \ldots, 6.$

The solution to this planning problem is shown in table 5.5.

5.3 THERMAL POWER

A thermal power plant[11] has variable operation cost which is depending on the used fuel and how the power plant is operated (the efficiency can vary with the loading of the unit). The generation cost is the most important factor when planning the schedule of these power plants, but in we must also consider that fuel is used when starting a thermal power plant and that the power plants cannot increase or decrease their generation at any rate.

5.3.1 General Description of Thermal Power Plants

A thermal power plant generates electricity by combustion of a fuel. The heat generated at the combustion is converted to mechanical energy, which is then transformed to electric energy in a generator.

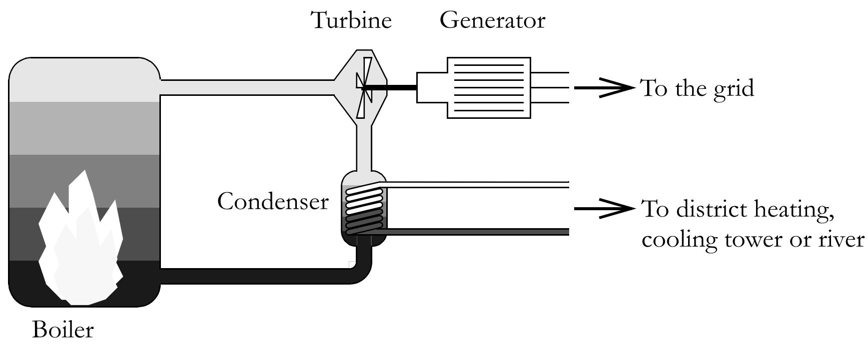
Thermal power plants can be designed in several ways. In the smallest thermal power plants, the generator is powered directly by a regular petrol or diesel engine. This type of thermal power plant is referred to as petrol and diesel generator sets respectively, and may have a rated capacity of about 500 W up to about 50 MW. Another type of thermal power plant is gas turbines. Also these power plants power the electric generator directly by an engine, but in this case it is of the same kind as the jet engines used in aircraft. Gas turbines for electricity generation typically have a rated capacity of about 10 MW to 100 MW. Finally, we have the largest thermal power plants, which are based on a steam cycle (see figure 5.6). The fuel is combusted in the boiler. Water is pumped through the boiler and heated to steam. The hot steam passes a turbine, which powers the electric generator. The steam from the turbine goes back to condenser, where it is cooled down to water again and pumped back to the boiler. A nuclear power plant works according to the same basic principle, but there are a few variants. In a boiling water reactor the water is heated by the warmth of the nuclear fission and then goes directly to the turbine. In a pressurised water reactor the water passes the nuclear reactor under pressure, which causes the water to be heated to several hundred degrees centigrade without vaporising. The hot water is then passing a steam generator, which simply acts as a large heat exchanger; the heat of the water from the reactor transfer is used to vaporise the water in the steam generator. The steam is then passing the turbine, in the same manner as for the other steam cycles (see figure 5.6b).

All thermal power plants generate heat as part of the process. A so-called condensing unit only generates electricity. In a combined heat and power plant (CHP) part of the heat is utilised, for example as district heating. By this means a larger share of the heat content of the fuel is utilised, which results in a higher total efficiency of than for a condensing power plant.

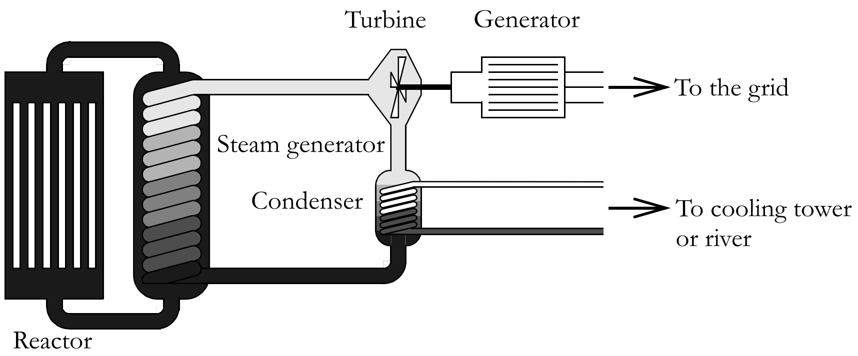
5.3.2 Generation Cost

The electricity generation in a thermal power plant depends on the fuel input. If the efficiency would be constant then a given fuel price would result in a constant operation cost per MWh. However, in practice the efficiency of a thermal power plant is varying depending on the fuel input, which thus means that the generation cost becomes a non-linear function of the power out-

11. This footnote has just been added to maintain the footnote numbering in pace with the Swedish edition.



a) Condensing or CHP plant. A boiling water reactor works in the same way, but the water is passing a nuclear reactor instead of a boiler.



b) Pressurised water reactor.

Figure 5.6 Overview of the layout of larger thermal power plants.

put. To derive a formula for the generation cost of a thermal power plant we start by calculating the necessary fuel input for a particular electricity generation:

$$F(G) = \frac{G}{h \cdot \eta(G)}, \tag{5.12}$$

where

$F(G)$ = fuel input as a function of the generation [tonnes/h or m^3/h],
G = electricity generation [MWh/h],
h = heat contents [MWh/tonnes or MWh/m^3] (cf. table 5.6),
$\eta(G)$ = efficiency at the generation G [%].

If the fuel price is ϕ then we get the following expression for the generation cost:

$$C(G) = \phi F(G) = \phi\frac{G}{h \cdot \eta(G)}. \tag{5.13}$$

A linear model of the generation cost could be obtained by using a piecewise linear approximation of (5.13)—using similar methods as we used in section 5.2.2 to linearise the electricity generation as function of the discharge in a hydro power plant—but it is more common to approximate (5.13) by a polynomial of degree two:

$$C(G) = \alpha + \beta G + \gamma G^2. \tag{5.14}$$

Table 5.6 Data of some common fuels.

Fuel	Heat contents [kWh/kg]	Density [kg/m^3]
Coal, coke	4.5 - 9.0	450 - 800
Diesel oil	11.9	840
Petrol	11.9	730
Fossil gas	14.4	0.75
Ethanol	7.5	790
Wood fuels	5.3[a]	310 - 470
Uranium	880 000	18 680

a. Per kg dry fuel.

In a LP problem we cannot model the generation cost by a second degree function, but as the coefficient of the quadratic term, γ, usually is far smaller than the linear coefficient, β, it is in many cases to approximate γ by zeros, which results in a linear function.

5.3.3 Operation Constraints

When planning the operation of a thermal power plant it is necessary to consider that ta change in the fuel input does not immediately affect the electricity generation. How fast a thermal power plant can respond to changes depend primarily on its size. Petrol and diesel generator sets are as fast as similar engines in cars; the time scale is thus about seconds. Gas turbines and steam cycle power plants are slower. In particular, it takes a certain time before these power plants can start their generation; gas turbines may take some ten minutes to be started, whereas the largest thermal power plants may require some hours.

In this section we will give a few examples of how operation constraints can be included in the short-term planning problem of thermal power plants.

Start-up Costs

Thermal power plants require a certain fuel input to start the process or to prevent that the plant is damaged. It is therefore common that a thermal power plant in addition to the upper limit of the electricity generation (which is decided by the installed capacity) also have a lower limit when the power plant has been committed. This requirement must be considered in the model of the electricity generation of a thermal power plant.

Consider a thermal power plant during a certain hour. If the power plant is committed in this hour then the upper limit of the generation is equal to the installed capacity; otherwise, it is equal to zero. This requirement can be expressed as

$$G_{g,t} \le u_{g,t} \cdot \overline{G}_g, \tag{5.15}$$

where

$G_{g,t}$ = generation in power plant g during hour t,
$u_{g,t}$ = unit commitment in power plant g during hour t (1 if the power plant is committed, 0 otherwise),
$\overline{G}_g$ = maximal generation when power plant g is committed.

Hence, to determine whether or not the power plant has been committed during a certain period of the short-term planning problem, we have to introduce a binary integer variable, $u_{g,t}$. Using this variable we may also represent the lower limit of the electricity generation:

$$G_{g,\,t} \geq u_{g,\,t} \cdot \underline{G}_g, \tag{5.16}$$

where

$\underline{G}_g$, = minimal generation when power plant g is committed.

When the power plant has been committed the lower limit is equal to $\underline{G}_{g,\,t}$ and when the power plant is off-line the lower limit is equal to zero. Combined, the two equations (5.15) and (5.16) forces the generation $G_{g,\,t}$ to the interval $[\underline{G}_{g,\,t}, \overline{G}_{g,\,t}]$ when $u_{g,\,t}$ equals one (i.e., when the power plant is committed). When $u_{g,\,t}$ is equal to zeros, the only possibility to fulfil both constraints is to set $G_{g,\,t} = 0$.

Starting up a thermal power plant results in a certain cost, as some fuel is used to heat the power plant to operational temperature. In small thermal power plants, this cost is low, but in larger units it might be significant. The amount of fuel necessary to heat the boiler depends on its current temperature, which in its turn is depending on how long it was since the power plant was taken off-line. The start-up cost can be written as t

$$C_{start\text{-}up}(t) = C_{cold\text{-}start}(1 - e^{-t/\tau}) + C_{fixed}, \tag{5.17}$$

where

$C_{start\text{-}up}(t)$ = start-up cost after an off-line period of t hours,
$C_{cold\text{-}start}$ = start-up cost when the boiler is cooled to room temperature,
τ = thermal time constant of the boiler,
C_{fixed} = fixed start-up cost.

In (5.17) we also have a fixed cost, which for example can include extra staff and maintenance costs for start-up of the power plant.

Example 5.7 (start-up cost in a thermal power plant): Assume that the cold-start cost of a particular thermal power plant is 1 000 ¤, that the thermal time constant of the boiler is 2 hours and that there are no fixed start-up costs. How many hours after the unit has been taken off-line will it take before the start-up cost is 90% of the cold-start cost? How large is the start-up cost for the intermediate hours? (Assume that the power plant always is started after a number of complete hours.)

Solution: We want the time t such that $C_{cold\text{-}start}(1 - e^{-t/\tau}) \geq 0.9C_{cold\text{-}start}$, i.e.,

$$t \geq -2\ \ln(0.1) = 4.6,$$

The result is therefore that the start-up cost exceeds 900 ¤ after five hours.

The start-up cost when the power plant is started after 1, …, 5 hours is given by

$$C_{start\text{-}up}(1) = 1\,000(1 - e^{-1/2}) \approx 393 ¤,$$
$$C_{start\text{-}up}(2) = 1\,000(1 - e^{-2/2}) \approx 632 ¤,$$
$$C_{start\text{-}up}(3) = 1\,000(1 - e^{-3/2}) \approx 777 ¤,$$
$$C_{start\text{-}up}(4) = 1\,000(1 - e^{-4/2}) \approx 865 ¤,$$
$$C_{start\text{-}up}(5) = 1\,000(1 - e^{-5/2}) \approx 918 ¤.$$

If the power plant only will be off-line for a short time i might be profitable to not stop the combustion in the boiler even though no electricity is generated (so-called banking). In this case, the fuel input is kept just above what is necessary to maintain the operation temperature of the boiler. This fuel input can be assumed to be constant, which makes the start-up cost directly proportional to the time the power plant has been off-line:

$$C_{start\text{-}up} = C_{banking} \cdot t \tag{5.18}$$

Example 5.8 (banking): Consider the same power plant as in example 5.7 and assume that banking costs 250¤/h. How long must the power plant be off-line to make it more profitable to stop the combustion in the boiler and let it start cooling down?

Solution: We want the time t such that $1\,000(1 - e^{-t/2}) = 250t$. This equation must be solved numerically; then we find that $t \approx 3.19$ h, i.e., if the power plant is to stay off-line at most three hours, banking is preferable, but if it should be off-line four hours or more then it is better to let the boiler start cooling.

To include the start-up cost in the objective function, we need to introduce more binary variables, which are equal to one the hour when the power plant is started and equal to zero the other hours. As we seen above, the start-up cost depends on the time the power plant has been off-line and we therefore need to introduce separate start-up variables for the off-line times we wish to consider. For the sake of simplicity, let us assume that we only need to differentiate between off-line times of one hour or at least two hours.[12] Introduce the following variables:

$s^*_{g,t}$ = start-up of power plant g during hour t after one hour down time (1 if the power plant is started after one hour down time, otherwise 0),

$s^{**}_{g,t}$ = start-up of power plant g during hour t after at least two hours down time (1 if the power plant is started after at least two hours down time, otherwise 0).

The total start-up cost during the planning period is then expressed as

$$\sum_{g \in \mathcal{G}} \sum_{t \in \mathcal{T}} (C^*_g s^*_{g,t} + C^{**}_g s^{**}_{g,t}), \tag{5.19}$$

where

C^*_g = start-up cost when power plant g is started after one hour down time,

C^{**}_g = start-up cost when power plant g is started at least two hours down time.

As there is a cost associated to non-zero start-up variables, the LP solver will try to set all start-up variables to equal zero. Therefore, special constraints are necessary to guarantee that the start-up variables are given the right value whenever the power plant is started:

$$s^{**}_{g,t} \ge u_{g,t} - u_{g,t-1} - u_{g,t-2}, \tag{5.20}$$

$$s^*_{g,t} \ge u_{g,t} - u_{g,t-1} - s^{**}_{g,t}. \tag{5.21}$$

Table 5.7 demonstrates that these constraints provide the desired results. (Please note that we must know the state of the power plants for the two hours before the first hour of the planning period.) The right hand side of (5.20) is the lower limit of the start-up variable $s^{**}_{g,t}$ and this lower limit must equal one only if the power plant is committed in hour t and has been off-line during the previous two hours. As $s^{**}_{g,t}$ is a binary variable, the start-up variables must equal one in these cases. In all other cases the lower limit is zero or negative, and then the start-up variable will equal zero in the optimal solution.

In a similar manner is the right hand side of (5.21) the lower limit for the start-up variable $s^*_{g,t}$. In this case the lower limit is equal to one only if the power plant is committed in hour t and has been off-line the previous hour. Moreover, the start-up variable $s^{**}_{g,t}$ must equal zeros, as both start-up variables of a power plant cannot be equal to one during the same hour.

12. The reasoning that follows is of course easy to modify to take other possible off-line times into account.

Table 5.7 Verification of the constraints (5.20) and (5.21).

	Hour					
	–1	0	1	2	3	4
Case I (one hour down time)						
State, u_t	1	1	0	1	1	1
Lower limit for start-up after at least two hours off-line, $u_t - u_{t-1} - u_{t-2}$			–2	0	0	–1
Lower limit for start-up after one hour off-line, $u_t - u_{t-1} - s_t^{**}$			–1	1	0	0
Case II (two hours down time)						
State, u_t	1	1	0	0	1	1
Lower limit for start-up after at least two hours off-line, $u_t - u_{t-1} - u_{t-2}$			–2	–1	1	0
Lower limit for start-up after one hour off-line, $u_t - u_{t-1} - s_t^{**}$			–1	0	0	0
Case III (three hours down time)						
State, u_t	1	1	0	0	0	1
Lower limit for start-up after at least two hours off-line, $u_t - u_{t-1} - u_{t-2}$			–2	–1	0	1
Lower limit for start-up after one hour off-line, $u_t - u_{t-1} - s_t^{**}$			–1	0	0	0

Minimal Up and Down Time

It is not unusual in unit commitment problems to use simpler models of the start-up costs, where the start-up costs are assumed to be constant. As a compensation we can introduce a requirement of the down time of a decommitted unit and the up time of a committed unit. Introduce the following variables:

$s_{g,t}^+$ = start-up variable for power plant g during hour t (1 if the power plant is started before this hour, otherwise 0),
$s_{g,t}^-$ = stop variable for power plant g during hour t (1 if the power plant is stopped before this hour, otherwise 0).

The total start and stop cost can then be written as

$$\sum_{g \in \mathcal{G}} \sum_{t \in \mathcal{T}} (C_g^+ s_{g,t}^+ + C_g^- s_{g,t}^-), \tag{5.22}$$

where

C_g^+ = start-up cost of power plant g,
C_g^- = stop cost of power plant g.

The requirement of minimal up and down time are formulated as follows:

$$s_{g,t}^+ + \sum_{k=t}^{t+t_g^+-1} s_{g,k}^- \le 1, \tag{5.23}$$

$$s_{g,t}^- + \sum_{k=t}^{t+t_g^--1} s_{g,k}^+ \le 1, \tag{5.24}$$

where

t_g^+ = minimal up time of power plant g,
t_g^- = minimal down time of power plant g.

If a power plant is started in hour t, i.e., when $s_{g,t}^+ = 1$, the constraint (5.23) forces $s_{g,t}^-$ to become zero for the next t_g^+ hours; thus, the power plant cannot be taken off-line during this period. The constraint (5.24) works in a similar way.

If the requirement of minimal up and down time should have any effect then we must also have a constraint which forces the start-up and stop variables to get the correct values whenever the state of a unit is changed. This is simply achieved by the following constraint:

$$u_{g,t} - u_{g,t-1} = s_{g,t}^+ - s_{g,t}^-. \tag{5.25}$$

It is easy to verify that this constraint forces $s_{g,t}^+$ and $s_{g,t}^-$ to behave as we desire. If the power plant is committed then $u_{g,t} = 1$ and $u_{g,t-1} = 0$, i.e., the left hand side is equal to one. The only way for the right hand side to become one is to let $s_{g,t}^+ = 1$; hence, the start-up variable is forced to become one when the power plant is started. When the power plant is taken off-line the left hand side equals minus one and the only solution then is to choose $s_{g,t}^- = 1$. If the state of the unit remains the same then the left hand side is zero and (5.25) can be fulfilled either by setting $s_{g,t}^+ = s_{g,t}^- = 1$ or by setting $s_{g,t}^+ = s_{g,t}^- = 0$. As start-up and stop of a power plant is associated to a cost according to (5.22), it will be optimal to choose the latter alternative.

Limited Generation Changes

A large thermal power plant cannot increase or decrease its generation at any rate. To increase the generation the boiler must be supplied more thermal energy (to increase the steam flow), which requires some time. It is neither possible to reduce the generation at any rate, as the steam is cooling the boiler and if the flow is interrupted to abruptly then there is a risk of damaging the boiler. In the planning problem it might therefore be necessary to include ramping constraints, which limit how much the generation may change from one hour to another.

Let us start by considering a generation increase. A limitation of the permissible generation increase from one hour to the next means that the difference in generation between one hour and the previous one, i.e., $G_{g,t} - G_{g,t-1}$, must be below a certain maximal generation increase, Δ_g^{G+}. Unfortunately, this requirement is only valid for those hours when the unit has been committed also the previous hour. The first hour after the power plant has been started the generation increase can exceed Δ_g^{G+} if the minimal generation, $\underline{G}_g$, is larger than Δ_g^{G+}. Assume that the maximal generation during the first hour the power plant is being committed is $\overline{G}_g^1$. We may then express the above reasoning as the following constraint:

$$G_{g,t} - G_{g,t-1} \le \Delta_g^{G+} + s_{g,t}^+(\overline{G}_g^1 - \Delta_g^{G+}). \tag{5.26}$$

This constraint is verified in table 5.8. In those cases when the power plant is off-line in hour t, i.e., when $u_{g,t} = 0$, the maximal generation increase is of no interest, as the power plant cannot increase its generation when it is off-line. In the other two cases we see that the maximal generation increase is equal to Δ_g^{G+} if the power plant was committed the previous hour too, and that it becomes equal to maximal generation capacity the first hour after start-up if the power plant was off-line the previous hour.

A similar constraint is applied for generation decrease:

$$G_{g,t-1} - G_{g,t} \le \Delta_g^{G-} + s_{g,t}^-(\overline{G}_g - \Delta_g^{G-}). \tag{5.27}$$

Table 5.8 Verification of the constraints (5.26) and (5.27).

State of the previous hour, $u_{g,\,t-1}$	State of the hour at hand, $u_{g,\,t}$	Maximal generation increase, $\Delta_g^{G+} + s_{g,\,t}^{+}(\overline{G}_g^1 - \Delta_g^{G+})$	Maximal generation decrease, $\Delta_g^{G-} + s_{g,\,t}^{-}(\overline{G}_g - \Delta_g^{G-})$
0	0	Δ_g^{G+}	Δ_g^{G-}
0	1	$\overline{G}_g^1$	Δ_g^{G-}
1	0	Δ_g^{G+}	$\overline{G}_g$
1	1	Δ_g^{G+}	Δ_g^{G-}

5.3.4 Some Unit Commitment Problems

In the previous sections we have shown how to model the generation cost of thermal power plants, and the operation constraints of larger thermal power plants. With that, it is time to give a few examples how these models can be applied in unit commitment problems:

Example 5.9 (planning of sales to a power pool). Kraftbolaget AB owns the thermal power plant Sotinge. The power plant has three blocks, i.e., separate boilers and turbines; data for these are given in table 5.9. The electricity generation of Sotinge is sold to the power pool ElKräng. Kraftbolaget AB are now about to submit bids from midnight to midday the following day. A forecast of the electricity prices for this period is given in table 5.10. Assume that the start-up cost is independent of the down time and that the stop cost is negligible. According to the plan for the current day, block I will be committed until midnight, whereas the other two blocks are taken off-line at 8 pm. Suggest a operation plan for the three blocks during the first twelve hours of the next day.

Table 5.9 Data for the fictitious thermal power plant Sotinge.

	Sotinge I	Sotinge II	Sotinge III
Installed capacity [MW]	240	110	170
Minimal generation when committed [MW]	80	40	60
Generation cost [SEK/MWh]	300	310	320
Start-up cost [SEK/start]	25 000	13 000	20 000
Minimal up time [h]	3	3	3
Minimal down time [h]	3	3	3

Table 5.10 Expected electricity prices at ElKräng.

Hour	Price [SEK/MWh]	Hour	Price [SEK/MWh]	Hour	Price [SEK/MWh]
1	316	5	309	9	359
2	313	6	318	10	355
3	306	7	342	11	348
4	299	8	353	12	344

Solution: In words the problem can be formulated as

maximise *income of sold electricity – generation costs – start-up costs*

subject to *limitations in generation capacity,*
minimal up and down times.

Before we start formulating this problem as a MILP problem we make sure that we define the variables and parameters that are used in the problem. We start by defining the blocks as unit 1, 2 and 3 respectively.

The following parameters are given in the problem text:

$$u_{g,0} = \text{commitment of unit } g \text{ at the beginning of the planning period} = \begin{cases} 1 & g = 1, \\ 0 & g = 2, \\ 0 & g = 3, \end{cases}$$

$$\lambda_t = \text{expected electrictiy price hour } t = \begin{cases} 316 & t = 1, & 342 & t = 7, \\ 313 & t = 2, & 353 & t = 8, \\ 306 & t = 3, & 359 & t = 9, \\ 299 & t = 4, & 355 & t = 10, \\ 309 & t = 5, & 348 & t = 11, \\ 318 & t = 6, & 344 & t = 12, \end{cases}$$

$$\overline{G}_g = \text{installed capacity of power plant } g = \begin{cases} 240 & g = 1, \\ 110 & g = 2, \\ 170 & g = 3, \end{cases}$$

$$\underline{G}_g = \text{minimal generation when power plant } g \text{ is committed} = \begin{cases} 80 & g = 1, \\ 40 & g = 2, \\ 60 & g = 3, \end{cases}$$

$$\beta_{Gg} = \text{variable generation cost in power plant } g = \begin{cases} 300 & g = 1, \\ 310 & g = 2, \\ 320 & g = 3, \end{cases}$$

$$C_g^+ = \text{start-up cost in power plant } g = \begin{cases} 25\ 000 & g = 1, \\ 13\ 000 & g = 2, \\ 20\ 000 & g = 3. \end{cases}$$

Then we state the optimisation variables used in the problem:

$G_{g,t}$ = generation in power plant g, hour t, $g = 1, 2, 3$, $t = 1, \ldots, 12$,
$u_{g,t}$ = unit commitment of power plant g during hour t, $g = 1, 2, 3$, $t = 1, \ldots, 12$,
$s_{g,t}^+$ = start-up variable for power plant g hour t, $g = 1, 2, 3$, $t = 1, \ldots, 12$,
$s_{g,t}^-$ = stop variable for power plant g hour t, $g = 1, 2, 3$, $t = 1, \ldots, 12$.

The objective of the planning is to maximise the income of sold electricity minus the variable generation costs and the start-up costs, which gives us the following objective function:

$$\text{maximise} \quad \sum_{t=1}^{12} \sum_{g=1}^{3} ((\lambda_t - \beta_G)G_{g,t} - C_g^+ s_{g,t}^+).$$

The constraints of this problem should limit the generation in each power plant as well as force the power plants to remain on-line for at least three hours after being started and remain off-line for at least three hours after the being stopped. The generation limitations are managed by the following constraints:

$$G_{g,t} - u_{g,t}\overline{G}_g \leq 0, \qquad g = 1, 2, 3, t = 1, \ldots, 12,$$

$$u_{g,t}\underline{G}_g - G_{g,t} \leq 0, \qquad g = 1, 2, 3, t = 1, \ldots, 12.$$

The next constraint sets the relation between unit commitment, start and stop:

$$u_{g,t} - u_{g,t-1} - s^+_{g,t} + s^-_{g,t} = 0, \qquad g = 1, 2, 3, t = 1, \ldots, 12.$$

Then we have the constraints which represents the minimal up and down times respectively:

$$s^+_{g,t} + s^-_{g,t+1} + s^-_{g,t+2} \leq 1, \qquad g = 1, 2, 3, t = 1, \ldots, 10,$$

$$s^+_{g,t} + s^-_{g,t+1} \leq 1, \qquad g = 1, 2, 3, t = 11,$$

$$s^-_{g,t} + s^+_{g,t+1} + s^+_{g,t+2} \leq 1, \qquad g = 1, 2, 3, t = 1, \ldots, 10,$$

$$s^-_{g,t} + s^+_{g,t+1} \leq 1, \qquad g = 1, 2, 3, t = 11.$$

Finally we state the variable limits. Notice that the upper and lower limits of $G_{g,t}$ is controlled by the constraints for generation limitations; hence, $G_{g,t}$ can be considered as a free variable.

$$u_{g,t} \in \{0, 1\}, \qquad g = 1, 2, 3, t = 1, \ldots, 12,$$

$$s^+_{g,t} \in \{0, 1\}, \qquad g = 1, 2, 3, t = 1, \ldots, 12,$$

$$s^-_{g,t} \in \{0, 1\}, \qquad g = 1, 2, 3, t = 1, \ldots, 12.$$

Solving the optimisation problem above yield the operation plan displayed in table 5.11.

Table 5.11 Operation plan in example 5.9.

Hour, t	1	2	3	4	5	6	7	8	9	10	11	12
$G_{1,t}$ [MW]	240	240	240	80	240	240	240	240	240	240	240	240
$G_{2,t}$ [MW]	110	110	40	40	40	110	110	110	110	110	110	110
$G_{2,t}$ [MW]	0	0	0	0	0	0	170	170	170	170	170	170
$u_{1,t}$	1	1	1	1	1	1	1	1	1	1	1	1
$u_{2,t}$	1	1	1	1	1	1	1	1	1	1	1	1
$u_{3,t}$	0	0	0	0	0	0	1	1	1	1	1	1
$s^+_{1,t}$	0	0	0	0	0	0	0	0	0	0	0	0
$s^+_{2,t}$	1	0	0	0	0	0	0	0	0	0	0	0
$s^+_{3,t}$	0	0	0	0	0	0	1	0	0	0	0	0
$s^-_{1,t}$	0	0	0	0	0	0	0	0	0	0	0	0
$s^-_{2,t}$	0	0	0	0	0	0	0	0	0	0	0	0
$s^-_{3,t}$	0	0	0	0	0	0	0	0	0	0	0	0

In the example above we see that two of the blocks are committed even when the electricity price is less than the variable generation cost. The explanation is of course that it is so expensive to restart the units after decommitting them that it is cheaper to generate some electricity at a loss. Therefore, the generation is reduced as much as possible without stopping the unit.

Example 5.10 (planning of bilateral trade and sales to power pool): Consider Sotinge power plant again, but assume that Kraftbolaget AB also has a bilateral contract to sell 200 MWh/h to AB Elleverantören.

In this example a variable start-up cost should be used instead of minimal up- and down times. Assume that the start-up costs given in example 5.9 correspond to the cold-start cost and that the thermal time constants are 2.5 hours in block I, 1.5 hours in block II, and 2 hours in block III. After three hours of down time, the start-up cost is assumed to equal the cold-start cost.

Solution: The income of the firm power contract does not influence the operation of the power plants; therefore, it can be omitted from the objective function. In words the planning problem may then be expressed as

maximise	*income of sold electricity – generation costs – start-up costs*
subject to	*limitations in generation capacity,*
	start-up constraints,
	load balance.

Most of the parameters are the same as in example 5.9, but we also need to introduce a few new ones. The firm sales and the thermal time constants are given above:

D = firm power sales = 200,

$$\tau = \text{thermal time constants} = \begin{cases} 2.5 & g = 1, \\ 1.5 & g = 2, \\ 2 & g = 3. \end{cases}$$

In the previous example it was sufficient to know the state of the units the hour before the beginning of the planning period. In this case we must know the state for three hours last three hours when calculating the start-up costs of the power plants.

$u_{g,t}$ = unit commitment of power plant g the hours before the start of the planning

$$\text{period} = \begin{cases} 1 & g = 1,\ t = -2, -1, 0, \\ 0 & g = 2, 3,\ t = -2, -1, 0, \end{cases}$$

The cold-start cost and the start-up cost after at least three hours down time is assumed to be the same as the cost in table 5.9. The start-up cost for shorter down times are calculated according to (5.17):

$$C_g^{***} = \text{start-up cost after at least three hours down time} = \begin{cases} 25\,000 & g = 1, \\ 13\,000 & g = 2, \\ 20\,000 & g = 3. \end{cases}$$

C_g^{**} = start-up cost after two hours down time =

$$= (1 - e^{-2/\tau})C_g^{***} \approx \begin{cases} 13\,767 & g = 1, \\ 9\,573 & g = 2, \\ 12\,642 & g = 3. \end{cases}$$

C_g^* = start-up cost after one hour down time =

$$= (1 - e^{-1/\tau})C_g^{***} \approx \begin{cases} 8\,242 & g = 1, \\ 6\,326 & g = 2, \\ 7\,869 & g = 3. \end{cases}$$

The start and stop variables of example 5.9 are not used here, but we will introduce three separate start-up variables for start-up after a given down time. We must also introduce variables for the trade at ElKräng. Hence, we get the following optimisation variables in this problem:

$G_{g,\,t}$ = generation in power plant g, hour t, $g = 1, 2, 3$, $t = 1, \ldots, 12$,
$u_{g,\,t}$ = unit commitment in power plant g during hour t, $g = 1, 2, 3$, $t = 1, \ldots, 12$,
$s_{g,\,t}^{***}$ = start-up of power plant g in hour t after at least three hours down time, $g = 1, 2, 3$, $t = 1, \ldots, 12$,
$s_{g,\,t}^{**}$ = start-up of power plant g in hour t after two hours down time, $g = 1, 2, 3$, $t = 1, \ldots, 12$,
$s_{g,\,t}^{*}$ = start-up of power plant g in hour t after one hour down time, $g = 1, 2, 3$, $t = 1, \ldots, 12$,
p_t = purchase from ElKräng hour t, $t = 1, \ldots, 12$,
r_t = sales to ElKräng hour t, $t = 1, \ldots, 12$.

The objective function is now to maximise the income of sold electricity minus the costs of purchase, generation and start-ups:

$$\text{maximise} \qquad \sum_{t=1}^{12}\left(\lambda_t(r_t - p_t) - \sum_{g=1}^{3}(\beta_{Gg}G_{g,\,t} + C_g^{***}s_{g,\,t}^{***} + C_g^{**}s_{g,\,t}^{**} + C_g^{*}s_{g,\,t}^{*})\right).$$

The generation limitations are the same as before:

$$G_{g,\,t} - u_{g,\,t}\overline{G}_g \le 0, \qquad g = 1, 2, 3,\ t = 1, \ldots, 12,$$

$$u_{g,\,t}\underline{G}_g - G_{g,\,t} \le 0, \qquad g = 1, 2, 3,\ t = 1, \ldots, 12.$$

Then we have the constraints that make sure the start-up variables are given the correct values:

$$u_{g,\,t} - u_{g,\,t-1} - u_{g,\,t-2} - u_{g,\,t-3} - s_{g,\,t}^{***} \le 0, \qquad g = 1, 2, 3,\ t = 1, \ldots, 12,$$

$$u_{g,\,t} - u_{g,\,t-1} - u_{g,\,t-2} - s_{g,\,t}^{***} - s_{g,\,t}^{**} \le 0, \qquad g = 1, 2, 3,\ t = 1, \ldots, 12,$$

$$u_{g,\,t} - u_{g,\,t-1} - s_{g,\,t}^{***} - s_{g,\,t}^{**} - s_{g,\,t}^{*} \le 0, \qquad g = 1, 2, 3,\ t = 1, \ldots, 12.$$

We also need load balance constraints to guarantee that the company's own generation plus purchase are equal to the firm power contract plus sales to the power pool:

$$\sum_{g=1}^{3} G_{g,\,t} + p_t - r_t = D_t, \qquad t = 1, \ldots, 12.$$

Finally we state the variable limits. Notice that the upper and lower limits of $G_{g,\,t}$ is controlled by the constraints for generation limitations; hence, $G_{g,\,t}$ can be considered as a free variable.

$u_{g,\,t} \in \{0, 1\}$, $g = 1, 2, 3, t = 1, \ldots, 12$,

$s^{***}_{g,\,t} \in \{0, 1\}$, $g = 1, 2, 3, t = 1, \ldots, 12$,

$s^{**}_{g,\,t} \in \{0, 1\}$, $g = 1, 2, 3, t = 1, \ldots, 12$,

$s^{*}_{g,\,t} \in \{0, 1\}$, $g = 1, 2, 3, t = 1, \ldots, 12$,

$0 \le p_t$, $t = 1, \ldots, 12$,

$0 \le r_t$, $t = 1, \ldots, 12$.

Solving the optimisation problem above yield the operation plan displayed in table 5.12.

Table 5.12 Operation plan in example 5.10.

Hour, t	1	2	3	4	5	6	7	8	9	10	11	12
$G_{1,\,t}$ [MW]	240	240	240	80	240	240	240	240	240	240	240	240
$G_{2,\,t}$ [MW]	110	110	40	40	40	110	110	110	110	110	110	110
$G_{2,\,t}$ [MW]	0	0	0	0	0	0	170	170	170	170	170	170
$u_{1,\,t}$	1	1	1	1	1	1	1	1	1	1	1	1
$u_{2,\,t}$	1	1	1	1	1	1	1	1	1	1	1	1
$u_{3,\,t}$	0	0	0	0	0	0	1	1	1	1	1	1
$s^{***}_{1,\,t}$	0	0	0	0	0	0	0	0	0	0	0	0
$s^{***}_{2,\,t}$	1	0	0	0	0	0	0	0	0	0	0	0
$s^{***}_{3,\,t}$	0	0	0	0	0	0	1	0	0	0	0	0
$s^{**}_{1,\,t}$	0	0	0	0	0	0	0	0	0	0	0	0
$s^{**}_{2,\,t}$	0	0	0	0	0	0	0	0	0	0	0	0
$s^{**}_{3,\,t}$	0	0	0	0	0	0	0	0	0	0	0	0
$s^{*}_{1,\,t}$	0	0	0	0	0	0	0	0	0	0	0	0
$s^{*}_{2,\,t}$	0	0	0	0	0	0	0	0	0	0	0	0
$s^{*}_{3,\,t}$	0	0	0	0	0	0	0	0	0	0	0	0
p_t	0	0	0	80	0	0	0	0	0	0	0	0
r_t	150	150	80	0	800	150	320	320	320	320	320	320

5.4 DUAL VARIABLES

Usually when solving an LP problem, the solution also includes the value of the so-called *dual variables* (see appendix A, examples A.8-A.10). One dual variable can be obtained for each constraint and for each variable limit. The dual variables show how much the optimal value would change if a small change is made to the right hand side of the corresponding constraint[13] and for small changes of the variable limits. In practice it is not very convenient to use the dual variables in this manner, as it is hard to know in advance whether or not a change is "large" or "small"—if the problem is not extremely large and difficult to solve it is generally preferable to simply solve the

13. This requires that the problem is formulated in standard form, i.e., all terms including optimisation variables should be in the left hand side, while the right hand side only includes known parameters.

new problem directly.

The interesting about dual variables is rather that they do not just have a mathematical meaning, but they also have practical interpretation in reality. The objective function of a short-term planning problem corresponds to the profit or costs during the planning period; hence, the dual variables state how much the profit or costs will change for small changes of the prerequisites of the planning problem. We may therefore consider the dual variables as a sort of derivatives of benefit and cost functions. This means that the dual variables can be interpreted as various prices,[14] as demonstrated in the following examples:

Example 5.11 (dual variables of hydrological balance constraints). Which unit has the dual variables of a hydrological balance constraint? Are these dual variables going to be larger than zero, equal to zero or lesser than zero? What practical interpretation can be given these dual variables? Assume that the short-term planning problem is formulated as maximisation of the profits.

Solution: The unit of the objective function is currency (¤) and the unit of the hydrological balance constraint is HE. The dual variables thus states how the profits will change for a change in the right hand side of the problem when stated in standard form, which means that only the local inflow will be found in the right hand side. The unit must therefore be ¤/HE.

During normal conditions an increased inflow will result in increased income (either by generating more electricity or by having more water stored in the reservoirs at the end of the planning period), which means that the dual variables must be larger than zero. However, this is not valid if the inflow is so large that water has to be spilled; in this case the income is not affected by increased inflow and the dual variables get the value zero.

In practice these dual variables correspond to a price which might be associated to the water in a certain reservoir. This price is referred to as *water value*.

Example 5.12 (dual variables of load balance constraints). Which unit has the dual variables of a load balance constraint? Are these dual variables going to be larger than zero, equal to zero or lesser than zero? What practical interpretation can be given these dual variables? Assume that the short-term planning problem is formulated as maximisation of the profits.

Solution: The unit of the objective function is currency (¤) and the unit of the hydrological balance constraint is MWh. The dual variables thus states how the profits will change for a change in the right hand side of the problem when stated in standard form, which means that only the load will be found in the right hand side. The unit must therefore be ¤/MWh.

If the load increases then the value of the objective function decreases. Admittedly, the income of sold electricity increases during the planning period, but if the load is given in advance then this income are also given in advance and will not be accounted for in the objective function. On the other hand is the generation cost increasing (in thermal power plants) or there will be a less amount of water stored at the end of the period. The dual variable is therefore negative or—in some rare cases—equal to zero.

In practice these dual variables correspond to the marginal production cost of the player performing the planning. If there is perfect competition in the market then this value also corresponds to the electricity price.

14. Therefore, economic literature frequently uses the notion “shadow prices” instead of dual variables.

The dual variables are also useful for checking that a correct optimal solution has been found for the planning problem. The dual variables of interest are in general less than the primal variables of interest; hence, it is less work to study the dual variables. Moreover, it is easier to check that the prices in the system are reasonable than to study the relation of the primal variables in detail. This is illustrated in the following examples:

Example 5.13 (checking the operation plan). Consider the planning problem in example 5.4, but assume that the electricity prices of the six hours are forecasted to 213, 225, 229, 227, 230 and 232 SEK/MWh respectively and that the future electricity generation is assumed to be sold for 228 SEK/MWh. Assume that a computer program has been written to solve this problem and that the program suggests the following discharge plan:

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
$M_{1,t}$ [HE]	857.4	1 020.4	883.4	1346.4	1 388.9	1 388.9
$M_{2,t}$ [HE]	556.6	557.6	537.0	538.0	269.0	0
$Q_{1,t}$ [HE]	0	0	0	0	120.6	163.0
$Q_{2,t}$ [HE]	0	0	21.6	0	270.0	270.0
$S_{1,t}$ [HE]	0	0	0	0	0	0
$S_{2,t}$ [HE]	0	0	0	0	0	0

The dual variables of the hydrological constraints have the following values:

	$t = 1$	$t = 2$	$t = 3$	$t = 4$	$t = 5$	$t = 6$
Degerforsen	47.5	47.5	47.5	47.5	47.5	47.5
Edensforsen	53.4	53.4	53.4	53.4	53.4	53.4

The optimal value of the objective function is 183 136.8. Is the suggested discharge plan really optimal?

Solution: No, there is something wrong with the discharge plan. The water value of Degerforsen should be higher than the water value in Edensforsen, as the water in the reservoir of Degerforsen can be used for electricity generation in both power plants, whereas the water in the reservoir of Edensforsen only can be used in that particular power plant.

Where is then the error? We can try to substitute the variable values of the discharge plan into the objective function and a few of the constraints as they were formulated in example 5.4:

Objective function:

$$\sum_{t=1}^{6} \lambda_T \sum_{i=1}^{2} \gamma_i Q_{i,t} + \lambda_f((\gamma_1 + \gamma_2)M_1(6) + \gamma_2 M_2(6)) = \dots = 183\ 136.8.$$

Hydrological balance of Degerforsen hour 1:

RHS: $M_{1,1} - M_{1,0} + Q_{1,1} + S_{1,1} = 857{,}4 - 694{,}4 + 0 + 0 = 163.$

LHS: $V_1 = 163.$

Hydrological balance of Edensforsen hour 1:

RHS: $M_{2,1} - M_{2,0} + Q_{2,1} + S_{2,1} - Q_{1,1} - S_{1,1} = 556.6 - 555.6 + 0 + 0 - 0 - 0 = 1.$

LHS: $V_2 = 1.$

Hydrological balance of Degerforsen hour 6:

RHS: $M_{1,6} - M_{1,5} + Q_{1,6} + S_{1,6} = 1\,388.9 - 1\,388.9 + 163.0 + 0 = 163.$

LHS: $V_1 = 163.$

Hydrological balance of Edensforsen hour 6:

RHS: $M_{2,6} - M_{2,5} + Q_{2,6} + S_{2,6} - Q_{1,6} - S_{1,6} =$
$= 0 - 269.0 + 270 + 0 - 163.0 - 0 = 162.$

LHS: $V_2 = 1.$

Apparently something is wrong in the hydrological balance constraint of Edensforsen. It would be a good idea to check that the water discharged in Degerforsen really is accounted for in the reservoir of Edensforsen.

EXERCISES

5.1 At 10:00 a certain reservoir holds 1 800 000 m^3. The mean local inflow to the reservoir is 40 m^3/s between 10:00 and 11:00. No electricity is generated in the power plant during this time and there is no spillage. How much is stored in the hydro reservoir at 11:00. The answer should be given in HE!

5.2 A particular hydro power plant has its best efficiency at the discharge 125 HE and when the power plant generates 80 MW. Which is the maximal production equivalent?

5.3 A hydro power plant has the installed capacity 80 MW, which is generated at maximal discharge (200 HE). The power plant has two turbines, which means that for some discharge it achieves its best efficiency and there is another discharge representing a local maxima in the relative efficiency curve. The best efficiency is reached for the discharge 100 HE and the production equivalent is then 0.42 MWh/HE. The local best efficiency point is reached for the discharge 160 HE, when the production equivalent is 0.4125 MWh/HE. Create a piecewise linear model with three segments for the power generation in this power plant.

5.4 In a particular hydro power plant the maximal discharge is 125 HE. At this discharge the power plant generates its installed capacity, which is 40.5 MW. The best efficiency is reached when the discharge is 80% of the maximal discharge. Assume that for higher discharges the marginal production equivalent is 10% less than for discharges below 80% of the maximal discharge. Create a piecewise linear model of the power generation in the power plant.

5.5 A given power plant has its best efficiency for the discharge 80 HE. The maximal discharge in the power plant is 200 HE and the relative efficiency is then 94%. This is enough to generate the installed capacity, which is 94 MW. Create a piecewise linear model of the power generation in the power plant.

5.6 Consider the hydro power plants in figure 5.7. The system has four reservoirs and three power plants. From Vattnet water can be discharged through the underground power plant Språnget; this water is then released into the lake Sjön. If necessary, water from Vattnet can also be spilled through the original river bed and will then reach the reservoir of the Fallet power plant.

The following symbols have been introduced in a short-term planning for these

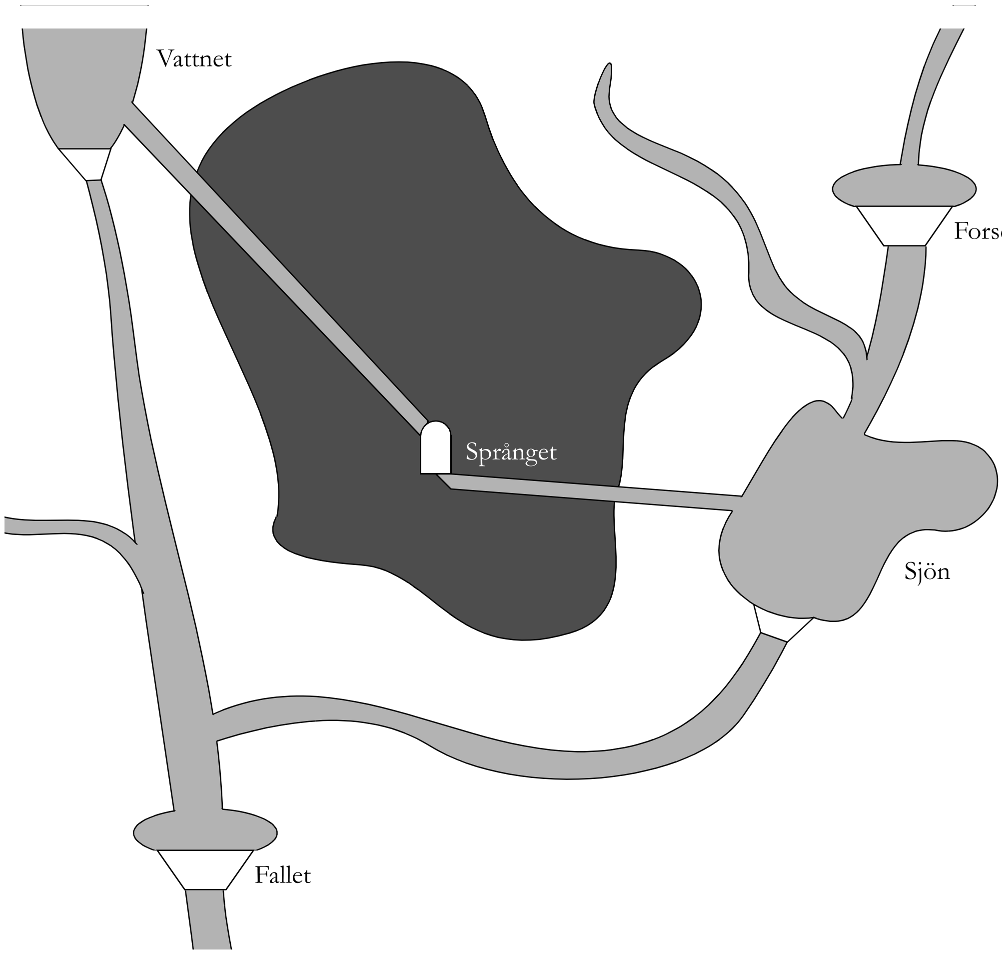


Figure 5.7 The river system in exercise 5.6.

power plants:

Indices for power plants and reservoirs: Vattnet/Språnget 1, Forsen 2, Sjön 3, Fallet 4.

γ_i = expected future production equivalent for water stored in reservoir i, i = 1, 2, 3, 4,
λ_t = expected electricity price during hour t, t = 1, …, 24,
λ_{25} = expected electricity price after the end of the planning period,
$M_{i,\,0}$ = contents of reservoir i at the beginning of the planning period, i = 1, 2, 3, 4,
$M_{i,\,t}$ = contents of the reservoir i at the end of hour t, i = 1, 2, 3, 4, t = 1, …, 24,
$\mu_{i,\,j}$ = marginal production equivalent for power plant i, segment j, i = 1, 2, 3, 4, j = 1, 2,
$Q_{i,\,j,\,t}$ = discharge in power plant i, segment j, during hour t, i = 1, 2, 3, 4, j = 1, 2, t = 1, …, 24.
$S_{i,\,t}$ = spillage from reservoir i during hour t, i = 1, 2, 3, 4, t = 1, …, 24,
$V_{i,\,t}$ = local inflow to reservoir i during hour t, i = 1, 2, 3, 4, t = 1, …, 24.

a) Which symbols denote optimisation variables and parameters respectively?

b) Formulate the objective function if the purpose of the planning is to maximise the income of generated hydro power plus the value of stored water.

c) Formulate the hydrological constraints of this system. The delay time between the power plants can be neglected.

5.7 AB Vattenkraft owns two hydro power plants, located as shown in figure 5.8. Data of the hydro power plants are given in table 5.13. The company expects that the electricity price the coming five hours will be 200, 210, 210, 220 and 210 SEK/MWh respectively. After this period the average electricity price is estimated to 200 SEK/MWh and stored water is assumed to be used for generation at best efficiency. The reservoirs are half-filled at the beginning of the planning period.

Since the river where the power plants are located is an important breeding area for salmon, the Environment Court has judged that the company must build fish ladders which allow fish to pass the dams. The minimum water flow in each fish ladder must be 1 m^3/s; this water can thus not be used for power generation. Moreover, the Environment Court has decided that for tourism reasons the river flow in the river sections downstream each power plant may not be less than 10 m^3/s.

Formulate the planning problem of AB Vattenkraft as a an LP problem. The water delay time between the power plants can be neglected.

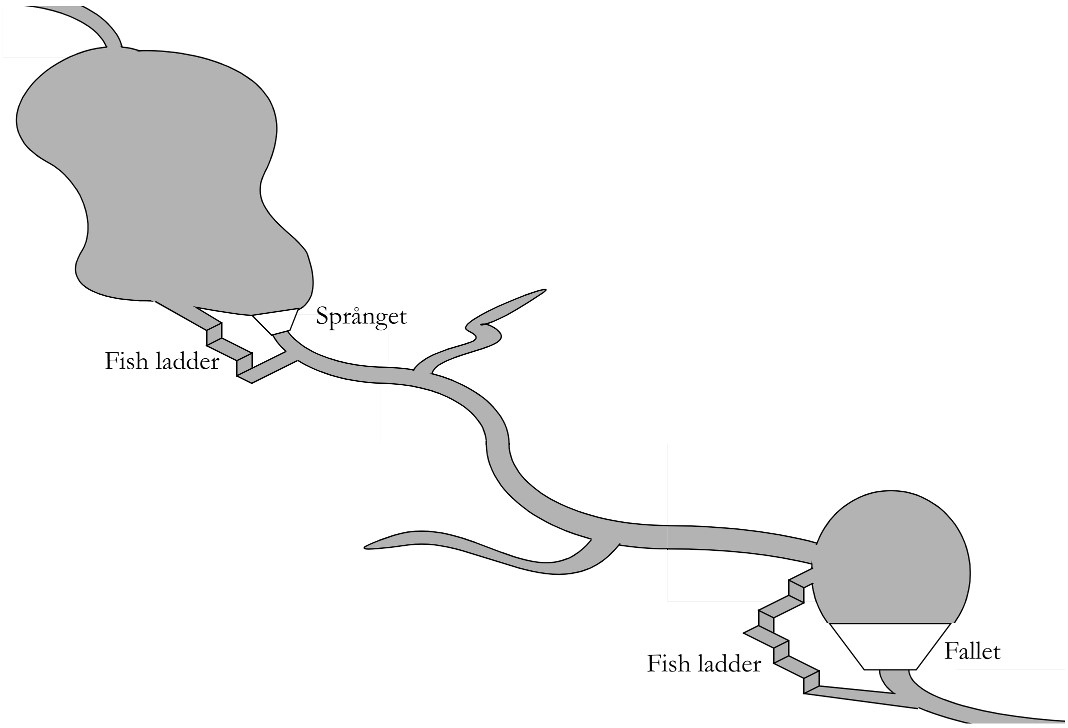


Figure 5.8 The hydro power plants in exercise 5.7.

Table 5.13 Data of the hydro power plants in exercise 5.7.

Power plant	Maximal reservoir contents [HE]	Marginal production equivalents [MWh/HE]	Local inflow [m^3/s]
Spränget	10 000	0.36 for discharges between 0 and 300 HE 0.32 for discharges between 300 and 400 HE	150
Fallet	2 000	0.48 for discharges between 0 and 100 HE 0.46 for discharges between 100 and 200 HE	10

5.8 The thermal power plant Flisinge is fuelled by biomass. The fuel costs 200 ¤/m^3 and has a density of 400 kg/m^3. The heat contents of the fuel is 5 MWh/ton and the efficiency of the power plant is 40%. How large is the variable generation cost in Flisinge?

5.9 Assume that a certain nuclear power plant has an efficiency of 35% and that the fuel costs 27 720 SEK/kg. The heat contents of uranium is given in table 5.6. How large is the variable generation cost?

5.10 A particular thermal power plant has a start-up cost of 50 000 SEK. The thermal time constant is 3 hours and the cost of banking is10 000 SEK/h. Calculate the start-up cost when the power plant is started after a down time of 1, …, 5 hours.

5.11 Consider a short-term planning problem for a thermal power plant. The following optimisation variables have been introduced:

u_t = unit commitment of the power plant during hour t (1 if the power plant is committed during hour t, 0 otherwise),

s_t^+ = start-up of the power plant in the beginning of hour t (1 if the power plant is started before hour t, 0 otherwise).

Formulate the constraints that sets the relation between u_t, u_{t-1} and s_t^+ for hour t. Notice that the constraint must be formulated without using any additional optimisation variables!

5.12 AB Elverket owns a biofuel-fired power plant with three blocks in Stad, as well as the nuclear power plant Strålinge with two reactors. Data of the power plants are given in table 5.14. The start-up costs are assumed to be independent of the down time, but on the other hand it is required that the down time should not be less than three hours.

The electricity generated by the company is sold to the multinational power pool ElKräng and it is assumed that all bids delivered to the pool also will be accepted to the prices stated in table 5.15. Normally all accepted bids receive the same price at ElKräng, but in those cases when there is transmission congestion, the market is divided in so-called price areas. All accepted bids then receive the electricity price valid in the price area where the generation will be fed into the grid. The power plant in Stad is in price area Nord, while the nuclear power plant is in price area Öst.

Formulate the planning problem of AB Elverket as a MILP problem.

Table 5.14 Data of the power plants in exercise 5.12.

Power plant	Installed capacity [MW]	Minimal generation when committed [MW]	Variable generation cost [¤/MWh	Start-up cost [¤]	State before the beginning of the planning period
Stad I	200	50	150	20 000	Decommitted
Stad II	150	30	180	18 000	Decommitted
Stad III	300	80	160	32 000	Decommitted
Strålinge I	600	500	100	40 000	Committed
Strålinge II	650	500	100	40 000	Committed

5.13 Energibolaget AB owns three hydro power plants located as in figure 5.9, as well as the thermal power plant Viken. Data of the hydro power plants are given in table 5.16. The power plant in Viken has an installed capacity of 200 MW and must at least generate 80 MW when committed. The generation cost is 320 SEK/MWh and the start-up cost is assumed to be 15 000 SEK if the down time is one hour and 35 000 SEK if the down time is two hours or more. The carbon dioxide emissions are

Table 5.15 Expected power pool prices in exercise 5.12.

Hour	1	2	3	4	5	6	7	8	9	10	11	12
Price area Nord	116	118	108	108	115	160	210	290	291	291	290	273
Price area Öst	116	118	108	108	115	130	136	148	148	150	145	141
Hour	13	14	15	16	17	18	19	20	21	22	23	24
Price area Nord	272	163	138	137	140	142	137	134	130	120	98	105
Price area Öst	141	139	138	137	140	142	137	134	130	120	98	105

0.33 tonnes/MWh. During start-up the emissions are 15 tonnes if the down time was one hour and 36 tonnes otherwise. Energibolaget AB has an environmental policy which states that they should never emit more than 850 tonnes carbon dioxide a day.

Currently Energibolaget AB sells all of its generation to AB Elhandel. The contracted sales are 350 MWh/h between 8 am and 5 pm, and 150 MWh/h during the remaining hours.

Energibolaget AB now wants to plan the operation for the next 24 hours. Formulate the planning problem of the company as a n MILP problem. Assume that the electricity price after the planning period will be 330 SEK/MWh and that stored water will be used for electricity generation at the maximal marginal production equivalent. The reservoirs are filled to 75% at the beginning of the planning period. The water delay time can be neglected.

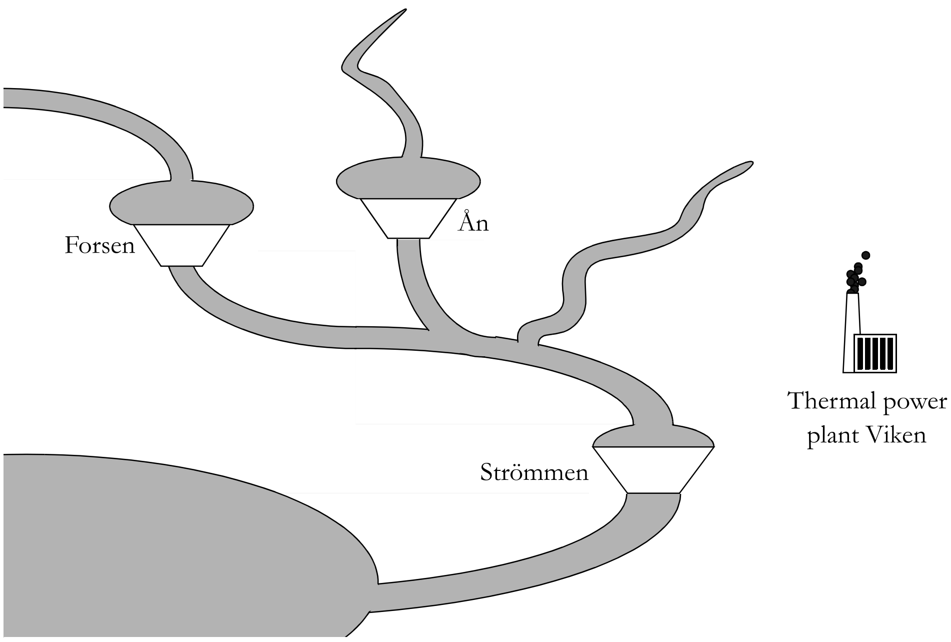


Figure 5.9 The power plants in exercise 5.13.

FURTHER READING

- J. F. Bard, "Short-term Scheduling of Thermal-electric Generators Using Lagrangian Relaxation", *Operations Research*, Vol. 36, No. 5, September-October 1988. — *Example of thermal power plant models.*
- G. W. Chang, Y. D. Tsai, C. Y. Lai & J. S. Chung, "A Practical Mixed Integer Linear Programming Based Approach for Unit Commitment", *2004 IEEE Power Engineering Society General Meeting Proceedings*, Denver 2004. — *Example of thermal power plant models.*

Table 5.16 Data of the hydro power plants in exercise 5.13.

Power plant	Maximal reservoir contents [m^3]	Marginal production equivalents [MWh/HE]	Local inflow [m^3/s]
Forsen	7 200 000	0.28 for discharges between 0 and 150 HE 0.21 for discharges between 150 and 310 HE	125
Ån	2 880 000	0.42 for discharges between 0 and 200 HE 0.32 for discharges between 200 and 400 HE	175
Strömmen	3 600 000	0.34 for discharges between 0 and 180 HE 0.28 for discharges between 180 and 370 HE	30

- O. Nilsson, L. Söder & D. Sjelvgren, "Integer Modelling of Spinning Reserve Requirements in Short Term Scheduling of Hydro Systems", *IEEE Transactions on Power Systems*, Vol. 13, No. 3, August 1998 — *Example of a mixed integer linear model for hydro power.*
- A. J. Wood & B. F. Wollenberg, *Power Generation Operation and Control*, second edition, John Wiley & Sons, 1996 — *General textbook which describes both hydro power and thermal power plants.*

Chapter 6

SIMULATION OF ELECTRICITY MARKETS

There are many contexts in which there is a need to predict the long-term behaviour of an electricity market. The most simple alternative is then to use rough estimates (as for example the simple price model in chapter 3), but to perform a deeper analysis it is necessary to *simulate* the electricity market. The starting point when simulating an electricity market is to make assumptions about the resources of the system (for example generation and transmission capacity), the demand and the regulation controlling the electricity trade. Given these inputs, we try to predict the function of the electricity market.

The choice of model and method in an electricity market simulation is of course depending on the factors which are to be studied. If we for example want to study the impact of various environment regulations then it is necessary to somehow model how the players of the electricity market respond to the various market designs. There is a wide range of possible models—in this compendium we will have to restrict ourselves to the most basic models for electricity market simulation.

We will describe two simulation methods. The first is probabilistic production cost simulation (PPC). This method is limited to a rather simplistic model of the electricity market, as it is based on analytical calculations. The other method is Monte Carlo simulation, which is a more versatile simulation method—the only limitation to the possible models to be simulated by Monte Carlo methods is the computation capacity of the computer used for the simulation. A complete description of Monte Carlo simulation would not fit into this compendium and we will therefore only provide an introduction to the subject.

6.1 PROBLEM DESCRIPTION

Before we address the actual simulation methods it is appropriate to investigate what kind of problems can be analysed by an electricity market simulation. The objective of simulating an electricity market is as mentioned above to predict the behaviour of the electricity market given a set of conditions (inputs). Let us introduce the designation *scenario* for a given situation, where available resources, demand as well as all other factors influencing the electricity market are known. As all these conditions vary more or less randomly in an electricity market, there will in general be an infinite number of possible scenarios. To predict the behaviour of the electricity market we would therefore like to know how the electricity market will work in each scenario.

The conditions in a particular scenario is described by a number of parameters, which we may refer to as *scenario parameters*. Each scenario parameter is a random variable, the probability distribution of which is *known* (consequently, the probability distribution is an input to the simulation).

Exactly which scenario parameters we have depend on which model we have chosen. For the moment, it is sufficient that we collect all scenario parameters in the random vector Y.

The simulation also requires a mathematical model of the electricity market. In this model there are several constants, the value of which is the same in all scenarios. We refer to these constants as *model constants*, and as the scenario parameters, they constitute inputs for the simulation. We may say that the model constants in combination with the structure of the mathematical model forms a mathematical function, g, which shows how the electricity market responds to a particular scenario. In many cases the electricity market is so complex that the function g only can be defined indirectly, for example from the solution of an optimisation problem (we will see an example of this later in this section).

The outputs of the simulation shows how the electricity market behaves in each possible scenario. It is possible to study several characteristics, as for example resource usage and prices. Depending on which properties we want to study, we will have to define a number of *result variables*. For the moment, we collect all result variables in a random vector, X. As the scenario parameters, the result variables are random variables, but the major difference is that the probability distribution of the result variables is *unknown* prior to the simulation.

Thus, the relation between the scenario parameters and the result variables is given by the mathematical model, i.e.,

$$X = g(Y). \tag{6.1}$$

To predict the behaviour of the electricity market is equivalent to calculating the probability distribution of the result variables. In many cases it is inconvenient to consider all possible events in detail (i.e., to consider all possible outcomes of the result variables). It is more appropriate to use simple, easily understood measures, which summarises the most important characteristics of an electricity market. Most straightforward is to define a number of key values—*system indices*—which can be used for comparison of alternative electricity markets. The system indices are in practice statistical measures (generally expectation values) of the result variables. Therefore, we may say that the actual objective of an electricity market simulation is to determine

$$E[X] = E[g(Y)]. \tag{6.2}$$

Below we will give an example of how an electricity market model can be designed and how the results of a simulation can be used for electricity market analysis.

6.1.1 Example of an Electricity Market Model

The electricity market model should, as described above, describe how the electricity market responds to a particular scenario. In many cases this means that we formulate an optimisation problem corresponding to the short-term planning problem the players of the electricity market would have to solve. The model suggested here is a multi-area model intended for Monte Carlo simulation. (However, as we will see later in this chapter, the model of probabilistic production cost simulation is a special case of the multi-area model.) In a Monte Carlo simulation a number of randomly chosen scenarios are investigated and it is observed how the electricity market behaves in those (see section 6.3). These observations are obtained by analysing the solution to the optimisation problem. A Monte Carlo simulation can require that a large number of scenarios are considered; therefore, it is desirable that each optimisation problem can be solved as quickly as possible. In practice this means that we should minimise the number of variables in the problem.

One way to reduce the number of variables, while still modelling different power plants and transmission losses is to use a multi-area model. The power system is then divided in several areas connected transmission lines (see figure 6.1). All power plants and all consumption must be located in a specific area. In the model we assume that there is perfect competition and that all

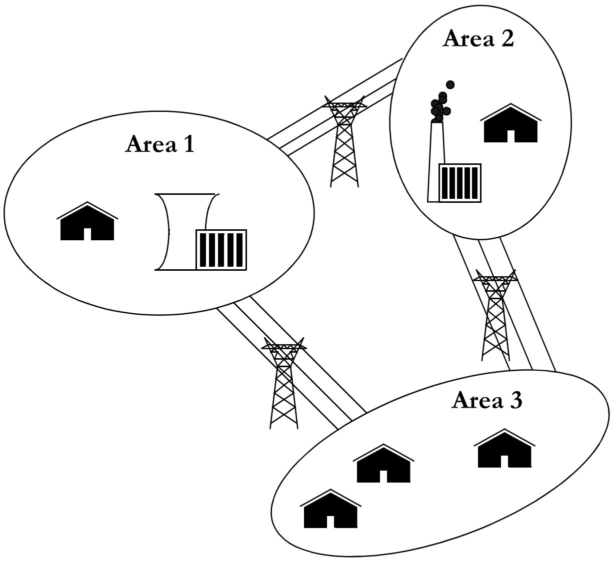


Figure 6.1 The power system in a multi-area problem.

players have perfect information. Moreover, we make the following assumption of the modelled players:

- **Thermal power plants.** The thermal power plants are modelled by a certain available generation capacity and a cost function. As we learned in section 5.3 it is necessary to use integer variables to model the start-up costs of a thermal power plant. Therefore we neglect the start-up costs, because including them would significantly increase the solution time of the multi-area problem. Hence, the cost function is just function of the generation cost. This function could be quadratic, i.e.,

$$C_{Gg}(G_g) = \alpha_{Gg} + \beta_{Gg}G_g + \gamma_{Gg}G_g^2. \tag{6.3}$$

  Several thermal power plants located in the same area might be combined into an equivalent power plant.

- **Non-dispatchable power plants.** By non-dispatchable power plants we refer to such power plants where the generation capacity depend on some factor beyond human control. Examples of non-dispatchable units are wind power plants, hydro power plants without reservoirs and photovoltaics. The designation non-dispatchable is actually slightly misleading, because it is always possible to reduce the generation of these power plants (for example by spilling water in a hydro power plant). However, in practice this is only done when absolutely necessary to maintain safe operation of the power system, as the variable generation cost in the non-dispatchable units generally is so low that it is optimal to run these units at the maximal available capacity.

  In the multi-area model we assume that the generation cost in the non-dispatchable power plants is negligible. The available generation capacity in all non-dispatchable units is summed per area.

- **The grid.** In the model the grid is only represented by the interconnections between the areas. Between two areas there can only be one interconnection in each direction and these interconnections are characterised by a certain available transmission capacity and a certain loss function. The loss function can be quadratic, i.e.,

$$L_{n,m}(P_{n,m}) = \beta_{Ln,m}P_{n,m} + \gamma_{Ln,m}P_{n,m}^2. \tag{6.4}$$

  We assume that the players of the electricity market have full control of the power flow between the areas, as long as the transmission capacity is not exceeded.

- **Load.** In the model it is assumed that the load is not price sensitive. Hence, the load can be modelled as a certain demand in each area. As there are no guarantee that there always is sufficient generation capacity, we will have to introduce a possibility to disconnect a part of the load. This possibility is represented in the model by a special variable for load shedding (unserved power).

Mathematical Formulation

The multi-area problem can in words be expressed as

minimise *generation cost + penalty cost of load shedding,* (6.5)

subject to *load balance in each area*
(i.e., *generation + import = load – load shedding + export*), (6.5a)

limitations in generation and transmission capacity. (6.5b)

The following parameters are included in the multi-area problem:

C_{Gg} = cost function for generation in thermal power plant g,
C_{Un} = penalty cost for unserved load in area n,
$\mathcal{G}$ = set of thermal power plants g,
$\mathcal{G}_n$ = set of thermal power plants g located in area n,
$\overline{G}_g$ = maximal generation in thermal power plant g,
D_n = load in area n,
$L_{n,m}$ = loss function for transmission from area n to area m,
$\mathcal{N}$ = set of areas n,
$\mathcal{P}$ = set of interconnections (n, m),
$\mathcal{P}_{n\leftarrow m}$ = set of areas m capable of exporting to area n,
$\mathcal{P}_{n\rightarrow m}$ = set of areas m capable of importing from area n,
$\overline{W}_n$ = maximal non-dispatchable generation in area n.

The following optimisation variables are included in the multi-area problem:

G_g = generation in thermal power plant g,
$P_{n,m}$ = transmission from area n to area m,
U_n = unserved power in area n,
W_n = non-dispatchable generation in area n.

Using these symbols we get the following mathematical formulation of the multi-area problem:

$$\text{minimise} \quad \sum_{g \in \mathcal{G}} C_{Gg}(G_g) + \sum_{n \in \mathcal{N}} C_{Un}(U_n) \tag{6.6}$$

$$\text{subject to} \quad \sum_{g \in \mathcal{G}_n} G_g + W_n + \sum_{m \in \mathcal{P}_{n\leftarrow m}} (P_{m,n} - L_{m,n}(P_{m,n})) = D_n - U_n + \sum_{m \in \mathcal{P}_{n\rightarrow m}} P_{n,m}, \quad \forall\, n \in \mathcal{N}, \tag{6.6a}$$

$$0 \le G_g \le \overline{G}_g, \quad \forall\, g \in \mathcal{G}, \tag{6.6b}$$

$$0 \le P_{n,m} \le \overline{P}_{n,m}, \quad \forall\, (n,m) \in \mathcal{P}, \tag{6.6c}$$

$$0 \le U_n \le D_n, \quad \forall\, n \in \mathcal{N}, \tag{6.6d}$$

$$0 \le W_n \le \overline{W}_n, \quad \forall\, g \in \mathcal{N}. \tag{6.6e}$$

Notice that the objective function has to include a penalty cost for unserved load, because otherwise we would get the trivial solution $U_n = D_n$ in each area, while the other optimisation variables are equal to zero. The cost of unserved power must be higher than the generation cost of the most expensive power plant. The penalty cost can be a fictitious cost, introduced just to model that all generation resources should be utilised before load shedding is used. Thus, the penalty cost does not have to be interpreted as a compensation to disconnected consumers.

Example 6.1 (formulation of multi-area model): The town Mji in East Africa is not connected to a national grid, but has a local power system of its own, operated by the Mji Electricity Consumers Cooperative (MECC). In Mji there is a diesel generator set, having a capacity of 250 kWh/h. Due to poor maintenance, failures are quite frequent in the diesel generator set. The operation cost is 10 ¤/kWh. There is also a wind power plant at Mlima, which is located about 15 km from Mji. Just next to Mlima, there is a village, Kijiji, which also is connected to the power system of MECC. The installed capacity of the wind power plant is 200 kW and the operation cost is negligible. The load in Mji and Kijiji can be considered normally distributed and independent of each other.

Mlima and Kijiji are connected to Mji via a transmission line. It can be assumed that the losses on the line are 2% of the injected power and that the line is never subject to failures. Formulate a two-area model of the MECC power system.

Solution: Let Mji be area 1 and Mlima/Kijiji area 2. Introduce the following parameters:

D_n = load in area n, n = 1, 2,
$\overline{G}$ = maximal generation in i diesel generator set g,
$\overline{W}$ = maximal generation in the wind power plant.

Moreover, introduce the following optimisation variables:

G = generation in diesel generator set g,
$P_{n,\,m}$ = transmission from area n to area m, $(n, m) = (1, 2), (2, 1)$,
W = generation in the wind power plant,
U_n = unserved load in area n, n = 1, 2.

Nothing is mentioned about reimbursement of consumers who are disconnected due to capacity deficit; hence, we have to assume a fictitious penalty cost. The only requirement is that this cost should be higher than the operation cost of the most expensive units. As the diesel generator sets cost 10 ¤/kWh we can for example choose the penalty cost 100 ¤/kWh. The multi-area model can then be written as

$$\begin{aligned}
&\text{minimise} && 10G + 100(U_1 + U_2), \\
&\text{subject to} && G + 0.98P_{2,\,1} = D_1 - U_1 + P_{1,\,2}, \\
& && W + 0.98P_{1,\,2} = D_2 - U_2 + P_{2,\,1}, \\
& && 0 \le G \le \overline{G}, \\
& && 0 \le P_{n,\,m}, && (n, m) = (1, 2), (2, 1), \\
& && 0 \le W \le \overline{W}, \\
& && 0 \le U_n \le D_n, && n = 1, 2.
\end{aligned}$$

Using the Multi-area Model in an Electricity Market Simulation

The conditions of the electricity market in the multi-area problem are decided by the parameters of the multi-area problem. These parameters can therefore all be considered scenario parameters.

However, many of the parameters of the multi-area problem, for example the area division, are such that they hardly can be considered random variables. In many cases it can be assumed that the scenario parameters are the available resources ($\overline{G}_g$, $\overline{P}_{n,m}$, and $\overline{W}_n$) as well as the demand (i.e., the load D_n). The other parameters of the multi-area problem are considered model constants instead.

The optimisation variables of the multi-area model shows how the electricity market will behave in a particular scenario. Thus, the optimisation variables can be used as result variables. Altogether, these result variables provide a detailed picture of the function of the electricity market, but as we already concluded, it is in many cases desirable to obtain system indices which provide an overview of the electricity market. It might therefore be appropriate to introduce some additional result variables which can be used as a foundation for the system indices.

In the simulation methods described in this compendium, it is primarily the cost and the reliability of supply that are of interest. Let us start by the operation cost:

> ***Definition 6.1.*** The Total Operation Cost, *TOC*,[1] is the sum of the operation cost in all power plants, as well as any other operation costs (as for example compensation for consumers who have been involuntarily disconnected).

If we in the multi-area model assume that no compensation is paid for unserved power (i.e., the penalty cost C_{Un} is a fictitious cost) the total operation cost is calculated by

$$TOC = \sum_{g \in \mathcal{G}} C_{Gg}(G_g). \tag{6.7}$$

The corresponding system index is the expected operation cost:

> ***Definition 6.2.*** The Expected Total Operation Cost, *ETOC*,[2] is the expected sum of the operation cost in all power plants, as well as any other operation cost, i.e.,
>
> $$ETOC = E[TOC].$$

Both *TOC* and *ETOC* are usually measured in cost per time unit.

To obtain a measure of the reliability of supply, we introduce the following result variable:

> ***Definition 6.3.*** The result variable *LOLO*[3] is a binary variable which is equal to one if load shedding occurs (i.e., if at least one consumer has been involuntarily disconnected due to capacity limitations in the system) and zero otherwise.

In the multi-area model *LOLO* is calculated by

$$LOLO = \begin{cases} 1 & \text{if } \sum_{n \in \mathcal{N}} U_n > 0, \\ 0 & \text{otherwise.} \end{cases} \tag{6.8}$$

The corresponding system index is the risk of power deficit:

> ***Definition 6.4.*** The risk of power deficit, *LOLP*,[4] is equal to the probability that at least one consumer is involuntarily disconnected due to capacity limitations in the system, i.e.,
>
> $$LOLP = E[LOLO].$$

Neither *LOLO* nor *LOLP* have any dimension. *LOLP* is therefore usually stated in per cent or in hours per year.

The system index *LOLP* only shows how common power deficit situations are, but does not tell

1. This footnote has just been added to maintain the footnote numbering in pace with the Swedish edition.
2. This footnote has just been added to maintain the footnote numbering in pace with the Swedish edition.
3. Short for "Loss Of Load Occasion".
4. Short for "Loss Of Load Probability".

us anything about the size of the problem; we have a power deficit even if just a 60 W light bulb has been disconnected. To get a measure of the extent of a power deficit situation we make the following definitions:

> ***Definition 6.5.*** The unserved energy, *ENS*,[5] is the energy that could not be delivered due to capacity limitations in the power system.

In the multi-area model *ENS* is calculated by

$$ENS = \sum_{n \in \mathcal{N}} U_n. \tag{6.9}$$

The corresponding system index is the expected unserved energy:

> ***Definition 6.6.*** The expected unserved energy, *EENS*,[6] is the expected energy that could not be delivered due to capacity limitations in the power system, i.e.,
>
> $$EENS = E[ENS].$$

Both *ENS* and *EENS* are usually measured as energy per time unit.

It must be observed that there is no simple relation between *LOLP* and *EENS*! It is not uncommon to believe that the unserved energy equals the *LOLP* multiplied by the maximal load or the *LOLP* multiplied by the average load. This could be the case at some occasions, bute it is absolutely no general rule (cf. exercise 6.7).

> ***Example 6.2 (scenario parameters and result variables in a multi-area model):*** Consider the multi-area model in example 6.1. Assume that the objective of the simulation is to determine the system indices *ETOC* and *LOLP*. Which scenario parameters are included in the model? Which result variables are relevant and how can values of those be obtained?
>
> ***Solution:*** The scenario parameters are the quantities that can vary randomly and which determine the conditions for operation of the system. In this case we have the maximal generation capacity in the power plants, $\overline{G}$ and $\overline{W}$, as well as the load in each area, D_1 and D_2.
>
> The relevant result variables in this simulation are *TOC* and *LOLO*, because the system indices *ETOC* and *LOLP* are the expectation values of these result variables. For a particular scenario, we can solve the multi-area problem from example 6.1 and then calculate the values of the result variables according to
>
> $$TOC = 10G,$$
>
> $$LOLO = \begin{cases} 0 & \text{om } U_1 + U_2 = 0, \\ 1 & \text{om } U_1 + U_2 > 0. \end{cases}$$

6.1.2 Examples of Applications

Electricity market simulation can be used to study how an electricity market can be expected to behave with various prerequisites. By comparing the results it is thus possible to conclude whether or not an investment is profitable, or which arrangement of the electricity trading will be best to achieve a specified goal. In this section, we will give some short examples of how the results of an electricity market simulation will be used.

5. Short for "Energy Not Served".
6. Short for "Expected Energy Not Served". Sometimes the abbreviation *UE* (for "Unserved Energy") is used.

The Value of an Investment

To determine the value of an investment it is necessary to find out how the above described system indices are affected by the investment. This requires at least two simulations: one simulation of the system before the investment is made and then a simulation of how the system behaves after the investment. The value of the investment is the difference of the system indices between the two simulations. The value of the investment can then be compared to the cost of the investment. The cost of the investment must consider interest rates and depreciation time. A simple way to accomplish this is to use the so-called *annuity formula*:

$$AC = I \cdot \frac{r/100}{1-(1+r/100)^{-n}}, \tag{6.10}$$

where

AC = annual cost [¤],
I = investment cost [¤],
r = real rate of interest [%],
n = depreciation time [years].

The annuity formula distributes the initial investment cost as an annual amount over the whole deprecation time. If the annual amounts are summed, the result is higher than the initial investment (if $r > 0$), due to the interest rate.

Example 6.3 (the value of a transmission line): Consider an electricity market divided into two regions. In region A the load is 150 kWh/h during peak periods and 100 kWh/h otherwise. In region B the corresponding values are 100 kWh/h and 50 kWh/h respectively. Workdays between 6 a.m. and 8 p.m., and weekends between 10 a.m. and 5 p.m. are considered peak periods. In region A there are two diesel generator sets with operation costs 10 ¤/kWh, and in region B there is one diesel generator set with an operation cost of 12 ¤/kWh. All units have an availability of 90% and a capacity of 100 kWh/h.

At present time the two regions are not interconnected; however, the construction of a transmission line is discussed. The transmission line can be considered loss free and having an unlimited capacity. The cost of the project, including capital costs and annual maintenance, is estimated to 275 000 ¤/year. Is it profitable to build the transmission line?

Solution: Two cases have to be simulated: the current situation with two separate system as well as the situation when the two regions are interconnected and can be treated as a single system. Later in this chapter we will describe methods to perform these simulations. For the moment it is sufficient to study the result, which are shown in the table below.

Table 6.1 System index with and without the transmission line.

System index	No transmission line	With the transmission line
$ETOC$ [¤/h]	2 002.50	1 972.80
$LOLP_A$ [%]	10	14.5
$LOLP_B$ [%]	10	

By comparing the two cases we can find that the transmission line results in an annual cost saving of 260 000 ¤/year, due to the reduced operation costs. If this was the only criteria to be evaluated the transmission line would seem unprofitable. However, without the transmission line the risk of power deficit for the market as a whole is 19%,[7] which is reduced to 14.5% if the transmission line is built. If the increased reliability is valued to at least 15 000 ¤/year then the transmission line is profitable. If the reduction in loss of load probability is less than that then the transmission line should not be built.

The Capacity Credit of a Power Plant

If the available generation capacity is less than the power demand then some consumption has to be disconnected. The risk of power deficit, *LOLP*, is in the normal case very low, but it can never be reduced to zero, i.e., $LOLP > 0$If we have a certain system and build another power plant, the probability of consumers having to be disconnected will decrease, as the generation capacity of the system increases. The ability of a power plant to increase the reliability of supply is referred to as its *capacity credit*. Of course, also the existing power plants in system have a capacity credit, because if they were removed the *LOLP* of the system would increase.

The capacity credit can be expressed in several ways. Two possible definitions are the following:

Definition 6.7. The capacity credit of a power plant *g* expressed as equivalent firm power is measured by comparing the ability of the power plant to decrease the *LOLP* to the corresponding ability of a 100% reliable power plant. Including the power plant *g* the system has certain risk of power deficit, $LOLP'$. Excluding power plant *g* and replacing it by a 100% reliable power plant with the installed capacity $\hat{G}$ yields another risk of power deficit, $LOLP''$. If $LOLP' = LOLP''$ then the power plants have the same ability to decrease the risk of power deficit, and then it said that power plant g has the capacity credit $\hat{G}$ expressed as equivalent firm power.

Definition 6.8. The capacity credit of a power plant *g* expressed as equivalent load increase is measured by studying the ability of power plant *g* to prevent the risk of power deficit from increasing as the load increases. A given mean load μ_D yields a certain risk of power deficit $LOLP'$ when power plant *g* is excluded from the system. If power plant *g* is included and the mean load is increased by Δ then the risk of power deficit becomes $LOLP''$. If $LOLP' = LOLP''$ then the power plant has prevented the risk of power deficit to increase for this load increase, and then it is said that power plant *g* has the capacity credit Δ expressed as an equivalent, equally distributed load increase.

There are many possible variations of these definitions. It is for example possible to replace the 100% reliable power plant in definition 6.7 by a standard power plant having a availability[8] less than 100%. In definition 6.8 we could have another distribution of the load increase, for example a load which is increasing more during peak load periods.

Example 6.4 (capacity credit): Table 6.2 shows how the *LOLP* is changed in a certain system when the load increases and new power plants are added. Determine the capacity credit of the wind power expressed as equivalent firm power and as equivalent load increase.

Solution: 100 MW wind power has been added in variant 1 and this results in the same risk of power deficit as in variant 2, where 22.2 MW of firm power has been

7. To obtain this figure it is necessary to consider that $LOLP_A$ is 19% during peak periods but only 1% otherwise, whereas $LOLP_B$ is the same regardless of which time it is. The probability of all consumers in both regions being supplied is therefore $0.5 \cdot 0.81 \cdot 0.9 + 0.5 \cdot 0.99 \cdot 0.9 = 81\%$.
8. See section 6.2.3 for a more thorough description of the reliability of a power plant.

Table 6.2 Data of various system configurations in example 6.4.

	Original system	Variant 1	Variant 2	Variant 3	Variant 4
Load (normally distributed) [MW]					
Mean	500.0	500.0	500.0	522.3	514.0
Standard deviation	100.0	100.0	100.0	100.0	105.0
Installed capacity in different power plants [MW]					
Coal condensing (95% availability)	5 × 200	5 × 200	5 × 200	5 × 200	5 × 200
Wind power[a]	0	100	0	100	100
Firm power (100% availability)	0	0	22.2	0	0
LOLP [%]	0.0787	0.0537	0.0537	0.0787	0.0787

a. At any given moment (independent of the load) the probability that the wind power can generate installed capacity is assumed to be 10%; the probability of half the installed capacity being available is 40% and the probability of no generation at all is 50%. This means that the wind power in average generates 30% of installed capacity.

added instead. This means, according to definition 6.7, that in this system 100 MW wind power has the capacity credit 22.2 MW expressed as equivalent firm power.

In variant 3 the mean load has increased to 522.4 MW, but the standard deviation of the load remains the same, which means that the load increase is equally large at all times. The *LOLP* is the same as in the original system, thanks to the 100 MW wind power which has been added. According to definition 6.8, we can conclude that 100 MW wind power in this system has the capacity credit 22.4 expressed as an evenly distributed load increase.

In variant 4 the mean load has increased to 514.0 MW. However, in this case the load increase is larger during peak load periods, which is reflected by the fact that the standard deviation of the load has increased to 105 MW. The *LOLP* is the same as in the original system, thanks to the 100 MW wind power which has been added. We can conclude that 100 MW wind power in this system has the capacity credit 14.0 expressed as an unevenly distributed load increase.

It can be noted that the example above counters the wide-spread misconception that "wind power has no capacity credit because it is not always windy". It is correct that the wind power has a lower capacity credit than the same installed capacity in a power plant which is not weather dependent, but as the example shows still wind power has the ability to decrease the risk of power deficit.

6.2 PROBABILISTIC PRODUCTION COST SIMULATION

In this section we will present the method of probabilistic production cost simulation (PPC). The method, which was developed in the late 1960s, can be used to simulate simple electricity markets. The main concept of the method is to use such a simple electricity market model that all scenario parameters can be combined into a single one-dimensional probability distribution. From this distribution it is possible to analytically calculate system indices as *ETOC* and *LOLP*.

In order to avoid too complex calculations it is necessary to make some simplifying assumptions. The most important assumption is to neglect all correlations[9] between random variables, as for example load and available generation capacity. This assumption is so fundamental for probabilistic production cost simulation that we state it formally:

> ***Assumption 6.9.*** All random variables in a probabilistic production cost simulation are assumed to be independent!

In practice it is sometimes possible to consider correlated variables in a simplified manner, which will be shown in the subsection on simulation of wind power.

Another simplification in PPC is that the transmission grid is more or less entirely neglected. The power system model in a probabilistic production cost simulation may therefore be seen as a variant of the electricity market model from section 6.1.1, but with just one area.

The description of probabilistic production cost simulation starts by a compilation of the most basic calculations in the method. Then we show how to obtain the necessary models of load and simple power plants. Later we add more advanced models of wind power and dispatchable hydro power. Finally we also describe a method to simplify the calculation of the *LOLP* by using the so-called normal approximation.

6.2.1 Basic Principles

Assume that we know the duration curve of the load in a system, $\tilde{F}_0(x)$,[10] and that this load must be supplied by a single power plant. The power plant is however not completely reliable; there is a possibility that the plant is failing and can not generate any electric power at all. The probability of this is called the availability. In this power plant we have the availability p_1 and the operation cost β_1. By comparing the load duration curve and the generation duration curve of the power plant it is comparatively easy to calculate the system reliability and operation cost, as shown in the following example:

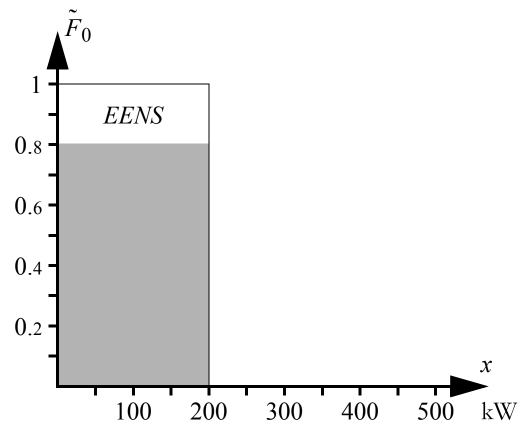


Figure 6.2 Calculation of system indices directly from the load duration curve.

> ***Example 6.5.*** In a simple system the load is constantly 200 kW. The system is supplied by a 200 kW power plant having an availability of 80% and the operation cost 1 ¤/kWh. Calculate the *EENS*, *LOLP* and *ETOC* of the system.
>
> ***Solution:*** Since the load is constant the probability is 1 that the load exceeds the level x if $x < 200$ kW and 0 if $x \geq 200$ kW, which results in the duration curve shown in figure 6.2. When the power plant is available, a probability which is stated to 80%, it will generate enough to meet the load. The duration curve of the power generation is therefore such that the probability is 1 that the generation exceeds the level x if x is negative. However, negative generation is obviously not possible, so that part of the duration curve is not interesting. The probability is 0.8 that x is exceeded if x is larger than or equal to zero but less than 200 kW and the probability is 0 then $x \geq 200$ kW. This duration curve is illustrated by the shadowed part of the figure.
>
> From the figure we can directly obtain the most important system indices. The

9. See appendix C.
10. Definitions of duration curves and other important notions concerning random variables are found in appendix C.

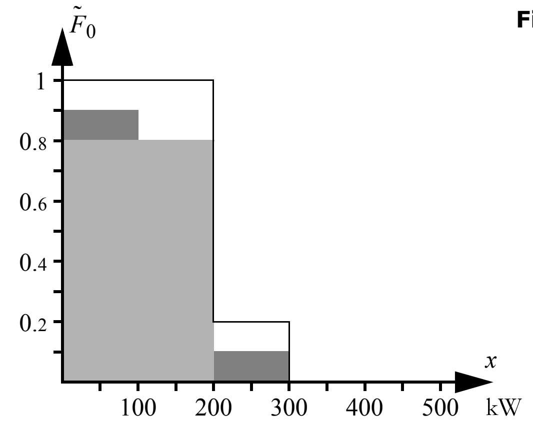


Figure 6.3 A system with two power plants and two possible load levels. The generation in the least expensive unit (capacity 200 kW, availability 80%, operation cost 1 ¤/kWh) is shown by the lighter shadowed area. There is also a backup unit (capacity 100 kW, availability 50%, operation cost 2 ¤/kWh) which is used when the base unit is unavailable or when the load is high (indicated by the two darker areas). The expectation value of unserved energy is indicated by the white areas below the load duration curve.

unserved energy corresponds to the white part below the load duration curve, which constitutes 20% of the total load, i.e., $EENS = 0.2 \cdot 200$ kWh/h = 40 kWh/h.Similarly, we can calculate the expected generation in the power plant to 160 kWh/h, which yields an expected total operation cost $ETOC = 160$ ¤/h. The loss of load probability is the same as the probability of the power plant not being available, i.e., $LOLP = 20\%$.

The example above is however almost trivial. If the load had more than one possible value and if the system also would include backup units which are uses when the base power plant is unavailable or for peak load periods (cf. figure 6.3)then it would be much more difficult to determine the value of the system indices just by reasoning. Apparently we need a smart method to simplify and systematise the calculations. One such smart method is to introduce *equivalent load duration curves.*

Equivalent Load

The consequence of a failure in a generating unit is that the power generation of that unit has to be replaced by increased generation in some other unit. If no other unit is available then there will be a power deficit. A failure in power plant can thus be seen as a load increase for the remaining units. From a system point of view, the failure is equivalent to a situation where the unit continues operation as before, but there is a load increase equal to the generation of the failing unit. We will therefore introduce the notion *equivalent load*:

Definition 6.10. The equivalent load is given by

$$E_g = D + \sum_{k=1}^{g} O_k,$$

where

E_g = equivalent load for the power plant next to be dispatched after unit g,
D = actual load,
O_k = outage in unit k.

The desired simplification of the calculation of the system indices can be achieved by studying the equivalent load duration curve. The question is then how to determine the equivalent load duration curve.

Example 6.6. Study the same system as in example 6.5. What does the equivalent load duration curve look like?

Solution: The actual load is always 200 kW. There is also a 20% probability that we have an outage in the power plant. Since the power plant always generates 200 kW when it is available, an outage in the power plant must be equivalent to a load increase by the same amount. Thus, the probability is 80% that the equivalent load is 200 kW and 20% that it is 400 kW. This results in the duration curve shown in figure 6.4.

In the example above it was not difficult to find the equivalent load duration curve by simple reasoning. In order to study more complex systems we will however need a more general formula. Suppose that we have a power plant with the availability p_1 (and hence the unavailability $q_1 = 1 - p_1$) and the installed capacity $\hat{G}_1$. We want to calculate the probability that the equivalent load is exceeding a given level x. There are two possibilities for this; either the power plant is available and the actual load is larger than x or the power plant is unavailable and the actual load is larger than $x - \hat{G}_1$. If we assume that actual load and outages in power plants are independent random variables, this may be expressed as

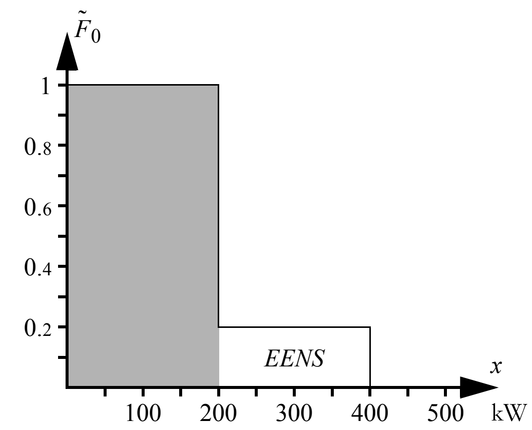


Figure 6.4 Equivalent load duration curve for the system in examples 6.5 and 6.6.

$$P(E > x) = p_1 \cdot P(D > x) + q_1 \cdot P(D > x - \hat{G}_1). \qquad (6.11)$$

The probability that a random variable exceeds a given level is equal to the value of the duration curve at the same level.[11] As before, the load duration curve is denoted $\tilde{F}_0(x)$ and we introduce the notation $\tilde{F}_1(x)$, for the equivalent load duration curve, which means that (6.11) can be written as

$$\tilde{F}_1(x) = p_1 \cdot \tilde{F}_0(x) + q_1 \cdot \tilde{F}_0(x - \hat{G}_1). \qquad (6.12)$$

This relation is true for all values of x. Using (6.12) it is possible to calculate the entire equivalent load duration curve.

Example 6.7. Calculate the equivalent load duration curve of the system in example 6.5 using (6.12).

Solution: The load duration curve can be written as

$$\tilde{F}_0(x) = \begin{cases} 1 & x < 200, \\ 0 & 200 \le x. \end{cases}$$

Applying the formula yields that

$$\tilde{F}_1(x) = 0.8\tilde{F}_0(x) + 0.2\tilde{F}_0(x - 200) =$$

$$= \begin{cases} 0.8 \cdot 1 + 0.2 \cdot 1 = 1 & x < 200, \\ 0.8 \cdot 0 + 0.2 \cdot 1 = 0.2 & 200 \le x < 400, \\ 0 & 400 \le x. \end{cases}$$

This is exactly the duration curve which was obtained by reasoning in example 6.6.

The equation (6.12) shows how the equivalent load is calculated in the case when the actual load

11. See definition C.3.

should be covered by one particular power plant. The same reasoning can also be applied to another power plant which "sees" a given equivalent load. Thus, we can generalise (6.12) to the following so-called convolution formula:

$$\tilde{F}_g(x) = p_g \cdot \tilde{F}_{g-1}(x) + q_g \cdot \tilde{F}_{g-1}(x - \hat{G}_g). \qquad (6.13)$$

The convolution formula shows how to calculate the equivalent load when considering outages in the g first units.

It must be noted that in addition to the assumption that the involved random variables are independent, we have made another important assumption when deriving the convolution formula; the power plants are dispatched in a given merit order. If it is possible to cover the load using only the first power plant then only power plant 1 will be used. If power plant 1 is unavailable or if the load is larger than the capacity of power plant 1 then power plant 2 is dispatched, etc. This is a reasonable assumption, as the least expensive way to cover the load is to primarily use the least expensive power plant, then the second cheapest power plant, etc. The convolution formula (6.13) should therefore be applied to the power plants in ascending cost order.

There are some more prerequisites if this model should be correct. Firstly, there may not be any transmission limitations which forces the load in a particular area to be covered by local generation, even though there is capacity available in less expensive units somewhere else in the system. Secondly, we have to neglect start-up times, which may cause more expensive units to be dispatched because the less expensive units have not been committed.[12] Finally, we assume that it is possible to in beforehand arrange the power plants in a merit order after ascending generation cost, i.e., that the generation cost per MWh is independent of the generation level. This implies that the power plants must have a cost function

$$C_{Gg}(G_g) = \alpha_{Gg} + \beta_{Gg}G_g, \qquad (6.14)$$

i.e., a fixed cost and a variable cost directly proportional to the power generation.

Calculation of System Indices

The difference between the duration curve of the actual load and the equivalent load duration curve is, as described above, that when studying the equivalent load duration curve it is assumed that all power plants are 100% available; an outage is included as an increase of the equivalent load instead. This means that the unserved energy remains the same when the equivalent load duration curve is studied. Compare figures 6.2 and 6.4; the white area is equally large in both cases. The difference is that it is much easier to find a general mathematic expression for the unserved energy when using equivalent load duration curves. The part of the equivalent load which exceeds the total installed capacity, $\hat{G}_g^{tot}$, cannot be covered by the g first power plants; hence, it will constitute unserved energy. We can express this as

$$EENS_g = T\int_{\hat{G}_g^{tot}}^{\infty} \tilde{F}_g(x)dx, \qquad (6.15)$$

where T is the duration of the time period for which we want to calculate the unserved energy.

Also the loss of load probability can directly be found in the equivalent load duration curve. Load shedding will be necessary if the equivalent load is larger than the installed capacity in the system. The probability for this is given by

$$LOLP_g = \tilde{F}_g(\hat{G}_g^{tot}). \qquad (6.16)$$

12. Cf. the discussion of short-term planning of thermal power plants section 5.3.

To be able to calculate the total operation cost it is necessary to know the expected generation in each power plant. Unlike the unserved power the expected generation is however not unchanged when the equivalent load duration curve is calculated. This is clearly visible when comparing figures 6.2 and 6.4; the shaded area is larger in the latter figure. The expected generation of a power plant can however by determined by studying how the power plant affects the unserved energy in the system. Before power plant g is added to the system we have the unserved energy $EENS_{g-1}$. After using (6.13) we get the unserved energy $EENS_g$. he difference between these values must of course be due to energy generation in power plant g; thus we can calculate the expected generation by

$$EG_g = EENS_{g-1} - EENS_g. \tag{6.17}$$

An alternative way of calculating the expected generation is to use the following formula:[13]

$$EG_g = T \cdot p_g \int_{\hat{G}_{g-1}^{tot}}^{\hat{G}_g^{tot}} \tilde{F}_{g-1}(x)dx. \tag{6.18}$$

Verification and derivation of this formula is left to the reader; see exercises 6.2 and 6.3.

As we know the expected generation in each power plant it is possible to calculate the total operation cost by summation of the expected costs in each power plant:

$$ETOC_g = \sum_{k=1}^{g} \beta_{Gk} EG_k, \tag{6.19}$$

where β_{Gk} is the variable operation cost in the cost function (6.14).

Example 6.8. Calculate the unserved energy, loss of load probability and expected operation cost per hour in the system described in figure 6.3.

Solution: The load duration curve is given in the figure:

$$\tilde{F}_0(x) = \begin{cases} 1 & x < 200, \\ 0.2 & 200 \le x < 300, \\ 0 & 300 \le x. \end{cases}$$

The larger power plant has the lower operation cost and should therefore be added first:

$$\tilde{F}_1(x) = 0.8\tilde{F}_0(x) + 0.2\tilde{F}_0(x-200) =$$

$$= \begin{cases} 0.8 \cdot 1 + 0.2 \cdot 1 = 1 & x < 200, \\ 0.8 \cdot 0{,}2 + 0.2 \cdot 1 = 0.36 & 200 \le x < 300, \\ 0.8 \cdot 0 + 0.2 \cdot 1 = 0.2 & 300 \le x < 400, \\ 0.8 \cdot 0 + 0.2 \cdot 0.2 = 0.04 & 400 \le x < 500, \\ 0 & 500 \le x. \end{cases}$$

Then the second power plant is added:

$$\tilde{F}_2(x) = 0.5\tilde{F}_1(x) + 0.5\tilde{F}_1(x-100) =$$

13. It should be noted that this formula cannot be applied to power plants which have more than one possible state of availability (see section 6.2.4).

$$= \begin{cases} 0.5 \cdot 1 + 0.5 \cdot 1 = 1 & x < 200, \\ 0.5 \cdot 0.36 + 0.5 \cdot 1 = 0.68 & 200 \le x < 300, \\ 0.5 \cdot 0.2 + 0.5 \cdot 0.36 = 0.28 & 300 \le x < 400, \\ 0.5 \cdot 0.04 + 0.5 \cdot 0.2 = 0.12 & 400 \le x < 500, \\ 0.5 \cdot 0 + 0.5 \cdot 0.04 = 0.02 & 500 \le x < 600, \\ 0 & 600 \le x. \end{cases}$$

To calculate the expected generation in the two power plants we must know the unserved energy before and after adding the power plants to the equivalent load duration curve. Since we want to calculate the operation cost per hour it is most practical to use the same period length in the calculations of unserved energy and expected generation:

$$EENS_0 = 1 \cdot \int_0^{\infty} \tilde{F}_0(x)dx = 1 \cdot 200 + 0.2 \cdot 100 = 220 \text{ kWh},$$

$$EENS_1 = 1 \cdot \int_{200}^{\infty} \tilde{F}_1(x)dx = 0.36 \cdot 100 + 0.2 \cdot 100 + 0.04 \cdot 100 = 60 \text{ kWh},$$

$$EENS_2 = 1 \cdot \int_{300}^{\infty} \tilde{F}_2(x)dx = 0.28 \cdot 100 + 0.12 \cdot 100 + 0.02 \cdot 100 = 42 \text{ kWh},$$

$$EG_1 = EENS_0 - EENS_1 = 160 \text{ kWh},$$

$$EG_2 = EENS_1 - EENS_2 = 18 \text{ kWh}.$$

It is now easy to calculate the expected operation cost per hour:

$$ETOC = \beta_1 \cdot EG_1 + \beta_2 \cdot EG_2 = 1 \cdot 160 + 2 \cdot 18 = 196 \text{ ¤/h}.$$

The loss of load probability is given by

$$LOLP = \tilde{F}_2(300) = 28\%$$

and the unserved energy is 42 kWh/h, as we already have calculated

6.2.2 Load Model

When using the method of probabilistic production cost simulation the load is described solely by a duration curve. This means that we either assumed that the producers are obliged to supply the load as long as it is technically possible or that the load is price independent (i.e., the consumers are willing to pay any price as long as their load is supplied). Obligation to supply is common in regulated electricity market, where it is common with so-called concession holders, which have monopoly on power delivery within a specified area, but also are obliged to supply all consumers within that area. In modern restructured electricity markets there is a tendency that the consumers are becoming more price sensitive, but it is still a reasonable simplification to assume price independent load.

In those cases when the load is assumed to follow some given probability distribution, e.g. the normal distribution, the load duration curve can easily be calculated using the definitions given in appendices C and D. Below we will show how the load duration curve can be determined starting from a specified typical load profile:

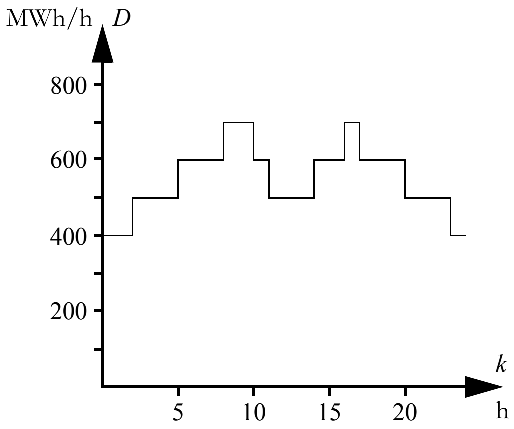


Figure 6.5 Load curve of the load in table 6.3.

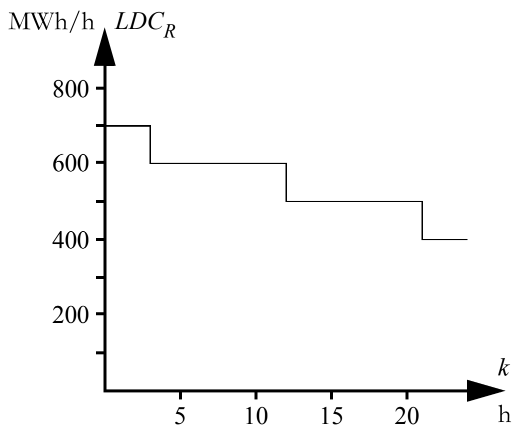


Figure 6.6 Real duration curve of the load in table 6.3.

Definition 6.11. The load curve, $D(k)$, states the mean load per hour during a specified time period: $k = 1, \ldots, T$.[14]

A given load curve can be transformed into the real load duration curve by sorting the load levels in descending order. This means that the real load duration curve is defined as follows:

Definition 6.12. The real load duration curve, $LDC_R(k)$, states the load level which is exceeded during k hours.

It should be noted that information is lost during the transformation from load curve to real load duration curve, as it is no longer possible to determine how fast the load is increasing or decreasing, or how often a given load level will be exceeded during a given time period.

Example 6.9. Table 6.3 shows the load during a typical day in a system. Draw the real load duration curve of this load.

Table 6.3 Load during a typical day in example 6.9.

Hour	Load [MWh/h]	Hour	Load [MWh/h]	Hour	Load [MWh/h]
1	400	9	700	17	700
2	400	10	700	18	600
3	500	11	600	19	600
4	500	12	500	20	600
5	500	13	500	21	500
6	600	14	500	22	500
7	600	15	600	23	500
8	600	16	600	24	400

Solution: In figure 6.5 the load curve corresponding to the values stated in table 6.3 is shown. The real load duration curve is obtained by simply sorting the load levels of the load curve in descending order, as shown in figure 6.6. The curve should be interpreted so that the peak load 700 MWh/h is exceeded during 3 hours (and hence also during 1 or 2 hours), etc.

The next step is to switch the axis of the real load duration curve:

Definition 6.13. The inverted load duration curve, $LDC(x)$, states how many hours

14. Some other period length than an hour can be used if desirable.

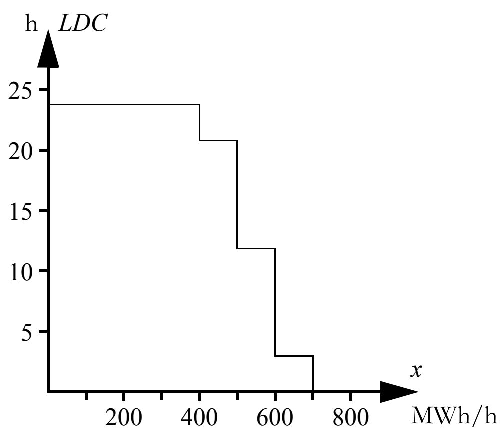


Figure 6.7 The inverted duration curve of the load in table 6.3.

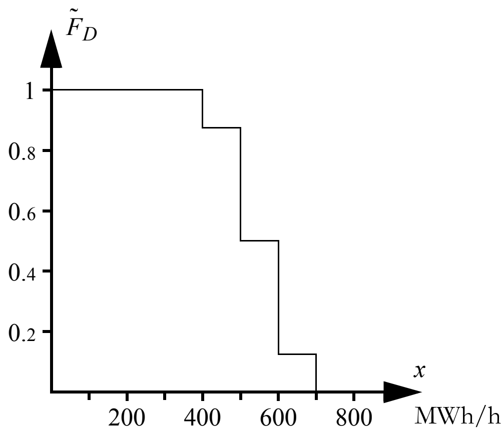


Figure 6.8 Duration curve of the load in table 6.3.

a certain load level *x* is exceeded.

The inverted load duration curve is valid for a specified time period. By normalising the vertical axis, i.e., dividing all values by the duration of the studied time period, *T*, a normalised load duration curve is obtained. This is the more formal term for what we in section 6.2.1 referred to as the load duration curve; however, we will continue to use the shorter term. The definition of a load duration curve is completely in accordance to the general definition of duration curves which is given in appendix C.

Example 6.10. Determine the load duration curve of the load in example 6.9.

Solution: Switching axis of the real load duration curve from figure 6.6 results in the inverted load duration curve shown in figure 6.7. The load duration curve is then obtained by scaling the inverted load duration curve.

Alternatively, the problem could have been solved using the data from table 6.3 directly. The frequency function, i.e., the probability of a specified load level is easily computed from the table:

$$f_D(x) = \begin{cases} 3/24 = 0.125 & x = 400, \\ 9/24 = 0.375 & x = 500, \\ 9/24 = 0.375 & x = 600, \\ 3/24 = 0.125 & x = 700. \end{cases}$$

Using theorem C.4 we get

$$\tilde{F}_D(x) = \sum_{t > x} f_D(t) = \begin{cases} 0.125 + 0.375 + 0.375 + 0.125 = 1 & x < 400, \\ 0.375 + 0.375 + 0.125 = 0.875 & 400 \le x < 500, \\ 0.375 + 0.125 = 0.5 & 500 \le x < 600, \\ 0.125 & 600 \le x < 700, \\ 0 & 700 \le x, \end{cases}$$

which is exactly the same duration curve as shown in figure 6.8.

In the examples above we have only considered simple loads which only have a few possible levels. In reality the load is of course a continuous random variable. The earlier stated formu-

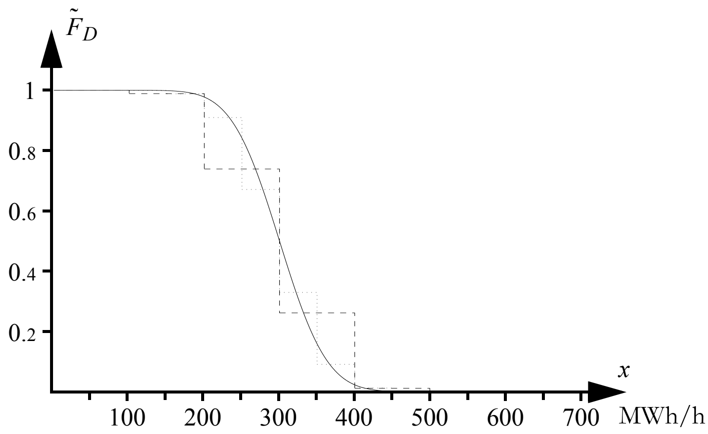


Figure 6.9 Approximation of the load duration curve in example 6.12.

lae—both concerning the load duration curve and calculation of system indices—are valid also then, but the calculations are harder. Therefore, it is appropriate to divide a continuous load duration curve into constant segments, i.e., to approximate the continuous function by a staircase-looking function, similar to the one obtained in example 6.10. The segment size (in MWh/h) correspond to the width of the "steps". For a given segment size it is then possible to represent the duration curve using a vector.

> ***Example 6.11.*** What is the vector representation of the load duration curve in example 6.10?
>
> ***Solution:*** In this case it is natural to use the step size 100 MWh/h, which means that the vector representation becomes [1 1 1 1 0.875 0.5 0.125 0]. The first element should be interpreted as the value of $\tilde{F}_D(x)$ in the first segment, i.e., $0 \le x < 100$, the second element is the value in the second segment $100 \le x < 200$, etc.

Given a specified segment size there are several possible ways to divide a continuous duration curve into segments. It is appropriate to select a segmentation which does not change the area below the duration curve, since this area is related to the expectation value of the corresponding random variable (according to theorem C.10). Performing a segmentation according to this principle may need quite a lot of work and is therefore not always worth the pains. Instead of putting a large effort on finding the best approximation it might be better to just reduce the segment size.

> ***Example 6.12 (segmentation of a load duration curve):*** Suppose that the load in a system is normally distributed and has a mean value of 300 MW and a standard deviation of 50 MW. Approximate the load duration curve using two segment sizes: 50 MW and 100 MW respectively.
>
> ***Solution:*** First we have to select a method to determine the value in each segment of the approximated load duration curve $\tilde{F}^*_D$. A simple method is to assume that the continuous load duration curve is more or less linear within each segment, and then choose $\tilde{F}^*_D$ as the mean of the linear approximation:
>
> $$\tilde{F}^*_D(x) = \frac{\tilde{F}_D((s-1)\cdot\Delta x) + \tilde{F}_D(s\cdot\Delta x)}{2} \quad \text{for } (s-1)\cdot\Delta x \le x < s\cdot\Delta x,$$
>
> where s denotes the s:th segment and Δx is the segment size. The result of applying this strategy is shown in figure 6.9. As can be seen in the figure, the assumption about linearity is reasonable around the mean load 300 MW, but is less good for low and

high load levels. The error is significantly larger in the approximation using the wider step length (dashed line in the figure).

6.2.3 Model of Thermal Power Plants

In section 6.2.1 we used a power plant model where the units were characterised by three properties: the installed capacity $\hat{G}$, generation cost β_G, and availability p. This model is suitable for thermal power plants.

The installed capacity of a power plant hardly offers any problem to determine, as it is defined by the technical performance of the equipment in the power plant. The operation cost of thermal power plants have already been treated in section 5.3.2. Notice that a probabilistic production cost simulation cannot consider start-up or stop costs, as all information about the time-line of events is lost when a chronological load curve is transformed into a load duration curve (cf. section 6.2.2).

The availability is difficult to calculate using a theoretical model and must in practice be based on assumptions, for example by comparison to operational statistics of similar power plants. A power plant can not be operated continuously; every now and then it has to be stopped for maintenance. We can distinguish between corrective and preventive maintenance. Corrective maintenance means that a vital component must be repaired in order to allow the power plant to generate power. Corrective maintenance is thus performed after failures and can not be anticipated. The purpose of preventive maintenance is to reduce the risk of failures and is performed both for economical and security reasons. The preventive maintenance must be planned in advance, so that the power plant is taken off-line at some occasion when the disturbance is as small as possible. The Swedish nuclear power plants are for example maintained during the summer, when the load in the Swedish system is low.

The convolution formula (6.13) is based on the assumption that load and outages in power plants are independent random variables. This is a reasonable assumption for failures in power plants. Preventive maintenance is however scheduled to specified occasions and there will therefore be a correlation between the load and preventive maintenance. This problem can be avoided (for example by performing separate calculations of the system indices during those periods when important power plants are stopped for maintenance), but to simplify the presentation we will restrict ourselves to studying corrective maintenance in this compendium.

During the time the power plant is repaired it is said to be *unavailable*; the remaining time it is *available*. Notice that an available power plant is not necessarily dispatched, as the load might be so low that the power plant is not needed. The availability of a power plant is the probability of the power plant being available. The most simple way to calculate the availability is to study statistics of the operation of the power plant (or a similar power plant). These calculations require the following notions:

Definition 6.14. The Mean Time To Failure is calculated by[15]

$$MTTF = \frac{1}{K}\sum_{k=1}^{K} t_u(k),$$

where K is the number of periods when the power plant is available and $t_u(k)$ is the duration of each of these periods.

Definition 6.15. The Mean Time To Repair is calculated by[16]

15. This footnote has just been added to maintain the footnote numbering in pace with the Swedish edition.
16. This footnote has just been added to maintain the footnote numbering in pace with the Swedish edition.

$$MTTR = \frac{1}{K}\sum_{k=1}^{K} t_d(k),$$

where K is the number of periods when the power plant is unavailable and $t_d(k)$ is the duration of each of these periods.

Definition 6.16. The failure rate λ is the probability that an available unit will fail. The failure rate can be estimated as

$$\lambda = \frac{1}{MTTF}.$$

Definition 6.17. The repair rate μ is the probability that an unavailable unit will be repaired. The repair rate can be estimated as

$$\mu = \frac{1}{MTTR}.$$

Using the above described definitions, we can express the availability as follows:

Definition 6.18. The availability is the probability that a power plant is available. This probability can be estimated as the part of a longer period that the unit is available:

$$p = \frac{\sum_{k=1}^{M} t_u(k)}{\sum_{k=1}^{M} (t_u(k) + t_d(k))} = \frac{M \cdot MTTF}{M(MTTF + MTTR)} = \frac{MTTF}{MTTF + MTTR} = \frac{\mu}{\mu + \lambda}.$$

Given the above definition of availability, the definition of unavailability is natural:

Definition 6.19. The unavailability is the probability that a power plant is unavailable, which can be estimated by

$$q = 1 - p = \frac{MTTR}{MTTF + MTTR} = \frac{\lambda}{\mu + \lambda}.$$

As we can see in the definitions, there are several ways to calculate the availability and unavailability. The available statistics determines which method should be chosen. Below follows a simple example.

Example 6.13 (availability in a power plant). Table 6.4 shows the operation log of a power plant. Calculate the failure rate, repair rate and unavailability of this unit.

Table 6.4 Example of operation log of a power plant.

Event	Time [week]								
Failure		20		60		70		101	
Repair	0		23		62		74		104

Solution: We start by calculating the mean time to failure and repair respectively:

$$MTTF = \frac{1}{4}((20 - 0) + (60 - 23) + (70 - 62) + (101 - 74)) = 23 \text{ weeks},$$

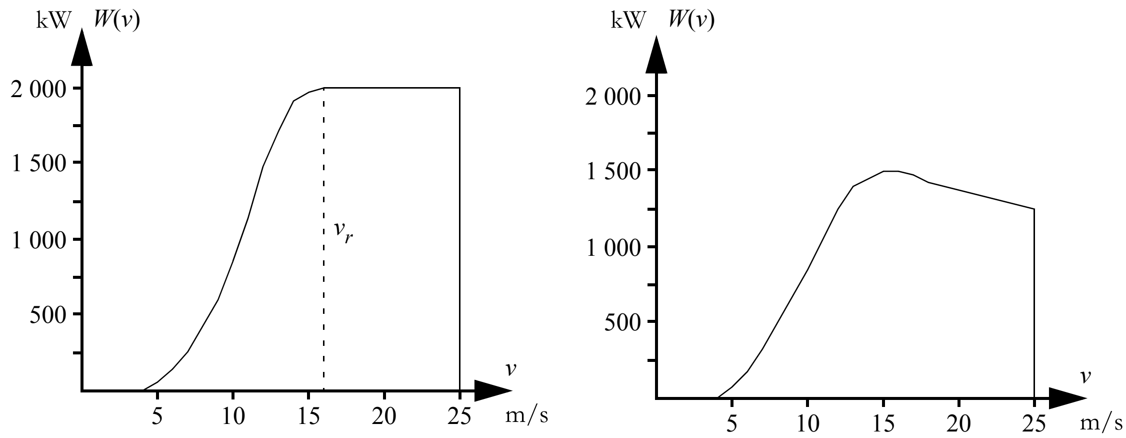


Figure 6.10 The relation between wind speed and electricity generation for two kinds of wind power plants (Vestas V80-2.0 MW and NEG Micon NM 1500C).

$$MTTR = \frac{1}{4}((23-20)+(62-60)+(74-70)+(104-101)) = 3 \text{ weeks.}$$

The failure and repair rates can now be calculated according to their definitions:

$$\lambda = \frac{1}{MTTF} \approx 0.0435 \text{ failures/week (when the power plant is available),}$$

$$\mu = \frac{1}{MTTR} \approx 0.3333 \text{ repairs/week (when the power plant is unavailable).}$$

When calculating the unavailable we assume that the two years recorded in the operation log are representative; hence, the unavailability is estimated to

$$q = \frac{MTTR}{MTTF + MTTR} = \frac{3}{3+23} \approx 0.1154.$$

6.2.4 Wind Power Model

A wind power plant converts the kinetic energy of the wind into electric energy. Since the wind speed is varying so will the electricity generation. Figure 6.10 shows the electricity generation as a function of the wind speed for two different types of wind power plants. The turbine in the left diagram is pitch controlled, which means that when the wind speed is higher than v_r the wind is “spilled” by feathering the blades. The motive for this solution is that it is not necessary to dimension important parts (as for example the gearbox and the generator) for power levels that are rather rare. The optimal level of rated power in this type of power plant is thus a trade-off between the costs of designing the power plant for high wind speeds and the income of the energy sales. The wind power plant in the right diagram is stall regulated, which have a fixed blade angle. However, at high wind speeds, some of the wind passes by due to turbulence. An advantage of this type of turbine compared to the pitch controlled units is that there is no need to invest in control equipment. A minor disadvantage is that the peak capacity is not exactly defined and the peak is only obtained for a relatively small wind speed interval. Both types of units are shut down at high wind speeds, because the wind gets too turbulent then, which causes unacceptable strains on the unit.

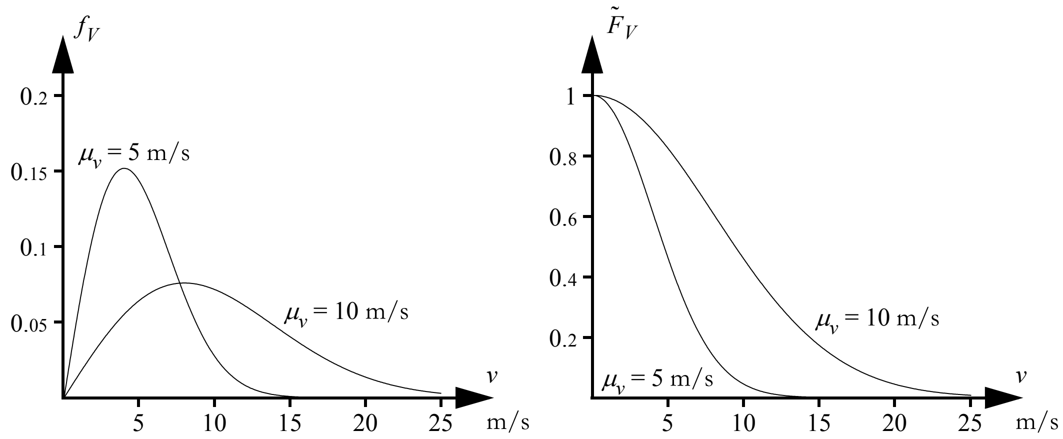


Figure 6.11 The density function and duration curves of Rayleigh distributed wind speeds.

Available Generation Capacity

The electricity generation in a wind power plant is depending on whether the plant is technically available or not, as well as the current wind speed. To obtain a density function of the available generation capacity of a wind power plant it is therefore necessary to include both the availability of the unit and the wind speed distribution.

If actual wind speed statistics for a site is available then these statistics should of course be used as basis for the wind power model. However, if such statistics are not available then it is common to assume that the wind is *Rayleigh distributed*. This means that the wind speed has the following duration curve:

$$\tilde{F}_V(v) = \begin{cases} 1 & v < 0, \\ e^{-(v/\alpha)^2} & v \geq 0, \end{cases} \tag{6.20}$$

where α is a scaling parameter. Figure 6.11 shows examples of the density function and duration curve of Rayleigh distributed wind speeds with the mean wind speed 5 m/s and 10 m/s respectively. As can be seen, the Rayleigh distribution resembles the normal distribution; the main difference is that a Rayleigh distributed random variable never can obtain negative values. The relation between α and the mean wind speed can be calculated using theorem C.10:

$$\mu_v = 0 + \int_0^\infty \tilde{F}_V(v)dv = \int_0^\infty e^{-(v/\alpha)^2}dv = \frac{1}{2}\alpha\sqrt{\pi} \approx 0.89\alpha. \tag{6.21}$$

Let us now study the probability that the available wind power generation capacity in a 100% reliable wind power plant is larger than W kW. If $v(W)$ is the inverse function of the electricity generation as function of the wind speed, W, and v_{co} is the wind speed for which the wind power plant is disconnected, then the wanted probability is equal to the probability that the wind speed belongs to the interval $[v(W), v_{co}]$, i.e.,

$$\tilde{F}_W(W) = \tilde{F}_V(v(W)) - \tilde{F}_V(v_{co}) = e^{-\left(\frac{(v(W))}{\alpha}\right)^2} - e^{-\left(\frac{v_{co}}{\alpha}\right)^2}. \tag{6.22}$$

It is of course only possible to apply (6.22) when the function $v(W)$ exists. This is the case for pitch controlled wind power plants, where $v(W)$ is the inverse function of $W(v)$ in the interval $v \in (0, v_r)$.

For stall regulated units there is no inverse function of $W(v)$ in that interval (cf. figure 6.10). In order to model stall regulated wind power plants it is necessary to approximate the generation function $W(v)$ by a non-decreasing function.

There will be a strong correlation between the generation in wind power units whenever several units are located in the same area, since the wind speed will be more or less the same for all units.[17] To fulfil the requirement that all random variables in a probabilistic production cost simulation should be independent it is simplest to create a single model for the total generation of the wind farm. In this model we will also consider that the turbines are not 100% available. In order to simplify the calculations, we will neglect the variations of the wind speed and assume that all turbines of the wind farm have exactly the same generation function $W(v)$ and availability p. The total wind power generation is then only depending on the current wind speed and the number of available wind power plants.

The density function of the total wind power generation, W_{tot}, is quite obvious when all turbines are unavailable:

$$f^{0}_{W_{tot}} = \begin{cases} 1 & W_{tot} = 0, \\ 0 & W_{tot} \neq 0. \end{cases} \tag{6.23}$$

If exactly n turbines are available ($n > 0$) then the probability that the total wind power generation equals W_{tot} is equal to the probability that each of the available turbines is generating W_{tot}/n, as we assume that the wind speed is the same for all units and all units are generating the same power for a given wind speed. Thus, we get the following probability density function:

$$f^{n}_{W_{tot}}(W_{tot}) = f_W\left(\frac{W_{tot}}{n}\right) = -\frac{d}{dW_{tot}}\tilde{F}_W\left(\frac{W_{tot}}{n}\right),\ n > 0, \tag{6.24}$$

where $\tilde{F}_W$ is given by (6.22).

To consider that the number of available turbines also is a random variable we can summarise the density functions according to (6.23) and (6.24), while weighting each density function by the probability of having exactly n units available. The result is the following probability density for a wind farm of N power plants:

$$f_{W_{tot}}(W_{tot}) = \sum_{n=0}^{N} p^n q^{(N-n)} \binom{N}{n} f^{n}_{W_{tot}}(W_{tot}), \tag{6.25}$$

where

p = availability of a single unit,
q = unavailability of a single unit = $1 - p$,
$\binom{N}{n}$ = binomial coefficient, i.e, the number of ways to select n elements from a total population of N elements

It should be noted that W_{tot} according to (6.25) is a continuous random variable. Those are, as already have been mentioned in section 6.2.2, impractical to use when performing probabilistic production cost simulations; therefore, it is appropriate to divide these density functions into segments too. It is recommended to use the same segment size as for the segmentation of the load duration curve.

Example 6.14 (model of a small wind farm): Assume that there is a wind farm with two identical units. From the wind conditions at the site the following

17. Small variations will be caused by the topology of the site. Moreover, windward units will "steal" some wind from leeward units.

model of the individual plants has been obtained (excluding the probability of failures):

$$f_W(W) = \begin{cases} 0.2 & W = 0, \\ 0.4 & W = 300, \\ 0.4 & W = 600. \end{cases}$$

Determine a model of the total power generation in the wind farm if the availability of the units are 99%.

Solution: If none of the units is available then we get the trivial frequency function $f^0_{W_{tot}} = 1$ when W_{tot} equals zero; otherwise we have $f^0_{W_{tot}} = 0$. If one of the units is available then $f^1_{W_{tot}}(W_{tot}) = f_W(W_{tot})$, and when both units are available the result is $f^2_{W_{tot}}(W_{tot}) = f_W(W_{tot}/2)$, i.e.,

$$f^2_{W_{tot}}(W_{tot}) = \begin{cases} 0.2 & W_{tot} = 0, \\ 0.4 & W_{tot} = 600, \\ 0.4 & W_{tot} = 1\,200. \end{cases}$$

The probability that none of the units is available is $0{,}01 \cdot 0{,}01 = 10^{-4}$. The probability that the first unit is available but not the second is $0{,}99 \cdot 0{,}01 = 0{,}0099$, which is as probable as the second unit being available but not the first. Hence, the probability that one unit is available is $2 \cdot 0{,}0099 = 0{,}0198$. Finally, the probability of both units being available is $0{,}99 \cdot 0{,}99 = 0{,}9801$. The frequency function of the total generation is then

$$f_{W_{tot}}(W_{tot}) = 0.0001 f^0_{W_{tot}}(W_{tot}) + 0.0198 f^1_{W_{tot}}(W_{tot}) + 0.9801 f^2_{W_{tot}}(W_{tot}) =$$

$$= \begin{cases} 0.0001 \cdot 1 + 0.0198 \cdot 0.2 + 0.9801 \cdot 0.2 = 0.20008 & W_{tot} = 0, \\ 0.0001 \cdot 0 + 0.0198 \cdot 0.4 + 0.9801 \cdot 0 = 0.00792 & W_{tot} = 300, \\ 0.0001 \cdot 0 + 0.0198 \cdot 0.4 + 0.9801 \cdot 0.4 = 0.39996 & W_{tot} = 600, \\ 0.0001 \cdot 0 + 0.0198 \cdot 0 + 0.9801 \cdot 0.4 = 0.39204 & W_{tot} = 1\,200. \end{cases}$$

Calculation of System Indices for Wind Power Systems

The convolution formula (6.13) is based on the assumption that a power plant either is generating its installed capacity or nothing at all. Moreover, it is assumed that load variations and failures in power plants are independent, i.e., failures are equally likely for all load levels. These assumptions are however not correct for wind power.

The wind power generation can, as described above, take any value between zero and installed capacity. Each of these levels have a given probability which is stated by the probability density $f_W(x)$. To calculate the equivalent load duration curve we can generalise (6.13) to the following equation:

$$\tilde{F}_g(x) = \sum_{i=1}^{N_g} p_{g,i} \tilde{F}_{g-1}(x - x_{g,i}), \tag{6.26}$$

where

N_g = number of states in power plant g,
$p_{g,i} = f_{W_g}(\overline{W} - x_{g,i})$ = probability of state i,

$x_{g,\ i}$ = outage (compared to installed capacity) in the i:th state.

It is easy to verify that (6.26) yields the same result as (6.13) for a two-state model where $p_{g,\ 1}$ equals the availability p_g and $p_{g,\ 2}$ equals the unavailability q_g.

When the equivalent load duration curve has been calculated, the system indices can be determined using the equations stated in section 6.2.1. It must however be noted that the expected generation in a power plant having more than two possible states can not be calculated using (6.18); it is only possible to use (6.17).

Before we study an example of how to apply (6.26), we will briefly comment upon the problem that where is often a correlation between the available generation capacity and the load. This correlation can be of three types:

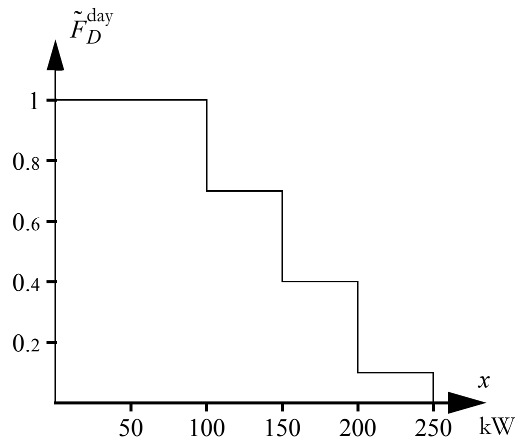
- **Yearly variation.** The wind power generation is to some extent depending on the season. In Sweden wind speeds are higher during the fall and winter than during the summer. The load also has a yearly variation, which among other things depend on temperature (affects electric heating), light variation (affects lighting) and vacations (affects the industrial load). This means that there is some correlation between wind power generation and the load. The convolution formula (6.26) is however based on the assumption that all random variables are independent. To consider the yearly variation it is therefore necessary to divide the year in different periods of similar wind and load conditions, so that it is reasonable to assume that wind power generation and load is independent within each period. The wanted system indices are then calculated separately for each period. The total value of the system indices are obtained by weighting together the results of the different periods, considering the period length.
- **Daily variation.** The load normally has a typical daily variation, where the load is higher during the day than during the night. The wind speed during the summer may also have a daily variation at some locations, resulting in higher wind speeds during the night. These correlations can be considered in a similar manner as the yearly variations, i.e., by dividing the day and night into separate periods and calculate separate system indices for these periods.
- **Special couplings.** In some situations there might be a special coupling between the wind power generation and the load. An example of such a special coupling, which has been discussed in Sweden, is the fact that strong winds cool houses. This means that houses with electric heating will increase at the same time as the wind power generation increases. However, it is not certain whether this really is the case, since the subject depends on complicated relations as how many electrically heated houses there are in an area, how well isolated they are, time constants (i.e., how long time it has to be windy before the houses are cooled down), etc.

One method to avoid these couplings is to divide the simulation in separate time periods, so that the assumption of independent scenario parameters is fulfilled *within each time period*. The following example illustrates how this can be done:

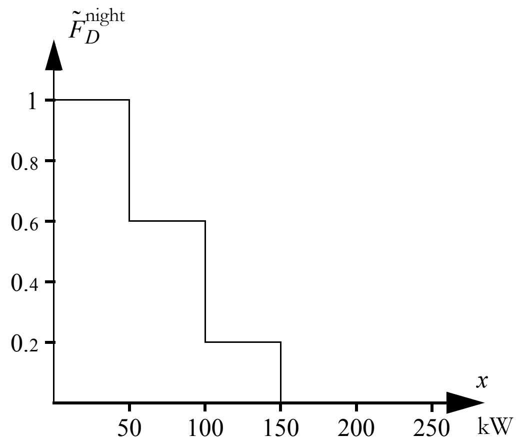
Example 6.15 (the power system in a small island): The small island Kobben is not connected to the national grid, but there is a local grid which is powered by a small wind power plant (installed capacity 200 kW) and a diesel generator set. The diesel generator set has a maximal capacity of 200 kW and the availability is 95%. A simplified model of the wind power plant is stated in table 6.5. The load duration curve is shown in figure 6.12. Calculate the risk of power deficit in this system.

Solution: We start by calculating the system $LOLP$ during day time. The total installed capacity is 400 kW; hence, we get

$$LOLP^{\text{day}} = \tilde{F}_2^{\text{day}}(400) = 0.95\tilde{F}_1^{\text{day}}(400) + 0.05\tilde{F}_1^{\text{day}}(400-200) =$$



a) Day time (7 a.m. to 9 p.m.)



b) Night time (9 p.m. to 7 a.m.).

Figure 6.12 The load for the small island in example 6.15.

Table 6.5 Model of the wind power plant in example 6.15.

Outage [kW]	Probability during the day [%]	Probability during the night [%]
0	30	10
100	40	50
200	30	40

$$= 0.95 \cdot (0.3\tilde{F}_D^{\text{day}}(400-0) + 0.4\tilde{F}_D^{\text{day}}(400-100) + 0.3\tilde{F}_D^{\text{day}}(400-200)) +$$

$$+ 0.05 \cdot (0.3\tilde{F}_D^{\text{day}}(200-0) + 0.4\tilde{F}_D^{\text{day}}(200-100) + 0.3\tilde{F}_D^{\text{day}}(200-200)) =$$

$$= 0.95 \cdot (0.3 \cdot 0 + 0.4 \cdot 0 + 0.3 \cdot 0.1) + 0.05 \cdot (0.3 \cdot 0.1 + 0.4 \cdot 0.7 + 0.3 \cdot 1) = 5.9\%.$$

In the same way we can calculate the night time *LOLP*:

$$LOLP^{\text{night}} = 0.95 \cdot (0.1 \cdot 0 + 0.5 \cdot 0 + 0.4 \cdot 0) +$$
$$+ 0.05 \cdot (0.1 \cdot 0 + 0.5 \cdot 0.2 + 0.4 \cdot 1) = 2.5\%.$$

To calculate the total system *LOLP* we use the weighted average of these two values:

$$LOLP = \frac{14}{24}LOLP^{\text{day}} + \frac{10}{24}LOLP^{\text{night}} \approx 4.5\%.$$

6.2.5 Model of Dispatchable Hydro Power

Hydro power plants can be divided into two types: run-of-the-river plants and dispatchable plants. A run-of-the-river plant has no reservoir, which means that the available generation capacity is determined by whether or not the power plant is technically available, and the water flow at the site of the plant. Thus, a run-of-the-river plant has similar properties as a wind power plant and can be included in a probabilistic production cost simulation using models resembling those described in section 6.2.4.

Dispatchable hydro power, i.e., hydro power plants having reservoirs, are not as dependent of the river flow, since water from the reservoir can be used when the natural flow is low. It is rather the available energy (i.e., the size of the inflow to the reservoirs during a year or some other appropriate time period) which limits the electricity generation in these plants. As hydro power plants in general have negligible or very small operation costs, the best way to utilise the available dispatchable hydro energy is to replace the generation in the most expensive thermal power plants. In this section we will describe a simple method to approximately determine which thermal power plant should be replaced given a particular energy limit of the dispatchable hydro power.

The Impact of Hydro Power on Operation Costs

The operation cost of a hydro power plant is generally smaller than the operation cost of thermal power plants. Therefore, it is obvious that in a hydro-thermal power system the total operation cost, $ETOC$, will be reduced the more hydro energy which is available during the studied period. What the need is a method to determine the total operation cost as a function of the available hydro energy, i.e., $ETOC(W_H)$.

Probabilistic production cost is as described earlier based on the assumption that the power plants are dispatched in a given merit order, which means that the g:th unit only is used when the equivalent load can not be covered by power plants 1, …, g – 1. The power plant which is first in the merit order will be generating as much as possible. If the first and second unit are switched then the former first power plant will decrease its generation, since a part of the load will be covered by the former second power plant instead. In other words, the expected generation of a power plant will decrease whenever the power plant is moved down in the merit order. This enables us to calculate the total operation cost for at least some levels of the expected energy generation in the hydro power. If a system has $\mathcal{G}$ thermal power plants and one dispatchable hydro power plant then we must perform $\mathcal{G} + 1$ separate simulations in order to try all positions of the hydro power, which undeniably would require quite a lot of calculations. Fortunately, there is a short cut.

Assume that we are given a merit order and then shifts power plant g and power plant $g + 1$. This will apparently not affect the expected energy generation of those power plants which are dispatched before the g:th unit, i.e., power plant 1, …, g – 1, because these power plants will see exactly the same equivalent load as before; the equivalent load is as we know equal to the real load plus outages in higher prioritised units. It is less obvious that the expected energy generation of the power plants positioned later than unit g + 1 in the merit order are not affected neither!

> ***Theorem 6.20.*** Assume that there are two power plants having the installed capacities $\hat{G}_A$ and $\hat{G}_B$ respectively, and the availability p_A and p_B respectively. Further, assume that these power plants are used to cover the equivalent load $\tilde{F}_g$. he same equivalent load duration curve (i.e., $\tilde{F}_{g+2}$), will then be obtained, regardless of in which order the units are dispatched.

Proof: We try to add unit A first:

$$\tilde{F}^{AB}_{g+1}(x) = p_A\tilde{F}_g(x) + q_A\tilde{F}_g(x - \hat{G}_A),$$

$$\tilde{F}^{AB}_{g+2}(x) = p_B\tilde{F}^{AB}_{g+1}(x) + q_B\tilde{F}^{AB}_{g+1}(x - \hat{G}_B) =$$
$$= p_B(p_A\tilde{F}_g(x) + q_A\tilde{F}_g(x - \hat{G}_A)) + q_B(p_A\tilde{F}_g(x - \hat{G}_B) + q_A\tilde{F}_g(x - \hat{G}_A - \hat{G}_B)).$$

Then we try to add unit B first:

$$\tilde{F}^{BA}_{g+1}(x) = p_B\tilde{F}_g(x) + q_B\tilde{F}_g(x - \hat{G}_B),$$

$$\tilde{F}^{BA}_{g+2}(x) = p_A\tilde{F}^{BA}_{g+1}(x) + q_A\tilde{F}^{BA}_{g+1}(x - \hat{G}_A) =$$

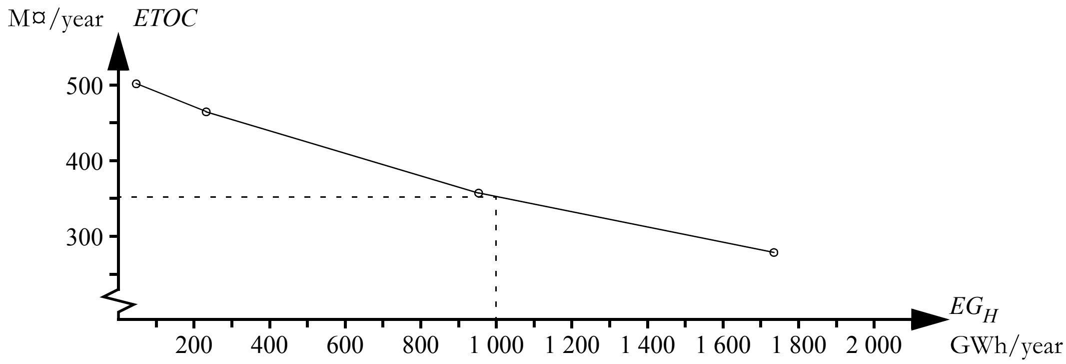


Figure 6.14 Estimation of the total operation cost in a system with dispatchable hydro power as function of the available hydro energy. The points obtained from the simulations are marked by small circles. Between these points it is assumed that the function is piecewise linear.

$$= p_A(p_B\tilde{F}_g(x) + q_B\tilde{F}_g(x - \hat{G}_B)) + q_A(p_B\tilde{F}_g(x - \hat{G}_A) + q_B\tilde{F}_g(x - \hat{G}_B - \hat{G}_A)).$$

The result is clearly that $\tilde{F}_{g+2}(x) = \tilde{F}^{AB}_{g+2}(x) = \tilde{F}^{BA}_{g+2}(x)$. ∎

Since the expected generation only is affected in the two power plants which are switched, there will only be two possible levels of the expected generation of the thermal power plants; one level if the power plant is dispatched before the hydro power and another—lower—level if the thermal power plant is dispatched after the hydro power. Thus, by performing two simulations, one having the hydro power as the first power plant in the merit order and one there it is last in the merit order, it will be possible to calculate both levels of the expected generation in all thermal power plants. With that, the expected generation of all thermal power plants can be calculated given a particular priority for the hydro power. Knowing the expected generation it is also possible to calculate the*ETOC* using (6.19). The result is a few points of the function $ETOC(W_H)$; or the remaining values of W_H the *ETOC can be determined using linear interpolation between the two closes known points*.

Example 6.16 (simulation of dispatchable hydro): The power plants in a certain electricity market are listed in table 6.6. The hydro power plant is dispatchable and the annual inflow is 1 TWh. The load duration curve is shown in figure 6.13. Determine the expected operation cost of the system.

Solution: The simulations are performed according to the equations derived in section 6.2.1. The calculations need no special considerations, but are comparatively space consuming and are therefore excluded. The result of four simulations are shown in table 6.6. Actually, it would have been sufficient to simulate the system with the merit orders H, A, B, C and A, B, C, H respectively; the other two positions were only simulated to illustrate the claim that the expected generation in the thermal power plants only has two possible levels, depending on whether the unit is dispatched before or after the hydro power. We can also notice that the *EENS* and *LOLP* is not affected by the merit order of the power plants.

Using the results from table 6.6 it is possible to estimate the operation cost as a function of the available hydro energy, as shown in figure 6.14. Using linear interpolation we obtain $ETOC \approx 351{,}82$ M¤/year in this case.

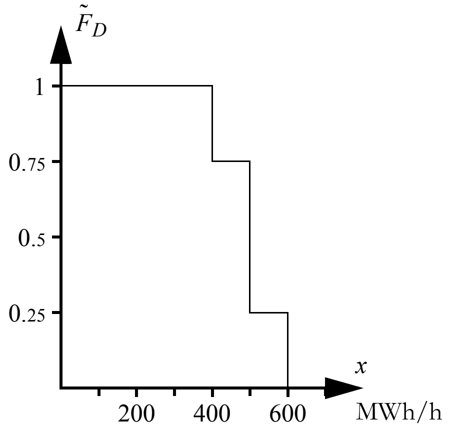


Figure 6.13 Load duration curve in example 6.16.

Table 6.6 Power plants in example 6.16.

Power plant	Capacity [MW]	Availa-bility [%]	Operation cost [¤/MWh]
Dispatchable hydro H	200	99	0
Thermal plant A	400	90	100
Thermal plant B	300	90	150
Thermal plant C	200	95	200

Table 6.6 Results of the simulations in example 6.16.

System index	Merit order			
	H, A, B, C	A, H, B, C	A, B, H, C	A, B, C, H
EG_A [GWh/year]	2 373.08	3 153.60	3 153.60	3 153.60
EG_B [GWh/year]	224.10	224.10	946.08	946.08
EG_C [GWh/year]	37.24	37.24	37.24	222.61
EG_H [GWh/year]	1 734.48	953.96	231.99	46.61
ETOC [M¤/year]	278.37	356.42	464.72	501.79
EENS [GWh/year]	11.09	11.09	11.09	11.09
LOLP [%]	0.9	0.9	0.9	0.9

The Water Value

The water stored in the reservoirs has a value, as it can be used to replace future generation in more expensive power plants. The marginal water value (often referred to just as the water value) is the decrease in the total operation cost given an extra MWh of available hydro energy. The unit of water value is therefore ¤/MWh. When considering the function $ETOC(W_H)$ the water value is corresponding to the slope of the linear segments, which means that we will get different water values depending on the amount of available hydro energy. The explanation is that the dispatchable hydro generation will be used to replace the most expensive thermal power plant. When the expected generation of that plant is decreased to zero, then the hydro power will replace the second most expensive thermal power plant, etc. As the water value is equal to the operation cost of the last replaced thermal power plant it will be highest when there is just so much water available that a part of the energy generation in the most expensive thermal power plant can be replaced.

6.2.6 Simplified Calculation of the Loss of Load Probability

As shown in the earlier parts of section 6.2 it is necessary to perform quite a lot of calculations during probabilistic production cost simulations. The large amount of calculations is primarily due to the convolution equations. For a real system with a large number of power plants of different sizes and using a relatively large accuracy in the segmentation of the load duration curve (i.e., a small segment size), the number of calculations become very large. In this section we will present a faster, simplified method, which can be used to estimate the *LOLP*.

As mentioned in section 6.2.2 the load is often modelled by one or more normal distributions. If that is not the case and a typical load profile is used instead, it is still possible to obtain a normal approximation of the load:

$$\mu_D = E[D], \tag{6.27a}$$

$$\sigma_D = \sqrt{Var[D]}, \tag{6.27b}$$

i.e., it is assumed that the load is normally distributed and having the same expectation value and standard deviation as the real load. The difference will hence be that the real load is likely to have another probability distribution. The expectation value and standard deviation can be calculated using the definitions and theorems stated in appendix C.

Example 6.17 (normal approximation of a load): Determine a normal approximation of the load in examples 6.9 and 6.10.

Solution: Since the frequency function of the load is known (from the solution to example 6.10), we can directly apply definitions C.6 and C.7:

$$\mu_D = \sum_x x f_D(x) = 400\cdot 0.125 + 500\cdot 0.375 + 600\cdot 0.375 + 700\cdot 0.125 =$$
$$= 550 \text{ MW},$$

$$\sigma_D = \sqrt{\sum_x (x - \mu_D)^2 f_D(x)} =$$
$$= \sqrt{(-150)^2 \cdot 0.125 + (-50)^2 \cdot 0.375 + 50^2 \cdot 0.375 + 150^2 \cdot 0.125} \approx$$
$$\approx 86.6 \text{ MW}.$$

Outages in power plants can also be considered to be approximately normally distributed. For a two-state model we get

$$\mu_{O_g} = E[O_g] = p_g \cdot 0 + q_g \cdot \hat{G}_g = q_g \cdot \hat{G}_g, \tag{6.28a}$$

$$\sigma_{O_g} = \sqrt{Var[O_g]} = \sqrt{E[(O_g - \mu_{O_g})^2]} = \sqrt{p_g(0 - \mu_{O_g})^2 + q_g(\hat{G}_g - \mu_{O_g})^2} =$$
$$= \sqrt{p_g \mu_{O_g}^2 + q_g \hat{G}_g^2 - 2q_g \mu_{O_g} \hat{G}_g + q_g \mu_{O_g}^2} = \{p_g = 1 - q_g\} =$$
$$= \sqrt{\mu_{O_g}^2 + q_g \hat{G}_g^2 - 2\mu_{O_g} q_g \hat{G}_g} = \{\mu_{O_g} = q_g \cdot \hat{G}_g\} = \sqrt{q_g \hat{G}_g^2 - (q_g \hat{G}_g)^2} =$$
$$= \hat{G}_g \sqrt{q_g - q_g^2} = \hat{G}_g \sqrt{q_g p_g}. \tag{6.28b}$$

For power plants having more than two possible states it is in most cases best to directly apply the definitions of expectation value and standard deviation.

Example 6.18 (normal approximation of a wind farm): Determine a normal approximation of the wind farm from example 6.14

Solution: The probability of having a given capacity available in the wind farm has already been calculated in example 6.14. The probability of an outage $x = 1\,200 - W_{tot}$ is

$$f_{O_v}(x) = \begin{cases} 0.39204 & x = 0, \\ 0.39996 & x = 600, \\ 0.00792 & x = 900, \\ 0.20008 & x = 1\,200. \end{cases}$$

According to definitions C.6 and C.7 we get

$$\mu_{O_v} = \sum_x x f_{O_v}(x) = 0.39204 \cdot 0 + 0.39996 \cdot 600 + 0.00792 \cdot 900 + 0.20008 \cdot 1\,200 =$$
$$= 487.2 \text{ kW},$$

$$\sigma_{O_v} = \sqrt{\sum_x (x - \mu_{O_v})^2 (f_{O_v}(x))} = \ldots \approx 315.7 \text{ kW}.$$

The basis of the convolution equation (6.13) and its generalised variant (6.26) is that the equivalent load is the sum of the actual load and outages in power plants. When both load and outages are modelled by normal distributions it becomes possible to replace the convolution by summarising normal distributions, which means that the equivalent load duration curve of g power plants can—according to theorem D.4—be written as

$$\tilde{F}_g^N \in N(\mu_g, \sigma_g) \tag{6.29}$$

where

$$\mu_g = \mu_D + \sum_{i=1}^{g} \mu_{O_g}, \tag{6.29a}$$

$$\sigma_g = \sqrt{\sigma_D^2 + \sum_{i=1}^{g} \sigma_{O_g}^2}. \tag{6.29b}$$

A power deficit occurs when the equivalent load exceeds the installed capacity. The probability of this is determined by

$$LOLP_g \approx \tilde{F}_g^N(\hat{G}_g^{tot}) = 1 - \Phi\left(\frac{\hat{G}_g^{tot} - \mu_g}{\sigma_g}\right). \tag{6.30}$$

It is important to remember that this approximation of the loss of load probability is based on the assumption that the equivalent load is at least approximately normally distributed. This is a reasonable assumption for large systems with approximately normally distributed load and power plants of similar size.

Example 6.19. Assume that the load in example 6.17 should be covered by $\mathcal{G}$ power plants of 200 MW each. The availability of each plant is 80%. Calculate the risk of power deficit for $\mathcal{G} = 1, \ldots, 12$.

Solution: he normal approximation of a single power plant can be obtained using (6.28a) and (6.28b):

$$\mu_{O_g} = 0.2 \cdot 200 = 40 \text{ MW},$$

$$\sigma_{O_g} = 200\sqrt{0.2 \cdot 0.8} = 80 \text{ MW}.$$

Table 6.7 The risk of power deficit in example 6.19.

Number of power plants	*LOLP*	
	Using the simplified method	Using the exact method
1	0.9995	1
2	0.9468	0.9200
3	0.6658	0.5040
4	0.3104	0.2128
5	0.1042	0.0771
6	0.0278	0.0253
7	$6.3435 \cdot 10^{-3}$	$7.7184 \cdot 10^{-3}$
8	$1.2933 \cdot 10^{-3}$	$2.2349 \cdot 10^{-3}$
9	$2.4316 \cdot 10^{-4}$	$6.2106 \cdot 10^{-4}$
10	$4.3046 \cdot 10^{-5}$	$1.6701 \cdot 10^{-5}$
11	$7.2789 \cdot 10^{-6}$	$4.3725 \cdot 10^{-6}$
12	$1.1878 \cdot 10^{-6}$	$1.1194 \cdot 10^{-6}$

We can now apply (6.30). The results are listed in table 6.7. For comparison, the results of the "exact" simulation[18] using the equations (6.13) and (6.16) are also shown.

Apparently it is possible to obtain results with rather large relative errors in the estimation of the *LOLP* using the simplified method, but the size of order is still correct.

6.3 MONTE CARLO SIMULATION

It is not always possible to directly calculate expectation values according to their definition (which is found inappendix C). The problem can either be that the integral itself is hard to solve or that the integrand is not known, for example because the density function of the random variable is unknown. An alternative for these cases is to use so-called Monte Carlo methods, which are based on estimation of the properties of the variable by random observations. In section 6.1 we showed that the system indices to be calculated in an electricity market simulation actually are the expectation values of various result variables. Hence, Monte Carlo methods seem suitable for estimating different system indices.

In this section we will describe basic methods which can be used for Monte Carlo simulation of electricity markets. These methods are applied on fairly simple examples.

6.3.1 Simple Sampling

By simple sampling we refer to a sample survey, where we completely randomly select observations—samples—of a random variable. From these samples it is possible to obtain an estimate of the probability distribution of the random variable.

18. Remember that the earlier described method also is based on simplifications and assumptions.

Theory

The basic idea of all Monte Carlo methods is that the expectation value of a random variable can be estimated by random observations of the variable. The expectation value can be interpreted as the mean of an infinite series of observation of the random variable.[19] It is—for natural reasons—impossible to perform an infinite series of experiments to determine the expectation value. However, the more observations that are made of a random variable, the more *likely* it will be that the mean of these observations is close to the expectation value. The expectation value can thus be estimated by calculating the mean of a large enough number of observations. This is referred to as *simple sampling*.

Theorem 6.21. If there are n independent observations, $x_1, \ldots, x_n$, of the random variable X then the mean of these observations, i.e.,

$$m_X = \frac{1}{n}\sum_{i=1}^{n} x_i$$

is an estimate of $E[X]$.

There is of course a formal proof of this theorem, but we will restrict ourselves to illustrating the principle in a simple example:

Example 6.20 (tossing a coin): Calculate the density function of the mean outcome when a coin is tossed 1, 2, 3, 10, 100 and 1 000 times respectively. Compare these density function to the expectation value of the coin toss.

Solution: Let the random variable C be equal to 0 if the result of tossing the coin was tails and 1 if it was heads. Introduce the symbol H_n for the mean of n coin tosses, i.e.,

$$H_n = \frac{1}{n}\sum_{i=1}^{n} c_i.$$

Notice that H_n is also a random variable, as it is a function of the sum of a number of random observations.

f the coin is tossed n times then it is possible to obtain 2^n different outcomes, provided that we take into account in which order the results were received. In this case the order does not matter, as it is only the total number of heads that affect the mean. It is therefore sufficient to know how many of the 2^n outcomes correspond to a given value of H_n.

Let h be the number of heads in a series of observations. The number of ways to select h elements out of a population of n elements is given by the binomial coefficient

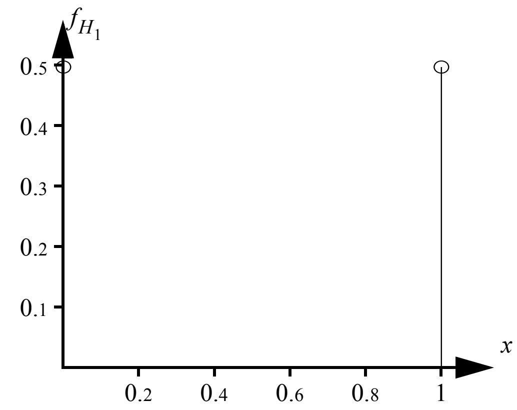
$$\binom{n}{h} = \frac{n!}{h!(n-h)!}.$$

For $n = 1$ we get the binomial coefficients

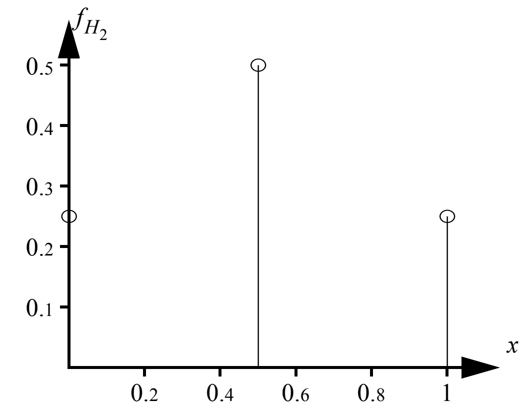
$$\binom{1}{0} = 1, \binom{1}{1} = 1.$$

The interpretation is that when one experiment is performed then there is only one way to have the outcome heads 0 times and 1 time respectively. For $n = 2$ we get

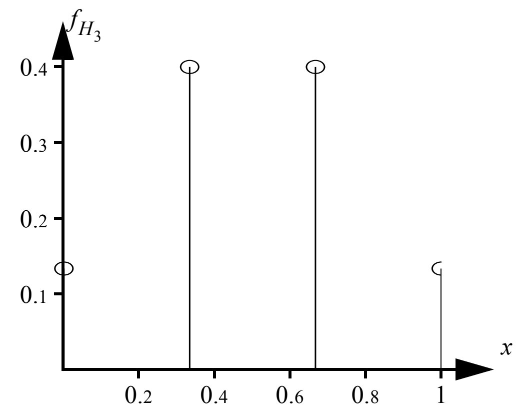
19. Cf. definition C.6.



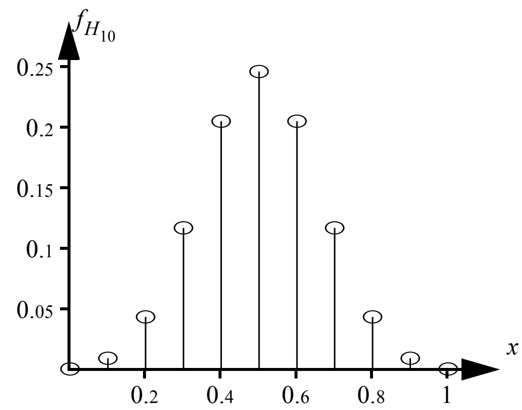
a) Density function after one trial.



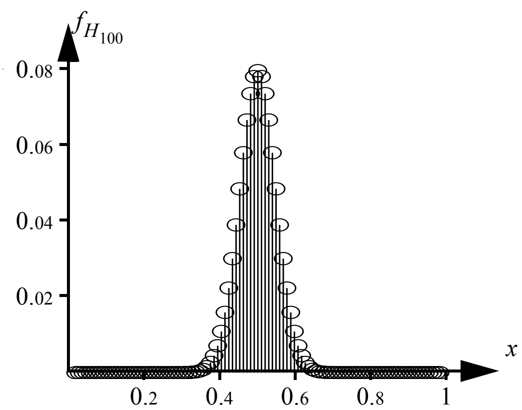
b) Density function after two trials.



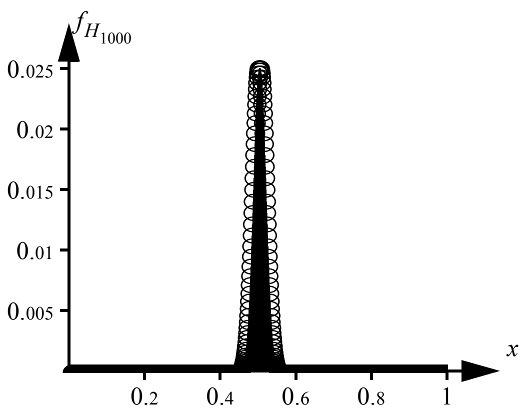
c) Density function after three trials.



d) Density function after ten trials.



e) Density function after 100 trials.



f) Density function after 1 000 trials.

Figure 6.15 Density function for the mean result tossing a coin.

$\binom{2}{0} = 1, \binom{2}{1} = 2, \binom{2}{2} = 1;$

when the experiment is performed twice there is of course one way to obtain tails both times ($h = 0$) or to obtain heads both times ($h = 2$), but there are two ways to get heads once ($h = 1$).

In the density function of H_n we have to consider the probability of each outcome (with regard to order) and the number of combinations which have the same result

$$f_{H_n}(x) = \frac{1}{2^n}\binom{n}{h} \text{ for } x \text{ such that } x = \frac{h}{n},\ h = 0, \ldots, n.$$

sing the formula above we can calculate the sought density functions, which are shown in figure 6.15.

The expectation value of a coin toss is $E[C] = 0.5$ (according to definition C.6). It is clearly seen in the figure that the probability of the mean result being close to 0.5 is increasing with the number of experiments.

The example of tossing a coin illustrates two interesting facts; the mean result is really an estimation of the expectation value of the studied random variable, and the estimate *is in itself a random variable*. We can also notice that it is not guaranteed that another experiment increases the accuracy of the estimate. If the coin for example is tossed twice then there is a 50% probability of obtaining an estimate which is exactly equal to the theoretical value 0.5. f the coin is tossed once more then it becomes impossible to obtain the theoretical value! However, the probability of getting an erroneous result becomes very low when the number of experiments is large enough. If the coin is tossed 1 000 times then the probability is more than 90% that the estimate will be somewhere in the interval between 0.475 and 0.525, i.e., a relative error less than 5%.

Since the estimate m_X is a random variable, it has a certain expectation value and a certain variance. The expectation value is equal to the expectation value of the studied variable, i.e., $E[m_X] = E[X]$. (If this was not the case then m_X would not be an estimate of $E[X]$.) The variance of the estimate is interesting, because the variance is a measure of how much a random variable can be expected to deviate from its expectation value; the larger the variance, the larger the deviations.[20] It is therefore desirable that the estimate has as small variance as possible. It can be shown that the following theorem is valid for the variance of the estimate:

Theorem 6.22. The variance of the estimate from simple sampling is

$$Var[m_X] = \frac{Var[X]}{n}.$$

The theorem shows that the more samples we take, the lower variance of the estimate we get. This confirms the observation from example 6.20 hat the probability of a good estimate increases as the number of samples increase.The question is then how to know when a sufficient number of samples have been collected to obtain a result with sufficient accuracy. In other words, we need to have some kind of simple test to decide whether the simulation can be stopped or if more samples should be taken. Such tests—or convergence criteria—can be designed in several ways. Below follows two simple methods which can be applied in most cases. However, first we must stress the fact that there are no guarantees that a correct result is obtained by a Monte Carlo method. Even if we toss a coin a million times, there is a possibility that we just get the outcome heads.[21] It is always necessary to accept that there is a certain probability that the result is not accurate enough.

20. Cf. definition C.7.

21. Such serious errors are of course *extremely* unlikely when the number of samples is large.

A very simple method to decide how long a simulation should be going on is to decide the number of samples in advance. The number of samples can be chosen intuitively (if similar systems have been simulated earlier) or calculated. The latter alternative requires some knowledge about the problem to be solved, as shown in the following example:

Example 6.21. An electricity market should be simulated using a multi-area model. The result $LOLP = 1.0\%$ was obtained from a probabilistic production cost simulation. The true value of the *LOLP* should be slightly higher, since the multi-area model includes the transmission losses. How man scenarios should be studied if it is desirable that there is a 95% probability that the estimate should be within ±0.05% of the true value?[22] Assume that no variance reduction techniques are applied and that the estimates m_{LOLO} can be assumed to be normally distributed around the true value.

Solution: It is 95% chance that a $N(\mu, \sigma)$-distributed random variable belongs to the interval $\mu \pm 1.96\sigma$ (cf. appendix D). The standard deviation of the estimate, $\sqrt{Var[m_{LOLO}]}$, must therefore be less than $0.0005/1.96 \approx 0.000255$. According to theorem 6.22 we get

$$Var[m_{LOLO}] = \frac{Var[LOLO]}{n}.$$

$Var[LOLO]$ is unknown, but can be approximated by

$$Var[LOLO_{PPC}] = \{\text{use definition C.7}\} = LOLP_{PPCS}(1 - LOLP_{PPC}).$$

Combining these formulae yields

$$n \geq \frac{0.01 \cdot 0.99}{0.000255^2} \approx 152\ 127.$$

To be on the safe side, we should study about 153 000 scenarios.[23]

An alternative method is to start with a limited number of samples and then try to estimate whether or not the accuracy is acceptable. If it is then the simulation can be ended, otherwise another batch of samples is taken. This approach requires that there is some kind of measure of the accuracy. Such a measure is the so-called coefficient of variation, which is defined as

$$a_X = \frac{\sqrt{Var[m_X]}}{m_X}. \tag{6.31}$$

The variance of the estimate, $Var[m_X]$, is not known when simulating. However, according to theorem 6.22 it is equal to $Var[X]/n$, where $Var[X]$ can be estimated by

$$s_X^2 = \frac{\sum_{i=1}^{n} (x_i - m_X)^2}{n - 1}, \tag{6.32}$$

for simple sampling.

If a_X is less than some relative tolerance, ρ, then the accuracy is assumed to be sufficient. The lower the value of ρ, the more accurate results can be expected. If variance reduction techniques are used efficiently for simulation of electricity markets then a relative tolerance about 10 to 20% should be sufficient for reasonable accurate results.

22. In the sense that if the true *LOLP* is 1.08% then we want the estimate to be in the interval 1.03% to 1.13%.
23. It should be noted that this number can be significantly reduced if variance reduction techniques would be applied.

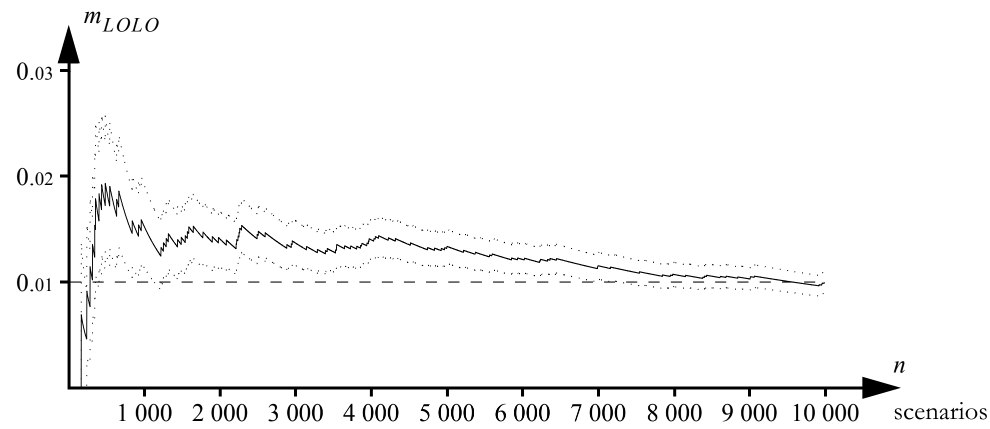


Figure 6.16 Example of how the estimate m_{LOLO} converges to the theoretical value, which in this case is $LOLP = 1\%$ (indicated by a dashed line). The dotted lines show the size of $\sqrt{Var[m_{LOLO}]}$. In this case the relative tolerance was chosen to $\rho = 0.1$.

Some care is necessary when using the coefficient of variation as convergence criterion. As the coefficient of variation is based on the estimate s_X there is a certain risk that an incorrect estimate causes a misleading result. When studying the risk of power deficit it is not unreasonable that the first scenarios all yield $LOLO = 0$, as in the example in figure 6.16. A long as all samples of $LOLO$ have the same value, we will get the estimate $s_{LOLO} = 0$ and hence $a_{LOLO} = 0$. Obviously, it is not sufficient to require that $a_{LOLO} < \rho$, but we will also need to have $s_{LOLO} > 0$. It is possible to imagine a system there $Var[LOLO]$ really is equal to zero, but in practice this means that the system has 100% available power plants with enough capacity to cover every possible load level. Such a system can of course not exist in reality. If $Var[TOC]$should be zeros then all power would have to be generated by power plants with negligible cost. Such a system could obviously exist, but in these cases there is no need to estimate $ETOC$, as we already known that $ETOC = 0$. We may therefore conclude that if $s_X = 0$ or $a_X > \rho$ then the simulation should continue, otherwise it is completed and the result can be displayed.

The problem is that the more samples we select, the longer time will it take to complete the simulation. Therefore, it seems justified to ask whether there is another possibility to increase the accuracy. Notice that theorem 6.22 applies to simple sampling, i.e., when each sample is selected completely at random from the whole population.[24] By using other methods to choose samples than simple sampling it is possible to obtain a lower variance. There are several such alternative methods of generating samples. These methods, which are referred to as *variance reduction techniques*, produce better estimates without increasing the number of samples. In sections 6.3.2-6.3.4 we will describe three methods that are fit for simulation of electricity markets.

Application to Electricity Markets

In the coin tossing example above we generated the random observations using a physical experiment. This method is for natural reasons not applicable when simulating an electricity market; we have to be able to perform the simulation using computer software. As we learned in section 6.1 there is generally an infinite number of possible scenarios in an electricity market. For each scenario we can calculate a value of the result variables which are included in the model. Thus, if we

24. To be more precise, the theorem applies to sampling with replacement, i.e., there is nothing which prevents a particular sample from being observed more than once.

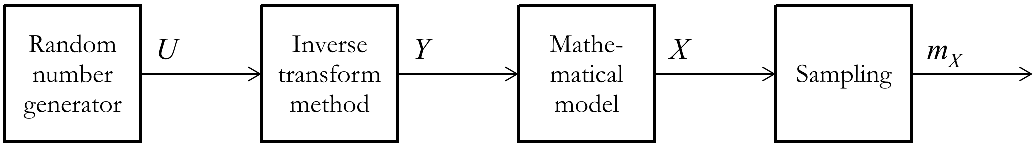


Figure 6.17 The principle of simple sampling.

randomly create a scenario, we can use the electricity market model to calculate a sample of each result variable. For simplicity, we collect the outcomes of all scenario parameters in a vector *Y*, while we in a similar manner collect the values of all result variables in a vector *X*, i.e.,

$$X = g(Y),$$

where *g* is the mathematical model of the electricity market (cf.section 6.1). Hence, simple sampling of an electricity market is equivalent to randomising *n* scenarios, $y_1, \ldots, y_n$, calculate the result variables for each scenario and form the mean of the observed outcomes of the result variables, i.e.,

$$m_X = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{n}\sum_{i=1}^{n} g(y_i). \tag{6.33}$$

To generate random scenarios we need a random number generator. A computer cannot generate truly random numbers, but by using clever functions it is possible to use a start value—a seed—to generate a sequence of pseudorandom numbers. This sequence is assuredly deterministic, but for an outsider it is impossible to predict the next number in the sequence by just observing the previous numbers. A good random number generates a sequence having approximately the same statistical properties as a sequence of *U*(0, 1)-distributed random number.[25] These random numbers can then be transformed to random number of the same probability distribution as the scenario parameters. It is recommended to use the inverse transform method (which is described in appendix E) for this transformation.

Assume that we have an electricity market simulation with *K* scenario parameters and that *n* scenarios are to be generated. The routine for applying simple sampling can then be summarised as follows (cf. figure 6.17):

Step 1. Generate a random number $u_{k,i}$ and transform it according to the probability distribution of the scenario parameter *k*, i.e., let $y_{k,i} = F_{Y_k}^{-1}(u_{k,i})$.

Step 2. Repeat step 1 for each of the *K* scenario parameters and combine the outcomes to a vector y_i. Calculate $x_i = g(y_i)$.

Step 3. Repeat step 1-2 for each of the *n* scenarios. Calculate

$$m_X = \frac{1}{n}\sum_{i=1}^{n} x_i.$$

Let us now study a small example of simple sampling of an electricity market:

Example 6.22 (simple sampling of a small power system). The South American village Pueblo is not connected to the national grid, but has a local grid of its own. The load in Pueblo is normally distributed with the mean 180 kW and standard deviation 40 kW. The system is supplied by three power plants: a hydro power

25. By *U*(0, 1)-distributed random number we mean random numbers which can get a value between 0 and 1, and where all outcomes are equally probable.

Table 6.8 Model of the power system operation in Pueblo.

D [kW]	*TOC* [¤/h]	Comment
0 - 150	0	The hydro power plant follows the load.
150 - 170	$2 \cdot (D - 150)$	The hydro power plant is operated at full capacity. The small diesel generator follows the load.
170 - 190	40	The hydro power plant is operated at full capacity. The large diesel generator set is operated at 40 kW. Surplus generation is consumed by the water heater.
190 - 250	$(D - 150)$	The hydro power plant is operated at full capacity. The large diesel generator set follows the load.
250 - 300	$100 + 2 \cdot (D\text{-}250)$	The hydro power plant is operated at full capacity. The large diesel generator set is operated at full capacity. The small diesel generator set follows the load.
> 300	200	All units are operated at full capacity. Some consumption is disconnected.[a]

a. As nothing else is stated in the assignment, we assume that no compensation is paid to those affected.

plant and two diesel generator sets. The hydro power plant is a run-of-the-river plant (i.e., there is no reservoir) and has an installed capacity of 150 kW. The least recorded water flow in the river where the hydro power plant is located, was large enough to run the plant at installed capacity. The operation cost is negligible. The two diesel generator sets are of 100 and 50 kW respectively. The larger generator has an operation cost of 1 ¤/kWh. However, the efficiency of this unit is very poor when it is only partly loaded, and it is therefore never operated at less than 40 kW. If it is necessary a water heater can be used to consume any surplus generation. The smaller diesel generator set has an operation cost of 2 ¤/kWh. All units have an availability of 100%. Determine the *ETOC* of the system and verify the result.

Solution: Both available capacity and operation cost of the power generators are fixed and can therefore be considered model constants.The only scenario parameter in this system is then the load, *D*. The objective of the simulation is to determine *ETOC*, which implies that we need to generate samples of the result variable *TOC*. The electricity market model *g* should therefore describe the relation between the load and the total operation cost. We assume that the system is operated so that the operation cost is minimised and that everything possible is done to avoid disconnection of consumers. This means that the system is operated as described in table 6.8. Using this strategy, we can draw the function $TOC = g(D)$ as in figure 6.18.

In this simple example of a Monte Carlo simulation we limit ourselves to creating 10 scenarios, which in this case corresponds to randomising 10 values of the load (refer to appendix E for a description of how random numbers of an arbitrary distribution can be created). Then we calculate the mean of the operation cost for these ten samples. Using appropriate software we obtain the ten scenarios listed in table 6.9. (The ten samples are also indicated in figure 6.18.) The estimate of the expectation value of the operation cost is therefore

$$m_{TOC} = \frac{1}{10}\sum_{i=1}^{10} TOC_i =$$

$$= (40.0 + 0 + 62.1 + 0 + 49.0 + 64.8 + 75.0 + 61.3 + 31.1 + 42.8)/10 \approx 42.61.$$

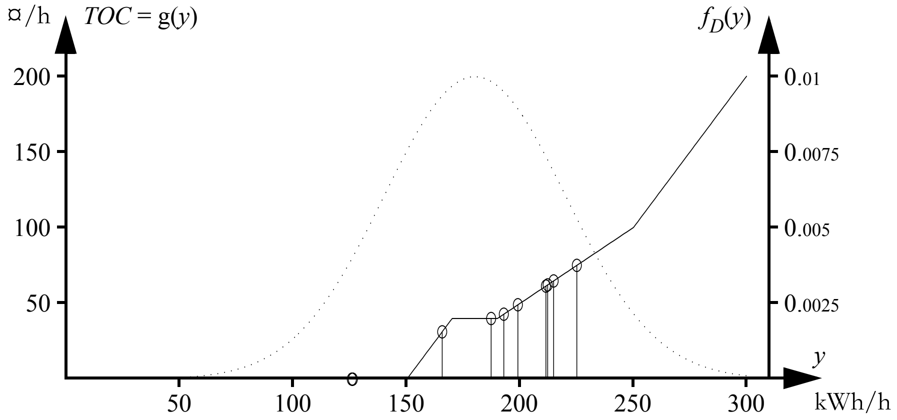


Figure 6.18 Cost function for the system in example 6.22. he figure shows cost function (solid line), the density function of the load (dotted line) as well as the ten chosen samples.

Table 6.9 Scenarios generated in example 6.22.

Scenario	*D* [kWh/h]	*TOC* [¤/h]
1	187.2	40.0
2	125.7	0
3	212.1	62.1
4	126.1	0
5	199.0	49.0
6	214.8	64.8
7	225.0	75.0
8	211.3	61.3
9	165.6	31.1
10	192.8	42.8

Hence, the estimated *ETOC* is equal to $ETOC \approx 42.61$ ¤/h.

In a small system, where the operation cost is well-defined function of one single scenario parameter, it is possible to calculate the exact expected operation cost using the definition of an expectation value:

$$ETOC = E[TOC] = \int_{-\infty}^{\infty} f_D(y)g(y)dy,$$

where $f_D(y)$ is the density function of the load, i.e., the density function of an $N(180, 40)$-distributed random variable (see appendix D), and $g(y)$ is our electricity market model, i.e., the relation between the result variable *TOC* and the scenario parameter *D*. The integral above can for example be solved numerically and the result is $ETOC \approx 39.66$ ¤/h. The result of the simple sampling is thus about 7.5% higher than the theoretical value.

6.3.2 Complementary Random Numbers

Complementary random numbers is a very simple variance reduction technique, which can be applied in all kinds of Monte Carlo simulations. The technique is based on modifying the random number generator so that it generates series of random values which are not independent. The consequence is that less samples are necessary to get a good spread of the samples.

Theory

Assume that m_{X1} and m_{X2} are two estimates of the expectation value, μ_X, of a certain random variable, i.e., $E[m_{X1}] = E[m_{X2}] = E[X] = \mu_X$. If we consider the average of m_{X1} and m_{X2} we find that the expectation value is

$$E\left[\frac{m_{X1} + m_{X2}}{2}\right] = \frac{1}{2}E[m_{X1}] + \frac{1}{2}E[m_{X2}] = \mu_X. \tag{6.34}$$

Hence, the average of m_{X1} and m_{X2} is also an estimate of μ_X, which maybe is not very surprising. However, the interesting is that the variance of the average is

$$Var\left[\frac{m_{X1} + m_{X2}}{2}\right] = \frac{1}{4}Var[m_{X1}] + \frac{1}{4}Var[m_{X2}] + \frac{1}{2}Cov[m_{X1}, m_{X2}]. \tag{6.35}$$

The variance of the average is, as can be seen, less than the variance of m_{X1} and m_{X2}respectively. Above all, it is possible to utilise that the covariance can become negative. A simple way of creating estimates with a negative covariance is to use complementary random numbers.

Random numbers of any probability distribution can be created by transformation of $U(0, 1)$-distributed random numbers (see appendix E). If U is $U(0, 1)$-distributed then the complementary random number of U is given by $U^* = 1 – U$. We can see that U^* also is a $U(0, 1)$--distributed random variable. If the outcome of U is high then U^* will receive a low value and vice versa, which obviously means that U and U^* are negatively correlated. If Y is a random variable from an arbitrary probability distribution and Y has been obtained by transforming U, and Y^* is obtained by applying the same transformation on U^*, then Y and Y^* will be negatively correlated.

> ***Example 6.23 (random complement of a normally distributed random variable):*** Randomise a sample from an $N(5, 2)$-distribution. Calculate the complementary random number of this sample.
>
> ***Solution:*** Normally distributed random variables can, as described in appendix E, be obtained by transformation of random number from a $U(0, 1)$-distribution. Assume that we start with $U = 0.15$. Using theorem E.2 we get $Y \approx –1.04$, which is an $N(1, 0)$-distributed random number. This value is transformed to the desired normal distribution using theorem D.3, which yields $5 + 2·(–1.04) \approx 2.93$.
>
> The random complement is obtained by performing the same transformations on $U^* = 1 – 0.15$, which corresponds to 7.07 rom the desired normal distribution. Considering that the normal distribution is symmetrical it is logical that Y^* is as much higher than the mean value as Y is smaller than the mean.

Application to Electricity Markets

Complementary random numbers are very straightforward to apply when simulating electricity markets. Assume that n scenarios y_i has been generated and that the complementary random number, y_i^* is calculated for each scenario. When the result variables are calculated both for the original scenarios and those based on complementary random numbers. Notice that as we in general will have more than one scenario parameter in an electricity market simulation, we will obtain more than one scenario based on complementary random numbers. Each scenario parameter will have two possible values: the original value and its complement. The original values and complements of K scenario parameters can be combined to 2^K scenarios.

Example 6.24 (complementary scenarios): Assume that a simple electricity market should be simulated. The market consists of a single power plant and a local load. The power plant has the installed capacity 100 MW and the availability is 95%. The load is normally distributed and the mean is 75 MW and the standard deviation is 10 MW. Randomise a scenario for this electricity market. Also state the complementary scenarios.

Solution: There are two scenario parameters in this electricity markets: the available generation capacity and the total load.

The available generation capacity can be determined by an $U(0, 1)$--distributed random variable, which is interpreted so that if $U_G \in U(0, 1) \leq 0.95$ then the power plant is available; otherwise it is not. Assume that we randomise $U_G = 0.2311$, which means that $U_G^* = 07689$. The power plant is available in both cases, i.e., $\overline{G} = \overline{G}^* = 100$ MW.

The load is randomised by transforming an $U(0, 1)$-distributed random variable (see appendix E). Assume that we randomise $U_D = 0.6068$, which yields $U_D^* = 0.3932$; then the transformation will produce the results $D \approx 77.7$ MW and $D^* \approx 72.3$ MW respectively.

The above random numbers and complementary random numbers can be combined to four scenarios according to table 6.10. Notice that the same scenario appears more than once, as $\overline{G} = \overline{G}^*$ in this case. This is completely normal and will not cause any error in the simulation.

Table 6.10 Complementary scenarios in example 6.24.

Available generation capacity	Total load
$\overline{G} = 100$ MW	$D = 77.7$ MW
$\overline{G} = 100$ MW	$D^* = 72.3$ MW
$\overline{G}^* = 100$ MW	$D = 77.7$ MW
$\overline{G}^* = 100$ MW	$D^* = 72.3$ MW

However, it should be noted that it is not necessary to apply complementary random numbers to all scenario parameters. Complementary random numbers will only result in a variance reduction if there is a negative correlation between the samples of the result variables. The method above results in a negative correlation between each y_i and y_i^*, but this is not necessarily reflected in the result variables, $x = g(y_i)$ and $x^* = g(y_i^*)$. If we for example consider a multi-area model, then the total load, $D_{tot} = \Sigma D_n$ will have much stronger correlation to *TOC* and *LOLO* than the load in a single area, D_n. It would therefore be unnecessary work to apply complementary random number on D_n. It is more efficient to randomise a value of the total load and then distribute this load among the areas. This can be done by randomising both a value of D_{tot} as well as a preliminary load in each area, D'_n. Then the preliminary load is scaled so that the load in the areas match the desired total load, while maintaining the relative distribution of the load among the areas:

$$D_n = \frac{D_{tot}}{\sum_{m \in \mathcal{N}} D'_m} D'_n. \tag{6.36}$$

Example 6.25. An electricity market can be divided into two areas. The load in the first area is $N(500, 40)$-distributed and the load in the second area is $N(450, 30)$-dis-

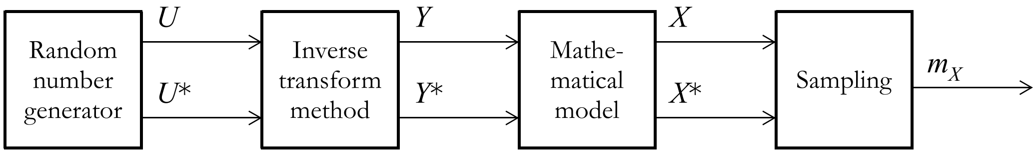


Figure 6.19 The principle of complementary random numbers.

tributed. The load level in the two areas are independent random variables. Generate two scenarios for this electricity market in such a way that the total load in the second scenario is equal to the complementary random number of the total load in the first scenario. Use the following five random numbers from an $N(0, 1)$-distribution: 1.65, –0.74, 0.27, –0.04 and 1.23.

Solution: We start by establishing the fact that the total load is the sum of two normally distributed random variables; hence, the total load is normally distributed and has a mean of 500 + 450 = 950 and the standard deviation $\sqrt{40^2 + 30^2} = 50$. We can now use theorem D.3 to transform the given random numbers to the appropriate normal distribution. We may start by transforming the first random number into a value of the total load, which results in $D_{tot} = 950 + 50 \cdot 1.65 = 1\,032.5$ MW. As the normal distribution is symmetrical, we then get the complementary random number $D_{tot}{}^* = 950 - 50 \cdot 1.65 = 867.5$ MW.

The next two random numbers are transformed into preliminary values of the load in the two areas for the first scenario: $D_1{}^{\mathrm{i}} = 500 + 40 \cdot (-0.74) = 470.4$ MW and $D_2{}^{\mathrm{i}} = 400 + 30 \cdot 0.27 = 408.1$ MW. Then these values are scaled to match the desired value of the total load in the first scenario:

$$D_1 = \frac{1\,032.5}{(470.4 + 408.1)} 470.4 \approx 552.9 \text{ MW},$$

$$D_2 = \frac{1\,032.5}{(470.4 + 408.1)} 408.1 \approx 479.6 \text{ MW}.$$

Finally, we use the last two random values to generate preliminary values of the load in the two areas for the second scenario: $D_1{}^{\mathrm{ii}} = 500 + 40 \cdot (-0.04) = 498.4$ MW and $D_2{}^{\mathrm{ii}} = 400 + 30 \cdot 1.23 = 436.9$ MW. This time the preliminary load is scaled so that the sum equals the complementary random number of the total load in the first scenario, which means that we get the following area loads:

$$D_1 = \frac{867.5}{(498.4 + 436.9)} 498.4 \approx 462.3 \text{ MW},$$

$$D_2 = \frac{867.5}{(498.4 + 436.9)} 436.9 \approx 405.2 \text{ MW}.$$

Assume that we have an electricity market simulation with K scenario parameters (out of which K_V scenario parameters are such that it makes sense to apply complementary random numbers) and that n scenarios are to be generated. The routine for applying complementary random numbers can then be summarised as follows (cf. figure 6.19):

Step 1. Generate a random number $u_{k,\,i}$ and transform it according to the probability distribution of the scenario parameter k, i.e., let $y_{k,\,i} = F_{Y_k}^{-1}(u_{k,\,i})$. If applicable to this scenario parameter, also transform the complementary random number i.e., let $y_{k,\,i}^* = F_{Y_k}^{-1}(1 - u_{k,\,i})$.

Step 2. Repeat step 1 for each of the K scenario parameters. Create 2^K complemen-

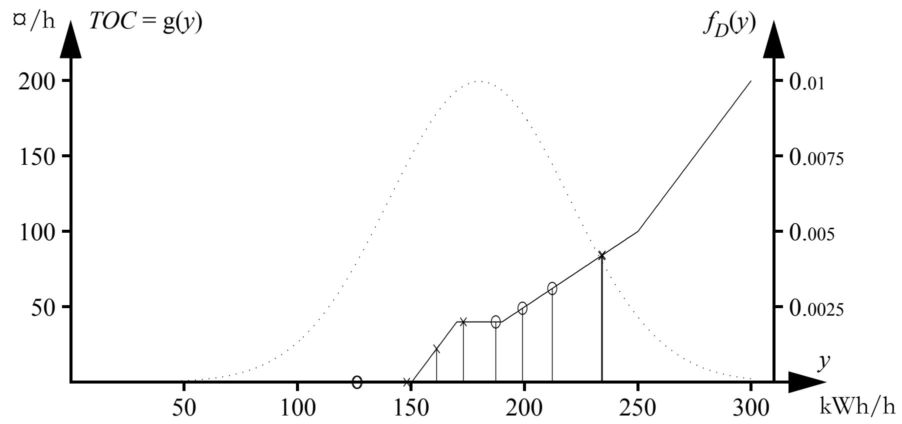


Figure 6.20 Cost function for the system in example 6.26. The figure shows the cost function (solid line), the density function of the load (dotted line) as well as the ten samples. The original samples are marked by circles and the complements are marked by crosses.

tary scenarios. Calculate $x_i = g(y_i)$ for all scenarios (original and complementary).

Step 3. Repeat step 1-2 for each of the *n* scenarios. Calculate

$$m_X = \frac{1}{n} \sum_{i=1}^{n} x_i.$$

The practical value of complementary random numbers is that we get a better spread of the samples. This is most value when the number of samples is small. If a lot of samples are taken then the spread will probably be rather good either way, which means that the impact of the complementary random numbers will not be so large.

A demonstration of the how complementary random numbers improve the results is given in the following example:

Example 6.26 (sampling using complementary random numbers): Consider the electricity market in example 6.22 again. Use complementary random numbers to improve the estimate of *ETOC*.

Solution: When this electricity market was simulated using simple sampling the resulting estimate of *ETOC* was about 7.5% higher than the true value. The explanation is indicated in figure 6.18. The samples are lop-sided, with a majority of high load scenarios, which causes the expected operation cost to be overestimated. One way to improve the estimate would of course be to take more samples than ten; hopefully this would result in a more even distribution.

A more simple way is to use complementary random numbers instead. If we take only five samples and then calculate the complement of these five, we get in total ten samples as in example 6.22. Table 6.11 shows the five first samples from example 6.22 and their complementary values (the samples are also indicated in figure 6.20). The mean of the ten marked samples results in the estimate $ETOC \approx 38.13$ ¤/h, which is about 3.8% less than the theoretical value, but yet a better estimate than was obtained with simple sampling.

Table 6.11 Scenarios generated in example 6.23.

Scenario	*D* [kWh/h]	*TOC* [¤/h]	*D** [kWh/h]	*TOC** [¤/h]
1	187.2	40.0	172.8	40.0
2	125.7	0	234.3	84.3
3	212.1	62.1	147.9	0
4	126.1	0	233.9	83.9
5	199.0	49.0	161.0	22.0

6.3.3 Control Variates

One of the most important reasons to use Monte Carlo methods is that they can manage models which are too complex to be treated analytically. However, using control variates, it becomes possible to use simplified analytical models to improve the result of the Monte Carlo-simulation.

Theory

Assume that X is a random variable with a certain expectation value μ_X, i.e., $E[X] = \mu_X$. Moreover, assume that there is another random variable, a *control variate* Z, the expectation value of which, $E[Z] = \mu_Z$, is known (through analytical calculations or from earlier investigations). Rather than estimating $E[X]$ we choose to estimate the expected difference between X and Z. An estimate of $E[X]$ is then calculated by

$$m_X = m_{(X-Z)} + \mu_Z, \tag{6.37}$$

because

$$E[m_{(X-Z)} + \mu_Z] = E[X-Z] + \mu_Z = E[X] - \mu_Z + \mu_Z = \mu_X. \tag{6.38}$$

The variance of the difference $X - Z$ is

$$Var[X-Z] = Var[X] + Var[Z] - 2Cov[X, Z]. \tag{6.39}$$

If we can find a control variate Z which has a strong positive correlation to X then it is possible that $2Cov[X, Z] > Var[Z]$; hence, $X - Z$ may have a lower variance than X. Simple sampling of $X - Z$ is then according to theorem 6.22 resulting in a smaller variance of the estimate than if X is sampled directly.

Application to Electricity Markets

The prerequisite of applying control variates is that there is an analytical model to compare with. For each scenario generated in the Monte Carlo simulation, we calculate the result variables of the detailed model, i.e.,

$$X = g(Y), \tag{6.40a}$$

as well as the analytical model, i.e.,

$$Z = \tilde{g}(Y). \tag{6.40b}$$

A suitable analytical model to be used in electricity market simulations is a single-area model, where the load is not price sensitive and the electricity market is in every scenario assumed to minimise the total operation cost while maintaining balance between generation and consumption.

Such a model can be defined from the solution to the following, simple LP problem:

$$\text{minimise} \quad \sum_{g \in \mathcal{G}} \beta_g \tilde{G}_g + \beta_U \tilde{U} \tag{6.41}$$

$$\text{subject to} \quad \sum_{g \in \mathcal{G}} \tilde{G}_g + \tilde{U} = D_{tot}, \tag{6.41a}$$

$$0 \le \tilde{G}_g \le \bar{G}_g, \qquad \forall g \in \mathcal{G}, \tag{6.41b}$$

$$0 \le \tilde{U}. \tag{6.41c}$$

In the optimisation problem (6.41) we have the following variables:

$\tilde{G}_g$ = generation in power plant g,
$\tilde{U}$ = unserved power.

The parameters are

D_{tot} = total load,
$\bar{G}_g$ = available generation capacity in power plant g,
β_g = generation cost in power plant g,
β_U = penalty cost for unserved power.[26]

From the solution to (6.41) we can define the control variates

$$\tilde{TOC} = \sum_{g \in \mathcal{G}} \beta_g \tilde{G}_g, \tag{6.42a}$$

$$\tilde{LOLO} = \begin{cases} 1 & \text{if } \tilde{U} > 0, \\ 0 & \text{otherwise}. \end{cases} \tag{6.42b}$$

It should be noted that the optimisation problem (6.41) is very straightforward to solve by hand calculations. The power plants are sorted in a merit order according to ascending cost. Then the generation in the least expensive power plant is increased until it either covers the load or reaches its available capacity. In the latter case we proceed and try with the next power plant, etc. As a last resort we use load shedding to balance the constraint (6.41a).

Since the model in (6.41) corresponds to the model used in probabilistic production cost simulation; thus, the expectation value of the control variates can be calculated using the methods described in section 6.2.

Assume that we have an electricity market simulation with K scenario parameters and that n scenarios are to be generated. The routine for applying control variates can then be summarised as follows (cf. figure 6.21):

Step 1. Generate a random number $u_{k,i}$ and transform it according to the probability distribution of the scenario parameter k, i.e., let $y_{k,i} = F_{Y_k}^{-1}(u_{k,i})$.

Step 2. Repeat step 1 for each of the K scenario parameters and combine the outcomes to a vector y_i. Calculate the result variables $x_i = g(y_i)$ and the control variates $z_i = \tilde{g}(y_i)$.

Step 3. Repeat step 1-2 for each of the n scenarios. Calculate

26. As in the multi-area problem from section 6.1.1, we are here forced to introduce a possibly fictitious cost for load shedding in order to prevent the load balance to be maintained by disconnecting all load, even though there is unused generation capacity available.

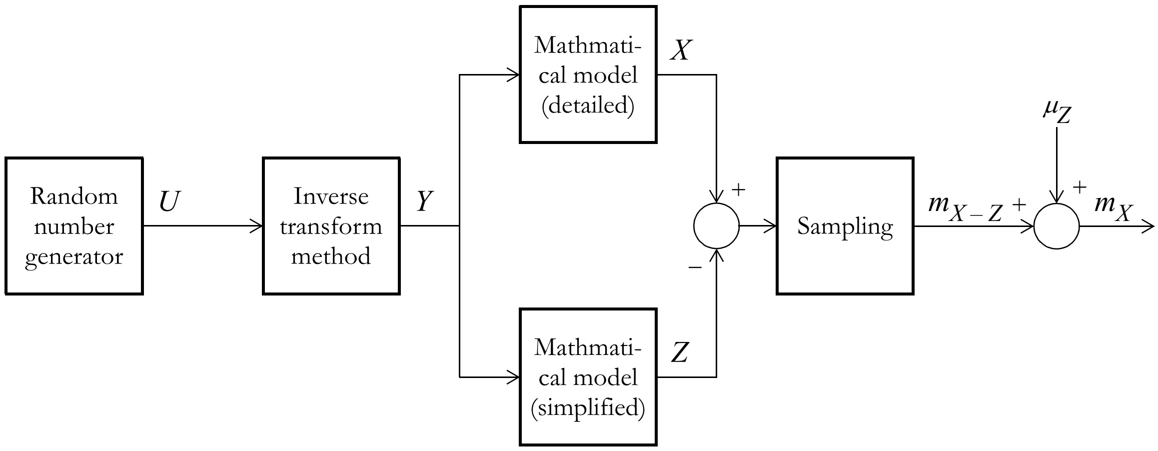


Figure 6.21 The principle of control variates.

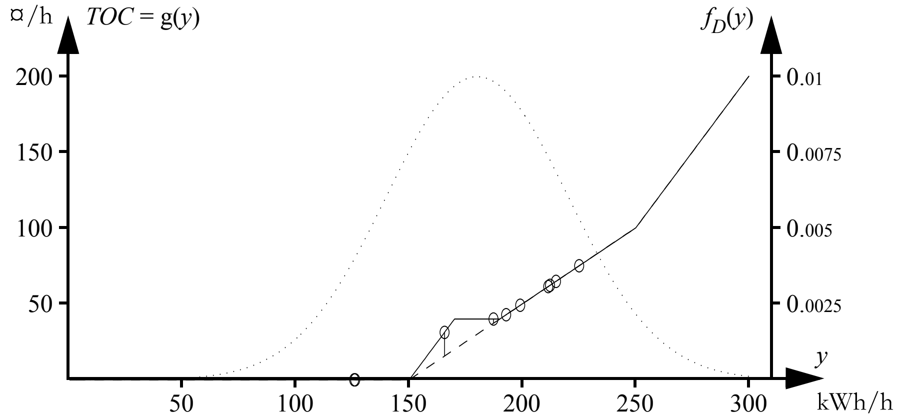


Figure 6.22 Cost function for the system in example 6.27. The figure shows cost function for the detailed model (solid line), cost function for a PPC model (dashed line)—notice that these lines coincide except for a small interval. The figure also shows the ten samples, which constitute the difference between the solid and the dashed lines. Finally, the density function of the load is shown (dotted line).

$$m_{X-Z} = \frac{1}{n}\sum_{i=1}^{n}(\boldsymbol{x}_i - \boldsymbol{z}_i).$$

Step 4. Calculate $\mu_Z = E[Z]$.

Step 5. Calculate $m_X = m_{X-Z} + \mu_Z$.

The principle of applying a control variate is illustrated in the following example:

Example 6.27 (sampling using a control variate). Consider the electricity market from example 6.22 once again. Use a control variate to improve the estimate of *ETOC*.

Solution: The same ten samples were used as in example 6.22. However, in this case the value of samples is equal to the difference between the detailed model, $g(D)$, and the simplified model, $\tilde{g}(D)$. Let us denote this difference by *TOCD*, i.e., *TOCD* = *TOC*

Table 6.12 Scenarios generated in example 6.27.

Scenario	D [kWh/h]	TOC [¤/h]	$\widetilde{TOC}$ [¤/h]	$TOCD$ [¤/h]
1	187.2	40.0	37.2	2.8
2	125.7	0	0	0
3	212.1	62.1	62.1	0
4	126.1	0	0	0
5	199.0	49.0	49.0	0
6	214.8	64.8	64.8	0
7	225.0	75.0	75.0	0
8	211.3	61.3	61.3	0
9	165.6	31.1	15.6	15.6
10	192.8	42.8	42.8	0

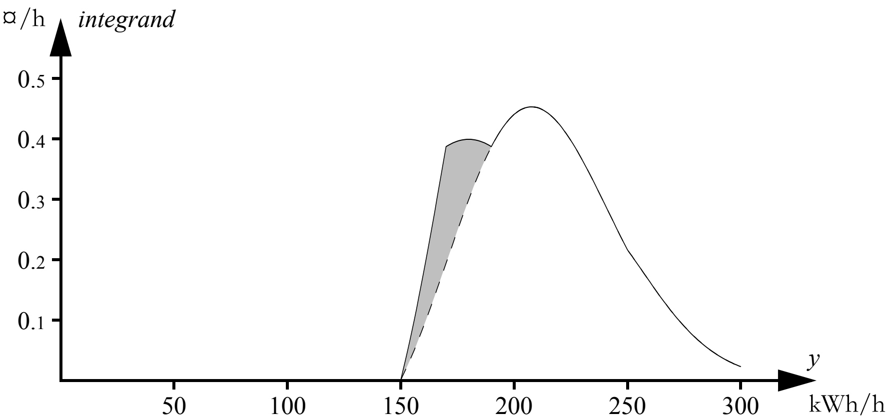


Figure 6.23 Illustration of the integrand for calculation of the expectation value with and without control variate The solid line shows the integrand $f_D(y)g(y)$ and the dashed line shows $f_D(y)\tilde{g}(y)$. The difference between them corresponds to the shaded area.

$-\ \widetilde{TOC}$. As we can see in table 6.12, the two models only differ in two cases. The mean of the ten samples is 1.84 ¤/h, which means that we obtain the estimate $m_{TOCD} = 1.84$ ¤/h. By performing a probabilistic production cost simulation of this system we get $\mu_{\widetilde{TOC}} \approx 36.27$ ¤/h. The estimate using a control variate is thus

$$m_{TOC} = m_{TOCD} + \mu_{\widetilde{TOC}} = 36.27 + 1.84 \approx 38.11 \text{ ¤/h}.$$

Hence, the control resulted in an improvement of the estimate which is about the same as when we were using complementary random numbers in example 6.26. To understand why the estimate is improved, we can study how the expectation values are calculated with and without control variate:

$$ETOC = E[TOC] = \int_{-\infty}^{\infty} f_D(y)g(y)dy,$$

$$ETOC = E[TOCD] + \mu_{\widetilde{TOC}} = \int_{-\infty}^{\infty} f_D(y)(g(y) - \tilde{g}(y))dy + \mu_{\widetilde{TOC}}.$$

The integrands of these two expressions are shown in figure 6.23. *ETOC* corresponds to the area below the solid line, i.e., both the shaded and the white area, whereas the expectation value of *TOCD* corresponds to just the shaded area. When using the control variate, we try to estimate the shaded area, while the white area is given using probabilistic production cost simulation. An error of 10% in the Monte Carlo estimate will therefore affect the whole result when *ETOC* is estimated directly, but only a small part of the result (the shaded area) when a control variate is used.

6.3.4 Stratified Sampling

The idea of stratified sampling is to divide the samples in different subpopulations which are referred to as *strata*,[27] and which are investigated separately. Some are practical (it may for example be so that different parts of a study must be performed using different methods), but there are also possibilities to gain efficiency. If it is possible to collect samples having similar properties in the same stratum then the variance within each stratum will be reduced. Since reduced variance means increased efficiency according to theorem 6.22, the result is that the expectation value of the samples within the stratum can be determined faster and/or more accurately. These expectation values can then simply be weighted together to obtain an expectation value of the entire population.

Theory

Assume that we have a random variable X with the sample space [28] $\mathcal{X}$. In stratified sampling the sample space is divided in L strata, where stratum h comprises the outcomes $\mathcal{X}_h$, which is a subset of $\mathcal{X}$. Strata may not overlap, i.e., each outcome must belong to exactly one stratum:

$$\bigcup_h \mathcal{X}_h = \varnothing, \tag{6.43a}$$

$$\mathcal{X}_h \cap \mathcal{X}_j = \varnothing \ \forall \ h \neq j. \tag{6.43b}$$

Each stratum is assigned a weight corresponding to how large part of the population which belongs to the stratum:

$$\omega_h = \frac{N_h}{N} = P(X \in \mathcal{X}_h), \tag{6.44}$$

where

ω_h = stratum weight of stratum h,
N_h = number of units (i.e., possible outcomes) in stratum h,
N = number of units in the entire population.

As can be seen above, the stratum weight can also be considered as the probability that a random observation of X will fall into stratum h.

Is is now possible to consider L separate random variables X_h, $h = 1, \ldots, L$, each with the sample space $\mathcal{X}_h$. The expectation value of each stratum is calculated separately. In some cases it might be possible to analytically determine $E[X_h]$; if not, we can estimate it by simple sampling:

27. This footnote has just been added to maintain the footnote numbering in pace with the Swedish edition.
28. The sample space is the set of possible outcomes of X.

$$m_{X_h} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{h,i}, \tag{6.45}$$

where

m_{X_h} = estimate of the expectation value of stratum h,
$x_{h,i}$ = value of the i:th sample from stratum h,
n_h = number of samples from stratum h.

The expectation value of the whole population, $E[X]$, can then be estimated by

$$m_X = \frac{\sum_{h=1}^{L} N_h m_{X_h}}{N} = \sum_{h=1}^{L} \omega_h m_{X_h}. \tag{6.46}$$

If it was possible to calculate $E[X_h] = \mu_{X_h}$ analytically then we would of course use $m_{X_h} = \mu_{X_h}$ instead of an estimate calculated according to (6.45).

In stratified sampling, the variance of the estimate m_X is

$$Var[m_X] = \sum_{h=1}^{L} \omega_h^2 \frac{Var[X_h]}{n_h}. \tag{6.47}$$

If strata have been chosen properly then the variance according to (6.47) can be less than the variance of simple sampling (cf. theorem 6.22). It must though be observed that the opposite is also possible; poorly chosen strata can result in a higher variance in the estimate of the expectation value $E[X]$!

If $E[X]$ is to be estimated using in total n samples, how should then these samples be divided between the different strata? It can be shown that (6.47) is minimised if the so-called Neyman allocation is used:

$$n_h = n \frac{\omega_h \sigma_{X_h}}{\sum_{k=1}^{L} \omega_k \sigma_{X_k}}, \tag{6.48}$$

where

σ_{X_h} = standard deviation of stratum h, i.e., $\sqrt{Var[X_h]}$.

If $Var[X_h]$ can be calculated analytically, it would also be possible to calculate $E[X_h]$, which means that there is no need at all for Monte Carlo simulation. Hence, we must assume that the σ_{X_h} are unknown. When using (6.48) we will have to use estimates instead:

$$s_{X_h} = \sqrt{\frac{1}{n_h} \sum_{i=1}^{n_h} (x_{h,i} - m_{X_h})^2}, \tag{6.49}$$

where

s_{X_h} = estimate of σ_{X_h}.

It should be noted that the Neyman allocation usually corresponds to a rather flat optimum, i.e., not much is lost if we deviate a little bit from the optimal allocation.

To calculate the coefficient of variation (see section 6.3.1) we need to estimate $Var[X]$. This is done by weighting together the estimates s_{X_h}:

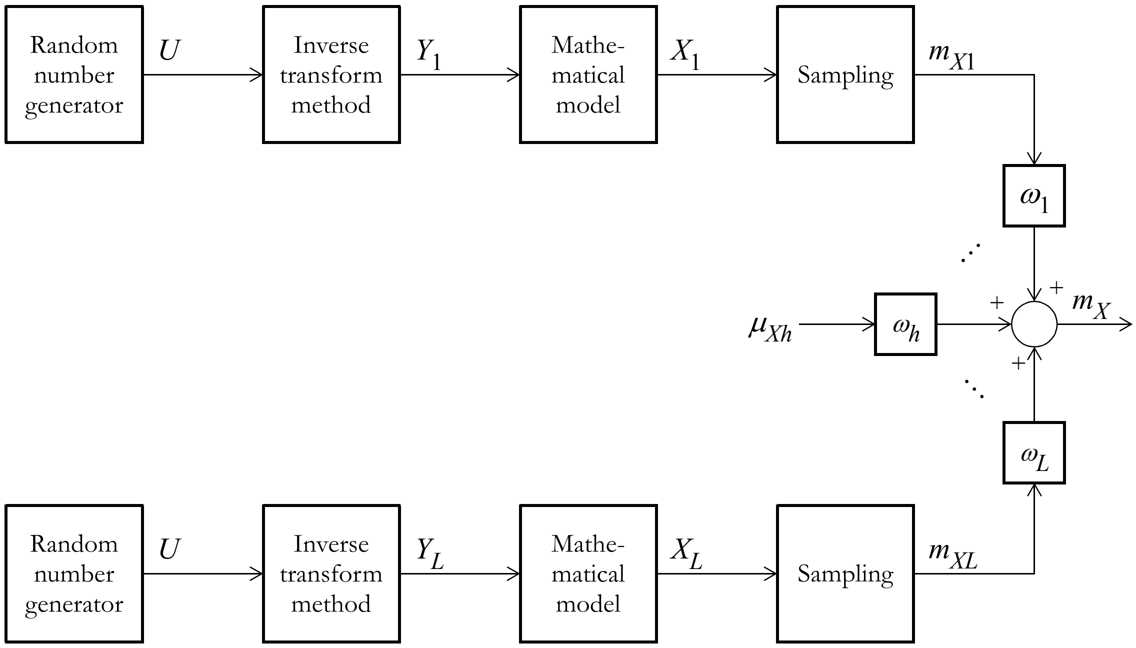


Figure 6.24 The principle of stratified sampling. The expectation value of each stratum is estimated by simple sampling (as in the top and bottom strata in the figure) or is known from analytical calculations (as in the middle stratum).

$$s_X^2 = \sum_{h=1}^{L} \omega_h^2 s_{X_h}^2. \tag{6.50}$$

Application to Electricity Markets

Before stratified sampling can be used in a simulation of an electricity market the strata have to be defined. As mentioned in the theory section above, an inappropriate stratification may result in a less accurate result than when using simple sampling; hence, strata need to be chosen carefully. We will return to this issue in section 6.3.5; for now we restrict ourselves to discussing how stratified sampling is applied given a specific stratification.

Assume that we have an electricity market simulation with K scenario parameters and L strata, and that n_h scenarios are to be generated from each stratum. The routine for applying stratified sampling can then be summarised as follows (cf. figure 6.24):

Step 1. Consider stratum h. If possible, calculate $E[X_h] = \mu_{Xh}$ using analytical methods, set $m_{Xh} = \mu_{Xh}$ and proceed to step 5.

Step 2. Generate a random number $u_{k,h,i}$ and transform it according to the probability distribution in stratum h of the scenario parameter k, i.e., let $y_{k,h,i} = F^{-1}_{Y_{k,h}}(u_{k,h,i})$.

Step 3. Repeat step 2 for each of the K scenario parameters and combine the outcomes to a vector $y_{h,i}$. Calculate $x_{h,i} = g(y_{h,i})$.

Step 4. Repeat step 2-3 for each of the n_h scenarios. Calculate

$$m_{Xh} = \frac{1}{n_h} \sum_{i=1}^{n_h} x_{h,i}.$$

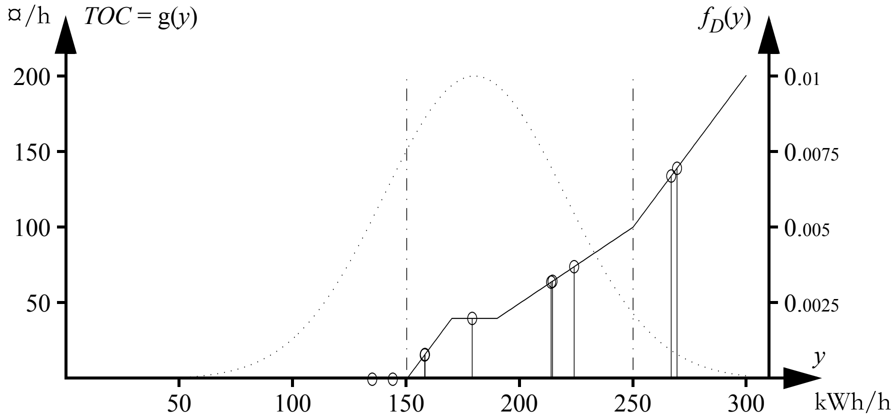


Figure 6.25 Cost function for the system in example 6.28. The figure shows the cost function (solid line) as well as the density function of the load (dotted line) and the ten chosen samples.The stratum boundaries are indicated by dash-dotted lines.

Step 5. Repeat step 1-4 for each of the L strata.

Step 6. Calculate $m_X = \sum_{h=1}^{L} \omega_h m_{X_h}$.

The following example demonstrates how stratified sampling is used in a simulation of an electricity market:

Example 6.28 (stratified sampling). Again we will study the electricity market from example 6.22. Use stratified sampling to improve the estimate of *ETOC*.

Solution: It is important to define strata and choose number of samples per stratum in a wise manner in order to get good results from stratified sampling. For the moment we will however make these decision based on engineer's intuition. Therefore, let stratum 1 be all scenarios where the load is less then or equal to 150 kW, stratum 2 is the scenarios where the load is higher than 150 kW but less than 250 kW and finally stratum 3 consists of all other load levels. Then we generate two scenarios each from stratum 1 and 3, while the remaining six samples are allocated to stratum 2. The chosen scenarios and the stratum boundaries are shown in table 6.13 and are also indicated in figure 6.25. The expectation value of each stratum is

$$ETOC_1 = 0,$$

$$ETOC_2 \approx 45{,}73,$$

$$ETOC_3 \approx 136{,}23.$$

In order to estimate *ETOC* we also need the stratum weights. According to (6.44) we get

$$\omega_1 = P(D \le 150) = \Phi((150 - 180)/40) \approx 0.23,$$

$$\omega_2 = P(150 < D \le 250) = \Phi((250 - 180)/40) - \Phi((150 - 180)/40) \approx 0.73,$$

$$\omega_3 = P(D > 250) = 1 - \Phi((250 - 180)/40) \approx 0.04.$$

The estimate of the whole population is then according to (6.46) $ETOC \approx 38.99$ ¤/h or about 1.7% less than the true value. Thus, also stratified sampling provided a defi-

Table 6.13 Scenarios generated in example 6.28.

Stratum	Scenario	D [kWh/h]	TOC [¤/h]
1	1	134.8	0
	2	143.9	0
2	1	158.0	15.9
	2	214.4	64.4
	3	158.1	16.2
	4	224.0	74.0
	5	213.8	63.8
	6	179.0	40.0
3	1	269.4	138.8
	2	266.8	133.6

nitely better estimate of the operation cost. There are two reasons why stratified sampling improves the result. One is that it is possible to define homogenous strata, ie.e., strata where the individual deviations between the units which belong to the strata are small; thereby, few samples are required to obtain a good estimate of the expectation value for that stratum. In this case TOC is equal to zero for all scenarios in stratum 1. The second explanation it is possible to use few sample in less important strata, which result in less accurate estimates for those strata, but that will not affect the final result very much. In the example above the estimate of $ETOC_3$ is rather poor (the true value is 131.54 ¤/h, i.e., the estimate above is 3.5% too high), but this error does not have a large impact on the final result, because stratum 3 has much lower stratum weight than the other two strata.

6.3.5 Monte Carlo Simulation of Electricity Markets

To complete our presentation concerning Monte Carlo simulation of electricity markets, we will address some further aspects. More precisely, we will discuss how several variance reduction techniques can be combined and we will provide some more details about how stratified sampling should be applied.

Combining Several Variance Reduction Techniques

It is perfectly possible to combine the three variance reduction techniques described in the previous sections. The procedure is demonstrated in the following example:

Example 6.29 (simulation using a combination of three variance reduction techniques). Consider the electricity market in example 6.22 again. What is the estimate of $ETOC$ when all three variance reduction techniques are applied simultaneously?

Solution: We use the same stratification as in example 6.28. From strata 1 and 3 three we create one original scenario and one complementary scenario each. In stratum 3 we create three original scenarios and thereby obtain three complementary scenarios. The selected samples are shown in table 6.14 and are also indicated in figure 6.26. The following estimates are obtained in each stratum:

$ETOCD_1 = 0,$

Table 6.14 Scenarios generated in example 6.29.

Stratum	Scenario	D [kWh/h]	TOC [¤/h]	$\widetilde{TOC}$ [¤/h]	$TOCD$ [¤/h]	D^* [kWh/h]	$\widetilde{TOC}^*$ [¤/h]	TOC_{SPS}^* [¤/h]	$TOCD^*$ [¤/h]
1	1	134.8	0	0	0	128.1	0	0	0
2	1	158.0	15.9	8.0	8.0	230.3	80.3	80.3	0
	2	214.4	64.4	64.4	0	167.9	35.8	17.9	17.9
	3	158.1	16.2	8.1	8.1	230.1	80.1	80.1	0
3	1	269.4	138.8	138.8	0	256.9	113.7	113.7	0

$$ETOCD_2 \approx 5.66,$$

$$ETOCD_3 = 0.$$

These results are weighted together, which results in the estimate $ETOCD \approx 4.15$. The final estimate of the operation cost is thus

$$ETOC = ETOC_{PPC} + ETOCD = 36.27 + 4.15 \approx 40.42 \text{ ¤/h},$$

which is about 1.9% too high compared to the theoretical value.

If we compare the result in example 6.29 to the earlier examples, we find that the estimate is somewhat less accurate when using a combination compared to just using stratified sampling! From this, we should however not conclude that a combination of the three variance reduction techniques is less efficient than just using simple sampling. Remember that Monte Carlo methods never provide exact answers; hence, we can never state that a certain method always is better than another. What we can claim is that a certain method (for example combining several variance reduction techniques) has a higher probability of providing accurate results than another method. To learn more about the precision of the different methods, we can perform the same simulation 1 000 times, but with different seeds to the random number generator. The result of such an experiment is shown in table 6.15. In the table we can see that an estimate based on simple sampling in the worst case can provide very misleading results, whereas the results using different var-

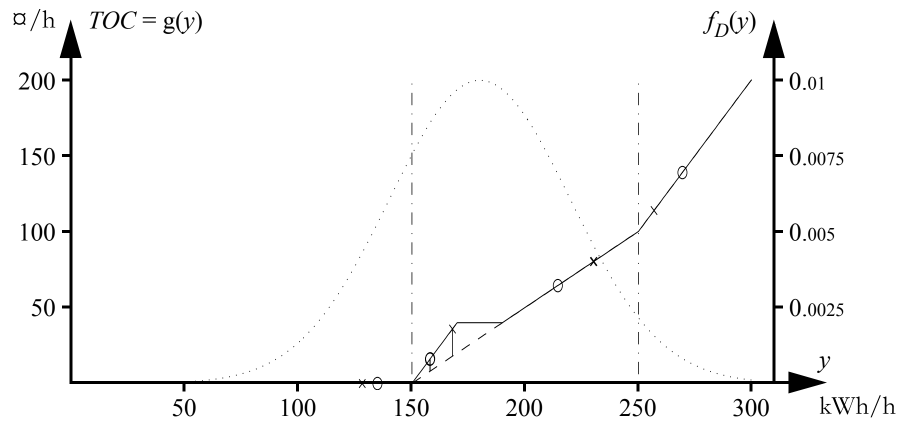


Figure 6.26 Cost function for the system in example 6.29. The figure shows the cost function of the detailed model (solid line), cost function for a PPC model (dashed line)—notice that these lines coincide except for a small interval. The figure also shows the disunity function of the load (dotted line) and the ten samples. The original samples are marked by circles and the complementary random numbers by crosses. The stratum borders are indicated by dash-dotted lines.

Table 6.15 Comparison of different variance reduction techniques.

Simulation method	Lowest estimate of *ETOC*	Average estimate of *ETOC*	Highest estimate of *ETOC*
Simple sampling	6.33	39.78	81.65
Complementary random numbers	31.48	39.85	64.22
Control variate	36.27	40.03	46.08
Stratified sampling	19.44	39.78	59.74
Combination	36.95	40.03	43.30

iance reduction techniques are more accurate.

However, the comparison above shows rather poor results for stratified sampling. This is explained by the fact that it is a little bit more difficult to obtain good results using stratified sampling compared to the other methods. The challenge is both to define appropriate strata and to decide how many scenarios should be generated in each of the strata. These issues were not properly addressed in example 6.28. Below we will describe some basic methods to manage these issues in an efficient manner.

Strata Trees

Defining a stratum means that we state which values of the scenario parameters which characterise the scenarios of the stratum. The scenarios of a stratum should preferably produce more or less the same results in order to make the stratification as efficient as possible. Let us consider a system there the available generation capacity in power plants with negligible operation costs (for example wind power) is $\overline{W}$ and the available generation capacity in the remaining power plants is $\overline{G}$. However, in some cases the complete generation capacity can not be utilised due to transmission congestion. We use the symbol $\overline{U}_W$ for the maximal unused capacity with negligible operation cost and $\overline{U}_{WG}$ for the maximal unused capacity in all power plants. Moreover, assume that we know how large the losses may become and that we use the symbol $\overline{L}$. In a real system it is practically impossible to exactly calculate $\overline{U}_W$, $\overline{U}_{WG}$, and $\overline{L}$, but in this compendium we will limit ourselves to system that are so simple that we avoid this kind of trouble.

Given the parameters above we can compare the available generation resources and the total load, D_{tot}. The result is that we can distinguish seven types of scenarios,[29] which are listed in table 6.16. In the first type of scenarios we always have enough generation capacity in power plants with negligible operation costs, which means that the operation cost, *TOC*, must be zero in all these scenarios, and no load shedding will be required, i.e., *LOLO* must also be zero. In scenarios of type II and III the generation capacity is still sufficient to cover the load, but some more costly power plants may have to be used. Thus, *TOC* will either be zero or larger than zero.[30] In scenarios of type IV we know that the load is larger than the generation capacity with negligible operation cost; hence, *TOC* must be larger than zero. The generation capacity is still large enough to prevent load shedding, so *LOLO* is equal to zero in these scenarios too. However, in scenarios of type V or VI transmission congestion or losses may cause the generation capacity to be insufficient, which means that *LOLO* will be zero in some cases, while in some—hopefully rare—scenarios load shedding is necessary and *LOLO* will equal unity. Finally, we have scenarios of type VII, in which the

29. Strictly speaking there are actually more than seven types of scenarios, as the seven main types may overlap in some cases (for example, if $\overline{G}$ = 10 MW while $\overline{L}$ = 20 MW), but to simplify this section we neglect such special cases.

30. The reason why we differentiated between scenarios where the operation cost is larger to zero due to congestion and losses respectively is that the probability of these events may differ significantly.

load is larger than the total generation capacity and hence load shedding is unavoidable.

To define strata based on the analysis above, we may use a so-called *strata tree*. A strata tree is a tree structure with the following properties:

- Each node specifies a subset of the sample space for one or more scenario parameters. The only exception is the root, which holds no information.
- Each node has a certain *node weight*. The node weight equals the probability that the scenario parameter belongs to the specified subset. The node weight of the root is always 1.
- The scenario parameters along a branch of the tree should be independent of each other.
- Each branch of the tree should specify a part of the sample space for each of the scenario parameters in the model. This means that each branch fulfils the requirements of defining a stratum; thus, each branch corresponds to a stratum. The stratum weight is calculated by multiplying all node weights along the branch.
- All nodes do not have to specify subsets for scenario parameters. It is possible to introduce various help variables, as long as it is possible to determine a proper node weight.

Since a stratification requires that all scenarios belong to a stratum and that there is no overlapping (6.43a, 6.43b). To guarantee that the strata tree really covers all scenarios it is possible to use the following two simple rules:

- The children of a certain node must define subsets for the same scenario parameter.
- The sum of the node weights of the children must always be exactly 1.

The idea of the strata tree is that we in one level of the tree put all possible states of the available generation capacity in separate nodes. Each of these nodes may have up to seven child nodes, which correspond to different levels of the load. By that means we can create strata, where all scenarios are of the same type. The principle is illustrated in the following example, which also illustrates how help variables can be used:

> ***Example 6.30 (strata tree):*** Consider the power system of MECC in example 6.1. Assume that the diesel generator set has an availability of 80%. The available generation capacity in the wind power plant can in reality vary between 0 and 200 kW, but let us for the sake of simplicity assume a model where there is a 50% probability that the wind power plant cannot generate anything at all and another 50% probability that the maximal generation of the wind power plant is 150 kW.

Table 6.16 Classification of scenarios.

Type	Load levels	*TOC*	*LOLO*
I	$D_{tot} \le \overline{W} - \overline{U}_W$	0	0
II	$\overline{W} - \overline{U}_W < D_{tot} \le \overline{W} - \overline{L}$	≥ 0*	0
III	$\overline{W} - \overline{L} < D_{tot} \le \overline{W}$	≥ 0**	0
IV	$\overline{W} < D_{tot} \le \overline{W} + \overline{G} - \overline{U}_{WG}$	> 0	0
V	$\overline{W} + \overline{G} - \overline{U}_{WG} < D_{tot} \le \overline{W} + \overline{G} - \overline{L}$	> 0	0 or 1*
VI	$\overline{W} + \overline{G} - \overline{L} < D_{tot} \le \overline{W} + \overline{G}$	> 0	0 or 1**
VII	$\overline{W} + \overline{G} < D_{tot}$	> 0	1

* Depending on impact of transmission limitations.
** Depending on impact of transmission losses.

Lighting constitutes a large share of the load in the MECC system; hence, the power demand has a peak in the evening between 6 p.m. and 12 midnight. The load is then $N(175, 48)$-distributed in Mji and $N(75, 20)$-distributed in Kijiji. During the rest of the day, the load in Kijiji is considerably lower—$N(30, 7)$-distributed—whereas there in Mji also is some industrial activities resulting in a relatively high load during off-peak hours—the load is then $N(120, 24)$-distributed.

a) Disregards of the transmission losses on the line between Mji and Kijiji. Suggest an appropriate strata tree for the system, and calculate the stratum weights.

b) How will the strata tree change if the losses on the transmission line are included, if it can be assumed that the maximal losses are 3 kW?

Solution: a) It would have been sufficient to create a strata tree with two levels below the root, if the probability distribution of both the available generation capacity and the load had been the same regardless of the time of the day. To manage the load in a simple way, we can however introduce another level representing the time of the day. This is actually not a scenario parameter (i.e., time is not included in the multi-area model), but is an example of a help variable.

We choose to place the time on the level just below the root, and on the next level we put the available generation capacity in the wind power plant ($\overline{W}$) and the diesel generator set ($\overline{G}$) respectively. At the bottom level of the tree we have the load (D).[31]

As there are no transmission limitations and we disregard the transmission losses, we have $\overline{U}_W = 0$, $\overline{U}_{WG} = 0$ and $\overline{L} = 0$. We will therefore not find any scenarios of type II, III, V or VI in this strata tree. However, we will find a special case reminding of type VII scenarios, but there $TOC = 0$, because $\overline{G} = 0$.

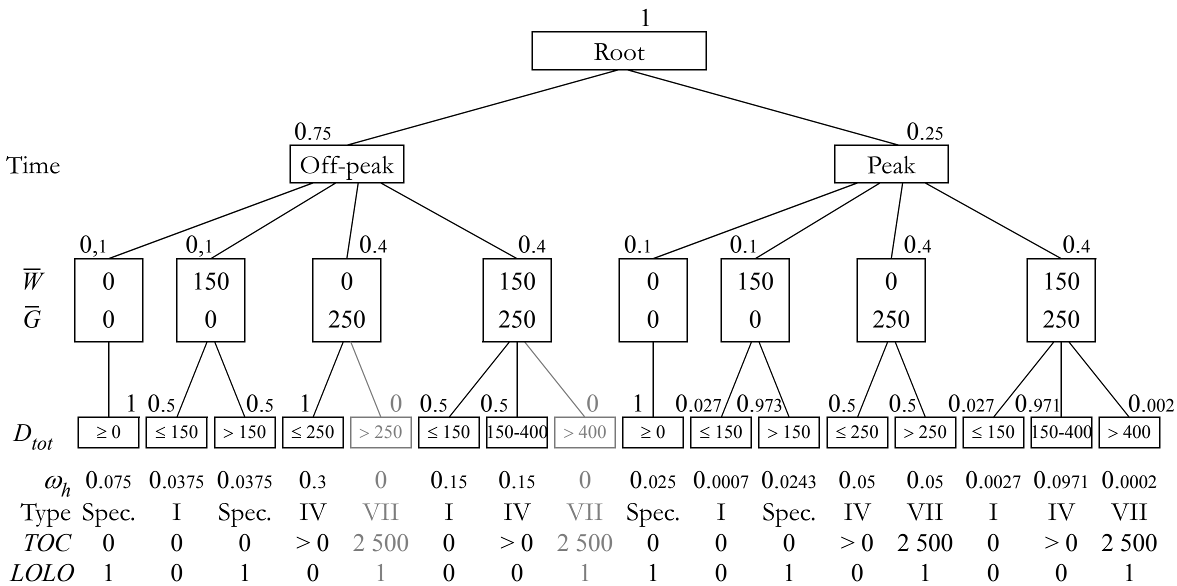


Figure 6.27 Strata tree for the system in example 6.30. All states of $\overline{W}$, $\overline{G}$ and D_{tot} are stated in kW.

31. An alternative approach would have been to place the time below the available generation capacity. The essential is that the time is above the load level of the tree, as the time determines which probability distribution should be used for calculation of the node weights on the load level.

We start by investigating the system during off-peak hours. As 75% of the time counts as off-peak, the node weight of the "off-peak" node is 0.75. There are now four possible states of the available generation capacity; either the wind power plant is capable of generating 0 or 150 kW, while at the same time the diesel generator set may either be available or unavailable. The probability of each these four states is easily calculated, which gives us the node weights for the four first nodes on the generation capacity level of the strata tree (see figure 6.27).

In the first of these four nodes, there is no generation capacity available and load shedding will therefore be necessary regardless of the current load. Thus, all these scenarios have the same properties and there is no need for a further separation.

In the next node we have 150 kW wind power available. In this case we need to child nodes: one for scenarios where the load does not exceed 150 kW and one for the scenarios there the load is larger than 150 kW. According to theorem D.4 the total load in the system is $N(150, 25)$-distributed, which means that the probability that the load is less than 150 kW is exactly 50%. The node weights of these two load nodes is therefore 0.5 each.

In the third node on the generation capacity level, only the diesel generator set is available; therefore it is desirable to differentiate between scenarios where the total load is lesser than or greater than 250 kW. However, when calculating the node weights, we find that the probability that the total load does not exceed 250 kW is 99.997%. Consequently, the node weight of the other node (where $D_{tot} > 250$) is only 0.00003, which may cause numerical problems when generating random numbers. We may therefore choose to exclude this node from the strata tree.

Then we have a node where both the wind power plant and the diesel generator set can be used for electricity generation. Here we need three load nodes to distinguish between scenarios with different properties. A load up to 150 kW can be covered using the wind power plant only, whereas both units are needed for a load between 150 and 400 kW, and if the load exceeds 400 kW then load shedding is necessary. The node weight of the third load node is however in practice zero, and therefore we exclude this node. The two remaining load nodes will then each have the node weight 0.5.

If we study the system during peak hours, we will find the same states of the available generation capacity. In the generation level, the right part of the strata tree is therefore identical to the left one. On the load level, it is the same intervals that are interesting, but as the total load during the evening is $N(250, 52)$-distributed, we will have other node weights (see figure 6.27).

Finally, the stratum weights are calculated by multiplying the node weights along each branch of the tree. The results are shown in figure 6.27.

b) When considering the losses it is not always possible to predict the properties of a scenario just by comparing the available generation capacity and the total load. If for example we investigate the last node on the generation level, we see that the available generation capacity with negligible operation cost is 150 kW. As it is stated in the problem that the losses will not exceed 3 kW, we can be certain that $TOC = 0$ and $LOLO = 0$ for all scenarios where the total load is less than 147 kW. When the total load is between 147 and 150 kW, the losses will determine if the wind power plant is sufficient or if the diesel generator set will be needed (which would result in $TOC > 0$). If the load is larger than 150 kW then the diesel generator set will definitely be needed, i.e., we will have $TOC > 0$. The question is now if there will be a power deficit

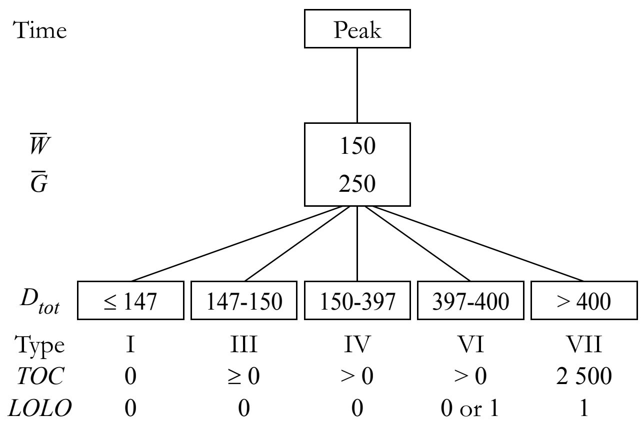


Figure 6.28 Part of the strata tree in example 6.30b.

or not. As long as the load plus the losses is not larger than the total generation capacity (400 kW) we will have *LOLO* = 0. We can be certain about this if the total load is between 150 and 397 kW. Between 397 and 400 kW, the losses will determine if we get *LOLO* = 0 or *LOLO* = 1. However, if the load exceeds 400 kW we can be certain that load shedding will be necessary. Hence, we can identify five intervals for the load, where the scenarios will have similar properties, and these five intervals are represented by five nodes in the strata tree, as illustrated in figure 6.28. A corresponding separation will be necessary also for the other generation capacity nodes in the strata tree from figure 6.27.

As we can see in the example above, strata trees can become quite big even for a small power system. For larger electricity markets, we must expect that there will be thousands of possible states for the available generation capacity, resulting in a huge strata tree having ten thousands of different branches. Much of the gain from applying stratified sampling will be lost if there would be so many strata. This problem can be managed by allowing one stratum to cover several branches of the strata tree. All scenarios of type I have the same characteristics; hence, all branches corresponding to type I scenarios could be merged into a single stratum, etc. It should however be noted that this method means that conditional probability distributions have to be used. How to manage these calculations is though beyond the scope of this presentation.

Sample Allocation

When strata have been defined, it is necessary to decide how many samples should be taken from each stratum. In general it is best to use the Neyman allocation, because it automatically allocates more samples to the strata where the uncertainty (i.e. the variance) is the largest, while strata with a large impact on the final result (i.e., high stratum weight) are prioritised. However, we are studying two—or maybe even more—result variables, while the Neyman allocation is based on comparing the variance in different strata for *one single* random variable. If the Neyman allocation is calculated with regard to *TOC* and when compared to the Neyman allocation for *LOLO*, we will in most cases get partly conflicting allocations. A simple solution to this dilemma is to simply use a compromise between the allocations which are optimal for each variable.

To use the Neyman allocation it is necessary to have estimates of the variance within each stratum.[32] To obtain such estimates we need to take a few samples from each stratum. Hence, at the

32. Cf. (6.48) and (6.49).

beginning of the simulation it is impossible to use the Neyman allocation. The solution is to divide a simulation into batches. In the first batch the number of samples per stratum is decided in advance. The appropriate number of samples varies from stratum to stratum. In for example a stratum with type I scenarios we will get the same value of *TOC* and *LOLO* for each studied scenario; hence, it is actually sufficient to study one single scenario (unless no other system indices as for example expected generation in different power plants, transmission between areas, etc., is to be studied). In strata with scenarios of type II we must in the first batch include enough samples to get both examples of scenarios where $TOC = 0$ and where $TOC > 0$. The appropriate number of samples in this kind of stratum is therefore depending on the probability of congestion scenarios. If the probability is 10% then it is almost certain that a series of 64 scenarios will include scenarios with $TOC = 0$ as well as scenarios where $TOC > 0$. If congestions is more unlikely then more scenarios are necessary; if the probability of congestion is 1% then it takes at least 500 scenarios. In a similar manner the number of scenarios per stratum necessary in the first batch can be analysed for the remaining types of strata.

When the result of the first batch has been compiled, it is possible to calculate how to allocate the samples in the next batch.

Example 6.31 (sample allocation): The first batch of a Monte Carlo simulation of an electricity market has just been completed. Ten samples were taken from each stratum. After analysing the results the estimates of the standard deviations recorded in table 6.17 were obtained. How should the 40 samples of the next batch be allocated?

Table 6.17 Results after the first batch of the simulation in example 6.31.

Stratum	Weight (ω_h)	Estimated standard deviation of the operation cost (s_{TOCh})	Estimated standard deviation of the risk of power deficit (s_{LOLOh})
1	0.1	0	0.08
2	0.2	750	0.007
3	0.3	1 500	0.002
4	0.4	1 000	0

Solution: First we determine how 80 samples should be distributed according to the is replaced bys_{TOCh}:

$$n_1^{TOC} = 80\frac{0.1 \cdot 0}{1\ 000} = 0,\ n_2^{TOC} = 80\frac{0.2 \cdot 750}{1\ 000} = 12,$$

$$n_3^{TOC} = 80\frac{0.3 \cdot 1\ 500}{1\ 000} = 36,\ n_4^{TOC} = 80\frac{0.4 \cdot 1\ 000}{1\ 000} = 32.$$

Then we apply the Neyman allocation to the *LOLO*:

$$n_1^{LOLO} = 80\frac{0.1 \cdot 0.08}{0.01} = 64,\ n_2^{LOLO} = 80\frac{0.2 \cdot 0.007}{0.01} = 11.2,$$

$$n_3^{LOLO} = 80\frac{0.3 \cdot 0.002}{0.01} = 4.8,\ n_4^{LOLO} = 80\frac{0.4 \cdot 0}{0.01} = 0.$$

A compromise between these two allocations is to take the mean of the desired number of samples in each stratum; after rounding we get

$$n_1 = 32,\ n_2 = 12,\ n_3 = 20,\ n_4 = 16.$$

From this allocation we subtract the number of samples that were made in the first batch, which yields that in the second batch we should take 22 samples from the first stratum, 2 from the second, 10 from the third and 6 from the fourth.

When the second batch is finished it is possible to calculate a new Neyman allocation based on the results of the first two batches. This procedure can be repeated after each batch, until an accurate enough result has been obtained.

A problem that might arise is that it might be detected that too many samples have been taken from a particular stratum. As the computation work of analysing these "unnecessary" samples already has been performed, it would be a waste of work to exclude these samples from the simulation. It is better to try to distribute the remaining samples as good as possible. We can say that those strata that have received too many samples have "stolen" samples from the remaining strata. The question is how the stolen samples should be divided between the strata; apparently there are several ways to solve this problem. An appropriate compromise is to reduce an equal share of samples from the affected strata. This can be done using the following algorithm:

Step 1. Determine the requested sample allocation for the b:th batch given by

$$n'_{h,b} = n_h^{\text{Neyman}} - \sum_{c=1}^{b-1} n_{h,c}, \tag{6.51}$$

where $n'_{h,b}$ denotes the number of samples in batch b from stratum h and n_h^{Neyman} is the total number of samples in stratum h according to the Neyman allocation.

Step 2. Let $\mathcal{H}^+$ be the index set of the strata which should be allocated more samples and $\mathcal{H}^-$ be the index set of those strata which have received too many samples, i.e.,

$$\mathcal{H}^+ = \{h: n'_{h,b} > 0\},\ \mathcal{H}^- = \{h: n'_{h,b} < 0\}. \tag{6.52}$$

Step 3. Calculate the total number of requested samples according to

$$\Delta^+ = \sum_{h \in \mathcal{H}^+} n'_{h,b} \tag{6.53}$$

and the total number of unnecessary samples according to

$$\Delta^- = -\sum_{h \in \mathcal{H}^-} n'_{h,b} \tag{6.54}$$

Step 4. The final sample allocation $n_{h,b}$ is obtained by choosing

$$n_{h,b} = 0\ \forall\ h \in \mathcal{H}^- \tag{6.55}$$

and

$$n_{h,b} = (1 - \Delta^-/\Delta^+)n'_{h,b}\ \ \forall\ h \in \mathcal{H}^+. \tag{6.56}$$

The algorithm is illustrated in the following example:

Example 6.32 (impossible sample allocation with several possible compromise solutions): The first batch of a Monte Carlo simulation of an electricity market has just been completed. Ten samples were taken from each stratum. After analysing the results the estimates of the standard deviations recorded in table 6.18 were obtained. How should the 40 samples of the next batch be allocated?

Solution: We start by examining the Neyman allocation of the operation cost:

$$n_1^{TOC} = 80\frac{0.1 \cdot 0}{1000} = 0,\ n_2^{TOC} = 80\frac{0.2 \cdot 500}{1000} = 8,$$

Table 6.18 Results after the first batch of the simulation in example 6.32.

Stratum	Weight (ω_h)	Estimated standard deviation of the operation cost (s_{TOCh})	Estimated standard deviation of the risk of power deficit (s_{LOLOh})
1	0.1	0	0.08
2	0.2	500	0.0055
3	0.3	1 600	0.003
4	0.4	1 050	0

$$n_3^{TOC} = 80\frac{0.3 \cdot 1600}{1000} = 38.4,\ n_4^{TOC} = 80\frac{0.4 \cdot 1050}{1000} = 33.6.$$

Then we study the Neyman allocation according to the *LOLO*:

$$n_1^{LOLO} = 80\frac{0.1 \cdot 0.08}{0.01} = 64,\ n_2^{LOLO} = 80\frac{0.2 \cdot 0.0055}{0.01} = 8.8,$$

$$n_3^{LOLO} = 80\frac{0.3 \cdot 0.003}{0.01} = 7.2,\ n_4^{LOLO} = 80\frac{0.4 \cdot 0}{0.01} = 0.$$

The compromise allocation after rounding is

$$n_1 = 32,\ n_2 = 8,\ n_3 = 23,\ n_4 = 17.$$

As 10 samples have been taken from each stratum we receive the following requested sample allocation in the second batch:

$$n_1^2 = 22,\ n_2^2 = -2,\ n_3^2 = 13,\ n_4^2 = 7.$$

The total number of unnecessary samples, Δ^-, is thus 2 and the total number of requested samples, Δ^+, is 42. The reduction in the strata which receive too few sample is therefore 2/42 ≈ 5%. After rounding we get the following sample allocation for the second batch:

$$n_1^2 = 21,\ n_2^2 = 0,\ n_3^2 = 12,\ n_4^2 = 7.$$

EXERCISES

6.1 The load in a certain power system is 100 MWh/h during workdays between 8 a.m. and 8 p.m., 80 MWh/h during weekends between 8 a.m. and 8 p.m., 60 MWh/h during evenings between 8 p.m. and 12 p.m., and 50 MWh/h during all other periods. Draw the load duration curve.

6.2 Show that (6.17) and (6.18) yields the same expected generation.

6.3 Prove equation (6.18).

Hint: A power plant g will be dispatched—provided that it is available—when the previous plant is available but does not have enough capacity to cover the load, or if the previous power plant is unavailable. How can the expected generation in these two cases be calculated using equivalent load duration curves?

6.4 Table 6.19 shows operation logs from 1999 of diesel generator set in the Tanzanian city Kigoma. Estimate the availability in this type of generator.

Table 6.19 Operation log from a diesel generator set in Kigoma.

Month	Operation time [h]	Forced outage time [h]
January	667.5	76.5
February	651	21
March	633	111
April	706	14

6.5 Figure 6.29 shows the duration curve of the total load in Rike. The installed capacity and the variable operation costs of the power sources uses in Rike are shown in i table 6.20. Assume that all power plants are 100% reliable.

Table 6.20 The power plants in Rike.

Power source	Total installed capacity [MW]	Variable operation cost [¤/MWh]
Hydro power	1 500	0
Nuclear power	1 500	100
Coal condensing	1 000	250

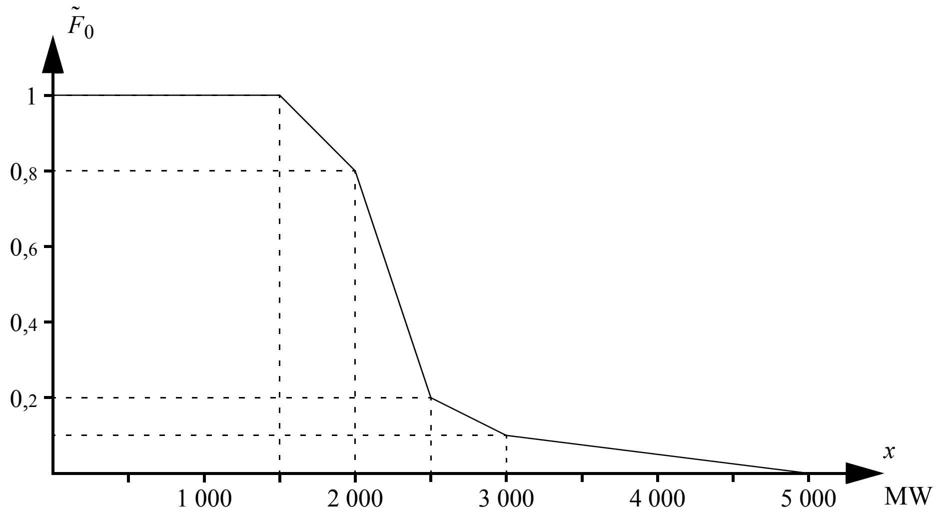


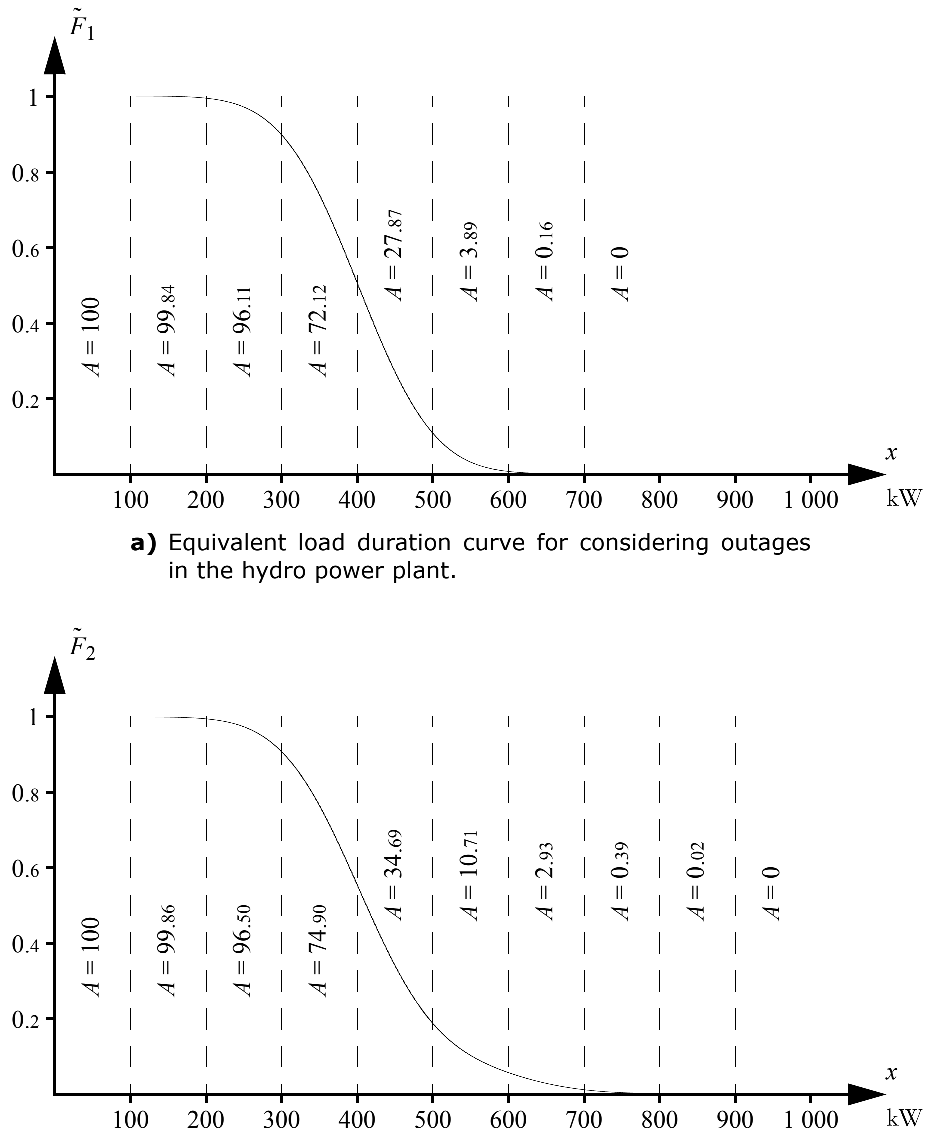
Figure 6.29 Total load duration curve for Rike.

a) Calculate the expected generation per hour in each power source.

b) Calculate the expected total operation cost per hour.

c) Calculate the risk of power deficit in Rike.

6.6 Ebbuga is a small town in Eastern Africa. The town is not connected to a national grid, but has a local system of its own. The local grid is supplied by a hydro power plant and a reserve capacity generator in the form of a diesel generator set. The hydro power plant is a run-of-the-river station and it has 500 kW capacity and the risk of failure is negligible. The natural flow in the river passing by the power plant is always sufficient to generate the installed capacity. The diesel generator set has an installed



a) Equivalent load duration curve for considering outages in the hydro power plant.

b) Equivalent load duration curve considering outages in both the hydro power plant and the diesel generator set.

Figure 6.30 Equivalent load duration curves in exercise 6.6.

capacity of 200 kW, the availability is 90% and the operation cost is 1 ¤/kWh. The load in Ebbuga, which includes the electrical losses in the local distribution grid, is normally distributed and has a mean of 400 kW and a standard deviation of 80 kW.

Figure 6.30 shows the equivalent load duration curves after adding the hydro power plant, $\tilde{F}_1$, and the diesel generator set, $\tilde{F}_2$. The figures also indicate the area below the duration curve for different intervals. What is the *ETOC* and *LOLP* of the system?

6.7 Assume that there is a load with the duration curve

$$\tilde{F}_0(x) = \begin{cases} 1 & x < 80, \\ 0.6 & 80 \le x < 100, \\ 0.1 & 100 \le x < 130, \\ 0 & 130 \le x. \end{cases}$$

a) Calculate *LOLP* and *EENS* if the system is supplied by a 120 MW unit, which is always available.

b) Calculate *LOLP* and *EENS* if the system is supplied by a 150 MW unit, which has 90% availability.

6.8 Large parts of the rural areas of Eggwanga have no access to electricity. To speed up the development, the Eggwangan authorities have decided to let private companies build and operate isolated power systems in rural areas. The condition is that the companies do not charge more than 1 ¤/kWh and that the risk of power deficit does not exceed 2%. In the small town Akabuga some local businessmen are considering to start the Akabuga Electricity Company Ltd. (AECL), which will build a local grid which can supply the town and its closes surroundings with electric power. One option is to supply the grid from diesel generator sets. It is planned to use a number of similar diesel generator sets, where each unit has a capacity of 200 kW, availability 90% and a generation cost of 0.50 ¤/kWh. The investment and maintenance costs of a diesel generator set is 50 000 ¤/year. The investment and maintenance costs of the distribution system is expected to be 2 M¤/year.

a) The equivalent load duration curve when four diesel generator sets is shown in table 6.21. Will AECL be profitable according to this calculation? (Disregard the *LOLP* requirement in this assignment.)

b) Would it be profitable to invest in another diesel generator set, so that the total installed capacity is 1 000 kW? (Continue to disregard the *LOLP* requirement.)

Table 6.21 Equivalent load duration curves in exercise 6.8.

$\tilde{F}_0(x)$	$\tilde{F}_1(x)$	$\tilde{F}_2(x)$	$\tilde{F}_3(x)$	$\tilde{F}_4(x)$	Interval
1	1	1	1	1	$x < 400$
0.6	0.64	0.676	0.7084	0.73756	$400 \le x < 600$
0.1	0.15	0.199	0.2467	0.29287	$600 \le x < 800$
0.01	0.019	0.0321	0.04879	0.068581	$800 \le x < 1\,000$
0	0.001	0.0028	0.00573	0.010036	$1\,000 \le x < 1\,200$
0	0	0.0001	0.00037	0.000906	$1\,200 \le x < 1\,400$
0	0	0	0.00001	0.000046	$1\,400 \le x < 1\,600$
0	0	0	0	0.000001	$1\,600 \le x < 1\,800$
0	0	0	0	0	$1\,800 \le x$

Table 6.22 Integral calculations of the equivalent load duration curve in exercise 6.9

	$\int_0^{200} \tilde{F}_g(x)dx$	$\int_{200}^{400} \tilde{F}_g(x)dx$	$\int_{400}^{600} \tilde{F}_g(x)dx$	$\int_{600}^{800} \tilde{F}_g(x)dx$	$\int_{800}^{1\,000} \tilde{F}_g(x)dx$	$\int_{1\,000}^{\infty} \tilde{F}_g(x)dx$
$g = 1$	200.00	199.66	119.11	21.17	1.73	0.01
$g = 4$	200.00	199.75	140.94	49.97	9.82	1.19

6.9 Another option for the electricity supply of Akabuga is to use diesel generator sets of the same sort as in exercise 6.8, as well as a small hydro power plant in Ekikko, which is located a short distance from the town. The hydro power plant has an installed capacity of 400 kW. Both operation cost and the risk of failures in the hydro power plant is negligible.

a) Figure 6.31 shows the load duration curve, $\tilde{F}_0$, as well as the equivalent load dura-

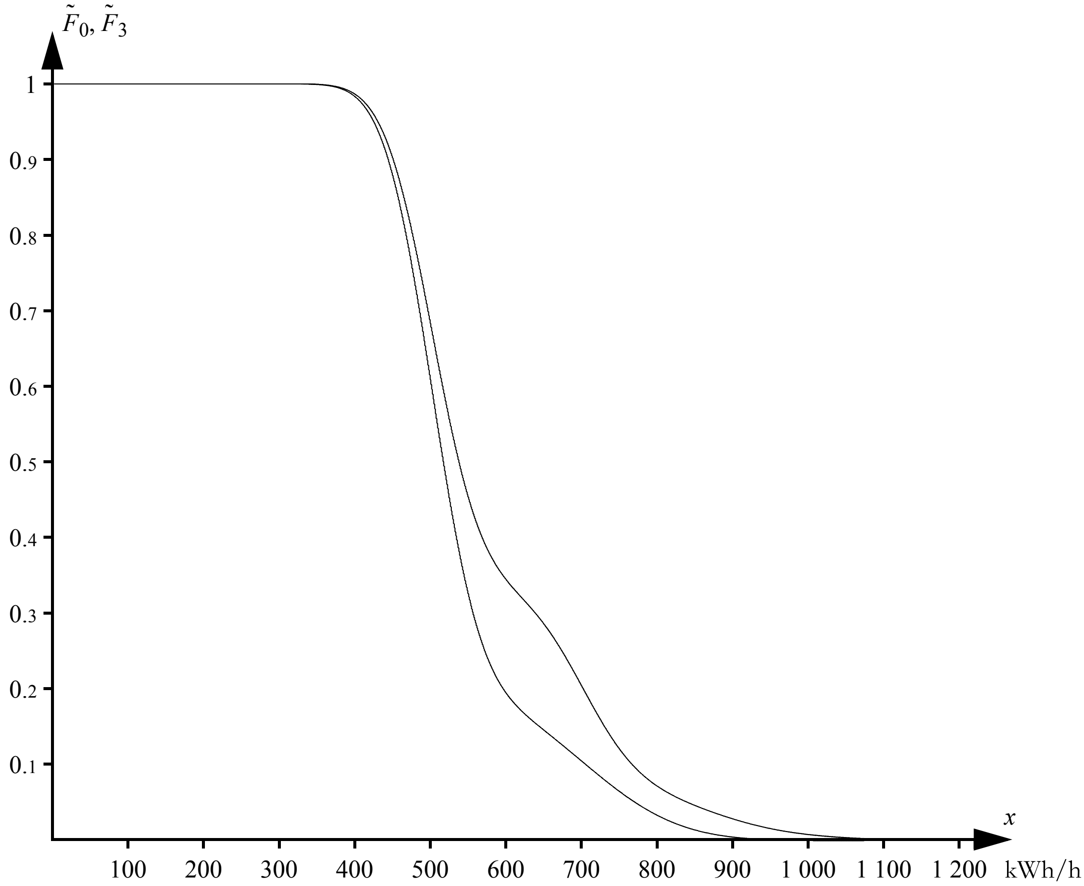


Figure 6.31 Equivalent load duration curves in exercise 6.9.

tion curve after adding the hydro power plant and two diesel generator sets. According to these calculations, will AECL fulfil the reliability of supply requirement, i.e., that the *LOLP* may not exceed 2%? If not, would it help to add a third diesel generator set?

Hint: It is not necessary to determine the entire equivalent load duration curve $\tilde{F}_4$ to calculate the *LOLP* including a third diesel generator set.

b) The water flow passing the power plant at Ekikko is assumed to always be sufficient to operate the power plant at its installed capacity. Assume that AECL chooses to install three diesel generator sets. Use the results displayed in table 6.22 to calculate the expected operation cost per hour.

6.10 Yet another option for Akabuga is to supply the grid using power form a hydro power plant at Ekiyira and a wind power plant in Olusozi, as shown in figure 6.32. The investment and maintenance costs to build the hydro power plant, the wind power plant as well as the transmission and distribution grid is estimated to 4 000 000 ¤/year. Data of the power plants is provided in table 6.23. Use the load duration curve $\tilde{F}_0$ from exercise 6.8. Use probabilistic production cost simulation to determine whether AECL will be profitable or not. (Disregard the *LOLP* requirement in this assignment too.)

6.11 In a certain power system the load is approximately normally distributed and the mean is 3 500 MW and the standard deviation is 300 MW. The system is supplied by five nuclear power plants having a capacity of 1 000 MW each and an availability of 95%.

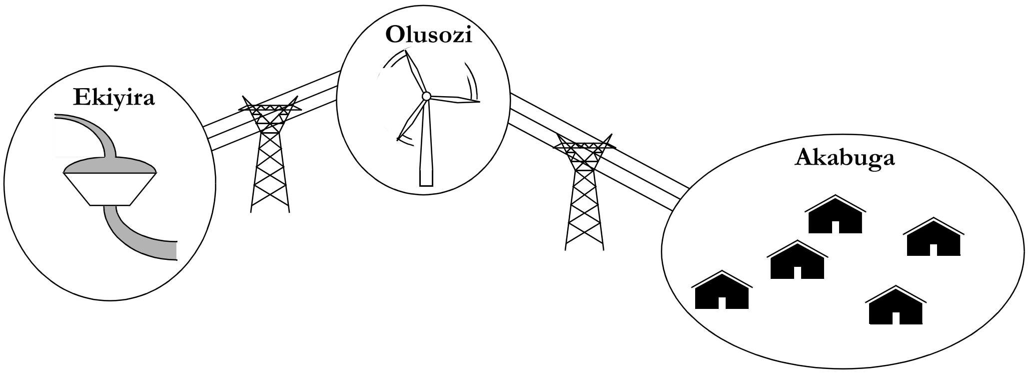


Figure 6.32 Suggested power system for Akabuga in exercise 6.10.

Table 6.23 Data of the planned power plants in exercise 6.10.

Power plant	Installed capacity [kW]	Generation cost [¤/kWh]	Possible states of the available generation capacity
Ekiyira	800	Negligible	800 kW (100%)
Olusozi	600	Negligible	600 kW (12%), 400 kW (18%), 200 kW (37%), 0 kW (33%)

a) Use the normal approximation to determine the system *LOLP*. The distribution function of the standardised normal distribution is found in appendix D.

b) Assume that *V* MW of wind power is added to this system and that the outage in these power plants can be assumed to be normally distributed with the mean $\mu_{O_v} = 0{,}6V$ and the standard deviation $\sigma_{O_v} = 0.1V$. Use the normal approximation of probabilistic production cost simulation to determine how much wind power is necessary to replace one of the nuclear power plants, without changing the system *LOLP* compared to part a.

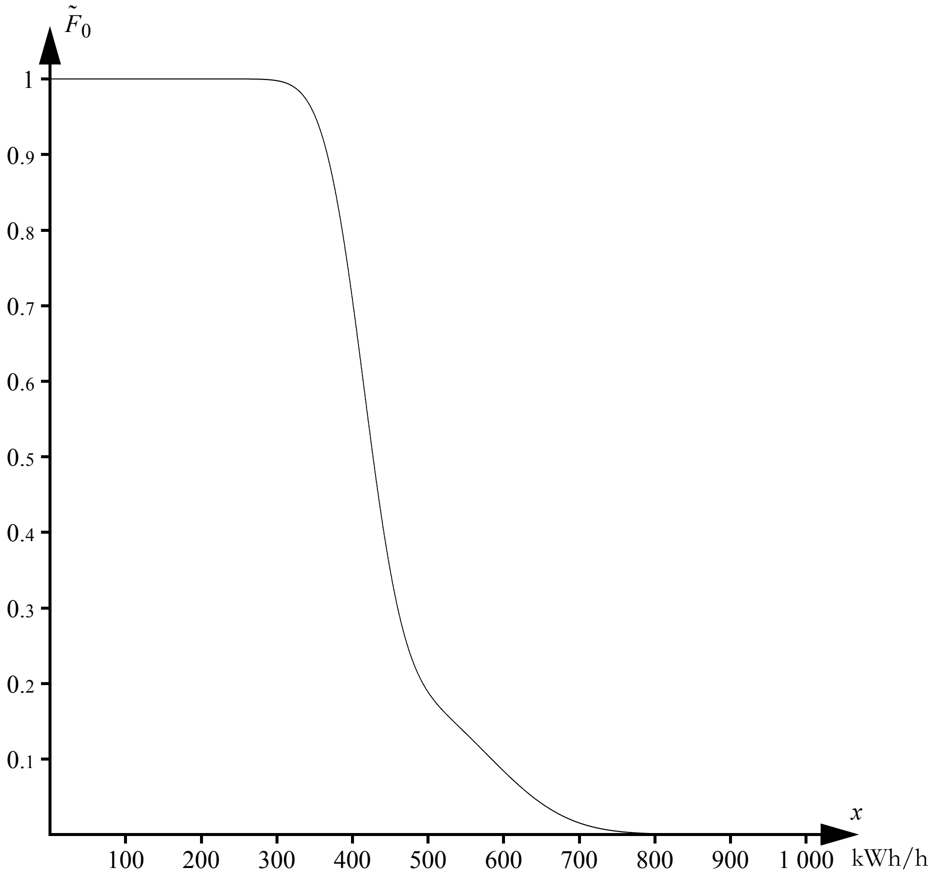
6.12 Assume that an electricity market has been simulated using Monte Carlo techniques and that the following results have been obtained:

$$\sum_{i=1}^{2\,000} toc_i = 512\,000 \text{ ¤/h}, \quad \sum_{i=1}^{2\,000} lolo_i = 16,$$

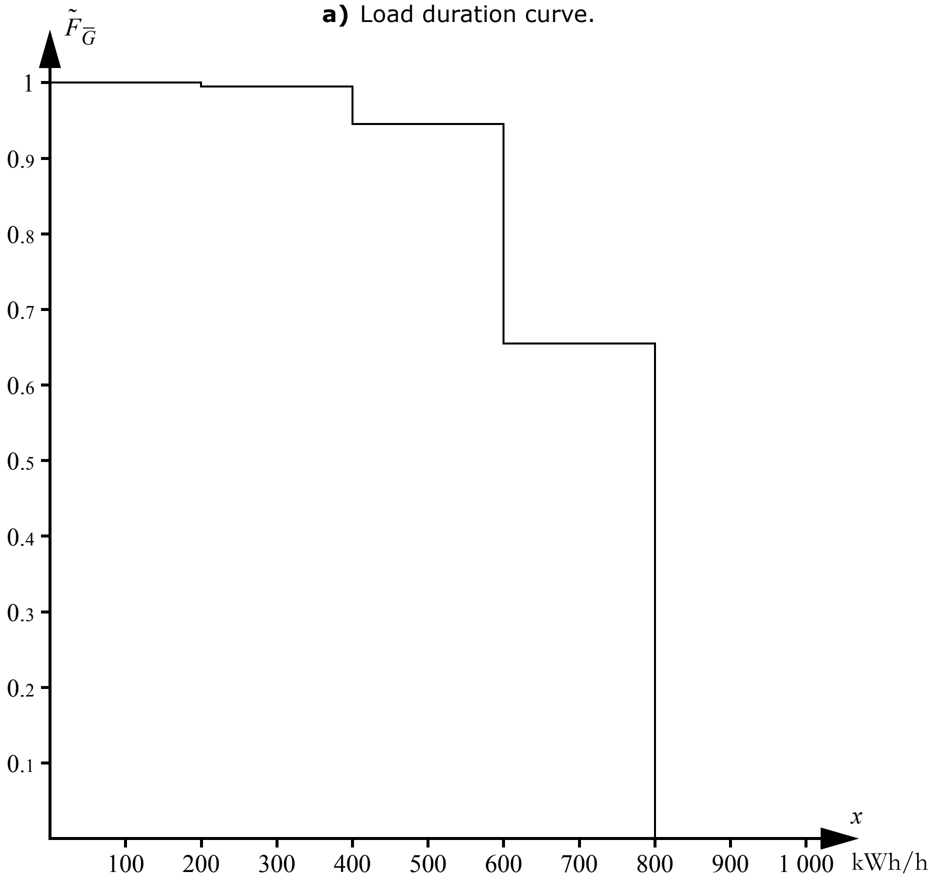
where toc_i and $lolo_i$ are the observations from scenario *i* of *TOC* and *LOLO* respectively. Which estimates of *ETOC* and *LOLP* are obtained from this simulation?

6.13 In a Monte Carlo simulation of an multi-area model, 10 000 scenarios have been generated. In 12 of these, the total unserved power, ΣU_n, was larger than zero. What is the *LOLP* estimate from this simulation?

6.14 Consider the suggested power system for Akabuga in exercise 6.8. Generate a scenario for a Monte Carlo simulation of the power system of Akabuga using the following random numbers from a *U*(0, 1)-distribution: 0.93 and 0.74. Use the duration curves of load as well as available generation capacity found in figure 6.33. Also generate the complementary scenarios. Which estimations of *ETOC* and *LOLP* are obtained from the four scenarios?



a) Load duration curve.



b) Duration curve of the total generation capacity in four diesel generator sets.

Figure 6.33 Duration curves of the scenario parameters in exercise 6.14.

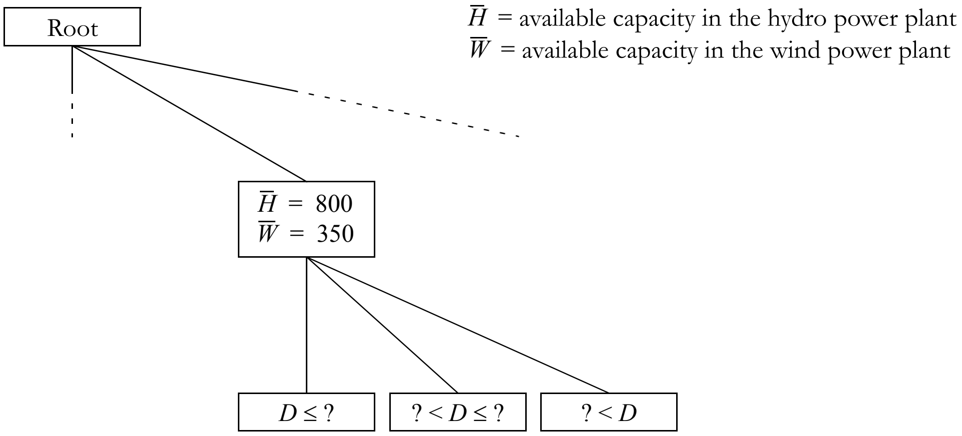


Figure 6.34 Part of the strata tree in exercise 6.16.

6.15 Consider the suggested power system for Akabuga in exercise 6.9.b. The losses on the line between Ekikko and Akabuga can be modelled as a quadratic function, $L(P) = 2\cdot10^{-5}P^2$. The following five scenarios have been generated in a Monte Carlo simulation of this system:

$\overline{H}$ [kW]	$\overline{G}_1$ [kW]	$\overline{G}_2$ [kW]	$\overline{G}_3$ [kW]	D
400	200	200	0	724
400	200	200	200	503
400	200	0	200	381
400	200	200	200	612
400	200	200	200	449

Which estimate of *ETOC* is obtained from these five scenarios if a control variate is used?

6.16 Consider the suggested power system for Akabuga in exercise 6.10. In a closer study of the suggested option, AECL has developed a more detailed model of the system, where for example the transmission losses are included. The losses between Ekiyira and Olusozi are modelled as a quadratic function, $L_{1,2}(P_{1,2}) = 3\cdot10^{-5}P_{1,2}^2$, and the losses on the line between Olusozi and Akabuga are assumed to be $L_{2,3}(P_{2,3}) = 8\cdot10^{-5}\cdot P_{2,3}^2$.

To make sure that the estimate of the risk of power deficit is accurate, it is recommended to use stratified sampling.

a) Figure 6.34 shows a part of the strata tree in this simulation. (The strata tree is very large—2 726 branches—and therefore just a small part is shown.) Suggest appropriate intervals for the nodes representing the load.

b) If each branch of the strata tree is treated as a separate stratum, there would be 2 726 strata. This is somewhat unnecessary; therefore, all strata of the same type has been merged together. Only three strata remain after the merging.

Three result variables are interesting in this simulation: *LOLO* (power deficit), *H* (hydro power generation) and *W* (wind power generation). After some simulation the results shown in tabell 6.24 have been collected. Which estimates do we get of the risk of power deficit and the generation in the power plants?

6.17 The first batch of a Monte Carlo simulation of an electricity market has just been completed. 16 samples were taken from each stratum. After analysing the results the

Table 6.24 Simulation results in exercise 6.16.

Stratum, h	Weight, ω	Number of scenarios, n_h	$\sum_{i=1}^{n_h} LOLO_{h,i}$	$\sum_{i=1}^{n_h} H_{h,i}$ [kWh/h]	$\sum_{i=1}^{n_h} W_{h,i}$ [kWh/h]
1	0.9560	110	0	41 905	20 548
2	0.0332	330	58	255 087	30 982
3	0.0108	8	8	6 400	210

estimates of the standard deviations recorded in table 6.25 were obtained. How should the samples of the next batch be allocated?

Table 6.25 Results after the first batch of the simulation inexercise 6.17.

Stratum	Weight (ω_h)	Estimated standard deviation of the operation cost (s_{TOCh})	Estimated standard deviation of the risk of power deficit (s_{LOLOh})
1	0.25	0	0.2
2	0.25	100	0.2
3	0.5	150	0

6.18 The first batch of a Monte Carlo simulation of an electricity market has just been completed. 12 samples were taken from each stratum. After analysing the results the estimates of the standard deviations recorded in table 6.26 were obtained. How should the samples of the next batch be allocated?

Table 6.26 Results after the first batch of the simulation in exercise 6.18.

Stratum	Weight (ω_h)	Estimated standard deviation of the operation cost (s_{TOCh})	Estimated standard deviation of the risk of power deficit (s_{LOLOh})
1	0.25	0	0.2
2	0.25	100	0
3	0.5	150	0

FURTHER READING

- M. Amelin, "Comparison of Capacity Credit Calculation Methods for Conventional Power Plants and Wind Power", *IEEE Transactions on Power Systems*, Vol. 24, No. 2, May 2009. — *Further details about capacity credit values.*
- H. Baleriaux, E. Jamoulle & F. Linard de Guertechin, "Simulation de l'explitation d'un parc de machines thermiques de production d'électricité couplé à des stations de pompage", *Extrait de la revue E* (édition S.R.B.E), Vol. 5, No. 7, 1967. — *One of the original references of probabilistic production cost simulation.*
- R. Billinton & W. Li, Reliability Assessment of Electric Power Systems Using Monte Carlo Methods, Plenum Press, New York 1994. — *Textbook which describes the basics of Monte Carlo simulation (random number generators, sampling and variance reduction techniques) and applications to reliability calculations in power systems.*
- R. R. Booth, "Power System Simulation Model Based on Probability Analysis", *IEEE Transactions on Power Apparatus & Systems*, vol. PAS-91, nr 1, januari/februari 1972. — *One of the original references of probabilistic production cost simulation.*
- W. G. Cochran, *Sampling Techniques*, third edition, John Wiley & Sons, 1977. — *Textbook which among other things provides a thorough description of stratified sampling.*

Appendix A

LINEAR PROGRAMMING

In this appendix it is briefly described what an optimisation problem is. The objective is tor introduce important notions, which should be known when formulating optimisation problems. The examples in this appendix have been kept small enough to be solved graphically. More general algorithms for solving optimisation problems are not described here—the interested reader should please refer to literature concerning optimisation theory and mathematical programming (see for example the literature mentioned at the end of this appendix).

A.1 OPTIMISATION THEORY

Optimisation theory (which is sometimes referred to as mathematical programming) is a branch of mathematics concerned with the maximisation or minimisation of a function defined on a set of feasible solutions. A general optimisation problem can be written as

$$\text{minimise} \quad f(\boldsymbol{x}), \tag{A.1a}$$

$$\text{subject to} \quad \boldsymbol{x} \in \mathcal{X}. \tag{A.1b}$$

The function to be minimised, $f(\boldsymbol{x})$, is referred to as the *objective function*. The variables which can be controlled when solving the problem are called *optimisation variables* and in the general form above denoted by the vector x. In most cases there are limitations on which values are allowed for x; this limitation is represented by a set of feasible solutions, $\mathcal{X}$. If these limitations are studied more closely it is possible to separate two kinds of limitations. The first is called *constraints*, and can in a general form be written as

$$g(\boldsymbol{x}) \le \boldsymbol{b}, \tag{A.2}$$

where $g(\boldsymbol{x})$ is a vector of functions, i.e., $g(\boldsymbol{x}) = [g_1(\boldsymbol{x}) \ \dots \ g_M(\boldsymbol{x})]^T$ and $\boldsymbol{b}$ is a constant vector. It should be noted that the constraints do not have to be inequalities; it is also possible to have equality constraints. The other kind of limitation is *variable limits*, which can be written in several ways. A common way is

$$\underline{\boldsymbol{x}} \le \boldsymbol{x} \le \bar{\boldsymbol{x}}, \tag{A.3}$$

where $\underline{\boldsymbol{x}}$ and $\bar{\boldsymbol{x}}$ are constant vectors representing the lower and upper limits of the allowed values of $\boldsymbol{x}$. The difference between the constraints and limits is that the former normally include more than one optimisation variable, whereas the latter only concern one variable. Together, (A.2) and (A.3) defines the feasible solutions $\mathcal{X}$.

Linear Programming

A linear programming problem (LP problem) is a special class of optimisation problem, where the objective function as well as the constraints are linear functions. An LP problem in standard form having N constraints and M variables is formulated as follows:

$$\text{minimise} \quad c_1x_1 + c_2x_2 + \ldots + c_Mx_M \tag{A.4}$$

$$\begin{aligned}\text{subject to} \quad & a_{1,1}x_1 + a_{1,2}x_2 + \ldots + a_{1,M}x_M = b_1,\\ & a_{2,1}x_1 + a_{2,2}x_2 + \ldots + a_{2,M}x_M = b_2,\\ & \ldots\\ & a_{N,1}x_1 + a_{N,2}x_2 + \ldots + a_{N,M}x_M = b_N,\\ & x_i \geq 0,\ i = 1, \ldots, M.\end{aligned}$$

This problem can be formulated in a more compact manner using matrices:

$$\text{minimise} \quad \boldsymbol{c}^T\boldsymbol{x} \tag{A.5}$$

$$\text{subject to} \quad \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}, \tag{A.5a}$$

$$\boldsymbol{x} \geq 0, \tag{A.5b}$$

where

$$\boldsymbol{c} = \begin{bmatrix} c_1 \\ \vdots \\ c_M \end{bmatrix},\ \boldsymbol{x} = \begin{bmatrix} x_1 \\ \vdots \\ x_M \end{bmatrix},\ \boldsymbol{A} = \begin{bmatrix} a_{1,1} & \ldots & a_{1,M} \\ \vdots & \ddots & \vdots \\ a_{N,1} & \ldots & a_{N,M} \end{bmatrix},\ \boldsymbol{b} = \begin{bmatrix} b_1 \\ \vdots \\ b_N \end{bmatrix}.$$

An advantage of using linear programming is that there is commercial software available, which can provide fast solutions to problems having thousands or tens of thousands optimisation variables.

Formulating Optimisation Problems

When formulating optimisation problem is very important to be clear. Even relatively simple problems may require a large number of variables and parameters; hence, misunderstandings easily occur if the problem is not formulated so that it is easy to read and understand. What is easy to read and understand is of course depending on the context. A mathematician, who wants to analyse how the problem should be solved, will probably prefer to have the problem formulated in some kind of standard form, which often includes a very compact formulation using vectors, matrices, etc. (cf. the previous section).

This strict mathematical form is however not very practical for engineering purposes, when it is more interesting to see the coupling between the real problem and its mathematical formulation. It is for example impractical to collect all optimisation variables in a vector x; it is preferable choose such symbols for the optimisation variables so that it is possible to immediately realise which quantity they denote. For the same reason it is impractical to use a vector function to describe the constraints; it is better to write each constraint separately. If the problem includes several similar constraints then it is possible to avoid repeating all these constraints by introducing appropriate indices for the concerned variables and parameters. In this compendium, we will stick to this way of formulating optimisation problems.

Even though a problem is formulated in a less strict mathematical form it is recommended to try to keep the structure of the optimisation problem, i.e., it should be easy to see what is the objective function, which are the constraints and which are the variable limits. It is also important to clearly state which symbols denote optimisation variables and which denote problem parameters. An appropriate routine when formulating optimisation problems is therefore as follows:

- **Formulate the problem in words.** Start by thinking through the problem you are about to formulate and make sure that you understand it. Then describe your problem in words.
- **Define symbols.** Review which variables and parameters which are involved in the problem and state clear definitions of them. If you are using indices to distinguish similar variables it is recommended to already at this point state the possible values for each index and each variable.
- **Write the mathematical formulation.** Formulating the problem can now be considered as the problem of translating the problem formulated in words into equations and inequalities using the defined notation.

A.2 EXAMPLES OF LP PROBLEMS

We start by an example of how an LP problem is formulated in standard form:

Example A.1 (LP problem in standard form): Alice's mother will have some guests and wants to serve coffee and something light to eat. The mother says "Alice, can you please go to the shop? I would like to fill the fruit bowl (which can contain two litres) and my five guests should have at least two each of what you buy. Here is 100 SEK. You can keep the change."

Alice thinks that apples and pears will be nice. The pears cost 3 SEK each and apples 5 SEK each. It can be assumed that there are 6 pears to 1 litre; the apples are larger and have a volume of 0.3 litres each. Formulate Alice's planning problem as an LP problem in standard form.

Solution: Let x_1 denote the number of pears and x_2 the number of apples Alice is buying. The problem can be formulated Alice minimising the costs of the purchase wile fulfilling the requirements of her mother:

$$\text{minimise} \quad z = 3x_1 + 5x_2$$

$$\text{subject to} \quad \frac{1}{6}x_1 + 0.3x_2 \geq 2,$$

$$x_1 + x_2 \geq 10,$$

$$x_1 \geq 0, x_2 \geq 0.$$

The solution to this problem is shown in figure A.1. The borders of the constraints and the variable limits are marked by dashed lines and the area which does not fulfil the requirements is shaded. Different combinations of x_1 and x_2 which result in the same value of the objective function are shown by thin, straight lines. As can be seen in the figure, the minimal value of z is obtained in the intersection of the constraints $x_1/6 + 0.3x_2 \geq 2$ and $4x_1 + x_2 \geq 10$. In that point we have $x_1 = 7.5$ and $x_2 = 2.5$, i.e., Alice should by 7.5 pears and 2.5 apples.

To write Alice's problem in standard form we must introduce slack variables, to obtain equality constraints. Hence, we get the following constraints:

$$\frac{1}{6}x_1 + 0.3x_2 - x_3 = 2,$$

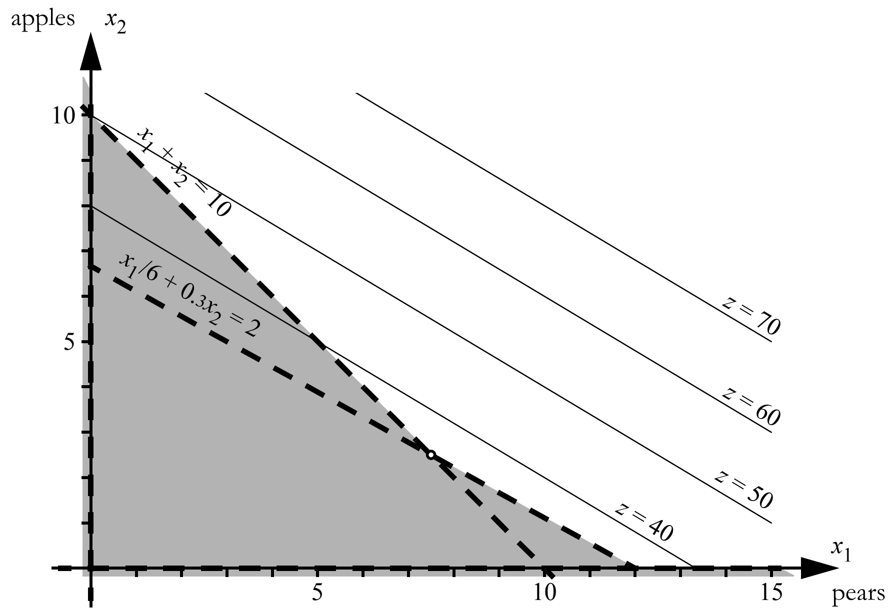


Figure A.1 Solution to example A.1: $x_1 = 7.5$, $x_2 = 2.5$, $z = 35$.

$$x_1 + x_2 - x_4 = 10,$$

where x_3 is the volume margin and x_4 is the number of pieces margin. Moreover, we have to make sure that the margins do not become negative:

$$x_3 \geq 0, x_4 \geq 0.$$

The matrix A and the vectors b and c in the standard form will then get the following values in this problem:

$$A = \begin{bmatrix} \frac{1}{6} & \frac{3}{10} & -1 & 0 \\ 1 & 1 & 0 & -1 \end{bmatrix}, \; b = \begin{bmatrix} 2 \\ 10 \end{bmatrix}, \; c = \begin{bmatrix} 3 \\ 5 \\ 0 \\ 0 \end{bmatrix}.$$

It is easily understood that the constraints will be very important for which solution we will get. The more constraints we add to the problem, the more limited the number of feasible solutions will become; hence, it is more likely that we have to accept a worse solution than if the freedom of choice had been larger. If too many constraints are added the problem might become impossible to solve, because there simply is no feasible solution:

Example A.2 (no feasible solution): Alice's father says: "Do not buy more than 1 kg". Assume that a pear weights 1/6 kg and an apple 0.3 kg. How will this instruction affect Alice's purchase?

Solution: The requirement of Alice's father can be formulated as

$$\frac{1}{6}x_1 + 0.3x_2 \leq 1.$$

This problem is shown in figure A.2. The figure shows that there is no point which fulfils all constraints. The problem has no solution.

Sometimes, however, a new constraint will have no impact on the solution at all:

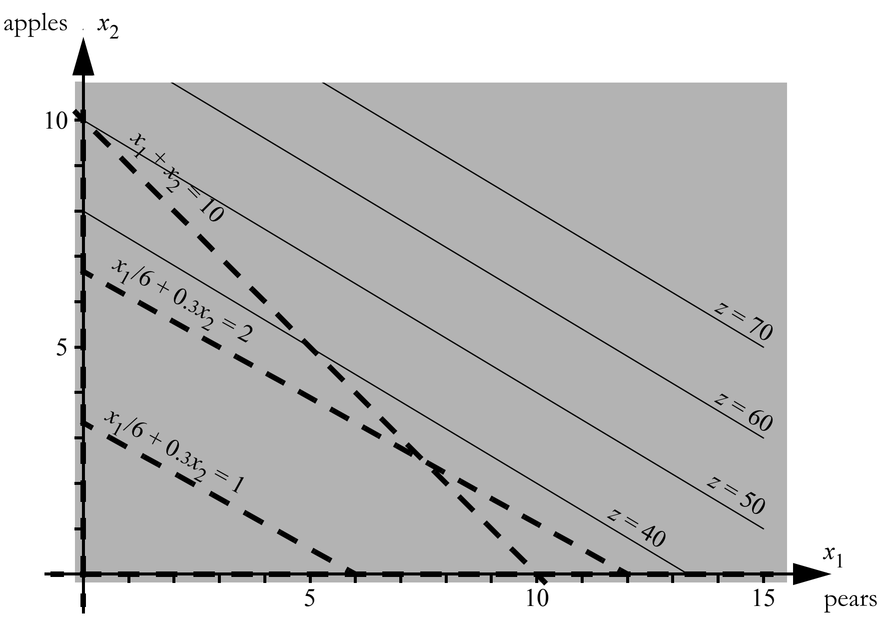


Figure A.2 Example A.2 has no feasible solution.

Example A.3 (non-binding constraint): When Alice arrives at the shop, she finds that there is only thirteen pears left. How will this affect Alice's purchase? (Disregard the instruction of Alice's father in example A.2).

Solution: Alice can apparently not buy more than thirteen pears, which means that we can add another constraint:

$$x_1 \leq 13.$$

The solution to this problem is shown in figure A.3 As can be seen in the figure, the solution is not affected by the new constraint, because the constraint is already fulfilled by the old solution. The new constraint is an example of a non-binding constraint.

In example A.3 there were still some constraints which established a lower limit on the value of the objective function. In some cases it may turn out that the constraints do not impose any limits at all and then it is not possible to distinguish an optimal solution:

Example A.4 (unbounded problem): The mother says: "Alice, here is my credit card. I will give you 1 SEK for each item you bring home." How will this affect Alice's purchase? (Disregards the limited amount of pears in example A.3.)

Solution: Alice's objective function now becomes

$$\text{maximise} \qquad z = x_1 + x_2,$$

which in standard form is expressed as

$$\text{minimise} \qquad z = -x_1 - x_2.$$

The problem is shown in figure A.4. The figure shows that the value of the objective function decreases when x_1 and x_2 increases. This means that Alice can buy an "infinite" amount of fruit and earn an "infinite" amount of money. (From Alice's point of view there is an infinite amount of money on the credit card). This implies that the problem does not have a finite solution.

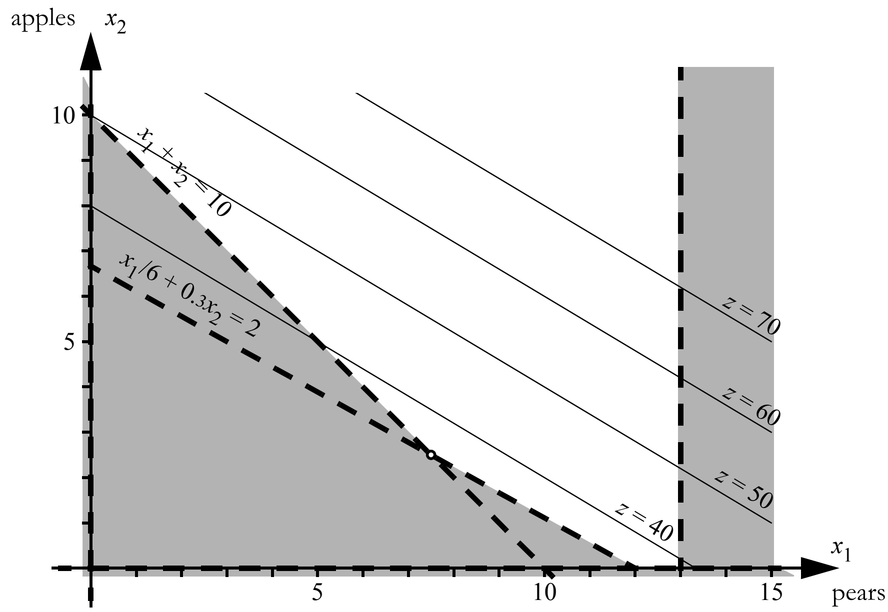


Figure A.3 Solution to example A.3: $x_1 = 7.5$, $x_2 = 7.5$, $z = 35$.

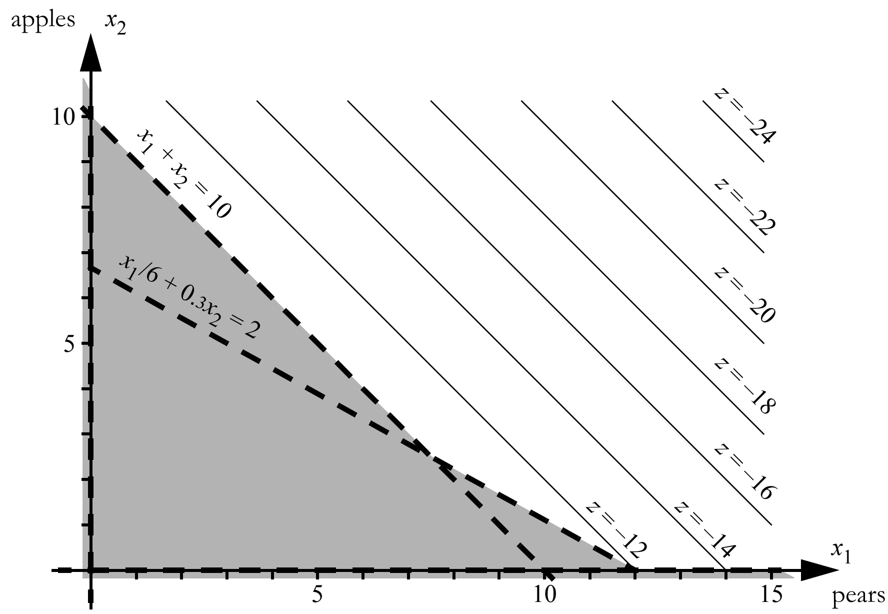


Figure A.4 Solution to example A.4: $x_1 = \infty$, $x_2 = \infty$, $z = -\infty$.

An interesting property of LP problems is that a small change in the objective function does not affect the optimal solution, but only the value of the objective function, as shown in the following example:

Example A.5 (new objective function): When Alice arrives at the shop she finds that pears cost 4 SEK each. How will this affect Alice's purchase?

Solution: The objective function now becomes

minimise $\quad z = 4x_1 + 5x_2$

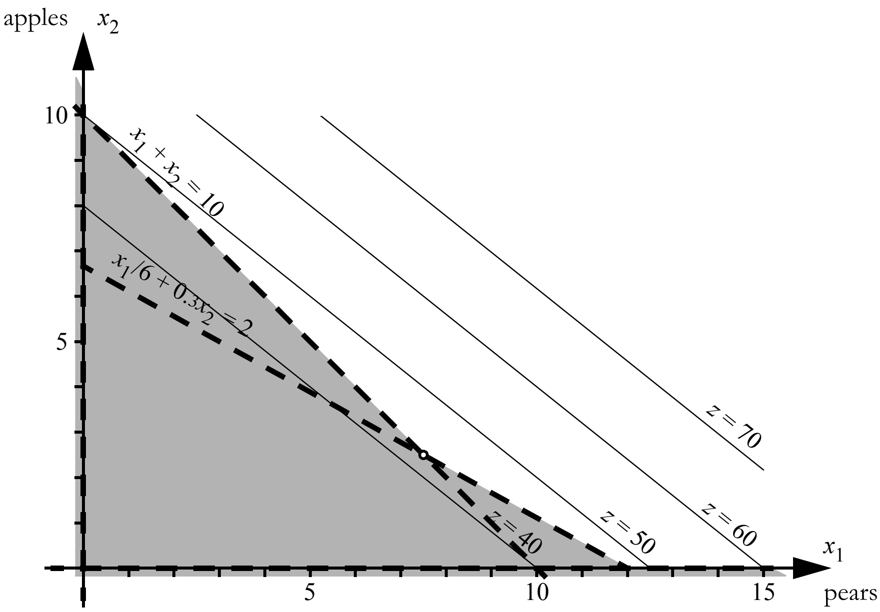


Figure A.5 Solution to example A.5: $x_1 = 7.5$, $x_2 = 2.5$, $z = 42.5$.

and the new problem is shown in figure A.5. As can be seen in the figure, the same point is optimal—Alice should still buy 7.5 pears and 2.5 apples, but the total cost is now 42.50 SEK.

In some cases there are several solutions which are equally good. This means that there in these cases are an finite optimal value of the objective function, but there is an infinite number of solutions.

Example A.6 (degenerated solution): Assume that Alice finds that the pears cost 5 SEK each. How will this affect Alice's purchase?

Solution: The objective function now becomes

minimise $z = 4x_1 + 5x_2$,

which is shown in figure A.6. In this case there is no unique solution, but an infinite number of solutions all providing the optimal value $z = 50$.

Some problems have several different solutions which all give a value of the objective function which is close to the optimal—it is said that these problems have a flat optimum. A flat optimum indicates that the problem is almost degenerated. In practice a flat optimum means that a some deviations from the optimal solution might be tolerable, as the impact on the objective function is small.

Example A.7 (flat optimum): Compare the following two cases:

a) Pears cost 4.90 SEK each and apples 5 SEK each,

b) Pears cost 5 SEK each and apples cost 4,90 SEK each.

Solution: The objective function in case a is

minimise $z = 4.9x_1 + 5x_2$,

and in case b we get

minimise $z = 5x_1 + 4.9x_2$.

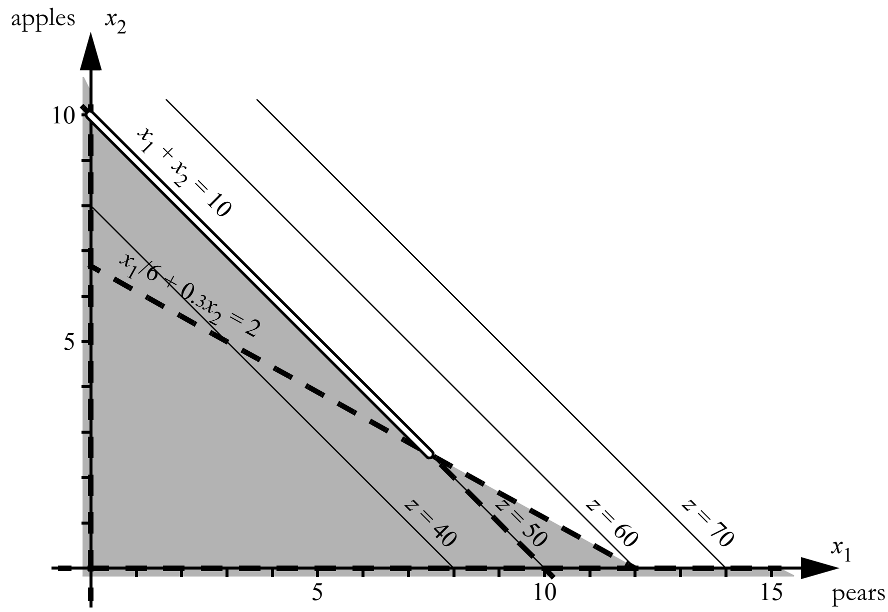


Figure A.6 Solution to example A.6: $x_1 \in [0, 7.5]$, $x_2 = 10 - x_1$, $z = 50$.

Both solutions are displayed in figure A.7. As can be seen in the figure, we get two very different solutions, but the optimal value of the objective function is approximately the same in both cases. Also notice that in both cases the value of the objective function would not deteriorate very much if we choose the wrong alternative. In case a it is optimal to buy 7.5 pears and 2.5 apples (cost 49.25 SEK), but if Alice would buy only apples instead, the cost would just increase by 0.75 SEK. Also in case b the cost would only increase by 0.75 SEK if she buys 7.5 pears and 2.5 apples instead of buying just apples.

Each LP problem has a corresponding so-called dual problem. Analysis of the dual problem can provide useful information about the original (the primal) problem. A more thorough explanation of the relation between primal and dual problems would be too complex for this appendix, but in the following two examples we will show how to obtain the dual problem and how the solution can be utilised.

Example A.8 (dual formulation of an LP problem): When Alice is passing by the candy in the shop, she recalls that her mother din not say anything about buying fruit. Alice could buy popcorn and bonbons instead, but would that be cheaper?

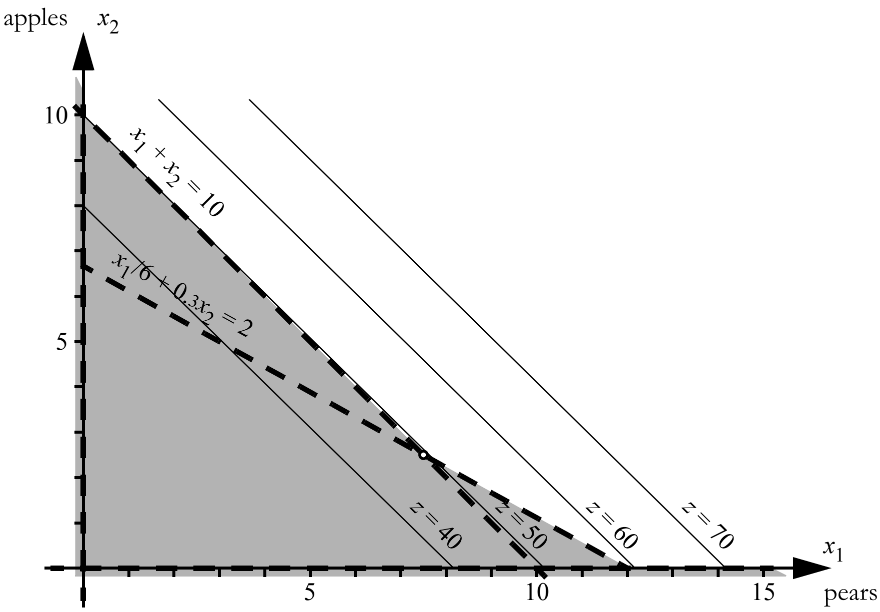
Assume that popcorn cost λ_1 SEK/litre and the bonbons cost λ_2 SEK each. Moreover, assume that the volume of the bonbons is negligible and that the number of popcorn is not considered. Let x_5 be the number of litres of popcorn and x_6 the number of bonbons. Which is the maximal prices of popcorn and bonbons which Alice can afford if she does not want to loose money on this alternative?

Solution: With these new conditions, Alice's problem can be written as

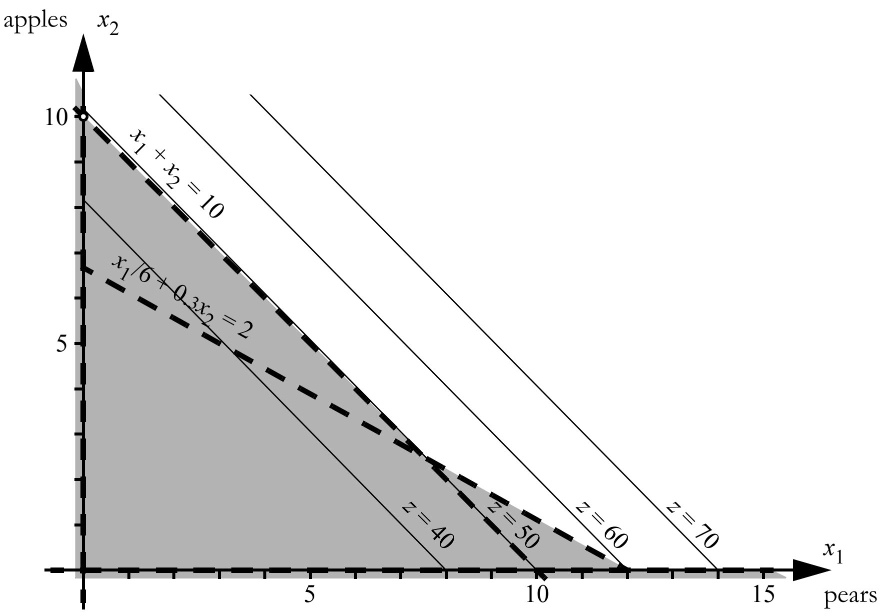
$$\Phi(\lambda_1, \lambda_2) = \text{minimise} \quad 3x_1 + 5x_2 + \lambda_1 x_5 + \lambda_2 x_6$$

$$\text{subject to} \quad \frac{1}{6}x_1 + 0.3x_2 - x_3 + x_5 = 2,$$

$$x_1 + x_2 - x_4 + x_6 = 10,$$



a) $x_1 = 7.5$, $x_2 = 2.5$, $z = 49.25$.



b) $x_1 = 0$, $x_2 = 10$, $z = 49$.

Figure A.7 Solutions to example A.7.

$$x_i \geq 0,\ i = 1, \ldots, 6,$$

where the prices λ_1 and λ_2 both must be larger than or equal to zero. If we rearrange the constraints to obtain expressions for x_5 and x_6, and substitute these expressions into the objective function, then the problem becomes

$$\begin{aligned} \Phi(\lambda_1, \lambda_2) = \text{ minimise} \quad & 3x_1 + 5x_2 + \lambda_1\left(2 - \frac{1}{6}x_1 - 0.3x_2 - x_3\right) + \\ & + \lambda_2(10 - x_1 - x_2 - x_4) \\ \text{subject to} \quad & x_i \geq 0,\ i = 1, \ldots, 4. \end{aligned}$$

This means that we have applied Lagrange relaxation to the constraints in example A.1. This has resulted in two new variables: λ_1 and λ_2. λ_1 is a dual variable connected to the first constraint (because the first constraint is a volume constraint and popcorn are assumed to only possess volume). Similarly, λ_2 is a dual variable of the second constraint (because it is a quantity constraint and bonbons are considered only by their numbers). We can say that λ_1 is a volume value and λ_2 a quantity value.

To simplify the analysis of Alice's problem we rewrite it once more:

$$\begin{aligned} \Phi(\lambda_1, \lambda_2) = \text{ minimise } \quad & x_1\left(3 - \frac{1}{6}\lambda_1 - \lambda_2\right) + x_2(5 - 0.3\lambda_1 - \lambda_2) + \\ & + 2\lambda_1 + 10\lambda_2 + \lambda_1 x_3 + \lambda_2 x_4 \\ \text{subject to } \quad & x_i \geq 0, i = 1, \ldots, 4. \end{aligned}$$

The expressions $\lambda_1 x_3$ and $\lambda_2 x_4$ must be larger than or equal to zero (as all involved variables are larger than or equal to zero) and are consequently minimised if Alice chooses $x_3 = x_4 = 0$. We can now investigate for which prices popcorn and bonbons are preferable for Alice. If

$$3 - \frac{1}{6}\lambda_1 - \lambda_2 \geq 0$$

then the product $x_3(3 - \lambda_1/6 - \lambda_2)$ must also be larger than or equal to zero. In that case the expression is minimised if Alice does not buy any pears, i.e., if she chooses $x_1 = 0$. By similar reasoning we can conclude that she will not buy any apples (i.e., she will choose $x_2 = 0$) if

$$5 - 0.3\lambda_1 - \lambda_2 \geq 0.$$

Which prices are then the highest which Alice can afford if she is to prefer popcorn and bonbons? We can formulate this as a new optimisation problem:

$$\begin{aligned} \text{maximise} \quad & z = 2\lambda_1 + 10\lambda_2 \\ \text{subject to} \quad & \frac{1}{6}\lambda_1 + \lambda_2 \leq 3, \\ & 0.3\lambda_1 + \lambda_2 \leq 5, \\ & \lambda_1 \geq 0, \lambda_2 \geq 0. \end{aligned}$$

This problem is referred to as the *dual* problem of Alice's original problem. In figure A.7 we see that the solution to the dual problem is $\lambda_1 = 15$ and $\lambda_2 = 0.5$. Hence, given these prices Alice will choose $x_1 = x_2 = x_3 = x_4 = 0$. To fulfil the constraints of the original problem she therefore chooses $x_5 = 2$ and $x_6 = 10$, i.e., she buys two litres popcorn and ten bonbons. Her cost for this purchase is 35 SEK, which is the same cost as for buying 7.5 pears and 2.5 apples. Thus, at these prices Alice does not ear anything on buying popcorn and bonbons instead of apples and pears. If the prices were just somewhat lower than 15 SEK/litre and 0.5 SEK each respectively, it would be profitable for Alice to fill the fruit bowl by popcorn and bonbons.

Example A.9 (application of dual variables): While Alice is walking to the shop she considers how much she could earn by deceiving her mother, for example by buying only 1,9 litres instead of 2.

Solution: The primal problem in this case is

$$\text{minimise} \quad z = 3x_1 + 5x_2$$

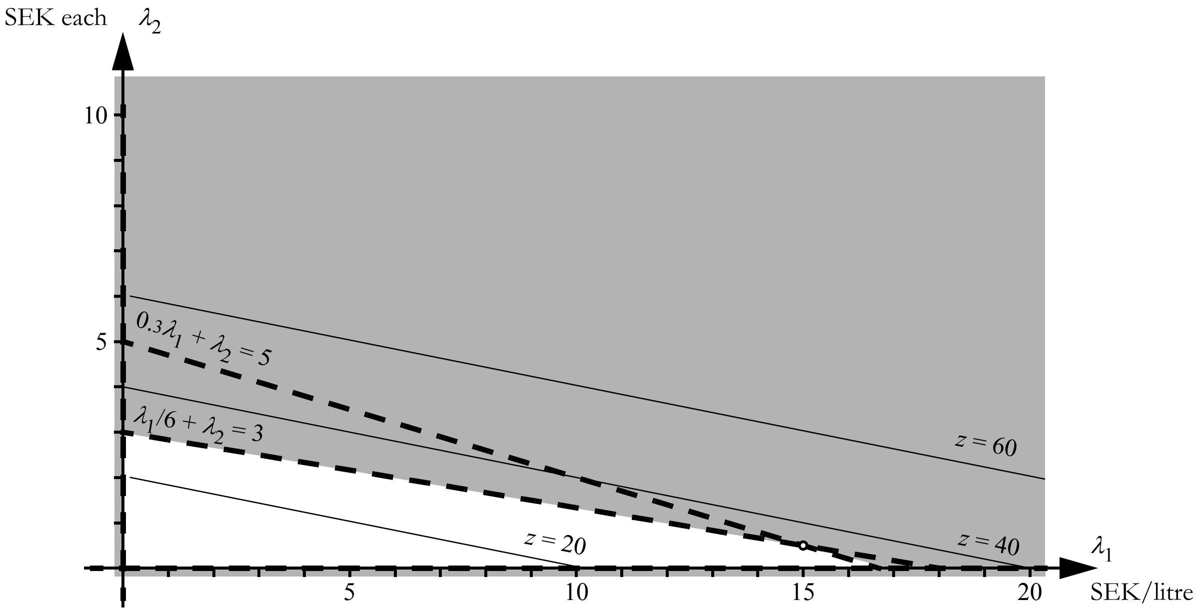


Figure A.8 Solution to the dual problem in example A.8: $\lambda_1 = 15$, $\lambda_2 = 0.5$, $z = 35$.

$$\text{subject to} \quad \frac{1}{6}x_1 + 0.3x_2 \geq 1.9,$$

$$x_1 + x_2 \geq 10,$$

$$x_1 \geq 0,\ x_2 \geq 0.$$

Assume that we solve the dual problem instead, which in this case becomes

$$\text{maximise} \quad z = 1.9\lambda_1 + 10\lambda_2$$

$$\text{subject to} \quad \frac{1}{6}\lambda_1 + \lambda_2 \leq 3,$$

$$0.3\lambda_1 + \lambda_2 \leq 5,$$

$$\lambda_1 \geq 0,\ \lambda_2 \geq 0.$$

This implies only a small modification of the objective function compared to the dual problem in example A.8. The solution is shown in figure A.9. As can be seen in the figure, the modification of the objective function is so small that the optimal solution stays in the same corner of the feasible region, i.e., $\lambda_1 = 15$ and $\lambda_2 = 0.5$. The optimal value of the objective function has however changed to

$$z = 1.9\lambda_1 + 10\lambda_2 = 33.50.$$

Rather than solving the dual problem again we could have assumed that the optimal solution would not move, which means that we can study the impact of the modification by using the dual variables:

$$\Delta z = (1.9 - 2)\lambda_1 = -0.1\lambda_1 = -1.5.$$

Hence, the dual variables can be seen as the marginal value of the limit imposed by a constraint. In this case, λ_1 is the marginal value of a volume change.

The dual variables can thus be used for quick estimations of how the optimal value of the objec-

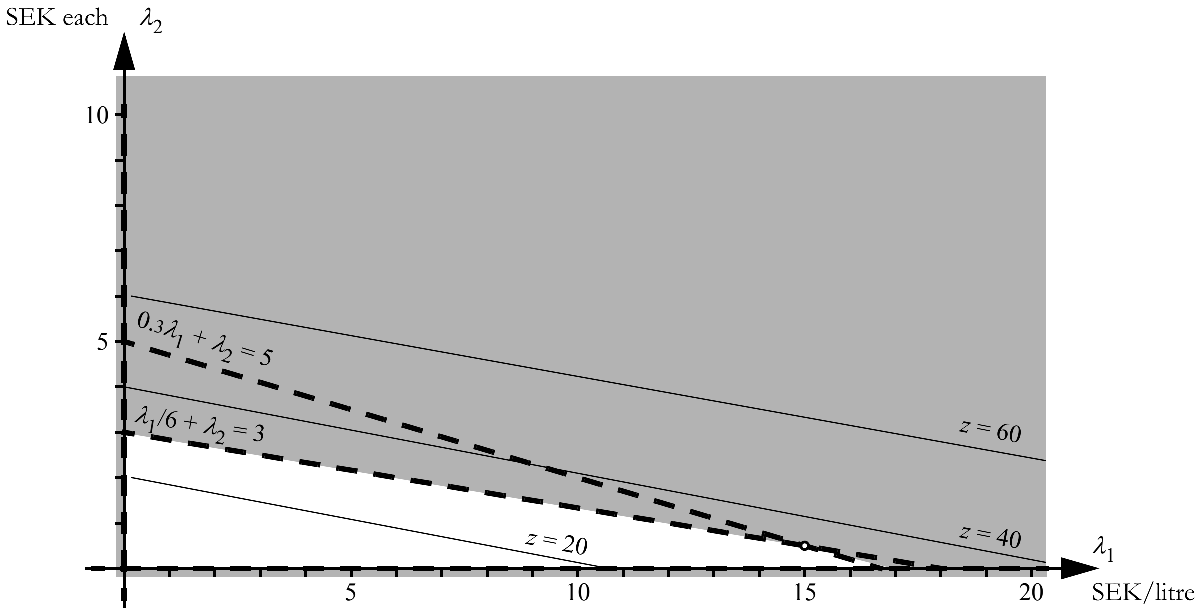


Figure A.9 Solution to the dual problem in example A.9: $\lambda_1 = 15$, $\lambda_2 = 0.5$, $z = 33.5$.

tive function will change when modifying the right hand side of a constraint. The estimate is however only correct if the optimal solution point of the dual problem does not change. Hence, in practice the dual variables can only be used to calculate the new optimal value of the objective function when small changes are made in the right hand side. Unfortunately, it is no possible to guarantee in advance whether or not a change is small; therefore, the results of this kind of calculations should not be trusted without consideration. Below follows an example of a change in the right hand side which is too large to allow the dual variables to be used for calculating the new optimal value:

Example A.10 (misleading application of dual variables): While Alice is walking to the store, she decides to surprise her mother by buying twice as many fruits as necessary. How much more will this cost compared to example A.1?

Solution: If we assume that the dual variables remain unchanged the optimal value of the objective function will change by

$$\Delta z = \Delta b_2 \lambda_2 = 10\lambda_2 = 5,$$

i.e., buying 20 fruits will cost 5 SEK more than just buying 10 fruits.

Is this correct? The dual problem in this case is

$$\text{maximise} \quad z = 2\lambda_1 + 20\lambda_2$$

$$\text{subject to} \quad \frac{1}{6}\lambda_1 + \lambda_2 \leq 3,$$

$$0.3\lambda_1 + \lambda_2 \leq 5,$$

$$\lambda_1 \geq 0,\ \lambda_2 \geq 0.$$

The solution to the dual problem is shown in figure A.10. As can be seen in the figure, the optimal solution is no longer found in the same corner of the feasible area as in examples A.8 and A.9. The optimal solution is $\lambda_1 = 0$ and $\lambda_2 = 3$, which results in the optimal value $z = 60$. As the optimal values of the primal and dual problem are the

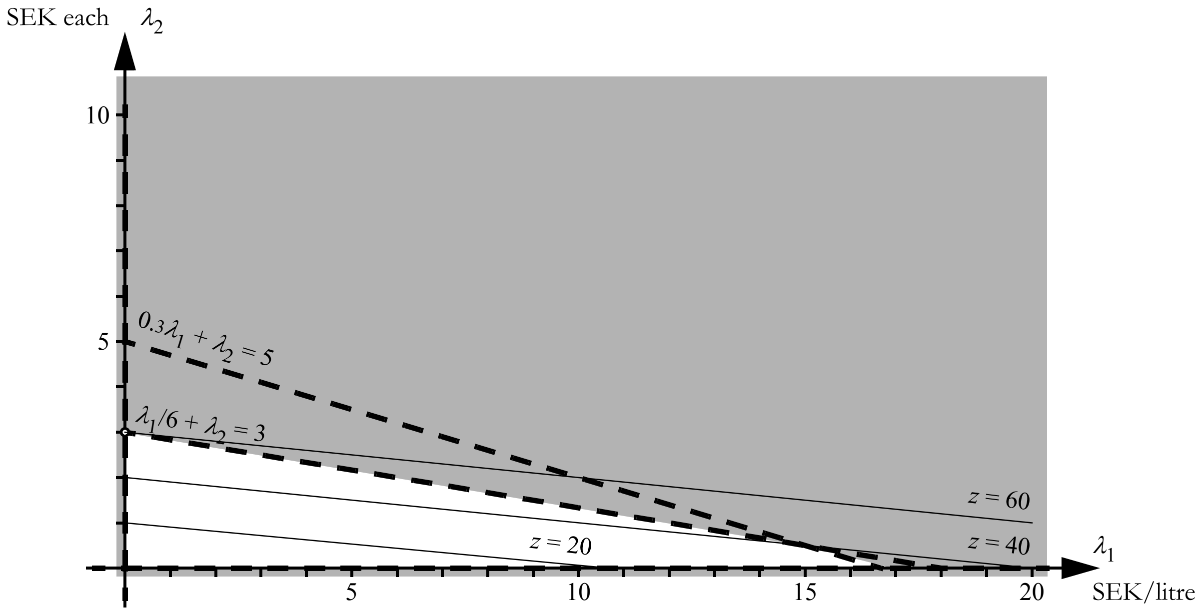


Figure A.10 Solution to the dual problem in example A.10: $\lambda_1 = 0$, $\lambda_2 = 3$, $z = 60$.

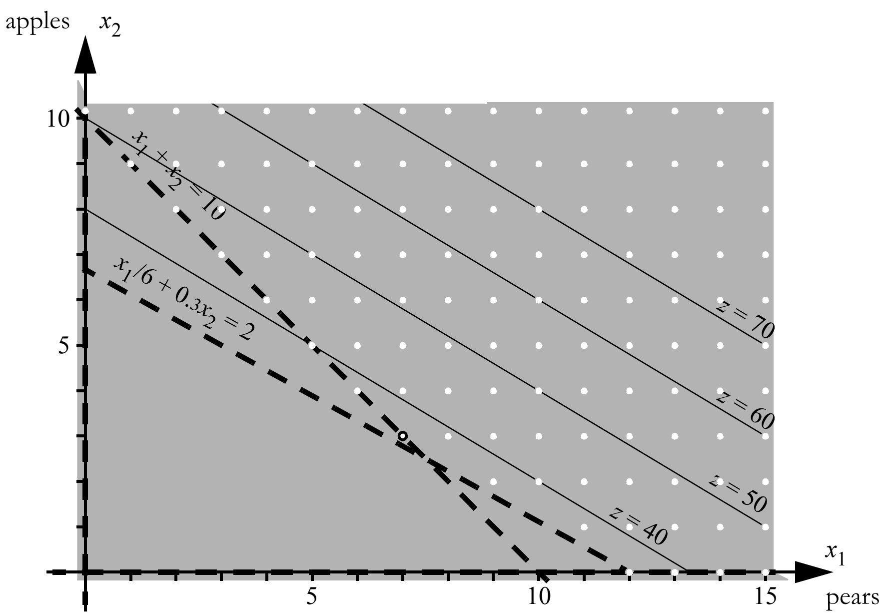


Figure A.11 Solution to example A.11: $x_1 = 7$, $x_2 = 3$, $z = 36$.

same, we can conclude that it will cost Alice 60 SEK to buy at least twenty fruits, i.e., the real cost increase is 25 SEK.

In an ordinary LP problem the variables may have any value within the variable limits. However, it is also possible to introduce further restrictions to the allowed values of the optimisation variables, namely that some variables may only assume integer values. These problems are called MILP problems.[1]

Example A.11 (integer solution): Most shops does not allow the customers to

1. Short for “Mixed Integer Linear Programming”.

buy half pears or apples. How does this affect Alice's purchase?

Solution: The problem formulation is more or less the same, but we have to add a requirement that x_1 and x_2 should be integers. The complete problem is then formulated as

$$\begin{aligned} &\text{minimise} && z = 3x_1 + 5x_2 \\ &\text{subject to} && \frac{1}{6}x_1 + 0.3x_2 \geq 2, \\ &&& x_1 + x_2 \geq 10, \\ &&& x_1 \geq 0, x_2 \geq 0, \\ &&& x_1, x_2 \text{ integers.} \end{aligned}$$

The solution to this problem is shown in figure A.11.

The example above shows that it is not necessarily harder to formulate a MILP problem compared to an ordinary LP problem. However, one should be aware that a MILP problem might be considerably harder to *solve*. In a favourable case, good software might quickly find the optimal solution also to a problem with integer variables, but in the worst case the solution time is increasing exponentially to the number of integer variables. There is also a risk that the optimal solution cannot be found at all. It is therefore appropriate to keep the number of integer as small as possible—if possible, they should be avoided completely!

A special form of integer variables are binary variables (which obviously can only be equal to either zero or one). Binary variables can for example be used to manage piecewise linear functions, though it is not always necessary to use binary variables to represent a piecewise linear function. This is illustrated in the following two examples:

Example A.12 (quantity discount): The shop offers Alice a quantity discount if she buys more than five pears: for the first five pears she pays 5 SEK each and each additional pear costs 3 SEK each. How does this affect Alice's purchase?

Solution: The cost as a function of the amount of pears purchased is shown in figure A.12. Since there are different prices for the first five pears and the remaining, we must dived the number of pears, x_1, in two variables, $x_{1,1}$ and $x_{1,2}$ respectively. We can say that the cost function is divided into two segments. However, a problem is that it is preferable to buy pears from the second segment (as the cost then is 3 SEK per pear). Thus, somehow we must force the purchase in segment two to be zero until the purchase in segment one reaches its maximal value. This can be done by introducing a binary variable:

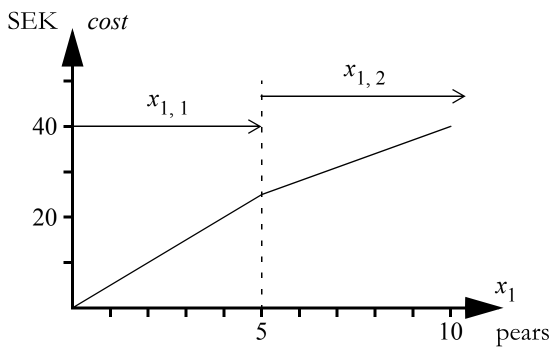


Figure A.12 The cost function of pears in example A.12.

$$\begin{aligned} &\text{minimise} && z = 5x_{1,1} + 3x_{1,2} + 5x_2 \\ &\text{subject to} && \frac{1}{6}x_{1,1} + \frac{1}{6}x_{1,2} + 0.3x_2 \geq 2, \\ &&& x_{1,1} + x_{1,2} + x_2 \geq 10, \end{aligned}$$

$$x_{1,1} - 5s \geq 0,$$

$$-x_{1,2} + M \cdot s \geq 0,$$

$$x_{1,1} \geq 0, x_{1,2} \geq 0, x_2 \geq 0, s \in \{0, 1\}.$$

The objective function now uses the piecewise linear cost function of pears. The first two constraints are more or less the same as in example A.1—the difference is that x_1 is substituted by $x_{1,1} + x_{1,2}$. The two new constraints are used to guarantee that $x_{1,2}$ is zero unless $x_{1,1}$ is equal to five. If the binary variable s is equal to zero then the fourth constraint (where M denotes an arbitrary large number) forces $x_{1,2}$ to be less than or equal to zero. Combined with the variable limit $x_{1,2} \geq 0$ the only possible solution is $x_{1,2} = 0$. On the other hand, if s is equal to one, the fourth constraint states that $x_{1,2}$ should be less than or equal to the arbitrary large number M, which in practice is no limitation. In this case it is the third constraint that is binding, as it forces $x_{1,1}$ to be at least equal to five.

Example A.13 (limited offer): Assume that the shop has a special offer instead of a quantity discount. The special offer is that each customer may buy at most five pears for 3 SEK each and that each additional pear costs the full price, which is 5 SEK each. How does this affect Alice's purchase?

Solution: Alice's cost function in this case is shown in figure A.13. We have to divide x_1 in two segments in this case too, but the difference is that now it is purchase in the first segment that is most profitable. In this case it does not matter if a solution is feasible even though $x_{1,1} < 5$ while $x_{1,2} > 0$, because such a solution will not be optimal. Hence, it is unnecessary (but not incorrect) to introduce a binary variable; everything that is needed is a limitation that $x_{1,1}$ may not exceed five. The most simple way of formulating Alice's problem is then as follows:

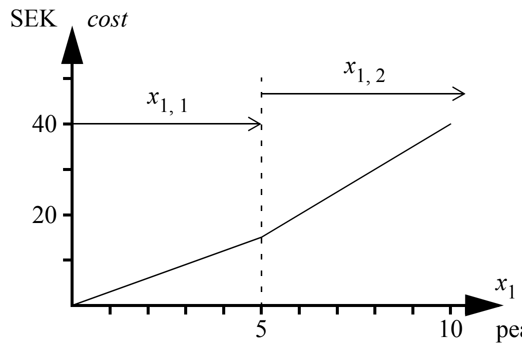


Figure A.13 The cost function of pears in example A.13.

$$\text{minimise} \quad z = 3x_{1,1} + 5x_{1,2} + 5x_2$$

$$\text{subject to} \quad \frac{1}{6}x_{1,1} + \frac{1}{6}x_{1,2} + 0.3x_2 \geq 2,$$

$$x_{1,1} + x_{1,2} + x_2 \geq 10,$$

$$5 \geq x_{1,1} \geq 0, x_{1,2} \geq 0, x_2 \geq 0.$$

FURTHER READING

- M. S. Bazaraa, H. D. Sherali & C.M. Shetty, *Nonlinear Programming: Theory and Algorithms*, second edition, John Wiley & Sons, 1993. — *Complete review of non-linear optimisation. Definitions of many important concepts in optimisation theory.*
- S. P. Bradley, A. C. Hax & T. L. Magnanti, *Applied Mathematical Programming*, Addison-Wesley Publishing Company, 1977. — *Basic textbook with good examples.*

- F. S. Hillier & G. J. Lieberman, *Introductions to Operations Research,* sixth edition, McGraw-Hill, 1995. *— Textbook in operations research with good examples.*
- S. G. Nash & A. Sofer, *Linear and Nonlinear Programming*, McGraw-Hill, 1996. — *Complete, modern textbook.*

Appendix B

SOLVING SHORT-TERM PLANNING PROBLEMS

In this compendium we focus on formulating short-term planning problems, but in practise it is of course also necessary to actually solve them. In a real short-term planning problem there will be hundreds or thousands of optimisation variables and then it is of course impossible to solve be problem by hand calculations. The solution is to use computers and today there is a variety of software to solve both linear programming problems and other kinds of optimisation problems.

The challenge when using computer software to solve optimisation problems is to ascertain that the intended problem has been solved—a small programming error may easily result in an incorrect solution or an infeasible problem. It is therefore important to be careful both when formulating the optimisation problem and then entering it into the computer. In this appendix we will provide an example of how to set about.

In example 5.6 we formulated a short-term planning problem for two hydro power plants. Below we will show how this problem can be solved using different programming languages. However, let us start with a compilation of the original problem formulation:

Indices of power plants

Degerforsen 1, Edensforsen 2.

Parameters

The following parameters are given:

$$\overline{H}_i = \text{installed capacity in power plant } i = \begin{cases} 62 & i = 1, \\ 63 & i = 2, \end{cases}$$

$$\overline{Q}_i = \text{maximal discharge in power plant } i = \begin{cases} 300 & i = 1, \\ 270 & i = 2, \end{cases}$$

$$\overline{M}_i = \text{maximal contents of reservoir } i = \begin{cases} 5\,000\,000/3\,600 & i = 1, \\ 4\,000\,000/3\,600 & i = 2, \end{cases}$$

$$\lambda_f = \text{expected future electricity price} = 185,$$

$$w_i = \text{mean annual flow} = \begin{cases} 163 & i = 1, \\ 164 & i = 2, \end{cases}$$

$$D_t = \text{contracted load for hour } t = \begin{cases} 90 & t = 1, \\ 98 & t = 2, \\ 104 & t = 3, \\ 112 & t = 4, \\ 100 & t = 5, \\ 80 & t = 6. \end{cases}$$

The following parameters are calculated based on the given parameters:

$M_{i,0}$ = start contents of reservoir $i = 0.5\overline{M}_i$, $i = 1, 2$,

$$V_i = \text{local inflow to reservoir } i = \begin{cases} w_1 & i = 1, \\ w_2 - w_1 & i = 2, \end{cases}$$

$\overline{Q}_{i,1}$ = maximal discharge in power plant i, segment 1 = $0.75\overline{Q}_i$, $i = 1, 2$,

$\overline{Q}_{i,2}$ = maximal discharge in power plant i, segment 2 = $\overline{Q}_i - \overline{Q}_{i,1}$, $i = 1, 2$,

$\mu_{i,1}$ = marginal production equivalent of power plant i, segment 1 =

$$= \frac{\overline{H}_i}{\overline{Q}_{i,1} + 0.95\overline{Q}_{i,2}}, \ i = 1, 2,$$

$\mu_{i,2}$ = marginal production equivalent of power plant i, segment 2 = $0.95\mu_{i,1}$.

Optimisation variables

$Q_{i,j,t}$ = discharge in power plant i, segment j, during hour t, $i = 1, 2, j = 1, 2, t = 1, \ldots, 6$,

$S_{i,t}$ = spillage from reservoir i during hour t, $i = 1, 2$, $t = 1, \ldots, 6$,

$M_{i,t}$ = contents of reservoir i at the end of hour t, $i = 1, 2$, $t = 1, \ldots, 6$.

Objective function

$$\text{maximise} \quad \lambda_f((\mu_{1,1} + \mu_{2,1})M_{1,6} + \mu_{2,1}M_{2,6}).$$

Constraints

$$M_{1,t} - M_{1,t-1} + Q_{1,1,t} + Q_{1,2,t} + S_{1,t} = V_1, \qquad t = 1, \ldots 6,$$

$$M_{2,t} - M_{2,t-1} + Q_{2,1,t} + Q_{2,2,t} + S_{2,t} - Q_{1,1,t} - Q_{1,2,t} - S_{1,t} = V_2, \qquad t = 1, \ldots 6,$$

$$\sum_{i=1}^{2} \sum_{j=1}^{2} \mu_{i,j} Q_{i,j,t} = D_t, \qquad t = 1, \ldots 6.$$

Variable limits

$$0 \le Q_{i,j,t} \le \overline{Q}_{i,j}, \qquad i = 1, 2, j = 1, 2, t = 1, \ldots, 6,$$

$$0 \le S_{i,t}, \qquad i = 1, 2, t = 1, \ldots, 6,$$

$$0 \le M_{i,t} \le \overline{M}_i, \qquad i = 1, 2, t = 1, \ldots, 6.$$

B.1 GAMS

GAMS is a programming language especially dedicated to formulation of optimisation problems. The idea is that problems and inputs are formulated in a standardised way—regardless of whether the problem at hand is linear or non-linear—and then the problem is passed on to special solvers. A GAMS program is written in an ordinary text file (it is most straightforward to use the built-in editor). When executing the program (choose Run in the menu File) GAMS will process the file and check that the problem syntax is correct. If that is the case then the problem is passed to a suitable solver. The solution (or an error message if no solution was found) is then returned to GAMS, where it is possible to choose how to present the results. Finally, the result is written to a text file having the same name as the program file, but with the suffix .lst. This listing file can be viewed in the built-in editor. Any errors encountered during the run results in error messages in the listing file.

We start by defining the indices to be used in the program. GAMS requires that all possible index values are states, in order to allow the program to check that other statements only use allowed values. The syntax for defining index values is

```
set symbol ”explanatory text” /element1, element2/;
```

The keyword `set` (we may also use the plural form `sets`) states that an index is going to be defined, and is followed by the symbol used to denote the index. Then an explanatory text follows, which GAMS will echo when listing the results. The explanatory text may not exceed 80 characters. If no slashes, commas or semi-colons are used in the text, there is no need for the quotation marks. Finally, the possible index values are listed within the slashes. It is allowed to define more than one index in the same `set` statement; the most straightforward is then to define each index in a row of its own.

Indices of power plants

Degerforsen 1, Edensforsen 2.

```
Sets
  i power plant
    /Degerforsen, Edensforsen/
  j segment /segment1*segment2/
  t hour /hour1*hour6/
;
```

Notice that we may define a sequence of numbered index values using `*`; the allowed values of the index `t` are thus `"hour1"`, `"hour2"`, `"hour3"`, `"hour4"`, `"hour5"` and `"hour6"`.

The next step is to declare the parameters of the problem. The syntax for declaring parameters is similar to the one for index values:

Parameters

The following parameters are given:

$\overline{H}_i$ = installed capacity in power plant i =

$$= \begin{cases} 62 & i = 1, \\ 63 & i = 2, \end{cases}$$

$\overline{Q}_i$ = maximal discharge in power plant i =

$$= \begin{cases} 300 & i = 1, \\ 270 & i = 2, \end{cases}$$

```
Parameters
Hmax(i)   installed capacity in
          power plant i
          /Degerforsen 62,
           Edensforsen 63/
Qtotmax(i)maximal discharge in
          power plant i
          /Degerforsen 300,
           Edensforsen 270/
```

$\overline{M}_i$ = maximal contents of reservoir i =

$$= \begin{cases} 5\,000\,000/3\,600 & i = 1, \\ 4\,000\,000/3\,600 & i = 2, \end{cases}$$

λ_f = expected future electricity price = = 185,

w_i = mean annual flow =

$$= \begin{cases} 163 & i = 1, \\ 164 & i = 2, \end{cases}$$

D_t = contracted load for hour t =

$$= \begin{cases} 90 & t = 1, \\ 98 & t = 2, \\ 104 & t = 3, \\ 112 & t = 4, \\ 100 & t = 5, \\ 80 & t = 6. \end{cases}$$

```
Mmax(i)   maximal contents of
          reservoir i
          /Degerforsen 5e6,
           Edensforsen 4e6/
lambdaf   expected future
          electricity price /185/
w(i)      mean annual flow at
          power plant i
          /Degerforsen 163,
           Edensforsen 164/
D(t)      contracted load hour t
          /hour1 90, hour2 98,
           hour3 104, hour4 112,
           hour5 100, hour6 80/
Mstart(i) start contents of
          reservoir i
V(i)      local inflow to
          reservoir i
Qmax(i,j) "maximal discharge in
          power plant i,
          segment j"
mu(i,j)   "marginal production
          equivalent for power
          plant i, segment j"
;
```

When a parameter has more than one index, a dot is used to separate different index values. The maximal discharge in each power plant and segment could thus be stated in the following manner:[1]

```
Parameters
Qmax(i,j) "maximal discharge in power plant i, segment j"
          /Degerforsen.segment1 225,
           Edensforsen.segment1 202.5,
           Degerforsen.segment2 75,
           Edensforsen.segment2 67.5/
;
```

However, in order to avoid rounding errors it might be better to let GAMS perform the calculations. In that case, it is sufficient to declare that the parameter exists and then assign values later:

The following parameters are calculated based on the given parameters:

$M_{i,0}$ = start contents of reservoir $i = 0.5\overline{M}_i$, $i = 1, 2$,

V_i = local inflow to reservoir i =

$$= \begin{cases} w_1 & i = 1, \\ w_2 - w_1 & i = 2, \end{cases}$$

```
Mmax(i) = Mmax(i)/3600;
Mstart(i) = 0.5*Mmax(i);
V(i) = w(i) - w(i-1);
Qmax(i,"segment1") =
  0.75*Qtotmax(i);
Qmax(i,"segment2") = Qtotmax(i) -
  Qmax(i,"segment1");
mu(i,"segment1") =
  Hmax(i)/(Qmax(i,"segment1") +
  0.95*Qmax(i,"segment2"));
mu(i,"segment2") =
  0.95*mu(i,"segment1");
```

1. An alternative approach would be to use the key word `Table`. See *GAMS Users Guide* (which can be accessed from the help menu in the GAMS window) for further details.

$\overline{Q}_{i,1}$ = maximal discharge in power plant i, segment 1 = $0.75\overline{Q}_i$, $i = 1, 2$,

$\overline{Q}_{i,2}$ = maximal discharge in power plant i, segment 2 = $\overline{Q}_i - \overline{Q}_{i,1}$, $i = 1, 2$,

$\mu_{i,1}$ = marginal production equivalent of power plant i, segment 1 =

$$= \frac{\overline{H}_i}{\overline{Q}_{i,1} + 0.95\overline{Q}_{i,2}},\ i = 1, 2,$$

$\mu_{i,2}$ = marginal production equivalent of power plant i, segment 2 = $0.95\mu_{i,1}$.

Notice that GAMS automatically performs assignments over all possible values of the indices involved in the expression. Thus, the first row in the code above will be interpreted by GAMS as

```
Mmax("Degerforsen") = Mmax("Degerforsen")/3600;
```

and

```
Mmax("Edensforsen") = Mmax("Edensforsen")/3600;
```

Another feature of GAMS is that expressions which contain non-existing index values will automatically be removed from the calculations. When the index `i` is equal to `"Edensforsen"` the row

```
V(i) = w(i) - w(i-1);
```

is read as

```
V("Edensforsen") = w("Edensforsen") - w("Degerforsen");
```

i.e., `i - 1` is automatically replaced by the previous index value. But when `i` is equal to `"Degerforsen"` there is no previous index value (as `"Degerforsen"` is the first power plant in the definition of the set); hence this part is removed and then the row reads

```
V("Degerforsen") = w("Degerforsen");
```

We must also declare which optimisation variables we have in the problem. Notice that we already now can state some variable limits by using the keywords `positive`, `negative`, `binary`, `integer` and `free`. GAMS requires that there is at least one free variable in each problem, as the variable to be maximised or minimised must be able to assume any value. The declarations are similar to the index values and parameters, i.e., a symbol is stated followed by an explanatory text (however, no values are stated):

Optimisation variables

$Q_{i,j,t}$ = discharge in power plant i, segment j, during hour t, $i = 1, 2$, $j = 1, 2$, $t = 1, \ldots, 6$,

$S_{i,t}$ = spillage from reservoir i during hour t, $i = 1, 2$, $t = 1, \ldots, 6$,

$M_{i,t}$ = contents of reservoir i at the end of hour t, $i = 1, 2$, $t = 1, \ldots, 6$.

```
Positive variables
  Q(i,j,t)   "discharge in power
             plant i, segment j,
             during hour t"
  S(i,t)     spillage from
             reservoir i
             during hour t
  M(i,t)     contents of reservoir
             i at the end of
             hour t
```

Table B.1 GAMS equation types.

Type of equations	Mathematical symbol	Sense
=e=	=	Left hand side must equal right hand side
=l=	≤	Left hand side must be less than or equal to the right hand side
=g=	≥	Left hand side must be greater than or equal to right hand side

```
;

Free variable
  z         value of stored water
;
```

The relation between the optimisation variables are described by equations in GAMS. An equation must first be declared; this is done in a similar manner as for variables. Then, the actual equation is defined according to the following syntax:

```
symbol .. LHS type RHS;
```

Here, `symbol` is the name the equation was given when declared. The name of the equation is separated from the equation definition by two dots. Both the left hand side (`LHS`) and right hand side (`RHS`) are mathematical expressions. The relation between the left and right hand sides are depending on the `type` of the equation (see table B.1). There is no difference in syntax between the objective function and the constraints; the objective function is simply the equation stating the value of the free variable to be minimised or maximised:

Objective function

maximise $\lambda_f((\mu_{1,1} + \mu_{2,1})M_{1,6} + \mu_{2,1}M_{2,6})$.

```
Equation
  objfnc  objective function;

objfnc..  z =e= lambdaf*
            ((mu("Degerforsen",
                 "segment1") +
              mu("Edensforsen",
                 "segment1"))
            *M("Degerforsen",
               "hour6")
          + mu("Edensforsen",
               "segment1")
            *M("Edensforsen",
               "hour6"));
```

The constraints are also stated as equations. Once again we can utilise that GAMS removes expressions if the index value does not exist. However, in the hydrological constraints there is a complication, which cannot be managed by GAMS itself. In the first hour there is no optimisation variable $M_{i,\,t-1}$, but there is a parameter $M_{i,\,0}$, which in the program has been given the symbol `Mstart(i)`. This parameter should only be included in those hydrological constraints for which `t` is equal to `"hour1"`. This is implemented by using a so-called dollar condition, which means that after an expression we may type a `$`-sign, followed by a condition. If this condition is not fulfilled the expression will be removed, in the same way as non-existing index values are removed. The function `ord` can be used to obtain the position of an index value in a set. As `"hour1"` is the first value in the set `t`, the condition `ord(t) = 1` will only be fulfilled for the hydrological balances during the first hour.

Table B.2 GAMS default variable limits.

Variable type	Lower limit (`.LO`)	Upper limit (`.UP`)	Other limitations
`free`	$-\infty$	$+\infty$	
`positive`	0	$+\infty$	
`negative`	$-\infty$	0	
`binary`	0	1	Integer variable
`integer`	0	100	Integer variable

Constraints

$$M_{1,t} - M_{1,t-1} + Q_{1,1,t} + Q_{1,2,t} + S_{1,t} = V_1, \quad t = 1, \dots 6,$$

$$M_{2,t} - M_{2,t-1} + Q_{2,1,t} + Q_{2,2,t} + S_{2,t} - Q_{1,1,t} - Q_{1,2,t} - S_{1,t} = V_2, \quad t = 1, \dots 6,$$

$$\sum_{i=1}^{2} \sum_{j=1}^{2} \mu_{i,j} Q_{i,j,t} = D_t, \quad t = 1, \dots 6.$$

```
Equations
  hydbal(i,t)  "hydrological
               balance in power
               plant i, hour t"
  loadbal(t)   load balance of
               hour t
;

hydbal(i,t).. M(i,t) =e= M(i,t-1)
  + Mstart(i)$(ord(t)
                         = 1)
  - sum(j,Q(i,j,t))
  - S(i,t)
  + sum(j,Q(i-1,j,t))
  + S(i-1,t)
  + V(i);
loadbal(t)..  sum((i,j),mu(i,j)*
                      Q(i,j,t))
              =e= D(t);
```

All optimisation variables in GAMS are given variable limits according to their type (see table B.2). It is simple to change the default values in those cases when a variable has another limit. The upper limit of a variable is accessed by the symbol of the variable followed by the suffix `.UP` and similarly, the lower limit has the suffix `.LO`. In this problem all variables have the lower limit zero and therefore do not need to be changed. The spillage has no upper limit; hence, the spillage uses the default upper value too. The only thing that needs to be done is to set the upper limits of discharge and reservoir contents:

Variable limits

$$0 \le Q_{i,j,t} \le \overline{Q}_{i,j}, \quad i = 1, 2, j = 1, 2, t = 1, \dots, 6,$$

$$0 \le S_{i,t}, \quad i = 1, 2, t = 1, \dots, 6,$$

$$0 \le M_{i,t} \le \overline{M}_i, \quad i = 1, 2, t = 1, \dots, 6.$$

```
loop(t,Q.up(i,j,t) = Qmax(i,j);
       M.up(i,t) = Mmax(i));
```

We have already seen that assignments automatically are performed for all possible values of an index. However, in this case we have more indices in the parameter that is to be assigned than in the parameter from which the values are taken. Therefore, we use a `loop`-statement to perform the same assignment for every hour. Thus, as a result of the loop, GAMS is performing the assignments

```
Q.UP(i,j,"hour1") = Qmax(i,j);
```

```
M.UP(i,"hour1") = Mmax(i);
Q.UP(i,j,"hour2") = Qmax(i,j);
M.UP(i,"hour2") = Mmax(i);
```

etc.

We can now solve the problem. First we give a name to the model, while stating which equations are included in the model (in this case it is all equations, which is stated by `/all/`). Then we ask GAMS to solve the optimisation problem. First, we must specify which type of solver that is to be used. In this case we have an LP problem; therefore, we state that the problem should be solved "`using lp`". It we would like to solve a MILP problem, we would have stated "`using mip`", and if we would like to solve a non-linear problem, we would state "`using nlp`". GAMS will automatically select the solver which is most appropriate for problems of the specified type. Finally, we must state which variable is to be minimised or maximised. In this case it is `z` that shall be maximised. Remember that the variable to be maximised or minimised must be declared as a free variable.

```
model hydroplanning /all/;

solve hydroplanning using lp
  maximizing z;
```

Finally we instruct GAMS to write the most important results at the end of the listing file. At the same time, we can also calculate the total discharge and electricity generation in each power plant and hour (cf. the discharge plan in table 5.5). To calculate the total discharge and electricity generation we need the optimal values of the discharge $V_{i,j,t}$. When GAMS is solving the optimisation problem the optimal solution is stored in parameters having the same name as the optimisation variable followed by the suffix `.L`. Hence, the optimal discharge is accessed through `Q.L`. We can also obtain the dual variables of both constraints and variable limits. The dual variables are stored in parameters which have the same name as the constraint or the optimisation variable followed by the suffix `.M`.

```
Parameters
  Qtot(i,t) "total discharge in
            power plant i,
            hour t"
  H(i,t)    "generation in power
            plant i, hour t"
;

loop((i,t),Qtot(i,t) =
            sum(j,Q.L(i,j,t));
           H(i,t) = sum(j,
            mu(i,j)*Q.L(i,j,t)));

display M.L, Q.L, Qtot, H, S.L;
display hydbal.M, loadbal.M;
```

B.2 MATLAB

Matlab is a general programming language for technical computing. There is a large range of ready-made functions for many areas of computing. With the exception of the most basic mathematical functions—which are included in the software—all functions are written in ordinary text files (it is most straightforward to use the built-in editor).

Functions to solve optimisation problems are found in the so-called Matlab Optimization Toolbox, but there are also freeware functions to be found on the Internet. Below it is demonstrated how our problem can be solved using `linprog` (which is included in Optimization Toolbox). As most Matlab functions for solving LP problems, `linprog` requires that the problem is formulated using matrices. In this case the problem should be on the form

$$\text{minimise} \quad f^T x \tag{B.1}$$

$$\text{subject to} \quad Ax \le b, \tag{B.1a}$$

$$A_{eq} x = b_{eq}, \tag{B.1b}$$

$$l_b \le x \le u_b. \tag{B.1c}$$

The function `linprog` can be called with different inputs and outputs, depending on how we wish to use the function.[2] Below follows a description of what is needed to solve the example in this appendix.

Assume that a problem has N_{ineq} inequality constraints, N_{eq} equality constraints and M variables. In that case `linprog` uses the following inputs:

- `f` coefficients of the objective function (column vector with M elements),
- `A` coefficients of the inequality constraints (matrix with N_{ineq} rows and M columns),
- `b` right hand side of the inequality constraints (column vector with N_{ineq} elements),
- `Aeq` coefficients of the equality constraints (matrix with N_{eq} rows and M columns),
- `beq` right hand side of the equality constraints (column vector with N_{eq} elements),
- `lb` lower bound of the optimisation variables (column vector with M elements),
- `ub` upper bound of the optimisation variables (column vector with M elements).

The result of a call is

- `x` optimal solution (column vector with M elements).

The difficulty when solving the planning problem is to correctly define the inputs necessary to call `linprog`. To make sure that no mistakes are made in this process, it is appropriate to write a Matlab program which follows the original problem formulation as closely as possible. Hence, we start by clarifying the index values we use for the power plants:

Indices of power plants

Degerforsen 1, Edensforsen 2.

```
% Indices of the power plants:
% Degerforsen 1, Edensforsen 2
```

In Matlab everything on a row after a `%`-sign is considered a comment. Hence, no calculations are performed by the two rows above—they are just intended to make the program easier to understand.

The next step is to define all parameter values:

2. Please refer to the built-in help function of Matlab for a more detailed description.

Parameters

The following parameters are given:

$\overline{H}_i$ = installed capacity in power plant i =

$$= \begin{cases} 62 & i = 1, \\ 63 & i = 2, \end{cases}$$

$\overline{Q}_i$ = maximal discharge in power plant i =

$$= \begin{cases} 300 & i = 1, \\ 270 & i = 2, \end{cases}$$

$\overline{M}_i$ = maximal contents of reservoir i =

$$= \begin{cases} 5\,000\,000/3\,600 & i = 1, \\ 4\,000\,000/3\,600 & i = 2, \end{cases}$$

λ_f = expected future electricity price = = 185,

w_i = mean annual flow =

$$= \begin{cases} 163 & i = 1, \\ 164 & i = 2, \end{cases}$$

D_t = contracted load for hour t =

$$= \begin{cases} 90 & t = 1, \\ 98 & t = 2, \\ 104 & t = 3, \\ 112 & t = 4, \\ 100 & t = 5, \\ 80 & t = 6. \end{cases}$$

```
Given parameters
npowerplants = 2;
nhours = 6;
Hmax = [62; 63]; % Hmax(i)
Qtotmax = [300; 270];
  % Qtotmax(i)
Mmax = [5e6; 4e6]/3600; % Mmax(i)
lambdaf = 185;
w = [163; 164]; % w(i)
D = [90 98 104 112 100 80];
    % D(t)
```

The following parameters are calculated based on the given parameters:

$M_{i,0}$ = start contents of reservoir $i = 0.5\overline{M}_i$, $i = 1, 2$,

V_i = local inflow to reservoir i =

$$= \begin{cases} w_1 & i = 1, \\ w_2 - w_1 & i = 2, \end{cases}$$

$\overline{Q}_{i,1}$ = maximal discharge in power plant i, segment 1 = $0.75\overline{Q}_i$, $i = 1, 2$,

```
% Calculated parameters
Mstart = 0.5*Mmax; % Mstart(i)
V = [w(1); w(2)-w(1)]; % V(i)
% Qmax(i,j)
Qmax(:,2) = Qtotmax - Qmax(:,1);
Qmax(:,1) = .75*Qtotmax;
% mu(i,j)
mu(:,1) = Hmax./(Qmax(:,1) +
  0.95*Qmax(:,2));
mu(:,2) = 0.95*mu(:,1);
```

$\overline{Q}_{i,2}$ = maximal discharge in power plant i, segment 2 = $\overline{Q}_i - \overline{Q}_{i,1}$, $i = 1, 2$,

$\mu_{i,1}$ = marginal production equivalent of power plant i, segment 1 =

$$= \frac{\overline{H}_i}{\overline{Q}_{i,1} + 0.95\overline{Q}_{i,2}}, \; i = 1, 2,$$

$\mu_{i,2}$ = marginal production equivalent of power plant i, segment 2 = $0.95\mu_{i,1}$.

Notice that we try to choose the same parameter name in Matlab as when we originally formulated the problem. It is also recommended to insert a comment explaining how the Matlab matrices are indexed. For example, we have above chosen to put all marginal production equivalents in a matrix `mu`, where the element on row `i`, column `j`, correspond to the marginal production equivalent of power plant i, segment j, i.e., $\mu_{i,j}$.

When an optimisation problem is formulated in matrix form, all optimisation variables are collected in a column vector x, i.e., each element in x should correspond to a unique optimisation variable in the original problem formulation, as in the following example:

$$x = [Q_{1,1,1}\; Q_{1,2,1}\; S_{1,1}\; M_{1,1}\; Q_{2,1,1}\; Q_{2,2,1}\; S_{2,1}\; M_{2,1}\; \ldots\; M_{3,6}]^T.$$

It is arbitrary in which order the variables area sorted into the vector x, but it is of course necessary to be consequent throughout the Matlab code. It is simplifying to use a “dictionary”, where it is easy to identify which element in x corresponds to a certain variable. To implement this, we use the following code:

Optimisation variables

$Q_{i,j,t}$ = discharge in power plant i, segment j, during hour t, $i = 1, 2$, $j = 1, 2$, $t = 1, \ldots, 6$,

$S_{i,t}$ = spillage from reservoir i during hour t, $i = 1, 2$, $t = 1, \ldots, 6$,

$M_{i,t}$ = contents of reservoir i at the end of hour t, $i = 1, 2$, $t = 1, \ldots, 6$.

```
% Variables
pos = 0;
for t = 1:nhours
  for i = 1:npowerplants
    pos = pos + 1;
    Qpos(i,1,t) = pos;
    pos = pos + 1;
    Qpos(i,2,t) = pos;
    pos = pos + 1;
    Spos(i,t) = pos;
    pos = pos + 1;
    Mpos(i,t) = pos;
  end
end
nvariables = pos;
```

In the code above the variable `pos` is just a counter, which tracks the last used position in i x. When the loops of the program has been performed, each variable in the problem will have been assigned a position and these positions have been stored in the three matrices `Vpos`, `Spos` and `Mpos`. These matrices are indexed in the same manner as the corresponding optimisation variable. Hence, if we want to know where in the vector x we will find $M_{2,3,4}$we should check the value of `Mpos(2,3,4)`.

Now we can define the objective function. The objective function in (B.1) is read

$$f^T x = f_1 x_1 + f_2 x_2 + \ldots + f_M x_M. \tag{B.2}$$

When calling `linprog` we just need to state the coefficients, f. If we look at the problem formulation we find that only two optimisation variables are included in the objective function, which means that most of the coefficients in f are equal to zero. To shorten the Matlab code we can start by a vector of the right size, which just hold zeros (such a vector is generated by the function `zeros`). Then we change just those coefficients that are non-zero. For example, the coefficient before $M_{2,6}$ is equal to $\mu_{2,1}$ and if $M_{2,6}$ is the n:th element of the vector x then the n:th element of f should equal $\mu_{2,1}$. To know the position of $M_{2,6}$ we use the earlier defined "dictionary" `Mpos`:

Objective function

maximise $\lambda_f((\mu_{1,1} + \mu_{2,1})M_{1,6} + \mu_{2,1}M_{2,6})$.

```
% Objective function:
f = zeros(nvariables,1);
f(Mpos(1,nhours)) = ...
  lambdaf*(mu(1,1) + mu(2,1));
f(Mpos(2,nhours)) = ...
  lambdaf*mu(2,1);
```

The constraints are built in a similar manner as the objective function. Each constraint corresponds to a certain row in the matrices A and A_{eq} respectively, and a certain row in the column vectors b and b_{eq} respectively. We use a counter `cnstr` to keep track of which constraint we currently define; each time we start stating the coefficients of a new constraint we first increase the value of `cnstr` by one.

Notice that in the hydrological constraints we are forced to use an `if`-clause to manage different cases. The first hour (i.e., then `t = 1`) the reservoir contents at the end of the previous hour is a known parameter, which should be added to the right hand side of the constraint, whereas $M_{i,t-1}$ is an optimisation variable for the other hours. Moreover, in the hydrological balance of Degerforsen (which we defined as power plant 1) there is no incoming water from discharge and spillage in the power plant upstream; hence, these coefficients should only be included if `i > 1`.

Constraints

$$M_{1,t} - M_{1,t-1} + Q_{1,1,t} + Q_{1,2,t} + S_{1,t} = V_1, \quad t = 1, \dots 6,$$

$$M_{2,t} - M_{2,t-1} + Q_{2,1,t} + Q_{2,2,t} + S_{2,t} - Q_{1,1,t} - Q_{1,2,t} - S_{1,t} = V_2, \quad t = 1, \dots 6,$$

$$\sum_{i=1}^{2}\sum_{j=1}^{2} \mu_{i,j} Q_{i,j,t} = D_t, \qquad t = 1, \dots 6.$$

```
% Constraints
nconstraints = 3*nhours;
Aeq = zeros(nconstraints,...
            nvariables);
beq = zeros(nconstraints,1);
cnstr = 0;
for t = 1:nhours
  for i = 1:npowerplants
    contrnr = cnstr + 1;
    Aeq(cnstr,Mpos(i,t)) = 1;
    if k == 1
      beq(cnstr) = beq(cnstr) ...
                   + Mstart(i);
    else
      Aeq(cnstr,Mpos(i,t-1)) ...
        = -1;
    end
    Aeq(cnstr,Qpos(i,1,t)) = 1;
    Aeq(cnstr,Qpos(i,2,t)) = 1;
    Aeq(cnstr,Spos(i,t)) = 1;
    if i > 1
      Aeq(cnstr,Qpos(i-1,1,t))...
        = -1;
```

```
          Aeq(cnstr,Qpos(i-1,2,t))...
            = -1;
          Aeq(cnstr,Spos(i-1,t))...
            = -1;
        end
        beq(cnstr) = beq(cnstr) + ...
                     V(i);
      end
      cnstr = cnstr + 1;
      for i = 1:npowerplants
        for j = 1:2
          Aeq(cnstr,Qpos(i,j,t)) =
            ...
            mu(i,j);
        end
      end
      beq(cnstr) = D(t);
    end
```

Finally we have to define the upper and lower limits on all optimisation variables. The lower limits are all zeros; therefore, the vector `lb` is easily defined directly by the function `zeros`. The upper limits are different for different variables. The most common limit is +∞ and we may therefore start from a vector of the correct size, where all elements are equal to +∞ (such a vector can be generated by the function `inf`) and then we change the upper limits that are not infinite:

Variable limits

$0 \le Q_{i,j,t} \le \overline{Q}_{i,j}$, $i = 1, 2, j = 1, 2, t = 1, \ldots, 6,$

$0 \le S_{i,t}$, $i = 1, 2, t = 1, \ldots, 6,$

$0 \le M_{i,t} \le \overline{M}_i$, $i = 1, 2, t = 1, \ldots, 6.$

```
% Variable limits:
lb = zeros(nvariables,1);
ub = inf(nvariables,1);
for t = 1:nhours
  for i = 1:npowerplants
    for j = 1:2
      ub(Qpos(i,j,t)) = ...
         Vmax(i,j);
    end
    ub(Mpos(i,t)) = Mmax(i);
  end
end
```

Now that all inputs have been defined (we have no inequality constraints, so there is no need to define `A` and `b`; they are replaced by empty matrices in the call to `linprog`) and are ready to solve the optimisation problem. As `linprog` solves a minimisation problem and we have formulated a maximisation problem we must remember to negate the objective function! The function `linprog` can in addition to the optimal solution also return additional information about the problem. For example is the value of the dual variables of both constraints and variable limits calculated. The dual variables are the fifth value returned by `linprog`. Hence, we are only interested of the first and fifth value, and the values in between are stored in the waste parameters `waste1`, `waste2` and `waste3`:

```
% Solve the optimisation problem:
[x,waste1,waste2,waste3,y] =
  linprog(-f,[],[],...
       Aeq,beq,lb,ub);
```

The function returns the optimal solution stored in the vector `x`. To get a better overview of the optimal solution we can extract the individual optimisation variables and put them in separate matrices for each kind of variable (discharge, spillage and reservoir contents). At the same time, we can also calculate the total discharge and electricity generation in each power plant and hour (cf. the discharge plan in table 5.5). The dual variables are returned in a Matlab structure. A dot followed by the field name in question is used to access a field in the structure., The structure storing the dual variables has the fields `.ineqlin` (dual variables of the inequality constraints), `.eqlin` (dual variables of the equality constraints), `.upper` (dual variables of the upper variable limits) and `.lower` (dual variables of the lower variable limits).

```
% Sort the results:
Qres = [];
Sres = [];
Mres = [];
for t = 1:nhours
  for i = 1:npowerplants
    for j = 1:2
      Qres(i,j,t) = ...
            x(Vpos(i,j,t));
    end
    Sres(i,t) = x(Spos(i,t));
    Mres(i,t) = x(Mpos(i,t));
    Qtot(i,t) = sum(Qres(i,:,t));
    H(i,t) = ...
      mu(i,1)*Qres(i,1,t) + ...
      mu(i,2)*Qres(i,2,t);
  end
end
optvalue = round(f'*x)
Mres, Qres, Qtot, H, Sres
dualvariables = y.eqlin
```

Appendix C

RANDOM VARIABLES

In this appendix some brief explanations are given about the most basic concepts of random variables.

Probability Distributions

The definitions below are generally given for a continuous random variable X. If X is a discrete random variable then all integrals should be replaced by sums.

> ***Definition C.1.*** The probability that an observation of X belongs to a given set $\mathcal{X}$ is given by the density function $f_X(x)$, i.e.,
>
> $$P(X \in \mathcal{X}) = \int_{\mathcal{X}} f(x)dx.$$

It should be noted that for discrete random variables $f_X(x)$ is often referred to as the probability function (or frequency function), because in this case $f_X(x)$ is actually equal to $P(X = x)$. Except for this difference in the interpretation of $f_X(x)$ density functions and probability functions have the same properties.

> ***Example C.1 (die).*** State the frequency function for the result of throwing a fair die.
>
> ***Solution:*** The probability is the same to get the results 1, 2, 3, 4, 5 and 6 respectively. This means that the result of throwing the die can be described by a random variable

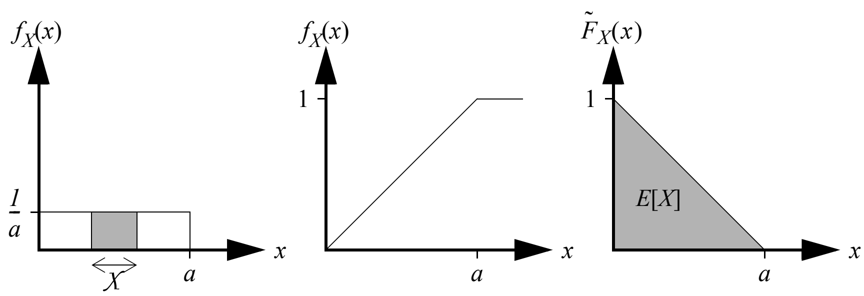


Figure C.1 Example of the density function, the distribution function and the duration curve of a random variable. The variable X is uniformly distributed on the interval $[0, a]$.

X, where the probabilities of each of the six possible outcomes are equal, i.e.,

$$f_X(x) = \begin{cases} \dfrac{1}{6} & x = 1, 2, 3, 4, 5, 6 \\ 0 & \text{all other } x. \end{cases}$$

Definition C.2. The probability that $X \leq x$ is given by the distribution function $F_X(x)$, i.e., $F_X(x) = P(X \leq x)$.

Definition C.3. The probability that $X > x$ is given by the duration curve $\tilde{F}_X(x)$, i.e., $\tilde{F}_X(x) = P(X > x)$.

Theorem C.4. The density function, the distribution function and the duration curve are related through

$$\tilde{F}_X(x) = 1 - F_X(x) = 1 - \int_{-\infty}^{x} f_X(t)dt = \int_{x}^{\infty} f_X(t)dt.$$

Proof: The first two equalities follow directly from definitions C.1-C.3. The last equality is obtained by observing that the probability of $X \in \boldsymbol{R}$ is exactly 100%, which can be written as

$$\int_{-\infty}^{\infty} f_X(t)dt = 1 \Leftrightarrow \int_{-\infty}^{x} f_X(t)dt + \int_{x}^{\infty} f_X(t)dt = 1. \blacksquare$$

Example C.2. What is the probability that the result of throwing a die is not larger than 3?

Solution: The probability that $X \leq 3$ is given by

$$F_X(3) = \sum_{i=1}^{3} f_X(i) = \frac{3}{6} = 0.5.$$

When adding two random variables the result will be a new random variable having a probability distribution of its own. The following applies to independent variables:[1]

Theorem C.5. The density function of the sum, Z, of two independent random variables X and Y can be obtained by a convolution formula. For independent discrete random variables the convolution formula is written as

$$f_Z(x) = \sum_{i} f_X(i) f_Y(x - i)$$

and for independent continuous random variables it reads

$$f_Z(x) = \int_{-\infty}^{\infty} f_X(t) f_Y(x - t)dt.$$

Convolution can also be applied to distribution functions and duration curves.

Example C.3. Assume that you have two fair dice and that X is the outcome of one die and Y is the outcome of the other die. What is the probability that the sum of both outcomes is equal to 9?

Solution: Both dice have the same frequency function as in example C.1, i.e.,

1. See the section *Correlations* on page 188.

$f_Y(x) = f_X(x)$. Let $Z = X + Y$. The probability that $Z = 9$ is given by

$$f_Z(9) = \sum_{i=3}^{6} f_X(i) f_Y(9-i) = f_X(3) \cdot f_Y(6) + f_X(4) \cdot f_Y(5) + f_X(5) \cdot f_Y(4) + f_X(6) \cdot f_Y(3) =$$

$$= \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} + \frac{1}{6} \cdot \frac{1}{6} = \frac{4}{36}.$$

Expectation Value and Variance

The density and distribution functions defined above provide detailed information about a random variable. In many cases it is of larger practical importance to just know the expected average and a measure of how much a single sample can diverge from the average. This information is given by the expectation value and the variance:

Definition C.6. The expectation value of a random variable X is given by

$$E[X] = \int_{-\infty}^{\infty} x f_X(x) dx.$$

Definition C.7. The variance of a random variable X is equal to the expected square of the deviation from the average, i.e.,

$$Var[X] = E[(X - E[X])^2] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx.$$

Another common measure is the standard deviation, which simply is the square root of the variance.

Example C.4. Calculate the expectation value, variance and standard deviation of a fair die.

Solution:

$$E[X] = \sum_{x=1}^{6} x f_X(x) = \frac{1}{6} \sum_{x=1}^{6} x = \frac{1+2+3+4+5+6}{6} = \frac{21}{6} = 3.5.$$

$$Var[X] = \sum_{x=1}^{6} (x - E[X])^2 f_X(x) = \sum_{x=1}^{6} x^2 f_X(x) - (E[X])^2 = \frac{1}{6} \sum_{x=1}^{6} x^2 - 3{,}5 =$$

$$= \frac{1+4+9+16+25+36}{6} - \left(\frac{21}{6}\right)^2 = \frac{35}{12} \approx 2.92.$$

$$\sigma_X = \sqrt{Var[X]} = \sqrt{\frac{35}{12}} \approx 1.71.$$

The following two theorems show how the expectation value and variance are affected by linear operations:

Theorem C.8. If a random variable X is multiplied by a scalar a then the expectation value and variance of aX is given by

$$E[aX] = aE[X],$$

$$Var[aX] = a^2 \, Var[X].$$

Theorem C.9. The expectation value and variance of the sum of two random variables X and Y are given by

$$E[X + Y] = E[X] + E[Y],$$

$$Var[X + Y] = Var[X] + Var[Y] + 2Cov[X, Y].$$

In this compendium we will often use duration curves to describe random variables. One of the reasons why duration curves are useful is that it is possible to calculate the expectation value of the corresponding random variable by studying the area below the duration curve:

Theorem C.10. If $\underline{x}$ is the smallest possible outcome of X then the expectation value of X is given by

$$E[X] = \underline{x} + \int_{\underline{x}}^{\infty} \tilde{F}_X(x)dx.$$

In power system analysis the smallest possible outcome $\underline{x}$ s often equal to zero, which means that the expectation value is equal to the area below the part of the duration curve which belongs to the first quadrant. This is illustrated in figure C.1.

Correlations

When studying more than one random variable it is generally very important to know how their probability distributions are related to each other. A very important notion in this context is independent random variables. There are several definitions of independent random variables, but simply speaking, two random variables are X and Y are independent if the outcome of X does not affect the outcome of Y and vice versa. For example, if a dice is rolled twice, and X is the first outcome and Y the second, then X and Y should be independent.[2] If X is the first outcome and Y is the sum of the two results, then X and Y are dependent; Y will have different probability functions when the outcome of X is known.

One measure of how the probability distributions of two random variables are correlated to each other is the covariance:

Definition C.11. The covariance of two random variables X and Y is given by

$$Cov[X, Y] = E[(X - E[X])(Y - E[Y])].$$

A related notion is the correlation coefficient:

Definition C.12. The correlation coefficient of two random variables X and Y is given by

$$\rho_{X, Y} = \frac{Cov[X, Y]}{\sqrt{Var[X]Var[Y]}}.$$

The correlation coefficient will always be in the interval [–1, 1]. If $\rho(X, Y) = 0$ the variables X and Y are considered to be uncorrelated. Independent variables are always uncorrelated, but the converse is not always true.[3] If $\rho(X, Y)$ is positive, then the variables are positively correlated, which means that if we for example know that the outcome of X was higher than the average, then it is more likely that also Y was over the average. If the variables instead are negatively correlated ($\rho(X, Y) < 0$) then a high value of X result in an increased probability of a low value of Y.

2. Unless there is something fishy about the dice.
3. Consider for example a $U(-1, 1)$-distributed random variable X and $Y = X^2$. Although Y is clearly depending on X, the correlation coefficient $\rho(X, Y) = 0$.

Appendix D

THE NORMAL DISTRIBUTION

The normal distribution (also referred to as the Gaussian distribution) is a probability distribution which have a lot of uses. A general normal distribution is characterised by two parameters: the mean μ and the standard deviation σ.

Definition D.1. A random variable X is normally distributed if it has the following density and distribution functions:

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

and

$$F_X(x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(y-\mu)^2}{2\sigma^2}} dy.$$

This is denoted $X \in N(\mu, \sigma)$.

Using the definitions of expectation value and standard deviation it is possible to control that the density function and distribution functions above correspond to the expectation value μ and standard deviation σ.

An important special case is when the mean is zero and the standard deviation is equal to one:

Definition D.2. The standardised normal distribution has $\mu = 0$ and $\sigma = 1$. The density function of an $N(0, 1)$-distributed random variable is denoted $\varphi(x)$ and the distribution function is denoted $\Phi(x)$.

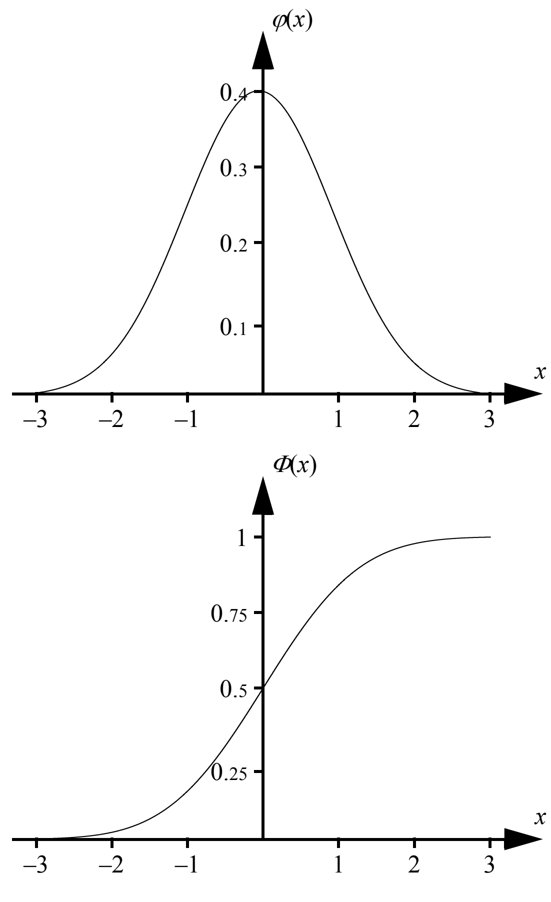


Figure D.1 The density function and distribution function of the standardised normal distribution.

Figure D.1 shows $\varphi(x)$ and $\Phi(x)$. Notice the symmetry!

Theorem D.3. If $X \in N(\mu, \sigma)$ then $Y = (X - \mu)/\sigma \in N(0, 1)$.

From theorem D.3 we can conclude that if $Y \in N(0, 1)$ then $X = \mu + \sigma Y \in N(\mu, \sigma)$. The practical importance of the theorem is that the standardised normal distribution can be used for calculations of a general normal distribution.

Example D.1. The random variable X is normally distributed with $\mu = 100$ and $\sigma = 50$. Calculate the probability that X is larger than 90.

Solution: According to the definition we have

$$P(X > 90) = 1 - P(X \leq 90) = 1 - F_X(90) = 1 - \Phi\left(\frac{90 - \mu}{\sigma}\right) = 1 - \Phi(-0.2) =$$

$$= \{\text{symmetry}\} = 1 - (1 - \Phi(0.2)).$$

The value of $\Phi(0.2)$ can be calculated on a computer or found in table D.1; the result is that the probability of $X > 90$ is about 57.93%.

An advantage of the normal distribution is that it is easy to add normally distributed variables:

Theorem D.4. Assume that there are independent normally distributed random variables, $X_1 \in N(\mu_1, \sigma_1), \ldots, X_n \in N(\mu_n, \sigma_n)$. Then it holds that the sum is

$$\sum_{i=1}^{n} c_i X_i \in N\left(\sum_{i=1}^{n} c_i \mu_i, \sqrt{\sum_{i=1}^{n} c_i^2 \sigma_i^2}\right).$$

Example D.2. Assume that $X \in N(100, 30)$ and $Y \in N(200, 40)$ and that X and Y are independent. Calculate the probability that the sum of these variables is larger than 350.

Solution: If $Z = X + Y$ then Z is, according to theorem D.4, normally distributed with the mean

$$\mu_Z = 100 + 200 = 300$$

and the standard deviation

$$\sigma_Z = \sqrt{30^2 + 40^2} = 50.$$

Hence we get,

$$P(X + Y > 350) = 1 - \Phi\left(\frac{350 - 300}{50}\right) = 1 - \Phi(1) \approx 0{,}1587.$$

Table

A table of the distribution function of the standardised normal distribution, i.e.,

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{y^2}{2}} dy$$

is given below. For $x < 0$ we can calculate $\Phi(x)$ using $\Phi(-x) = 1 - \Phi(x)$.

Table D.1 The distribution function of the standardised normal distribution.

x	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.50000	0.50399	0.50798	0.51197	0.51595	0.51994	0.52392	0.52790	0.53188	0.53586
0.1	0.53983	0.54380	0.54776	0.55172	0.55567	0.55962	0.56356	0.56749	0.57142	0.57535
0.2	0.57926	0.58317	0.58706	0.59095	0.59483	0.59871	0.60257	0.60642	0.61026	0.61409
0.3	0.61791	0.62172	0.62552	0.62930	0.63307	0.63683	0.64058	0.64431	0.64803	0.65173
0.4	0.65542	0.65910	0.66276	0.66640	0.67003	0.67364	0.67724	0.68082	0.68439	0.68793
0.5	0.69146	0.69497	0.69847	0.70194	0.70540	0.70884	0.71226	0.71566	0.71904	0.72240
0.6	0.72575	0.72907	0.73237	0.73565	0.73891	0.74215	0.74537	0.74857	0.75175	0.75490
0.7	0.75804	0.76115	0.76424	0.76730	0.77035	0.77337	0.77637	0.77935	0.78230	0.78524
0.8	0.78814	0.79103	0.79389	0.79673	0.79955	0.80234	0.80511	0.80785	0.81057	0.81327
0.9	0.81594	0.81859	0.82121	0.82381	0.82639	0.82894	0.83147	0.83398	0.83646	0.83891
1.0	0.84134	0.84375	0.84614	0.84849	0.85083	0.85314	0.85543	0.85769	0.85993	0.86214
1.1	0.86433	0.86650	0.86864	0.87076	0.87286	0.87493	0.87698	0.87900	0.88100	0.88298
1.2	0.88493	0.88686	0.88877	0.89065	0.89251	0.89435	0.89617	0.89796	0.89973	0.90147
1.3	0.90320	0.90490	0.90658	0.90824	0.90988	0.91149	0.91309	0.91466	0.91621	0.91774
1.4	0.91924	0.92073	0.92220	0.92364	0.92507	0.92647	0.92785	0.92922	0.93056	0.93189
1.5	0.93319	0.93448	0.93574	0.93699	0.93822	0.93943	0.94062	0.94179	0.94295	0.94408
1.6	0.94520	0.94630	0.94738	0.94845	0.94950	0.95053	0.95154	0.95254	0.95352	0.95449
1.7	0.95543	0.95637	0.95728	0.95818	0.95907	0.95994	0.96080	0.96164	0.96246	0.96327
1.8	0.96407	0.96485	0.96562	0.96638	0.96712	0.96784	0.96856	0.96926	0.96995	0.97062
1.9	0.97128	0.97193	0.97257	0.97320	0.97381	0.97441	0.97500	0.97558	0.97615	0.97670
2.0	0.97725	0.97778	0.97831	0.97882	0.97932	0.97982	0.98030	0.98077	0.98124	0.98169
2.1	0.98214	0.98257	0.98300	0.98341	0.98382	0.98422	0.98461	0.98500	0.98537	0.98574
2.2	0.98610	0.98645	0.98679	0.98713	0.98745	0.98778	0.98809	0.98840	0.98870	0.98899
2.3	0.98928	0.98956	0.98983	0.99010	0.99036	0.99061	0.99086	0.99111	0.99134	0.99158
2.4	0.99180	0.99202	0.99224	0.99245	0.99266	0.99286	0.99305	0.99324	0.99343	0.99361
2.5	0.99379	0.99396	0.99413	0.99430	0.99446	0.99461	0.99477	0.99492	0.99506	0.99520
2.6	0.99534	0.99547	0.99560	0.99573	0.99585	0.99598	0.99609	0.99621	0.99632	0.99643
2.7	0.99653	0.99664	0.99674	0.99683	0.99693	0.99702	0.99711	0.99720	0.99728	0.99736
2.8	0.99744	0.99752	0.99760	0.99767	0.99774	0.99781	0.99788	0.99795	0.99801	0.99807
2.9	0.99813	0.99819	0.99825	0.99831	0.99836	0.99841	0.99846	0.99851	0.99856	0.99861
3.0	0.99865	0.99869	0.99874	0.99878	0.99882	0.99886	0.99889	0.99893	0.99896	0.99900
3.1	0.99903	0.99906	0.99910	0.99913	0.99916	0.99918	0.99921	0.99924	0.99926	0.99929
3.2	0.99931	0.99934	0.99936	0.99938	0.99940	0.99942	0.99944	0.99946	0.99948	0.99950
3.3	0.99952	0.99953	0.99955	0.99957	0.99958	0.99960	0.99961	0.99962	0.99964	0.99965
3.4	0.99966	0.99968	0.99969	0.99970	0.99971	0.99972	0.99973	0.99974	0.99975	0.99976
3.5	0.99977	0.99978	0.99978	0.99979	0.99980	0.99981	0.99981	0.99982	0.99983	0.99983
3.6	0.99984	0.99985	0.99985	0.99986	0.99986	0.99987	0.99987	0.99988	0.99988	0.99989
3.7	0.99989	0.99990	0.99990	0.99990	0.99991	0.99991	0.99992	0.99992	0.99992	0.99992
3.8	0.99993	0.99993	0.99993	0.99994	0.99994	0.99994	0.99994	0.99995	0.99995	0.99995
3.9	0.99995	0.99995	0.99996	0.99996	0.99996	0.99996	0.99996	0.99996	0.99997	0.99997

Appendix E

RANDOM NUMBERS

Without algorithms to generate random numbers it would be impossible to perform Monte Carlo simulations on computers. Several methods for creating so-called pseudorandom numbers have been developed. The term pseudorandom is used to indicate that these numbers are not truly random, but based on a seed. Given the same seed a pseudorandom number generator will produce the same sequence of numbers. A good generator will produce a long sequence before it start repeating itself. Further, it is desirable that the generated numbers are as close as possible to a uniform distribution and that the correlation between the numbers is negligible.

Software which can generate good random numbers is readily available on all modern computer systems and we will therefore not study how a random number generator is working. Instead of that, we will show how to create random numbers with a certain distribution by transformation of a $U(0, 1)$-distributed random numbers, and how to select random numbers from a specified interval.

Generating Random Numbers from an Arbitrary Distribution

The common random number generator creates random numbers which are uniformly distributed between 0 and 1, i.e., random numbers from a $U(0, 1)$-distribution. The following theorem shows how random numbers of most other distributions can be created:

> ***Theorem E.1.*** If a random variable U is $U(0, 1)$-distributed then the random variable $Y = F_Y^{-1}(U)$ has the distribution function $F_Y(x)$.

$F_Y^{-1}(x)$ in the theorem is the inverse function of $F_Y(x)$. The theorem is therefore referred to as the *inverse transform method.*

Table E.1 Simplified model of the available generation capacity in a certain wind power plant.

Available generation capacity [kW]	Probability [%]
0	45
500	35
1 000	20

> ***Example E.1.*** Assume a simplified wind power plant model according to table E.1. Calculate a value of the available generation capacity starting from the random number $U_0 = 0.75$ from a $U(0, 1)$-distribution.

Solution: The table gives us the density function

$$f_{\overline{W}}(x) = \begin{cases} 0.45 & x = 0, \\ 0.35 & x = 500, \\ 0.2 & x = 1\,000, \\ 0 & \text{otherwise.} \end{cases}$$

According to theorem C.4 we get

$$F_{\overline{W}}(x) = \sum_{t \le x} f_{\overline{W}}(t) = \begin{cases} 0 & x < 0, \\ 0.45 & 0 \le x < 500, \\ 0.8 & 500 \le x < 1\,000, \\ 1 & 1\,000 \le x. \end{cases}$$

Although it is possible to state the inverse function $F_{\overline{W}}^{-1}(x)$, with a formal mathematical expression, it is more simple to solve the problem graphically. Figure E.1 shows the distribution function $F_{\overline{W}}$. As we all know, the inverted function is obtained by switching the *x*- and *y*-axis. $F_{\overline{W}}^{-1}(0.75)$ may hence be found by starting at 0.75 on the *y*-axis and then draw a horizontal line until the curve $F_{\overline{W}}$ is reached. From there a vertical line is drawn to the *x*-axis and there the result can be read. In this case we obtain $\overline{W}$ = 500 kW.

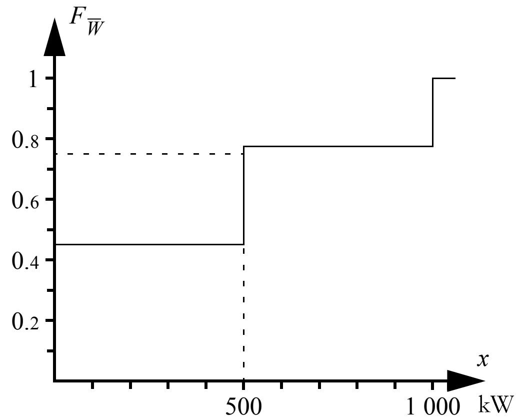


Figure E.1 Distribution function of the wind power model in example E.1.

It can be noted that when a graphic method is uses—as in the example above—it does not matter whether the distribution function or duration curve is used.

Generating Normally Distributed Random Numbers

The normal distribution does not have an inverse function $\Phi^{-1}(x)$! It is however possible to use an approximation of the inverse function for the transformation:

Theorem E.2. If *U* is a *U*(0, 1)-distributed random number then *X* is a *N*(0, 1)-distributed random number, if *X* is calculated according to

$$Q = \begin{cases} U & \text{if } 0 \le U \le 0.5, \\ 1 - U & \text{if } 0.5 < U \le 1, \end{cases}$$

$$t = \sqrt{-2\ln Q},$$

$$c_0 = 2.515517, \quad c_1 = 0.802853, \qquad c_2 = 0.010328,$$

$$d_1 = 1.432788, \quad d_2 = 0.189269, \qquad d_3 = 0.001308,$$

$$z = t - \frac{c_0 + c_1 t + c_2 t^2}{1 + d_1 t + d_2 t^2 + d_3 t^3}$$

and finally

$$X = \begin{cases} -z & \text{if } 0 \le U < 0.5, \\ 0 & \text{if } U = 0.5, \\ z & \text{if } 0.5 < U \le 1. \end{cases}$$

It should be noted that each value of U corresponds to a unique value of X!

Generating Random Numbers from a Part of a Distribution

When using stratified sampling it is necessary to be able to direct the outcome of a random variable to specified interval. For uniformly distributed random numbers this is a trivial task. If we want the outcome to belong to the interval $[a, b]$ then we just need to scale a $U(0, 1)$-distributed random number U_0:

$$U = (b - a)U_0 + a. \tag{E.1}$$

Since random numbers of an arbitrary distribution can be obtained by transforming $U(0, 1)$-distributed random numbers, it is possible to generate random numbers belonging to a specified interval by first determining which interval of the $U(0, 1)$ distribution which after transformation will correspond to the desired interval.

Example E.2. Assume that the load in a power system is $N(950, 50)$-distributed and that a random load larger than 1 000 MW is needed. Calculate a load level given the random number $U_0 = 0.4908$ from a $U(0, 1)$-distribution.

Solution: First we determine which value of an $N(0, 1)$-distribution which corresponds to 1 000 in an $N(950, 50)$-distribution:

$$x = \frac{1\,000 - 950}{50} = 1.$$

Then we calculate which value this corresponds to when a $U(0, 1)$-distribution random variable is transformed according to theorem E.2:

$$a = \Phi(x) \approx \{\text{use table D.1}\} \approx 0.84134.$$

Thus, a $U(0.84134, 1)$-distributed random number will after transformation correspond to a value within the desired interval. The calculations are then

$U = a + U_0 \cdot (1 - a) \approx 0.91921,$	{transformation to a $U(0.84134, 1)$-distribution}
$X \approx$ {apply theorem E.2} $\approx 1.4000,$	{transformation to an $N(0, 1)$-distribution}
$D = 950 + 50X \approx 1\,020.0.$	{transformation to an $N(950, 50)$-distribution}

ANSWERS AND SOLUTIONS

2.1

See page 9.

2.2

See page 9.

2.3

See page 9.

2.4

See the introduction to section 2.2 and figure 2.2.

2.5

See the introduction to section 2.2 and figure 2.2.

2.6

See the introduction to section 2.2 and figure 2.2.

2.7

See section 2.2.1.

2.8

See page 12.

2.9

A block bid is a purchase or sell bid which comprises several trading periods and which must be accepted as a whole.

2.10

The customer buys the same amount of energy in each trading period as long as the contract is valid.

2.11

During the time the contract is valid, the customer is allowed to consume as much energy they want each trading period, provided that the maximal power is not exceeded.

2.12

See section 2.2.2.

2.13

See page 14.

2.14

See page 14.

2.15

See section 2.2.3.

3.1

At an electricity price of 100 ¤/MWh the combined heat and power (CHP) having a variable operation cost less than the electricity price will be used. As the relation between the variable operation cost and the used potential is assumed to be linear, we must have

$$\frac{\textit{current price interval}}{\textit{total price interval}} = \frac{\textit{used potential}}{\textit{total potential}},$$

i.e., we use 35 · (100 – 80)/(150 – 80) = 10 TWh CHP. This is not enough.

If the electricity price is higher than 100 ¤/MWh but less than 120 ¤/MWh both CHP and nuclear power is used. Assume that the electricity price is λ. The total used potential is then

$$\underbrace{35 \cdot \frac{\lambda - 80}{150 - 80}}_{\text{used CHP}} + \underbrace{40 \cdot \frac{\lambda - 100}{120 - 100}}_{\text{used nuclear}}.$$

Setting this expression to be equal to 72.5 and solving for λ yields the electricity price $\lambda = 125$ ¤/MWh. This is however more than 120 ¤/MWh, i.e., the price when the entire nuclear potential is used. Thus, we can conclude that 40 TWh of the load is covered by nuclear generation and the remaining 32.5 TWh must be generated by the CHP:

$$35 \cdot \frac{\lambda - 80}{150 - 80} = 32.5 \Rightarrow \lambda = 145 ¤/\text{MWh}.$$

3.2

The electricity price is not changed, but remains at 145 ¤/MWh.

3.3

a) The electricity price is determined by the intersection of the supply and demand curves. The intersection can be found graphically by drawing both curves in the same figure. An alternative method of solution is to assume an electricity price, λ, between 25 and 40 ¤/MWh. The supply at these price levels can be written as 30 (hydro & nuclear) + (λ – 25) (fossil fuels) and the demand can be written 40 – (λ – 20)/2. These two expressions should be equal, which results in the electricity price λ = 30 ¤/MWh.

b) AB Vattenkraft generates 20 TWh at a variable cost of 5 ¤/MWh, which means that the total variable generation cost is 100 M¤/year. Everything is sold for the price 30 ¤/MWh, which generates an income of 600 M¤/year. Hence, the profit is 600 (income) – 100 (variable costs) – 100 (fixed costs) = 400 M¤/year.

3.4

a) The hydro generation and nuclear generation are both maximal at the electricity price 220 ¤/MW. The contribution from the bio fuelled power plants equals (220 – 100)/(260 – 100) · 16 = 12 TWh and the fossil fuelled power plants supply (220 – 190)/(390 – 190) · 20 = 3 TWh. Thus, the total generation is 120 TWh/year and the electricity consumption is obviously equally large.

b) The income of AB Vattenkraft amounts to 10 TWh/year · 220 ¤/MWh = 2 200 M¤/year. From the income we should subtract the total variable costs (10 TWh/year · 5 ¤/MWh = 50 M¤/year) and the fixed costs, which results in the profit 1 500 M¤/year.

3.5

Assume that the electricity price is 240 SEK/MWh. The total generation capacity is then 98 TWh/year. However, at this electricity price AB Skog is a net consumer of electricity; hence, the demand is larger than 100 TWh/year. The electricity price must therefore be higher in order to get a balance between supply and demand.

If the electricity price is 320 SEK/MWh then the total generation capacity is 101 TWh/year. The consumption of AB Skog at this price is 0.8 TWh/year, which results in a total consumption of 100.8 TWh/year. Consequently, it will be possible to obtain balance between supply and demand at this price.

3.6

a) 128.21 SEK/MWh.

b) Sweden imports approximately 4.2 TWh, Norway exports 18 TWh, Finland imports approximately 10.1 TWh and Denmark imports approximately 3.7 TWh.

c) If the reservoirs are filled then a part of the hydro energy must be used earlier than desired (otherwise the water has to be spilled). Hence, there will be a surplus of water before the 31st of July, which implies that the generation in the most expensive thermal power plants can be reduced; thus, the electricity price will be less than 128,21 SEK/MWh during this time. After 31st of July there is a shortage of water, which must be compensated by increased generation in more expensive thermal power plants, i.e., the electricity price will be higher than 128,21 SEK/MWh.

3.7

a) Assume that the company reduces nuclear generation having a variable operation cost 75 SEK/MWh by 1 TWh. This results in a reduced profit approximately equal to 1 TWh · (128.21 – 75) SEK/MWh ≈ 53.5 MSEK. On the other hand, the profit on sales from hydro power and the remaining nuclear power is increased, because for each TWh generation reduction by Voimajättiläs OY. results in a price increase of approximately 0.42 SEK/MWh. The extra profit is thus 34.5 TWh ·0.42 SEK/MWh ≈ 14.5 MSEK, which is not even close to the lost profit. The company is therefore better of not trying to manipulate the Nordic electricity price.

b) Similar reasoning as above yields a lost profit of 53.5 MSEK, but an extra profit of 55.6 MSEK. Kraftjätten AB could therefore gain of increasing the Nordic electricity price by voluntarily reducing the nuclear generation.

3.8

a) Assume an electricity price λ between 230 and 390 ¤/MWh. At this price all wind power and hydro power will be utilised, i.e., 70 TWh. The contribution from coal condensing and fossil gas can be written as

$$\frac{\lambda - 200}{320 - 200}30 + \frac{\lambda - 230}{390 - 230}20.$$

By setting this expression equal to 30 and solving for λ, we get the electricity price 290 ¤/MWh.

b) Naturally, the consumers' purchase costs corresponds to the producers' income of sold electricity, i.e., these two terms will cancel each other. The total surplus is therefore equal to the value of the consumption minus the coal production cost minus the cost of the emissions. At the electricity price 290 ¤/MWh we use 90/120 = 75% of the coal condensing, i.e., 22.5 TWh, and 60/160 = 37.5% of the fossil gas, i.e., 7.5 TWh. Thus, we get the following total surplus:

$$TS = 100 \cdot 500 - 10 \cdot 1 - 60 \cdot 10 - 22.5 \cdot (290 + 200)/2 - 7.5 \cdot (290 + 230)/2$$
$$- 22.5 \cdot 1 \cdot 50 - 7.5 \cdot 0.4 \cdot 50 \text{ TWh/year} \cdot \text{¤/MWh} = 40\,652.5 \text{ M¤/year}.$$

c) Including the emission rights, the cost of all coal condensing is increased by 1·40 = 40 ¤/MWh, while the costs of fossil gas is increased by 0.4·40 = 16 ¤/MWh. The two power sources must still produce in total 30 TWh; hence, we get the equation

$$\frac{\lambda - 240}{360 - 240}30 + \frac{\lambda - 246}{406 - 246}20 = 30 \Rightarrow \lambda = 322 \text{ ¤/MWh}.$$

The coal condensing generation is decreased to 20.5 TWh, whereas the fossil gas generation is increased to 9.5 TWh. The total surplus is now

$$TS = 100 \cdot 500 - 10 \cdot 1 - 60 \cdot 10 - 20.5 \cdot (322 + 240)/2 - 9.5 \cdot (322 + 246)/2$$
$$- 20.5 \cdot 1 \cdot 10 - 9.5 \cdot 0.4 \cdot 10 = 40\,668.05 \text{ M¤/year}.$$

4.1

$R = \Delta G/\Delta f = 10/0.05 = 200$ MW/Hz.

4.2

a) The increase in electricity consumptions results in a frequency decrease 100/2 000 = 0.05 Hz, i.e., the new frequency is 50 – 0.05 = 49.95 Hz.

b) The decrease in electricity consumptions results in a frequency increase 80/2 000 = 0.04 Hz, i.e., the new frequency is 50 + 0.04 = 50.04 Hz.

c) The increase in electricity generation results in a frequency increase 80/2 000 = 0.04 Hz, i.e., the new frequency is 50 + 0.04 = 50.04 Hz.

d) Due to the failure of the power plant, the new system gain is 1 800 MW/Hz. The decrease in electricity generation results in a frequency decrease 400/1 800 ≈ 0.22 Hz, i.e., the new frequency is 50 – 0.22 = 49.78 Hz.

4.3

a) $f = 49.89$ Hz, i.e., not within the interval.

b) $f \approx 49.90$ Hz (rounded upwards), i.e., not within the interval.

c) $f = 49.94$ Hz, i.e., within the interval.

d) $f = 50.09$ Hz, i.e., within the interval.

e) The specified reduction of the time error requires $f = 50.50$ Hz, i.e., not within the interval.

4.4

a) $G = G_0 - R(f - f_0) = 70 - 200(49.82 - 50) = 106$ MW. This is however more than the installed capacity; therefore, Fors will generate as much as possible, i.e., 100 MW.

b) $G = G_0 - R(f - f_0) = 70 - 200(49.94 - 50) = 82$ MW.

c) $G = G_0 - R(f - f_0) = 70 - 200(50.06 - 50) = 58$ MW.

d) $G = G_0 - R(f - f_0) = 70 - 200(50.18 - 50) = 34$ MW. This is however less than the least possible generation; therefore, Fors will generate as little as possible, i.e., 40 MW.

4.5

a) The gain in each power plant is about 1% of the total gain in the synchronous grid; hence, the power plants should supply about 1% of the generation increase, i.e., 10 MW. However, the first power plant can only increase its generation by 6 MW. The other power plants of the system must then altogether supply 1 000 MW. To achieve this generation increase the frequency must decrease by 0.2 Hz. This means that the other power plant on Ön increases its generation by 10 MW, i.e., to 72 MW (which is possible as it is less than the installed capacity). The total generation increase on Ön is therefore 16 MW.

b) The earlier import of 390 MW is decrease to an import of 374 MW, which does not cause any problem for the transmission line.

c) The new frequency of the system is 49.780 Hz.

4.6

$f = 49.8$ Hz, $G_1 = 200$ MW, $G_2 = 370$ MW, $G_3 = 580$ MW.

4.7

a) The necessary gain is given by $R_{tot} = \Delta G/\Delta f = 800$ MW/0.2 Hz = 4 000 MW/Hz.

b) If the fault occurs in area B then area A may not increase its export by more than 300 MW. As $\Delta G_A/\Delta G = R_A/R_{tot}$ it is not possible to have more than 300/400, i.e., 75%, of the gain in area A.

If the fault occurs in area A then area B may not increase the export by more than 700 MW, which means that at most 700/800, i.e., 87.5%, of the gain can be located in area B.

4.8

The generation increase is 80 MW in each area; hence, the flow on the transmission lines AB and BC must increase by 80 MW, which is possible. The new system frequency is approximately 49.7 Hz.

4.9

All transmission lines are disconnected. The frequencies become $f_A \approx 50.54$ Hz, $f_B \approx 49.35$ Hz and $f_C \approx 49.94$ Hz respectively.

4.10

The total gain is 2 500 + 2 000 + 1 250 + 250 = 6 000 MW/Hz. When the coal condensing unit is stopped the generation decreases by 200 MW; thus, the frequency decreases by 200/6 000 ≈ 0.033 Hz. This frequency lasts for five minutes resulting in a time deviation of –5·60·0.033·/50 = –0.2 s.

After five minutes the load increases by 100 MW; thus, the frequency decreases by and 100/6 000 ≈ 0.017 Hz. The total frequency decrease is now –0.033 – 0.017 = 0.05 Hz. This frequency also lasts for five minutes resulting in a further time deviation of –5·60·0.05·/50 = –0.3 s.

Ten minutes after the fault in the coal condensing unit the generation is increased by 300 MW; thus the frequency increases by 300/6 000 = 0.05 Hz. By that, the frequency is restored to 50 Hz and the time error will not change any more. The total time error is now +1 {the start value} – 0.2 – 0.3 = 0.5 s.

As the frequency has returned to 50 Hz no primary controlled generation is activated, but the neighbouring countries of Sweden have changed their net generation, which affects the transmission between the countries. As there is a circular flow between Norway, Sweden and Finland there is an infinite number of solutions to how the transmission changes. Let us therefore assume that the flow on the smallest connection (i.e., the one between Finland and Norway) remains unchanged. The new transmission flows then become as follows:

- Sweden → Denmark: 500 {start value} + 200 = 700 MW,
- Norway → Sweden: 2 000 {start value} – 100 = 1 900 MW,
- Sweden → Finland: 200 {start value} – 300= –100 MW.

4.11

$f = 49.833$ Hz, $t_i = -3$ s, $P_{N\to S} = -66.67$ MW, $P_{F\to N} = 100$ MW {asssumed constant},

$P_{S \to F} = -8.33$ MW, $P_{S \to D} = 458.33$ MW.

5.1

1 800 000 m^3 water corresponds to 1 800 000/3 600 = 500 HE. As the inflow is 40 HE, whereas no water is released, the reservoir must contain 540 HE at the end of the hour.

5.2

The production equivalent is given by $\gamma(Q) = H(Q)/Q$. The maximal production equivalent is obtained at best efficiency, i.e., $\gamma_{max} = 80/125 = 0.64$ MWh/HE.

5.3

For start, we know that at the discharge zero the generation is zero. Let us denote this point by $Q_0 = 0$ and $H_0 = 0$. Moreover, we know that at maximal discharge the installed capacity is generated; this point we denote $Q_3 = 200$ and $H_3 = 80$. To calculate the marginal production equivalents we also need the electricity generation at the two breakpoints best efficiency and local best efficiency. These are given by the formula $H = \gamma(Q) \cdot Q$:

$$Q_1 = 100,\ \gamma(Q_1) = 0.42 \Rightarrow H_1 = 0.42 \cdot 100 = 42 \text{ MW},$$

$$Q_2 = 160,\ \gamma(Q_2) = 0.4125 \Rightarrow H_2 = 0.4125 \cdot 160 = 66 \text{ MW}.$$

The marginal production equivalents can now be calculated according to

$$\mu_j = \frac{H_j - H_{j-1}}{Q_j - Q_{j-1}},$$

which results in the following linear model of the power plant:

$$\mu_1 = 0.42 \text{ MWh/HE},\ \overline{Q}_1 = 100 \text{ HE},$$

$$\mu_2 = 0.40 \text{ MWh/HE},\ \overline{Q}_2 = 60 \text{ HE},$$

$$\mu_3 = 0.35 \text{ MWh/HE},\ \overline{Q}_3 = 40 \text{ HE}.$$

5.4

$\mu_1 = 0.33$ MWh/HE, $\overline{Q}_1 = 100$ HE,
$\mu_2 = 0.30$ MWh/HE, $\overline{Q}_2 = 25$ HE.

5.5

$\mu_1 = 0.50$ MWh/HE, $\overline{Q}_1 = 80$ HE,
$\mu_2 = 0.45$ MWh/HE, $\overline{Q}_2 = 120$ HE.

5.6

a) Parameters: γ_i, λ_t, λ_{25}, $M_{i,0}$ and $V_{i,t}$. Optimisation variables: $M_{i,t}$, $S_{i,t}$ and $Q_{i,j,t}$.

b) maximise $\sum_{t=1}^{24} \lambda_t \sum_{i=1}^{4} \sum_{j=1}^{2} \mu_{i,j} Q_{i,j,t} +$

$$+ \lambda_{25}((\gamma_1 + \gamma_3 + \gamma_4)M_{1,24} + (\gamma_2 + \gamma_3 + \gamma_4)M_{2,24} + (\gamma_3 + \gamma_4)M_{3,24} + \gamma_4 M_{4,24}).$$

c) Hydrological balance for Vattnet/Språnget:

$$M_{1,t} = M_{1,t-1} + V_{1,t} - Q_{1,t} - S_{1,t}.$$

Hydrological balance for Forsen:

$$M_{2,t} = M_{2,t-1} + V_{2,t} - Q_{2,t} - S_{2,t}.$$

Hydrological balance for Sjön:

$$M_{3,t} = M_{3,t-1} + V_{3,t} + Q_{1,t} + Q_{2,t} + S_{2,t} - S_{3,t}.$$

Hydrological balance for Fallet:

$$M_{4,t} = M_{4,t-1} + V_{4,t} + S_{1,t} + S_{3,t} - Q_{4,t} - S_{4,t}.$$

5.7

In words the planning problem can be formulated as

maximise	*value of sold electricity + value of stored water*,
subject to	*hydrological balance for Språnget and Fallet*,
	minimal flow in the river.

Indices for power plants

Språnget 1, Fallet 2.

Parameters

The following parameters are stated in the problem:

$$\overline{M}_i = \text{maximal contents of reservoir } i = \begin{cases} 10\,000 & i = 1, \\ 2\,000 & i = 2, \end{cases}$$

$\mu_{i,j}$ = marginal production equivalent in power plant i, segment j =

$$= \begin{cases} 0.36 & i = 1, j = 1, \\ 0.32 & i = 1, j = 2, \\ 0.48 & i = 2, j = 1, \\ 0.47 & i = 2, j = 2, \end{cases}$$

$$\overline{Q}_{i,j} = \text{maximal discharge in power plant } i\text{, segment } j = \begin{cases} 300 & i = 1, j = 1, \\ 100 & i = 1, j = 2, \\ 100 & i = 2, j = 1, \\ 100 & i = 2, j = 2, \end{cases}$$

$$Q_V = \text{local inflow to reservoir } i = \begin{cases} 150 & i = 1, \\ 10 & i = 2, \end{cases}$$

$\underline{S}_i$ = minimal spillage through the fish ladder of power plant i = 1, i = 1, 2,

$\underline{\Theta}_i$ = minimal total water flow downstream power plant i = 10, i = 1, 2,

λ_f = expected future electricity price = 185,

$$\lambda_t = \text{expected electricity price during hour } t = \begin{cases} 200 & t = 1, \\ 210 & t = 2, \\ 210 & t = 3, \\ 220 & t = 4, \\ 210 & t = 5. \end{cases}$$

The only parameters which have to be calculated are the start contents of the reservoirs:

$$M_{i,0} = \text{start contents of reservoir } i = 0.5\overline{M}_i = \begin{cases} 5\ 000 & i = 1, \\ 1\ 000 & i = 2. \end{cases}$$

Optimisation variables

$Q_{i,j,t}$ = discharge in power plant i, segment j, during hour t,
i = 1, 2, j = 1, 2, t = 1, …, 5,
$S_{i,t}$ = spillage from reservoir i during hour t, i = 1, 2, t = 1, …, 5,
$M_{i,t}$ = contents of reservoir i at the end of hour t, i = 1, 2, t = 1, …, 5.

Objective function

$$\text{maximise} \quad \sum_{t=1}^{5} \lambda_t \sum_{i=1}^{2} \sum_{j=1}^{2} \mu_{i,j} Q_{i,j,t} + \lambda_f((\mu_{1,1} + \mu_{2,1})M_{1,5} + \mu_{2,1}M_{2,5}).$$

Constraints

Hydrological balance for Språnget:

$$M_{1,t} - M_{1,t-1} + Q_{1,1,t} + Q_{1,2,t} + S_{1,t} = V_{1,t}, \qquad t = 1, \ldots 5.$$

Hydrological balance for Fallet:

$$M_{2,t} - M_{2,t-1} + Q_{2,1,t} + Q_{2,2,t} + S_{2,t} - Q_{1,1,t} - Q_{1,2,t} - S_{1,t} = V_{2,t}, \qquad t = 1, \ldots 5.$$

Minimal flow in the river:

$$Q_{i,1,t} + Q_{i,2,t} + S_{i,t} \geq \underline{\Theta}_i, \qquad i = 1, 2, t = 1, \ldots 5.$$

Variable limits

$$0 \leq Q_{i,j,t} \leq \overline{Q}_{i,j}, \qquad i = 1, 2, j = 1, 2, t = 1, \ldots, 5,$$

$$\underline{S}_i \leq S_{i,t}, \qquad i = 1, 2, t = 1, \ldots, 5,$$

$$0 \leq M_{i,t} \leq \overline{M}_i, \qquad i = 1, 2, t = 1, \ldots, 5.$$

5.8

One m^3 of the fuel corresponds to 0.4 ton. As the heat content is 5 MWh/ton and the efficiency is 40% we get 0,8 MWh electricity per m^3 fuel. If the fuel costs 200 ¤/m^3 then the variable generation cost is 200/0.8 = 250 ¤/MWh.

5.9

90 SEK/MWh.

5.10

$C_{start\text{-}up}(1) = 10\,000$,
$C_{start\text{-}up}(2) = 20\,000$,
$C_{start\text{-}up}(3) = 30\,000$,
$C_{start\text{-}up}(4) = 36\,820$,
$C_{start\text{-}up}(5) = 40\,566$.

5.11

$u_t - u_{t-1} \le s_t^+$.

5.12

In words the planning problem can be formulated as

maximise *value of sold electricity – generation costs – start-up costs*,
subject to *limitations in generation capacity,*
relation between start-up and unit commitment,
requirement on minimum down time in the power plants.

Indices for power plants

Stad I 1, Stad II 2, Stad III 3, Strålinge I 4, Strålinge II 5.

Indices for price areas

Nord 1, Öst 2.

Parameters

The following parameters are stated in the problem:

$$\overline{G}_g = \text{installed capacity in power plant } g = \begin{cases} 200 & g = 1, \\ 150 & g = 2, \\ 300 & g = 3, \\ 600 & g = 4, \\ 650 & g = 5, \end{cases}$$

$$\underline{G}_g = \text{minimal generation when power plant } g \text{ is committed} = \begin{cases} 50 & g = 1, \\ 30 & g = 2, \\ 80 & g = 3, \\ 500 & g = 4, \\ 500 & g = 5, \end{cases}$$

$$\beta_{Gg} = \text{variable generation cost in power plant } g = \begin{cases} 150 & g = 1, \\ 180 & g = 2, \\ 160 & g = 3, \\ 100 & g = 4, \\ 100 & g = 5, \end{cases}$$

$$C_g^+ = \text{start-up cost in power plant } g = \begin{cases} 20\,000 & g = 1, \\ 18\,000 & g = 2, \\ 32\,000 & g = 3, \\ 40\,000 & g = 4, \\ 40\,000 & g = 5, \end{cases}$$

$u_{g,\,t}$ = commitment of unit g at the beginning of the planning period =

$$= \begin{cases} 0 & g = 1, 2, 3,\ t = -1, 0, \\ 1 & g = 4, 5,\ t = -1, 0, \end{cases}$$

$\lambda_{n,\,t}$ = expected electricity price in area n during hour t =

$$= \begin{cases}
116 \quad n = 1, 2, t = 1, & 291 \quad n = 1, t = 9, 10, & 138 \quad n = 1, 2, t = 15, \\
118 \quad n = 1, 2, t = 2, & 148 \quad n = 1, t = 9, & 137 \quad n = 1, 2, t = 16, \\
108 \quad n = 1, 2, t = 3, 4, & 150 \quad n = 2, t = 10, & 140 \quad n = 1, 2, t = 17, \\
115 \quad n = 1, 2, t = 5, & 290 \quad n = 1, t = 11, & 142 \quad n = 1, 2, t = 18, \\
160 \quad n = 1, t = 6, & 145 \quad n = 2, t = 11, & 137 \quad n = 1, 2, t = 19, \\
130 \quad n = 2, t = 6, & 273 \quad n = 1, t = 12, & 134 \quad n = 1, 2, t = 20, \\
210 \quad n = 1, t = 7, & 141 \quad n = 2, t = 12, 13, & 130 \quad n = 1, 2, t = 21, \\
136 \quad n = 2, t = 7, & 272 \quad n = 1, t = 13, & 120 \quad n = 1, 2, t = 22, \\
290 \quad n = 1, t = 8, & 163 \quad n = 1, t = 14, & 98 \quad n = 1, 2, t = 23, \\
148 \quad n = 2, t = 8, & 139 \quad n = 2, t = 14, & 105 \quad n = 1, 2, t = 24,
\end{cases}$$

t_g^- = minimal down time of power plant $g = 3$, $g = 1, \ldots, 5$.

Optimisation variables

$G_{g,\,t}$ = generation in power plant g, hour t, $g = 1, \ldots, 5$, $t = 1, \ldots, 24$,
$u_{g,\,t}$ = unit commitment of power plant g during hour t, $g = 1, \ldots, 5$, $t = 1, \ldots, 24$,
$s_{g,\,t}^+$ = start-up variable for power plant g, hour t, $g = 1, \ldots, 5$, $t = 1, \ldots, 24$,
$s_{g,\,t}^-$ = stop variable for power plant g, hour t, $g = 1, \ldots, 5$, $t = 1, \ldots, 24$.

Objective function

$$\text{maximise} \sum_{t=1}^{24} \left(\sum_{g=1}^{3} \lambda_{1,\,t} G_{g,\,t} + \sum_{g=4}^{5} \lambda_{2,\,t} G_{g,\,t} \right) - \sum_{t=1}^{24} \sum_{g=1}^{5} (\beta_{Gg} G_{g,\,t} + C_g^+ s_{g,\,t}^+).$$

Constraints

Maximal generation:

$$G_{g,\,t} - u_{g,\,t} \overline{G}_g \le 0, \qquad g = 1, \ldots, 5,\ t = 1, \ldots, 24.$$

Minimal generation:

$$u_{g,\,t}\underline{G}_g - G_{g,\,t} \le 0, \qquad g = 1, \ldots, 5,\ t = 1, \ldots, 24.$$

Relation between start-up, stop and unit commitment:

$$u_{g,\,t} - u_{g,\,t-1} - s^+_{g,\,t} + s^-_{g,\,t} = 0, \qquad g = 1, \ldots, 5,\ t = 1, \ldots, 24.$$

Requirement on down time:

$$s^-_{g,\,t} + \sum_{k=t+1}^{t+\underline{t}^-_g-1} s^+_{g,\,k} \le 1, \qquad g = 1, \ldots, 5,\ t = 1, \ldots, 22,$$

$$s^-_{g,\,t} + s^+_{g,\,t+1} \le 1, \qquad g = 1, \ldots, 5,\ t = 23.$$

Variable limits

$$u_{g,\,t} \in \{0, 1\}, \qquad g = 1, \ldots, 5,\ t = 1, \ldots, 24,$$

$$s^+_{g,\,t} \in \{0, 1\}, \qquad g = 1, \ldots, 5,\ t = 1, \ldots, 24,$$

$$s^-_{g,\,t} \in \{0, 1\}, \qquad g = 1, \ldots, 5,\ t = 1, \ldots, 24.$$

5.13

In words the planning problem can be formulated as

maximise *value of stored water – generation costs – start-up costs*,

subject to *hydrological balance for Forsen, Ån and Strömmen*,

limitations in generation capacity,

relation between start-up and unit commitment in Viken,

delivery of contracted load,

limitations in daily carbon dioxide emissions.

Indices of hydro power plants

Forsen 1, Ån 2, Strömmen 3.

Parameters

The following parameters are stated in the problem:

$\mu_{i,\,j}$ = marginal production equivalent in power plant i, segment j =

$$= \begin{cases} 0{,}28 & i = 1, j = 1, \\ 0{,}21 & i = 1, j = 2, \\ 0{,}42 & i = 2, j = 1, \\ 0{,}32 & i = 2, j = 2, \\ 0{,}34 & i = 3, j = 1, \\ 0{,}48 & i = 3, j = 1, \end{cases}$$

$$\overline{Q}_{i,j} = \text{maximal discharge in power plant } i\text{, segment } j = \begin{cases} 150 & i = 1, j = 1, \\ 160 & i = 1, j = 2, \\ 200 & i = 2, j = 1, \\ 200 & i = 2, j = 2, \\ 180 & i = 3, j = 1, \\ 190 & i = 3, j = 1, \end{cases}$$

$$V_i = \text{local inflow to reservoir } i = \begin{cases} 125 & i = 1, \\ 175 & i = 2, \\ 30 & i = 3, \end{cases}$$

$\overline{G}$ = maximal generation in Viken = 200,
$\underline{G}$ = minimal generation in Viken = 80,
β_G = variable operation cost in Viken = 320,
C^* = start-up cost after one hour down time in Viken = 15 000,
C^{**} = start-up cost after at least two hours down time in Viken = 35 000,
β_E = carbon dioxide emissions per generated MWh = 0.33,
E^* = carbon dioxide emissions when starting after one hour down time = 15,
E^{**} = carbon dioxide emissions when starting after at least two hours down time = = 36,

$$D_t = \text{contracted load during hour } t = \begin{cases} 150 & t = 1, \ldots, 8, 18, \ldots 24, \\ 350 & t = 9, \ldots, 17, \end{cases}$$

λ_f = expected future electricity price = 330,
u_t = unit commitment in Viken before the start of the planning period = 0, $t = -1, 0$.

The maximal reservoir contents must be converted to HE:

$$\overline{M}_i = \text{maximal contents of reservoir } i = \begin{cases} 7.2 \cdot 10^6/3\,600 = 2\,000 & i = 1, \\ 2.88 \cdot 10^6/3\,600 = 800 & i = 2, \\ 3.6 \cdot 10^6/3\,600 = 1\,000 & i = 3. \end{cases}$$

The start contents are 75% of the maximal contents:

$$M_{i,0} = \text{startcontents of reservoir } i = 0.75\overline{M}_i = \begin{cases} 1\,000 & i = 1, \\ 400 & i = 2, \\ 500 & i = 3. \end{cases}$$

Optimisation variables

$Q_{i,j,t}$ = discharge in power plant i, segment j, during hour t, $i = 1, 2, 3, j = 1, 2, t = 1, \ldots, 24$,
$S_{i,t}$ = spillage from reservoir i during hour t, $i = 1, 2, 3, t = 1, \ldots, 24$,
$M_{i,t}$ = contents of reservoir i at the end of hour t, $i = 1, 2, 3, t = 1, \ldots, 24$,
G_t = generation in Viken during hour t, $t = 1, \ldots, 24$,
u_t = unit commitment in Viken during hour t, $t = 1, \ldots, 24$,
s_t^* = start-up of Viken in hour t (after one hour down time), $t = 1, \ldots, 24$,
s_t^{**} = start-up of Viken in hour t (after at least two hours down time), $t = 1, \ldots, 24$.

Objective function

maximise $\lambda_f((\mu_{1,1} + \mu_{3,1})M_{1,24} + (\mu_{2,1} + \mu_{3,1})M_{2,24} + \mu_{3,1}M_{3,24})$

$$- \sum_{t=1}^{24} (\beta_G G_t + C^* s_t^* + C^{**} s_t^{**}).$$

Constraints

Hydrological balance for Forsen and Ån:

$$M_{i,t} - M_{i,t-1} + Q_{i,1,t} + Q_{1,2,t} + S_{i,t} = V_{i,t}, \qquad i = 1, 2, t = 1, \dots 24.$$

Hydrological balance for Strömmen:

$$M_{3,t} - M_{3,t-1} + Q_{3,1,t} + Q_{3,2,t} + S_{3,t} - Q_{1,1,t} - Q_{1,2,t} - S_{1,t} - Q_{2,1,t} - Q_{2,2,t} - S_{2,t} = V_{3,t}, \qquad t = 1, \dots 24.$$

Maximal generation in Viken:

$$G_t - u_t\overline{G} \le 0, \qquad t = 1, \dots, 24.$$

Minimal generation in Viken:

$$u_t\underline{G} - G_t \le 0, \qquad t = 1, \dots, 24.$$

Relation between start-up and unit commitment:

$$u_t - u_{t-1} - u_{t-2} - s_t^{**} \le 0, \qquad t = 1, \dots, 24,$$

$$u_t - u_{t-1} - s_t^{**} - s_t^* \le 0, \qquad t = 1, \dots, 24.$$

Delivery of contracted load:

$$\sum_{i=1}^{3} \sum_{j=1}^{2} \mu_{i,j} Q_{i,j,t} + G_t = D_t, \qquad t = 1, \dots, 24.$$

Limitation of carbon dioxide emissions:

$$\sum_{t=1}^{24} (\beta_E G_t + E^* s_t^* + E^{**} s_t^{**}) \le \overline{E}.$$

Variable limits

$$0 \le Q_{i,j,t} \le \overline{Q}_{i,j}, \qquad i = 1, 2, 3, j = 1, 2, t = 1, \dots, 24,$$

$$0 \le S_{i,t}, \qquad i = 1, 2, 3, t = 1, \dots, 24,$$

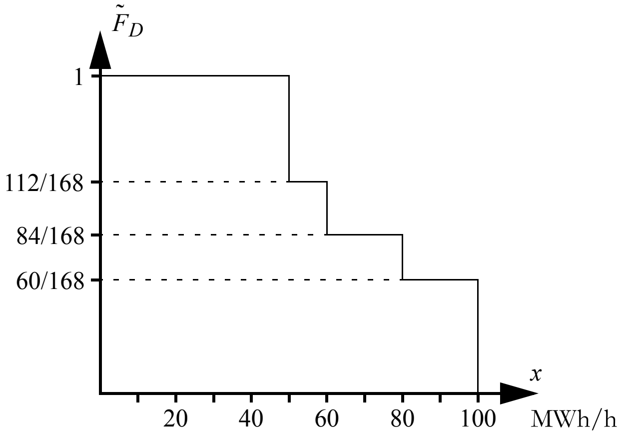
$$0 \le M_{i,t} \le \overline{M}_i, \qquad i = 1, 2, 3, t = 1, \dots, 24,$$

$$u_t \in \{0, 1\}, \qquad t = 1, \dots, 24,$$

$$s_t^* \in \{0, 1\}, \qquad t = 1, \dots, 24,$$

$$s_t^{**} \in \{0, 1\}, \qquad t = 1, \dots, 24.$$

6.1



6.2

$$EG_g = EENS_{g-1} - EENS_g = \{\text{use (6.15)}\} = T\int_{\hat{G}_{g-1}^{tot}}^{\infty} \tilde{F}_{g-1}(x)dx - T\int_{\hat{G}_{g}^{tot}}^{\infty} \tilde{F}_{g}(x)dx =$$

$$= T\left((p_g + q_g)\int_{\hat{G}_{g-1}^{tot}}^{\infty} \tilde{F}_{g-1}(x)dx - \int_{\hat{G}_{g}^{tot}}^{\infty} (p_g\tilde{F}_{g-1}(x) + q_g\tilde{F}_{g-1}(x - \hat{G}_g))dx \right) =$$

$$= \ldots = T \cdot p_g \int_{\hat{G}_{g-1}^{tot}}^{\hat{G}_{g}^{tot}} \tilde{F}_{g-1}(x)dx = EG_g, \text{ QED.}$$

6.3

If the previous power plant is available then power plant g will be used to cover the part of the load which exceeds the total capacity of the power plants with lesser operation costs, i.e., $\hat{G}_{g-1}$, but which is less than the total capacity including power plant g, i.e., $\hat{G}_g^{tot}$. The load "seen" by power plant g equals the real load plus outages in the less expensive power plants, except the previous unit (as we assumed that unit to be available), i.e., the equivalent load $\tilde{F}_{g-2}$. Therefore, the expected generation in this case is

$$T\int_{\hat{G}_{g-1}^{tot}}^{\hat{G}_{g}^{tot}} \tilde{F}_{g-2}(x)dx.$$

On the other hand, if the previous power plant is not available then power plant g must be dispatched already when the load exceeds the installed capacity of the less expensive units excluding power plant $g-1$ (which is not available), i.e., $\hat{G}_{g-2}^{tot}$. The load exceeding $\hat{G}_{g-2}^{tot} + \hat{G}_g$ cannot be covered by power plant g. Also in this case the load seen by power plant g is equal to the equivalent load n $\tilde{F}_{g-2}$. Hence, in this case we get the expected generation

$$T\int_{\hat{G}_{g-2}^{tot}}^{\hat{G}_{g-2}^{tot}+\hat{G}_g} \tilde{F}_{g-2}(x)dx.$$

This expression can be simplified by adding $\hat{G}_{g-1}$ to both the upper and lower limit of the integral, while subtracting the same amount from the integration variable. This yields

$$T\int_{\hat{G}_{g-1}^{tot}}^{\hat{G}_g^{tot}} \tilde{F}_{g-2}(x-\hat{G}_{g-1})dx.$$

The expected generation in power plant *g* is the sum of the generation in the two cases described above. If we consider the probability of each case then we obtain the following expression for the total expected generation:

$$EG_g = T\cdot p_g\left(p_{g-1}\int_{\hat{G}_{g-1}^{tot}}^{\hat{G}_g^{tot}} \tilde{F}_{g-2}(x)dx + q_{g-1}\int_{\hat{G}_{g-1}^{tot}}^{\hat{G}_g^{tot}} \tilde{F}_{g-2}(x)dx\right) =$$

$$= T\cdot p_g\int_{\hat{G}_{g-1}^{tot}}^{\hat{G}_g^{tot}} (p_{g-1}\tilde{F}_{g-2}(x) + q_{g-1}\tilde{F}_{g-2}(x-\hat{G}_{g-1}))dx.$$

The function to be integrated is clearly on the same form as the right hand side of (6.13). Hence, the formula above can be simplified to

$$EG_g = T\cdot p_g\int_{\hat{G}_{g-1}^{tot}}^{\hat{G}_g^{tot}} \tilde{F}_{g-1}(x)dx,$$

which is the formula we should derive.

6.4

Since we do not know the number and duration of each failure we cannot calculate the *MTTF*. However, the table shows the sums

$$\sum_{k=1}^{M} t_u(k) = 667.5 + 651 + 633 + 706 = 2\,657.5 \text{ h}$$

and

$$\sum_{k=1}^{M} t_d(k) = 76.5 + 21 + 111 + 14 = 222.5 \text{ h}.$$

This is enough estimation to estimate the availability using the expression

$$p = \frac{\sum_{k=1}^{M} t_u(k)}{\sum_{k=1}^{M} (t_u(k) + t_d(k))} = \frac{2\,657.5}{2\,657.5 + 222.5} \approx 0.92.$$

6.5

a) Since all power plants are 100% reliable, we get $\tilde{F}_3(x) = \tilde{F}_2(x) = \tilde{F}_1(x) = \tilde{F}_0(x)$; hence,

$$EG_1 = EENS_0 - EENS_1 = \int_0^{\infty} \tilde{F}_0(x)dx - \int_{1\,500}^{\infty} \tilde{F}_1(x)dx = \int_0^{1\,500} \tilde{F}_0(x)dx = 1 \cdot 1\,500 =$$

$= 1\,500$ MWh/h.

$$EG_2 = EENS_1 - EENS_2 = \int_{1\,500}^{\infty} \tilde{F}_1(x)dx - \int_{3\,000}^{\infty} \tilde{F}_2(x)dx = \int_{1\,500}^{3\,000} \tilde{F}_0(x)dx =$$

$= (1 + 0.8) \cdot 500/2 + (0.8 + 0.2) \cdot 500/2 + (0.2 + 0.1) \cdot 500/2 = 775$ MWh/h.

$$EG_3 = EENS_2 - EENS_3 = \int_{3\,000}^{\infty} \tilde{F}_2(x)dx - \int_{4\,000}^{\infty} \tilde{F}_3(x)dx = \int_{3\,000}^{4\,000} \tilde{F}_0(x)dx =$$

$= (0.1 + 0.05) \cdot 1\,000/2 = 75$ MWh/h.

b) $ETOC = 100EG_2 + 250EG_3 = 96\,250$ ¤/h.

c) $LOLP = \tilde{F}_3(4\,000) = \tilde{F}_0(4\,000) = 5\%$.

6.6

The risk of load shedding is given directly by figure 6.30b as $LOLP = \tilde{F}_2(700) \approx 1\%$.

The expected generation in the diesel generator is

$$EG_2 = EENS_1 - EENS_2 = \int_{500}^{\infty} \tilde{F}_1(x)dx + \int_{700}^{\infty} \tilde{F}_2(x)dx = (3.89 + 0.16) - (0.39 + 0.02) =$$

$= 3.64$ kWh/h.

Hence, we get $ETOC = 1 \cdot EG_2 = 3.64$ ¤/h.

6.7

a) The equivalent load duration curve including a power plant of 120 MW having 100% availability is given by

$$\tilde{F}_1(x) = 1 \cdot \tilde{F}_0(x) + 0 \cdot \tilde{F}_0(x - 120) = \tilde{F}_1(x).$$

Hence, we get

$$LOLP = \tilde{F}_1(120) = 10\%, EENS = \int_{120}^{\infty} \tilde{F}_1(x)dx = 0.1 \cdot 10 = 1 \text{ MWh/h.}$$

b) The equivalent load duration curve including a power plant of 150 MW having 90% availability is given by

$$\tilde{F}_1(x) = 0{,}9 \cdot \tilde{F}_0(x) + 0{,}1 \cdot \tilde{F}_0(x-150) = \begin{cases} 1 & x < 80, \\ 0.64 & 80 \le x < 100, \\ 0.19 & 100 \le x < 130, \\ 0.10 & 130 \le x < 230, \\ 0.06 & 230 \le x < 250, \\ 0.01 & 250 \le x < 280, \\ 0 & 280 \le x. \end{cases}$$

Hence, we get

$$LOLP = \tilde{F}_1(150) = 10\%, \; EENS = \int_{150}^{\infty} \tilde{F}_1(x)dx = 0.1 \cdot 80 + 0.06 \cdot 20 + 0.01 \cdot 30 = 9.5 \text{ MWh/h.}$$

6.8

a) The unserved energy with no diesel generator sets is

$$EENS_0 = 8\,760 \int_0^{\infty} \tilde{F}_0(x)dx = 8\,760(400 \cdot 1 + 200 \cdot 0.6 + 200 \cdot 0.1 + 200 \cdot 0.01) = 4\,747\,920 \text{ kWh/yr.}$$

The unserved energy with four diesel generator sets is

$$EENS_4 = 8\,760 \int_{800}^{\infty} \tilde{F}_4(x)dx =$$

$$= 8\,760 \cdot 200(0.068581 + 0.010036 + 0.000906 + 0.000046 + 0.000001) \approx 139\,407 \text{ kWh/yr.}$$

The difference between the two equals the energy produced and—as we neglect the losses—sold to the customers. If the highest price is charged the profit is 1.00 – 0.50 ¤/kWh. In a year, AECL will gain a total income of about 2.3 M¤. The total investment cost of the gird and the four diesel generator sets is 2.2 M¤/yr.; hence, the investment yields a small profit.

b) To calculate $EENS_5$ we need $\tilde{F}_5(x)$ for $x \ge 1\,000$:

$$\tilde{F}_5(x) = 0.9 \cdot \tilde{F}_4(x) + 0.1 \cdot \tilde{F}_4(x-200) =$$

$$= \begin{cases} 0.9 \cdot 0{,}010036 + 0.1 \cdot 0.068581 = 0.0158905 & 1\,000 \le x < 1\,200, \\ 0.9 \cdot 0{,}000906 + 0.1 \cdot 0.010036 = 0.0018190 & 1\,200 \le x < 1\,400, \\ 0.9 \cdot 0{,}000046 + 0.1 \cdot 0.000906 = 0.0001320 & 1\,400 \le x < 1\,600, \\ 0.9 \cdot 0{,}000001 + 0.1 \cdot 0.000046 = 0.0000055 & 1\,600 \le x < 1\,800, \\ 0.9 \cdot 0 + 0.1 \cdot 0.000001 = 0.0000001 & 1\,800 \le x < 2\,000, \\ 0 & 2\,000 \le x. \end{cases}$$

$$EENS_5 = 8\,760 \int_{1\,000}^{\infty} \tilde{F}_5(x)dx =$$

$$= 8\,760 \cdot 200(0.0158905 + 0.0018190 + 0.0001320 + 0.0000055 + 0.0000001) \approx 31\,268 \text{ kWh/yr.}$$

The expected generation in the fifth unit is $EENS_4 - EENS_5 \approx 108\,139$ kWh/yr., which generates an additional income of about 54 000 ¤/yr. The investment is then profitable, as the investment cost is slightly less than that.

6.9

a) In the figure we can read $LOLP = \tilde{F}_3(800) \approx 7\%$, which apparently does not fulfil the requirement. Adding another diesel generator set results in

$$LOLP = \tilde{F}_4(1\,000) = 0.9\tilde{F}_3(1\,000) + 0.1\tilde{F}_3(1\,000 - 200) \approx 0.9 \cdot 0.01 + 0.1 \cdot 0.07 \approx 1.6\%.$$

Hence, a third diesel generator set seems to be sufficient to fulfil the requirement on reliability of supply.

b) The hydro power is assumed to be 100% reliable, which means that $\tilde{F}_1(x) = \tilde{F}_2$. As the hydro power has negligible operation cost, the *ETOC* will only depend on the total generation in the three diesel generator sets. First, we calculate the unserved energy with and without the three diesel generator sets:

$$EENS_1 = \int_{400}^{\infty} \tilde{F}_1(x)dx = 119.11 + 21.17 + 1.73 + 0.01 = 142.02 \text{ kWh/h},$$

$$EENS_4 = \int_{1\,000}^{\infty} \tilde{F}_4(x)dx = 1.19 \text{ kWh/h}.$$

Thus, we get

$$ETOC = 0.5 \cdot (EG_2 + EG_3 + EG_4) = 0.5 \cdot (EENS_1 - EENS_4) \approx 70.4 \text{ ¤/h}.$$

6.10

First we consider the expected load during one year, which corresponds to the unserved energy when no power plants have been included:

$$EENS_0 = 8\,760 \int_{0}^{\infty} \tilde{F}_0(x)dx = 8760 \cdot (400 \cdot 1 + 200 \cdot 0{,}6 + 200 \cdot 0{,}1 + 200 \cdot 0{,}01) = 4\,747\,920 \text{ kWh/yr}.$$

The unserved energy when both power plants have been added is calculated using

$$EENS_2 = 8\,760 \int_{1\,400}^{\infty} \tilde{F}_2(x)dx.$$

We see that it is sufficient to determine $\tilde{F}_2(x)$ for $x \geq 1\,400$. As both power plants have negligible operation costs it does not matter in which order we add them to the equivalent load. Let us start with the hydro power:

$$\tilde{F}_1(x) = 1 \cdot \tilde{F}_0(x) + 0 \cdot \tilde{F}_0(x - 800) = \tilde{F}_0(x).$$

Then we add the wind power:

$$\tilde{F}_1(x) = 0.12 \cdot \tilde{F}_0(x) + 0.18 \cdot \tilde{F}_0(x - 200) + 0.37 \cdot \tilde{F}_0(x - 400) + 0.33 \cdot \tilde{F}_0(x - 600) =$$

$$= \begin{cases} \dots & \dots \\ 0.12 \cdot 0 + 0.18 \cdot 0 + 0.37 \cdot 0 + 0.33 \cdot 0.01 = 0.0033 & 1\,400 \leq x < 1\,600, \\ 0 & 1\,600 \leq x. \end{cases}$$

We get

$$EENS_2 = 8760 \cdot 200 \cdot 0{,}0033 = 5\ 782 \text{ kWh/yr.}$$

The generation in the two power plants is equal to the difference between $EENS_0$ and $EENS_2$, i.e., 4 742 138 kWh/yr. If all energy is sold for 1 ¤/kWh then the income will be larger than the annual cost. The investments seems to be profitable.

6.11

a) According to the problem we have $\mu_L = 3\ 500$ and $\sigma_L = 300$. Using the definitions we get the following mean and standard deviation of outages in the nuclear power plants: $\mu_O = 0.95 \cdot 0 + 0.05 \cdot 1\ 000 = 50$ and $\sigma_O^2 = 0.95 \cdot (0 - 50)^2 + 0.05 \cdot (1\ 000 - 50)^2 = 47\ 500$. The equivalent load duration curve including the five nuclear power plants has the mean

$$\mu_5 = \mu_L + 5\mu_O = 3\ 750$$

and the standard deviation

$$\sigma_5 = \sqrt{\rho_L^2 + 5\rho_O^2} \approx 572.$$

The loss of load probability is then

$$LOLP \approx \tilde{F}_{5N}(5 \cdot 1\ 000) = 1 - \Phi\left(\frac{5\ 000 - \mu_5}{\sigma_5}\right) = 1 - \Phi\left(\frac{5\ 000 - 3\ 750}{572}\right) =$$

$$= 1 - \Phi(2.18) = \{\text{use table}\} = 0.0146.$$

b) With four nuclear power plants and wind power we get an equivalent load duration curve with the mean

$$\mu_{4V} = \mu_L + 4\mu_O + 0.6V = 3\ 700 + 0.6V$$

and the standard deviation

$$\sigma_{4V} = \sqrt{\sigma_L^2 + 4\sigma_O^2 + (0.1V)^2} = \sqrt{280\ 000 + 0.01V^2}.$$

If the *LOLP* is to remain unchanged we must have

$$\Phi\left(\frac{4\ 000 + V - \mu_{4V}}{\sigma_{4V}}\right) = \Phi(2.18)$$

$$\Rightarrow$$

$$4\ 000 + V - 3\ 700 - 0{,}6V = 2.18 \cdot \sqrt{280\ 000 + 0.01V^2}.$$

If we take the square of both sides of the equation and rearrange the expression we get

$$0.11V^2 + 240V - 1\ 245\ 878 = 0 \Rightarrow V \approx 2\ 430.$$

Thus, it is necessary to build about 2 430 MW wind power if the risk of power deficit should be the same as before.

6.12

ETOC is estimated by $m_{TOC} = \dfrac{1}{2\ 000} \sum_{i=1}^{2\ 000} toc_i = 256$ ¤/h.

LOLP is estimated by $m_{LOLO} = \frac{1}{2\,000}\sum_{i=1}^{2\,000} lolo_i = 0.8\%$.

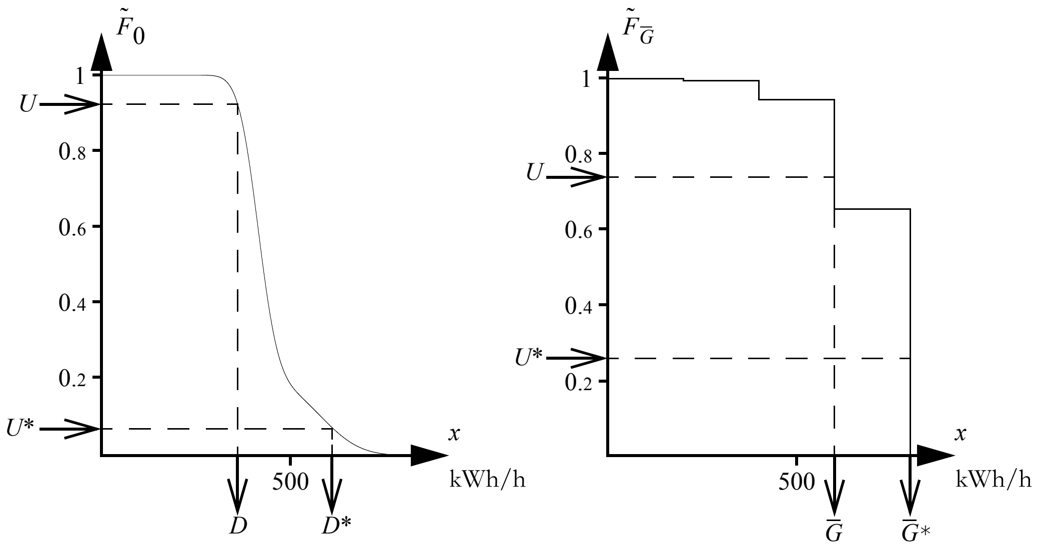
6.13

LOLP is estimated to 0.12%.

6.14

The $U(0, 1)$-distributed random numbers are transformed to random number with the distribution $F_Y(x)$ using the inverse transform method, i.e., by calculating $Y = F_Y^{-1}(U)$. The inverse distribution function can be found in the figures, which gives the following random numbers and complementary random numbers:

$$D \approx 360,\ D^* \approx 620,\ \overline{G} = 600,\ \overline{G}^* = 800.$$



We get the following samples of *TOC* and *LOLO*:

	$\overline{G}$	D	*TOC*	*LOLO*
	600	360	180	0
	600	620	300	1
	800	360	180	0
	800	620	310	0
Mean			242,50	0,25

Hence, *ETOC* is estimated to 242.50 ¤/h and *LOLP* to 25%.

6.15

ETOC is estimated to about 71.7 ¤/h when using the control variate method.

6.16

a) The maximal losses in this system are obtained when both hydro and wind power generates their installed capacity:

$L_{1,2} = 3 \cdot 10^{-5} \cdot 800^2 = 19.2$ kW, {800 = installed hydro capacity}

$L_{2,3} = 8 \cdot 10^{-5} \cdot (780.8 + 600)^2 \approx 152.5$ kW, {720.8 = maximal import from Ekiyira, 600 = installed wind capacity}

Hence, the maximal losses are $\bar{L} = 171.7$ kW. This means that there are three interesting intervals for the load:

	TOC	*LOLO*
$D \leq 978.3$	0	0
$978.3 < D \leq 1\,150$	0	0 or 1
$1\,150 < D$	0	1

b) First, we compute the mean of the observations for each stratum:

$$m_{LOLO_1} = \frac{1}{n_1}\sum_{i=1}^{n_1} LOLO_{1,i} = \frac{0}{110} = 0,$$

$$m_{LOLO_2} = \frac{1}{n_2}\sum_{i=1}^{n_2} LOLO_{2,i} = \frac{58}{330} = 0.1758,$$

$$m_{LOLO_3} = \frac{1}{n_3}\sum_{i=1}^{n_3} LOLO_{3,i} = \frac{8}{8} = 1.$$

The final estimate of *LOLP* is calculated by weighting the results according to their stratum weights:

$$LOLP \approx \sum_{h=1}^{3} \omega_h m_{LOLO_h} = 0.9560 \cdot 0 + 0.0332 \cdot 0.1758 + 0.0108 \cdot 1 = 0.0166 \approx 1.7\%.$$

The estimates of the expected generation in the power plants is calculated in a similar manner:

$$E[H] \approx 0.9560 \cdot 41\,905/110 + 0.0332 \cdot 255\,087/330 + 0.0108 \cdot 6\,400/8 \approx 398 \text{ kWh/h},$$

$$E[W] \approx 0.9560 \cdot 20\,548/110 + 0.0332 \cdot 30\,982/330 + 0.0108 \cdot 210/8 \approx 182 \text{ kWh/h}.$$

6.17

$n_1^2 = 8$, $n_2^2 = 20$, $n_3^2 = 20$.

6.18

$n_1^2 = 22$, $n_2^2 = 0$, $n_3^2 = 14$.

INDEX