Modeling the total inflow energy to hydropower plants

- a study of Sweden and Norway

Fredrik Olsson and Mark Pearson

Department of Water Resources Engineering
Lund Institute of Technology
Lund University

Modeling the total inflow energy to hydropower plants

- a study of Sweden and Norway

Modellering av den totala tillrinningsenergin till vattenkraftverk

- en studie av Sverige och Norge

Fredrik Olsson and Mark Pearson

Copyright © Fredrik Olsson and Mark Pearson

Institutionen för Teknisk Vattenresurslära
Lunds Tekniska Högskola, Lunds Universitet
Box 118
221 00 Lund
046-222 00 00

Department of Water Resources Engineering
Lund's Institute of Technology, Lund's University
Box 118
SE-221 00 Lund
(+46)46-222 00 00

Thesis number 2005:4

KFS AB
Lund 2005
Printed in Sweden

Abstract

Title: Modeling the inflow energy to hydropower plants – a study of Sweden and Norway

Authors: Fredrik Olsson and Mark Pearson

Supervisors: Dr. Cintia Bertacchi Uvo, Department of Water Resources Engineering, Lund University and Stefan Söderberg, SMHI.

Problem: The electricity price changes each day in the Nordic power exchange market, Nord Pool. These changes are closely linked with the amount of water entering the hydropower plants, also called inflow energy, since hydropower is the electricity source primary used. By in advance knowing the total weekly inflow energy, players dealing with electricity can anticipate minor changes in the electricity price.

Purpose: The main purpose of this study is to find a model for Sweden and Norway based on Neural Network or Multiple Regression which simulates the total weekly inflow energy from runoff data. The model should use as few stations as possible while still performing a simulation with high skill. The selected model should be integrated in a user-friendly software for operational use for Sweden.

Method: By using runoff for Swedish and Norwegian stations provided by SMHI and NVE a model for each country was selected. The selected stations were chosen through trial and error after evaluating the model's simulations. Forecasted runoff data was further used in the Swedish model for performing simulations of the total weekly inflow energy days in advance. The selected model was made operational by integrating it into a user-friendly software.

Conclusions: The software, named TWh-Simulator, performs simulations of such quality that the results can be relied upon. This conclusion is based upon the good results achieved when validating the simulations over three different time periods. In order for the model to retain the good results it has to be updated at some time.

Keywords: Nord Pool, hydropower, inflow energy, runoff, Multiple Regression, Neural Network

Sammanfattning

Titel:	Modellering av den totala tillrinningsenergin till vattenkraftverk - en studie av Sverige och Norge
Författare:	Fredrik Olsson and Mark Pearson
Handledare:	Dr. Cintia Bertacchi Uvo, Institutionen för Teknisk Vattenresurslära vid Lunds Tekniska Högskola och Stefan Söderberg, SMHI.
Problempresentation:	Elpriset ändras varje dag på den nordiska elbörsen, Nord Pool. Dessa förändringar är nära relaterade med den mängd vatten som rinner in till vattenkraftverken, kallad tillrinningsenergi, eftersom vattenkraft är den sorts elektricitet som i första hand används. Genom att i förväg veta den totala veckotillrinningsenergin kan de som handlar med elektricitet förutspå mindre förändringar i elpriset.
Syfte:	Det främsta syftet med det här examensarbetet är att ta fram en model för Sverige och Norge baserad antingen på Neurala Nätverk eller Multipel Regression som kan simulera den totala veckotillrinningsenergin från avrinningsdata. Modellen ska använda sig av så få stationer som möjligt samtidigt som simuleringarna har bra kvalitet. Den valda modellen ska sedan integreras i en användarvänlig mjukvara för att kunna användas operativt för Sverige
Metod:	Genom att använda avrinning från svenska och norska stationer, tillhandahållna av SMHI och NVE, valdes en model för respektive land. Stationerna valdes genom “trial and error” efter att modellens resultat evaluerats. Prognostiserad avrinningsdata användes vidare i den svenska modellen för att kunna utföra prognoser av den totala veckotillrinningsenergin. Den valda modellen gjordes operativ genom att integrera den i en användarvänlig mjukvara.

Slutsatser:	Mjukvaran, kallad TWh-Simulator, utför simuleringar av sådan kvalitet att dess resultat är fullt tillförlitliga. Denna slutsats bygger på de goda uppnådda resultat när modellen validerades på tre olika tidsperioder. För att modellen ska upprätthålla dess goda resultat måste den uppdateras vid något tillfälle
Nyckelord:	Nord Pool, vattenkraft, tillrinningsenergi, avrinning, Multipel Regression, Neurala Nätverk

Preface

First of all we would like to thank our supervisors Cintia Bertacchi Uvo at the department of Water Resources Engineering at Lund Institute of Technology and Stefan Söderberg at SMHI for all the time and commitment that they have devoted to us during our work with this thesis. We would also like to thank Erik Holmqvist at NVE for providing us with the Norwegian data as well as for taking the time to answer all the questions we had about the Norwegian hydropower production.

Furthermore we would also like to thank Ron Pearson for proofreading the thesis along with our opponents Magnus Fredrikson and Johan Persson for valuable comments.

Lund, March 14, 2005

Fredrik Olsson and Mark Pearson

Table of contents

1 Introduction

In this chapter, the background to this thesis is described along with the goals and limitations.

1.1 Background

Nowadays electricity is a necessity. Without it, a whole community can stop functioning. Sweden has, with a consumption of 15 000 kWh per inhabitant and year, the fourth largest electricity consumption in the world, only beaten by Norway, Canada and Island. This high consumption is mainly due to the cold climate and energy craving industries. During 2003, Sweden consumed a total of 145,3 TWh electricity of which 42,3 TWh was used in households, 33-34 TWh in the service sector and 55,7 TWh in industries. (Svensk Energi, 2005 a) Due to the high electricity consumptions it is of common interest to buy electricity at the lowest price possible.

Norway was, in 1991, the first Nordic country to deregulate the power market. Since 1996, the Swedish power market is deregulated as well, making it possible for the user to choose its energy supplier. The electricity price changes each day and can make a huge financial difference for industries that have a high consumption of electricity. Nowadays, it is also possible to buy electricity weeks in advance, perhaps when the price is supposedly low. Since industries consume a lot of energy, the price at which it is purchased is of great importance. Several electricity suppliers provide, as a special service to their industrial costumers, customized electricity purchases that intend to minimize electricity costs to high electricity consumers.

Nord Pool is the Nordic power exchange market where these transactions are made. For people dealing with electricity in future markets, it can be helpful to know in advance the amount of water which enters the Nordic hydropower plant dams every week. The reason for this is that hydropower is the energy source primary used since it is the cheapest, as well as the most environmental friendly way of producing electricity. The inflow to the hydropower plants will thus influence the electricity price, making it lower when there is an abundance of water and higher in the opposite situation.

Each Wednesday at 13:00 hrs, Svensk Energi, which is the Swedish power suppliers' trade organization, releases the amount of water as inflow energy concerning the previous week which has entered the Swedish hydropower plants dams (Svensk Energi, 2005 b). The corresponding information for the Norwegian hydropower plants is released by NVE (The Norwegian Water Resources and Energy Directorate) each Wednesday (NVE, 2005 a). By applying models which use forecasted discharge data as input the inflow energy can be estimated at a prior date.

1.2 Goals

The goals to be reached by this thesis are to:

- develop a statistical model which can be used to forecast the weekly hydropower inflow energy, in Sweden and Norway, from discharge measurements, by means of Neural Network or Multiple Linear Regression. To meet operational requirements, the model should have high skill and use as few discharge stations as possible.

- develop a user-friendly software for operational purposes, which forecasts the total weekly inflow energy using daily forecasted runoff data. The software will be used by SMHI (The Swedish Meteorological and Hydrological Institute).

1.3 Limitations

- Due to data availability, modeling with forecasted data, was performed only for Sweden.

- Only SMHI's discharge measurement stations have been used for Sweden.

- Errors in the discharge data due to measurement errors have not been taken into consideration.

2 Hydrology

In this chapter, the different basic hydrological processes are introduced along with different hydrological concepts that are important for understanding the modeling of the total weekly inflow energy. Techniques used for measuring discharge are also described to improve the understanding of the data that is used in the modeling process. In the end the hydropower and electricity market in the Nordic countries are described.

2.1 Hydrology and runoff processes

Hydrology is the science of water on the continents such as streams, lakes, groundwater, snow and ice. Within the science of hydrology, the chemical and the physical properties of water are also studied along with its hydrological movement.

2.1.1 Hydrological cycle

Water can never disappear. It is constantly moving and restored. This process is called the hydrological cycle and can be seen in Figure 2.1. The driving force behind the cycle is the sun. To describe the hydrological cycle it is easiest to start the cycle in the oceans as they hold about 97% of the water on Earth.

The radiation from the sun heats the water which in turn evaporates and rises from the ocean surface. During the rising process, vapor gets colder, condensates and forms clouds. When the weight of the condensed vapor is heavier than the strength of the rising currents, the water falls down as precipitation. The precipitation can fall on land or back to the oceans. If the precipitation falls into the oceans the process starts all over again. (Hamill, 2001, p.436-437)

Precipitation that falls on land will eventually also find its way back to the ocean but depending on which way it takes, the time spent will be different. Precipitation that falls on land can be stored in different water storages such as lakes, marshes, groundwater, glaciers and snow.

Precipitation that reaches the ground can form surface runoff, get stored as snow and ice or infiltrate through the soil surface. Surface runoff is formed by the water that can not infiltrate into the soil or when the precipitation falls with a higher intensity than the water can infiltrate into the soil. Surface runoff will only occur during very heavy rainfall or during the melting of the snow storage. The chance of occurring surface runoff increases when the soil has a low permeability. A soil that is frozen has very low permeability. Urban environments have a high percentage of impermeable soil, thus generating high surface runoff.

Water that infiltrates the soil surface percolates down through the unsaturated zone of the soil. If the water is not taken up by the vegetation it will eventually reach the groundwater, reach a stream and then be transported back towards the ocean.

Some precipitation never reaches the ground, instead it is captured by the vegetation. This water is called intercepted water and evaporates back into the atmosphere. During summer, as much as 29 % of the rain in a Scandinavian forest is intercepted and never reaches the ground (Ward and Robinson, 2000, p.75). Water that is taken up by the vegetation through the roots is released into the atmosphere by transpiration. There is also some evaporation of the water that has reached the ground. These three processes are normally grouped together and called evapotranspiration, since they are difficult to separate.

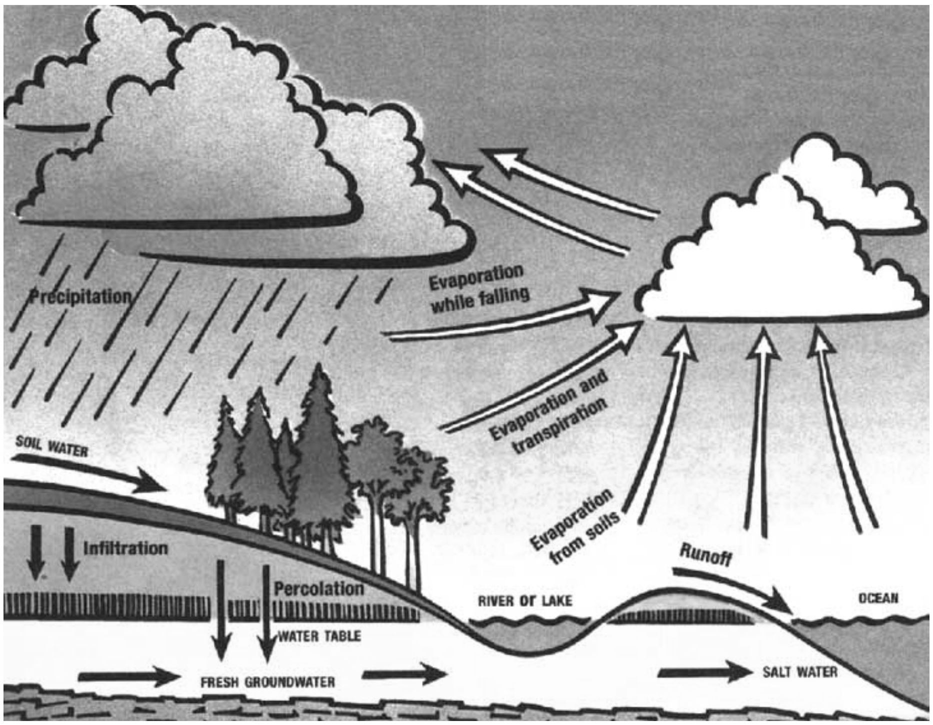

Figure 2.1. The hydrological cycle (FAO Document Repository, 2005).

2.1.2 Catchment area

In hydrology, the catchment area is a very important concept. To every stream a catchment area is connected. The catchment area consists of the area upstream a given point in the stream that contributes water to this certain point. The catchment area is limited by water dividers (Bergström, 2003, p.59). All precipitation that falls on the inside of the water divider will contribute with water to a certain point in the stream. Precipitation that falls outside the water divider will contribute with water to a point further down in the stream or to another stream. All points in the landscape belong to a catchment area (Grip and Rodhe, 1985, p.11).

2.1.3 Runoff

Runoff is the total amount of water that leaves a catchment area. The runoff can be either as surface runoff or as groundwater runoff. The runoff for a catchment area can be calculated from the water balance equation (e.g. Bergström, 2003, p.2):

$$P = Q + E \pm M \tag{2.1}$$

where P = the precipitation during a given time period
Q = the runoff during the same time period
E = the evapotranspiration during the same time period
M = the changes in the stored water in the catchment area

The unit is often given in mm and represents the water depth over an area, but other units such as m^3/s can also be used. In Sweden 6000 m^3 of water is transported into the surrounding oceans every second (Bergström, 2003, p.89).

2.1.4 Factors affecting the runoff

Many factors affect the runoff in a catchment. How the different factors affect the runoff is important to understand when analyzing the runoff pattern from a specific catchment. The most important factors are described below, together with the resulting effect they have on the runoff.

The topography decides the size of the catchment area by creating the different water dividers. The topography always creates the surface water divider and also influences the speed of the water movement. A steeper topography makes the water travel faster in the ground than a flat topography. This is why the groundwater often accumulates at the foot of a slope and makes the ground surface saturated. The topography will also influence the amount and type of precipitation that will fall in the catchment area. Orographic precipitation is very common along the western coast of northern Scandinavia and is a result of moist air from the ocean lifting over the mountain range. Often more precipitation falls on the windward side of the slope than on the leeward side. Since the dominating wind direction in Scandinavia is from the west, there is more precipitation falling in Norway than in Sweden because of the presence of the Scandinavian mountain range. (Ward and Robinson, 2000, p.20)

The shape of the catchment area also affects the runoff. A circular shaped catchment area has short but high runoff peaks after a rainfall since the runoff from different parts of the catchment area will reach the outlet at approximately the same time. A rectangular shaped catchment area has longer but lower peaks after a rainfall since the runoff from different parts in the catchment area will reach the outlet at different times. The time it takes for the rain, falling at the most remote part of the catchment, to reach the outlet is called the time of concentration. (Hamill, 2001, p.517)

The vegetation has a big influence on runoff as it affects the evapotranspiration, the infiltration capacity of the soil and the snowmelt. Different types of vegetations intercept different amounts of water from a rain or snowfall. In general, denser and taller vegetation results in higher interception losses than low and sparse vegetation. A forest has higher interception losses than grass or agricultural crops. (Ward and Robinson, 2000, p.74-79) Vegetation influences the evapotranspiration in both positive and negative ways. More vegetation is directly connected to an increase of the transpiration and interception losses. Evaporation is dependent on the wind velocity and the sun radiation. Vegetation reduces the wind velocity and blocks the sun radiation, thus reducing the evaporation. Vegetation will also keep the soil from getting compact and that will ease the infiltration of water through the soil. The snowmelt velocity will increase in an area with less vegetation since vegetation blocks the sun's radiation. (Ward and Robinson, 2000, p.103-124) Areas with little vegetation will therefore have higher and shorter runoff peaks during the snowmelt period.

The soil type is important for the percolation velocity of water through the soil in the unsaturated zone. If the soil is impermeable or thin, the rainfall will run quickly off the surface and very little percolation will occur. Grain size is the most important parameter for the percolation velocity. A soil that contains a high percentage of clay grains will have a slower percolation velocity than a soil with high percentage of larger grains. Deep and permeable soils can store large amounts of water and thus contribute to reduced runoff. (Ward and Robinson, 2000, p.185-187)

Lakes and marshes in the catchment area are important for the runoff because they have a great capacity to store water after a rainfall or during the snow melting period. If a large amount of the area in a catchment consists of lakes and marshes it will reduce the runoff by storing some of the water. (Bergström, 2003, p.72-73)

Glaciers located within the catchment will affect the runoff as they store water, in the form of ice, for a long time. Glaciers will start to contribute to the runoff at the same time as the snow starts to melt. The difference between the snow melting and the melting of the glaciers is that the glaciers will keep melting during the whole summer while the snowmelt period will only last for a few weeks in spring. During the summer, the contribution of melt water from a glacier can be a significant part of a catchment's runoff. The runoff from a glacier is not always even, outburst floods created by sudden release of large quantities of water stored within, under or alongside the glacier can occur. (Ward and Robinson, 2000, p.292-295)

Human impacts play an important role for the runoff. The hydrological cycle has evolved over millions of years, but recently the activities of humans have started to change the cycle. Human activities change the natural ways for the water movement which have caused a disturbance in the hydrological cycle. Urbanization has turned grass areas into concrete and roads. Dams have been built for hydropower, flood control or irrigation. This forces the water to diverge from its normal route. Artificial drainage creates new areas for agriculture where earlier marshes where located. Deforestation changes the conditions for evapotranspiration and the infiltration capacity of the soil. All these human activities have an influence on the catchment's runoff. (Hamill, 2001, p.440-443)

The Scandinavian hydrology conditions are characterized by Scandinavia's location in the temporized weather zone. Despite the closeness in location between Sweden and Norway there are differences in the hydrological factors that are affecting their runoff. Some of these hydrological factors are described below.

The weather in the northern parts of Sweden and Norway during the winter is cold and wet. In northern Sweden the snow storage is very important for the hydropower production. In Norway the snow storage, together with glaciers, play an important role for the hydropower production. Sweden has around 300 glaciers. The largest glacier in Sweden is called Stuorrajekna and has an area of about 13 km^2. The Swedish glaciers are too small to have any significant impact on the runoff. (Bergström, 2003, p.28-29) On the other hand, Norway has around 1600 glaciers. The largest, Jostedalsbreen, has an area of 487 km^2 (Till Topps, 2005). The glaciers in Norway are generally larger than in Sweden and in some areas they play an important role for the runoff.

The Norwegian topography is overall steeper than in Sweden and that makes most of the catchment areas small but the large altitude differences makes the landscape more favorable for producing hydropower. Another difference that the topography makes is what type of vegetation that will dominate the landscape. The lowlands of Sweden have a lot of forest compared to the highlands that covers most parts of Norway. Sweden also has more lakes than Norway. About 10% of Sweden is covered by lakes (Bergstöm, 2003, p.72). Lakes have a big effect on the runoff pattern, as can be read about above.

The northern parts of Sweden have a thin soil layer whereas the southern parts have a thicker soil layer. The soil layer in Norway is very thin all over the country. This means that the soil storage capacity in Norway is low and that will characterize the runoff pattern in most parts of the country.

2.2 Discharge measurement

This chapter describes several ways of measuring river discharge. Discharge is defined as the amount of water which is transported during a specific time period. In a river, it is the amount of water transported during a specific time period across a river section. The most commonly used units are m^3/s and l/s. Synonyms such as flow, flow rate and inflow are also used in this thesis.

Discharge measurement can be made directly or indirectly. Direct methods indicate that the discharge is measured at the site and indirect methods that discharge is inferred from another measurable variable. Figure 2.2 illustrates what a typical river cross section looks like and measurements typically made.

2.2.1 Direct methods

There are several different direct methods that can be used in order to determine the discharge.

Volumetric measurement is the most accurate method when measuring small flows in the range of 0 – 15 l/s, such as those from a spring. The water is collected in a beaker with known volume, V, the time T, which it takes to fill it is measured. The discharge, Q, is derived through the quotient V/T. (Gordon et al., 1999, p.157)

There are three different types of current meters which are normally used when measuring stream flow: propeller, cup and electromagnetic. The velocity is measured at various points of the river's cross section and then integrated along the cross sectional area to get the flow as can be seen in Figure 2.2. Both the propeller and the cup measurements use rotors that are turned by the passing fluid making it possible to calculate the water velocity. The *propeller-type meter* has a horizontal axis rotor. The size of the propeller can be changed to properly correspond to the ranges of the flow. The *cup-type meter* has a vertical rotor and is more sensitive to lower velocities and debris entanglement than the propeller-type meter.

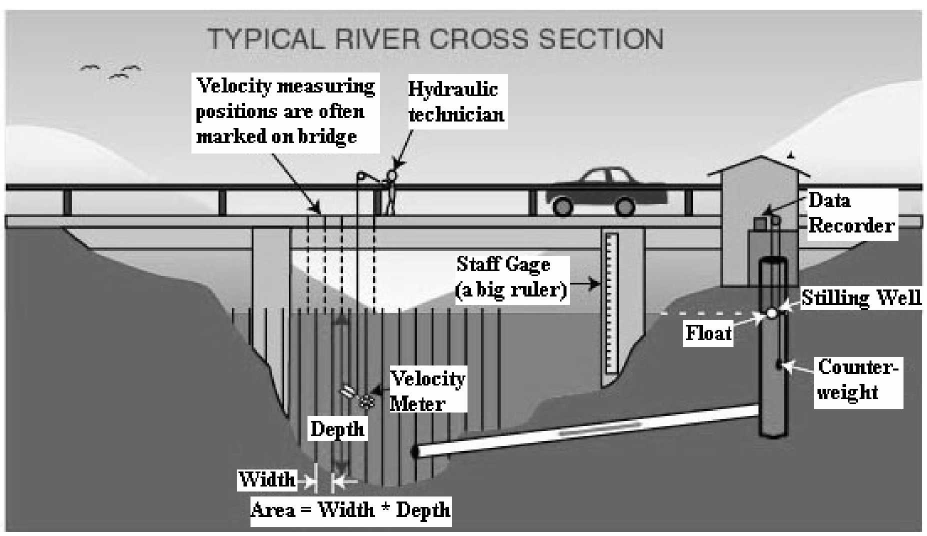

Figure 2.2. A typical river cross section (U.S. Geological Survey, 2005).

The principle of the *electromagnetic meters* is that the water flowing through them induces a voltage that is measured. These meters are most often used in oceanography since salty water has a higher conductivity than fresh water. If the electromagnetic meter is used in very pure water, the instrument has to be extremely sensitive. The advantage of these meters is that they can easily be used in places with vegetation, where the rotary meters can have problems. Current meters can be operated from a bridge, boat or from wading and are suitable when trying to determine the flow in rivers. (Gordon et al., 1999, p.159-161)

Dilution gauging methods can be used if the flow to be measured is very turbulent. Dye or salt with a known concentration is added and then monitored at some point downstream. By measuring the concentration of the dye or salt in the water at the monitoring point, the river flow can be calculated. (Gordon et al., 1999, p.163-164)

2.2.2 Indirect methods

One of the most frequently used methods when measuring discharge is the *stage-discharge method*. This method is used on sites that are frequently visited. The relationship between discharge and stream depth needs to be measured simultaneously for a large range of discharges. The discharge is measured with direct methods as described in Chapter 2.2.1. The stage is registered in a gauging station which is located close to the stream. This is illustrated as a ruler in Figure 2.2. The stage-discharge relationship is developed by plotting the measured stage against the measured discharge. When establishing the stage-discharge curve, logarithmic paper is preferred since the curves most often are parabolic and thereby will come out as a straight line as can be seen in Figure 2.3. This makes it easier to extrapolate the curve to points beyond the measured interval. The lower part of the rating curve may be plotted on arithmetic paper for improved accuracy and since the zero flow stage cannot be plotted on a logarithmic paper. When the stage-discharge curve has been made for a station along the stream, the discharge can be estimated for any given stage. The estimated discharge can contain errors due to measuring errors when establishing the curve and when determining the stage. (Bergström 2003, p.91-94 and Gordon et al., 1999, p.168-170)

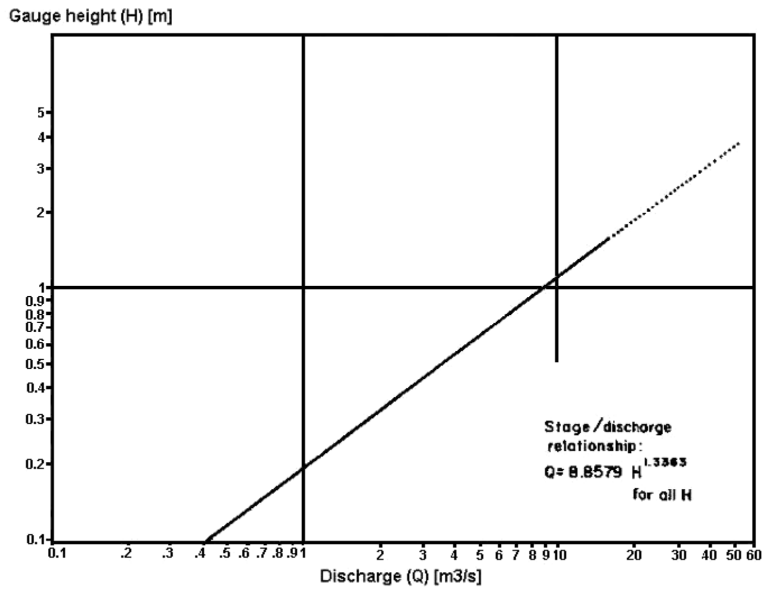

Figure 2.3. Stage-discharge curve (International Livestock Research Institute, 2005).

2.3 Hydropower

This chapter explains the fundamentals of hydropower, covering its history and the basic ideas behind it. This is followed by a description of some of the environmental consequences that a hydropower plant can cause as well as the extent of hydropower production in Sweden and Norway.

2.3.1 Short History

For thousands of years people have been using water to perform work of some kind. The Greeks used water driven wheels to grind wheat into flower about 2000 years ago. The power generated from water was also used to saw wood as well as powering textile mills and manufacturing plants.

The technology for creating hydropower from falling water has existed for more than a century. In the mid 1700s the French hydraulic and military engineer Bernard Forest de Bélidor started the evolution of the modern hydropower turbine. The breakthrough came in 1882 when an electric generator was coupled to the turbine which resulted in one of the world’s first hydropowered plants, located in Appleton, Wisconsin. (U.S. Department of Energy, 2005)

2.3.2 The principles behind energy extraction

The basic idea behind the hydropower production is to use the power of moving water to turn a turbine which is connected to a generator. Gravity is the force that makes this possible hence a height difference must be present to make the water flow. The theoretical power can be calculated according to the following equation (e.g., Bergström, 2001, p.106):

$$P = \rho \cdot g \cdot Q \cdot h \cdot \eta \tag{2.2}$$

where P = extracted mechanical power [W]
ρ = density of water [kg/m^3]
g = gravity constant [m/s^2]
Q = discharge [m^3/s]
h = pressure head [m H_2O]
η = efficiency of the turbine [%]

According to Equation 2.2, the power generation is directly proportional to the pressure head, h, as well as the discharge, Q. The pressure head is the elevation difference between the water level of the top reservoir and the inlet to the turbine. Due to the proportionality between the pressure head and the discharge the same amount of power can be extracted by having a moderate pressure head and a high discharge as by having a high pressure head and a moderate discharge (Bergström, 2001, p.106).

2.3.3 Hydropower regulation

Electricity can not be stored, it has to be consumed the instant it is produced. However, the water that generates the energy which creates the electricity can be stored. In order to properly utilize the hydropower plants the water has to be stored. This can be done in reservoirs or in lakes. The level in these will vary during the year since it is directly linked to the inflow. By storing water in the reservoir it is possible to control the energy production over time.

In Sweden, the peaks of the inflow occur during spring when the snow starts to melt. Another peak occurs in autumn, a period when rainfall is abundant. The situation for Norway is similar, except for streams that are glacier fed. In this case, the increase of runoff starts during spring, at the same time as the snow melts. The difference is that the glacier continues to feed water to the stream until autumn when the weather becomes colder and the glaciers freeze. Hydropower companies aim to manage their plants so that they can store water in the reservoir during periods of high inflow and extract it during winter, when the demands, as well as the prices are higher. (Svensk Energi, 2005 c and Vattenfall, 2005)

2.3.4 Consequences of hydropower

The establishment of a hydropower plant in a river has an environmental, as well as a social effect on its surroundings. A hydropower plant will change the water level and the flow both upstream and downstream of the dam construction as the flow becomes regulated by the opening or closure of the gates to the turbines and spillway. This will affect the people living in these places. Figure 2.4 shows the difference between the unregulated and the regulated flows of Luleälven in northern Sweden. The dam construction causes the peaks to be smaller and the flow to fluctuate more than in an unregulated river.

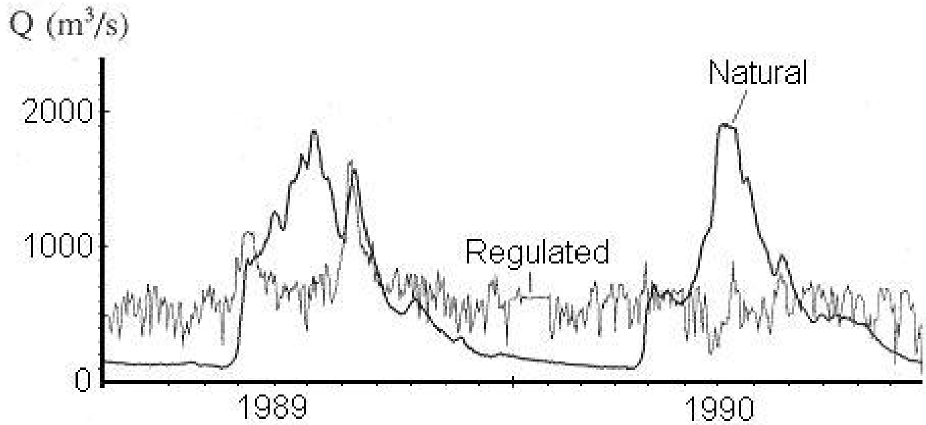

Figure 2.4. Comparison between the regulated- and reconstructed natural discharge in Luleälven between 1989 and 1990 (Bergström, 2001, p.112).

In some cases people living downstream of the dam have to move due to the risk of floods when the flow is very high. During periods with intense precipitation, it is sometimes not possible to store all the incoming water in the reservoir. The water that can not be stored has to be released to the river through the spillway without going over the turbines when the energy production can not be increased. This will cause the flow rate to increase and if this is more than the river can manage, the result will be flooding, affecting the people living close to the river (Bergström, 2001, p.112). These disasters can be mitigated by a well functioning dam management. Through runoff forecasts the operators can approximately know the magnitude of the inflowing water and can thereby release water from the dam earlier.

The environment is affected in many ways by a hydropower plant. The flow downstream the dam becomes smaller than before the regulation, causing problems for the river's habitat. Plants, insects, plankton and fish are all affected by water temperature and flow rate that the plant establishment will tend to change. A reservoir can act as very effective sediment trap making the water which is let out on the other side very clear. This water tends to be "sediment hungry" and can cause removal of fines immediately downstream of the dam. (Gordon et al., 1999, p.406-407) The dam construction makes it very difficult for the fish to wander upstream when they mate. Special arrangements such as fishways have to be made so that they can pass the construction.

However, there are positive aspects of having dams in the streams as well. During a period of heavy precipitation water can be stored in the dam's reservoir and thereby acting as a buffer by preventing all the water to flow downstream at once. The dam can thereby prevent possible flooding to occur. The dam can also be used as storage for water used for agricultural purposes.

2.3.5 Electricity production

Swedish and Norwegian rivers are generally suitable for hydropower production mainly due to their steepness. Many hydropower stations can be built in the rivers composing what is called a hydropower system.

The *Swedish* hydropower system produces about 65 TWh during an average year. The yearly deviation can however be quite high. During a dry year, with low precipitation, production can be as little as 50 TWh. On the other hand, in a wet year the production can be as high as 75 TWh. Most of the Swedish production is located in a small number of rivers which can be seen in Table 2.1. It shows the production distribution during 2003. Figure 2.5 shows the location of these rivers.

Table 2.1. Swedish energy production 2003 (Svensk Energi, 2005 c)

River	**Production [TWh]**
Lule älv	11,3
Skellefte älv	3,0
Ume älv	6,0
Ångermanälven	6,5
Faxälven	3,3
Indalsälven	8,3
Ljungan	1,5
Ljusnan	3,1
Dalälven	4,0
Klarälven	1,5
Göta älv	1,1
Other rivers	3,4
Total	**53,0**

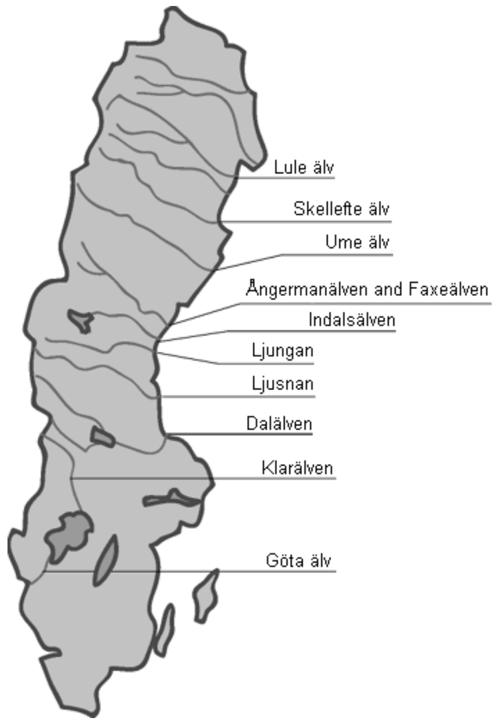

Figure 2.5. The Swedish rivers producing the most hydropower (Svensk Energi, 2005 d).

The Swedish system can still be further developed, with a total expansion potential of about 27 TWh. However, Torne älv, Kalix älv, Pite älv and Vindelälven are four rivers which are completely protected against further development. (Svensk Energi, 2005 c)

Norway has an average production of about 118 TWh per year. The production is however dependent on the amount of precipitation and can vary from year to year. During 1996 which was a very dry year, the Norwegian yearly production was as little as 104 TWh (Energibedriftenes Landsforening, 2005 a). On the other hand, in 1995, a wet year, the production was about 122 TWh (NVE, 2005 b). Table 2.2 shows the ten hydropower plants which had the highest production in Norway during 2001 (Energibedriftenes Landsforening, 2005 a).

Table 2.2. The ten most energy producing hydropower plants in Norway 2001 (Energibedriftenes Landsforening, 2005 a)

Station	**Production [MWh]**
Kvilldal	3 516,50
Tonstad	4 168,90
Aurland 1	2 406,80
Saurdal	1 291,00
Sy-Sima	2 074,70
Rana	2 122,90
Lang-Sima	1 328,90
Tokke	2 221,10
Svartisen	1 995,70
Brokke	1 407,00
Sum	**22 533,50**

The Norwegian system can, like the Swedish system be further developed. The potential which is economically and technically possible is about 187 TWh. (Enova, 2005)

2.4 Electricity exchange and Nord Pool

Electricity can not be stored and thereby has to be used directly after it has been produced. The electricity market is where this product is bought and sold making it possible to be properly distributed. Nord Pool is the Nordic power market where players from all countries can participate and buy electricity at the market price. The electricity prices are however often regulated within each country. This chapter explains what Nord Pool is and how it works. It also describes the benefits with having an integrated Nordic power market.

2.4.1 Power market deregulation – The beginning of Nord Pool

Norway was in 1991, the first Nordic country to deregulate their power market. In 1992 Statnett was established as a transmission system operator and shortly thereafter a tariff system was implemented. Statnett Marked, established by Statnett in 1993 was the first Norwegian power exchange market and made it possible for the Norwegian electricity consumer to freely pick their supplier.

In Sweden, the first step towards a deregulated power market was taken in 1993 when the state-owned Svenska Kraftnät was launched to manage the Swedish power network as well as foreign interconnects. On the 1st of January 1996, Sweden deregulated its power market and shortly after Svenska Kraftnät and Statnett Marked joined forces and Nord Pool was formed with the two parties as equal owners.

Finland was, in 1998, the third Nordic country to join the Nordic power exchange market. It was not until 2000 all of the Nordic countries were a part of the power exchange market when eastern Denmark joined, western Denmark had joined already in 1999.

At present it is not only the Nordic countries that are involved in the Nordic power exchange market but as well players from Great Britain, Germany, Belgium, USA, Switzerland, Italy and the Netherlands. (Nord Pool, 2005 c)

2.4.2 Spot market and financial market

Nord Pool mainly consists of two markets, the spot market and the financial market.

On the *spot market*, Elspot, the players are able to buy and sell physical power with hourly contracts for the following day. The deadline for submitting bids on Elspot for delivery the following day is at 12:00. The price calculation for Elspot is based on the balance between bids and offers from all the active market participants. There is a supplement for the Swedish and Finnish players called Elbas where bids can be placed up to one hour before delivery. The reason why the Elbas market was opened is that the time-span between the day's Elspot price-fixing and the actual delivery can be as much as 36 hours. During this time the consumption and sale situation might have changed and by using the Elbas market the player has a chance of making adjustments. The turnover for the spot market was 119 TWh in 2003 and 167 TWh in 2004. (Energibedriftenes Landsforening, 2005 b, Nord Pool, 2005 a and Svensk Energi, 2005 e)

Eltermin is the *financial market* where players are dealing with futures. A future is a contract between a seller and a buyer of delivery of a specified quantity of electricity at a certain price. This market gives the players a chance of setting the price for a purchase or a sale up to three years ahead of time. This can be a valuable tool when making calculations or budgeting for future revenues and expenses. The turnover for the financial market was 590 TWh in 2004. (Energibedriftenes Landsforening, 2005 b, Nord Pool, 2005 b and Svensk Energi, 2005 e)

2.4.3 An integrated Nordic power market

The Nordic power market has many advantages since the power plants in the integrated countries can be used in a more environmental and economic way. The reason for this is that plants that have a low variable cost for electricity production (cost of fuel, operation, maintenance, taxes and fees) such as hydropower and nuclear power are used before the plants with a higher cost. Since all electricity producing plants do not have to be used at the same time they are set into operation in a certain order. If the production from hydropower is sufficient to cover the demand, no other plants are used since it is the cheapest and most environmental method. If this is not enough, the nuclear power plants are used followed by condensation plants and gas turbines. (Svensk Energi, 2005 e) The environmental benefits are that the plants last taken into operation are condensation plants and gas turbines which are powered with fossil fuel which among other things contributes to global warming.

The electricity price on the market will reach a level which makes the most expensive electricity production at the moment to be beneficial. During periods when the production from the hydropower stations is not enough and thus other productions have started, the greatest profit margin lies for the hydropower producer. The spot price is thus closely linked to the availability of water as well as the demand of electricity. In Table 2.3 the average spot prices for the years 1996 to 2002 are displayed. (Svensk Energi, 2005 e) It is interesting to notice the higher prices in 1996, 2001 and 2002. These high prices are closely linked to the lack of water in the hydropower dams.

The winter 95/96 was very dry and cold. It resulted in a very high demand of electricity and no snow storage for reposition of the dams during spring of 1996. This combination made the spot price to rise. In 2001 and 2002 the prices were as well very high which was the result of drought conditions during the autumn that did not provide replacement of water in the dams in preparation for the high electricity demands during the winter.

Table 2.3. The average spot prices between 1996 and 2002 (Svensk Energi, 2005 e)

Year	**Spot Price [öre/kWh]**
1996	23,6
1997	14,6
1998	12,3
1999	11,8
2000	10,8
2001	21,4
2002	24,6

The electricity production differs considerably among the Nordic countries. Table 2.4 shows in what way the Nordic countries produced electricity in 2003. Almost all of Norway's electricity is generated through hydropower compared to Denmark where most of the electricity comes from thermal power such as combustion of gas and coal. The demand for wind power is, however, increasing since it is a more environmentally friendly way to produce electricity. Sweden and Finland have a similar combination of electricity sources using hydropower, nuclear power and thermal power. (Nord Pool, 2005 c)

Table 2.4. The Nordic countries' main electricity sources (Nord Pool, 2005 c)

Country	**Hydro Power (%)**	**Nuclear Power (%)**	**Wind Power (%)**	**Thermal Power (%)**
Norway	99	-	-	1
Sweden	40	49	-	11
Denmark	-	-	13	87
Finland	12	27	-	61

3 Data

In this chapter, the different types of data which are used for the modeling is described. The quality of the data considering measurement errors is not discussed since it was one of the limitations of the thesis. Differences between the Swedish and the Norwegian data are also described.

3.1 Data Description

The data used in this thesis consists of three different types. The first type is *daily runoff* measurements from different stations in Sweden and Norway. These measurements are performed with the stage-discharge method. The second type of data that has been used is *total weekly inflow energy* to all hydropower plants in Sweden and Norway, given in the energy unit TWh. The third and last type of data that has been used is *daily forecasted runoff* for different Swedish stations.

SMHI provided the daily runoff measurement*s* from their Swedish stations and NVE provided the daily runoff measurements from a selection of the Norwegian stations. The daily runoff data consist of measurements from 35 stations well distributed over Sweden and 34 stations over Norway. The measurements for Sweden are conducted over the period 1980 to 2004. The measurements for Norway are conducted over the period 1980 to 2003. For Sweden, the daily runoff was provided in both the unit of mm and m^3/s but for Norway only in m^3/s.

Data concerning the total weekly inflow energy to the Swedish hydropower plants is provided weekly by Svensk Energi. It is published on their internet site (www.svenskenergi.se) every Wednesday. A compilation of these data from 1980 to 2004 was provided to us by SMHI. The total weekly inflow data is transformed into the amount of electricity that can be produced by the Swedish hydropower stations using the amount of water that the inflow represents. In Norway, the equivalent data is calculated and provided by NVE (www.nve.no).

The Swedish measurements of the total weekly inflow energy follow the weeks of the year starting in the first week of 1980. That means that some weeks at the end of the year have some days from one year and some others from the following year. Due to this system each year contains either 52 or 53 complete weeks.

The only difference between the Swedish- and the Norwegian total weekly inflow energy data is that the Norwegian measurement system does not follow the weekly system. Between year 1980 and year 1996 the weekly measurements started on the 1st of January and then continued until 52 weeks had been measured. The remaining days of the year were not taken into account. In 1997, the Norwegian system changed and started to follow the weekly system but only using 52 weeks in every year. This means that during years which contain 53 weeks a whole week is not taken into account.

SMHI provided daily forecasted runoff for 12 stations in Sweden that were selected during the development of this thesis. Chapter 4.5 describes the details of the station selection process. The daily forecasted runoff data is simulated by using the HBV-model (Lindström et al., 1997). The HBV-model is a conceptual hydrological model used to simulate runoff from temperature-, precipitation-, and snow storage measurements as input for the areas where the runoff shall be simulated. The daily forecasted runoff was given in the unit m^3/s.

Before the data was used in the models, it was plotted and analyzed to determine its quality. Missing values were present in the daily runoff data for both Norway and Sweden. The total weekly inflow data had no missing values for the Swedish and Norwegian time series. The daily forecasted runoff data for Sweden had no missing values either.

For the development of the models, *daily runoff* data was used as *input data* and the *weekly total inflow* as the *target data* for the simulations. When the models should perform future simulations, *daily forecasted runoff* was used as *input data* for the model instead.

4 Modeling the total weekly inflow energy – Theory and Methodology

In this chapter different modeling processes and statistical concepts will be introduced, along with different evaluation parameters that will be of importance for understanding the modeling of total weekly inflow energy presented in this thesis. In order to easier understand the chapters concerning Neural Network and Multiple Linear Regression the methodology will be presented directly after the theory. Chapter 4.5 gives a detailed description about the methodology of the station selection.

4.1 Data pre-processing

This chapter describes different actions which should be performed before using the data for modeling. The reason for this pre-treatment is that the results for the simulations are improved.

The *Swedish runoff data* was available in two different units, mm and m^3/s. The data set given in m^3/s was more accurate and was thereby used. The available *Norwegian runoff data* was in m^3/s thus these values were used. In order to improve the results from the models, the data sets underwent a pre-processing.

The first step was to remove the seasonal fluctuation of the daily runoff data by extracting the daily mean from each value. After that, the runoff data was transformed into weekly values to fit the total inflow energy data. Following that, the data set was standardized and normalized. The following chapters give a detailed description of each step.

4.1.1 Seasonal Fluctuations

The deviations which the data shows from the normal seasonal values are called seasonal fluctuations. A good model is expected to model these fluctuations, i.e., how much the variable to be modeled is deviating from its average behavior. This concept can be extended also to other time scales. We applied it to the daily time scale of our runoff data.

The first step was calculating a mean runoff year, where each day of the year was given an average of all values available in our data set for that same day. This average year contains 366 days, including February 29. After that, the daily average was extracted from each day of the data set.

4.1.2 Daily values transformed into weekly values

The total inflow energy from hydropower to which the models' simulations are to be compared to is on weekly basis. Thus, the daily runoff data was transformed into weekly values following the weekly schemes described in Chapter 3.1, for the Swedish and Norwegian data. This was done by merely adding the daily runoff data into weekly values. Weeks containing missing values were removed, together with all the corresponding weeks in the other time series.

4.1.3 Standardization and normalization

Standardization of data makes comparison possible between different, unrelated data sets. A common way of standardizing a data set is by subtracting the mean of the data set from each data point and dividing by the standard deviation of the series. After standardization, a data set has a mean of zero and a standard deviation of one. The standardized data is sometimes referred to as Z-score and is calculated according to Equation 4.1.

$$Z_x = \frac{X_i - \overline{X}}{\sigma_x} \tag{4.1}$$

where

$$\overline{X} = \frac{\sum_i^n X_i}{n} \tag{4.2}$$

$$\sigma_x = \sqrt{\frac{\sum_i^n (X_i - \overline{X})^2}{n-1}} \tag{4.3}$$

and X_i = data point
$\overline{X}$ = mean of the data set
n = number of data points
σ_x = standard deviation

Normalization means that the data set is rescaled so that it will be normally distributed. Depending on which range is wanted the procedure can differ. Most commonly used procedures are the application of log or root to each of the data values.

A data range between minus one and plus one is often used in data modeling. This range is achieved by dividing each data point by the maximum value in the data set. Since the maximum value might be negative, the value has to be with an absolute sign (Internet FAQ Archive, 2005).

This procedure frequently improves statistical modeling as many of the modeling methods are developed considering that the data is normally distributed.

Each time series was standardized and normalized individually in order to have values between minus one and one, a standard deviation of one as well as a mean of zero.

4.2 Evaluation parameters

The following chapters treat different parameters which are used when interpreting the models' results. Often it is not enough to only use one evaluation parameter since it can be misleading. For the best result these parameters should be applied together with a visual inspection of the simulation.

4.2.1 Correlation coefficient

The Pearson-Moment Correlation Coefficient (r), or simply correlation coefficient as it is more often called, is a measure of the degree of linear relationship between two variables x and y. The correlation coefficient can take on any value between -1 and 1.

The sign before the correlation coefficient, plus or minus, indicates if the relationship is positive or negative. A positive correlation coefficient indicates that when one variable increases, the other variable increases as well. In the same way when one variable decreases the other one does the same. A negative correlation coefficient means that when one of the variables increase the other one decreases and vice-versa. In simple terms the correlation coefficient describes how well one variable follows the other one.

By taking the absolute value of the correlation coefficient the strength of the relationship can be established. A correlation coefficient of $r = 0.5$ indicates a stronger linear relationship than $r = 0.4$. In the same way a correlation coefficient of $r = -0.5$ shows a greater degree of linear relationship than $r = 0.4$. If the correlation coefficient is zero there is no linear relationship between the variables and correlation coefficients of $r = 1$ and $r = -1$ indicate a perfect linear relationship.

A scatterplot diagram can easily illustrate how the correlation coefficient changes as the linear relationship between the two variables is altered. When $r = 0$ the points scatter widely about the plot, almost forming a circle. As the linear relationship increases the points scatter in a more elliptical shape until they reach the limit, when $r = 1$ or $r = -1$ and all the points fall on a straight line. (Stockburger, 1996)

The correlation coefficient between x and y is commonly defined as:

$$r = \frac{\sum_{i}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i}^{n}(x_i - \bar{x})^2 \sum_{i}^{n}(y_i - \bar{y})^2}} \tag{4.4}$$

where n = number of values
y_i = observed value
$\bar{y}$ = observed mean value
x_i = observed value
$\bar{x}$ = observed mean value

4.2.2 R^2-value

The R^2-value is often called the coefficient of determination and is referred to as the proportion of variation of the variable explained by the model. The definition of R^2 is the ratio of the sum of squares explained by a regression model and the total sum of squares around the mean. The R^2-value can take any value between 0 and 1 where $R^2 = 1$ means that there is a perfect fit between the observed values and the values calculated by the model. The R^2-value can be calculated according to:

$$R^2 = 1 - \frac{SSE}{SST} = 1 - \frac{\sum_{i}^{n}(y_i - \hat{y}_i)^2}{\sum_{i}^{n}(y_i - \bar{y})^2} \tag{4.5}$$

where y_i = observed value
$\bar{y}$ = observed mean value
$\hat{y}_i$ = simulated value
n = number of values

4.2.3 Root mean square error

The root mean square error (RMSE) is a measure of the difference between two compared data sets. The RMSE can vary from zero to infinity with a perfect fit between the datasets having a RMSE of zero The RMSE between two data sets x and y can be calculated according to Equation 4.6.

$$RMSE = \sqrt{\frac{\sum_{i}^{n} (y_i - x_i)^2}{n}} \quad (4.6)$$

where n = number of values

y_i = observed value

x_i = observed value

4.2.4 Accumulated error and accumulated absolute error

The *accumulated error* is the summarized difference between the simulated value $\hat{x}$ and the true value x and is given by Equation 4.7.

$$\sum_{i}^{n} (\hat{x}_i - x_i) \quad (4.7)$$

The *accumulated absolute error* is the summarized difference between the simulated value $\hat{x}$ and the true value x, irregardless of positive and negative values and can be calculated according to Equation 4.8.

$$\sum_{i}^{n} \left|(\hat{x}_i - x_i)\right| \quad (4.8)$$

where $\hat{x}_i$ = simulated value

x_i = true value

n = number of values

4.3 Modeling methods - Neural Network

Neural Network is one of the tested modeling methods in this thesis. The following is a theoretical description of the modeling method and how it was applied.

The field of Neural Network has a history of some five decades but has only been used more frequently for the last 15 years, as the understanding for Neural Network along with the computer capacity has increased. Nowadays, Neural Networks are used in a numbers of different fields such as defense industry, manufacturing, entertainment, telecommunications, medical research and speech recognition along with many others.

Demuth and Beale (2004) present a comprehensive description of this modeling method, its use and applications based on MATLAB programming. We present here a simple compilation of this theory, based on their description. Demuth and

Beale (2004) is also the source of the figures presented in this chapter, unless specifically stated.

A Neural Network is composed by simple elements operating together. The inspirations to these elements come from the biological nervous systems. As in the biological nervous system, the network function is determined largely by the connection between the elements. A Neural Network can be adjusted to perform a different operation by changing the values of the connections between the elements. These values are called the *weights* and the elements are called *neurons*. Most Neural Networks are adjusted so that a particular input leads to a specific target output. The adjustment of the Neural Network is based on the comparison of the output from the Neural Network and the target. The adjustment of the weights in the Neural Network to achieve the target is called training. Today Neural Network can be trained to solve problems that are difficult for conventional computers or humans to solve. It should be pointed out that Neural Network performs better interpolations than extrapolations. This means that when Neural Network is presented with new input data which does not have the same characteristics as the one it has been trained with, it has difficulties to perform a good simulation.

4.3.1 Neuron Model

The elements that build up the Neural Network are called neurons. The simplest neuron contains a weight and a specific function. The operation of a neuron can be described as a scalar input p that is transmitted through a connection to the neuron, where it is multiplied by the weight w for this special connection. The scalar product wp is formed. In the simplest neuron construction wp is the only argument n for the transfer function f. The transfer function f produces the scalar output a. This procedure can be seen in Figure 4.1.

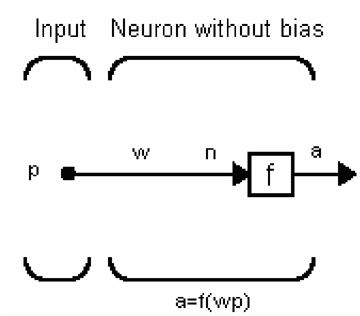

Figure 4.1. A neuron without a bias.

A bias b can also be added to the neuron. The bias b is simply a value added to the scalar product wp before the transfer function f. This sum now becomes the argument n for the transfer function f. This procedure can be seen in Figure 4.2.

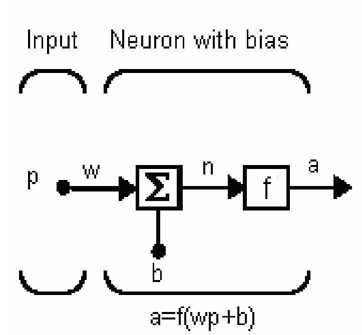

Figure 4.2. A neuron with a bias.

Both the weight *w* and the bias *b* are adjustable scalar parameters for the neuron. This is the central idea of the Neural Network, that such parameters as *w* and *b* can be adjusted so that the network performs in the way that is desired. Just as *w* and *b* can be adjusted to get the desired result, different transfer functions can be chosen. There are many different transfer functions to choose from. Three of the most commonly used and basic functions are described below.

The *hard-limit transfer function* limits the output from the neuron to either 0 or 1. If the argument *n* to the transfer function *f* is less then 0, the hard-limit transfer function's output *a* is 0. If the argument *n* is greater than or equal to 0 the transfer function's output *a* is 1. The hard-limit transfer function can be seen in Figure 4.3.

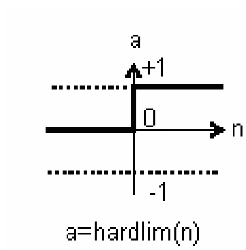

Figure 4.3. The hard-limit transfer function.

The *linear transfer function* is used when the problem that shall be solved has linear characteristics. The linear transfer function can be seen in Figure 4.4.

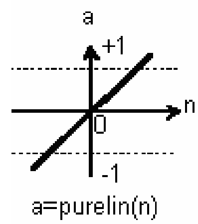

Figure 4.4. The linear transfer function.

The *tan-sigmoid transfer function* takes the argument *n*, which may consist of any values between plus and minus infinity and makes the output *a*, into the range of -1 to 1. The tan-sigmoid transfer function can be seen in Figure 4.5.

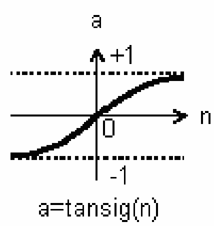

Figure 4.5. The tan-sigmoid transfer function.

One neuron can handle many different inputs at the same time. This will be easy to illustrate when the input is a vector. Presenting the input to a neuron as a vector is the most common way when using Neural Network. The individual input elements

$$p_1, p_2, p_3, \ldots\ldots p_R$$

are multiplied with the weight of its connection.

$$w_{1,1}, w_{1,2}, w_{1,3}, \ldots\ldots w_{1,R}$$

The sum of these products is called *Wp.*

$$Wp = w_{1,1} \cdot p_1 + w_{1,2} \cdot p_2 + w_{1,3} \cdot p_3 + \ldots\ldots + w_{1,R} \cdot p_R$$

If the neuron has a bias *b* it shall be added to *Wp* before entering the transfer function. This sum is the argument *n* for the transfer function.

$$n = Wp + b$$

This procedure can be seen in Figure 4.6.

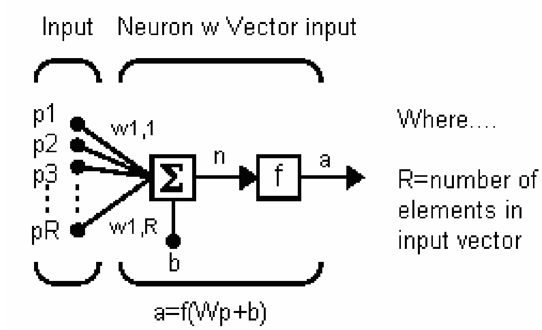

Figure 4.6. A neuron with a vector input.

4.3.2 Network Structure

A layer in a network includes the combination of weights, the multiplication and addition operations, the bias and the transfer function. However, the input to a network is not included in the layer. Two or more of the neurons can be combined in one layer and work together in parallel. In Figure 4.7 the structure of a layer is presented. In this network, each element of the input vector *p* is connected to each neuron through the weight matrix *W*. Every neuron has an adder that gathers the weighted inputs and bias *b* to form its own argument n_i. The different arguments n_i form an input vector *n* that is transformed by the transfer function *f*. At last, the neuron layer forms an output consisting of a column vector *a*.

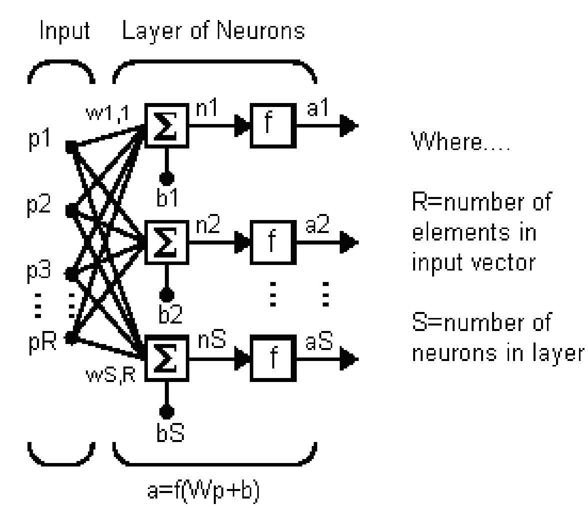

Figure 4.7. Layer structure.

To create a layer of neurons having different transfer functions *f*, simply put two of the networks shown above in parallel. Both networks will have the same input *p*.

A network can contain one or more layers. If a network consists of more than one layer, the different layers are located after each other. This means that the output *a* from one layer will be the input *p* for the following layer, this process is called *feed-forward*. Because of the feed-forward process, a distinction also has to be made between weight matrices that are connected to the input and weight matrices that are connected between the layers. Weight matrices connected to the input are often called *input weights* and weight matrices connecting layers are often called *layer weights*. If the layer consists of more than one neuron each layer has a weight matrix *W*, a bias vector *b* and an output vector *a*. A network structure, consisting of three layers with *S* neurons in each layer can be seen in Figure 4.8.

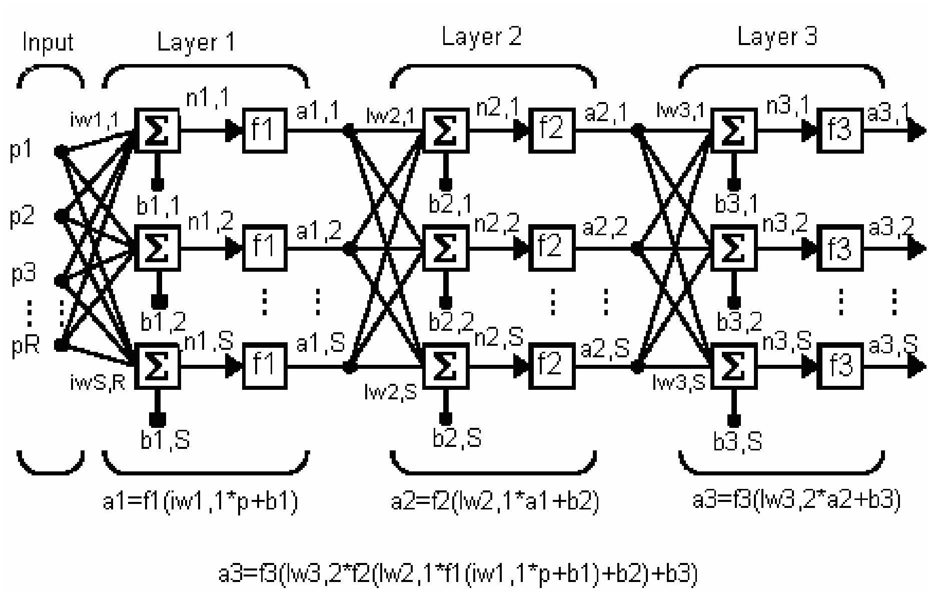

Figure 4.8. A three-layered feed-forward structure.

In a network it is common for the different layers to have different numbers of neurons. The layers in a multilayer network play different roles. The layer that produces the network output is called the *output layer*, the other layers are called the *hidden layers*. The three-layered network in Figure 4.8 therefore consists of two hidden layers and one output layer. Multiple-layer networks are powerful and can be used for many purposes in different fields.

4.3.3 Training the Neural Network

There are many different methods for training a Neural Network. They are all built on the same principal that by adapting weights and biases, the network can perform in a desired way. Basically all the different training methods can be divided into two different groups, *incremental* training and *batch* training. In incremental training the weights and bias are updated each time an input is presented to the network. In batch training the weights and bias are only updated after all of the inputs are presented to the networks.

The network training process has four steps.

1. Assemble the training data
2. Create the network object
3. Train the network
4. Simulate the network response to new inputs

One of the most used training methods, called *backpropagation*, is described below. The backpropagation training method is commonly used when training multilayer feed-forward networks. The reason backpropagation is often used is that Neural Networks trained with this method tend to give good results when presented to inputs that the Neural Network never has seen before, i.e., the Neural Network is able to do a good *generalization*. During the training, the weights and biases of the network are iteratively adjusted to minimize the mean square error between the network output and the target output. The equation to calculate mean square error is shown in Equation 4.9 (Mathworks, 2005).

$$mse = \frac{1}{R}\sum_{k=1}^{R}(t(k) - a(k))^2 \qquad (4.9)$$

where t = the target for a specific point
a = the output for a specific point
R = the number of inputs to the network

The updates of the network weights and biases during the backpropagation training are conducted in the direction in which the mean square error decreases most rapidly. The fastest decrease for the mean square error is given along the negative of the gradient. An iteration of this algorithm can be written according to Equation 4.10.

$$x_{k+1} = x_k - \alpha_k \cdot g_k \qquad (4.10)$$

where x_k = the vector of current weights and biases
g_k = the gradient
α_k = the learning rate

The learning rate can be set before the training starts and is a parameter that indicates how large every step between iterations shall be. If the learning rate is set too high the algorithm may become unstable. If the learning rate is too small the algorithm will take too many iterations to converge. There are several different backpropagation training algorithms to calculate the updates for the weights and biases, however there is no algorithm that is best suited for all types of problems.

A common problem that occurs when training the Neural Network is overfitting. This means that the Neural Network fits to the noise of the training set. As a consequence, the Neural Network does not perform well when a new set of data is provided to it, i.e., the Neural Network does not have a good generalization. This can be prevented by using *early stopping* as a part of the input data is set to be used for testing during the training. The training function can self-control its work and stop when the training starts to overfit.

The structure of a network is not completely based on the problem that shall be solved, as the amount of input data used for the simulation also affects the structure. A large amount of input data allows more neurons in the different layers to be used without risking overfitting. The exact number of neurons that gives the best result after the training is often found through *trial and error*. Also the amount of layers can differ depending on the amount of input data that is available. The number of neurons in the output layer is determined by the number of outputs required to solve the problem.

The efficiency of training increases if the input data is pre-processed before the training starts. In Chapter 4.1 the pre-processing of the input data is described.

4.3.4 Modeling with Neural Network

A Neural Network follows a specific routine when it performs a simulation. First the network conducts its training on a large part of the input data. When the training has stopped the Neural Network validates the training of its weights and biases by performing a simulation on the part of the input data that was not used for the training. The performance of the simulation is then checked by looking at how well the result from the Neural Network simulation for the validation period matches the real target data. Parameters used for the evaluation are correlation coefficient, R^2, RMSE, accumulated error and accumulated absolute error.

Before a good result can be simulated, the structure of the Neural Network has to be established, this is done by trail and error. There are three parameters that can be changed when the best structure of a Neural Network shall be established. The parameters are; type of transfer function, number of layers and number of neurons in each layer. The type of transfer functions that will be used is limited by the problem that is being solved. The amount of data gives an idea of how many layers and neurons that should be in the network. After these rough estimations the trail and error approach starts. For a Neural Network structure to be chosen it has to provide good and stabile simulation results.

The structure of the network that gave the best and the most stable results with our data sets was a feed-forward network with three layers. The first two layers contained five neurons in each layer and the transfer function for these neurons was a linear function. The output layer contained one neuron since the simulation only should give one result. The transfer function for the output layer was a tan-sigmoid function. The type of training function that gave the best result was backpropagation with adaptive learning rate. This network structure proved to be the best for both the Swedish and Norwegian Neural Network models.

4.4 Modeling methods - Multiple Linear Regression

Multiple Linear Regression is the other tested modeling methods in this work. The following is a theoretical description of the modeling method and how it was applied.

Multiple Linear Regression is another name for the *least squares* method. The purpose of Multiple Linear Regression is to model the relationship between two or more predictor variables (x_i) and a response variable (y) by fitting a linear equation to the observed data. All of the values in the independent variable x are associated with a value of the dependent variable y. In words the model can be described as:

$$\text{Data} = \text{Fit} + \text{Residual} \tag{4.11}$$

The "fit" term represents the expression $\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p$. The "residual" term represents the deviations of the observed values y from their estimation by the regression. The model deviation is denoted as ε. The common form of the Multiple Linear Regression is:

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + \ldots + b_p x_{ip} + \varepsilon_i \qquad \text{for i = 1, 2, ..., n observations} \tag{4.12}$$

In the least squares model, the best fit for the observed data is calculated by minimizing the sum of the square difference between the observed and the calculated value for each data point. Since the deviations are first squared and then added up, there will be no cancellations between positive and negative values. Like Neural Network, Multiple Linear Regression performs better interpolations than extrapolations.

The least squares estimates b_0, b_1, …., b_p, often referred to as regression coefficient, are determined following the principal that the difference between response variable y and the estimated response variable ŷ is minimum, i.e.

$$\sum (y_i - \hat{y}_i)^2 = \min \tag{4.13}$$

That means that its derivative in respect to each regression coefficient is zero.
Each of the regression coefficients is determined by solving the following matrix equation:

$$\begin{pmatrix} y_1 \\ y_2 \\ y_3 \\ : \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_{1,1} & .. & x_{1,k} \\ 1 & x_{2,1} & .. & x_{2,k} \\ 1 & x_{3,1} & .. & x_{3,k} \\ : & : & : & : \\ 1 & x_{n,1} & .. & x_{n,k} \end{pmatrix} \begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ : \\ b_k \end{pmatrix} \tag{4.14}$$

4.4.1 Modeling with Multiple Linear Regression

To perform a simulation with Multiple Linear Regression the input data has to be divided into two parts. The largest part is used to develop the model, i.e., to solve the linear system between the input data and the corresponding target data. A smaller part is reserved for the validation of the model. The solutions to the linear system are the regression coefficients that form the regression equation. By analyzing the regression coefficients it is possible to get an indication of how much each station contributes to the solution of the linear system.

The regression equation is then applied to the validation input set and its performance is checked against the validation output set using the evaluation parameters described in Chapter 4.2.

4.5 Selection of stations used as input data

A model is easier to maintain if it uses few stations as input data since this will make it easier to upgrade the model and less data has to be handled when running the model. With this in mind, a selection among the 35 runoff stations in Sweden and 34 in Norway was made so that a compromise between the number of input runoff stations and the quality of the model output could be found.

The selection was performed in two steps:

1) Comparing the correlation coefficients between the runoff data at each station and the total weekly inflow energy data. The stations that had correlation coefficient above 0.7 were selected to be used in the next step.

2) Running the Neural Network and Multiple Linear Regression models with different combinations of stations as input and analyzing their results.

The selection started by calculating the correlation coefficients between each runoff station and the total weekly inflow energy data series. The correlation coefficient provided a good indication of which runoff stations are representative for the total weekly inflow energy to the hydropower dams. This resulted in that 12 Swedish and 13 Norwegian stations were selected.

A *Neural Network* was trained separately for Sweden and Norway using the selected stations. The reduction was done using a sensitivity test that extracted one input station at a time, analyzed the changes in the output and compared the output to the target values. By comparing the evaluation parameters of each test, the stations that where the least representative for the weekly total inflow energy were eliminated. The elimination continued until a good compromise between the number of input stations and the skill of the model was found.

The Neural Network sensitivity test reduced the number of Swedish input stations to six. A similar process also reduced the number of input stations for Norway to six.

The station selection procedure for the *Multiple Linear Regression* model is easier than the one for the Neural Network model. When the linear system is solved during a simulation, the regression coefficients give an indication of the contribution of each input to the simulation. The station contributing the least to the simulation is eliminated and the result of the elimination can be seen by comparing the evaluation parameters before and after the elimination. This procedure is very straight forward. In the end a compromise is found between the number of stations that shall be a part of the simulation and the quality of the result of the simulation.

The Multiple Linear Regression model managed to reduce the number of Swedish stations to only six and still perform the simulations with good results. By further reducing the number of used stations the results degraded rapidly. The optimal balance between used stations and results was found by using six stations. Five of the Swedish stations that were chosen by the Multiple Linear Regression model were the same as for the Neural Network model, only one station was different. The reduction of Norwegian stations with the Multiple Linear Regression model resulted in the same stations as the Neural Network model.

5 Results and Discussion

This chapter starts with an analysis of the input stations selected for the modeling and is followed by the results from the modeling of Sweden and Norway. In the end, the modeling results using forecasted data are presented followed by a description of the software based on these results.

The modeling results presented in Chapters 5.2 and 5.3 are all from a validation period in the end of the time series. This period was not used in any step of the training of the Neural Network or in the calculation of the Multiple Linear Regression coefficients. In order to be certain about the model performance, over the entire time series other validation intervals were tested, one in the beginning of the time series and one using every fifth value. Since the differences in the results for the three validation periods are minimal, the results for the two last mentioned validation periods will not be presented.

5.1 Analysis of selected input stations

Through the station selection it became clear which areas in Sweden that were most representative for the total inflow to the hydropower plants. The total weekly inflow energy pattern to the hydropower plants is strongly characterized by a very prominent peak during the snowmelt period and low inflow during the winter when most of the precipitation is in the form of snow and is stored on the ground. Stations located south of Stockholm did not represent the inflow very well. The reason for this is that the most of the hydropower production is located in the north of Sweden which means that the total weekly inflow energy patterns over the year will be similar to the runoff pattern in this region. The runoff pattern between the south and north of Sweden is different because of the differences in the climate.

Stations in the north of Sweden that are located close to the coast were also not representative for the weekly total inflow energy. These stations had runoff patterns that were similar to the pattern of the total weekly inflow energy with peaks during the snowmelt period and little runoff during the winter, however, their snowmelt peak comes too early in the year to fully represent the inflow pattern that is significant for the whole hydropower production. Since the inflow peak during the snowmelt period is very significant for the total inflow, it is important that the runoff peak from the stations match the total inflow peak in time. The reason that the stations close to the coast do not match their peaks with the total weekly inflow energy is that these stations are located on a lower altitude and the closeness to the coast also provides a warmer climate than in the interior. This makes the snowmelt peak come too early in the spring. The closeness to the sea also makes the runoff from these stations go directly into the sea instead of into the dams.

The stations located in the interior of northern Sweden were the stations that best represented the total inflow to the hydropower production and can be seen in Figure 5.1. The main reason for this is that most hydropower plants are located here and the runoff pattern for these stations will thereby be similar to the total inflow pattern. In this part of Sweden there are many rivers suitable for hydropower production. Larger altitude changes in the topography in the interior rather than close to the coast also makes it more cost-effective to locate the hydropower plants in this region.

The possibility to produce hydropower in Norway is different than in Sweden. More precipitation falls in Norway and the topography is better suited for hydropower production. Another difference is that Norway has large glaciers that in some parts affect the runoff pattern throughout the year. The hydropower production in Norway is divided into three areas. These three areas can be seen in Figure 5.1. Most of the hydropower is produced in area 1 and 2 (NVE, 2005 c). Hydropower plants are located all over Norway, from the very south to the very north. Since the hydropower production in Norway is spread throughout the country the total inflow pattern is not as distinguished as in Sweden. Also the large glaciers contribute to the total inflow pattern since they provide runoff during the whole summer in the parts where they are located. Because of this, many stations are representative for the total inflow. Most of them are located in area 1 and 2 but there are also stations in the north that are representative for the total inflow. The stations most representatives for the Norwegian hydropower production can be seen in Figure 5.1. As can be seen here stations from all three areas are represented.

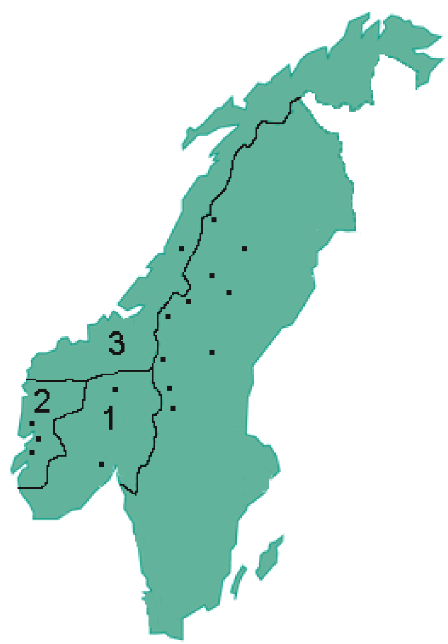

Figure 5.1. The locations of Swedish and Norwegian stations used for modeling.

5.2 Modeling

This chapter presents the results from the modeling of the total weekly inflow energy to the hydropower dams using Multiple Linear Regression and Neural Network. These results were used to determine which of the two models is most convenient to be used in the operational system developed in this thesis.

5.2.1 Sweden

The results of the Multiple Linear Regression and the Neural Network models for Sweden are shown in Figures 5.2 and 5.3.

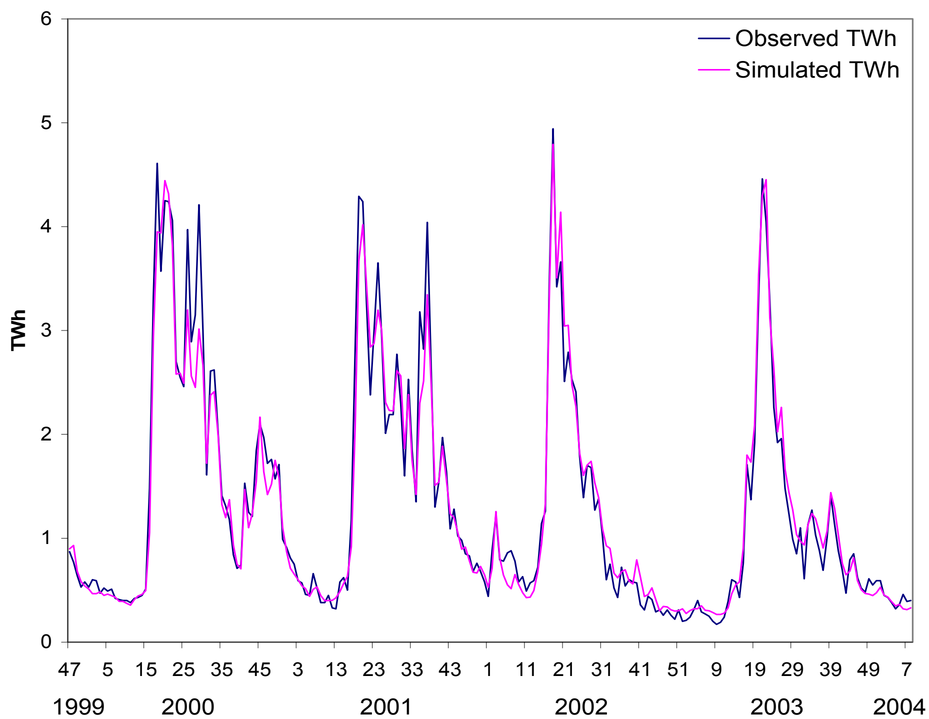

Figure 5.2. Validation of the Swedish Multiple Linear Regression model.

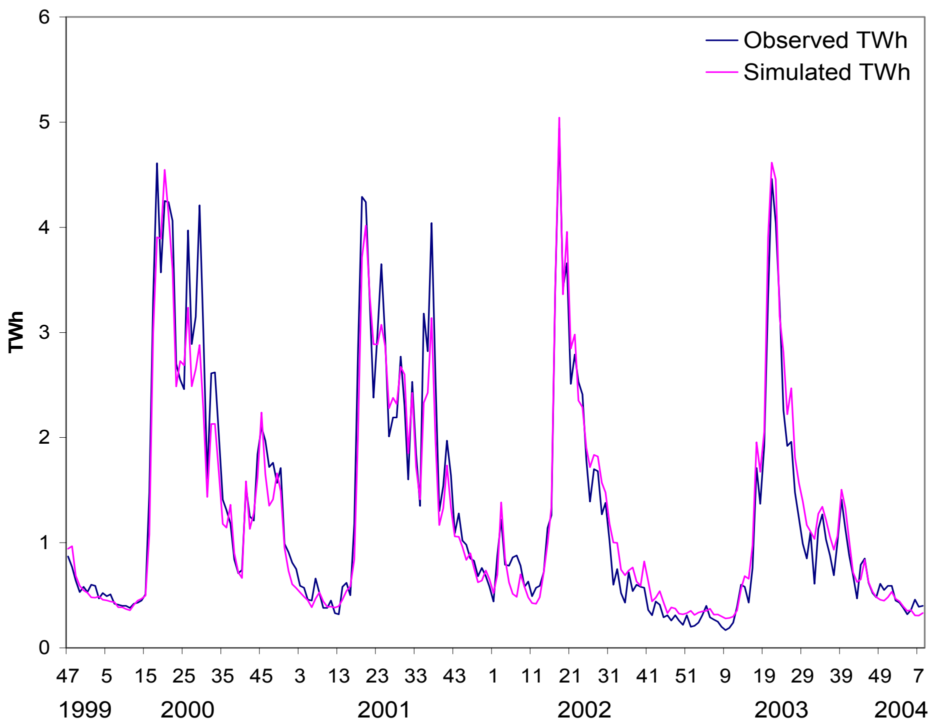

Figure 5.3. Validation of the Swedish Neural Network model.

A visual review of these figures shows little if any differences between the two models. In order to get an idea of their differences, it is easier to look at Table 5.1 which compares the evaluation parameters for the two models.

Table 5.1. Evaluation parameters for the Swedish Multiple Linear Regression- and Neural Network models, validated between 199947 and 200408

Model	**Multiple Linear Regression**	**Neural Network**
Correlation coefficient	0,98	0,97
R^2	0,96	0,95
RMSE	0,22	0,26
Average accumulated error [TWh]	-0,91	-1,06
Average accumulated absolute error [TWh]	7,91	9,74

As seen in Figures 5.2 and 5.3, the visible differences between the model results are very small. When evaluating the parameters in Table 5.1 it can be seen that the Multiple Linear Regression model gives better results than the Neural Network model. This indicates that the relationship between the runoff and the total weekly inflow energy is almost linear. This linearity is also evident if we consider that linear transfer functions were used in the Neural Network model.

The average accumulated and average accumulated absolute errors have been calculated by comparing the weekly simulations to the corresponding target. These weekly errors have been added up to yearly values. From these yearly values, an average value has been calculated. The years 1999 and 2004 have not been included since these years were not fully included in the validation period. However, these years have been used when calculating the correlation coefficient, R^2-value and RMSE.

Performing simulations with Neural Network is more complex than with Multiple Linear Regression. If the relationship between the runoff and the total weekly inflow energy is almost linear it is easier for the Multiple Linear Regression to perform a better simulation than the Neural Network. Since the chosen model also shall be performing simulations in an operational system, it is also an advantage if the model is easy to maintain and update. In the perspective of constructing a user-friendly software the Multiple Linear Regression has an advantage over the Neural Network.

Due to better simulation results and simplicity the Multiple Linear Regression model is chosen as the model to be used in the software developed in this thesis. These results are found in Chapter 5.3.

5.2.2 Norway

The outcome of the Multiple Linear Regression- and Neural Network model for Norway can be seen in Figures 5.4 and 5.5.

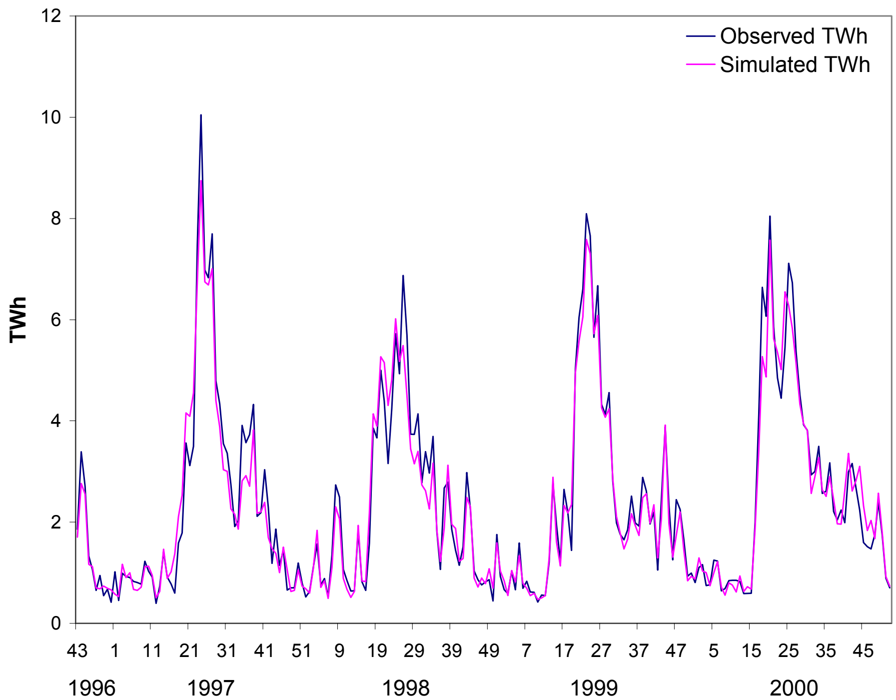

Figure 5.4. Validation of the Norwegian Multiple Linear Regression model.

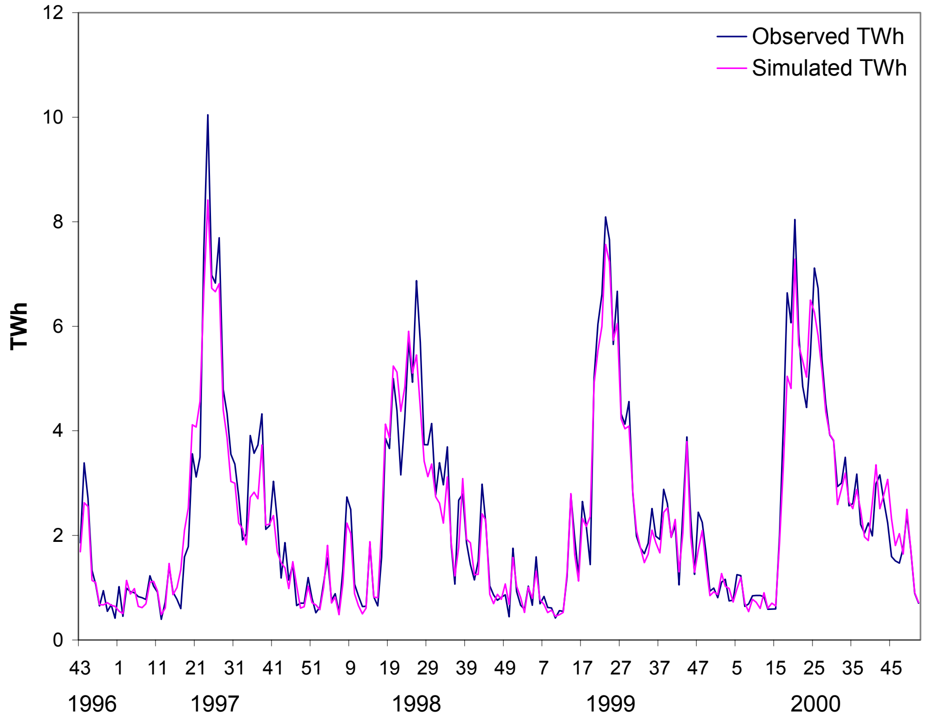

Figure 5.5. Validation of the Norwegian Neural Network model.

The two figures are extremely alike and differences are hard to see by looking at them. In order to really see the deviations it is necessary to look at the evaluation parameters which can be seen in Table 5.2.

Table 5.2. Evaluation parameters for the Norwegian Multiple Linear Regression- and Neural Network models, validated between 199643 and 200052

Model	**Multiple Linear Regression**	**Neural Network**
Correlation coefficient	0,98	0,98
$\mathbf{R^2}$	0,95	0,95
RMSE	0,42	0,42
Average accumulated error [TWh]	-3,85	-5,32
Average accumulated absolute error [TWh]	15,52	16,28

The goal with the modeling for Norway was never to perform simulations with forecasted data. This was one of the limitations for this thesis. The goal for the Norwegian investigation was to determine which model would perform the best simulations with only observed data. When closely looking at the parameters in

Table 5.2, the Multiple Linear Regression model gives slightly better results than the Neural Network model.

The average accumulated and average accumulated absolute errors have been calculated by comparing the weekly simulations to the corresponding target. These weekly errors have been added up to yearly values. From these yearly values, an average value has been calculated. The year 1996 has not been included since this year was not fully included in the validation period. However, it has been used when calculating the correlation coefficient, R^2-value and RMSE.

If simulations were to be performed with forecasted data for Norway, the Multiple Linear Regression model would have been chosen. The Multiple Linear Regression model is simpler than the Neural Network model and performs better results.

The conditions for hydropower production in Norway is different compared to the Swedish ones. This means that even if one type of model is the best in Sweden it does not necessary has to be the best in Norway as well. The topography and the larger amount of precipitation in Norway give different conditions for producing hydropower. Many areas in Norway where hydropower is produced are also influenced by glaciers. All these differences in the conditions may indicate that it is not the same model that performs the best result in both countries. A Neural Network model can be trained to learn any relationship between the input data and the target data whereas a Multiple Linear Regression model only can simulate a linear relationship.

5.3 Forecasting

In this chapter the results from the modeling performed with forecasted daily runoff data will be presented. This modeling was only performed for Sweden since this was one of the limitations for this thesis.

5.3.1 Input data

Since the Multiple Linear Regression model was chosen for the development of the operational software, the station selection using daily forecasted runoff was only performed using this model. The same selection procedure described in Chapter 4.5 was followed. The selection of stations using daily forecasted runoff resulted in seven stations. The six stations selected using daily runoff data were also selected in this case plus one additional station which further improved the result. The location of all the stations used in the Swedish and Norwegian models can be seen in Figure 5.1.

The forecast of the total weekly inflow energy was made by using daily forecasted runoff data as input to the Multiple Linear Regression model without changing the simulation procedure in any way. The difference is that instead of only using observed daily runoff, a combination of observed and forecasted daily runoff is used as well as only using forecasted daily runoff. There are thus eight different combinations of observed and forecasted runoff data that should be modeled depending on which day of the week the simulation should be performed. If the simulation shall be performed on a Monday, only forecasted runoff will be available which means that this simulation can only be performed using forecasted runoff. On Tuesday, observed runoff for Monday will be available and the simulation can be performed with one observed runoff value and six forecasted runoff values. For every day of the week that proceeds, one more observed runoff value can be used. On Monday the following week that the simulation shall be performed for, the simulation can be performed with only observed runoff. Figure 5.6 shows the combination of observed and forecasted values for a week of interest. The combination is dependent on which day the simulation is performed.

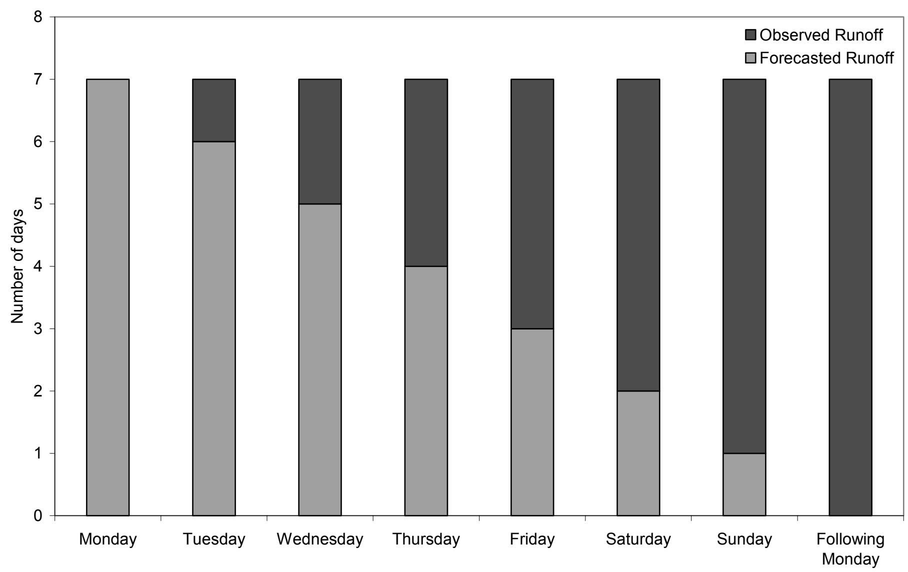

Figure 5.6. The eight combinations of observed- and forecasted days used when simulating the total weekly inflow energy.

5.3.2 Modeling the total weekly inflow energy

The outcome of two of the eight Multiple Linear Regression models can be seen in Figures 5.7 and 5.8. They use one respectively seven days of forecasted runoff data. The reason for just including two of these eight models is that the differences between the eight models' results are minimal. The eight models represent the different simulations that are performed depending on the number of forecasted days used.

The regression coefficients in the regression equation (Equation 4.14) were calculated using only observed values. The reason for this is that when the model has to be run in reality it will be much easier to perform a simulation for the software if it can use the same regression coefficients independently of how many days are forecasted or observed. Before it was chosen to use only the observed values' coefficients, simulations were performed using the coefficients corresponding to the number of forecasted days. The results from these simulations were compared and only minor differences were noticed and are thus not shown. Due to these minor differences and that by using only one set of coefficients the model would be simpler, it was chosen to only use the observed values' coefficients in the model.

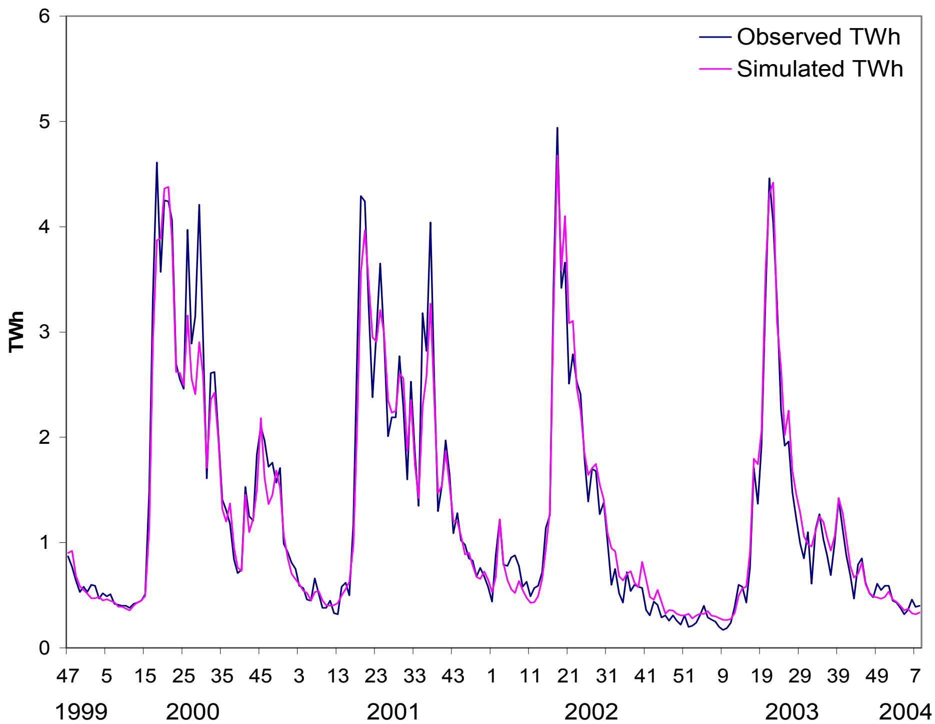

Figure 5.7. Validation of the Swedish Multiple Linear Regression model using one day's forecasted value.

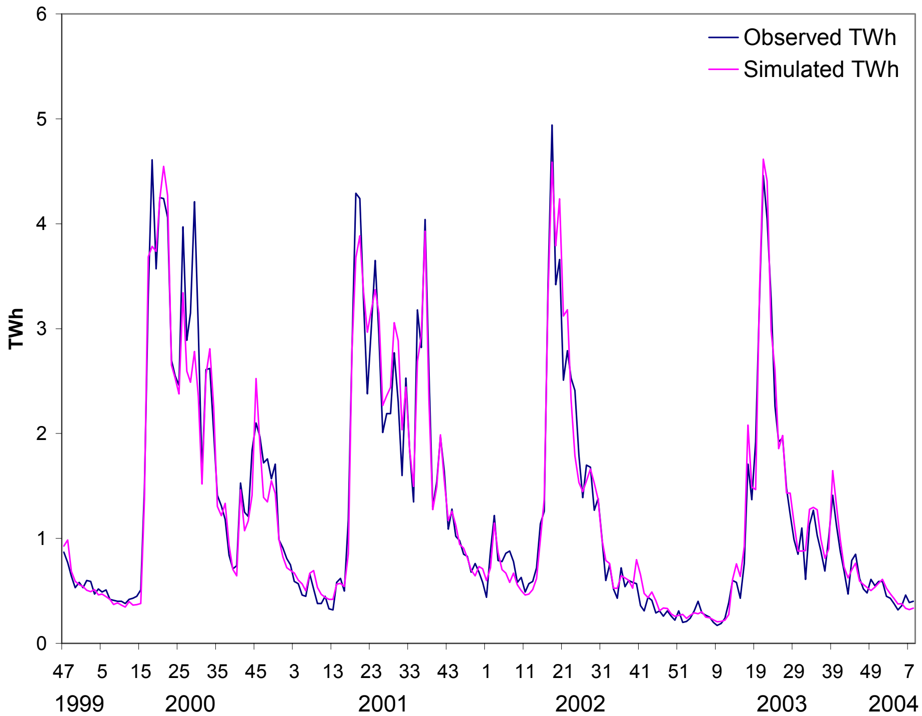

Figure 5.8. Validation of the Swedish Multiple Linear Regression model using merely forecasted values.

The easiest way of seeing the differences between the models is by looking at the evaluation parameters which can be seen in Table 5.3.

Table 5.3. Evaluation parameters for the eight Multiple Linear Regression models, validated between 199947 and 200408

Number of forecasted days	0	1	2	3	4	5	6	7
Correlation coefficient	0,98	0,98	0,98	0,98	0,98	0,98	0,98	0,98
$\mathbf{R^2}$	0,96	0,96	0,96	0,96	0,96	0,96	0,96	0,96
RMSE	0,22	0,23	0,23	0,23	0,22	0,22	0,22	0,23
Average yearly accumulated error [TWh]	-0,91	-0,90	-0,92	-0,87	-0,71	-0,56	-0,36	-0,21
Average yearly accumulated absolute error [TWh]	7,91	8,26	8,20	7,91	7,64	7,63	7,72	7,89

It can easily be explained why Figures 5.7 and 5.8 are so similar by looking at the parameters. The correlation coefficients and the R^2-values are exactly the same for all the eight models, furthermore they are very high. The RMSE is low for all the models and the minor deviations are insignificant. It is important to remember that the evaluation parameters should be analyzed together with a plot such as Figure 5.7 in order to be able to fully determine the result from the simulation. To only analyze the evaluation parameters is not enough since sometimes a bad simulation can have good evaluation parameters.

The average accumulated error tells us how much the models over- or underestimates on a yearly basis and is about one percent of the yearly electricity production from hydropower. This can be compared to the average accumulated absolute error which does not let the errors even each other out since it does not consider over- or underestimations.

The yearly average accumulated absolute error for the eight models is about ten percent of the yearly electricity production from hydropower. This means that around nine percent of the error is evened out every year for the average accumulated absolute error. Both these errors have been calculated by comparing the weekly simulations to the corresponding target. These weekly errors have been added up to yearly values. From these yearly values, an average value has been calculated. The years 1999 and 2004 have not been included since these years were not fully included in the validation period. However, these years have been used when calculating the correlation coefficient, R^2-value and RMSE.

When looking closely at the average accumulated error and the average accumulated absolute error values for the eight different models displayed in Table 5.3, it can be seen that they tend to improve with the number of simulated days used. Figure 5.9 shows the weekly error for the eight models during year 2002 and gives an indication why the errors tend to get better.

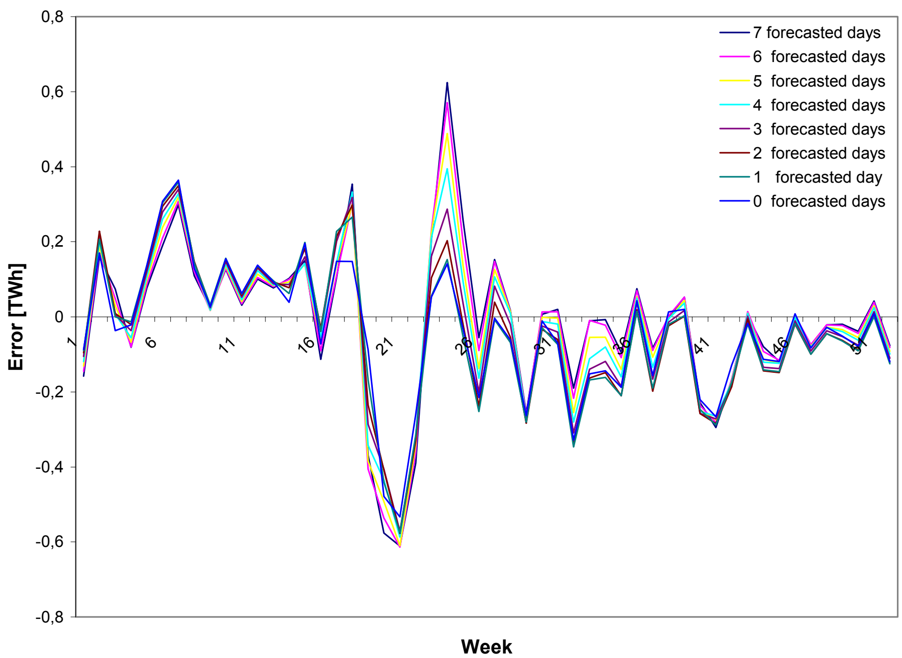

Figure 5.9. Weekly errors for the eight Swedish models for simulations errors performed year 2002.

This figure shows that the eight models more or less follow the same pattern throughout the year. However, there are a few deviations where the models using forecasted runoff seem to overestimate their results. On average, when the models overestimate their results, the overestimation is larger for the models using more days of forecasted runoff. This can easily be seen by looking at the large positive peak in approximately the middle of the time series. At this point the model using only observed runoff gives an error of about 0.15 TWh. The more forecasted runoff days used the larger this error gets, culminating at an error of about 0.6 when using merely forecasted runoff. This pattern can as well be seen at different points throughout the time series. Since the models on a yearly average underestimate their simulations, this can explain why the models using forecasted runoff give a better average yearly accumulated error. The reason for this is that when calculating the accumulated error the positive- and negative errors can eliminate each other resulting in a very low accumulated error.

5.4 TWh-Simulator

The TWh-Simulator is the software which was the final result of this thesis. The TWh-Simulator can make forecasts of the total weekly inflow energy to the Swedish hydropower plants. By using forecasts of the runoff at various locations in Sweden it instantly gives an estimate of what the total inflow energy to the hydropower plants will be at the end of the week. The only thing that constantly has to be updated is the database containing the forecasted runoff. The software has an interface which is easy to understand and use. The interface can be seen in Figure 5.10. By selecting the week for which the simulation should account for, selecting the day for which it should be run, a simulation for the week will be presented together with the true values for the nine previous weeks. A detailed user's manual can be found in Appendix.

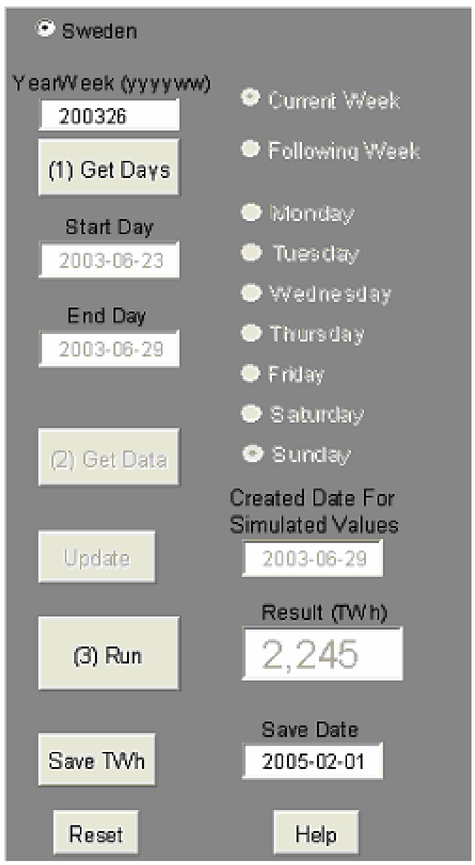

Figure 5.10. TWh-Simulator interface.

5.4.1 Using the TWh-Simulator

When the TWh-Simulator is run it uses as many observed values as possible. Depending on which day of the week it is run, the number of observed and forecasted values varies. However, the TWh-Simulator has to have values either observed or forecasted for every day in the week of interest. SMHI makes a 10-day forecast every day of the runoff which is used for the days where observed data is not available. Below it is shown in detail how the simulations for a week of interest are made where **k** indicates the week of interest, **k-1** the previous week and **k+1** the following week.

- *Thursday*$_{\mathbf{k-1}}$ through *Monday*$_{\mathbf{k}}$: The first five simulations for week **k** can be performed using solely forecasted data.
- *Tuesday*$_{\mathbf{k}}$: The sixth simulation for week **k** can be performed using observed data for Monday$_{\mathbf{k}}$ and forecasted data for the remaining days of the week.
- *Wednesday*$_{\mathbf{k}}$: The seventh simulation for week **k** can be performed using observed data for Monday$_{\mathbf{k}}$ and Tuesday$_{\mathbf{k}}$ and forecasted data for the remaining days of the week.
- *Thursday*$_{\mathbf{k}}$: The eighth simulation for week **k** can be performed using observed data for Monday$_{\mathbf{k}}$, Tuesday$_{\mathbf{k}}$ and Wednesday$_{\mathbf{k}}$ and forecasted data for the remaining days of the week.
- *Friday*$_{\mathbf{k}}$: The ninth simulation for week **k** can be performed using observed data for Monday$_{\mathbf{k}}$, Tuesday$_{\mathbf{k}}$, Wednesday$_{\mathbf{k}}$ and Thursday$_{\mathbf{k}}$ and forecasted data for the remaining days of the week.
- *Saturday*$_{\mathbf{k}}$: The tenth simulation for week **k** can be performed using observed data for Monday$_{\mathbf{k}}$, Tuesday$_{\mathbf{k}}$, Wednesday$_{\mathbf{k}}$, Thursday$_{\mathbf{k}}$ and Friday$_{\mathbf{k}}$ and forecasted data for Saturday$_{\mathbf{k}}$ and Sunday$_{\mathbf{k}}$.
- *Sunday*$_{\mathbf{k}}$: The eleventh simulation for week **k** can be performed using observed data for Monday$_{\mathbf{k}}$, Tuesday$_{\mathbf{k}}$, Wednesday$_{\mathbf{k}}$, Thursday$_{\mathbf{k}}$, Friday$_{\mathbf{k}}$ and Saturday$_{\mathbf{k}}$ and forecasted data for Sunday$_{\mathbf{k}}$.
- *Monday*$_{\mathbf{k+1}}$: The twelfth simulation for week **k** can be performed using only observed data.

5.4.2 Data scheme

When using the software the data goes through a scheme in which the data is treated in different ways. This scheme can be seen in Figure 5.11.

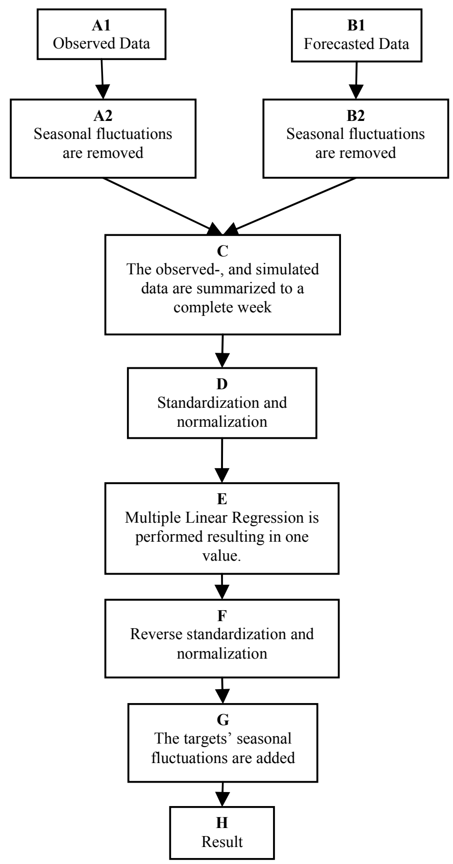

Figure 5.11. The data scheme for the TWh-Simulator.

A1: Current number of days' observed runoff data available is obtained from a database. This can vary between zero and seven days of data.

B1: The number of days' forecasted runoff data needed is obtained from a database. This can vary between zero and seven days of data.

A2 and B2: The seasonal fluctuations are subtracted from the observed and forecasted data. It is, however, not the same seasonal fluctuations that are used for the two datasets. They are thus calculated using observed as well as using forecasted values.

C: The observed and forecasted input values which add up to a complete week are made into one value for each station applying to the entire week.

D: The observed and forecasted input values are standardized and normalized.

E: Multiple Linear Regression is performed at which the regression coefficients are multiplied with the stations' values.

F: A reverse standardization and normalization is carried out using values calculated from the observed total weekly inflow energy.

G: The seasonal fluctuations for the observed total weekly inflow energy are added in order to transform the values back into the right unit, i.e. TWh.

H: The result is presented together with the nine previous weeks in a table as well as in a graph that can be seen in Figure 5.12.

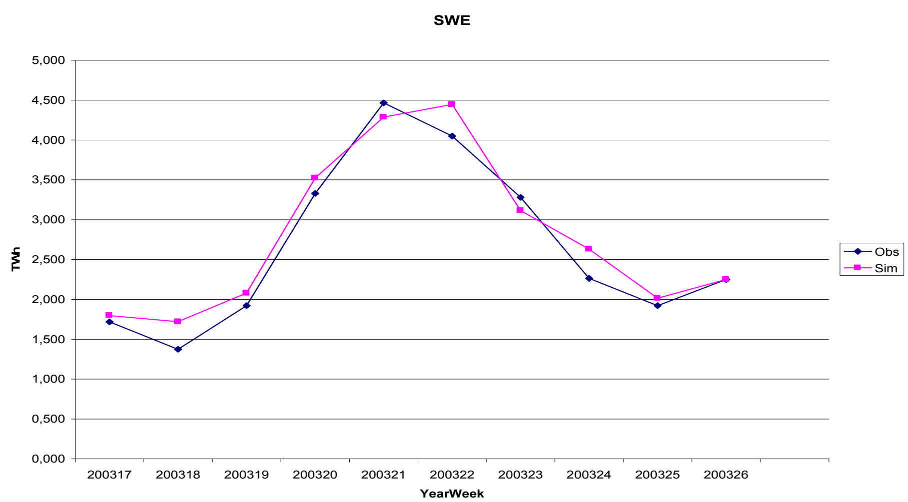

Figure 5.12. The result presented in a graph from a run with the TWh-Simulator.

6 Conclusions

The purpose of this thesis was to investigate which of the Neural Network and Multiple Linear Regression model that best can simulate the total weekly inflow energy to Swedish and Norwegian hydropower plants, by using daily runoff data from different measuring stations as input. After closely analyzing the results from these models Multiple Linear Regression was chosen since it performed better as well as it is easier to use and maintain. The Multiple Linear Regression model is further used to develop a user-friendly software that can be used operationally.

The Multiple Linear Regression model performed equal good simulations for the three different validation periods that were tested. This indicates that simulations made by the TWh-Simulator are reliable and thereby can be used for commercial purposes. In order to fully make use of the simulations it is however important to have knowledge about how the simulations are performed and the hydrological principles behind the model's development.

The locations of the selected stations coincide with the areas where the most hydropower is produced. Hence, the runoff measurements from these stations show similar patterns as the total inflow to the hydropower plants. This pattern is more obvious in Sweden since the hydropower production is concentrated to the northern parts. The inflow to the Norwegian hydropower plants is not as distinguished as the Swedish since the hydropower production is more evenly spread out over the country. The large Norwegian glaciers will also in some areas influence the runoff, thus also affecting the inflow to the hydropower plants.

In order for the TWh-Simulator to maintain its good simulation results its parameters should at some time be updated. This is because the parameters are based on a limited time series and might thereby not correspond to changes in the future.

9 References

Bergström, Sten (2001)
Sveriges Hydrologi – grundläggande hydrologiska förhållanden, fourth edition, SMHI and Svenska Hydrologiska Rådet, Norrköping.

Demuth, Howard and Beale, Mark (2004)
Neural Network Toolbox – for use with MatLab, User's Guide, the Mathworks Inc, Natick.

Gordon, Nancy D., McMahon, Thomas A. and Finlayson, Brian L. (1999)
Stream Hydrology – an introduction for ecologists, John Wiley & Sons Ltd, Chichester.

Grip, Harald and Rodhe, Allan (1985)
Vattnets väg från regn till bäck, Forskningsrådens förlagstjänst, Uppsala.

Hamill, Les (2001)
Understanding hydraulics, second edition, Palgrave publishers Ltd, New York.

Lindström, Göran, Gardelin, Marie, Johansson, Barbro, Persson, Magnus and Bergström, Sten (1996)
HBV-96 – En areellt fördelad modell för vattenkrafthydrologin, SMHI, Norrköping

Stockburger, David W. (1996)
Introductory Statistics: Concepts, Models and Applications WWW version 1.0
www.psychstat.smsu.edu/introbook/sbk00.htm (2005-01-12)

Ward, R.C and Robinson, M (2000)
Principles of hydrology, fourth edition, McGraw-Hill Publishing Company, Maidenhead.

WWW

Department of Statistics, Yale University, 2005
www.stat.yale.edu/Courses/1997-98/101/linmult.htm (2005-01-10)

Energibedriftenes Landsforening, 2005 a
www.energifakta.no/documents/vannkraft/prodkap.htm (2005-01-06)

Energibedriftenes Landsforening, 2005 b
www.energifakta.no/documents/oekonomi/kraftmarked/nordpool.htm (2005-01-06)

Enova, 2005
www.enova.no/?itemid=120 (2005-01-06)

FAO Document Repository 2005
www.fao.org/documents/show_cdr.asp?url_file=/docrep/U8480E/U8480E0c.htm (2005-02-09)

Gepsoft, 2005
www.gepsoft.com/gepsoft/APS3KB/Chapter09/Section1/SS04.htm (2005-01-13)

Henry, Neil W., 2005
www.people.vcu.edu/~nhenry/Rsq.htm (2005-01-13)

International Livestock Research Institute, 2005
www.ilri.org/publications/cdrom/integratedwater/IWMI/Documents/related_doucments/HTML/x5524e/x5524e0s.gif (2005-02-08)

Internet FAQ Archive – Online Education, 2005
www.faqs.org/faqs/ai-faq/neural-nets/part2/section-16.html (2005-01-14)

Linköpings Universitet, 2005
www.mai.liu.se/~annor/kurser/HSTA72/formel04.pdf (2005-01-13)

Mathworks, 2005
www.mathworks.com/access/helpdesk/help/toolbox/nnet/adapt107.html (2005-01-15)

NVE, 2005 a
www.nve.no/modules/module_109/publisher_view_product.asp?iEntityId=7842&iMenuId=1293&strBoxColor=FFFF99&strYAH=>NVE>%3Cb%3EEnergi%3C%2Fb%3E>Kraftsituasjonen (2005-01-10)

NVE, 2005 b
www.nve.no/modules/module_109/publisher_view_product.asp?iEntityId=1000&iMenuId=212&strBoxColor=FFFF99&strYAH=>NVE>%3Cb%3EEnergi%3C%2Fb%3E>Produksjon+og+forbruk>%C5rsoversikt (2005-01-06)

NVE, 2005 c
www.energistatus.no/el_produksjon/vannkraft_del1.htm (2005-02-22)

Nord, Pool 2005 a
www.nordpool.com/nordpool/spot/index.html (2005-01-06)

Nord, Pool 2005 b
www.nordpool.com/nordpool/financial/index.html (2005-01-06)

Nord, Pool 2005 c
www.nordpool.com/nordpool/group/index.html (2005-01-06)

Svensk Energi, 2005 a
www.svenskenergi.se/energifakta/elanvandning.htm (2005-01-12)

Svensk Energi, 2005 b
www.svenskenergi.se/kraftlaget/sverige.htm (2005-02-02)

Svensk Energi, 2005 c
www.svenskenergi.se/energifakta/vattenkraft.htm (2005-01-06)

Svensk Energi, 2005 d
service.svenskenergi.se/emma_demo/mapp2_produktion/2AVattenkraft/2Abild2_vattenkraft.gif (2005-02-08)

Svensk Energi, 2005 e
www.svenskenergi.se/marknad/fakta.htm (2005-01-06)

Till Topps, 2005
www.tilltopps.com/index.php?menu=9.1 (2005-02-11)

U.S. Department of Energy, 2005
www.eere.energy.gov/windandhydro/hydro_history.html (2005-01-04)

U.S. Geological Survey, 2005
ny.water.usgs.gov/pubs/posters/streamgagingposter.html (2005-03-01)

Vattenfall, 2005
www.vattenfall.se/om_vattenfall/energikunskap/vatten/ (2005-01-05)

Wikipedia, 2005
en.wikipedia.org/wiki/Climate (2005-01-14)

10 Appendix TWh-Simulator help file

1. Fill in the *YearWeek* (yyyyww) that the simulation should account for.

2. Press the *Get Days* button.

3. Select if the simulation should be performed the *current* or the *following* week.

4. If the simulation is performed the current week a *weekday* must be selected. If it is performed the following week this is not necessary since the model only uses observed values.

5. Press the *Get Data* button. If there are any values missing for the selected period these will be colored red. In order to run the model these values have to be updated. This is done by manually changing the values in the cells. Another option is to change the date for which the simulated values have been created. It is only possible to use prior dates. Change to a prior date and press the Get Data button again.

6. Press the *Update* button if the values have been changed manually. If no changes have been made this is not necessary.

7. Press the *Run* button. The result will be shown in the *Result Box* as well as in a table together with the nine previous weeks. These values are plotted and can be viewed in the Result sheet.

8. Press the *Save* button if the result is to be saved into the database.

9. Press the *Reset* button at any time to clear all the fields and start over.

Table A1. Color definition in the main page.

Color	Meaning
	Observed values
	Simulated values
	Missing values