CODEN:LUTEDX/(TEIE-5144)/1-73/(2000)

Industrial Electrical Engineering and Automation

An Initial Approach for Data Analysis of Production Losses in Nuclear Power Plants

Letizia Bagnoli

Dept. of Industrial Electrical Engineering and Automation
Lund University

Abstract

Analysis of production losses is useful for investigating and examining problems connected with a production process. It is a way to understand and improve the production process; to plan maintenance programs and to decrease the cost of production.

This report describes an initial approach to quantify and analyse the loss of production in nuclear power plants. The aim of this work is to improve the understanding of causes of production losses and to support analysis of the production as well as production planning at the plant. The work generates a basis for further development of the software that analyses and presents production data on the internet/intranet.

The work has been carried out in co-operation with Barsebäck Kraft AB, at a nuclear power utility for electricity generation. It consists of *maintenance and production data analysis*, *model development for different causes of production losses* and *software specifications of the data structure and user interface.*

Acknowledgements

First of all I want to thank the people that have been most engaged in this work, *Ulf Jeppsson (IEA), Pekka Skogberg (Barsebäck), Joan Dorrepaal (Bi-Cycle)* and *Gustaf Olsson (IEA).* You have helped me with all sorts of problems, showed a spirit of good will, and been a great support.

I also want to thank all people that I have met in Lund (professors, assistants and students), especially *Tobias* and *Per*. You were a great help in making feel at home and consequently in starting up this project, otherwise I would never have decided to stay one year in Sweden.

I want to thank all the people at Barsebäck. I have always been received with kindness and all information I sought has always been handed to me – one way or the other.

Last but not the least, I want to thank all my friends in Florence. Even if I am not with them, they never leave me alone.

Contents

1 Introduction

Traditional maintenance of large production facilities is to a large part carried out based on empirically defined time schedules and maintenance intervals. However, some surveys indicate that 70% of the failures that appear within industrial production processes are not related to failures of different subsystems (sensors, pumps, electrical drives etc.) but rather because the scheduled maintenance work itself is carried out in an erroneous way and thus creates failures. Consequently, one of the largest contributions to the need for maintenance is the scheduled maintenance itself (which in turn creates a need for emergency maintenance).

For large and complex industrial processes (power plants, chemical industries, paper- and pulp processes etc.) the costs for maintenance is highly significant and yet few efforts have been made to minimize these costs in a systematic way. There is a high potential to save resources by efficient and optimised maintenance. This can be achieved by using production data to determine when maintenance is to be carried out rather than basing such work on rigid time schedules. An integration of production- and maintenance data bases is a necessary step towards achieving such a goal.

1.1 Goal and strategy

The goal of the work was to develop a concept for efficient use of production data for a nuclear power plant to support the optimisation of the maintenance work and to determine what production data are the most relevant for the maintenance work. The concept should then be integrated into the existing software analysis tool Bi-Cycle Warehouse and Analysis. The work was carried out at the nuclear power plant in Barsebäck, Sweden.

This report describes an initial approach to quantify and analyse the loss of production in nuclear power plants. The aim is to improve the understanding of causes of production losses and to support analysis of the production as well as production planning at the plant. The work generates a basis for further development of the software that analyses and presents production data on the internet/intranet.

The work consists of *maintenance and production data analysis*, *model development for different causes of production losses* and *software specifications of the data structure and user interface.*

Production data (1994-1999) from the nuclear power plant (B2) in Barsebäck (Sweden) are used for the study. Barsebäck Kraft AB (BKAB), situated north of Malmö, is a nuclear power utility for electricity generation. BKAB consists of two ABB Atom designed boiling water reactor (BWR) units, B1 and B2. The 615 MW units were taken into operation 1975 and 1977, respectively. B1 was closed at the end of November 1998. B2 has been in operation for 23 years.

1.2 Nuclear power production in Sweden

The nuclear power plant at Barsebäck (Figure 1) was placed in commercial operation in 1972, marking the beginning of the expansion of nuclear power in Sweden.

Fig 1. The nuclear power plant at Barsebäck.

The choice of nuclear power was far from being a foregone conclusion. While there was a heavy demand for electricity, public opinion strongly opposed the further development of the rivers in northern Sweden to generate more hydroelectric power. They decided for the nuclear power because the power generation experts were seeking a future technology that would reduce dependence on oil.

The locations of nuclear power plants provoke much discussion. When the power plants were in the planning stage, Barsebäck on the coast north of Malmö, was regarded as a suitable site. Access to seawater for cooling purposes is essential. This location was also considered suitable by the licensing authorities. The locations for all Swedish nuclear reactors are shown in Figure 2 and Figure 3 presents the contribution of power and electricity from the Swedish nuclear power plants compared to other power sources.

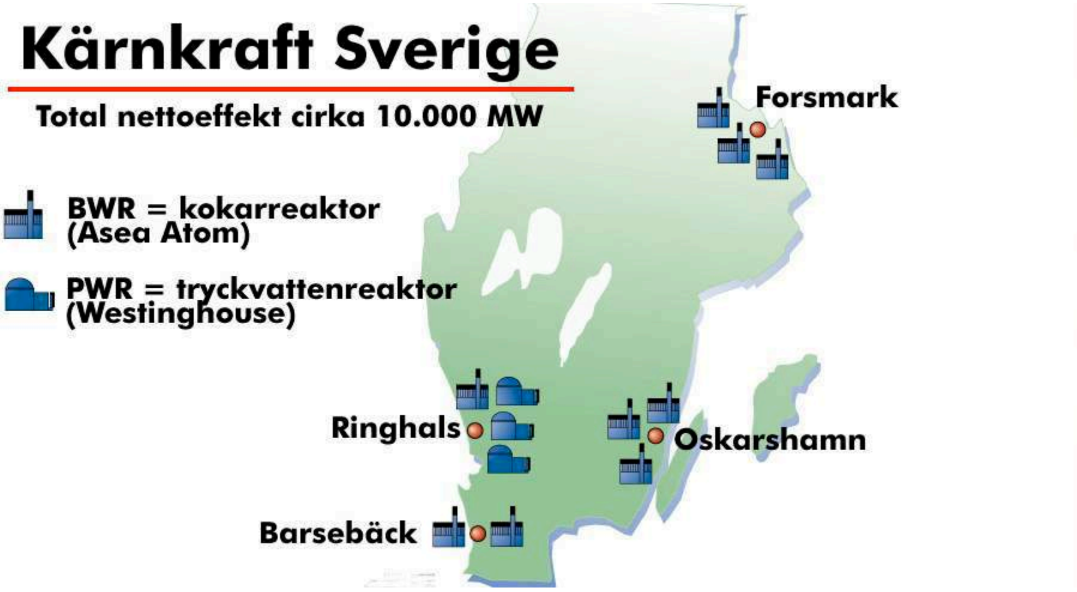


Fig 2. Location of nuclear power plants in Sweden.

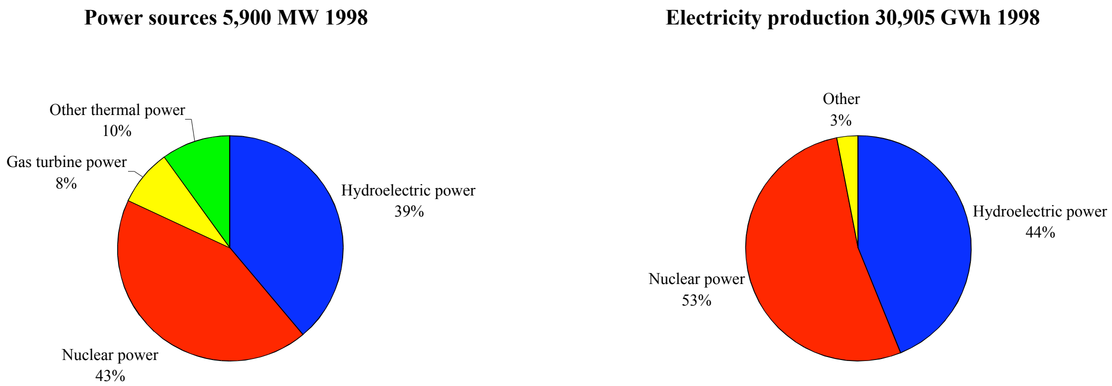


Fig 3. Power sources and electricity production in Sweden.

Nuclear power plants are relatively expensive to build, but have very low operating costs today. Even if the costs of extensive renovation and rebuilding work are taken into account, nuclear power provides cheap electrical energy compared with the cost of building new power plants. The service life of a nuclear power plant is determined primarily by the length of time for which it remains economically advantageous to replace or renovate parts that do not fulfil the requirements for safe operation. The nuclear power plants were originally designed for a service lifespan of 40 years.

1.3 The Sydkraft Group

This work was carried out at the nuclear power plant in Barsebäck, Sweden. The plant is owned and operated by the Sydkraft Group, which comprises of four business areas consisting of some 60 operating companies, half of which have their own personnel. The Group includes companies, which supply electricity, natural gas, heating power generation and solid fuel, and also provides computer, electrical installation, measurement, telecom and consulting services. Sydkraft produces electricity in its own plants and also purchases and sells power in the Nordic region and through its cable link with Germany.

It is in the power companies' interests to use resources as efficiently as possible. During the winter months, both nuclear power and hydropower are needed, which run at full capacity during this period (see Figure 4). In the summer, following the spring flood, an abundance of water coincides with a relatively low power demand, making this the best time to inspect the nuclear power plants and change the fuel. In a normal year, the Barsebäck plant produces approximately 9 TWh per year. This is the double the amount of electricity consumed each year by the cities of Malmö and Copenhagen combined.

In 1998, with the support of the "Act concerning the phasing out of nuclear power", the government decided to close one reactor at Barsebäck, unconditionally, on July 1, 1998, and to close the second reactor unit not later than July 1, 2001, provided that the shortfall in power production can be offset through power produced from renewable sources and a reduction of electricity consumption. Since then the decision concerning the second reactor has been altered and it is today not clear when the final shut-down is to take place.

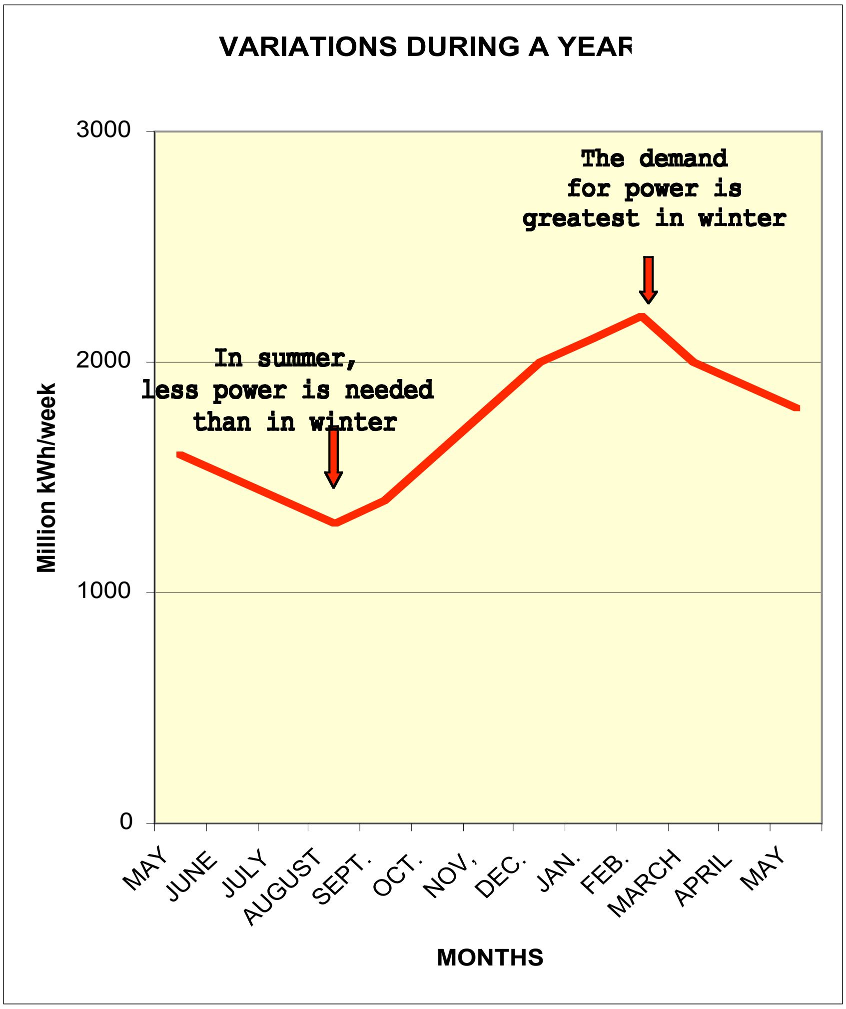


Fig 4. Electricity consumption in Sweden during a year.

2 Bi-Cycle – software for maintenance

Bi-Cycle is a maintenance data-warehouse and analysis software system, that uncover key maintenance problem areas and identifies potential cost reductions. Most industries store engineering, production and maintenance data. All these sources of information are instrumental to improve maintenance plans.

The Bi-Cycle Warehouse module allows the user to link all the maintenance-related information. This consolidated data is then scrutinised by the Bi-Cycle Analysis module, which provides sophisticated tools to analyse years of accumulated data on hundreds of thousands of components. Bi-Cycle was developed jointly with the Swedish nuclear power industry. It has been tried and tested, with major success, in rigorous use.

2.1 Maintenance data warehouse

Bi-Cycle Warehouse stores information from all the different sources in the company that are relevant to improve maintenance plans. Engineering information, maintenance reports and plant production data from one or more plants are placed in one data warehouse.

For example, it enables the user to search through all the systems, equipment and components. This hierarchy can be changed in for example: equipment type, subtype and manufacturer. For the selected systems, equipment or components the related maintenance work orders are easily accessible.

The user can navigate through this data with extreme ease of use, using the intuitive datatree navigator. The datatree arranges the components in the same way as the plant build-up, on the levels plant, equipment and components.

Bi-Cycle Warehouse allows the user to filter, search and sort more than 200,000 maintenance reports in seconds. Components with similar problems, for example valve leakage, can be identified and grouped together for analysis.

2.2 Maintenance data analysis

With the Analysis tool, the user is able to determine important maintenance indicators and sort items according to these indicators. For example sorting all centrifugal pumps by number of failures in the specific year, or sorting all valves by occurrence of a certain type of failure. Classifying maintenance reports based on free text keywords and/or predefined codes are possible.

Within minutes, the user is presented with all relevant maintenance, safety and costs indicators (see Table 1) for more than 100,000 components.

Bi-Cycle Analysis has three analysis functions: History, Watchdog and PM Fix. History produces maintenance history reports with clarifying charts. Watchdog calculates maintenance indicators, identifying essential problem areas such as components with high numbers of maintenanceinduced failures, or the component manufacturer with the highest mean repair time. PM Fix compares the results of different preventive maintenance (PM) programs on similar components to deduce optimal maintenance intervals. Further, it calculates maintenance statistics on different failure modes (e.g. FMECA - Failure Mode Effect and Criticality Analysis).

All results of an analysis in Bi-Cycle can be displayed graphically and exported to other programs (like Excel) or in the form of user defined reports. Reports can be automatically generated in Word format or publicised directly on intra- or Internet in HTML format.

Table 1. Example of available indicators in Bi-Cycle.

Preventive maintenance (PM), Corrective maintenance (CM)	**Trends**	**Reliability characteristics**
• jobs per year, %PM, %CM • costs per year, hours per year, %PM, %CM • costs per job, hours per job • failures after maintenance	• trends in failures, jobs and costs • trends in types of failures and jobs • trends in other maintenance indicators	• failure rate (Weibull shape) • repair hours per year • mean repair time

3 Analysis of production data

With the tool Bi-Cycle Warehouse and Analysis it is possible to make some analysis of the production data in order to study the causes of failure. In this program, the nuclear power plant is divided into three groups:

- systems;
- equipment;
- components.

In each of these, one can study the causes of production losses. For every group it is possible to see the type of failure, preventive maintenance (PM) and corrective maintenance (CM) reports and other adequate data like the description of the failure, the date, the repair work and the designation of the system, or equipment or component. Based on a thorough analysis of available production data, it was possible to identity the most relevant causes and their respective influence on the production losses.

3.1 Production data

Schematically, the production (or rather the production losses) in a nuclear power plant can be described according to Figure 5.

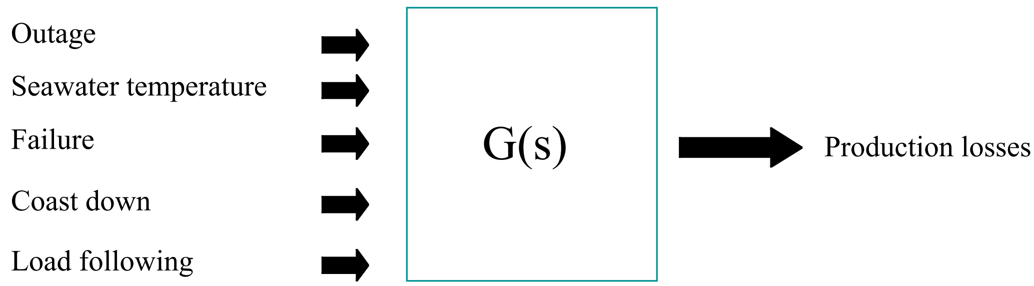


Fig 5. Input-output relationship for a nuclear power plant.

The major causes that influence the production are: *outages* (planned and unplanned revision), *failures*, *coast down*, *load following* and *seawater temperature*, see also Figure 6.

The study is based on production data from 1994 to 1999. A number of annually recurring events that cause loss of production have been identified. The data will be analysed using time series analysis since the main characteristic of a time series is that its observations have some form of dependency on time. In this section the most important causes of production losses are briefly described, see[7].

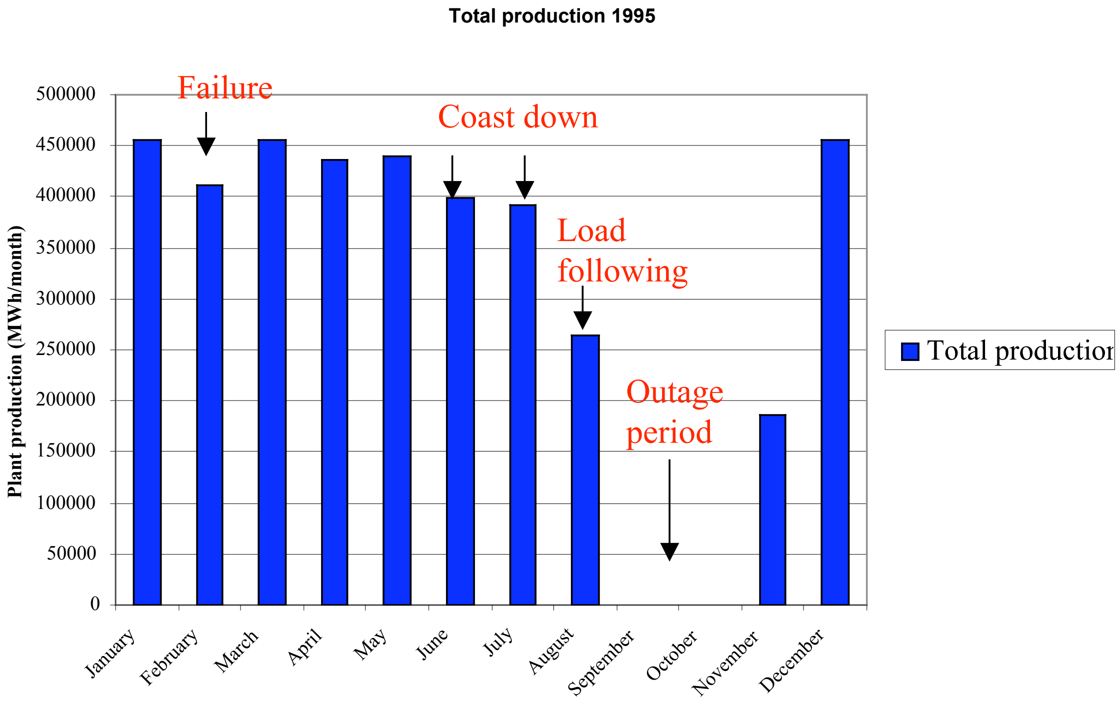


Fig 6. Total production during 1995 and the causes for production losses.

3.1.1 Outage

Outage is the most significant cause because the plant has to shut down for almost one month in order to perform planned activities such as maintenance, refuelling and modifications (see Table 2). Outages are undertaken during low electricity consumption periods (i.e. during the summer).

Table 2. Data regarding outage periods.

Year	Duration of outage period (days)	Production loss MWh/10000
1994	38	56.088
1995	86	126.936
1996	79	116.604
1997	53	78.228
1998	45	66.42
1999	50	73.8

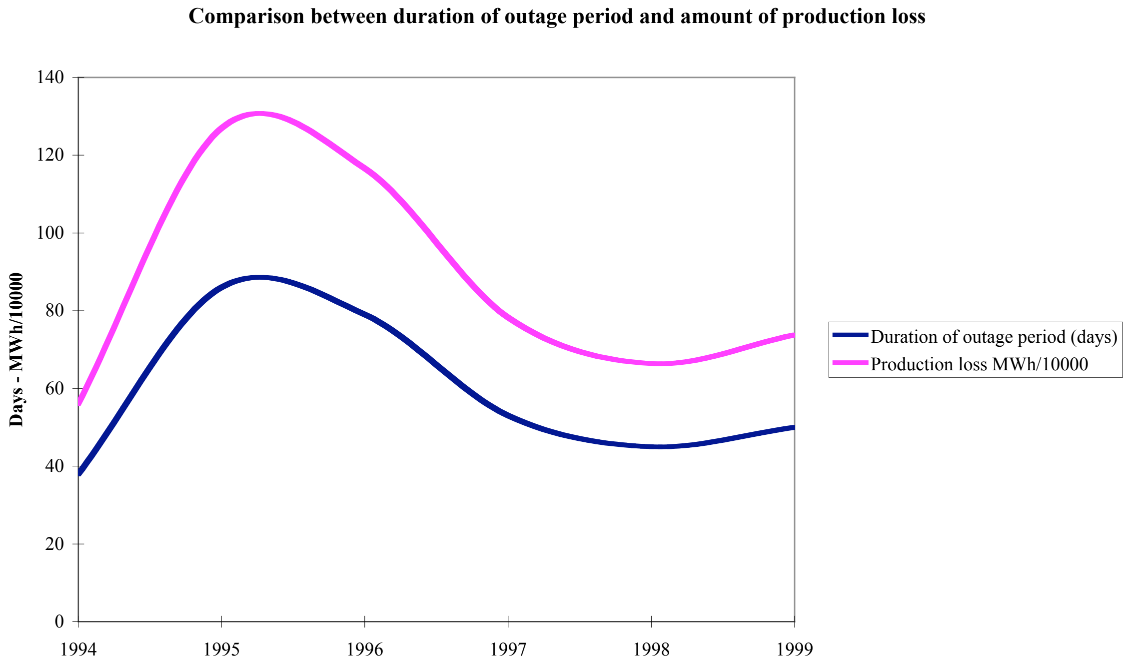


Fig 7. Data and amount of production loss due to outage.

Figure 7 shows the relation between the production loss during the outage period and the duration in days of the revision period. We can see that both of them have the same trend. In fact we can calculate that for each day the plant looses 14,760 MWh (in relation to the full capacity). Consequently, to avoid a waste of energy, the best thing is to concentrate all maintenance works in a few days. The production loss for outage has decreased over the last years, which may indicate that a better program for maintenance has been applied.

3.1.2 Coast down

The *coast down phenomena (reduction of fuel capacity)* occurs only in 1994 and 1995. This effect is not relevant for this study due to the limited amount of available data. Usually the coast down phenomena occurs after a long operational period and during coast down the plant losses are approximately 2% of the weekly production. Moreover, a model describing the effects of coast down already exists at Barsebäck.

The coast down occurs when certain pumps carry their maximum flow in combination with the use of control rods to reduce the plant production. These pumps have to carry the full flow as they try to compensate the reduction of reactor energy with more work.

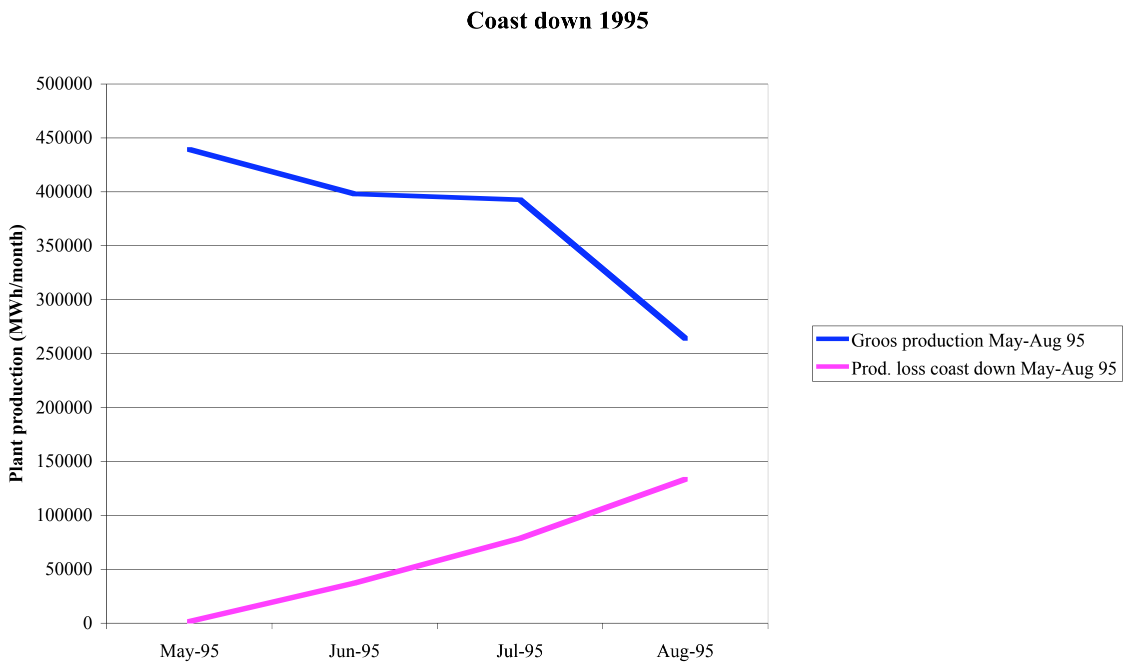
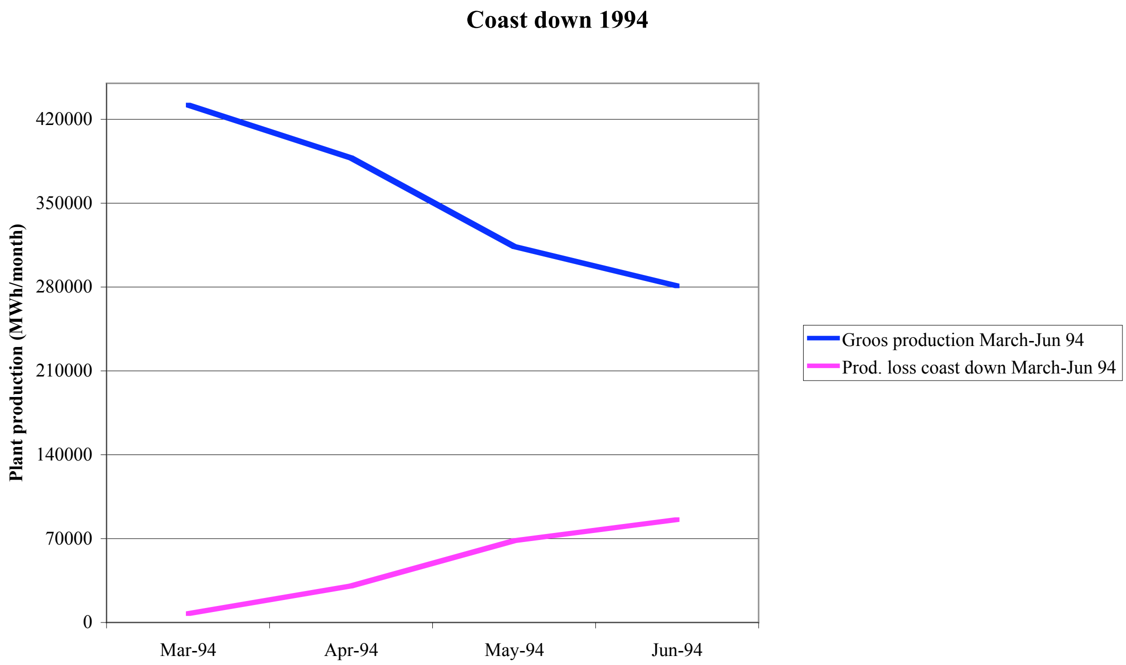


Fig 8. Production decrease due to coast down and the coast down loss increase in 1995.

In Figures 8 and 9, we can see how the production decreases during coast down. Due to other effects (e.g. failure and load following), the production decrease and the coast down increase do not coincide.



.
Fig 9. Production decrease due to coast down and the coast down loss increase in 1994.

3.1.3 Component failure

Failures are important, but this cause is not relevant to study by deterministic models because the failures occur more and less at random. However, stochastic models may in future works be used to describe such behaviour. Figure 10 demonstrates that the production loss due to component failures is rather constant over time (i.e. between 500,000 MWh and 1,000,000 MWh), which is due to the random behaviour of the failures. However, a limited decrease can be seen for 1998 and also an increase for 1999.

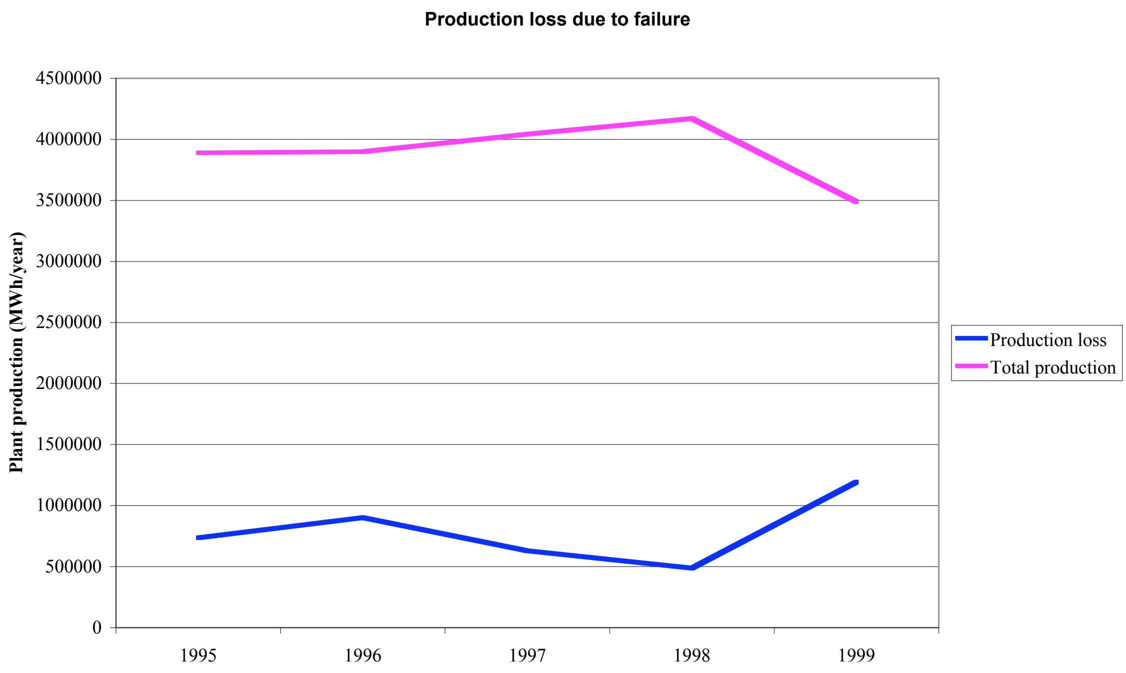


Fig 10. Trend of production loss due to component failures.

The mechanical failures are the most common ones and Figure 11 shows the components that fail most frequently. The system's classification considering the influence on reactor safety and availability (energy production) is presented in Table 3. The classification data, regarding the entity of the failures, are shown in Table 4. These are divided into reactor safety and availability (for electric power production).

The production loss for component failures is always related to the total production because normally a higher production means a higher probability for failures.

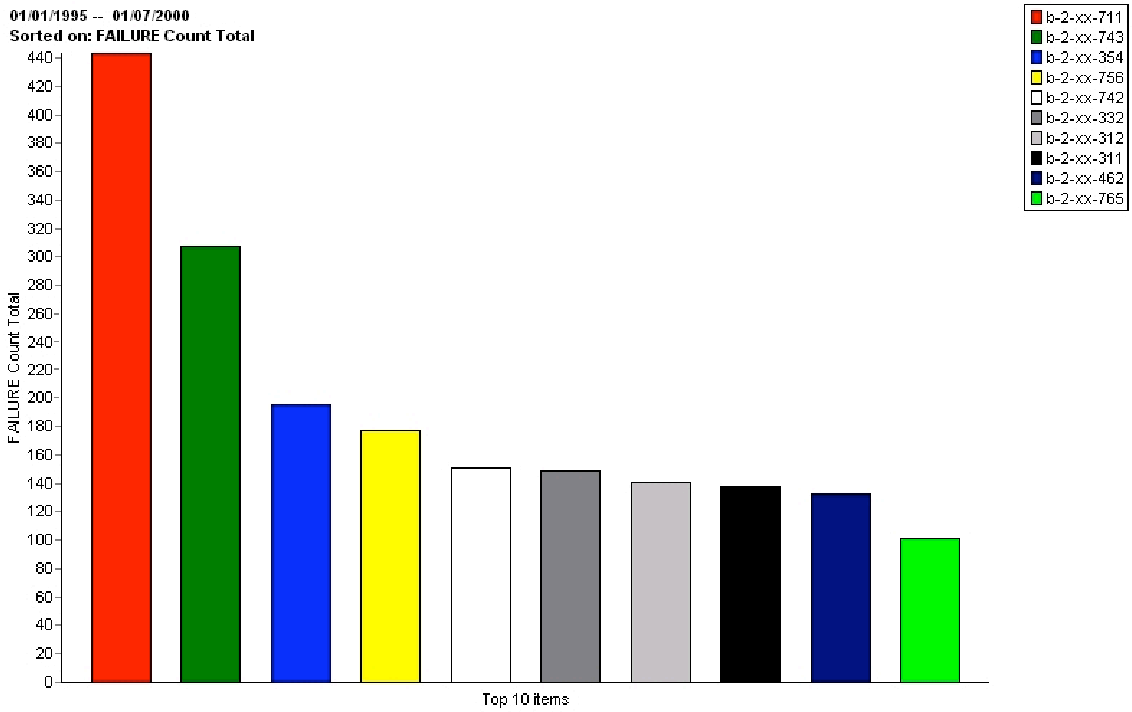


Fig 11. Most frequent components to fail from 1995 to 1999.

Table 3. System's classification of component failures (cmp. Figure 11).

Designation	Reactor-safety	Availability
b2-711 **-Seawater cooling system. A total system failure will automatically stop the plant. Degradations due to sea pollutions lead normally to reducing the power. This system has also safety influence.**	**1**	**2**
b2-743 **-Ventilation for non-controlled (low radiation activity) areas. No availability influence.**	**1**	**0**
b2-354 **-Hydraulic scram system. A safety system that is only used to stop the plant quickly (in emergency). No availability influence.**	**2**	**0**
b2-756 **-Nitrogen dosing. Prevents pipe corrosion. Influence on both safety and availability.**	**2**	**2**
b2-742 **-Ventilation for radiation active areas. Only safety influence.**	**2**	**0**
b2-332 **-Condensate cleaning system. Only availability influence**	**0**	**2**
b2-312 **-Feed water system. This system leads water into the reactor tank and a total system failure will lead to stop of the plant. Both safety and availability influence.**	**2**	**2**
b2-311 **-Mean steam pipes (this system leads steam to from the reactor tank to the turbine. A system failure will stop the plant. Both safety and availability influence.**	**1**	**2**
b2-462 **-Main condensate system (low-pressure – pre heater). A system failure will stop the plant also effect the reactor safety.**	**2**	**2**
b2-765 **-Water for extinguishing system. Only safety influence**	**2**	**0**

Table 4. Classification data.

Reactor safety:

Class

0 No STF (non safety requirements)

1 The failure shall be corrected within a time that allows **exceed 48h.**

2 The failure shall be corrected **within 48h.**

Availability (for electric power production):

Class

0 Failure has no influence on production capacity

1 Failure leads to reduced (automatically or manually)production capacity, but leads not to stop of the plant.

2 Failure leads to reduced (automatically or manually) production capacity and leads to stop of the plant.

3.1.4 Seawater temperature

The *seawater temperature* influences the production during the whole year, although the influence is not as significant as for the other causes, see Figure 12 below.

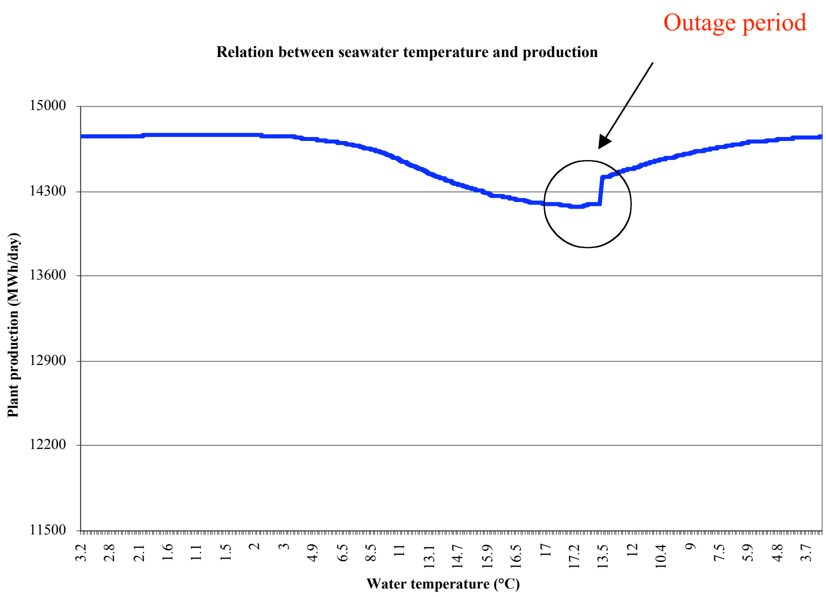


Fig 12. Relationship between seawater temperature and production during a year (outage period removed from the data).

Figure 13 represents the behaviour of the seawater temperature during an entire year. We can recognize that this behaviour is the normal behaviour of the seawater temperature but if we compare this with Figure 14, we can see that the two figures are complementary, thus, when the seawater temperature is low the plant produces more and the production decrease by the end of spring and in the summer when the seawater temperature is high.

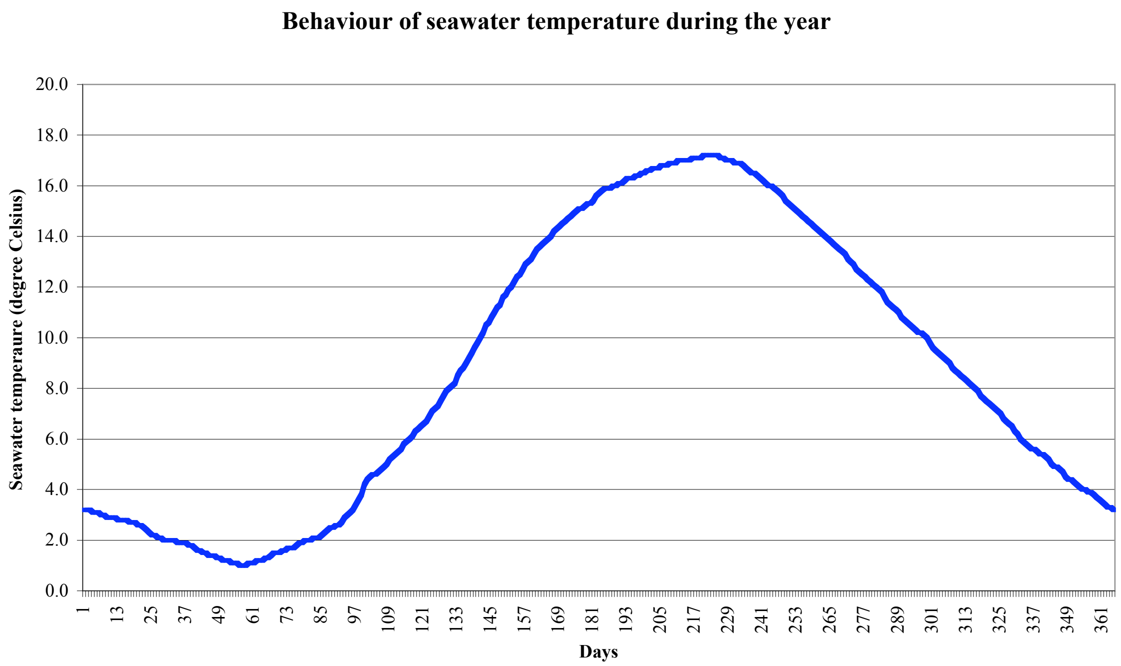


Fig 13. Seawater temperature during an entire year.

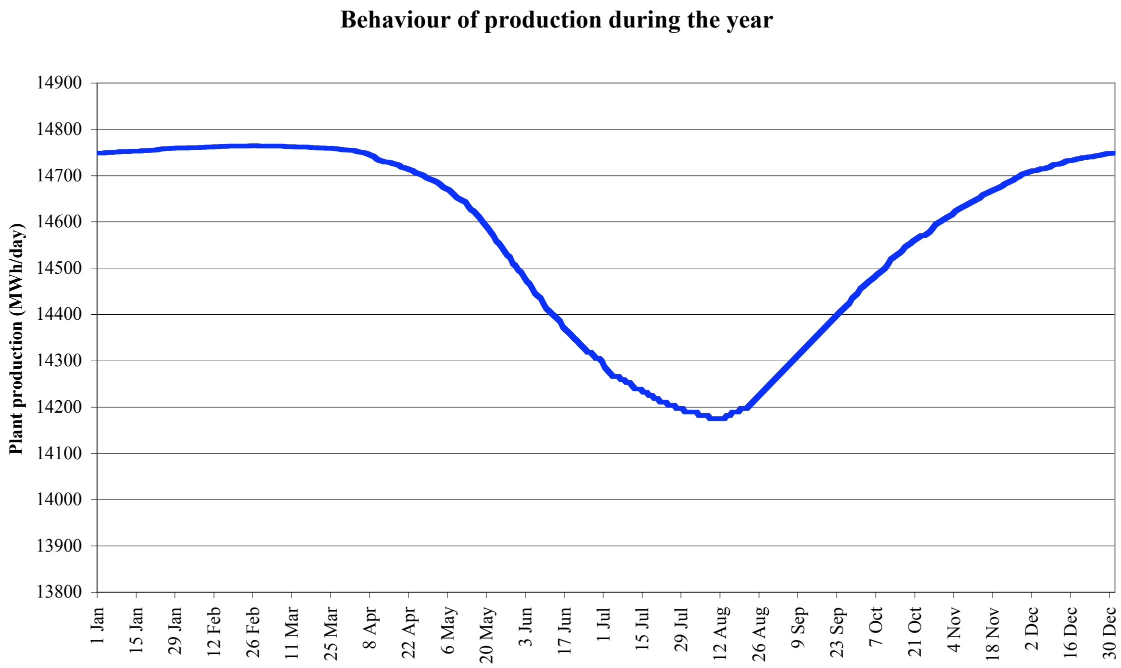


Fig 14. Production during an entire year.

3.1.5 Load following

The *load following,* due to low electricity consumption (and prices), usually starts by the end of spring, (i.e. April or May), and ceases by the end of summer.

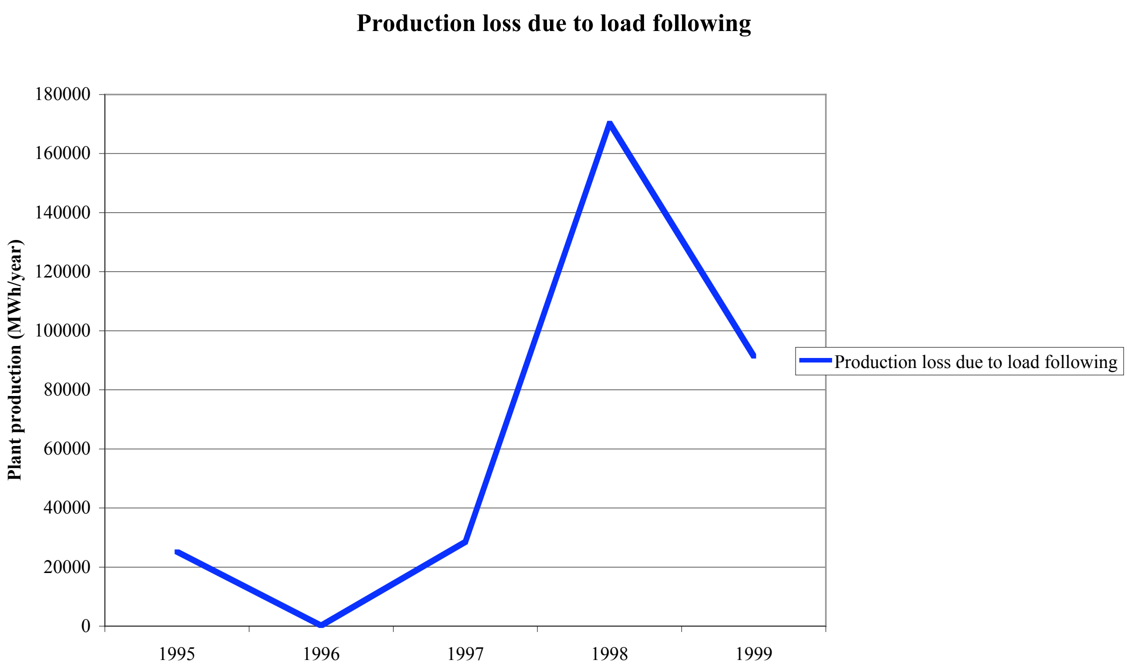


Fig 15. Production loss due to low prices.

Figure 15 shows the amount of production loss due to load following between 1995 and 1999. The total production loss due to load following has a strange trend. It started in 1995 with a low value and was almost zero in 1996 and then increased until the peak in 1998 (almost 170,000 MWh). Then is started to decrease during the end of 1998 and 1999. This behaviour is due to the shut-down of reactor 1, which led to that reactor 2 had to produce more energy also during the summer.

3.2 Data quality

The data are stored in data base files and they are hourly averages for every day. The maximum hourly production is 615 MWh and when the data are higher than this, it needs to be changed to the maximum value because the plant is not able to produce more. For missing data (only one or two time instances), we have chosen to calculate the average production based on data for seven hours before and after the problem occurred. Sometimes strange values appear (like –9999.99 indicating a data base problem) and those time instances have been treated similar to missing data (see Table 5). For some types of analysis it is not essential to have access to hourly values. For that reason and to speed up the analysis the hourly values have also been used to calculate daily averages.

Table 5. Real data from the production data base.

DATE	PRODUCTION	
1999-01-18	615	615
1999-01-18 01:00	616	615
1999-01-18 02:00	615	615
1999-01-18 03:00	616	615
1999-01-18 04:00	615	615
1999-01-18 05:00	616	615
1999-01-18 06:00	615	615
1999-01-18 07:00	615	615
1999-01-18 08:00	615	615
1999-01-18 09:00	616	615
1999-01-18 10:00	615	615
1999-01-18 11:00	616	615
1999-01-18 12:00	615	615
1999-01-18 13:00	616	615
1999-01-18 14:00	615	615
1999-01-18 15:00	-9999,99	
1999-01-18 16:00	615	615
1999-01-18 17:00	615	615
1999-01-18 18:00	616	615
1999-01-18 19:00	615	615
1999-01-18 20:00	616	615
1999-01-18 21:00	616	615
1999-01-18 22:00	615	615
1999-01-18 23:00	616	615

Figure 16 shows that the yearly production is always between 300,000 MWh and 350,000 MWh. Only for the year 2000 the prediction is somewhat higher but this does not mean that the plant will be able to produce this amount of energy.

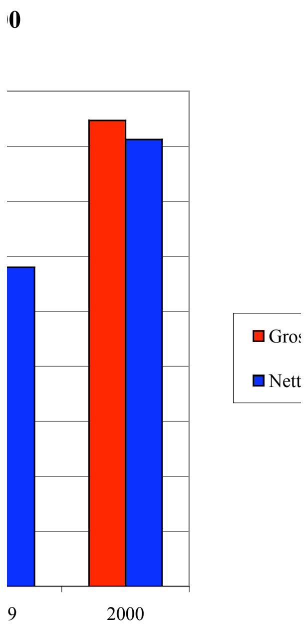


Fig 16. Total plant production from 1995 to 2000 (year 2000 predicted production).

The production data may also be used to calculate different indicators, such as the time and the energy utilisation factor, using the simple formulae given below. The values of the indicators from 1995 to 1999 are presented in Tables 6 and 7.

- Energy utilisation factor: $$\frac{Gross\,Production}{CalendarTime \times Max\,Production(615MWh)} \times 100$$
- Time utilisation factor: $$\frac{Power.to.the.grid}{CalendarTime} \times 100$$

Table 6. Data regarding time utilisation factor.

Year	Calendar Time	Power to the grid	Time utilisation factor
1995	8760	6724	76.75799087
1996	8783	6477	73.74473415
1997	8760	6826	77.92237443
1998	8760	7371	84.14383562
1999	8784	6180	70.35519126

Table 7. Data regarding energy utilisation factor.

Year	Calendar Time	Gross production	Energy utilisation factor
1995	8760	3890487	72.21455619
1996	8783	3900155.1	72.20443595
1997	8760	4042937	75.04430709
1998	8760	4171528	77.4311913
1999	8784	3600392	66.64726702

These indicators, together with the total production show that the year 1998 was the best. Considering all five years both the time and energy utilisation factors are highest for 1998 (see Figures 17 and 18).

The remaining part of this section will devoted to show various aspects of the available production data and maintenance data for each individual year between 1999 and 1994.

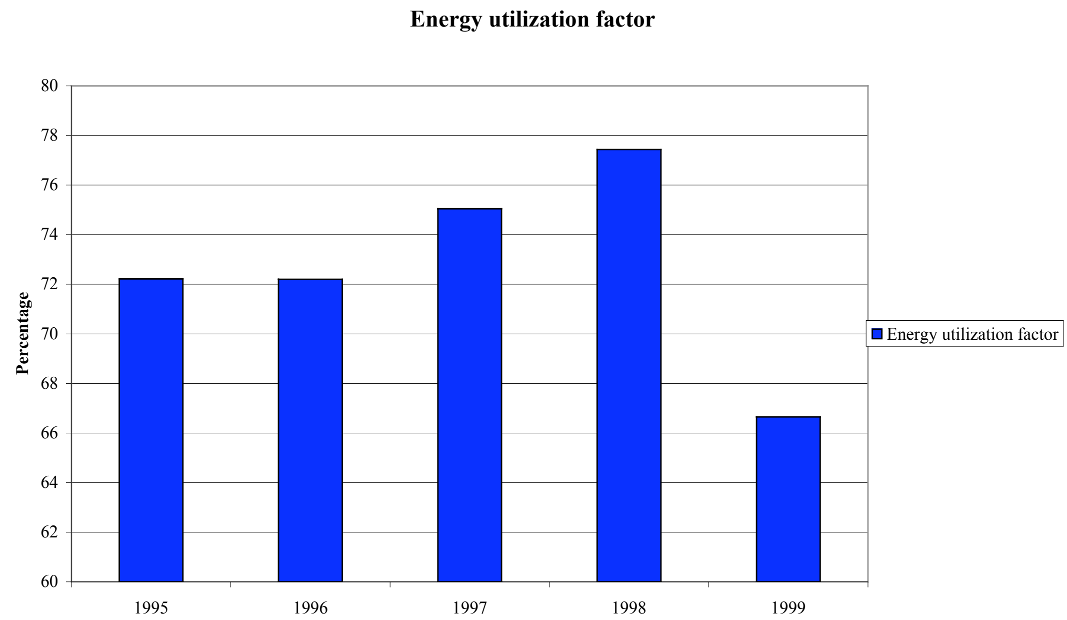


Fig 17. Energy utilisation factor.

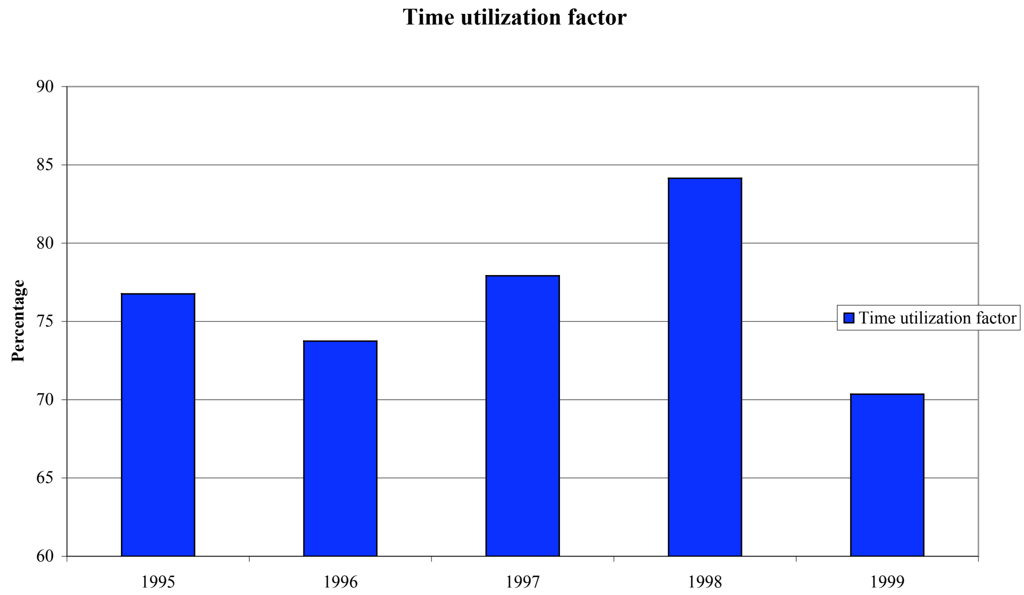


Fig 18. Time utilization factor.

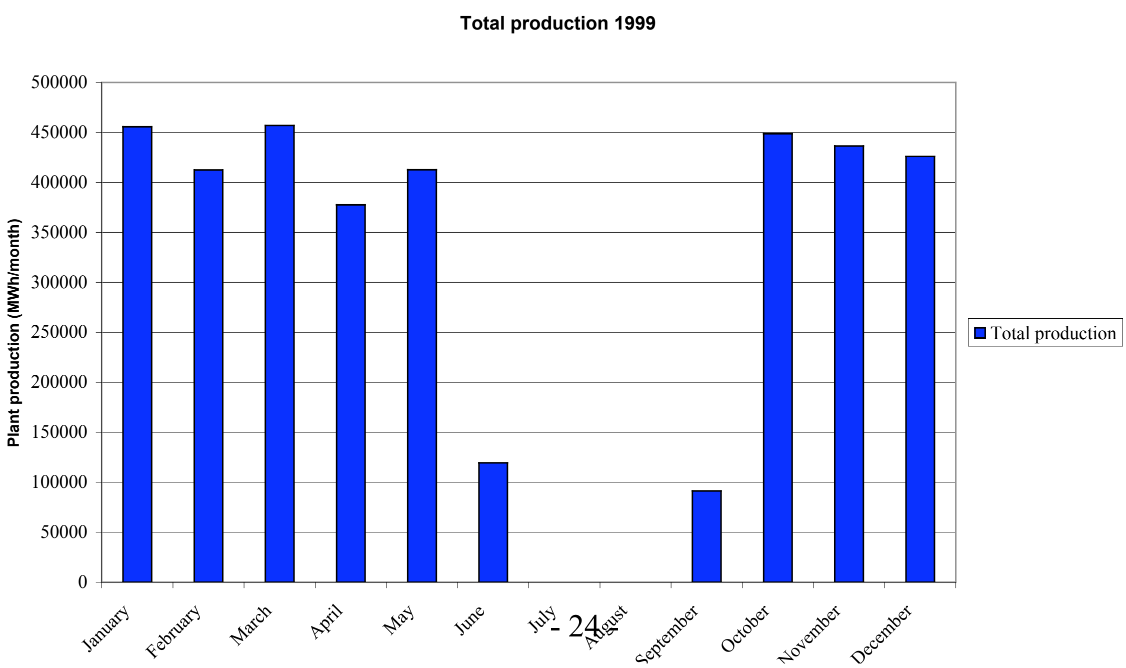


Figure 19 demonstrates the production for each month during 1999 and it is clearly seen that from 10th June to 23rd September the production was zero because the plant was stopped for the annual revision period. Actually the declared revision period was only from 3rd August to 23rd September but prior to the revision different types of repairs were needed.

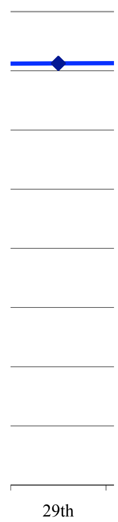


Fig 20. Trend of regaining production after a shut-down.

Figure 20 represents the trend of production after the outage period. It can be seen that the plant needs almost six days to arrive at the normal production after a shut-down. The sum of production loss due to failure was 1,193,803 MWh and the sum of production loss due to load following was 91,340.2 MWh during 1999.

Table 8. Failures grouped by different causes (1999).

Class	Total	Per Item	MTBF	99
Break/Crack	25	0.095	15400	25
Internal leakage	69	0.261	5570	69
External leakage	138	0.523	2790	138
Mechanical failure	294	1.11	1310	294
Electrical failure	145	0.549	2650	145
Calibre	90	0.341	4270	90
Other failure	152	0.576	2530	152
No failure	102	0.386	3770	102
No Class	823	3.12	467	823

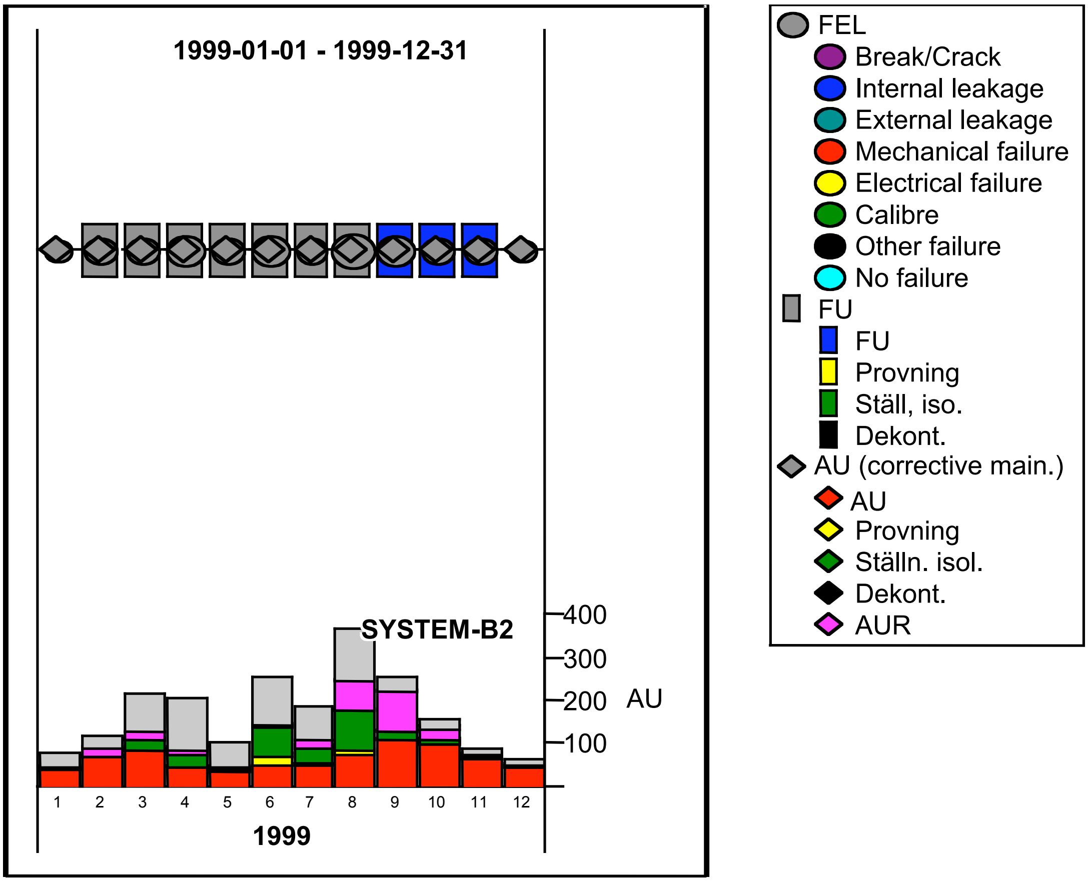


Fig 21. Bi-Cycle chart for 1999.

In Table 8 the different types of failures that occurred during 1999 are given. They have been divided into different groups depending on the characteristics of the failures. Figure 21 shows the same type of information (on a monthly basis) in a graphical format, using colour coding to represent different failure classes. Both the tabular format and the graphical format are available from the maintenance software program Bi-Cycle. The most common type of failure is the one that cannot be incorporated into a specific class (i.e. no classified failure) followed by different sorts of mechanical failures.

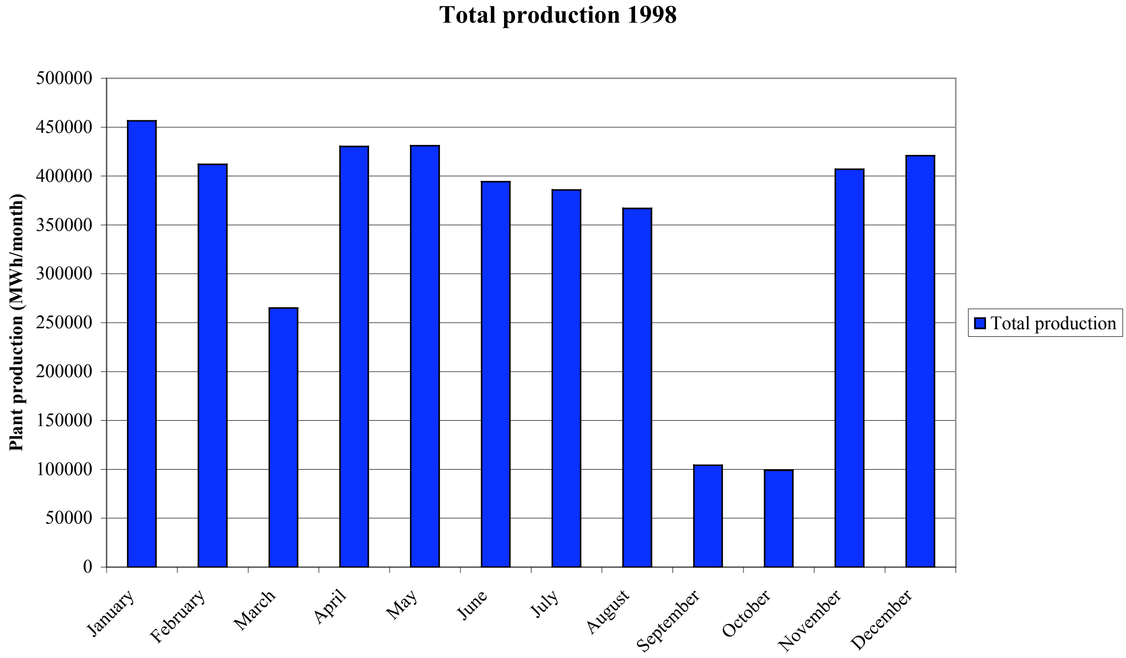


Fig 22. Total plant production during 1998.

In 1998 the production has increased compared to 1999. In fact this year is the best for the total energy production of all the studied years. From Figure 22 it seems almost as if there was no outage period; but this is not true because if we check the real date we see that from 11th September to 19th October there was no production. However, in the graph, since it is monthly averages it may be difficult to see this. Also the production loss is less than in 1999. This is to a large degree due to the significant reduction of the so called no classified failures (see Table 9 and Figure 23). The number of mechanical failures remains the most significant also this year. The loss of production due to failures is 487,674 MWh and 170,317 MWh of production is lost due to load following.

Table 9. Failures grouped by different causes (1998).

Class	Total	Per Item	MTBF	98
Break/Crack	50	0.189	5770	50
Internal leakage	49	0.186	5890	49
External leakage	175	0.663	1650	175
Mechanical failure	616	2.33	468	616
Electrical failure	233	0.883	1240	233
Calibre	90	0.341	3210	90
Other failure	523	1.98	552	523
No failure	520	1.97	555	520
No Class	34	0.129	8490	34

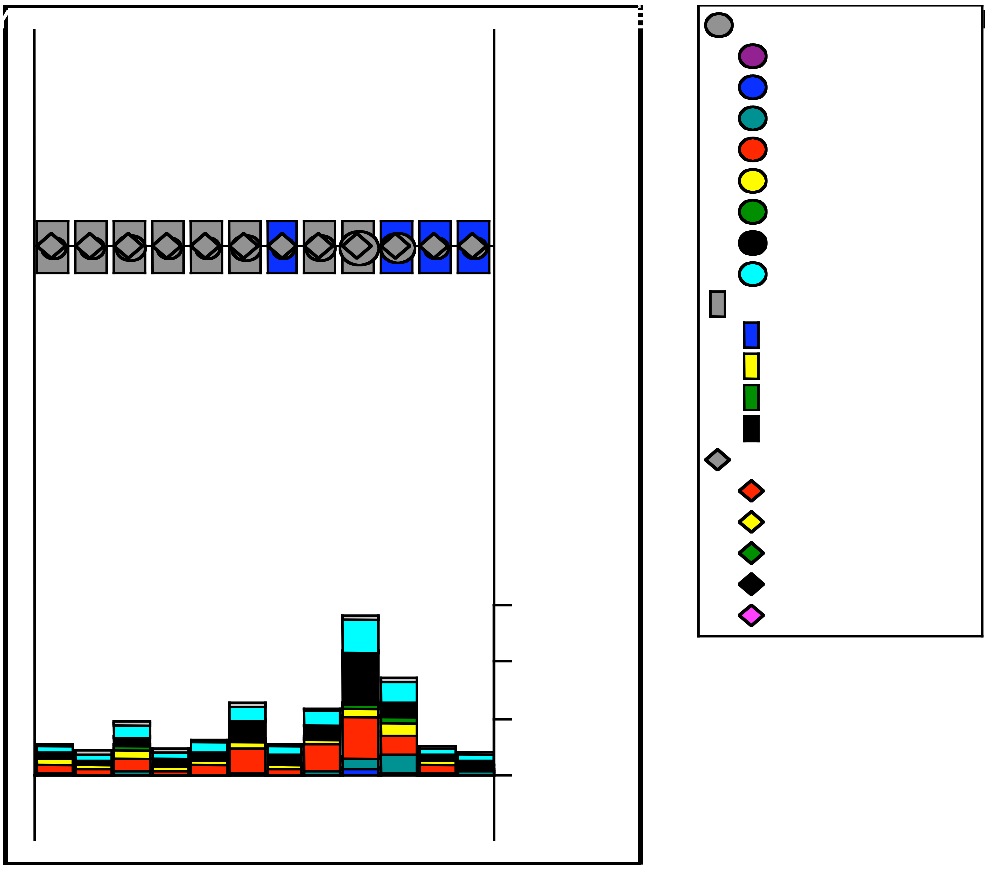


Fig 23. Bi-Cycle chart for 1998.

In 1997 the outage period was from 18th July to 8th September; but also during parts of February and March there was no production due to failures (see Figure 24). But also if the production was zero during these periods the total production was satisfactory. In fact the energy utilisation factor was 77.92%, which is a very good result. The outage period was fairly short and consequently also the production loss due to revision was only 78,228 MWh.

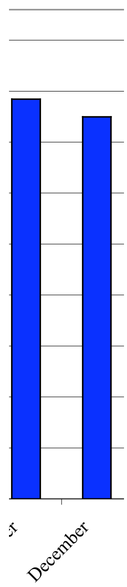


Fig 24. Total plant production during 1997.

Table 10. Failures grouped by different causes (1997).

Class	**Total**	**Per Item**	**MTBF**	**97**
Break/Crack	88	0.333	2190	88
Internal leakage	65	0.246	2960	65
External leakage	179	0.678	1080	179
Mechanical failures	638	2.42	302	638
Electrical failure	247	0.936	779	247
Calibre	79	0.299	2440	79
Other failure	499	1.89	386	499
No failure	548	2.08	351	548
No Class	59	0.223	3260	59

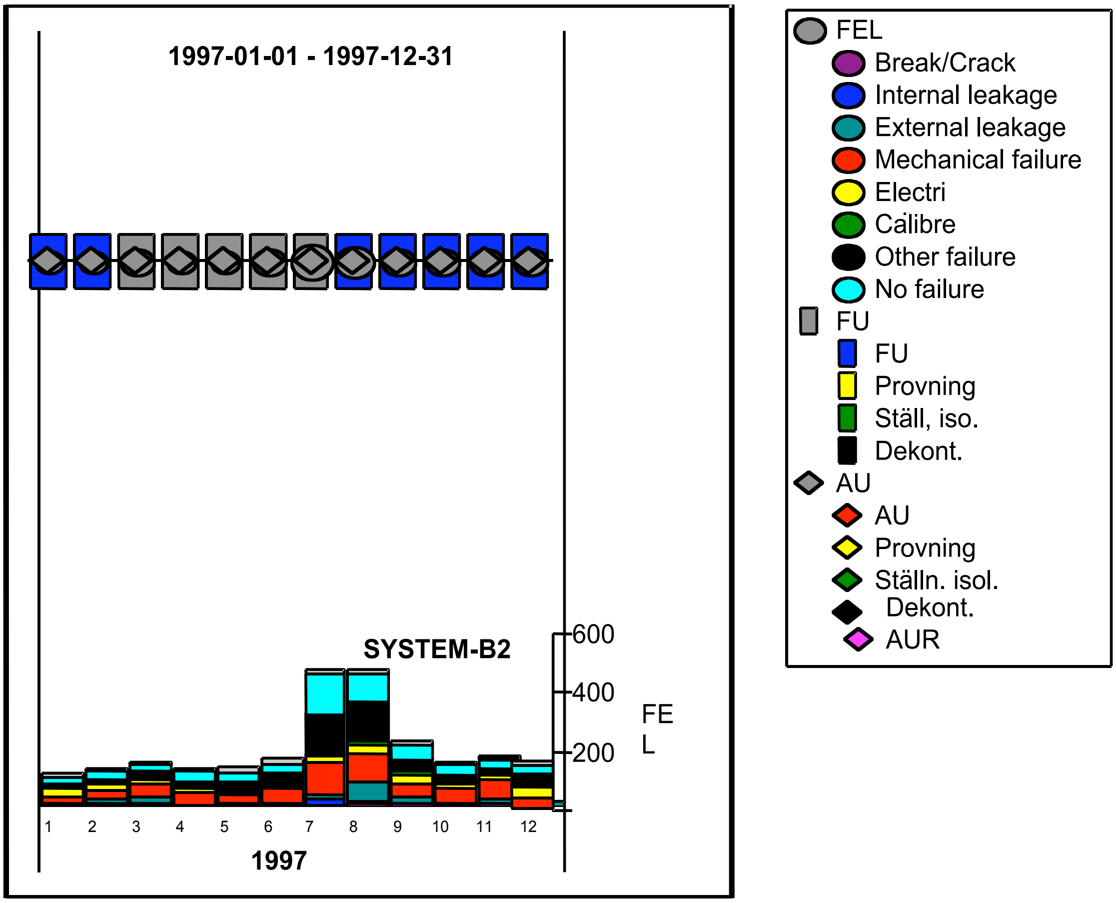


Fig 25. Bi-Cycle chart for 1997.

The problems with the mechanical failures are similar to the other years and also there is a significant amount of no classified failures (see Table 10 and Figure 25). The production loss due to failures were 630,528 MWh and only 28,550 MWh due to load following.

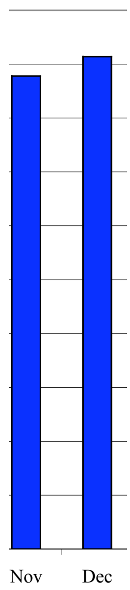


Fig 26. Total plant production during 1996.

In 1996 the production was zero for more than two months, in fact the outage period started 26th of July and finished 10th of October (see Figure 26). The production loss due to outage was consequently large, 1,107,000 MWh. Production loss due to other causes was 901,888 MWh and almost all of this was due to failures. This year is the first in which no load following is present. To explain this, we have tried to study the air temperature with the help of SMHI (Swedish Meteorological and Hydrological Institute) since a cold summer could be a reason why the electricity prices remained high over the entire year. We looked at time series for the daily temperatures in the meteorological station closest to Barsebäck, but we have not found any significant differences compared to the other years. The outage period was one of the longest in the last fifteen years.

Table 11. Failures grouped by different causes (1996).

Class	Total	Per Item	MTBF	96
Break/Crack	57	0.216	1690	57
Internal leakage	34	0.129	2830	34
External leakage	101	0.383	954	101
Mechanical failure	527	2	183	527
Electrical failure	265	1	364	265
Calibre	88	0.333	1100	88
Other failure	193	0.731	499	193
No failure	620	2.35	155	620
No Class	776	2.94	124	776

For the total component failures it is clear that the number of no classified failures is high as well as the number of mechanical failures (see Table 11 and Figure 27). The high number of no classified failures is a general problem with regard to maintenance and analysis of maintenance plans. When the failures are not classified they cannot be analysed in a traditional way since the reasons for such failures are completely unknown.

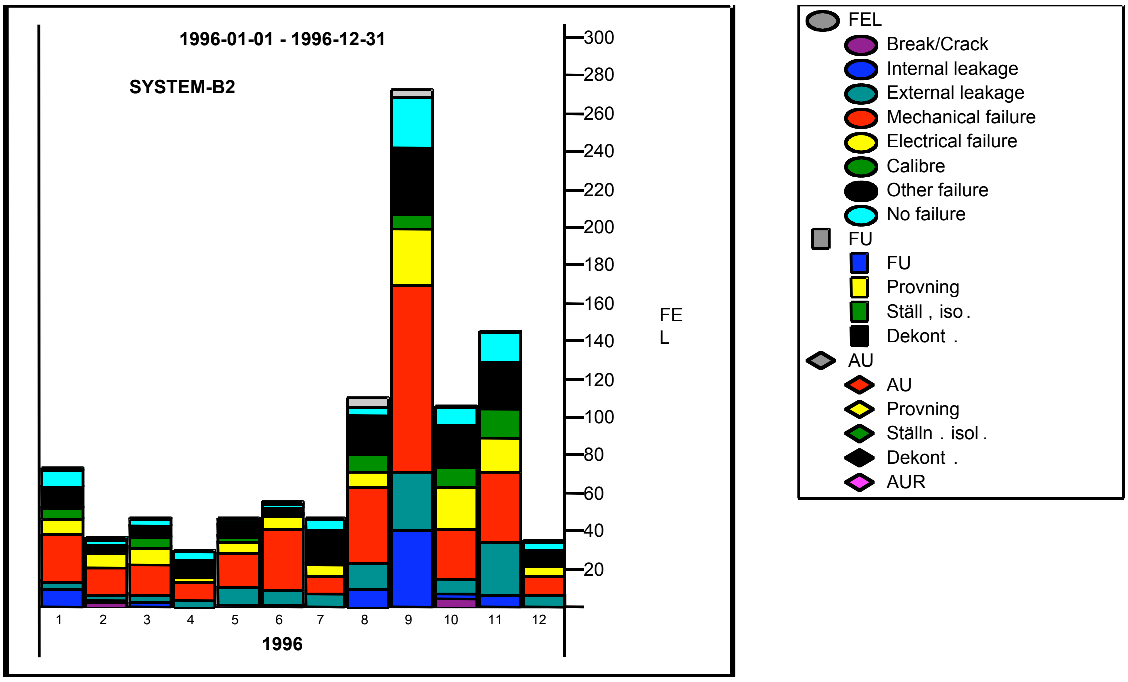


Fig 27. Bi-Cycle chart for 1996.

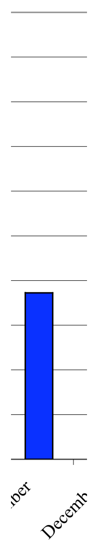


Fig 28. Total plant production during 1995.

In 1995 the production was zero for almost three months. The outage period started 23rd of August and finished 17th of November (see Figure 28). The production loss due to outage was consequently large, 1,279,000 MWh. This year, for the first time, we observe the coast down problem. The production loss due to coast down was 133,580 MWh. Due to failures and load following, the plant lost 762,244 MWh. If we compare Figures 28 and 29, we can see that when the production is zero during the outage, we have the maximum of failures. The reason for this is that during the revision a large number of failures are detected and repaired, which is an obvious goal of the revision work. From Table 12 it can also be noted that the number of no class failures is small, especially compared to years 1996 and 1999.

Table 12. Failures grouped by different causes (1995).

Class	Total	Per Item	MTBF	95
Break/Crack	17	0.064	11300	17
Internal leakage	72	0.273	2670	72
External leakage	124	0.47	1550	124
Mechanical failure	337	1.28	570	337
Electrical failure	129	0.489	1490	129
Calibre	63	0.239	3050	63
Other failure	167	0.633	1150	167
No failure	87	0.33	2210	87
No Class	8	0.03	24000	8

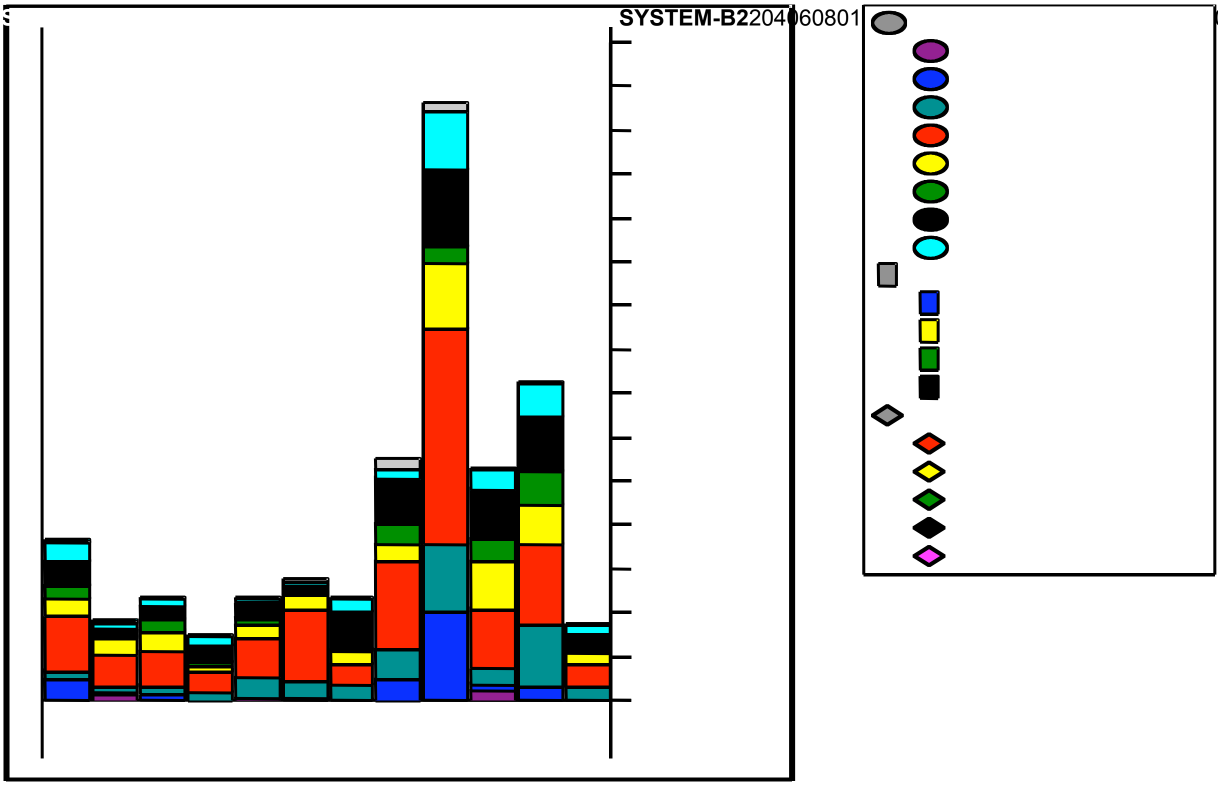


Fig 29.   Bi-Cycle chart for 1995.

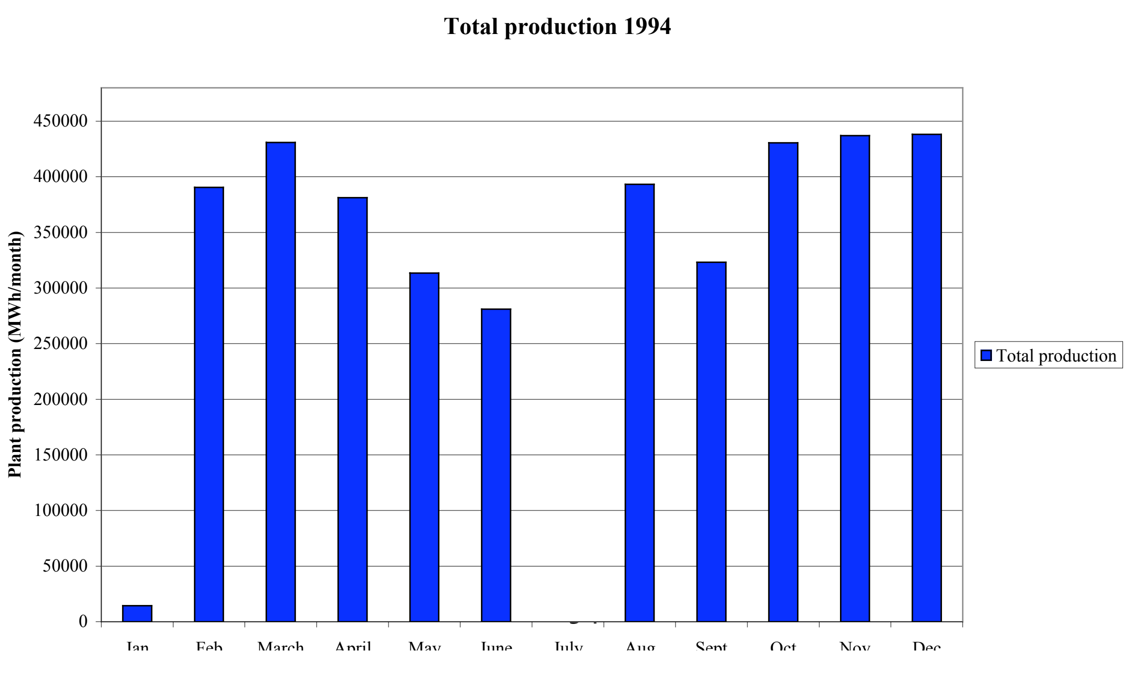


Fig 30. Total plant production during 1994.

Figure 30 shows that the outage period during 1994 was between 24th June and 1st August and that in January the plant had a lot of problems due to failures. The outage period was the shortest during the five investigated years and the production loss due to outage was only 560,880 MWh. The loss of production due to failures was 612,735 MWh and 20,908 MWh of production is lost due to load following.

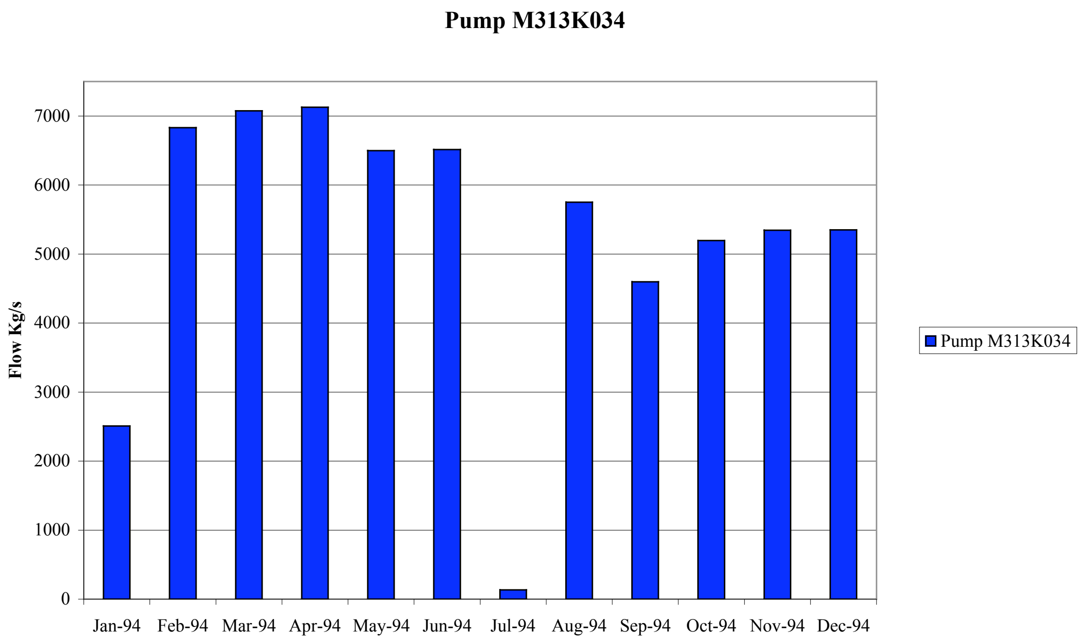


Fig 31. Behaviour of the flow in a pump 313K034.

During 1994 the coast phenomena is present. Coast down is related to the flow produced by a number of pumps at the plant. For instance, we can look at the maximum flow from pump 313K034. For the pump 313K034 the normal operating range is between 2500 Kg/s and 6000 Kg/s and we can see from Figure 31 that it have to pump the maximum flow from February to June and it is during this period that the coast down is present. The loss of production due to coast down in 1994 was 192,024 MWh.

Based on the data from 1994 to 1999 we have found a lot of things that appear repeatedly during all years, e.g. revision, low electricity prices (load following) and component failures. Coast down was only present during 1994 and 1995 and appear to be a problem with which the plant has learned to deal with. The number of failures varies significantly over the years but the most prominent reason for failures appear to be various mechanical problems. The data has been used to explain the behaviour of a nuclear power plant and in the next chapter the data will be used to identify mathematical models for various causes of production losses.

4 Modelling production loss

4.1 Introduction

Models describe relationship between measured signals. It is convenient to distinguish between input signals and output signals. The outputs are then partly determined by the inputs. In most cases the outputs are also affected by more signals than the measured input. Such "unmeasured inputs" will be called disturbance signals or noise. If we denote inputs and output by u and y, respectively, the relationship can be depicted as in the figure below.

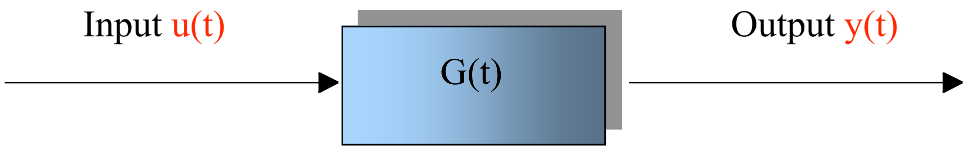


Fig 32. Relationship between input and output.

All these signals are functions of time. Often, in the identification context, only discrete-time points are considered, since the measurement equipment typically records the signals just at discrete-time instants, often equally spread in time with a sampling interval of T time units. The basic relationship is the linear difference equation. An example of such an equation is the following one:

$$y(t) + a_1 y(t-1) + \ldots + a_{na} y(t-na) = b_1 u(t-nk) + \ldots + b_{nb} u(t-nk-nb+1)$$

which relates the current output $y(t)$ to a finite number of past outputs $y(t\text{-}k)$ and inputs $u(t\text{-}k)$. The structure is thus entirely defined by the three integers na, nb and nk. na is equal to the number of poles and $nb\text{-}1$ is the number of zeros, while nk is the pure time-delay in the system. For a system under sampled-data control, typically nk is equal to 1 if there is no dead time. The output at time t is thus computed as a linear combination of past outputs and past inputs. It follows that the output at time t depends on the input signal at many previous time instants. This is what the word dynamic refers to.

The identification problem is then to use measurements of u and y to figure out:
- the coefficients in the equation $(a_1 \ldots a_{na}, b_1 \ldots b_{nb})$;
- how many delayed outputs to use in the description ($y(t\text{-}T) \ldots y(t\text{-}nT)$);
- how many delayed inputs to use.

4.1.1 Principles for identification

The system identification problem is to estimate a model of a system based on observed input-output data. Several ways to describe a system and to estimate such descriptions exist.

The procedure to determine a model of a dynamical system from observed input-output data involves three basic ingredients:

- the input-output data;
- a set of candidate models (the model structure);
- a criterion to select a particular model in the set, based on the information in the data (the identification method).

The identification process amounts to repeatedly selecting a model structure, computing the best model in the structure, and evaluating this model's properties to see if they are satisfactory. The cycle can be itemized as follows:

1. Design an experiment and collect input-output data from the process to be identified.
2. Examine the data. Polish it so as to remove trends and outliers, select useful portions of the original data, and apply filtering to entrance important frequency ranges.
3. Select and define a model structure (a set of candidate system description) within which the model is to be found.
4. Compute the best model in the model structure according to the input-output data and given criterion of fit.
5. Examine the obtained model's properties.
6. If the model is good enough, then stop; otherwise go back to step 3 to try another model set. Possibly also try other estimation methods (step 4) or work further on the input-output data (step 1 and 2).

4.1.2 Estimation methods

One can distinguish between two different types of estimation methods:

- Direct estimation of the impulse or the frequency response of the system. These methods are often also called non-parametric methods and do not impose any structure assumptions about the system, other then that it is linear.
- Parametric model. A specific model structure is assumed, and the parameters in this structure are estimated using data. This open up a large variety of possibilities, corresponding to different ways of describing the system. Dominating ways are state-space and several variants of difference equation descriptions.

Having estimated a model is just a first step. It must now be examined, compared with other models, and tested with new data sets. All linear models that are estimated can be written in the form:

$$y(t) = G(z) \cdot u(t) + v(t)$$

where $G(z)$ is the (discrete-time) transfer function of the system and $v(t)$ is an additive disturbance, see [2].

A good way of obtaining insight into the quality of a model is to simulate it with the input from a fresh dataset, and compare the simulated output with the measured one. This gives a good feel for which properties of the system the model has picked up, and which have not been picked up.

4.2 Modelling production loss

The production in a nuclear power plant, as we have already seen, can be schematically outlined as in Figure 33.

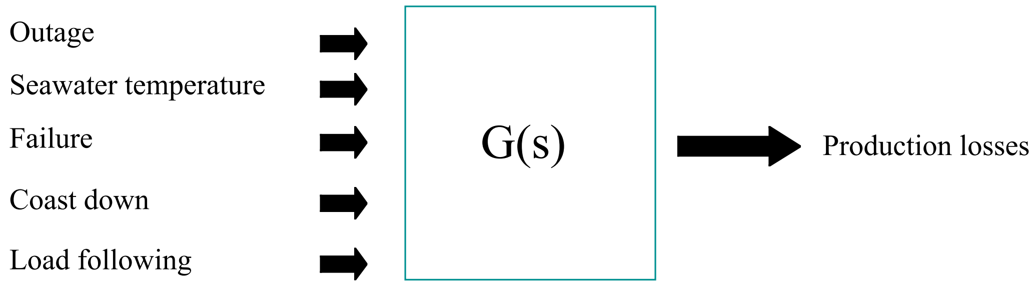


Fig 33. Input-output relationship in a nuclear power plant.

The inputs that determine the production are in the left and any problems in one of these can have repercussions on the production. The only causes that it is not determined by human intervention are the seawater temperature and failures; the others are decided by company strategies. There is someone that decides when to start the outage, the coast down and load following period. Naturally, for the component failures nobody can know when they will appear and they are not relevant to study by using deterministic modelling.

In this work we had access to the real seawater temperature data from 1994 to 1999 measured at Barsebäck (at eight different locations) and stored in the same data base as the production. These data are hourly averages. As with the production data, daily averages were also calculated based on the hourly values. In this case we are interested in making a model of how the production changes as a function of the seawater temperature. Thus not all data are of interest because if the production goes to zero for some other reason, that type of behaviour should not be represented by the model. Consequently, the data must be pre-processed in order to remove outliers and different types of disturbances.

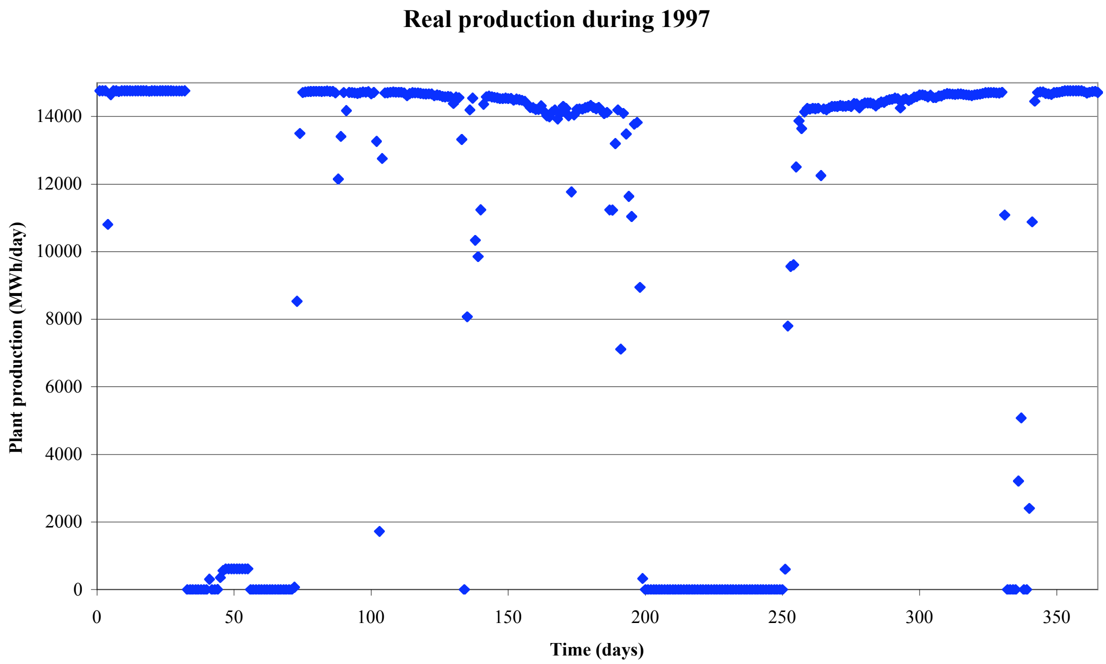


Fig 34. Raw daily production data for 1997.

4.2.1 Data pre-processing

As shown in Figure 34, we have a lot of values of energy production that do not depend on the seawater temperature. For instance, there is the revision period during which the production is zero and also various types of failures that lead to zero production. Before the data can be used to identify the relationship between seawater temperature and production such data must be removed. Based on other sources of information from the plant it was decided to only include production values between the maximum production (14760 MWh/day and 615 MWh/hour) and 14000 MWh for the daily production and 550 MWh for the hourly data.

Basically, the daily data are only a low-pass filtered version of the hourly data (calculated as an average). When we look at the number of data that is thrown away using the limits given above there is a significant difference. For the daily production data between 23% and 56% are removed (see Table 13), whereas for the hourly data only between 6.6% and 30% are thrown away (see Table 14). Due to this difference it was also decided to base one possible model on a stricter lower limit of the hourly data, in this case 600 MWh/hour. Using this limitation approximately the same number of data points are removed from the hourly data as for the daily data. Furthermore, the data for some years also contain strange values of the seawater temperature (e.g. –5), which actually indicate sensor failures. Therefore we have also imposed a limitation on the temperature data, which means that all temperature values (and the associated production values) less than zero degrees have been removed from the data series, see [1].

Table 13. Amount of data available to identify the daily model.

Year	1994	1995	1996	1997	1998	1999
Total no. of daily data after revision	327	279	286	312	320	315
No. of days used	143	192	220	223	198	209
% data used	43.73	68.82	76.92	71.47	61.87	66.35
% data removed	56.27	31.18	23.07	28.53	38.13	33.65

Table 14. Amount of data available to identify the hourly model (low limit 550 MWh/hour).

Year	1994	1995	1996	1997	1998	1999
Total no. of hourly data after revision	6688	6621	5889	5817	7114	5914
No.of hours used	5035	5084	5683	5435	4967	5005
% data used	75.28	76.79	91.41	93.43	69.82	84.63
% data removed	24.72	23.21	8.59	6.57	30.18	15.37

In Figure 35 the hourly production data is plotted as a function of the seawater temperature (using the limitations discussed above). If we compare this figure with the daily production data shown in Figure 36 (only data values higher than 14,000 MWh), we can observe the obvious similarities. However, we can also see some data that appear unrealistic. For example, in the hourly data there are a lot of production data close to the maximum value also when the seawater temperature is high, especially during 1999. However, there is a reason for this: when reactor B1 was closed (in 1998), B2 had to produce energy close to its maximum also during the summer of 1999. This demonstrates the need for good process knowledge when interpreting production data.

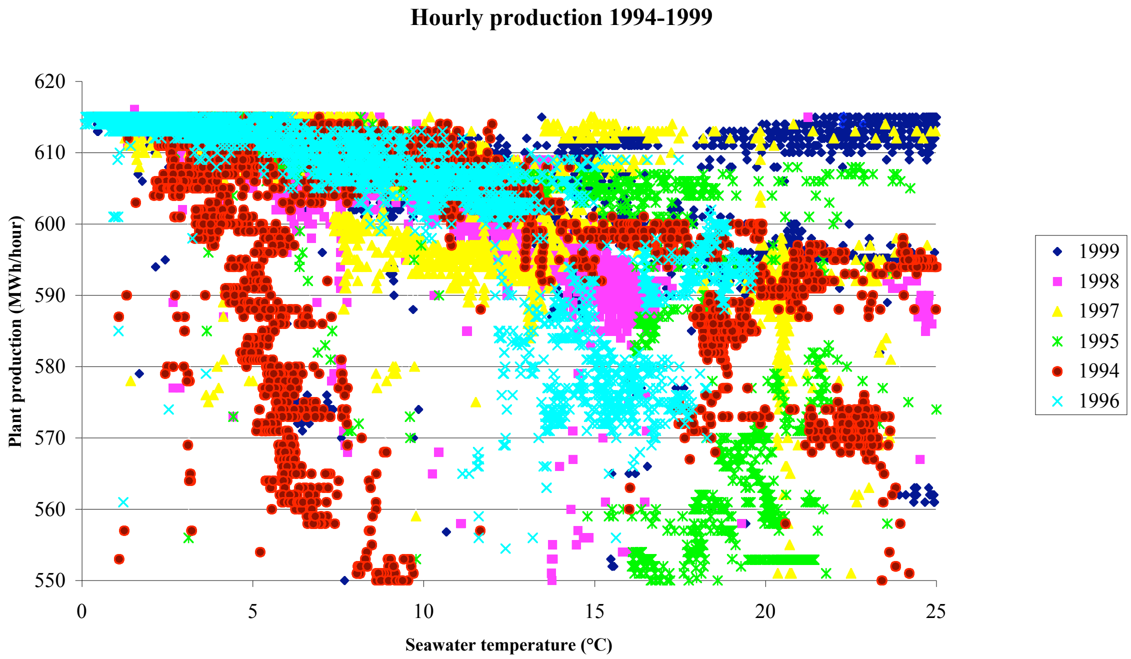


Fig 35. Hourly production from 1994 to 1999 as a function of seawater temperature.

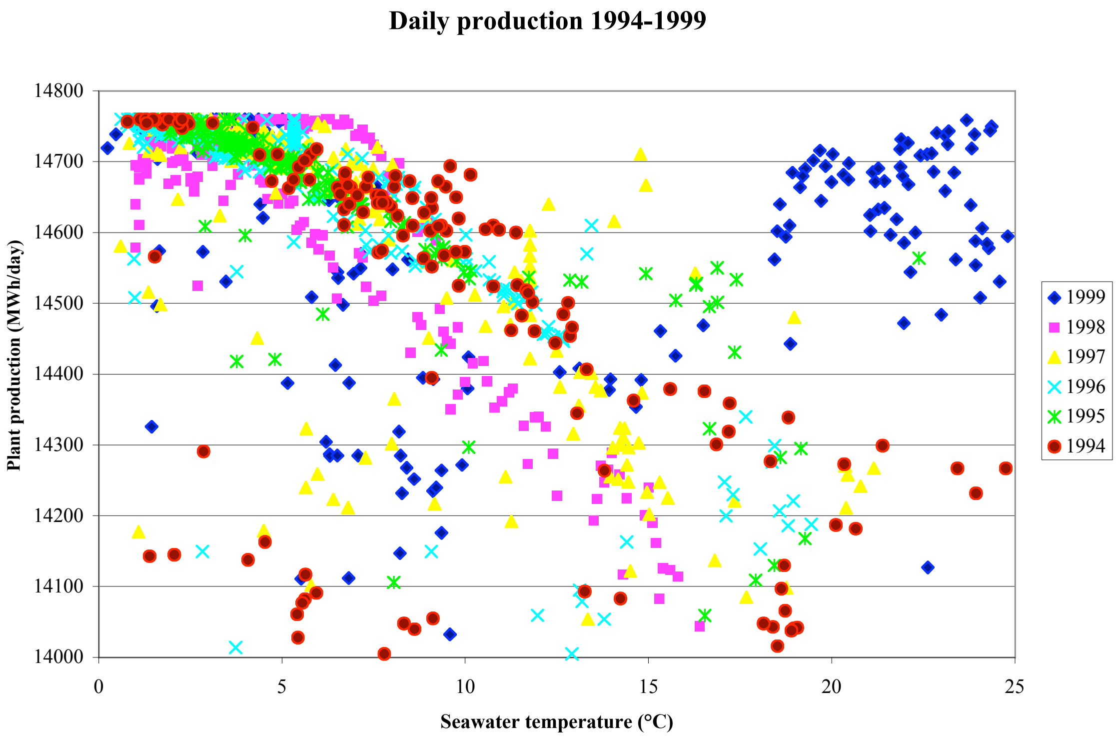


Fig 36. Daily production from 1994 to 1999 as a function of seawater temperature.

The software program Matlab, see [7] was used to pre-process the data further and identify a model. To eliminate various types of disturbances from the data, digital filters were used. A digital filter's output *y(n)* is related to its input *x(n)* by convolution with its impulse response *h(n):*

$$y(n) = (h * x)(n) = \sum_{m=-\infty}^{\infty} h(n-m)x(m)$$

In general, the *z-transform Y(z)* of a digital filter's output *y(n)* is related to the *z-transform X(z)* of the input by:

$$Y(z) = H(z)X(z) = \frac{b(1) + b(2)z^{-1} + \ldots + b(nb+1)z^{-nb}}{1 + a(2)z^{-1} + \ldots + a(na+1)z^{-na}} X(z)$$

where *H(z)* is the filter's transfer function. Here, the constants *b(i)* and *a(i)* are the filter's coefficients in two vectors: one for the numerator and one for the denominator. Many standard names for filters reflect the number of *a* and *b* coefficients present:

- when *nb* = 0 (that is, *b* is a scalar), the filter is an Infinite Impulse Response (IIR), all pole, recursive, or autoregressive (AR) filter;
- when *na* = 0 (that is, *a* is a scalar) the filter is a Finite Impulse Response (FIR), all-zero, non-recursive or moving average (MA) filter;
- if both, *na* and *nb* are greater than zero, the filter is an IIR, pole-zero, recursive, or autoregressive moving average (ARMA) filter.

For the models in this work we have chosen a moving average filter, because in the case of FIR filters, it is possible to design linear phase filters that, when applied to data, simply delay the output by a fixed number of samples. Instead, for IIR filters, the phase distortion is usually highly non-linear. The applied filtfilt function of Matlab, uses the information in the signal at points before and after the current point, in essence looking into the future to eliminate phase distortion. As the filter is used in an off-line application the non-causal implementation is not a limitation rather it avoids the general problem of time delay, which is an unavoidable drawback for all causal filters.

To see how filtfilt does this, recall that if the z-transform of a sequence *x(n)* is *X(n)*, the z-transform of the time reversed sequence *x(n)* is *X(1/z)*. The principle is shown in Figure 37.

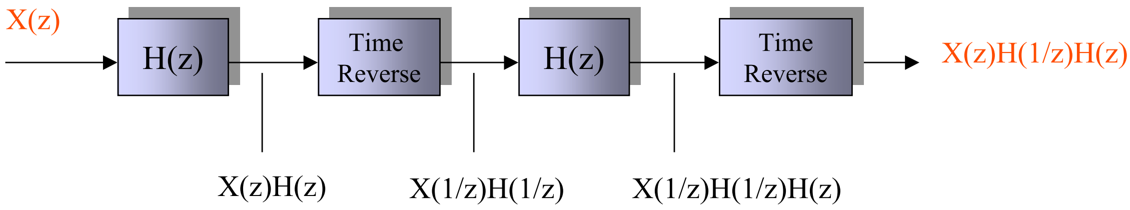


Fig 37. Non-causal filtering process.

When $|z| = 1$, that is $z=e^{j\omega}$, the output reduces to $X(e^{j\omega})\left|H(e^{j\omega})\right|^2$. Given all the samples of the sequence *x(n)*, a doubly filtered version of *x* that has zero-phase distortion is possible. We can use different number of points to calculate the average. The applied filtering reported in this report are based on 5, 10 and 15-point averaging FIR filters for the daily model and 150, 300 and 400-point averaging FIR filters for the hourly model. First we create the specific filter that we want in Matlab and then apply the function to the production and temperature vectors, see [8].

Looking at the effects of the filters we see that the phase does not change but remains the same as in the original signal. In the following figures (numbers 38 to 41) we demonstrate the effects of the different filtering approaches. The daily production and seawater temperature data are filtered using a 5- and 15-point MA filter.

The solid black line represents the behaviour of the signal after filtering and the other line shows the original signal. In this first case, the effect of the filter is less emphasised than in the case in which we base the average on 15 points. This is simply because when we use a 15-point averaging filter we base the average on more values and, consequently, the effect of the filtering is more dramatic.

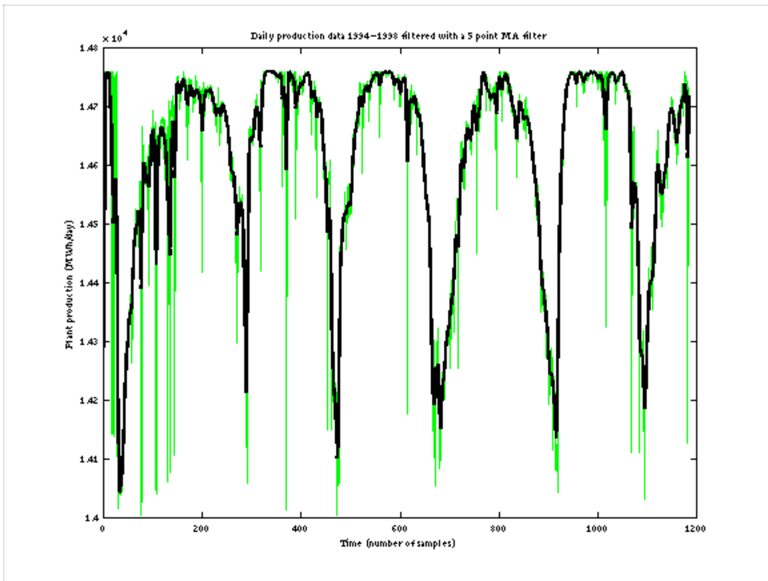


Fig 38. Daily production data filtered using a 5-point MA filter.

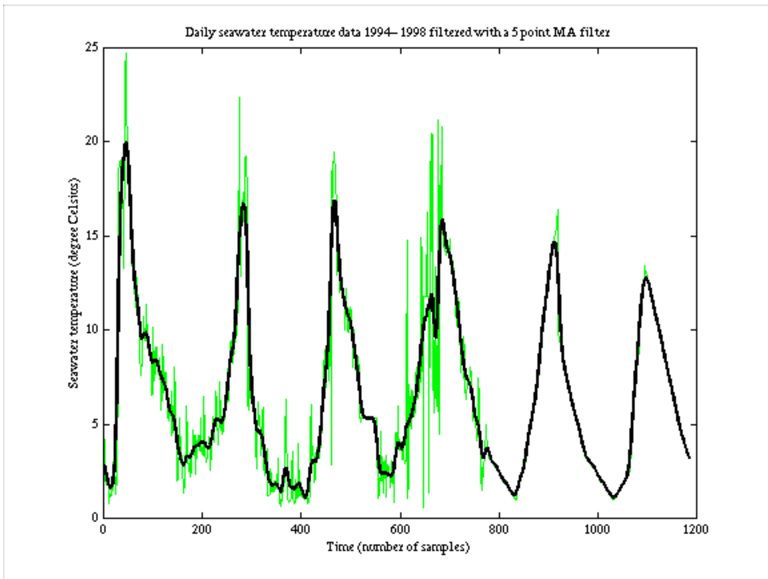


Fig 39. Daily seawater temperature data filtered using a 5-point MA filter.

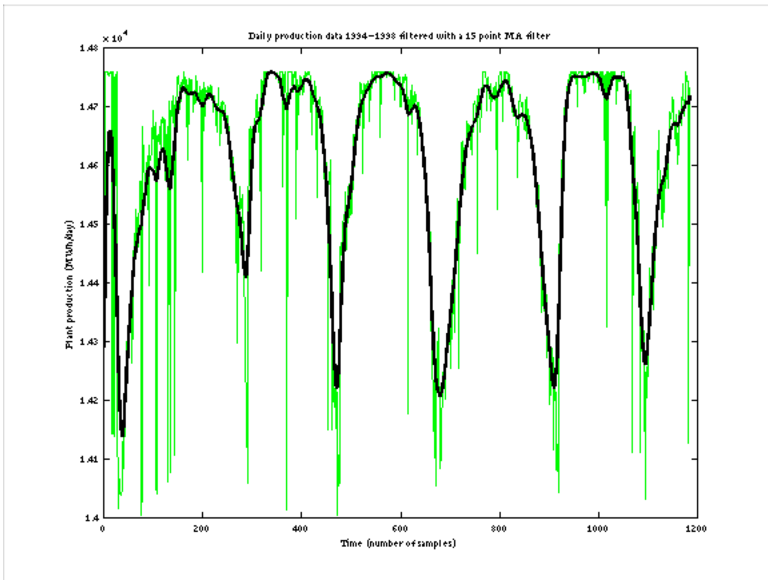


Fig 40. Daily production data filtered using a 15-point MA filter.

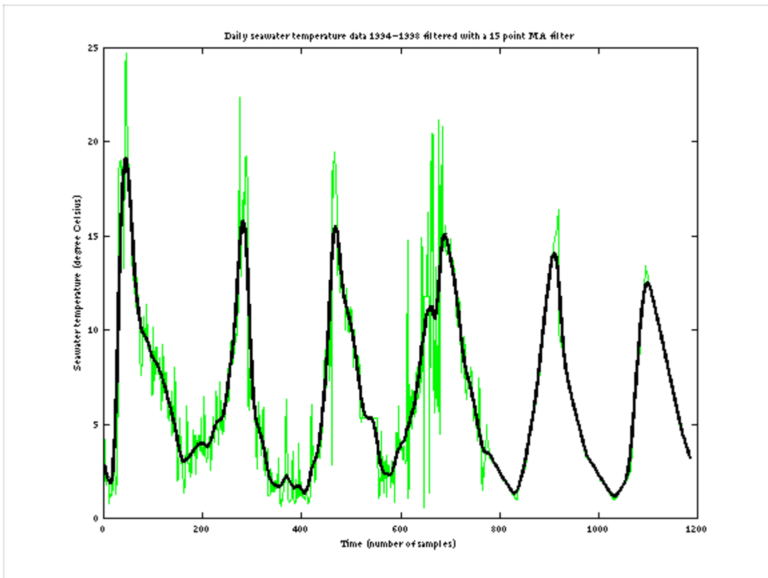


Fig 41. Daily seawater temperature data filtered using a 15-point MA filter.

From Figures 39 and 41 we can see that for the last two years the seawater temperature data have been measured with much higher accuracy, in fact the filtered signal more or less follows the behaviour of the raw data. In principle, the raw data are so accurate that no filtering of the temperature data would have been necessary if the model identification would be based only data from the years 1997 and 1998. Whether this measurement improvement is due to new and better sensors or some other reason is unknown.

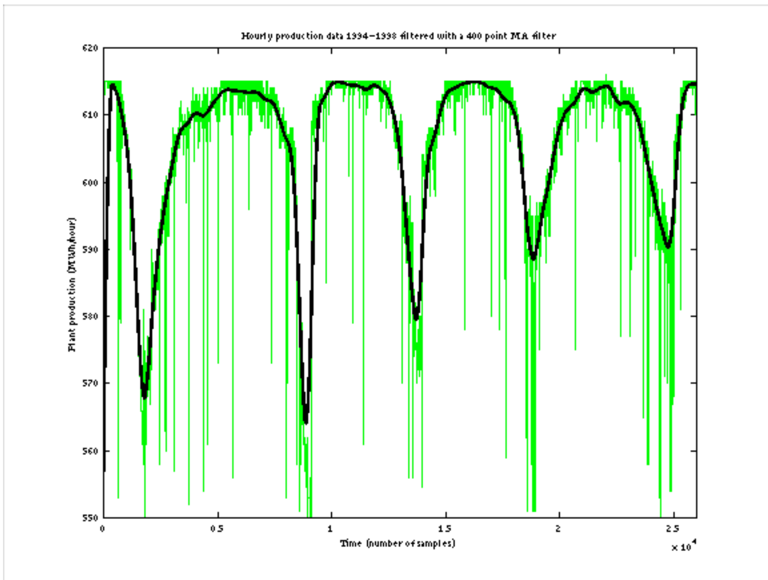


Fig 42. Hourly production data filtered using a 400-point MA filter (prod > 550 MWh).

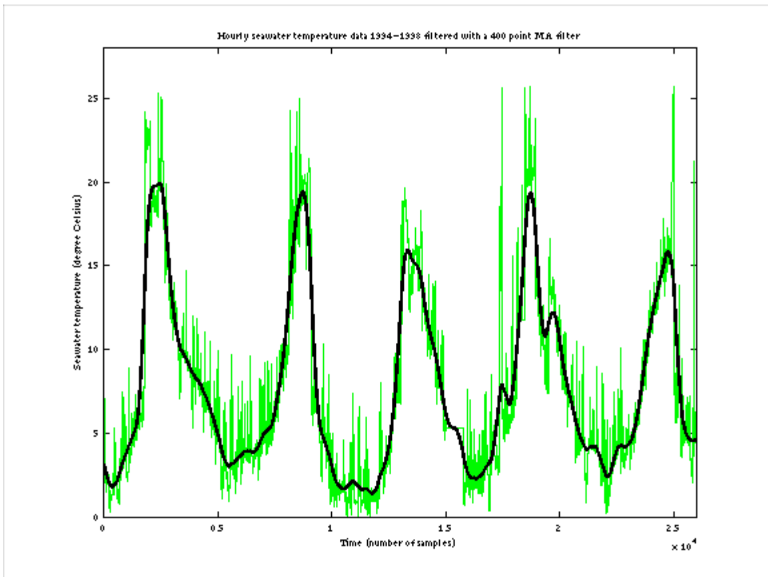


Fig 43. Hourly seawater temperature data filtered using a 400-point MA filter.

Figures 42 and 43 show the hourly production and seawater temperature data. In this case, we have used a 400-point averaging MA filter with a low limit for production of 550 MWh. The number of data points is high (due to the chosen limit) and the effect of the filtering is more significant than in the case when the signal was filtered with a 150-point MA filter, as is shown in Figures 44 and 45.

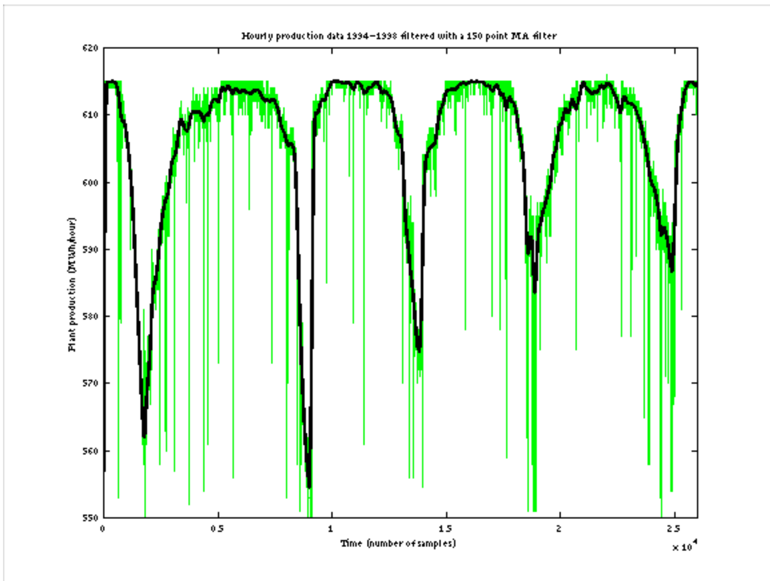


Fig 44. Hourly production data filtered using a 150-point MA filter (prod > 550MWh).

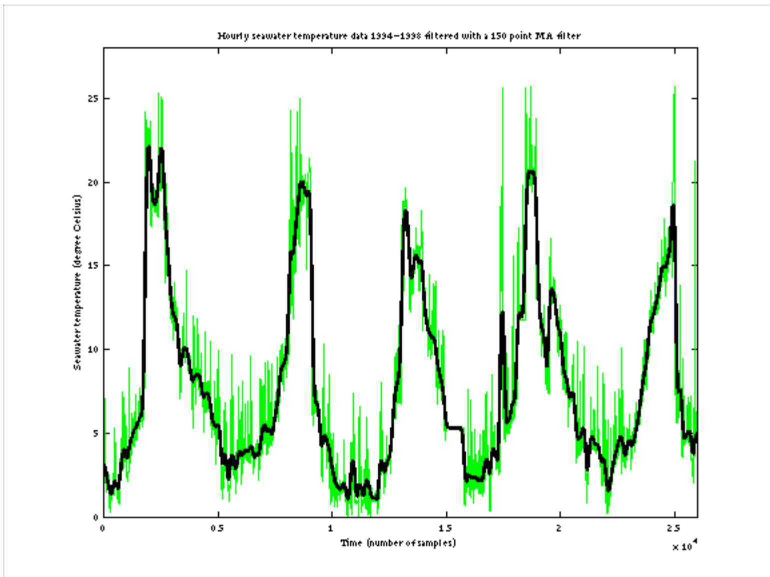


Fig 45. Hourly seawater temperature data filtered using a 150-point MA filter.

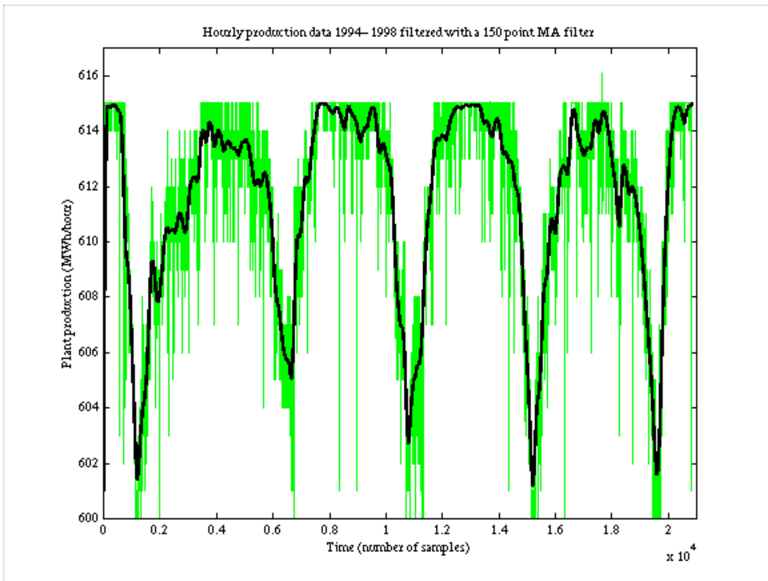


Fig 46. Hourly production data filtered using a 150-point MA filter (prod > 600 MWh).

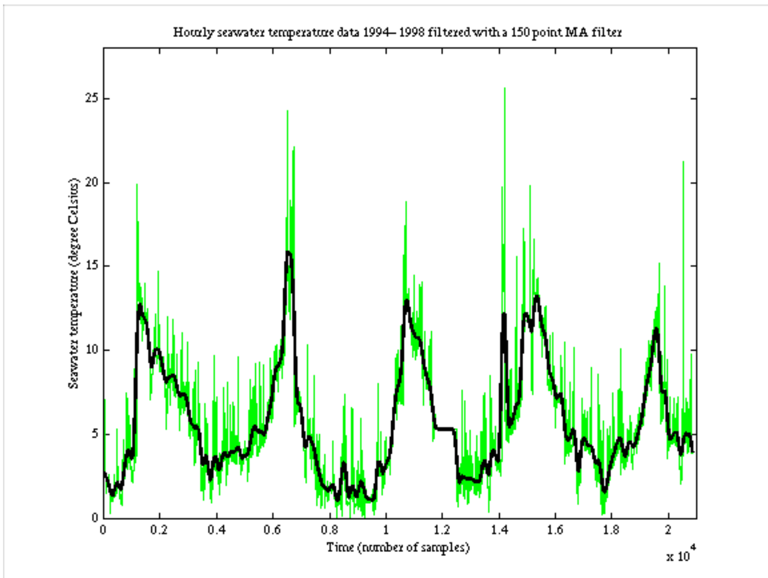


Fig 47. Hourly seawater temperature data filtered using a 150-point MA filter.

In Figures 46 and 47 we have changed the limitation for the production data: 600 MWh is the new value, but we still filter with an identical 150-point averaging filter. The disturbances in the raw data are naturally more prominent in the case of a low limit of 550 MWh. In the figures below we have also applied a 400-point MA filter to the raw data using a low limit of 600 MWh (Figures 48 and 49). The conclusions are the same at discussed above.

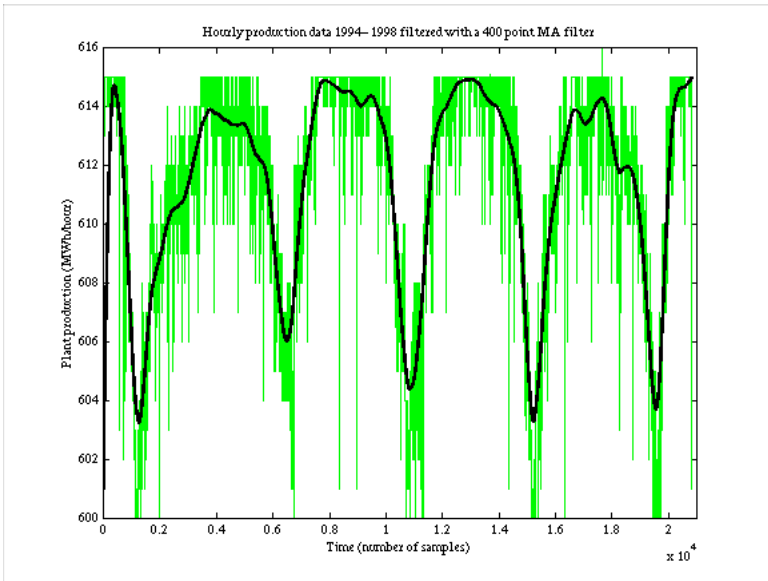


Fig 48. Hourly production data filtered using a 400-point MA filter (prod > 600 MWh).

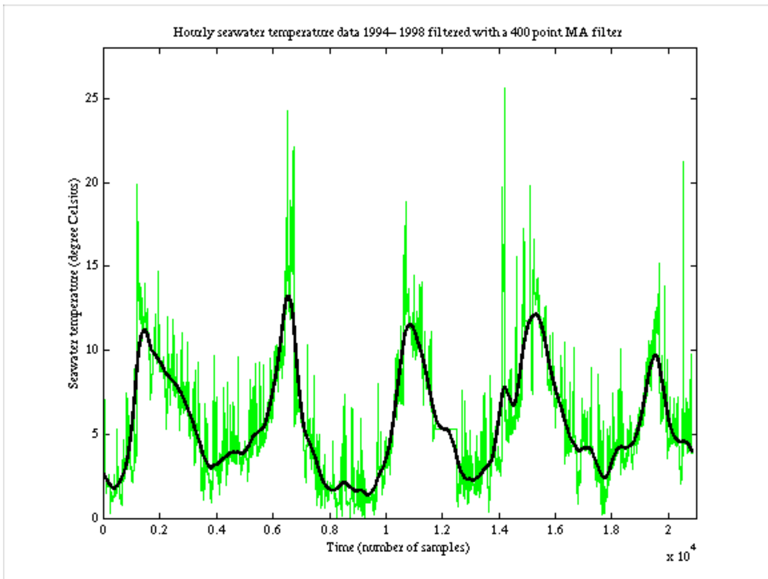


Fig 49. Hourly seawater temperature data filtered using a 400-point MA filter.

As we can see from the figures, the behaviour of the signals is strongly affected by the type of filter we apply. The more we increase the number of points for calculating the averages the stronger is the effect of the filter.

To further investigate the influence of different filters for the data pre-processing we have also tested a Butterworth low-pass filter. The characteristics of a n:th order Butterworth filter is:

$$H(z) = \frac{b(1) + b(2)z^{-1} + \ldots + b(n+1)z^{-n}}{a(1) + a(2)z^{-1} + \ldots + a(n+1)z^{-n}}$$

In the figures below (50 and 51) a 5th order filter with a cut-off frequency of approximately 1/20 day^{-1} has been applied to the daily data. In this specific case, the effect of the filter is not as significant as for the previously applied MA filters. By modifying the cut-off frequency a similar type of behaviour as shown before can be achieved. However, as the differences between the filter types are quite limited, the moving average filter was selected due to its simplicity.

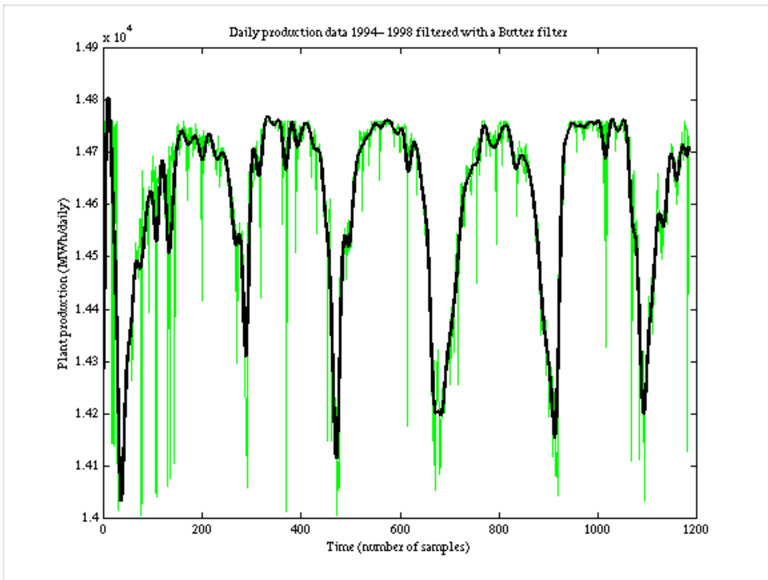


Fig 50. Daily production data filtered using a 5th order Butterworth filter.

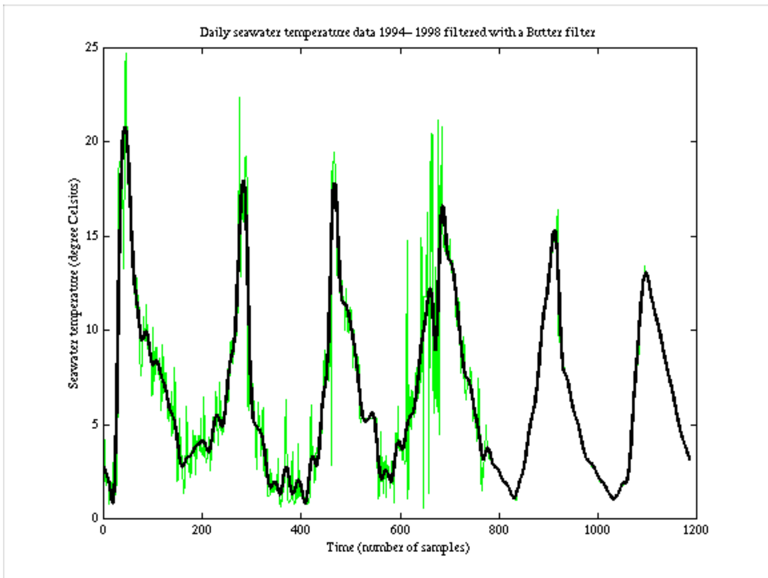


Fig 51. Daily seawater temperature data filtered using a 5th order Butterworth filter.

Based on the behaviour of the graphs shown in Figures 38 to 49 the final selection of the number of averaging points was done. To identify the models we base the averages on 10 point for the daily data and 300 points for the hourly data.

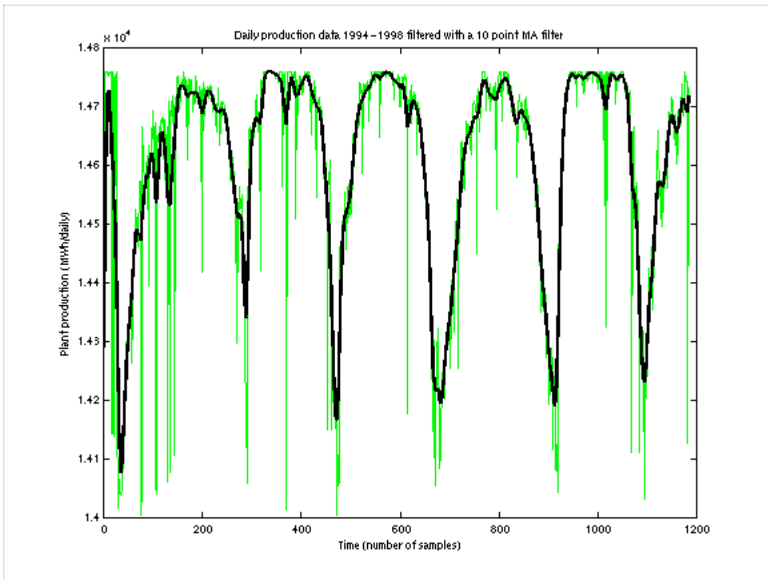


Fig 52. Daily production data filtered using a 10-point MA filter.

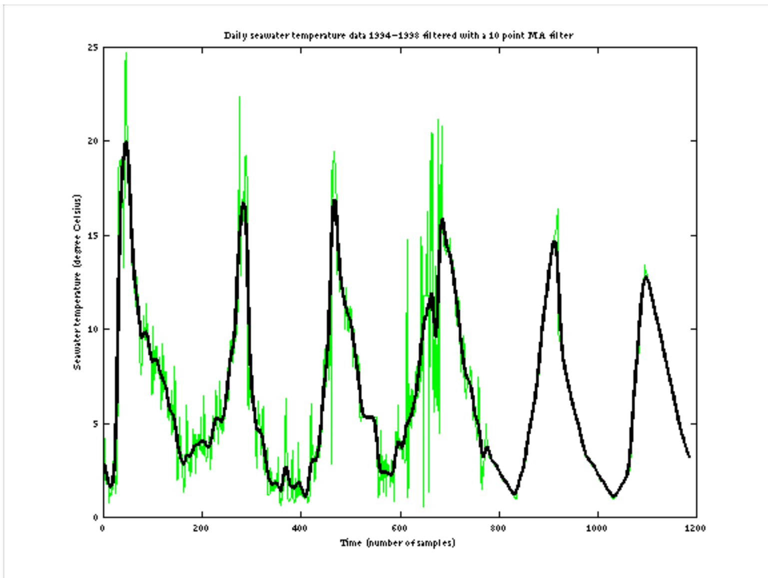


Fig 53. Daily seawater temperature data filtered using a 10-point MA filter.

Figures 52 and 53 show that the effect of this filter is more or less a compromise between the previously shown filters. For the hourly data we show only the case when the low limit of the production is set to 600 MWh (see Figures 54 and 55).

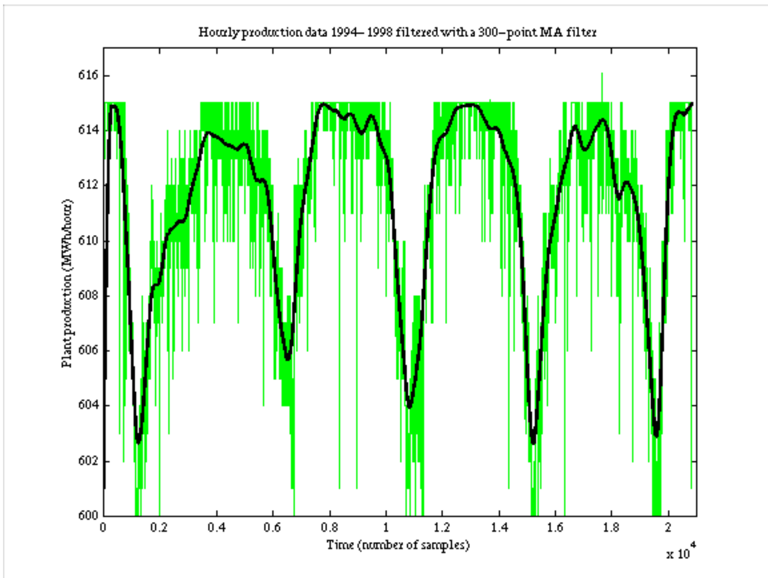


Fig 54. Hourly production data filtered using a 300-point MA filter (prod > 600 MWh).

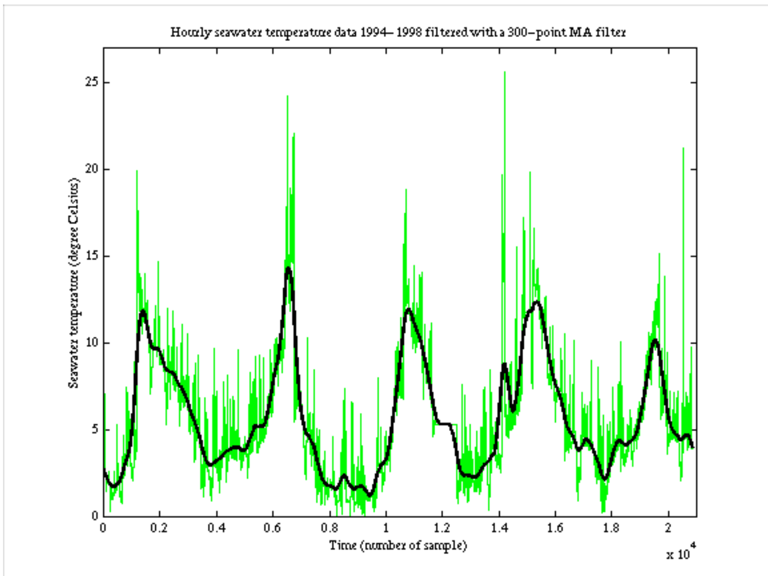


Fig 55. Hourly seawater temperature data filtered using a 300-point MA filter.

4.2.2 Model identification

In order to identify models of how the production is related to the seawater temperature (both on a daily and hourly basis) it was decided to use data from 1994 to 1998 for the actual identification, and use data from the year 1999 for validation of the models. After the data pre-processing (i.e. removal of 'unrealistic' data, limiting the data region and filtering the data), we are ready to identify the models. In this case we are only interested in identifying a static model (i.e. no dynamics). We use the Matlab function polyfit for this purpose, see [7]. Polyfit identifies a polynomial model of a specified degree in a least-square sense based on the input (i.e. temperature) and output (i.e. production) data. Furthermore, error bounds are a useful tool for determining if the data are reasonably modelled. Using the polyfit function an optional second parameter can also be obtained (in addition to the identified coefficients of the polynomial). In order to obtain the errors bounds, this second parameter is passed as an input parameter to polyval (another Matlab function). Using these functions, we identify a model and produce error bounds for a second-order polynomial model. The calculated error bounds correspond to a 95% confidence interval. Naturally, a higher-order model could be used but tests have shown that the improvement of such an approach is limited. We have chosen a polynomial model because of is quite easy and it gives good results. Of course, there is a wide choice of models: polynomial, exponential, and logarithmic.

The models describe how the production of energy varies as a function of the seawater temperature and a confidence interval of 95% has also been calculated. This means that when the production is within this interval, the possible variation is most likely due to effects caused by the seawater temperature; see Figures 56, 57 and 58.

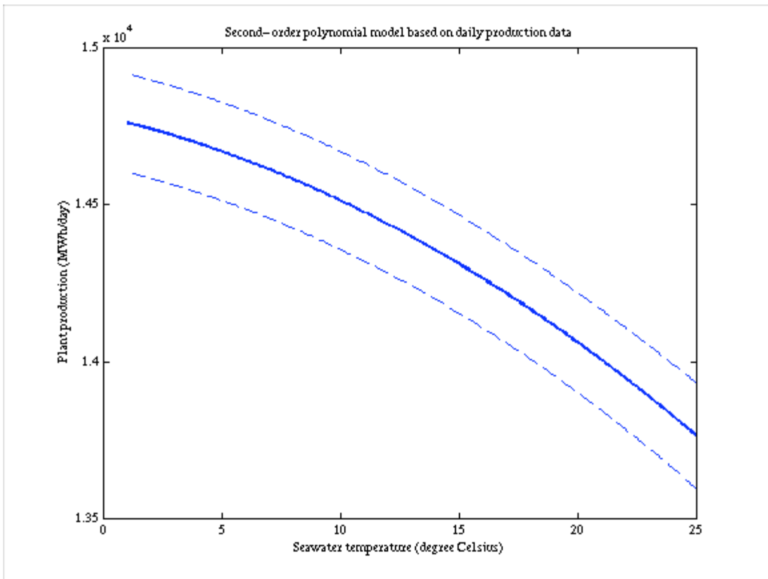


Fig 56. Model based on daily production data with a confidence interval of 95%.

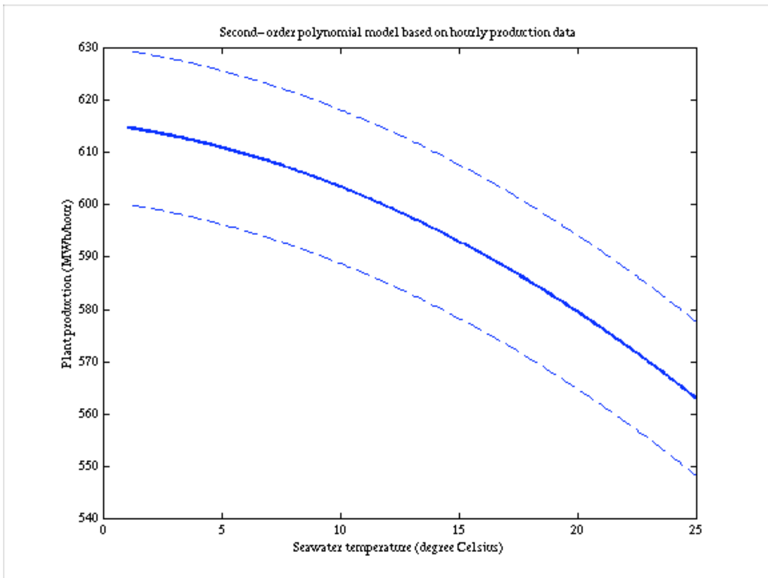


Fig 57. Model based on hourly production data with a confidence interval of 95% (production low limit equal to 550 MWh).

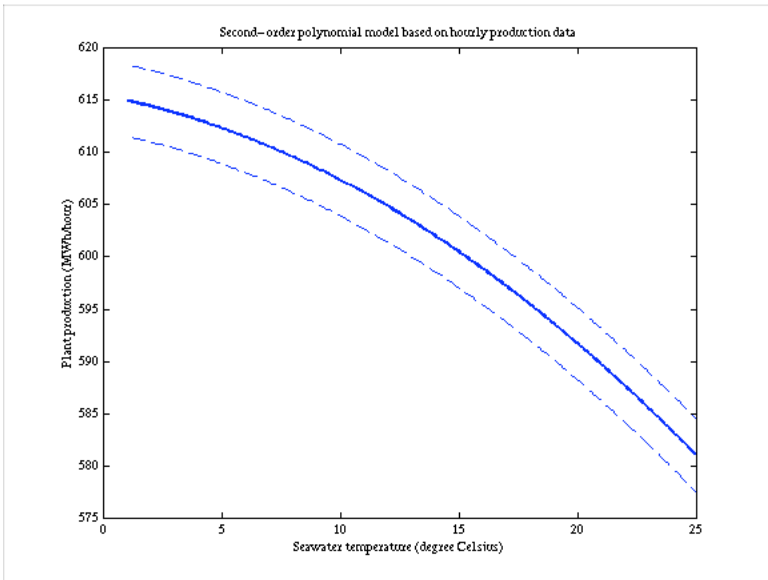


Fig 58. Model based on hourly production data with a confidence interval of 95% (production low limit equal to 600 MWh).

These three figures (56, 57 and 58) show the different models that we have obtained when using daily data and hourly data with different limitations for the amounts of production. The identified *equations* of the models are given below (*x* represents the seawater temperature in Celsius and *y* is the energy production in MWh/day and MWh/hour, respectively):

- daily production model (based on production data greater than 14000 MWh/day) $y = 14779.17 - 17.3188x - 0.9327x^2$;
- hourly production model (based on production data greater than 600 MWh/hour) $y = 615.4125 - 0.4308x - 0.0379x^2$;
- hourly production model (based on production data greater than 550 MWh/hour) $y = 615.4186 - 0.5995x - 0.0601x^2$.

The figures show that the trend is the same for both hourly and daily production data. This is an indicator that the model is reasonable, because it demonstrates that the relationship between the seawater temperature and the hourly and daily production is similar. If we consider the confidence intervals and compare the two different hourly models, we can see that the model identified from the data with a low limit of 600 MWh has a smaller confidence interval. This is an immediate effect of that more of the disturbances are removed from the data when the low limit is increased.

4.3 Model validation

Model validation is the process of gaining confidence in a model. Essentially this is achieved by "twisting and turning" the model to scrutinize all aspects of it. Of particular importance is the model's ability to reproduce the behaviour of the validation data sets.

Now we need to decide if the models are adequate descriptions for our purposes. Model validation is the heart of the identification problem, but there is no absolute procedure for approaching it. It is wise to be equipped with a variety of different tools with which to evaluate model qualities. The data from year 1999 are used for the validation. The raw data of production and temperature related to this year are plotted together with the model predictions. The results are shown in Figures 59 and 60.

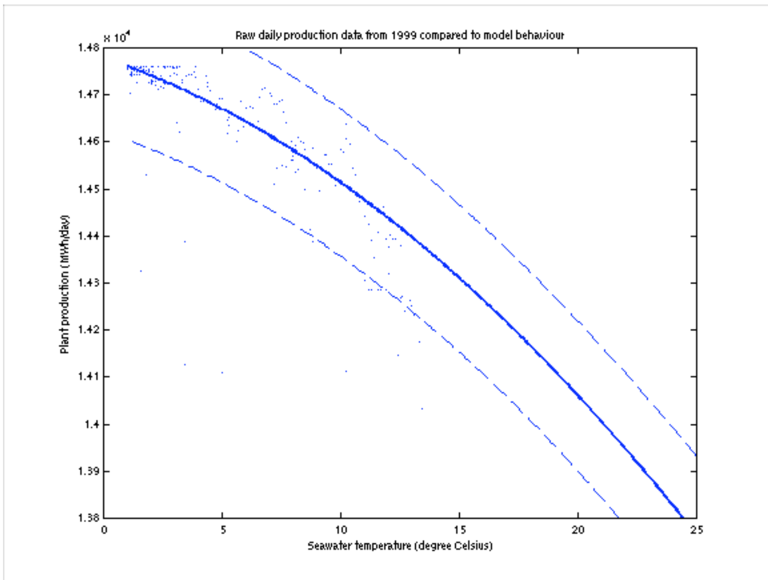


Fig 59. Validation of the daily model using production data from 1999.

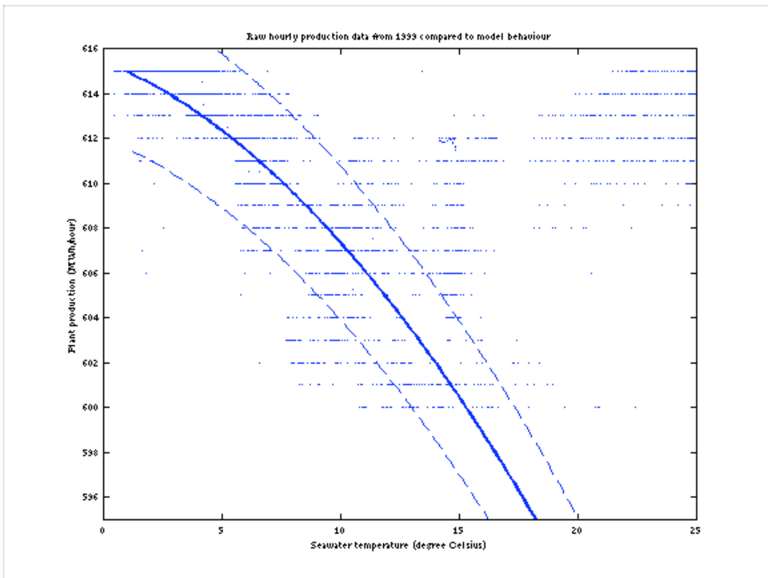


Fig 60. Validation of the hourly model using production data from 1999 (note the truncation effect of the production values in the data base, which appear as 'straight lines').

Figures 59 and 60 show that the models appear reasonable. It should be noted that all data points do not have to be inside the confidence interval (both due to the statistical variation of the raw data and because there are some disturbances that can be related to other types of events).

In order to further investigate the influence of the data pre-processing, we have also tried to identify a model based on the raw daily data, see Figures 61 and 62. Thus we have used all data (production and temperature including the outage period, outliers etc.), and applied a filter of the same type as before but with a moving average based on 25 data points (because the disturbances present in the raw data are much more severe). After the filtering we have removed unreasonable values and decided to apply the same limitations for temperature and production as before. This approach is referred to as procedure 2 in the text below. The model equation obtained using this principle is given below.

- Daily production model (based on the procedure 2)
  $y = 14653.60 - 0.6594x - 1.7876x^2$.

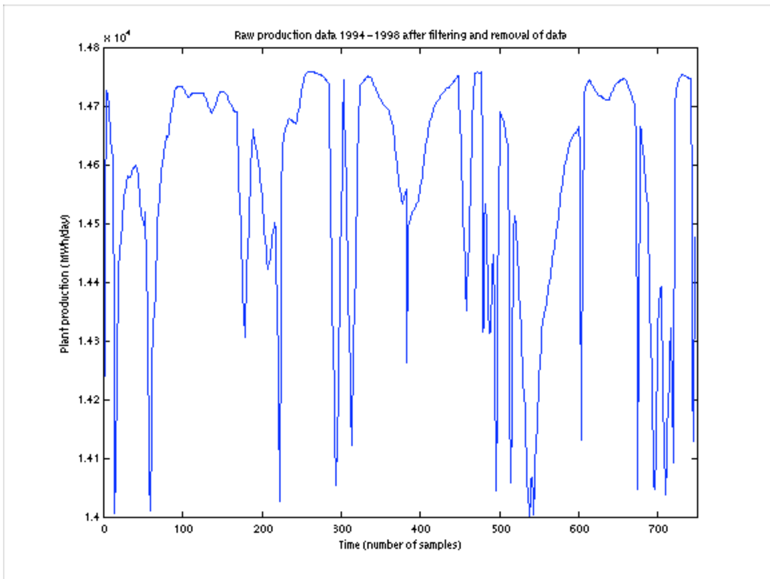


Fig 61. Production data after filtering and removal of outliers (procedure 2).

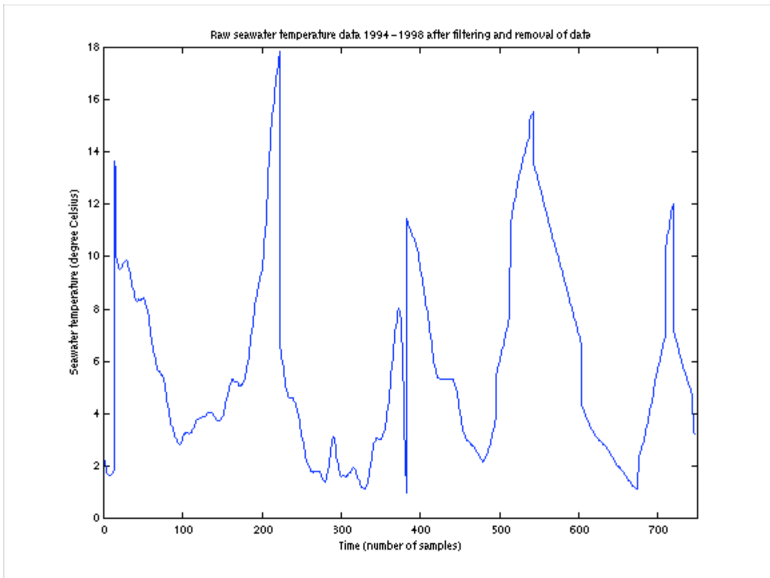


Fig 62. Seawater temperature data after filtering and removal of outliers (procedure 2).

The behaviour of the second-order polynomial model that we obtain based on these daily data is shown in Figure 63.

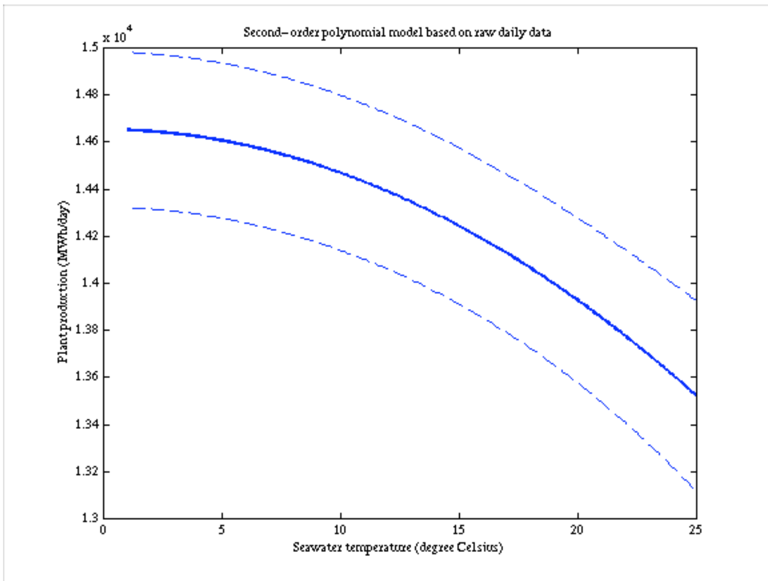


Fig 63. Model based on data using procedure 2 and corresponding confidence interval of 95%.

The general behaviour of the different types of models is almost identical; the main difference is related to the size of the confidence interval. This is logical because when the raw data are pre-processed in a filter the effects due to other causes than the seawater temperature will have a more prominent influence and the confidence interval will become larger. In Figure 64 the model behaviour is validated using raw data from 1999.

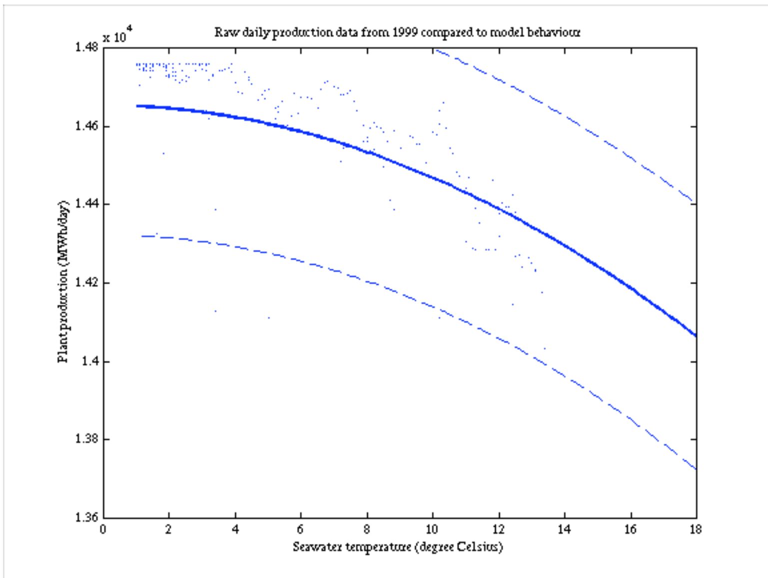


Fig 64. Validation of the model (based on procedure 2) using production data from 1999.

In this case almost all data points are within the confidence interval. However, this does not mean that the model is better (instead the uncertainty of the model is higher as is indicated by the larger span of the confidence interval). Filtering the raw data and removing outliers afterwards represents another approach to the traditional pre-processing of data and it was interesting to investigate how such an approach affected the model behaviour.

We have built these models to investigate the importance of the seawater temperature for the energy production in nuclear power plants. As have already been stated, the influence of the seawater temperature is only responsible for small variations of the overall energy production. Other factors, such as revision and component failures, are the primary reasons for the major variations of the energy production.

5 Software design

This chapter briefly describes the main problems of the software design process and presents a motivation for the use of abstract data types and how such data types should be specified.

5.1 The software design process

Modifying a software system is often called maintenance. Maintenance of large software system is not only difficult but also error-prone. Therefore, the design of reliable software should fulfil the following conditions:

- it should lead to adequate programs, i.e. programs that solve the customer's problem; this requires a correct and complete understanding of the problem by the programmer;
- it should lead to correct programs, i.e. programs that are free of bugs and thus behave the way the programmer wants them to behave;
- it should lead to programs whose maintenance is easy, i.e. programs that can be easily corrected or modified without introducing new errors.

5.1.1 Conventional software design

A conventional but naive methodology for software design consists in writing a program that is supposed to solve a given problem. Repeating the different design steps then starts the process of maintenance. The so-called software lifecycle model of Figure 65 may illustrate this methodology. The programming language is assumed to be a classical high-level programming language, such a Pascal, C, ML or LISP. The compiler is assumed to detect any syntactical errors and to provide the user with sufficient information on how to correct them.

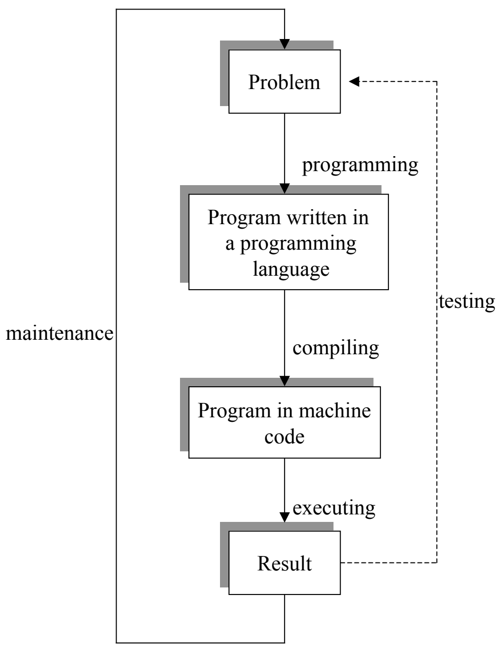


Fig 65. The different steps for the software maintenance process.

The rest of the discussion therefore excludes syntactical errors. Testing consists in running the program for ‘typical’ as well as for ‘critical’ input data that are supposed to cover the cases that may imaginably occur. The methodology described has at least the following two deficiencies. Firstly, being based on testing, it can only confirm the existence of errors, not their absence. Hence, when testing uncovers no errors the program may nevertheless be erroneous. Secondly, a deficiency of the methodology described is the fact that results are compared with ‘expectations’, i.e. the results of a program run on test data are compared with the expectations resulting from ‘one’s own understanding of the problem’. Hence testing may fail to unveil inadequacies of the program, see [4].

5.1.2 Abstraction and formalisation

The goal of a less naive methodology for software design is to avoid errors and inadequacies as far possible, or at least, to try to detect and correct them in an early stage of the design. The first step towards this goal is abstraction and formalisation, which means that the problem to be solved is described in an abstract and formal way. Being abstract, the description avoids mentioning unnecessary details; being formal, it avoids imprecision. In this way, the description may in particular help to avoid inadequacies. Again, a software life cycle model (Figure 66) may illustrate the methodology. It is obtained by extending the process of Figure 65 and refers to concepts that are briefly discussed below.

A specification constitutes an abstract description of the problem to be solved. It is abstract in the sense that it merely states the required properties of the software system to be designed. Hence, a specification concentrates on the demands of the customer. Therefore a specification is concerned with “what has to be done” but not with “how it is done”. A formal specification is a specification expressed in some formal language.

Program verification consists of mathematically proving that a program satisfies its specification. Unlike testing, program verification proves the absence of programming errors, since it proves that the program yields the correct result for any input data. Like testing, program verification cannot prove the adequacy of a program. The reason is that a mathematical proof is feasible only if the property to be proved can be stated with mathematical precision. Some formal specification techniques lead to specifications that constitute “abstract” although possibly very inefficient programs. The execution of these specifications for testing purposes is called rapid prototyping. The advantage of rapid prototyping over classical testing is that the possible detection of inadequacies occurs at an earlier stage in the software design process.

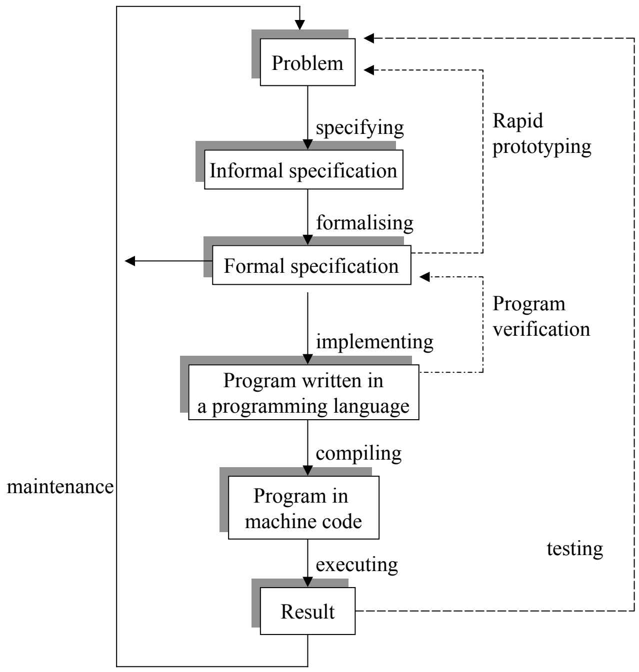


Fig 66. Extended software design process.

5.1.3 Modularisation

Modularisation consists in decomposing a problem into sub-problems and in repeating this decomposition on each of the sub-problems until the resulting sub-problems are of "manageable" size. The solution to such a sub-problem is called a module. When applied to software design, modularisation has obvious advantages. Being of small size each sub-problem may be easy to understand and solve. Different software designers may solve the different sub-problems independently from each other. Maintenance is easier because the different modules may be maintained independently. Finally, modules may be reused for other problems, consequently avoiding the repeated solution of the same sub-problem. Of course, the full benefit of these advantages is obtained only if the different subproblems are sufficiently independent. This requires that the modularisation reflect the inherent structure of the problem.

Conceptually, modularised software design suggests a top-down design or "design by stepwise refinement", decomposing a problem into sub-problems. On the other hand the use of modules suggests a bottom-up design, composing (solutions to) sub-problems into a (solution to a) "larger" problem. In practice, modularised software design generally uses a mixture of top-down and bottom-up design. "Learning by past experience" is a first step towards prevention of accidents, see [5].

The production data from the Barsebäck plant have been organised into temporal groups. As the raw data are hourly averages, daily and monthly values have also been calculated. Consequently, it is now possible to visualise the production trends based on monthly, daily and hourly data. The work with regard to the maintenance software program should also involve several aspects with regard to failure curves. In fact the goal is to be able to visualise the trends of production and simultaneously obtain information to promote an easy understanding of which causes are responsible for different production losses visible in the production data. To manage this aspect we need to connect the Bi-Cycle data base, in which there are the reporter for accidents and failures, with the conventional production data base. The production data base is structured in a similar fashion as Bi-Cycle 2000. We can use the same principles to select the appropriate time window, measurements and other choices in the program.

The chart shown in the centre of the program display (see also Figure 68) will be the production trend using different colours to indicate when the production decreases fast towards zero. Moreover, a special colour is used to show the revision period. Then it will be possible select any part of the production graph, to get information about the production trend during that specific time period including all failures that have occurred (if any). On the right side of the display, we can find all data and charts related to the causes of production loss and model predictions of these.

As discussed before, the raw hourly production data were used to calculate daily and monthly averages as well. This was done to enhance the understanding when displaying the data. The user can select which type of averages is most interesting for the current purpose and display those. In fact it is interesting to see when the production is zero for just few hours or if it is zero during several days. If the production is zero for weeks, this means that the plant is in a revision period and for this time interval it is shut down. Since the revision period is the most significant cause of production loss, it is important to clearly indicate this interval. Consequently, the production trend during such intervals are displayed using different colours.

5.1.4 Data types

The data are stored in a data base and are available as hourly averages (see Table 15). The data are structured in a matrix format where the rows represent time and the columns contain the production and signals from sensors and actuators.

Table 15. Example of data base contents for production data.

DATE	M211K116	M211K126	M211K127	M211K405
12/1/99 0:00	70,29	70,19	70,1	4,01
12/1/99 1:00	70,3	70,23	70,12	4
12/1/99 2:00	70,33	70,23	70,14	4
12/1/99 3:00	70,3	70,19	70,12	4
12/1/99 4:00	70,31	70,21	70,12	4
12/1/99 5:00	70,29	70,16	70,12	3,99
12/1/99 6:00	70,34	70,23	70,16	4
12/1/99 7:00	70,32	70,26	70,12	4,01
12/1/99 8:00	70,3	70,23	70,1	3,99
12/1/99 9:00	70,32	70,19	70,14	3,99

Outage

The revision period is characterised by zero production and because of this it is the largest cause of production loss. The most important thing is to determine the duration of this period. However, this data are not directly available from the production data base, rather it has to looked for in different production sheets which are updated by the staff at Barsebäck. Therefore it was necessary to create an Excel file with these data to be able to show the duration trend for revisions. Moreover, during these periods basically no data are available. In order to calculate the production loss, we used the information that the plant each day can produce at maximum of 615 MWh*24h, i.e. the maximum daily production loss is 14,760 MWh. This value was used for the analysis. It is then possible to show that the duration of the revision period and the production loss have identical trends. When the plant is restarted after the revision period the production does not immediately regain its maximum value instead it takes a few days. If we want to know how the production increases this information is available from the production data base.

Coast down

The production loss due to coast down is also not directly available from the data base. The information may be found in manual production sheets in which there are the sums of production losses divided by each cause for each month. An Excel file was created based on these data and the trend of coast down could then be represented by graphs (see Table 16).

Table 16. Example of production data related to coast down.

Gross production May-Aug 95	439271.4	398116.4	392713	264014.5
Prod. loss coast down May-Aug 95	1562.55	37211.55	78636.25	133580.3
Gross production March-Jun 94	431986	387769	313848	280852
Prod. loss coast down March-Jun 94	7422	30448.27	68208.81	85945.47

For the second indicator for coast down (i.e. the flows from certain pumps discussed in Section 3.1.2), we have compared the flows when coast down is not present with when it is present. In the example below, we have chosen data from the year 1994 when coast down appeared and from 1998 when no coast down occurred. Based on the available hourly pump data daily and monthly averages were created (similar as discussed for the production data), see Table 17. This data are then used in a histogram to show the differences.

Table 17. Example of data regarding the flows of a coast down related pumps.

	January	February	March	April	May
pump M313K034 1998 (kg/s)	5,362.198	5,384.74	4,422.66	5,479.47	5,119.34
pump M313K034 1994 (kg/s)	2,507.841	6,832.787	7,074.503	7,128.028	6,499.899
pump 512K104 1998 (kg/s)	10,268.95	10,234.59	6,909.41	10,131.62	10,328.92
pump 512K104 1994 (kg/s)	5,025.938	10,768.78	10,666.7	10,717.02	9,420.677

Component failures

Component failures are difficult to analyse because they appear more or less at random. However, with the help of the Bi-Cycle we can show which causes are more recurrent. In order to do this the Bi-Cycle data base is essential. For example, we can select a specific year and get access to all failures divided by different classes (see e.g. Table 9). Then this information can be exported as an Excel file and used in various diagrams. However, this is not always interesting because a lot of failures are without specific classification (i.e. no class). More interesting is often to show the ten components that are subject to the highest number of failures for each year. In addition, we can use Bi-Cycle to find the correlation between the components and the failures for each year.

Load following

The problem of load following is interesting because it is present almost every year. Also in this case the information is not directly available from the data base. The information may be found in manual production sheets in which there are the sums of production losses divided by each cause for each month. An Excel file was created based on these data (see Table 18) and the trend of load following can then be visualized.

Table 18. Example of data related to load following.

	Production loss due to low prices (MWh)	Total production (MWh)
1995	25,203.3	3,890,487
1996	198	3,900,155
1997	28,550	4,042,937
1998	170,317	4,171,528
1999	913,40.21	3,488,644

5.2 Data structure and user interface

The principal aim of this work is to use only one data base where the engineer registrers production loss data every day and where he can immediately see the trend of the production. The production data base has links to BiCycle. BiCycle has links to IDUN (operation and maintenance information system), which makes it possible to present for instance failure data together with the production data. When the engineer requests a graph on the production trend, he gains knowledge about the causes that have influenced the losses of production.

Barsebäck has already a production data base PDB (based on ORACLE), which will be used to link the raw production data to Bi-Cycle. The principle data base structure is presented in Figure 67.

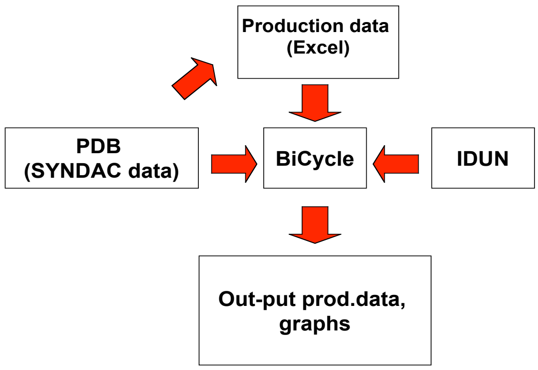


Fig 67. Principal data base structure.

The production losses data base (based on Excel) includes the real production values and planned production for every day, see Table 19. The causes of production losses are divided into:

- causes that do not effect the availability of the gross production (coast down, external failure, load following, seawater temperature);
- causes that have an influence on the gross production (outage, test…).

To further enhance the understanding, the following types of indicators can be used to measure the production capacity performance including the losses:

- Energy utilisation factor: $\dfrac{Gross\,Production}{CalendarTime \times Max\,Production(615MWh)} \times 100$
- Time utilisation factor: $\dfrac{Power.to.the.grid}{CalendarTime} \times 100$
- Energy availability factor: $\dfrac{PossibleGross\,Production}{CalendarTime \times Max\,Production(615MWh)} \times 100$
- Time error factor: $\dfrac{TotalHourStopForFailure}{TotalHourStopForFailure \times TotalHour\,Production} \times 100$

The Excel file shown in the table below is really useful because we can directly use this data in Bi-Cycle to achieve a better view of the production behaviour. In Figure 68 the solid line represents the total production (taken from the Excel file). It is plotted together with the number of failures shown in the corresponding histogram (divided into functional and non-functional failures).

Table 19. Production data in Excel format.

		not affect the availability gross prod				*affect the availability gross prod*			
Date	**Real prod**	**Coast down**	**Load following**	**Ext failure**	**Seawater temp**	**Revision**	**Failure efficiency rate**	**Test**	**Failure**
01-jan	14736				24				
02-jan	14699				61				
03-jan	14694				6			60	
04-jan	10116				4			4640	
05-jan	8770				10			5980	
06-jan	14567				4			189	
07-jan	14707				53				
08-jan	14686				74				
09-jan	14711				49				
10-jan	14700				60				
11-jan	14698				62				
12-jan	14702				58				
13-jan	14699				61				
14-jan	14710				50				
15-jan	14714				48				
16-jan	14705				55				
17-jan	14682				78				
18-jan	14716				44				
19-jan	14720				40				
20-jan	14713				47				
21-jan	14737				23				
22-jan	14747				13				
23-jan	14743				17				
24-jan	14753				7				
25-jan	14753				7				
26-jan	11931				9			2820	
27-jan	10920				15			3825	
28-jan	13486				4			1270	
29-jan	14731				29				
30-jan	14712				48				
31-jan	14743				17				
Total	**437701**	**0**	**0**	**0**	**1077**	**0**	**0**	**18784**	**0**

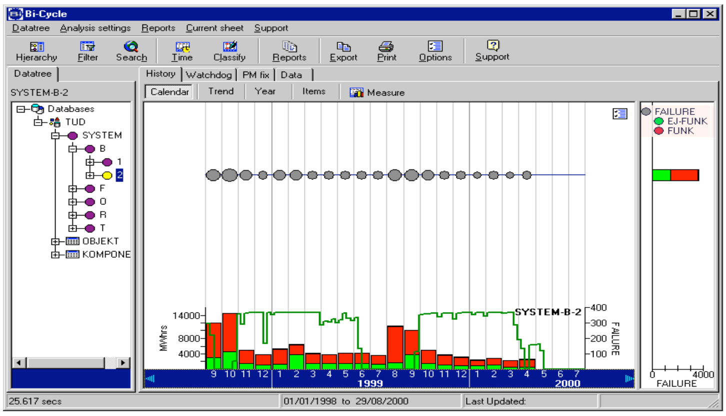


Fig 68. Representation of total production related to the number of failures (from Bi-Cycle).

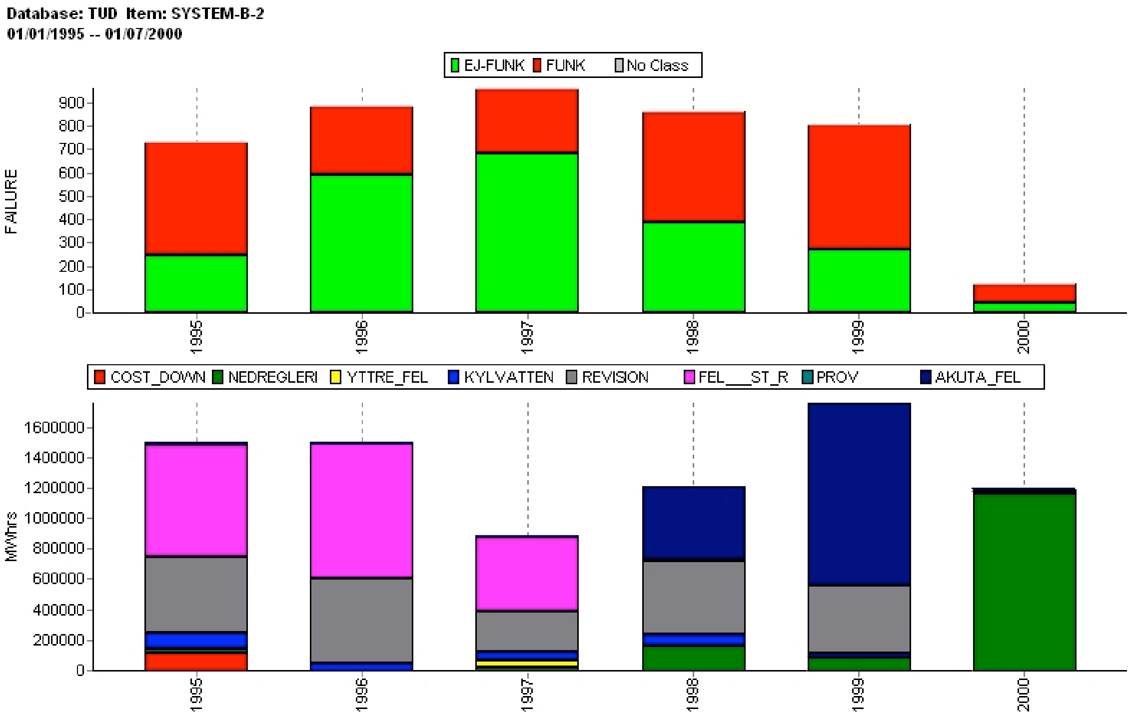


Fig 69. Yearly amount of production losses related to the type of failure (from Bi-Cycle).

Figure 69 is an example of how to work with the combined data bases. The first chart shows, for each year, the division between functional and non-functional failures (both these types of failures cause production loss but represent different classes of failures). The data are coming from the IDUN failure report data base. In the second chart we see the total amount of production loss for each year. Moreover, the production loss is now divided into different sections, where each section is related to a specific type of failure. For example, when examining the information for 1995 we can see that the largest part of the production loss is due to component failures, followed by outage, load following and coast down (in descending order).

Figure 70 represents information from the years 1995 to 1999 and shows the behaviour of the real production (solid line) compared and plotted together with the number of failures shown in the corresponding histogram (divided into functional and non-functional failures).

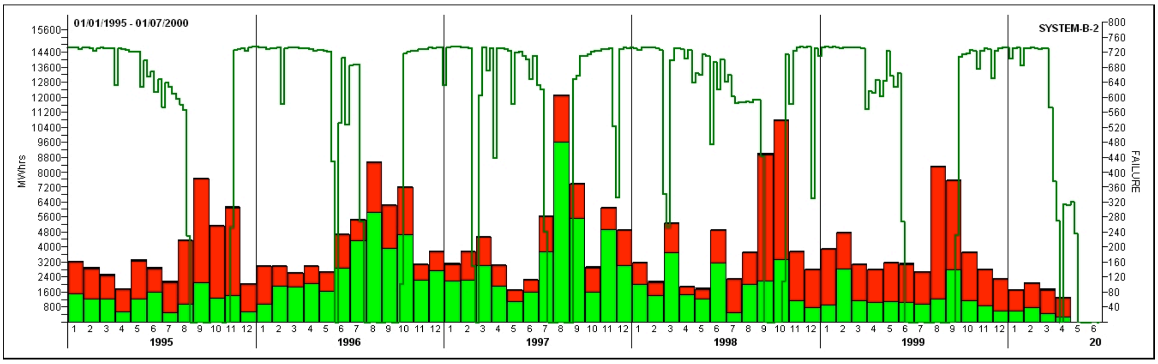


Fig 70. Representation of total production related to the number of failures from 1995 to 1999 (from Bi-Cycle).

6 Conclusion

6.1 The data concept

In order to understand the essential features of a system or a process it is necessary to make measurements. These measurements are called data and are collected to understand the properties of the process. It is important to realise that data contain both systematic information about a given system and unwanted variations (noise). The noise is a mixture of measurement errors, sampling errors and other sources of variability. Obviously, the mere fact that we have done a lot of measurements does not per se imply that we know the properties of the system. All data must be processed and analysed by appropriate methods to reveal the systematic information. The conclusion is that data is not synonymous with information; but that data must be processed with data analytical tools to extract the information, see [3].

6.2 The model concept

It is of the utmost importance to recognize that a model is an approximation, which simplifies the study of reality. A model will never be 100% perfect, but may still be useful. Models are not reality, but approximate representations of some important aspects of reality. Provided that a model is sound – there are tools to test this – it constitutes an excellent tool for understanding important mechanisms of reality, and for manipulating parts of reality towards a wanted outcome.

Certain classes of mathematical models are discernible, i.e., empirical models, semi-empirical models and theoretical models. A theoretical model, also called a hard or fundamental model, is usually derived from a well established and accepted theory building within a field. Theoretical models are often regarded as fundamental laws of natural science even though the label model would certainly be more appropriate in experimental discipline. However, in most cases, the mechanisms of a system or a process are usually not understood well enough, or may be too complicated, to permit an exact model to be postulated from theory. In such circumstances, an empirical model based on experiments might be a valuable alternative. A semi-empirical model is a local model, which describes the situation within the investigated interval.

6.3 Summary of results

The seawater temperature is independent of the company’s decisions. On the contrary, there is someone who decides when outage and load following are initiated and the effects of coast down is also a fairly well-known phenomena. On the other hand, component failures are more or less random and much more difficult to study with traditional models. Consequently, the only available independent variable is the seawater temperature, which has an influence on the production during the entire year.

At this time the models are primarily developed to be used for investigations of the historical behaviour of the production. However, they can rather easily be modified for on-line purposes and in that sense be used to identify deviations in the plant production and determine whether these deviations are related to changes of the seawater temperature or if the causes are related to some other type of event.

In Barsebäck, there is a software program with which the production is planned and the production losses predicted including those related to seawater temperature. This model is based on more theoretical and ideal assumptions and is probably more suited as a planning tool than for on-line purposes. The models developed in this work are entirely empirical (i.e. only based on an input-output relationship) and the work also suggests methods of how to deal with disturbances and non-valid data. Consequently, they are more related to on-line applications. However, it is encouraging that the Barsebäck planning model and the models identified in this work demonstrate a similar behaviour and that the output from the ideal planning model is well within the 95% confidence interval suggested in this work. Since the two types of models have been developed independently the validity of the models for their respective purposes is made more credible.

6.4 Generalisation of the problem

The kind of study carried out in this work is not only relevant for nuclear power plant, but the same principle may be used for various production companies where different events and phenomena cause production losses.

The analysis of experimental and process data consists of three primary stages:

- evaluation and pre-processing of raw data;
- model derivation and interpretation;
- model validation and use.

The work may be continued and extended in several ways, especially the modelling part. For example, the need to predict various aspects related to failures may arise (frequency of failures for different components, probability of a failure occurring within a certain time horizon, the effect of explicit failure types on the production etc.). From a modelling point of view the type of static and deterministic models used to describe the relationship between seawater temperature and production is not suitable. Instead such an approach should be based on stochastic models or models that are based on probability distributions (since failures occur in a more or less random fashion and it is only possible to determine a probability that a certain type of failure will appear). Another interesting approach would be to create models where all types of factors that are related to production losses could be included simultaneously. Such multiple input models are much more complex than traditional single input/single output models. However, multivariate analysis and multivariate projection methods represent a highly interesting approach to deal with such types of systems. Therefore, a short introduction to multivariate analysis is given below, which may be inspirational for various continuations of this work.

Today, the reality for experiment has changed. Data matrices are no longer typically long and lean, but rather short and fat. This causes problems for the classical methods of statistics and raises new demands on the data analytical techniques. Short and fat data structures arise because it is no longer difficult and time-consuming to measure variables. Due to the introduction of modern electronics, a vast array of technical instruments have been devised, which are capable of outputting hundreds or thousands of variables within a short period of time. The existence of large data tables in general, and short and fat matrices in particular, necessitates the use of multivariate projection models like PCA (Principal Components Analysis) and PLS (Projections to Latent Structures). For multivariate projection methods, such as PCA and PLS, the basic conceptual model is such that the variable correlations are modelled as arising from a small set of latent variables, where all the measured variables are modelled as linear combinations of these latent variables. This is illustrated in Figure 71 (compare also with Figure 5, which represents the system in this work). A process is characterised by registering six signals (variables). Note that many more signals could be used. In the ensuing multivariate analysis, the information in these six variables is contracted to a few informative variables, latent variables.

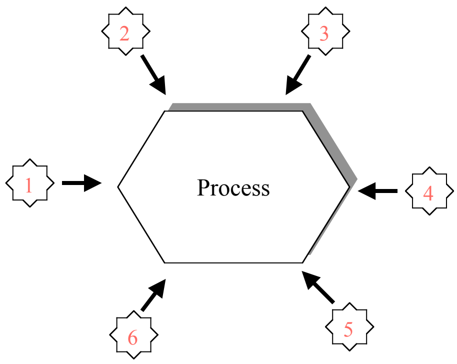


Fig 71. Example where six variables are measured to characterise the behaviour of a process.

PCA is a multivariate projection method that is designed to:

- extract and
- highlight the systematic variation in a multivariate data matrix.

This means that the primary objectives of PCA are:

- to evaluate the underlying dimensionality (complexity) of the data;
- to get an overview of the dominant patterns and major trends in the data.

Therefore, PCA summarises the information residing in the initial data matrix into a form, which may be more easily overviewed and used. The original multi-dimensional space, defined by the number of measured variables, is contracted into a few descriptive dimensions, denoted principal components, which represent the main variation in the data. Each principal component can be displayed graphically and analysed separately, and its meaning may often be interpreted according to simple technical fundamental factors.

Prior to PCA data are usually pre-processed by means of mean centring and scaling to unit variance. Plots of PCA scores are invaluable for viewing relationships among the observations, e.g., for finding outliers. It is here appropriate to make a distinction between strong and moderate outliers. Strong outliers that are found in score plots, conform with the overall correlation structure, whereas moderate ones, which are found in residual plots, break the general correlation structure. Moderate outliers do not show the same profound effect on the model building, as does a strong outlier.

Partial least square projection is a regression extension of PCA, which is used when it is of interest to connect the information in two blocks of variables, x (inputs) and y (outputs), to each other. The explanation of PLS is made both for the situation when only one response variable is modelled, and when several responses are analysed at the same time. PLS derives its usefulness from its ability to analyse data with many, noise, collinear, and even incomplete variables in both x and y. PLS has the desirable property that for the parameters regarding the observations, the precision improves with the increasing number of relevant variable. PLS can be seen as a certain technique of generalised regression to model the association between x and y, but it can also be seen as a philosophy of how to deal with complicated and approximate relationships. As in any data analysis application, data are usually pre-processed prior to using PLS. PLS modelling works best when the data are fairly symmetrically distributed and have a fairly constant error variance. In addition, data are usually centred and scaled to unit variance before the analysis. This is because in PLS a given variable has an influence on the model parameters that increases with the variance of the variable. Scaling all variable to unit variance corresponds to the assumption that all variables are equally important a priori. PLS provides many parameters and diagnostics, which are of utility for model interpretation, and assessment of model performance and relevance. Whenever one wishes to model one or several response variables, y, by a linear model based on a set of correlated x-variables, the PLS method is a good choice.

The rather natural assumption underlying the method, namely that the predictor and response variables are correlated and possibly also noisy and incomplete, is more in line with reality than those of classical regression. In PLS, the variable correlations are modelled as arising from a smaller set of latent variable, where all the measured variables are modelled as linear combinations of these latent variables. PLS has the ability to model and analyse several y-variables together, which has the advantage of giving a simpler picture than separate models for each response. In general, when the y-variables are strongly correlated, one can recommend than they are analysed together, since the correlations stabilise the model. Another attractive property of PLS lies in its ability to cope with almost any type of data matrix. For instance, the precision and reliability of the PLS parameters related to the observations, is enhanced by increasing the number of relevant variables. Also PLS work wells with short and fat matrices.

Multivariate Data Analysis is used when we need to measure many things, many variables, many properties of the systems and processes. Hence, data collected in science, technology, and almost everywhere else are multivariate, a data table with multiple variables measured on multiple observations. Multivariate data, well measured on intelligently selected variables contain much more information than univariate data, and hence an adequate multivariate characterization of samples, systems and processes of interesting, is a necessary first step in their investigation. However, to use a multivariate data set to reach insight about the studied system, it is not enough to just look at the data table. Rather, the data must be analysed so that the desired information in the data is expressed in a way that we can grasp, for instance a graphs or two, or a few information-rich parameters. This approach is useful in science and technology for a wide variety of applications. This information content in collected multivariate data can be expresses in term of plots and list of parameters resulting from a multivariate data analysis. These results help to improve processes and productions, as well as improving the efficiency of research and development. In industry this may lead to great savings of cost as well as increase income from higher yields and higher quality.

References

[1] T. W. Anderson (1958), *The Statistical Analysis of Time Series*, John Wiley & Sons, Inc., Stanford University, USA.

[2] David R. Brillinger (1975), *Time Series Data Analysis and Theory*, International Series in Decision Process, USA.

[3] L. Eriksson, E. Johansson, N. Kettaneh-Wold and S. Wold (1999), *Introduction to Multi- and Megavariate Data Analysis using Projection Methods (PCA & PLS)*, UMETRICS, Sweden.

[4] B. V. Gnedenko, Yu. K. Belyayed and A. D. Solovyev (1965), *Mathematical Methods of Reliability Theory*, Academic Press, New York, USA.

[5] E. J. Henley, H. Kumamoto (1985), *Designing for Reliability and Safety Control*, Prentice-Hall INC., New Jersey,USA.

[6] R. L. Horton (1978), *The General Linear Model*, McGraw-Hill International Book Company, Lehigh University, USA.

[7] MathWorks (1992), *Matlab – User's Guide*, The Math Works, Inc., Natick, Massachusetts, USA.

[8] R. L. Thomas (1997), *Modern Econometrics*, Addison-Wesley, Essex, England.