VTT PUBLICATIONS 289

Load research and load estimation in electricity distribution

Anssi Seppälä

VTT Energy

Dissertation for the degree of Doctor of Technology to be presented with due permission for public examination and debate in Auditorium S1 at Helsinki University of Technology (Espoo, Finland) on the 29th of November, 1996, at 12 o'clock noon.

TECHNICAL RESEARCH CENTRE OF FINLAND
ESPOO 1996

ISBN 951–38–4947–3 (soft back ed.)
ISSN 1235–0621 (soft back ed.)

ISBN 951–38–5200–8 (URL: http://www.inf.vtt.fi/pdf/)
ISSN 1455–0849 (URL: http://www.inf.vtt.fi/pdf/)

JULKAISIJA – UTGIVARE – PUBLISHER

Valtion teknillinen tutkimuskeskus (VTT), Vuorimiehentie 5, PL 2000, 02044 VTT
puh. vaihde (09) 4561, faksi (09) 456 4374

Statens tekniska forskningscentral (VTT), Bergsmansvägen 5, PB 2000, 02044 VTT
tel. växel (09) 4561, fax (09) 456 4374

Technical Research Centre of Finland (VTT), Vuorimiehentie 5, P.O.Box 2000, FIN–02044 VTT, Finland
phone internat. + 358 9 4561, fax + 358 9 456 4374

VTT Energia, Energiajärjestelmät, Tekniikantie 4 C, PL 1606, 02044 VTT
puh. vaihde (09) 4561, faksi (09) 456 6538

VTT Energi, Energisystem, Teknikvägen 4 C, PB 1606, 02044 VTT
tel. växel (09) 4561, fax (09) 456 6538

VTT Energy, Energy Systems, Tekniikantie 4 C, P.O.Box 1606, FIN–02044 VTT, Finland
phone internat. + 358 9 4561, fax + 358 9 456 6538

Technical editing Leena Ukskoski

VTT OFFSETPAINO, ESPOO 1996

Seppälä, Anssi. Load research and load estimation in electricity distribution. Espoo 1996, Technical Research Centre of Finland, VTT Publications 289. 118 p. + app. 19 p.

UDC 621.316:681.5.012(083.7)
Keywords electric power generation, electric power distribution, electric loads, load research, load estimation, electricity, distribution systems, customers, measurement, models, variations, analyzing

ABSTRACT

The topics introduced in this thesis are: the Finnish load research project, a simple form customer class load model, analysis of the origins of customer's load distribution, a method for the estimation of the confidence interval of customer loads and Distribution Load Estimation (DLE) which utilises both the load models and measurements from distribution networks.

These developments bring new knowledge and understanding of electricity customer loads, their statistical behaviour and new simple methods of how the loads should be estimated in electric utility applications. The economic benefit is to decrease investment costs by reducing the planning margin when the loads are more reliably estimated in electrc utilities. As the Finnish electricity production, transmission and distribution is moving towards the de-regulated electricity markets, this study also contributes to the development for this new situation.

The Finnish load research project started in 1983. The project was initially coordinated by the Association of Finnish Electric Utilities and 40 utilities joined the project. Now there are over 1000 customer hourly load recordings in a database.

A simple form customer class load model is introduced. The model is designed to be practical for most utility applications and has been used by the Finnish utilities for several years. There is now available models for 46 different customer classes. The only variable of the model is the customer's annual energy consumption. The model gives the customer's average hourly load and standard deviation for a selected month, day and hour.

The statistical distribution of customer loads is studied and a model for customer electric load variation is developed. The model results in a

lognormal distribution as an extreme case. The model is easy to simulate and produces distributions similar to those observed in load research data. Analysis of the load variation model is an introduction to the further analysis of methods for confidence interval estimation.

Using the `simple form load model´, a method for estimating confidence intervals (confidence limits) of customer hourly load is developed. The two methods selected for final analysis are based on normal and lognormal distribution estimated in a simplified manner. The simplified lognormal estimation method is a new method presented in this thesis. The estimation of several cumulated customer class loads is also analysed.

Customer class load estimation which combines the information from load models and distribution network load measurements is developed. This method, called Distribution Load Estimation (DLE), utilises information already available in the utility's databases and is thus easy to apply. The resulting load data is more reliable than the load models alone. One important result of DLE is the estimate of the customer class' share to the distribution system's total load.

PREFACE

This study is one consequence of the load research project of Finnish electric utilities started at the Association of Finnish Electric Utilities (AFEU) in 1983. Forty utilities joined the project and over 1000 customers' hourly loads have been recorded since then. The work for this thesis started while I was working at the AFEU in 1993 and continued at VTT Energy from 1994 as a part of the distribution automation research programme EDISON.

The work has been supervised by professor Jorma Mörsky. I am grateful to him for the co-operation and support during the academic process.

I owe many thanks to Dr Matti Lehtonen in VTT Energy for research management, enthusiasm and support while studying these new matters of electric power systems and distribution automation. Also I want to thank Mr Tapio Hakola and Mr Erkki Antila in ABB Transmit Oy for giving the industrial perspective to this study and associate professor Mati Meldorf from Tallinn Technical University for very important comments. For an inspiring work environment I want to thank all my superiors and colleagues at VTT Energy.

The Finnish load research project has been a huge team work of many people working in different organisations. While the number of people is too large to mention individually I want to send thanks to all those who took part in the project and took responsibility for many important tasks in the electric utilities and in the AFEU.

For the financial support I want to thank VTT Energy, the Association of Finnish Electric Utilities, TEKES Technology development centre, ABB Transmit Oy and Imatran Voima foundation.

Regaarding the English language I want to thank Mr. Harvey Benson for his fast and good service in checking the manuscript. The fine chart figures of the analysis of load data were possible thanks to Adrian Smith's Rain PostScript graphics package.

The warmest thanks I want to address to my family. The writing of this work took much of my time at home. I am grateful for the patience and understanding from my wife Ruut and daughters Anna and Pihla. Their support and engouragement made this work possible.

Helsinki 12.9.1996

Anssi Seppälä

CONTENTS

SYMBOLS

AFEU	Association of Finnish Electric Utilities
APL	A Programming Language
DLE	Distribution Load Estimation
DSM	Demand Side Management
δ_k	condition (0 or 1) if the time of use τ_k of appliance k exceeds T
$(\Delta\tau_k)_i$	change of τ_k in step i of a sequence of random changes
$\Delta(W_T)_i$	change of W_T in step i of a sequence of random changes
$d(t)$	day type at time t
ε, e, v	symbols for random error of time, energy, etc.
$E\{X\}$	expected value of random variable X
$\varphi(X)$	normal distribution density function
$F(X)$	normal distribution function
G	a function representing the weighted sum of errors in DLE
$g(X)$	transformation function of sample data
$h(t)$	hour of day at time t
k_1, k_2	coefficients of Velander's formula
LNE	LogNormal distribution Estimation method for confidence interval
LNEA	LogNormal distribution Estimation method for confidence interval, variation A
LNEB	LogNormal distribution Estimation method for confidence interval, variation B
L_α	model of confidence interval α
$\hat{L}_\alpha$	estimated model parameter of confidence interval α, normal distribution
$L_c(m,d,h)$	ratio of hourly load to annual energy of class c, month = m, day = d, hour = h
$\Lambda(X)$	lognormal distribution density function
$m(t)$	season (month) at time t
m_1, m_2,...	distribution parameter, mean: m_1 = normal distribution, m_2 = lognormal distribution, m_{3a} and m_{3b} lognormal distribution, m_4 = simplified lognormal distribution
NE	Normal distribution Estimation method for confidence intervals
N(0,1)	normal distribution with $\mu = 0$ (mean) and $\sigma = 1$ (standard deviation)
$\Pr\{\wp\}$	probability of event $\wp$
$\overline{P}$	average active power load
P	active power load
$P_{N,k}$	installed (nominal) active power of an electric appliance k
P_α	α percentile of power $\Pr\{P \le P_\alpha\} = \alpha/100$
q_1	error of α[%] in confidence interval estimation

q_2	error of L_α in confidence interval estimation expressed in [%]
$r(W_i)$	Kapteyn's reaction function
σ	parameter of normal distribution and lognormal distribution
$\sigma\{X\}$	standard deviation of random variable X
s_1, s_2,..	distribution parameter, standard deviation: s_1 = normal distribution, s_2 = lognormal distribution, s_{3a} and s_{3b} lognormal distribution, s_4 = simplified lognormal distribution
$s_k(t)$	state of appliance k
SCADA	Supervisory Control And Data Acquisition
SLNE	Simplified LogNormal distribution Estimation method for confidence intervals
T	time interval of the integration of energy W_T consumed in time T to calculate load $P = W_T / T$
t	time (point)
τ_k	time of use of an appliance k (time period)
θ	temperature
U_α	α percentile of unit normal distribution
w	energy
W_a	annual energy
ξ	parameter of lognormal distribution

1 INTRODUCTION

The electric load in electricity distribution varies with time and place (See examples of load variation of three types of customers in Fig.1) and the power production and distribution system must respond to the customers' load demand at any time. Therefore modern electricity distribution utilities need accurate load data for pricing and tariff planning, distribution network planning and operation, power production planning, load management, customer service and billing and finally also for providing information to customers and public authorities.

The load information mostly needed is how a customer or a group of customers uses electric energy at different hours of the day, different days of the week and seasons of the year and what their share of the utility's total load is and how loads of different customers aggregate in different locations of a distribution network.

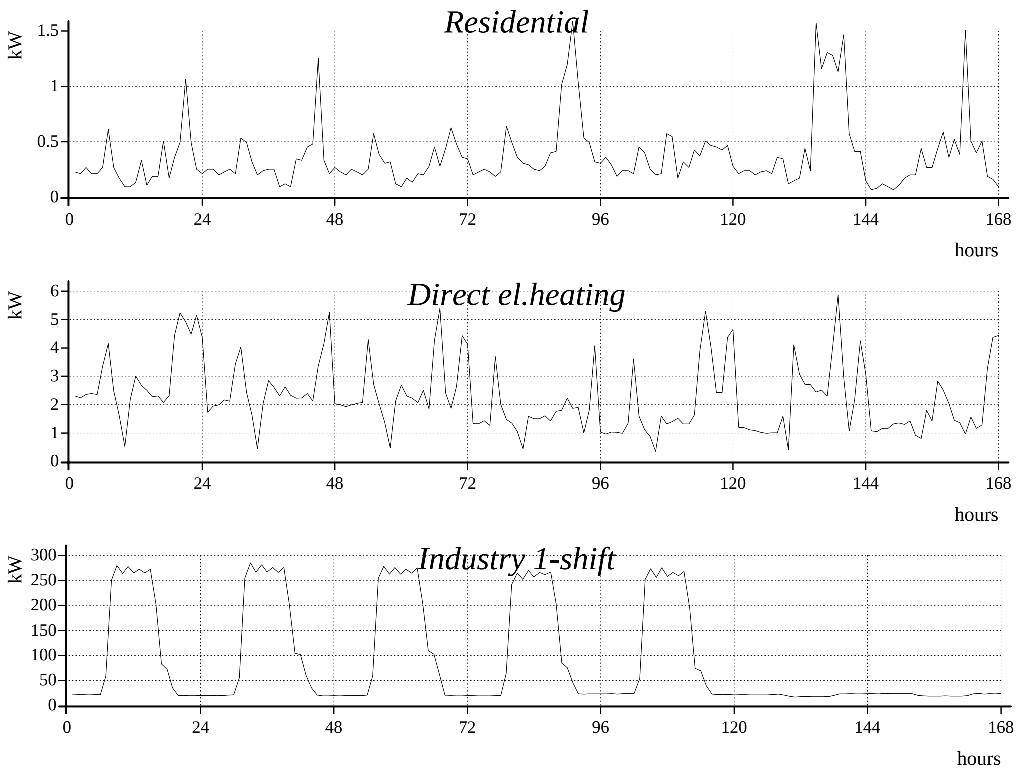


Fig. 1. Examples of customer load variation over one week for three different types of customers.

This study concentrates on two problems: estimation of customer's hourly load using statistics from load research measurements and the distribution load estimation based both on load models and the direct load measurements from the distribution network.

2 LOAD INFORMATION IN ELECTRICITY DISTRIBUTION

2.1 GENERAL

The mission of the electric power utilities is to service the customer's needs of electric energy at optimal costs. The most important thing characterising the service is the load supplied to customers. Other factors are reliability, number and length of outages, the quality of voltage and mechanical and electrotechnical security of installations.

The load data is needed for defining the requirements of the network's transmission capacity, approximating the transmission losses or estimating the existing network's capability to transfer increasing loads. The planning of new generation capacity or energy purchase requires knowledge of customers' load variation (Fig. 2).

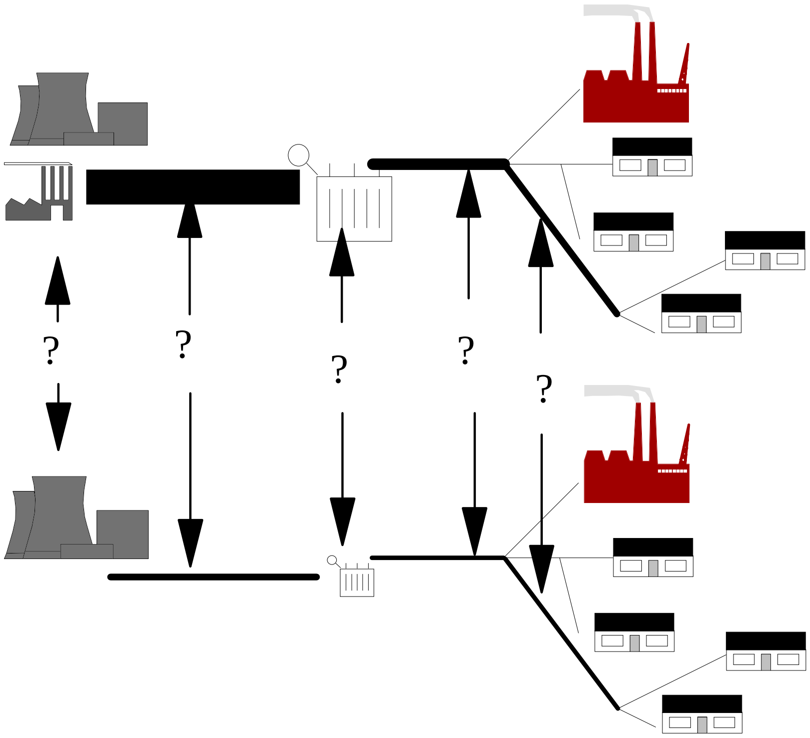


Fig. 2. Load data is needed for planning and dimensioning of electricity production, transmission and distribution.

The physical properties of network components are usually far more accurately known than the load, and the accuracy of load estimates and forecasts is the main factor determining the overall accuracy of several power systems' computations. There is a continuous need to improve the knowledge of loads in electric power systems by collecting and analysing more load information, developing better load models and developing new applications utilising all the new information available (Lakervi & Holmes 1995 pp. 209 - 221).

2.2 THE MEANING OF LOAD

The load data may be formulated in several ways according to the requirements of applications. The most important specifications for load data are

- System location: customer site, low voltage network, transformer, etc.
- Customer class: industry, service, residential, electric heating, etc.
- Time: time of year, day of week, time of day.
- Dimension: A, kW, cos ϕ.
- Time resolution of the load recording: 5 min, 15 min, 30 min, 60 min, etc.

The load influences the distribution network causing energy losses and voltage drop. While the voltage U is approximately constant, the current and the load factor alternate with the load. The relation between load current I, active power load P and load factor cos ϕ is defined in a three phase distribution system by the equation

$$P = \sqrt{3} \cdot U \cdot I \cdot \cos\phi \qquad (1)$$

The load current causes thermal losses in electrical components (conductors, breakers, transformers). The thermal losses are proportional to the resistance of the component and square of the load current. The heat causes ageing and damage to the components. In some components, like power transformers, such phenomena is critical. On the other hand the energy losses increase the transmission costs in the distribution network. Transmission losses may grow to over 10 % of the total transmitted energy.

For example, the thermal loss load of a power transformer is defined by equation (2) where power loss P_θ is the thermal loss load, P_N the thermal loss in nominal current, I_N is nominal current and I is the load current (for example 500 kVA transformer's P_N = 5 kW):

$$P_\theta \approx P_N \left(\frac{I}{I_N}\right)^2 \qquad (2)$$

The loading capability of a transformer is determined by the thermal ageing of the transformer's coil's insulators. The durability of a transformer can be estimated when the load of the transformer is known (Erhiö 1991). Therefore load data is essential in calculations finding the most economical targets for network reinforcements.

With energy business the pricing of electricity is determined by the customer's energy use at different times and the amount of incremental power demand the customer causes to the energy selling company's energy purchase. The planning of the time of day tariffs and seasonal tariffs requires knowledge of the energy shares for different time/price categories. These values depend on the customer's load variation.

The electricity market in Finland calculates energy sales on a one hour basis.

2.3 FACTORS INFLUENCING THE ELECTRIC LOAD

Usually all the needed load data is not available directly and the load values must be estimated and forecasted using other available information. The load calculations for different locations in the radially operated distribution network are rather straightforward when the customers' loads are known.

The load modelling and forecasting is based on knowledge of several factors influencing the customer's load. The most important factors are:

- Customer factors: type of consumption, type of electric heating, size of building, electric appliances, number of employees, etc..
- Time factors: time of day, day of week (+ special days) and time of year.
- Climate factors: temperature, humidity, solar radiation, etc.
- Other electric loads correlated to the target load.
- Previous load values and load curve patterns.

The relation of the factors to the electric loads are handled by various modelling techniques. A wide range of research of modelling electric loads by mathematical methods have been reported. In Finland mathematical modelling studies were done in Helsinki University's System Analysis laboratory by Karanta & Ruusunen (1991) for electric utility's total electric load and Räsänen (1995) for single customers' loads.

Load modelling and modelling applications for Finnish power companies have also been studied by Meldorf (1995) who also presents a complete utility level load modelling application software.

2.3.1 Customer factors

The customer factors of electricity consumption are primarily the number, type and size of the electrical equipment of the customer. While the electrical equipment and installations vary from customer to customer there are recognised types of customers which have similar properties. Such customer types are for example: residential, electric heating, agriculture, small industry and service.

2.3.2 Time factor

The electric load varies with time depending on human and economic activity. There is more load in the day time and less load at night. Also the load varies between week days and usually the load is lower at weekend than on week days. The cyclic time dependency leads to analysing the loads: on hour of day basis, day of week basis and time of year basis.

The time factor is important in the Finnish power system because the production capacity is limited and the price of the incremental power to maximum load is sometimes very high. The customer load’s coincidence with the energy seller’s own purchase is a very important pricing factor.

2.3.3 Climate factors

The weather factors like out-door temperature, wind speed, sun radiation etc. influence the load. The out door temperature mainly influences customers with electric heating. The temperature varies over a wide range in the Finnish winter (about 20 degrees C change in a few days is normal!). This causes a lot of variation in temperature dependent loads, especially electric heating.

Temperature is not the only factor, as the demand for heating energy is also dependent on sun radiation, wind speed and humidity. Also the automatic control of different heating equipment reacts to the temperature changes in different ways. However in practice only the out door temperature is taken into account as knowledge of the values of the other factors is limited.

Although the temperature correlation is obvious for total heating energy use, the interaction between hourly load and out-door temperature is more complicated (Räsänen 1995). This is because of the automatic thermostat control of the heating equipment, which among other things, also interacts with the other uses of electricity. For example, heat storage is designed to store the heating energy at night and transfer it to day time use.

2.3.4 Other electric loads

Electric loads are sometimes influenced by each other. A good example is how the use of other electrical appliances in a building reduces the demand for electric heating. The use of one appliance also generates the need to use other appliances. This interaction is not well known and will be analysed with the analysis of statistical distributions of customer's electricity consumption in chapter 4.

2.3.5 Previous load values

The electric loads have many periodic patterns. The load variation includes autocorrelation. When there is knowledge of previous load values e.g. from the previous day and from the previous hour, the load is usually very easy to predict with good accuracy. This property has been successfully utilised with forecasting of the utility's total load. However the previous load data recordings are seldom available for a customer or a customer class.

2.4 AVAILABLE DATA IN ELECTRIC UTILITIES

Usually the only measurements from customer loads is the energy consumption from the billing meters. From the bigger customers there might also be hourly meter recordings or maximum load values. The customer billing databases usually include some kind of classification and naturally the pricing information: size of the main fuse and annual energy.

The annual energy is the most important factor used in this study. The annual average load is equal to the annual average hourly loads, and therefore a reasonable factor explaining the hourly load differences between customers of the same class.

The new electricity market will promote new metering techniques and the number of hourly load recordings is growing. However small residential customers will not be under direct hourly recordings for many years.

2.5 THE SIMPLE FORM CUSTOMER CLASS LOAD MODEL FOR DISTRIBUTION APPLICATIONS

Most mathematical load models developed for forecasting purposes are so far too complicated to be directly applied to studies of distribution networks (See Fig. 3). The number of calculated network nodes is high and the knowledge from the loads and load measurements is limited. Therefore simple form load models are needed which are easy to adminster and use only such information that is available directly from utilitiy customer billing systems.

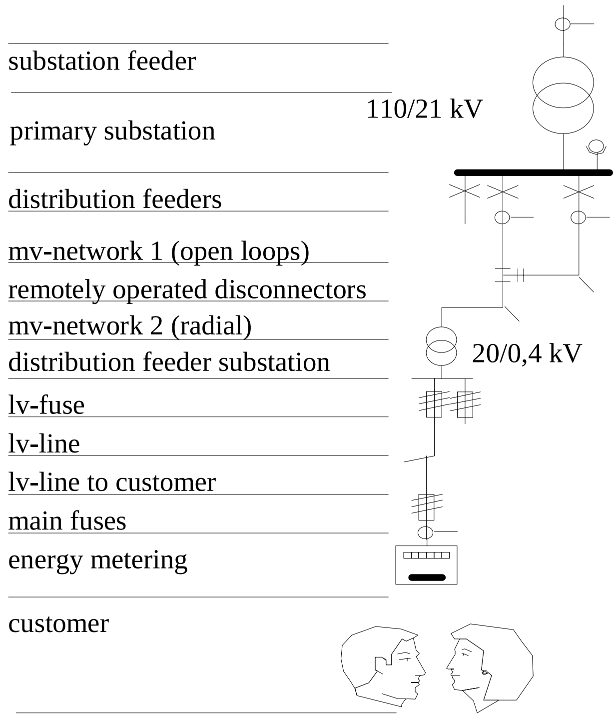


Fig. 3. For planning and monitoring purposes the electric utility needs to estimate the loading of the distribution network. The readings from the customer billing meters are the best and usually the only source of information of the customer's energy use.

In the Nordic countries the traditional method to estimate peak load in distribution network from customer's annual energy W_a has been Velander's formula (3)

$$P_{\max} = k_1 W_a + k_2 \sqrt{W_a} \tag{3}$$

The coefficients k_1 and k_2 studied from the load recording data from the Finnish load research project have been published in the network planning recommendations by the AFEU.

Velander's formula has been quite reliable in medium voltage network (mv-network) load calculations when the number of customers has been large. However the load estimates of small numbers of customers in low-voltage networks (lv-network) have been quite unreliable.

The simple form load model used nowadays in electricity distribution applications of most Finnish electric utilities represents the customer's average hourly load $\overline{P}(t)$ and standard deviation $s_P(t)$ as a linear function of the annual energy consumption W_a in eq. (4) .

$$\begin{cases} \overline{P}(t) = L_c(m(t), d(t), h(t)) \cdot W_a \\ s_P(t) = s_{Lc}(m(t), d(t), h(t)) \cdot W_a \end{cases} \tag{4}$$

$m(t)$, $d(t)$, $h(t)$ are classifying functions resulting in a category where a specific hour t belongs. Their definition may vary among applications, but in general:

- The value of $m(t)$ is season, time of year, usually month, but may be a week or a two week period.
- The value of $d(t)$ is day type, usually day of week or working day/holiday.
- The value of $h(t)$ is hour 1…24.

The parameters L_c and s_{Lc} are estimated from load research data (see chapter 3) from the average and standard deviation of the hourly load recordings divided by the customer's annual energy consumption

$$\begin{cases} L_c(m,d,h) = E\left\{\dfrac{W_{h,c}(m,d,h)}{W_{a,c}}\right\} \\ s_{Lc}(m,d,h) = \sigma\left\{\dfrac{W_{h,c}(m,d,h)}{W_{a,c}}\right\} \end{cases} \tag{5}$$

where $W_{h,c}(m,d,h)$ is class c customer's hourly energy in month (season) m, day (day type) d and hour h. $W_{a,c}$ is the class c customer's annual energy.

Examples from the data and how the division by the annual energy affects the variation of data is shown in the figures in Appendix 4.2.

The several factors affecting the loads (chapter 2.3) are not taken into the model. Their impact is now cumulated in the mean and the standard deviation of the model. The practical motivation for this simple model is that there is usually no data or previous load measurements available for calculations where this model is applied. This simple model is a straightforward statistic of consumption of electricity of a specific customer class in a specific time range compared to the customer's annual energy use.

2.6 ELECTRICITY DISTRIBUTION APPLICATIONS UTILISING LOAD MODELS

The Finnish electric utilities now use various applications in network planning, tariff planning and production planning which use the load models from the national load research project (Fig. 4).

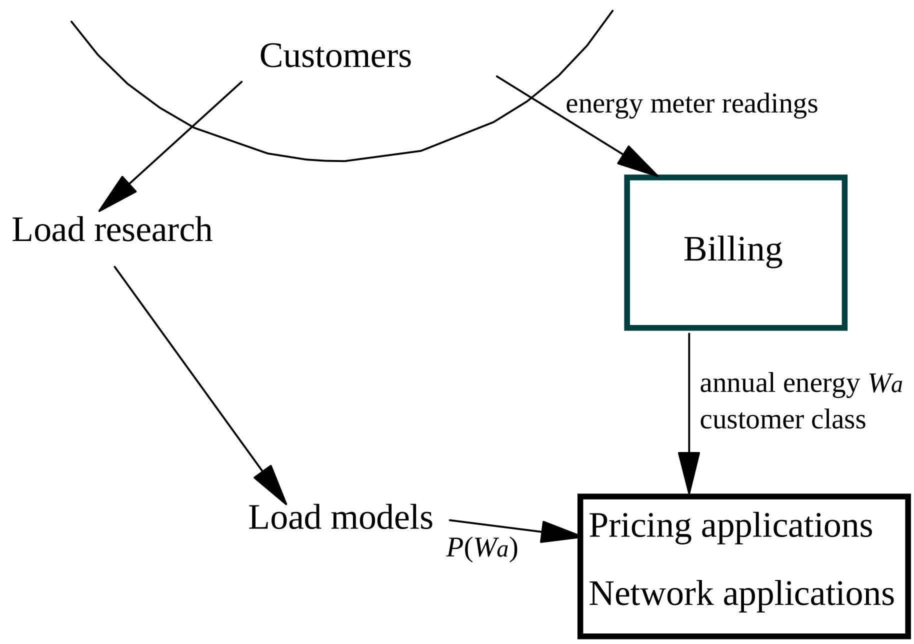


Fig. 4. Load research produces simple load models to be used in applications where the only available data is the customer's annual energy use and customer class. Using the load models the applications can estimate the load for one year on hourly basis.

Distribution load flow software based on load curves was introduced by Rossinen (1982). Since the first load models were published in (STYV 1985) more applications for network load computation, network planning (Juuti et al. 1987), (Kohtala & Koivuranta 1991), (Partanen 1991) and electricity pricing based on load curve data (Ojala 1992) were introduced.

The Finnish software companies, for example Tekla Oy, Tietosavo Oy and Versoft Oy, have produced commercial network information systems and distribution network load flow calculation software products which utilise the load models from the Finnish load research project.

The model parameters are usually presented in watts [W] when the annual energy W_a = 10 MWh. The parameter values are also sometimes prepared for every hour of the year and organised as two 365 x 24 matrixes, one for average L_c and one for standard deviation s_{Lc}.

Model (4) written with dimensions is then:

$$\begin{cases} \overline{P}(t)[\mathrm{W}] = L_c(m(t), d(t), h(t)) \left[\dfrac{\mathrm{W}}{10\ \mathrm{MWh}} \right] \cdot W_a [10\ \mathrm{MWh}] \\ s_P(t)[\mathrm{W}] = s_{Lc}(m(t), d(t), h(t)) \left[\dfrac{\mathrm{W}}{10\ \mathrm{MWh}} \right] \cdot W_a [10\ \mathrm{MWh}] \end{cases} \qquad (6)$$

The number of different seasons $m(t)$ and types of day $d(t)$ may vary according to the accuracy required. See Table 1.

Table 1. Different configurations of load models and their applications.

Configuration of a model	Application
24 hourly values for 7 days a week for 12 months a year: m = 1…12, d = 1…28(29)/30/31, h = 1…24	The most complete form and is used mostly with pricing applications where the complete year's load data in needed. New applications for load forecasting and network load monitoring require this model. Specific for one year's calendar.
24 hourly values for 3 days(working day, Saturday and Sunday) for 12 months a year: m = 1…12, d = 1…3 , h = 1…24	Suitable for simple pricing applications. No specific calendar.
24 hourly values for 3 days (working day, Saturday and Sunday) for 26 two-week periods of the year: m = 1…26, d = 1…3 , h = 1…24	Traditionally used in long term production planning applications and also network planning and load flow applications. No specific calendar.

The experience of using the load models has been positive. The distribution network load flow applications give much better load estimates than the conventional methods and, for example, utilities have therefore been able to reduce their investment plans. The wide use of load models and positive feedback has encouraged the continuation of the study.

2.7 STATISTICAL ANALYSIS OF LOAD MODEL PARAMETERS

2.7.1 Sampling and classification

The parameters L_c and s_{Lc} of the load model are statistically estimated from the load research data. Because of the large number of customers, sampling is the only possible way to collect data and estimate the parameters. The problem is how the selection and analysis of the sample of customers should be made to finally get the most accurate load estimates for practical network calculations. See Fig. 5.

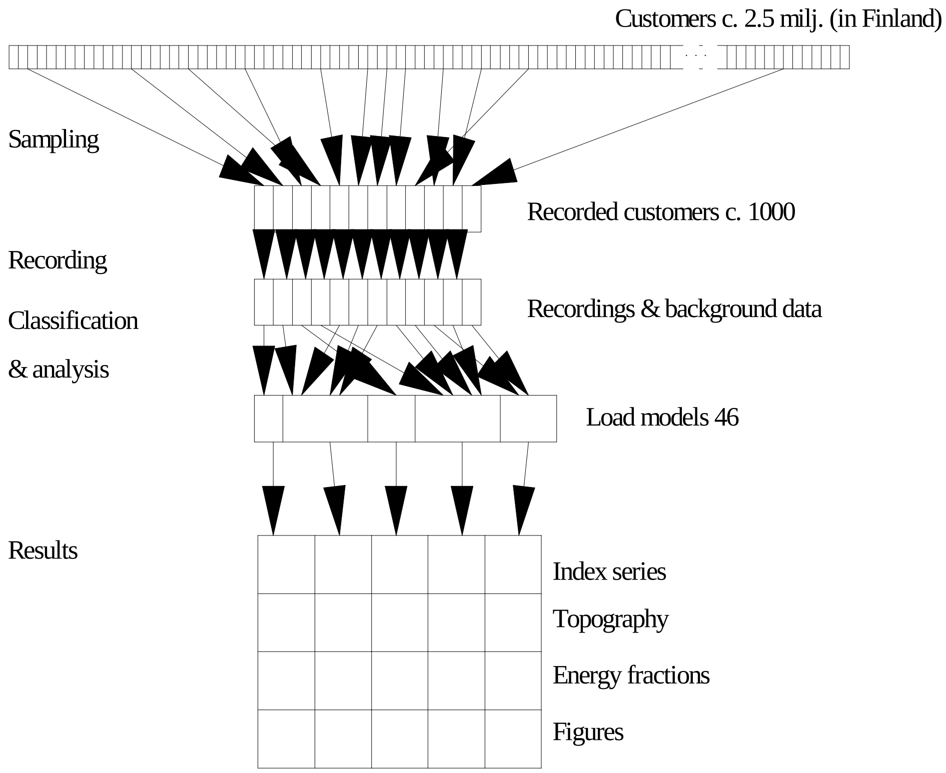


Fig. 5. Load research as a process from sampling to final results for applications. Different types of results are required: relative load index series, hour/season load topography, energy fractions, figures, etc.

The way to minimise the sample size and research costs is to make stratified sampling where the population is divided into some strata where the variance is known to be small compared with the variance between strata (Pahkinen & Lehtonen 1989). Instead of terms *strata* and *stratification* the terms *class* and *classification* are used with the load research.

The utilities' applications require a set of load models to represent all the customer classes. Deciding the optimal number of classes and the type of load model for one class is a complicated problem. The practical criteria for load data classification are according to experience:

1. The load variance in one class of customers should be as small as possible.
2. The number of classes should not be too large.
3. The classes should be representative.
4. The classes should be easily linked with the utility's databases.

Load research classification has also helped the utilities to classify their own customers. Because of the many requirements load classification is said to be more an art than a computation and is best done by an experienced analyst.

For classification of the load research data, automatic classification methods, i.e. cluster analysis, were also considered but not completely applied (Seppälä 1984). During the latest analysis the data was first manually split to 77 customer classes (Paananen 1991). After verifying the results the classification finally resulted in the 46 classes presented in Appendix 2 (Seppälä & Paananen 1992). Räsänen (1995) developed methods for load analysis of load classification based on the correlation between load curves, but application of the method did not change the manual classification.

2.7.2 Generalisation and bias

The application of the model (estimated from a sample) to the whole population is called generalisation. Usually the generalisation is done with the sample ratio, which is the relation of the number of items in the sample to the number of items in the whole population. For example, assuming a population of 1000 we study a sample of 100 and find 5 items. Generalising with the sample ratio of 1000/100 = 10 we expect the total number of items in a population of 1000 is 50.

With the `simple form load models´ the generalisation is not done with the number of customers in a sample. The generalisation is done with the annual energy consumption, where the load of a class is estimated by multiplying average customer's load per annual energy consumption with the whole class' total annual energy consumption. While there are, no doubt, many benefits, such an estimation is biased when the customer's load variation is different between customers with different annual energy consumption. This problem has been studied by Särndal & Wright (1984) and they call simple load models (6) "cosmetic" estimators.

The bias of the simple form load model is an acceptable drawback of a practical and relatively cheap method. The load models are known to correspond quite reliably to the total load of the utilities. However, one method to remove bias from the load models is for each utility to make its own load models based on sampling from the utility's own customer population.

Another method to improve the load estimators is to utilise the direct measurements from the network. This method called Distribution Load Estimation (DLE) is introduced in chapter 7.

3 LOAD RESEARCH

3.1 GENERAL

The method of load research, in general, is to collect and analyse load data from different locations of the distribution system (usually at the customer's energy meters) to support the needs of load data presented in chapter 2.2. Load research usually requires special metering instruments and human work when the meterings are done at the customer's site. Thus load research is regarded to be expensive.

The benefits of load research come from improved accuracy of the decisions made in utilities using more reliable load information. Two examples from electricity production planning and demand side management (DSM) are analysed in (Gellings & Swift 1988). They give examples where a given reduction of uncertainty in load data could reduce the total costs of 1000 MW production or a DSM investment by about $40 million.

3.2 HISTORY

In the early days the load data collecting technique was simply to read energy meters regularly and analyse the information. Devices which automatically printed or plotted the kWh value on paper were also used. These data collecting methods were expensive, limiting comprehensive studies. On the other hand the ability to handle and collect large amounts of load research data was also limited. Anyway, the need for load research was recognised in the industry and many methods to improve the work were developed (Wolf 1959 pp. 212 - 252).

The first load data analysing methods were mostly numerical simplifications of the representation of load data. Wolf (1959 pp. 61 - 137) reviews methods of analysis of symbolic load duration curves. Most of those methods are trivial for modern calculators or computers and no longer relevant research topics.

In the 1970's magnetic tape recorders and in the 1980's low cost electronic recorders became available to collect load data, making it possible to conduct wide range load research covering hundreds of customers. Also the development of computers made it possible to store and manipulate large amounts of data to make comprehensive data-analyses[1]. UNIPEDE (International Union of Producers and Distributors of Electrical Energy) published a book in 1973 (UNIPEDE 1973) where the methods of regression analysis of load data were reviewed.

Computer based statistical load analysis was first done by the regression method using measurement data from substations together with total energy consumption data from customers. This method is described for example in (UNIPEDE 1973 pp.89-101). Also in Finland at least two such studies are reported (Puromäki 1959) and (Leino 1974).

Fikri studied the statistical properties of loads and their applications for network planning in (Fikri 1975). The study was based on some recorded data and development of calculations assuming that loads were normally distributed.

Load research projects have been reported in the 70's and 80's from many countries. Some projects are listed in the UNIPEDE congress report (Kofod et al. 1988). Load research projects are referenced from Germany, Denmark, Spain, France, Norway, Sweden and the United Kingdom. In the United States load research has had a special position because of the Public Utility Regulatory Policies Act of 1978 that has set high expectations for the quality of the statistical data and analysis behind a utility's proposals for rate increases and system expansion.

Nowadays load research is a normal activity in electric utilities. The collection and handling of data is no longer a problem. The focus is on analysis and utilisation of the load research data.

3.3 RECENT LOAD RESEARCH PROJECTS IN SOME OTHER COUNTRIES

3.3.1 The United Kingdom

In the UK the responsible organisation for load research co-operation is the Electricity Association (EA). The EA has studied loads in England for a long time and so far they have produced analyses for 250 customer groups (Allera 1994). They are also actively reporting their results (EA 1994). In the EA, load research has been a continuous activity for many years.

[1] For example, hourly load recording over one year produces 8760 measurements. In four byte memory and approximating some overhead we get 40 kbytes per one year of recordings. Thus in one megabyte, 25 one year’s recordings can be stored. Modern PC computers can easily manage over 1000 megabytes data storage.

3.3.2 Sweden

The Swedish Association of Electric Utilities SEF (Svenska Elverksföreningen) organised a load research project to get load data for network calculations. About 400 customers were recorded and analysed in 45 categories. The recordings were done in 15 minute intervals. The results were analysed and published in (SEF 1991). This analysis differs from others by its way of adjusting temperature dependent load data with degree-day figures (graddagtal) to standardise the circumstances of load data from different locations and temperatures. A software package "Betty" has been developed to give load values and estimates for single and aggregated loads utilising the results of load research projects.

3.3.3 Norway

In Norway, load recordings are organised by the Electricity Research Centre EFI (Electrisitets Forsknings Institutt) in Norway. The report from Feilberg & Livik (1993) describes how the load research results are integrated into a software package ”PMAX”. The results are based on 15 minute load recordings from 100 recorders. The results are reported for eight customer categories.

3.4 THE FINNISH LOAD RESEARCH PROJECT

3.4.1 General

The Finnish electric utilities started to co-operate with load research in 1983. Most of the recordings were done using a specific electronic load data recorder produced by a Finnish company Mittrix Oy (Fig. 6). Most of the recordings were done on the customer level. The author was working with the project at the beginning and the first steps of this project are described in the author’s M.Sc. thesis (Seppälä 1984).

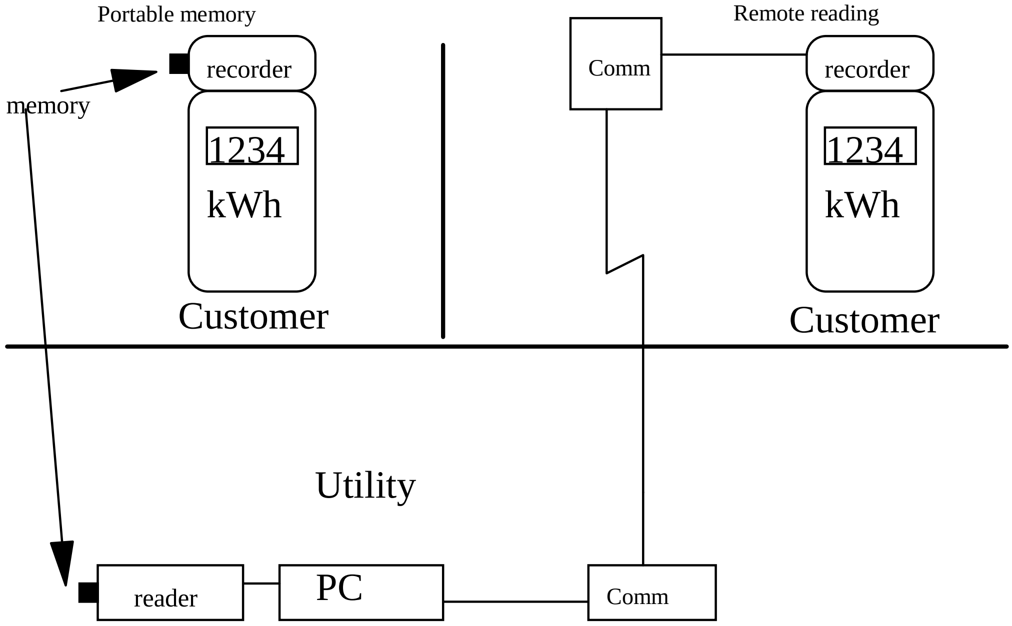


Fig. 6. The load research recording system. The portable memory (left) has been the most used type of recorder. Remote meter reading with telephone communication (right) is now more popular due to the needs of the electricity market.

About 1000 consumer load recordings have been collected. The latest results of the analysis were published in 1992 (Seppälä & Paananen 1992). The results including load models for 46 customer classes were published in several data formats. For the complete list of publications of the load research project see Appendix 1.

The load research project was originally conducted by the Association of Finnish Electric Utilities (AFEU) from 1983 to 1994. Since 1994 the research has been VTT Energy's responsibility (Fig. 7). The project has regularly employed one half time employee and, in addition, temporarily two to three other persons.

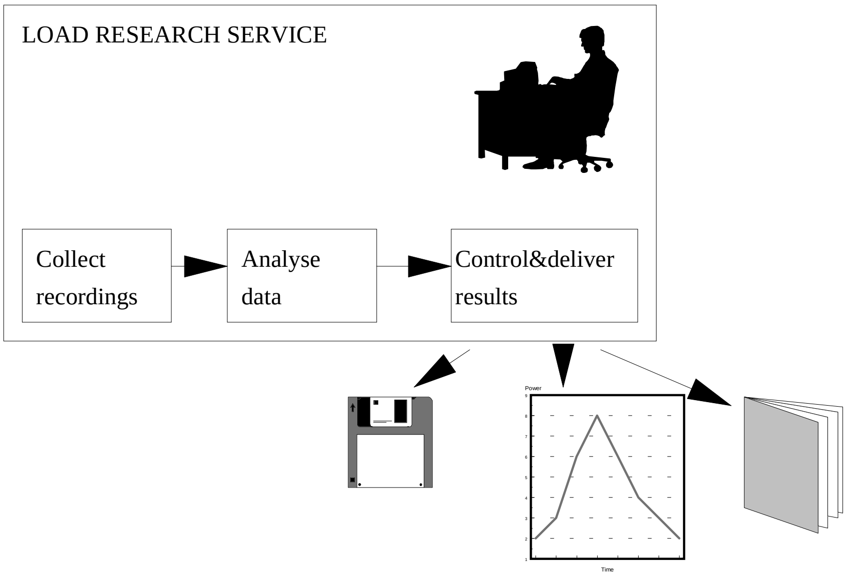


Fig. 7. Load research is a service to collect and analyse load research data and then deliver the results to be used in the utilities' applications.

3.4.2 Load research data management

Since the beginning of the load research project the greatest challenge has been to keep the load data in order and available to the analysis software. Most of the data analysis and data manipulating software was written during the project and by the people working with the project (See 8). During the years from 1983 the platform of load research data storage and manipulation moved from mainframe computer to a desktop computer.

The data management of load research now utilises modern computer technology. The load data is stored in a Relational Data Base. The applications are connected to the database through ODBC (Open Database Connectivity) using SQL (Structured Query Language). The applications include load data management, calendar, reporting, import of data from load recorders and statistical analysis.

Most of the analysis programming is done with APL (A Programming Language). APL is an array oriented programming language with a special mathematical notation. APL was found to be a very suitable tool for calculations, data manipulation, graphical presentations and creating user interfaces for load research.

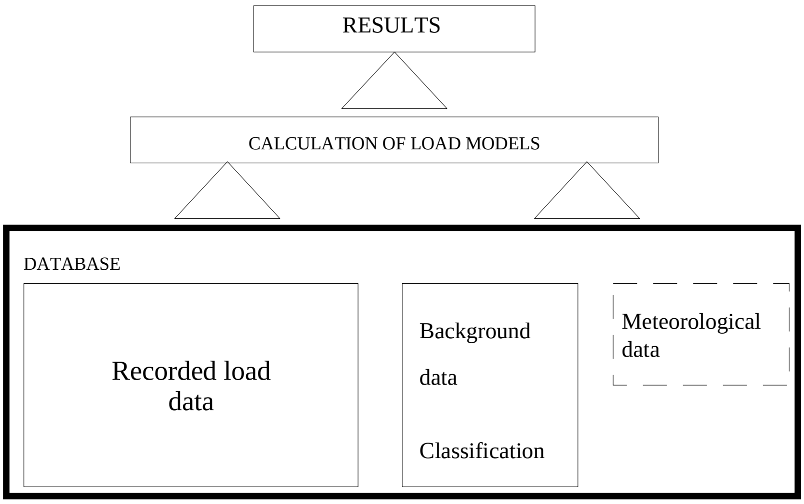


Fig.8. Load research utilises modern data management systems. The collected data is stored in a relational database. The calculation software and graphics is mostly done using the programming language APL.

3.4.3 Years of the Finnish load research project 1983 - 1996

Start 1983
Forty utilities joined the project and ordered a total of 556 load data recorders for this research. To read the EPROM memories of the load recorders, a special data translation and collection computer station was maintained to feed the load data to the mainframe computer of AFEU.

The selection of customers for the research was the utilities' responsibility. The initial classification included five classes of customers and five types of residential electric heating. The classes were residential, buildings (non-residential), agriculture, industry and service. The types of electric heating were direct electric heating, partly storage electric heating, full storage electric heating, dual heating (electricity and oil/wood), heat pumps and no electric heating. See Table 2.

Table 2. The initial classification of customers in the beginning of the load research project.

Class	Recorders
Direct electric heating	79
Partly storage electric heating	77
Full storage electric heating	30
Dual heating	85
Heat pump	48
Residential without electric heating	81
Agriculture	50
Industry	61
Service	62
Total	573

1983 - 1985

The results of the recordings were first analysed by the author and published in 1985 in the form of so called index series (STYV 1985). The analysis was done for 18 customer classes following the tradition of the national production planning applications. See Appendix 2.2.

1986 - 1989

The recorders were transferred to new customers during 1986 - 1988. The focus was then on industry and service class customers. The study of load modelling for distribution network planning was done and published by Härkönen (1987) and in the network planning recommendations of the AFEU. Also the overall average load curves from various categories were published in 1988. The study of temperature dependence of electric loads was published by Siirto (1989).

1989 - 1991

A statistical load model analysis software package LoadLab by System Analysis Laboratory in the Helsinki University of Technology (Räsänen 1995) was developed. The development work was jointly financed by the AFEU and Imatran Voima Oy.

1991 - 1992

The data management of load research data was transferred to a relational database, and load data manipulation software was developed for the PC. The complete analysis was done with 667 different customer recordings in 46 customer classes. The basics of the computation and the use of LoadLab is described in (Paananen 1991). The flow of the estimation process of the load models is presented in (Fig. 9).

The publication (Seppälä & Paananen 1992) consisted of descriptions for 46 load classes. The data was also made available on data disk in different formats for uploading to applications software. The files consisted of the parameters L_c and s_{Lc} for the simple form load model (4).

As an example, graphs of one of the analysed load classes is shown in Fig. 10 - Fig. 12.

The final classifications of the analysis are presented in Appendix 2.

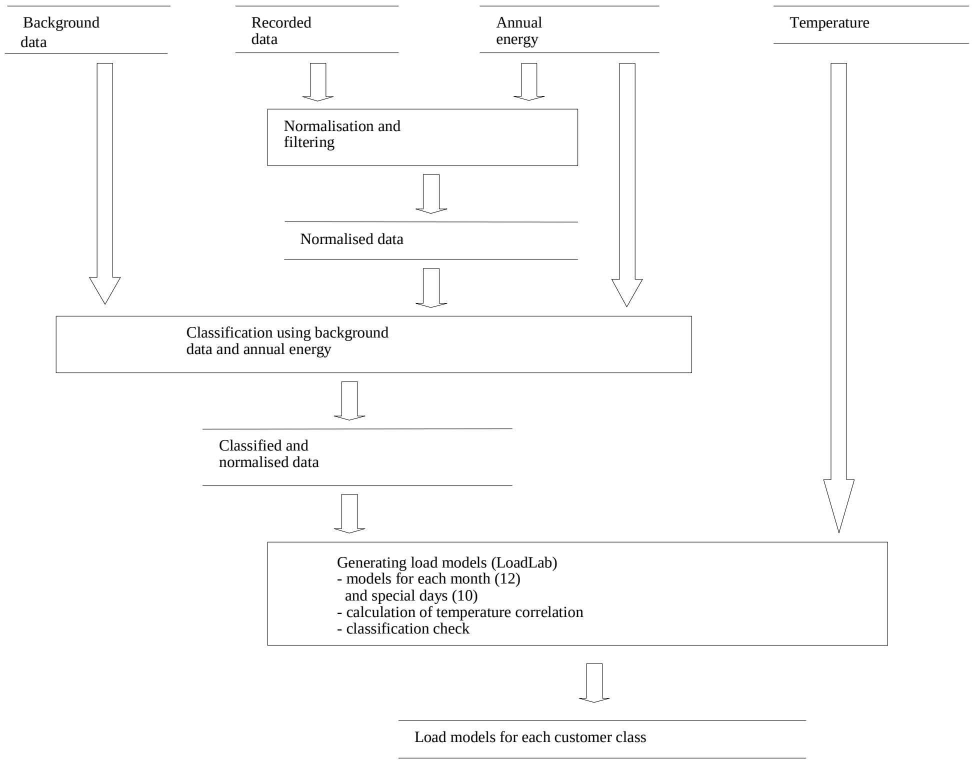
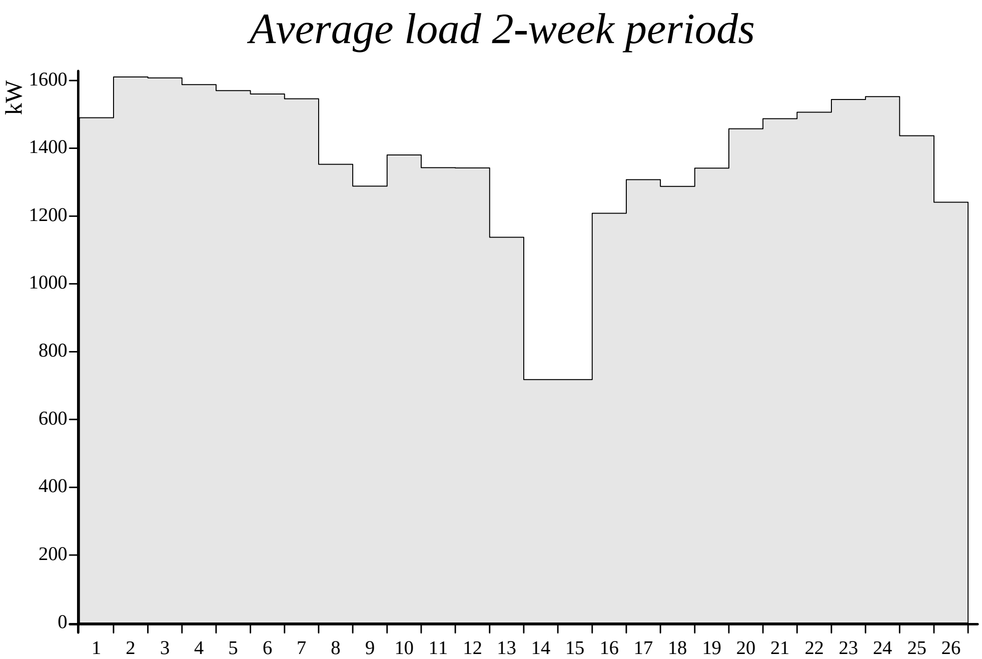


Fig. 9. Estimation of the load model parameters (Seppälä & Paananen 1992).



2-week indexes

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
109	118	117	116	115	114	113	99	94	101	98	98	83	52	52	88	95	94	98	106	109	110	113	113	105	91

Fig. 10. Example of load representation for one calendar year in two week periods. Industry 1-shift, annual energy 10,000 MWh. Below: the data in index form where the average is set at 100.

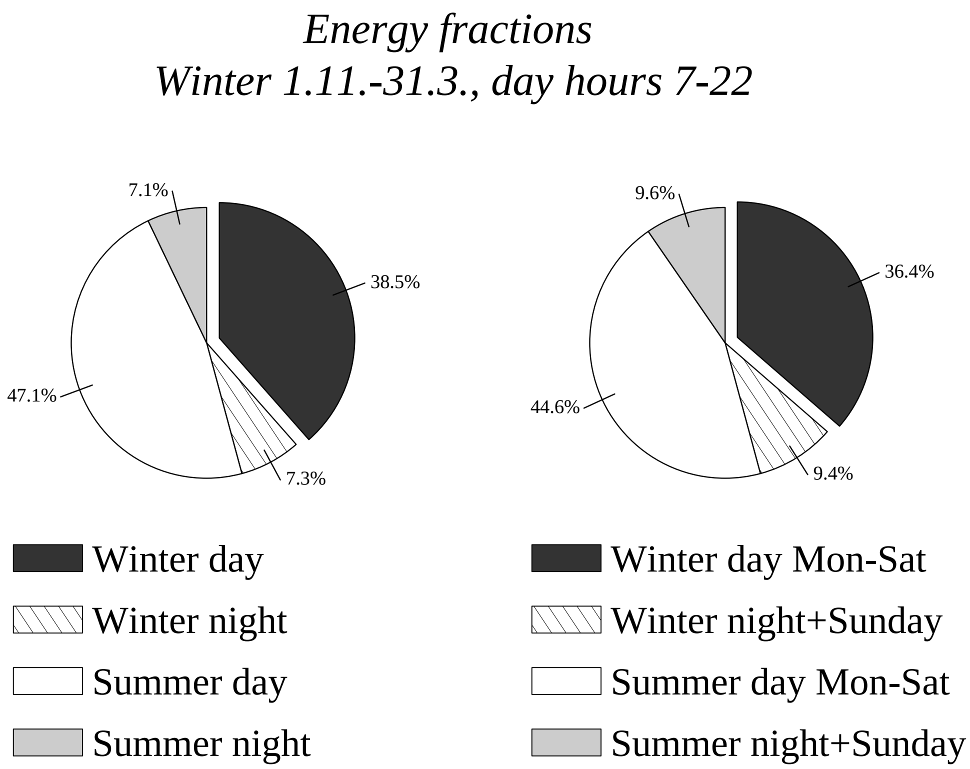


Fig. 11. Example of load representation for one calendar year. Energy fractions. Winter time 1.11. - 31.3. and day time 7 - 22. Industry 1-shift, annual energy 10,000 MWh.

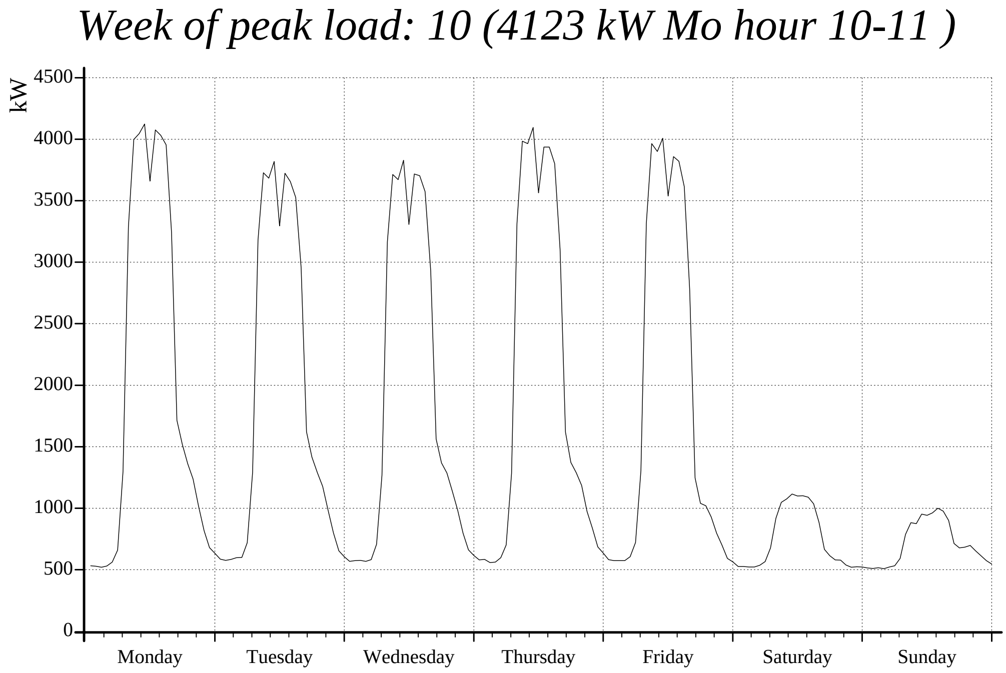


Fig. 12. Example of load representation: Hourly load curve for the week where the maximum load value exists. Week nr. 10 Monday 10 - 11 hour peak 4123 kW. Industry 1-shift, annual energy 10,000 MWh.

1993 -

The continuing work of load research focused on the verification of the previous results and planning the study for the future. It was clearly seen that the number of different customer classes was sufficient for most applications. The greater problem was to determine how reliable these results actually are. The feedback from utilities was in general positive, but some minor errors were also reported. Also the possibilities of using remote measurements and other distribution automation data had to be analysed. The load recordings continued on a small scale, studying some special groups according to the utilities' interests. The preparation of this thesis started. The goal was to develop the load research to better meet the utilitys' needs and make some theoretical basic research.

3.5 THE EXPERIENCE OF THE FINNISH LOAD RESEARCH PROJECT

3.5.1 General

Load models are now used in many applications in electric utilities. The planning staff require simple and easy-to-use methods and they have no resources to handle all the statistical and probabilistic problems involved. This means more responsibility on the researcher to formulate the results so that they are easy to use and also easy to understand. This chapter summarises some of the experience from the years of the Finnish load research program indicating what kind of problems have been encountered and if any solution was found. In general the experience has been positive.

3.5.2 Temperature standardisation

Experience has shown that, in the applications where the simple form load models are used, only electric heating has such a degree of temperature dependency that it needs to be taken into account. Temperature standardisation was made for electric heating in studies (STYV 1985) and (Seppälä & Paananen 1992). The load models were standardised to a long term average monthly temperature.

The simple method of temperature standardisation is that a 1 °C change in outdoor temperature makes, on average, a 4 % change in electric heating load. This well known rule of thumb was also confirmed when analysing load research data by Siirto (1989). Applying this rule we can transform the electric heating load P_1 from out-door temperature θ_1 to desired temperature θ_2 by the equation

$$P_2 = P_1 \cdot (1 + 0.04 \cdot (\theta_2 - \theta_1)) \qquad (7)$$

3.5.3 Unspecified load distribution caused by load control

The loads that are influenced by load management control are not regularly distributed. This is well seen from the load data from electric storage heating. Electric storage heaters are coupled from a few 0.5 ... 3 kW resistors controlled by a clock and thermostat. The resistors themselves have fixed installed power, but the way the load recorders collect hourly energy consumption lead to load values which are randomly distributed from zero to maximum demand with high variance.

3.5.4 Linking the load models with the utility's customer data

The linking of load models to customer and network data is a critical phase before most of the calculations can be run. This work is usually done with the help of the utility's customer billing system. For each customer a load model is selected with special linking rules. In these rules the available information of the customer's annual energy, tariff and the utility's own categorisation is utilised. The rules are specific for every utility.

After each customer has its load model linked, the application to network information system is straightforward. The identifier of the customer's point of delivery joins the network node to customer billing data and load model.

The correspondence of the utility's customer classification to the categories of load research depends on how well the utility people understand the background of each sample of load models. The publication (Seppälä & Paananen 1992) explaining the background of each customer category is the handbook for applying the load models in utility applications. The results of load models linked to some network feeders are presented in the examples in the end of this section (Chapter 3.5.6).

By verifying the total of the load models with the utility's total load, the accuracy of linking of the models with customers can be checked. In the case of a single feeder the errors caused by wrong network topology or bad metering data may lead to poor results, but in general the results have been good.

The utilities' customer and network computation applications include tools for designing the linking rules between customer data and load models. To what extent these rules are similar between utilities is not known, but some utilities have been co-operating in Finland to develop these rules together. The overall analysis of these rules and verification with load models should be further studied.

3.5.5 Problems with seasonal variation in some classes

Some loads have no regular seasonal variation because of the irregular Finnish spring and autumn climate. In practice, in agriculture and in summer cottages, the beginning and ending of the season may shift one month depending on the weather conditions. Calculating average load from different years where the seasons vary, results in a flat load profile which does not correspond to any real year's load. To find a solution to this problem requires further studies and load recordings.

3.5.6 Examples of load models compared with network measurement data

The following four figures (Fig. 13 - Fig. 16) and corresponding tables (Table 3 - Table 6) are examples of how the simple form load models from the load research project correspond to some feeder measurements from substations. The feeder current measurements are transformed from amperes to active power using $\cos\phi = 0.9$ and $U = 21$ kV. The measurement data was collected from substations of two Finnish electric utilities Lounais-Suomen Sähkö Oy and Hämeen Sähkö Oy.

From Lounais-Suomen Sähkö Oy three feeder current measurements from two substations, Meriniitty and Perniö, are represented here. The customer classification for each feeder has been collected from the utility's network information system. The data is over one year's period starting from summer 1993 and ending autumn 1994. Examples of how the models and measurements fit are shown in the following four examples.

The differences between the measurements and load models, according to these examples, can be quite big. The reasons for the differences between measurement and model load level can be numerous, for example errors in scaling of the measurements, incomplete customer data, etc. However the shape of daily and weekly load variation seem quite similar as seen from the hourly plotted weeks.

The reader must notice the data presented here are randomly selected examples from systems that are under continuous development, and these examples also show one method of checking the accuarcy of the information.

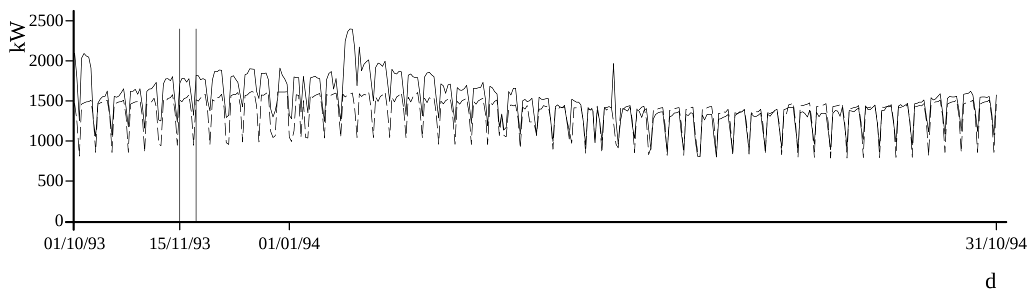
3.5.7 Experience of the Finnish load research project compared to other countries

In general the load research activity is similar in every country, but there are also some differences which should be noted here. The organisation of the Finnish load research project has been very small compared to similar organisations in bigger countries. Therefore there have been limited resources to make load analysis. However the several successful applications have shown that the project has succeeded to serve the utilities' needs.

The close co-operation between Finnish utilities, application software vendors, universities and research institutes has resulted in advanced load research, load modelling and load model utilisation in electricity distribution. The applications of load research data in Finnish distribution utilities might be regarded as one of the most advanced in the world.

Table 3. Example 1. Data from the Meriniitty Keskusta feeder. For description of classes refer to Appendix 2.1.

Class	Nr of customers	W_a [kWh/a]
110	12	249060
120	29	425510
220	10	251800
601	38	199960
611	288	500120
1010	13	146880
1020	3	144730
1030	3	55910
810430	5	72850
820430	3	242680
820480	1	680910
910820	18	327240
920622	229	8309180
920623	5	222090
Total	657	11828920



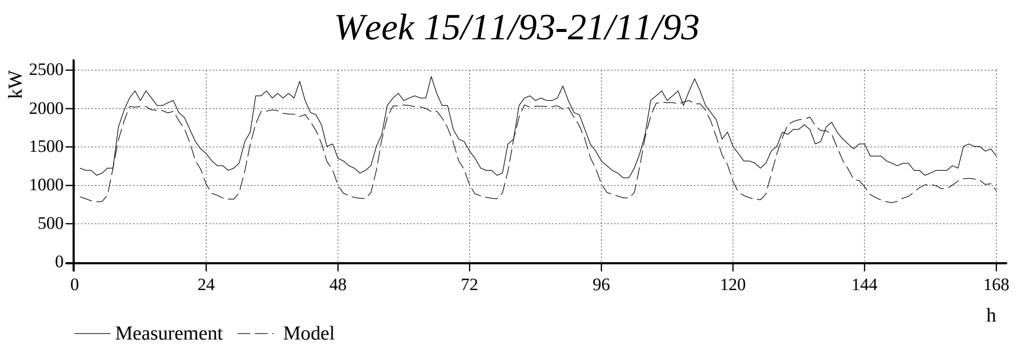


Fig. 13. Example 1. Measurement and load models for the Meriniitty Keskusta feeder. Average daily load (above) and hourly curves over one week (below).

The overall daily variation of the models is quite similar with the measurements. The influence of the cold winter of 1994 is seen in the measurements increasing the difference in the models.

Table 4. Example 2. Data from the Meriniitty Myllyojantie feeder.

Class	Nr of customers	W_a [kWh/a]
810430	3	2057920
810480	3	12641310
820430	1	124590
820480	1	145400
920622	1	146310
920623	2	97000
Total	11	15212530

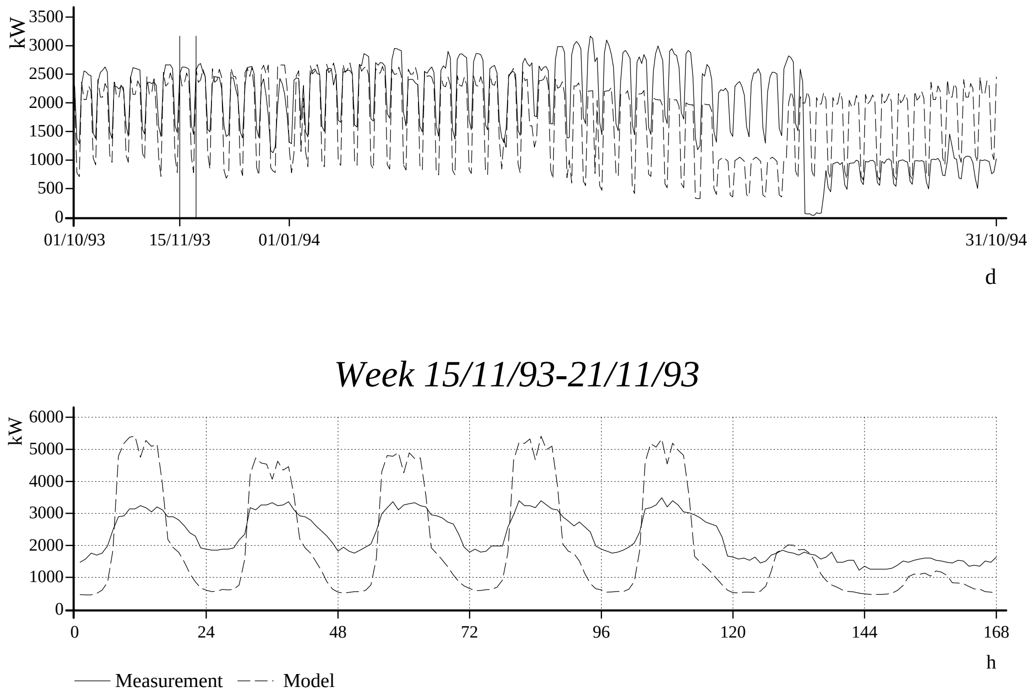


Fig. 14. Example 2. Measurement and load models for the Meriniitty Myllyojantie feeder. Average daily load (above) and hourly curves over one week (below).

The daily energies match well, but the measured daily load curve is very different because of class 810480 with large annual energy use. The industry class 810480 is obviously not 1-shift as it is classified by the utility. Also the summer holidays do not affect the load as seen in the model. At the end of the measurement some switching operations have occurred causing a big error. This shows how the topology of the network is essential for the reliability of the calculations.

Table 5. Example 3. Data from the Perniö Kirkonkylä feeder.

Class	Nr of customers	W_a [kWh/a]	Class	Nr of customers	W_a [kWh/a]
120	52	926350	1010	20	274950
120	3	178790	1030	3	58120
220	7	159940	810430	6	42400
300	3	146820	820430	4	55670
520	2	35060	820480	6	63020
601	168	946090	910820	18	742420
611	418	701670	910830	14	572950
612	5	34840	920623	1	10510
612	1	8050	Total	735	5081750
712	1	22960			
713	1	6440			
733	2	94700			

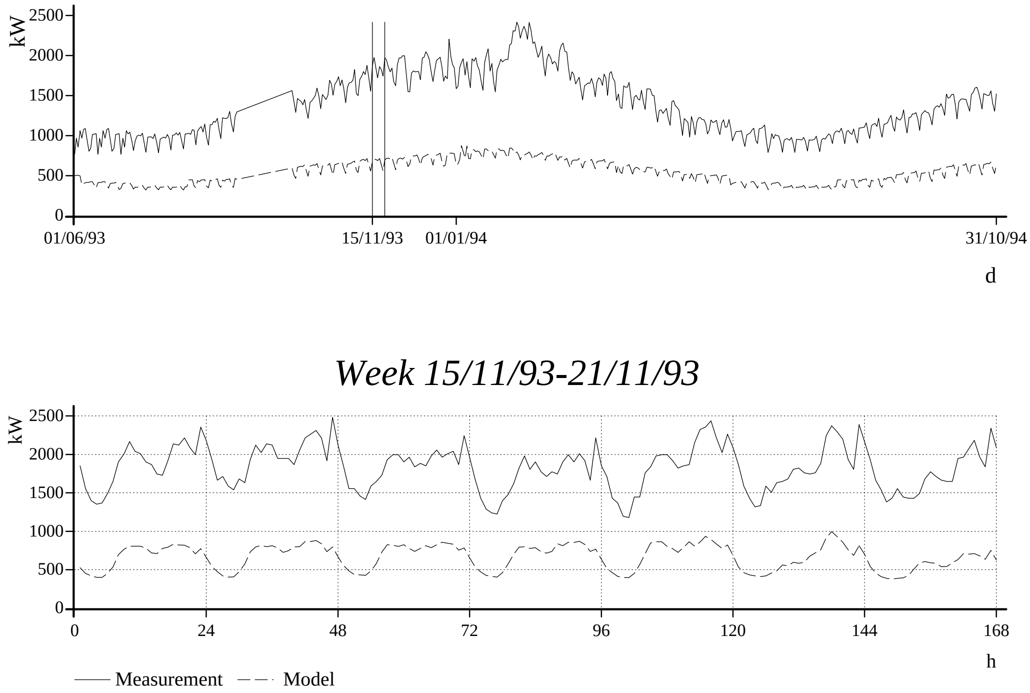


Fig. 15. The Perniö Kirkonkylä feeder. Average daily load (above) and hourly curves over one week (below).

The shapes of the daily load curves are very similar, but the energy level is about 40 % from the measured. This is obviously an error in data collecting and should be fixed. These kinds of errors are quite common and one must always check how accurate the information is from different sources.

Table 6. Example 4. Data from the feeder of the Kulju primary substation. Note different classification. Refer to Appendix 2.2.

Class	Nr of customers	W_a [kWh/a]
1	9	120417
3	52	2905894
4	74	1717852
5	87	2842655
6	18	1510319
7	1577	7633621
8	56	1060872
9	77	157594
11	1171	18313903
12	53	636423
13	576	1384983
16	2	32410
17	246	132420
22	37	1100713
Total	4035	39550076

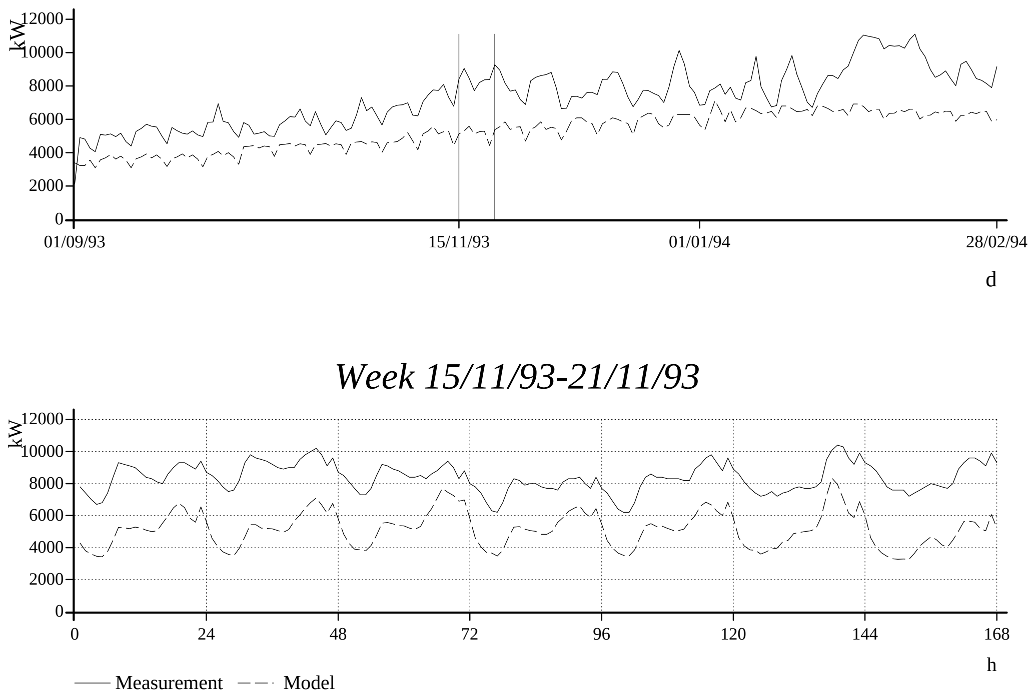


Fig. 16. Example 4. Total energy measurement from the feeder of the Kulju substation. Average daily load (above) and hourly curves over one week (below).The load models are from the alternative classification which includes 18 different customer classes categories. (See Appendix 2.2). The shapes of load models are similar although the level is again too low.

4 DERIVATION OF STATISTICAL DISTRIBUTION FUNCTIONS FOR CUSTOMER LOAD

4.1 INTRODUCTION

The loads of customers, even of the same type, are not likely to be the same at the same time. The load of one specific customer is usually different at different times. The load of a customer can't be exactly predicted. We say that the load is a randomly distributed variable or a random variable. The most usual parameters describing a random variable are *mean*, *standard deviation* and *variance*.

More information on the probabilities of a random variable is presented in the *distribution* of the variable. For most of the computations we need the *distribution function* where the probability is approximated as a function of the variable. In statistics usually, and when there is no knowledge of the statistical distribution function, the normal distribution is assumed (Pahkinen & Lehtonen 1989 p. 12). Also with electric loads the most usual assumption has been normal distribution (Fikri 1975 p. 3.07). Especially when it comes with cumulated independent loads the Central Limit Theorem states that the distribution function converges to normal distribution.

In a report on statistical methods for load research data analysis the conclusion is that the statistical distribution of electric load variation does not follow any common probability density functions (SRC 1983 p.3 - 47). The goodness of fit of estimated distributions is in general very low.

The Weibull distribution was fitted to consumer billing data in (Irwin et.al. 1986). The Weibull distribution was found flexible enough to explain the distribution of customers' annual energy use in different areas in Northern Ireland. The analysis covered only customer billing data and annual energy.

In 1993 Herman & Kritzinger published results from fitting statistical distribution functions to grouped domestic loads. The distributions tested were Weibull, normal, Erlang and beta. As a result they propose the use of the beta distribution function. No applications using the beta function are described. The goodness of fit is in general low for all distributions.

The estimation of statistical distribution functions with poor results was also experienced in the first analysis of the Finnish load research project (Seppälä 1984 pp. 45 - 60). From the load data we can see, that the loads are distributed around the mean and in many cases the distribution has a bell shape. Some loads at some time don't seem to have any regular distribution at all. This is especially the case with automatically controlled loads, like loads of storage heating during night hours.

However, looked at another way, two examples of experienced load distributions from the load research data shown in Fig. 17 give some new ideas for finding the distribution function for hourly electric customer class loads. These figures were produced by cumulating several hours' data to one set by scaling each hour's sample data between zero and one. We can see the average distribution over several hours. These figures show a good fit to normal distribution on day hours (high load) and lognormal distribution on night hours (low load).

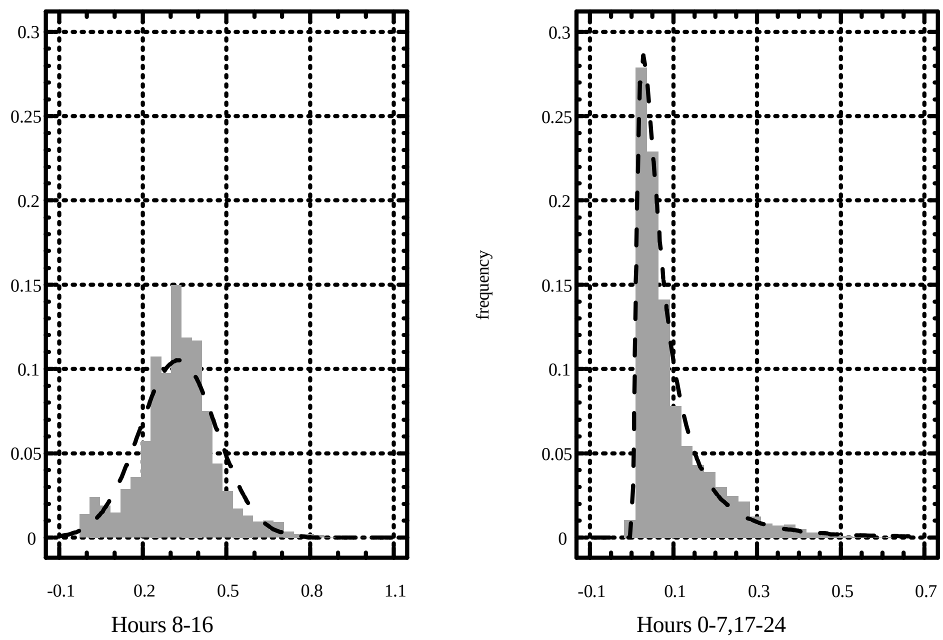


Fig. 17. Examples of load distributions of one customer class with estimated normal (left) and lognormal (right) distribution functions. The data is obtained by scaling each hour's load values between 0 *and* 1*. The maximum value of each hour gets the value* 1 *and other values proportionally less than* 1*. Customer class is Private Service during January working days.*

4.2 NORMAL DISTRIBUTION AND LOGNORMAL DISTRIBUTION FUNCTIONS

Normal distribution is the limit of distribution of random sums and also it is the limit distribution of many other distribution functions. A random variable x with density

$$\varphi(x) = \frac{1}{\sqrt{2\pi\sigma}} e^{-\frac{1}{2}\left[\frac{x-\mu}{\sigma}\right]^2} \tag{8}$$

is said to have *normal* distribution with parameters μ (mean) and σ (standard deviation).

The normal distribution with μ = 0 and σ = 1 is called a *unit normal distribution* N(0,1). A normal distribution function is defined by the mean and standard deviation of the population, which makes it easy to estimate.

It is, in many circumstances, possible to determine the function[2] $g(x)$ which will transform the skew distribution into a normal one.

$$F\left(\frac{g(x)-\xi}{\sigma}\right) = \mathrm{N}(0,1) \tag{9}$$

In practical work the interest is to find a function of the type where $g(x)$ does not include any unknown parameters. According to Johnson & Kotz (1970) the only transformation with statistical importance is a logarithm transformation. The data is said to be *lognormal* distributed when the logarithm of the data is normally distributed

$$F\left(\frac{\ln x-\xi}{\sigma}\right) = \mathrm{N}(0,1) \tag{10}$$

ξ and σ can be estimated from the observations by taking the logarithm of the data and estimating the mean and standard deviation. The distribution density function of lognormal distribution is

$$\Lambda(x) = \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left[\frac{\ln x-\xi}{\sigma}\right]^2} \tag{11}$$

Some aspects make lognormal distribution interesting in the study of electric loads. Lognormal distribution is used widely in statistics as a distribu-

[2] In general it holds for all monotony functions of g

$$\Pr\{x < a\} = \Pr\{g(x) < g(a)\}$$

tion of consumption, and customer electric hourly load can also be regarded as hourly electric energy consumption. Another interesting point is that the lognormal density function tends to normal density function when the ratio between standard deviation and the mean becomes small (Aitchison & Brown 1957).

4.3 THE PHYSICAL BACKGROUND OF LOAD VARIATION

The variability of electric loads is a result of complicated processes and interactions between electrical appliances, environmental factors and human behaviour. The origin of load variation within the customer loads in a customer class can be split into three categories:
- variation in customer's human behaviour
- variation in environmental conditions
- variation in electrical appliances and installations.

The customer load is limited between zero and the total power of installed equipment. Thus the load variation is limited from both directions. The use of the appliances is connected to each other by wiring or by "logical connection". The "logical connection" means that the appliances are usually used together, like kitchen lightning which is usually switched on at the same time as electric cooking appliances.

Customer's behaviour is in general the background to the load variation. The peoples' daily living rhythm and quite strictly regulated daily working hours bring a very regular portion to load variation. The variation of the residential load is a result of people's varying activities at home. Industry and service have day/night and workday/holiday schedules which are also clearly observed in load variation.

We shall now study the properties of electric loads as a result of a physical random process. Certain processes produce a known statistical distribution. The most common example is the sum of independent random variables which results in normal distribution according to the Central Limit Theorem. In the following chapters we develop models of processes which generate similar distributions as observed in load research data.

We assume the appliances, installations and environmental conditions to be similar. Thus the variability of electric load is caused by the variability of human behaviour. The human variability here is approximated as a set or sequence of small random actions altering the use of the electrical appliances.

The value of load P is observed from the energy W_T used in a time interval T (typically T = 15, 30 or 60 min).

$$P = \frac{W_T}{T} \tag{12}$$

The length of time T determines how much randomness in the load values is included. The shorter the interval T is, the more random that load value is.

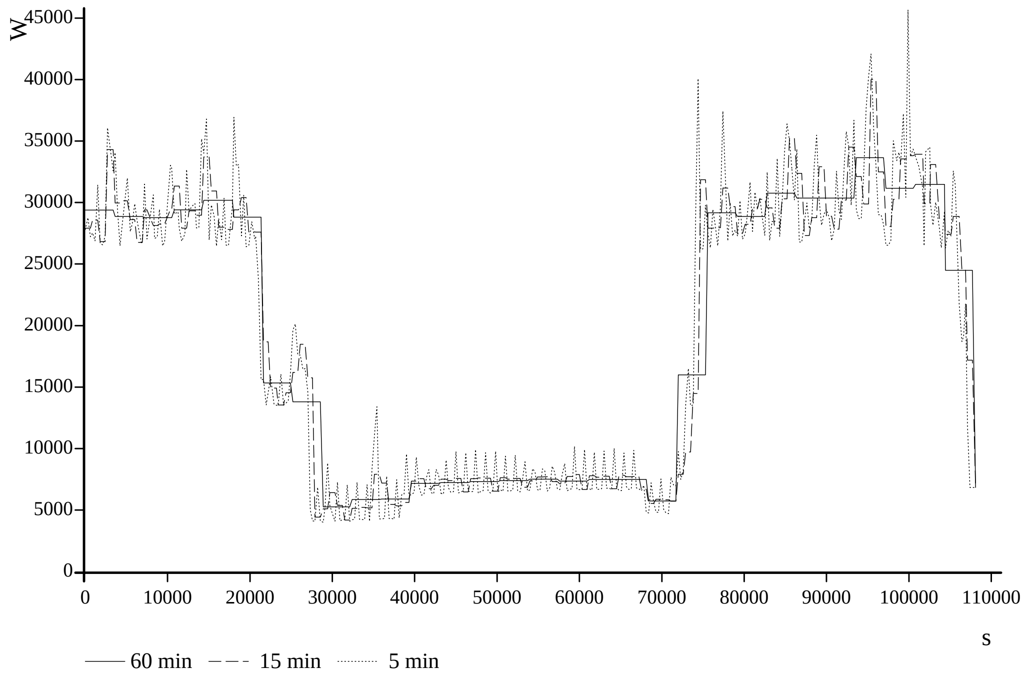


Fig. 18. Example of a load over two days observed with different time intervals T. The measurement is the total load of an office building including different service companies.

In Fig. 18 one customer's load P measured with different time intervals is presented. The longer the interval is, the more stable the load is. The time interval in the Finnish load research project is 60 min.

4.4 DERIVATION OF CUSTOMER LOAD DISTRIBUTION - BINOMIAL PROCESS

4.4.1 General

The customer is seen here as a system of many appliances switched on or off (see Fig. 19), which leads us first to study how the load distribution could be represented as a result of a binomial distribution process. There are two possibilities: an additive and a multiplicative binomial process (Hald 1967 p.31 - 33).

If the probability of a certain appliance to be switched on is θ then the probability of k appliances out of n being switched on is

$$\Pr\{n,k\} = \binom{n}{k}\theta^{k}(1-\theta)^{n-k} \qquad (13)$$

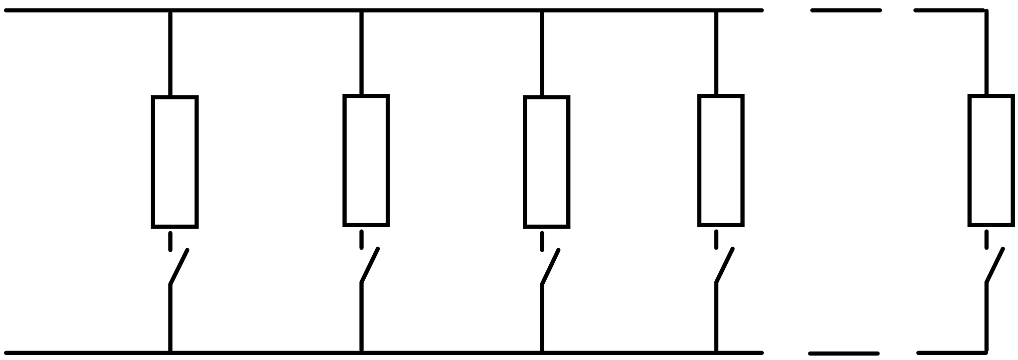


Fig. 19. Customer load as a combination of small loads.

4.4.2 Independent small loads - additive binomial process

We define the random variable as the demand of energy w over a time period T. The randomness of w is caused by small random deviations ε_i added to w_0 in a random sequence $w_n = w_0 + \varepsilon_1 + ... + \varepsilon_n$, where $\Pr\{\varepsilon_i = \varepsilon\} = \theta$ and $\Pr\{\varepsilon_i = 0\} = 1 - \theta$. The probability of the value of w_n after n steps follows in Table 7.

Table 7. The probability of deviations of the additive process.

Value of w_n	Probability
w_0	$(1-\theta)^n$
$w_0+\varepsilon$	$\binom{n}{1}\theta^1(1-\theta)^{n-1}$
$w_0+2\varepsilon$	$\binom{n}{2}\theta^2(1-\theta)^{(n-2)}$
$\vdots$	$\vdots$
$w_0+n\varepsilon$	θ^n

The distribution tends to become symmetrical, approximating normal distribution when $n \to \infty$ (DeMoivre-Laplace theorem) (Papoulis 1965 p. 66). The distribution of w tends to become normal even if the deviations (ε) are not the same in every phase, as long as all *deviations are of the same order of magnitude*. See also Galton's distribution machine in Appendix 5.

4.4.3 Interdependent loads - multiplicative binomial process

We assume that the loads are in such a connection that they all react together. Thus the changes are relative instead of additive and the process is multiplicative $w_n = w_0\,(1 + \varepsilon_1)...(1 + \varepsilon_n)$, where $\Pr\{\varepsilon_i = \varepsilon\} = \theta$ and $\Pr\{\varepsilon_i = 0\} = 1 - \theta$. Then we have a similar Table 8.

Table 8. Probability of deviations of the multiplicative process.

Value of w_n	Probability
w_0	$(1-\theta)^n$
$w_0\,(1+\varepsilon)$	$\binom{n}{1}\theta^1(1-\theta)^{n-1}$
$w_0\,(1+\varepsilon)^2$	$\binom{n}{2}\theta^2(1-\theta)^{(n-2)}$
$\vdots$	$\vdots$
$w_0(1+\varepsilon)^n$	θ^n

Writing

$$\begin{cases} \ln w_n = y_n \\ \ln(1+\varepsilon) = \delta \end{cases} \qquad (14)$$

the distribution in Table 8 takes the same form as in Table 7. See Table 9.

Table 9. Logarithm of the multiplicative process.

Logarithm of w_n	Probability
y_n	$(1-\theta)^n$
$y_n+\delta$	$\binom{n}{1}\theta^1(1-\theta)^{n-1}$
$y_n+2\delta$	$\binom{n}{2}\theta^2(1-\theta)^{(n-2)}$
$\vdots$	$\vdots$
$y_n+n\delta$	θ^n

Thus, the logarithm of the variable is approximately normally distributed. See also Kapteyn's skew curve machine in Appendix 5.

Next we want to get closer to the technical properties of the loads of electrical appliances and develop a more complete load variation model.

4.5 DERIVATION OF CUSTOMER LOAD DISTRIBUTION - KAPTEYN'S DERIVATION

4.5.1 General

In this chapter the form of load distribution is derived using some selected simplifying assumptions and Kapteyn's derivation. The goal of this chapter is to give background on how a model of process of electrical appliances leads to lognormal distribution as a special case (Seppälä 1996).

4.5.2 Definition of customer load

The customer load is the total energy W_T consumed during a time interval T of certain length (5, 15, 30, 60 min etc.) (See Fig. 20). The value of the average power is then

$$P = \frac{W_T}{T} \tag{15}$$

The energy W_T is a sum of the energies of the customer's appliances

$$W_T = \sum_k w_{Tk} \tag{16}$$

where the energy consumed in T by an appliance k is w_{Tk} . This energy depends on the fixed nominal power $P_{N,k}$ of the appliance and the time τ_k the appliance is used in T

$$w_{Tk} = P_{N,k} \cdot \tau_k \tag{17}$$

Also we define the status $s_k(t) = 1$ when an appliance is switched on and $s_k(t) = 0$ when the appliance is switched off.

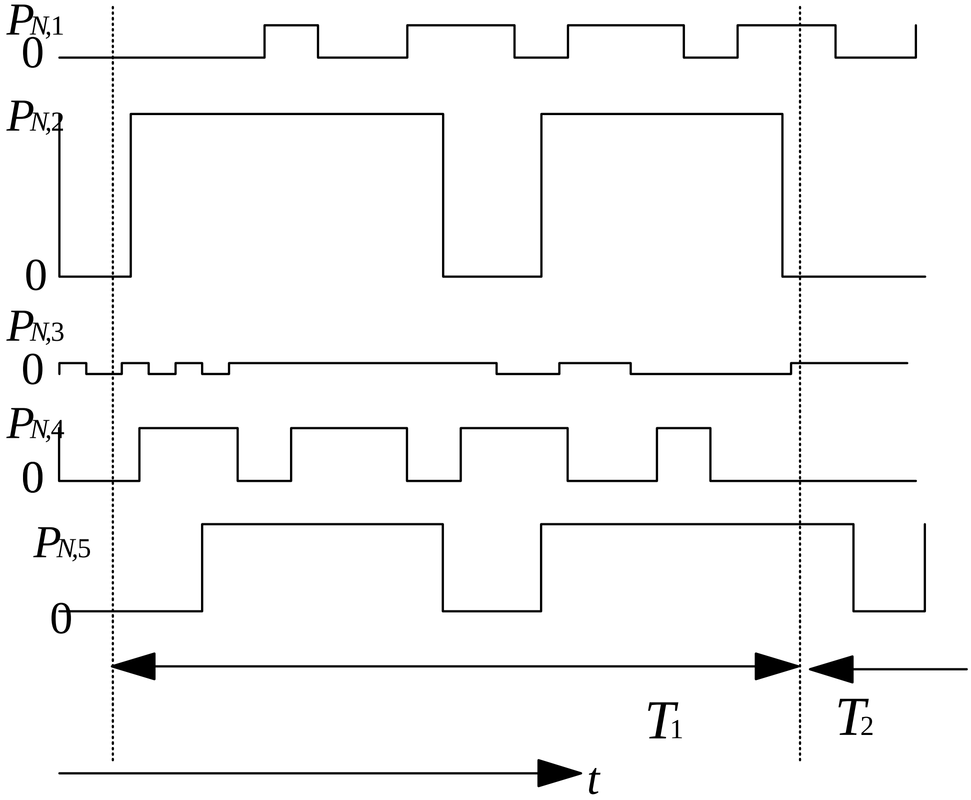


Fig.20. Customer's total load consists of several appliances, which have fixed nominal power $P_{N,1}$, $P_{N,2}$, $P_{N,3}$, $P_{N,4}$, $P_{N,5}$. The total energy demand over T varies when the time of use of the distinct appliances varies.

4.5.3 Customer's random action and reaction of electric appliances

The customer's influence on the electric load is an action changing the time of use of the appliances by a small random value ε_i ($\ll T$).

We assume that an action ε_i affects the time of use of appliance k by value $\Delta(\tau_k)_i$ according to the following equation (18). See also Fig. 21.

$$\begin{cases} \Delta(\tau_k)_i = s_k(t)\varepsilon_i & , 0 \le \tau_k < T \\ \Delta(\tau_k)_i = 0 & , \tau_k = T \end{cases} \qquad (18)$$

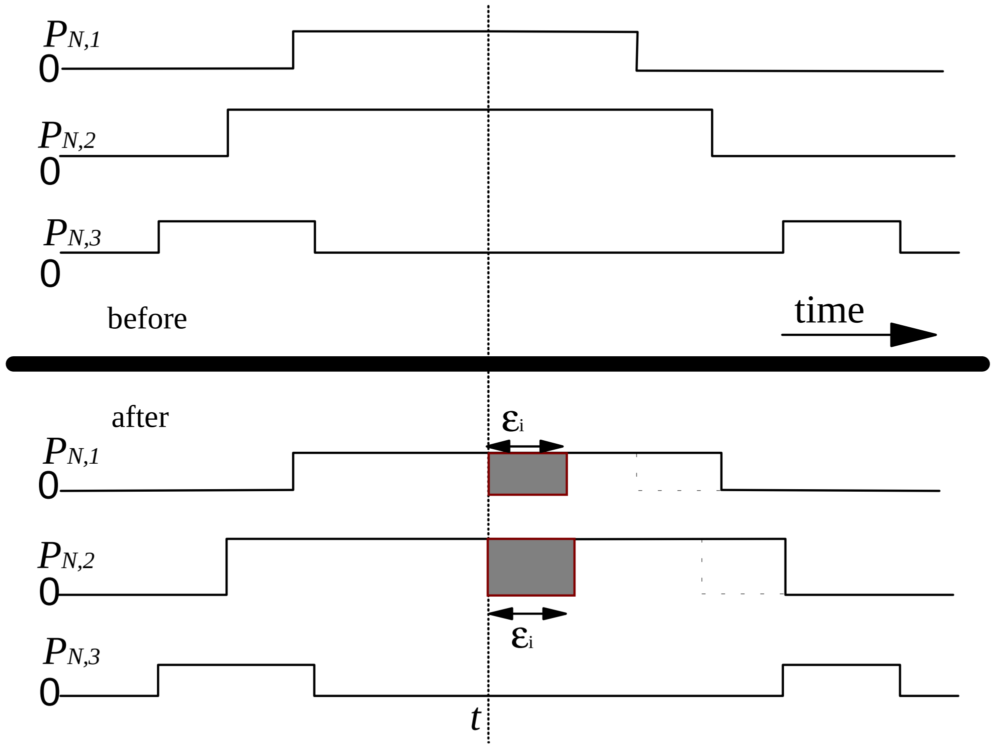


Fig. 21. The reaction of appliances after the customer alters the time of use of the appliances by the value ε_i . Time of use of load $P_{N,1}$ and $P_{N,2}$ changes by the amount of ε_i . $P_{N,3}$ remains unchanged (below the thick line).

The assumptions of this model (18) are

1. The action of ε_i takes place at a random time t.
2. The influence of this random action to any appliance k depends on the momentary status $s_k(t)$ of the appliance.
3. The appliance's time of use τ_k is uniformly distributed over time interval T.
4. The resulting influence $\Delta(\tau_k)_i$ is observed only from the varying length of time the appliance is used. When $\tau_k = T$ the load is at maximum and no change will be observed.

The expected value of the state of an appliance k is according to the assumption 3.

$$s_k(t) = \frac{\tau_k}{T} \tag{19}$$

Combining the two previous equations (18) and (19) we get the expected change of the time of use of one appliance k

$$\Delta(\tau_k)_i = \frac{\tau_k}{T}\varepsilon_i \delta_k \qquad (20)$$

where

$$\delta_k = \begin{cases} 1 & when\ 0 \le \tau_k < T \\ 0 & when\ \tau_k = T \end{cases} \qquad (21)$$

δ_k is the parameter for each appliance according to assumption 4. If the time of use of appliance k is already T, it can not grow (δ_k = 0) any more. Then the appliance is being used to its maximum capacity.

4.5.4 Customer's random actions and reaction of customer's total load

The customer's total energy use over a time interval T is W_T . Substituting (16) to (17) we get

$$W_T = \sum_k \tau_k \cdot P_{N,k} \qquad (22)$$

The difference of the load $\Delta(W_T)_i$ is a sum of the differences of time of use of distinct appliances

$$\Delta(W_T)_i = \sum_k \Delta(\tau_k)_i \cdot P_{N,k} \qquad (23)$$

And applying the previous result (20)

$$\Delta(W_T)_i = \sum_k \frac{\tau_k}{T}\varepsilon_i \delta_k P_{N,k} \qquad (24)$$

To eliminate δ_k from the equation we assume the value of the average power to be far lower than the total installed power

$$P << \sum_k P_N \qquad (25)$$

when we can "safely" assume $\tau_k < T$ for almost all k and approximate $\delta_k \approx 1$ for all k (No appliance is used at its full capacity). Then the change of load gets a value

$$\Delta(W_T)_i \approx \frac{\varepsilon_i}{T} W_T \qquad (26)$$

4.5.5 Definition of the reaction function with low load

Assuming the previous model, the total energy of a customer over an interval T (for example one hour) is subject to a process which successively alters the load magnitude from the expected value W_0 to W_1, W_2,..W_i.... The origin of this process is the customer successively varying the time of use of the electrical appliances. Each step i corresponds to one small random change $(\Delta\tau_k)_i$.

Assuming the average load P is far from the maximum so that $\tau_k < T$ for almost all k, the expected difference between two phases $\Delta(W)_i = W_i - W_{i-1}$ is, applying the previous result (26)

$$W_{i+1} - W_i = (\Delta W_T)_i = \varepsilon_i \frac{W_i}{T} = r(W_i)\varepsilon_i \quad , \quad \tau_k < T \quad , \quad P = \text{"}low\text{"} \qquad (27)$$

r is *reaction function* needed in the following Kapteyn’s derivation.

$$r(W_i) = \frac{W_i}{T} \qquad (28)$$

4.5.6 Kapteyn’s derivation of a skew distribution

Now we study how a sequence of small random changes taking place in a certain order affects the value of W. The sequence of n changes (ε_1,…, ε_n) of the load W expressed with the help of the reaction function r will be

$$\begin{aligned} W_1 &= W_0 + \varepsilon_1 r(W_0) \\ W_2 &= W_1 + \varepsilon_2 r(W_1) \\ &\vdots \\ W_n &= W_{n-1} + \varepsilon_n r(W_{n-1}) \end{aligned} \qquad (29)$$

Adding these equations and solving for ε_i, we get the following result

$$\sum_{i=1}^{n} \varepsilon_i = \sum_{i=1}^{n} \frac{W_i - W_{i-1}}{r(W_{i-1})} \approx \int_{w_0}^{w_n} \frac{dW}{r(W)} = g(W) \qquad (30)$$

According to the Central Limit Theorem $\sum_{i=1}^{n} \varepsilon_i$ will be normally distributed when $n \to \infty$. Also then the function $g(W)$ is normally distributed.

$$g(W) = T \int_{W_0}^{W_n} \frac{dW}{W} = T(\ln W_n - \ln W_0) \qquad (31)$$

The conclusion, according to the assumptions of the model, is that the customer load distribution is lognormal when the average load level is “low” and technical and environmental conditions are similar within the customer class.

4.5.7 Simulation of the customer load distribution

The simulation of the previous load distribution model (18) is now studied. The simulation was done with a computer with APL. The representation of a system of distinct loads over a time interval T was done using matrices of bits (bitmaps).

Each row of the bitmap represents one electrical appliance (1...k) and each column a short fixed slice of time ($t_1...t_n$) of length $\Delta t = t/n$. When the appliance is switched on the corresponding state bit $s(t)$ has the value 1 and when the appliance is switched off the corresponding bit has the value 0 (see Table 10). The time slices approximate the small time interval $\varepsilon_i = \Delta t$.

Table 10. An example of a bitmap representing the use of appliances at consecutive time slices $t_1,...,t_n$.

	$s(t_1)$	$s(t_2)$	$s(t_3)$	$s(t_4)$	...	$s(t_n)$
Appl. 1	1	1	1	0	...	1
Appl. 2	0	1	0	1	...	0
Appl. 3	0	0	1	0	...	0
...	...	...	...	...	...	...
Appl. k	0	0	1	1	...	1

The total energy is then related to the sum of the elements of the bitmap. If the corresponding bit is 1 for an appliance at time t_i, the next time slice having value 0 will be turned to 1. If the value at t_i is 0 nothing will be changed. For example the previous table would look like the following if the increment takes place at t_2. Appl. 1 and Appl. 2 are switched on and their use will be increased by turning the next 0 values to 1 (Table 11).

*Table 11. The increment of use at time t_2. The changed values are printed as bold italic (**1**).*

	$s(t_1)$	$s(t_2)$	$s(t_3)$	$s(t_4)$	...	$s(t_n)$
Appl. 1	1	**1**	1	***1***	...	1
Appl. 2	0	**1**	***1***	1	...	0
Appl. 3	0	**0**	1	0	...	0
...	...	**...**	...	...	...	...
Appl. *k*	0	**0**	1	1	...	1

4.5.8 An example of the results of the simulation

Frequency histograms of a simulated customer's load variation are presented here in Fig. 22.

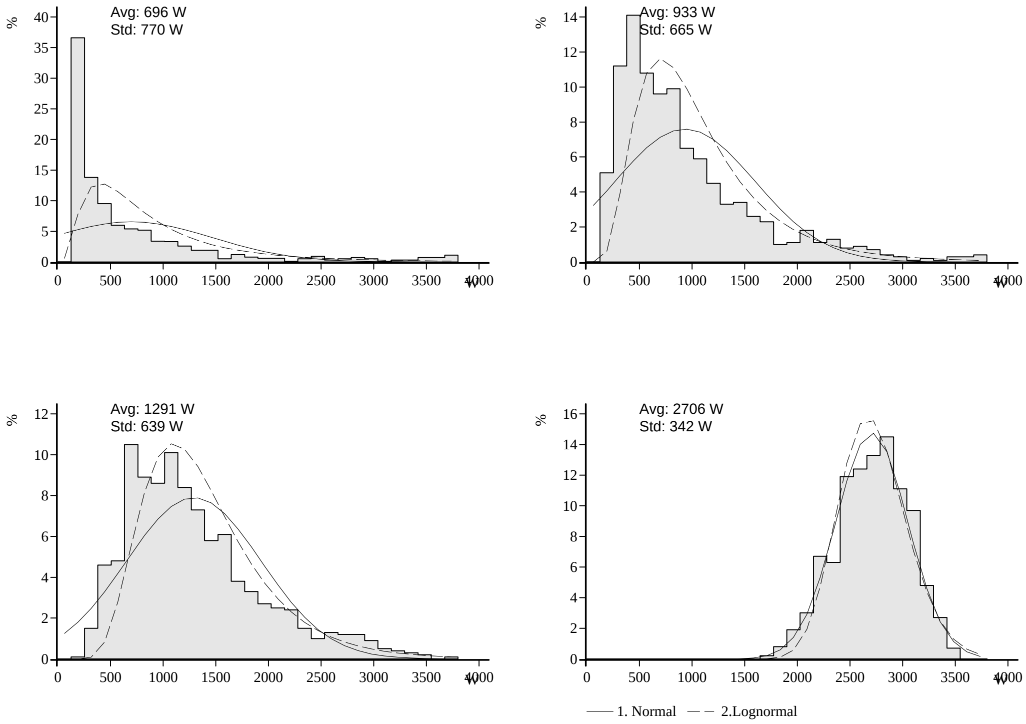


Fig. 22. Result of simulation of load variation of an example household with different load levels and standard deviations.

The customer is assumed to have the following electrical appliances installed:

Appliance	P_N/W	number
light	60	5
light	100	5
heat	500	4
cook	1000	1

The maximum hourly load is then 3800 W. Four simulations, where the expected load is 696 W, 933 W, 1291 W and 2706 W, were calculated.

The result of the simulation confirms that the load distribution remains lognormal with small average loads. Also we see how the distribution is close to normal when average load is larger, which confirms the result in Fig. 17. The fitting of these distributions to distribution functions will be studied in the next chapter.

The simulation model is interesting also for practical applications because here the influence and interaction of customers with distinct electrical appliances would be quite easy to define. As we see, with the computation capability of computers continuously increasing, the simple form of the simulation program makes it attractive to use in applications. The applications of Demand Side Management (DSM) and customer load control may find this method useful in studying the influence of the DSM and load control actions on the customers' load variation.

4.5.9 Discussion

The previous approach is an application of methods first published in 1903. These methods were applied to the study of botanical science phenomena. The question stated by Kapteyn (1916) was why skew distributions exist in nature? The conclusion was roughly that, causes which are independent of the size of the individuals, produce normal curves and causes which are dependent on size, produce skew (in special case lognormal) curves.

This also seems natural for electricity demand. The causes changing the customer's use of electrical appliances are dependent on the use of the appliances. It is possible to define the distribution function at least in one special case. This leads us to assume that there is a tendency to skew distribution somewhere between normal and lognormal distribution.

The model of interacting electrical appliances can be explained in many ways. When the customer's activity increases or decreases it influences all the appliances in use (TV, cooking, lightning). In industry the use of the machines is usually linked together and in offices the use of lightning and ventilation and computers is related to each other. When the time of use of one appliance changes, similar changes in the use of other appliances can be expected.

This model is an approximation of the situation when the load level is far lower than the maximum load and explains the skewness of the load distribution and the selection of lognormal distribution. When the load P grows and some of the appliances are used in maximum capacity, the distribution obviously becomes closer to normal distribution.

The real interactions between electric loads are much more complicated than the previous model. The loads may interact with negative correlation while the increase in use of one appliance decreases the use of another appliance. However the model is easy to simulate which opens possibilities to further study the properties of load variations. The generalisation of this model should be a subject for further study.

Kapteyn’s approach is open to mathematical criticism. Also when these ideas were published, some mathematicians, especially Pearson, criticised the conclusions of how transformation of data to normal distribution was obtained. However this method was appraised by those who found it giving better insight into the systems which obviously lead to skewed distributions (Baart de la Faille 1915 and Aitchison & Brown 1957 pp. 20 - 22).

From the older statistical literature ”distribution machines” can be found. They were simple apparatus to simulate certain distributions based on binomial processes. One famous one is Galton’s normal curve apparatus from 1889 (Hald 1965 p. 32). A corresponding skew curve machine was made by Kapteyn (1916 fig 7.) See Appendix 5. The appearance of these apparatus may help also the modern reader to understand the physical origins of random distributions.

5 ESTIMATION OF CONFIDENCE INTERVALS OF CUSTOMER LOADS

5.1 GENERAL

In distribution applications the main problem is determining the dimensions for the network components and monitoring the loading of the installed network. Therefore the most interesting information from the electric loads is the highest values and the probability of their occurrence.

In statistical terms the question is to estimate confidence intervals (or confidence limits) for the load variation. The confidence intervals are related to the statistical distribution of the load. There are two possible strategies for estimating confidence intervals: calculating them directly from the observed data or estimating them with the help of a suitable distribution function.

In distribution load flow calculations the normal distribution function has been practically the only approximation (Fikri 1975) (Juuti et al. 1987). Normal distribution is the best choice for cumulated independent customer loads. When the load of one or two customers should be estimated the normal distribution tends to be unreliable. According to the previous chapter other interesting distribution functions are the lognormal distribution and distributions somewhere between normal and lognormal distribution.

There are also plenty of methods which are based on the re-use of data, i.e. bootstrap (Pahkinen & Lehtonen 1989, pp. 227 - 243) and (Räsänen 1995). However the preference of this study is to find simple parameterised confidence interval estimators, which are easier to adopt to the electric utilities' current applications.

This chapter covers the case of estimating confidence intervals for one customer from a customer class. The estimation of confidence intervals of several customers' cumulated load from the same customer class is studied in chapter 6. The most general case of cumulated loads of several customers from different classes is more complicated while the correlation between and within the customer classes should be estimated. This case will be left for further study.

In network load flow calculations these results are most applicable when calculating the low voltage network where the number of customers is small. Also the results of this analysis are suitable for energy sales where one wants to analyse the risk of one big customer increasing the total load. Because of the different requirements of applications, the target is also to find the most general method which is independent of time and customer class.

5.2 INTRODUCTION

The confidence interval of load $P_{\alpha\%}$ is defined here as a positive load value under which the load remains with a given probability (see Fig. 23 and Fig. 24):

$$\Pr\{P \leq P_{\alpha\%}\} = \frac{\alpha}{100} \qquad (32)$$

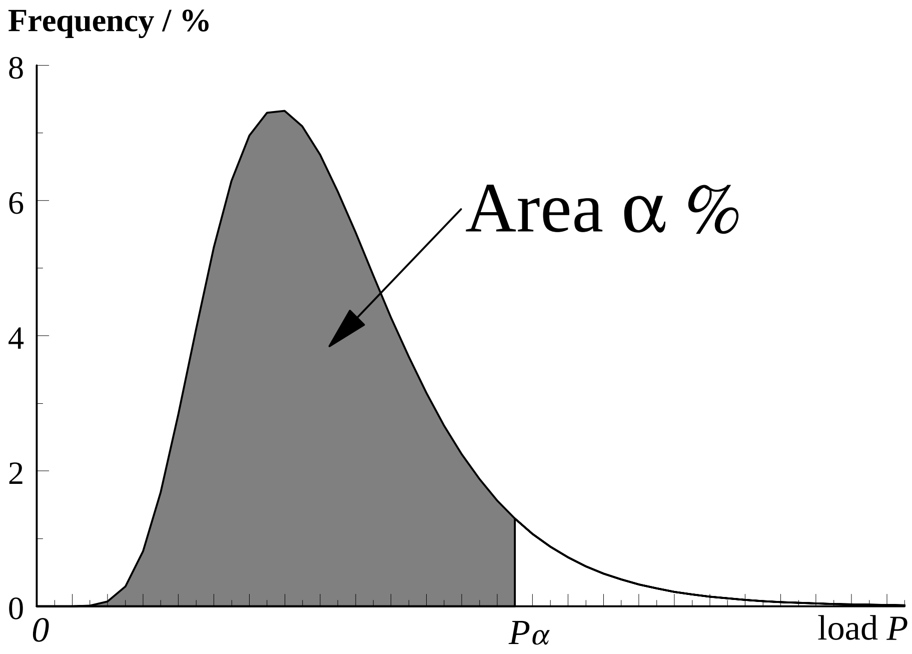


Fig.23. The distribution frequency and the confidence interval P_{α}.

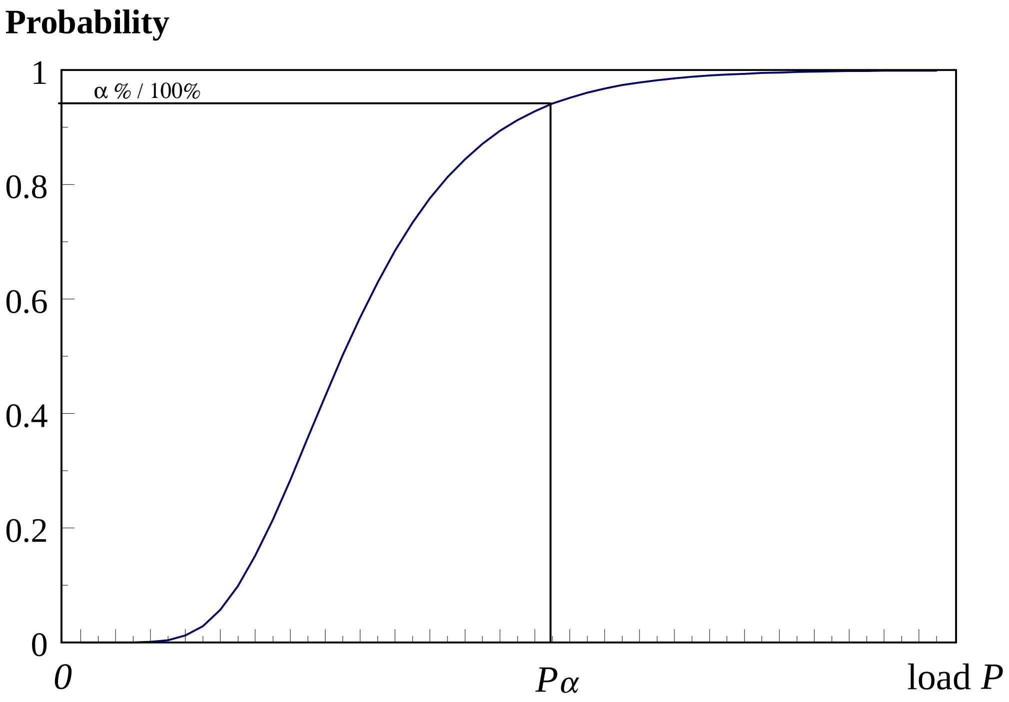


Fig. 24. The distribution of load and confidence interval P_α.

Applying the simple form load models in equation (4) the confidence interval will be estimated by estimating the confidence interval for parameter *L*.

$$\hat{P}_\alpha = \hat{L}_\alpha W_a \tag{33}$$

5.2.1 The measure for the accuracy of confidence interval estimation

To verify the different estimation methods, two error values q_1 and q_2 are calculated for each hour. They are defined from the estimated and observed percentiles and confidence levels according to Fig. 25.

$$\begin{aligned} q_1[\%] &= \alpha_{estimated}[\%] - \alpha_{observed}[\%] \\ q_2[\%] &= 100\,\frac{\hat{L}_\alpha - L_{\alpha\ observed}}{L_{\alpha\ observed}} \end{aligned} \tag{34}$$

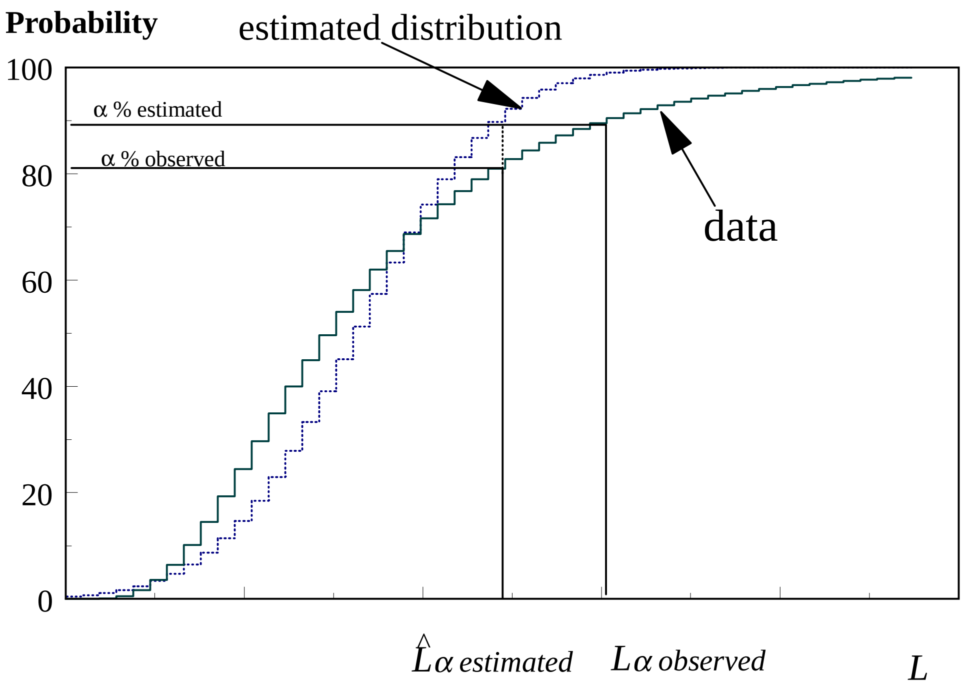


Fig. 25. Verification of the estimated confidence interval.

q_1 defines what per cent of the observed loads are actually below the estimator. q_2 defines what per cent the estimator $\hat{L}_\alpha$ differs from the percentile L_α observed from the data. q_2 represents the value and direction of how much the confidence interval estimator deviates from the value observed from the sample data.

While q_1 and q_2 are calculated separately for each hour, for each estimation method and for each estimated confidence interval, the average of q_2 is selected to represent the overall error of the method over several hours and customer classes.

5.2.2 The customer classes selected for this study

The selected customer classes for this analysis are shown in Table 12. These classes are the most important for distribution utilities. The number of classes is also limited to nine to keep the computation data and time within reasonable limits.

Table 12. Selected customer classes for analysis of confidence interval estimation.

Class	Description
810	Industry, 1-shift, all branches
820	Industry, 2-shift, all branches
910	Service, public, all branches
920	Service, private, all branches
110	Residential, one family house, direct electric heat
120	Residential, one family house, direct electric heat, water boiler at night
602	Residential, one family house, no electric heat, electric sauna
220	Residential, one family house, partly electric storage heat
712	Agriculture, milk production, residence included

Overall figures of the models of these classes are presented in Appendix 4.1.

5.3 DESCRIPTION OF THE CONFIDENCE INTERVAL ESTIMATION METHODS

According to the previous results in chapter 4 this analysis concentrates on normal and lognormal distribution and their approximations. Five different confidence interval estimation methods will be studied:

- Normal distribution Estimation method: NE
- LogNormal distribution Estimation method: LNE
- LogNormal distribution Estimation method, variation A: LNEA
- LogNormal distribution Estimation method, variation B: LNEB
- Simplified LogNormal distribution Estimation method: SLNE

Recalling equation (5) the estimates of the simple form load model parameters are briefly m_1 for average and s_1 for standard deviation.

$$\begin{cases} L_c(m,d,h) = E\left\{\dfrac{W_h(m,d,h)}{W_a}\right\} = m_1 \\ s_{Lc}(m,d,h) = \sigma\left\{\dfrac{W_h(m,d,h)}{W_a}\right\} = s_1 \end{cases} \tag{35}$$

The number of items in a sample is N. The n:th item of load data of the specific class c, month m, day d and hour h is briefly L_n

$$L_n = \frac{(W_c(m,d,h))_n}{(W_{a,c})_n} \text{ , } n = 1 \ldots N \tag{36}$$

5.3.1 Normal distribution Estimation method: NE

The parameters for normal distribution are estimated using formulas

$$\begin{cases} m_1 = \dfrac{1}{N} \displaystyle\sum_{n=1}^{N} L_n \\ s_1 = \sqrt{\dfrac{1}{N-1} \displaystyle\sum_{n=1}^{N} (L_n - m_1)^2} \end{cases} \tag{37}$$

The estimators for percentiles are selected from respective percentiles of unit normal distributions U_α. See Table 13.

$$\hat{L}_{\alpha 1} = m_1 + U_\alpha s_1 \tag{38}$$

Table 13. Selected percentiles for unit normal distribution are found, for example, from (Milton & Arnold 1990, Table V pp. 637 - 638).

α	50 %	84.13 %	95 %	99 %	99.5 %	99.9 %
U_α	0	1	1.65	2.33	2.58	3.10

5.3.2 LogNormal distribution Estimation method: LNE

The parameters of log-normal distribution function are estimated in a like manner to normal distribution taking the logarithm of the data.

$$\begin{cases} m_2 = \dfrac{1}{N} \displaystyle\sum_{n=1}^{N} \ln L_n \\ s_2 = \sqrt{\dfrac{1}{N-1} \displaystyle\sum_{n=1}^{N} (\ln L_n - m_2)^2} \end{cases} \tag{39}$$

$$\hat{L}_{\alpha 2} = \exp(m_2 + U_\alpha s_2) \tag{40}$$

5.3.3 LogNormal distribution Estimation method variations

The confidence intervals remain in logarithm transformation (the order of the values is not affected). Thus we can estimate the parameters using corresponding percentiles of the observed distribution and the previous equation (40) and table Table 13. This method is tested with the median and $L_{84.13\,\%}$ when $U_\alpha = 1$ (41) and $L_{95\,\%}$ percentiles when $U_\alpha = 1.65$ (43):

$$\begin{cases} m_{3a} = \ln(\hat{L}_{50\%}) \\ s_{3a} = \ln(\hat{L}_{84.13\%}) - m_{3a} \end{cases} \tag{41}$$

$$\hat{L}_{\alpha 3a} = \exp(m_{3a} + U_\alpha s_{3a}) \tag{42}$$

In the following text this method is abbreviated LogNormal Distribution method A, LNEA.

$$\begin{cases} m_{3b} = \ln(\hat{L}_{50\%}) \\ s_{3b} = \dfrac{1}{1.65}(\ln(\hat{L}_{95\%}) - m_{3b}) \end{cases} \tag{43}$$

$$\hat{L}_{\alpha 3b} = \exp(m_{3b} + U_\alpha s_{3b}) \tag{44}$$

In the following text this method is abbreviated LogNormal Distribution method B, LNEB.

5.3.4 Simplified LogNormal distribution Estimation method: SLNE

The parameters of the lognormal distribution function are estimated here in a simplified way using the ordinary mean (m_1) and standard deviation (s_1) estimated in (37). The ”simplified” estimators m_4 and s_4 of the lognormal distribution are:

$$\begin{cases} m_4 = \ln m_1 \\ s_4 = \ln(m_1 + s_1) - m_4 = \ln(m_1 + s_1) - \ln m_1 \end{cases} \tag{45}$$

$$\begin{aligned}\hat{L}_{\alpha 4} &= \exp(m_4 + U_\alpha s_4) \\ &= \exp(\ln m_1 + U_\alpha (\ln(m_1 + s_1) - \ln m_1) \\ &= m_1 \left(1 + \frac{s_1}{m_1}\right)^{U_\alpha}\end{aligned} \tag{46}$$

The background for selecting this "ad hoc" method is the practical reason that the best known parameters of loads are the mean and the standard deviation and they are already available in the electric power distribution applications. Thus applying this method requires no additional data to be distributed to the current computing systems.

5.3.5 Properties of SLNE

SLNE here, represents the distribution which is "somewhere between the normal and lognormal distributions". This is explained by the known properties of the normal and lognormal distribution. In the following we assume that the data comes from lognormal distribution with parameters ξ and σ.

The relation between the parameters of lognormal distribution (ξ and σ) and mean and standard deviation (m_1 and s_1) can be derived from the definition of lognormal distribution (Johnson & Kotz 1970 p.115, Lokki 1980 pp. 436 - 438):

$$\begin{cases} m_1 = e^{\xi + \frac{1}{2}\sigma^2} \\ s_1 = e^{\xi} \sqrt{e^{\sigma^2} (e^{\sigma^2} - 1)} \end{cases} \tag{47}$$

The parameters of lognormal distribution are as in equation (39)

$$\begin{cases} m_2 = \xi \\ s_2 = \sigma \end{cases} \tag{48}$$

By substituting equation (47) to the equations of simplified lognormal distribution parameters in equation (45) we get

$$\begin{cases} m_4 = \xi + \frac{1}{2}\sigma^2 \\ s_4 = \ln\left(1 + \dfrac{e^{\xi}\sqrt{e^{\sigma^2}(e^{\sigma^2} - 1)}}{e^{\xi + \frac{1}{2}\sigma^2}}\right) = \ln\left(1 + \sqrt{e^{\sigma^2} - 1}\right) \end{cases} \tag{49}$$

In the following we now study the estimation error q_2 of SLNE compared to NE. Recalling the confidence interval estimators for normal distribution (NE),

$$\hat{L}_{\alpha 1} = m_1 + U_\alpha s_1 \tag{50}$$

and simplified lognormal distribution (SLNE) according to eq. (46)

$$\hat{L}_{\alpha 4} = m_1 \left(1 + \frac{s_1}{m_1}\right)^{U_\alpha} \tag{51}$$

and assuming the lognormal distribution as the actual distribution of the data

$$L_{\alpha 2} = \exp(\xi + U_\alpha \sigma) \tag{52}$$

we can study the relative estimation error value q_2 when the parameters of the distribution varies. We calculate for NE method

$$q_2[\%] = 100\,\frac{\hat{L}_{\alpha 1} - L_{\alpha 2}}{L_{\alpha 2}} \tag{53}$$

and for SLNE method

$$q_2[\%] = 100\,\frac{\hat{L}_{\alpha 4} - L_{\alpha 2}}{L_{\alpha 2}} \tag{54}$$

Setting the parameter $\xi = 0$ and varying the parameter σ the result is plotted as a function of s_1/m_1 in Fig. 26.

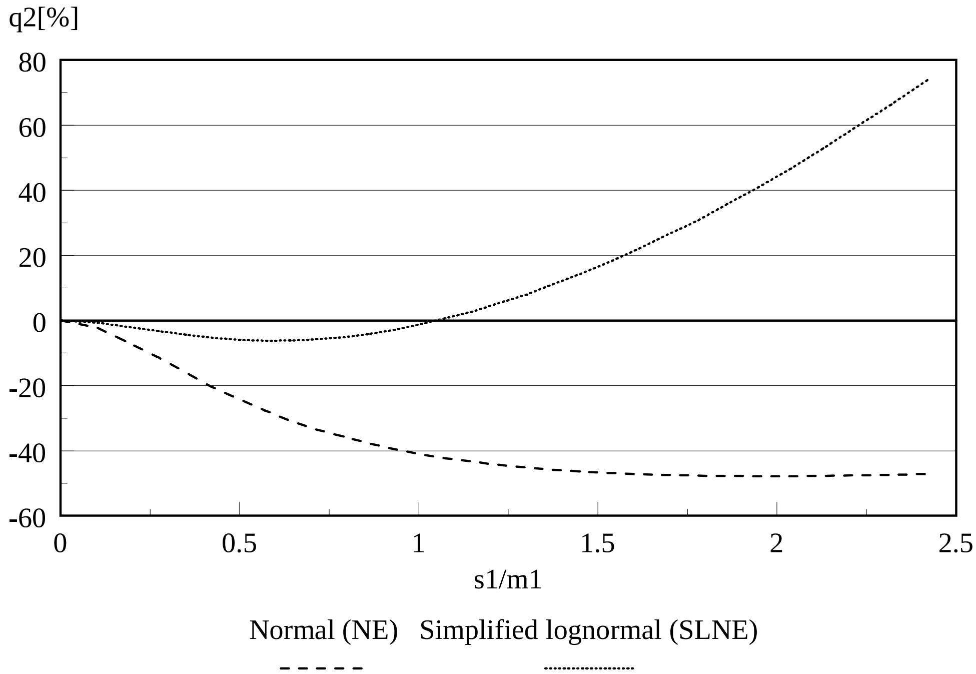


Fig. 26. The theoretical estimation error q_2 of 99.5 % estimators of NE and SLNE applied to lognormal distribution.

The curves in Fig. 26 show how the error of the NE method grows to a negative direction when the relation s_1/m_1 grows. The error of the simplified lognormal estimation is quite small when $s_1/m_1 < 1.5$. The error then grows rapidly when $s_1/m_1 > 1.5$. This explains why the SLNE is below the LNE but above the NE. Also it is accurate enough in load estimation when the s_1/m_1 of load data is mostly well below 1.5 as seen from Fig. 28 - Fig. 30.

5.3.6 The flow of computation estimating and verifying the estimators

The flow of the computation is presented in Fig. 27. The overall results from the estimation of the parameters are shown for three customer classes in Fig. 28, Fig. 29 and Fig. 30.

Comparing these figures with earlier load models presented in Appendix 4.1. we find them similar. The different monthly load curves, especially in 1-shift Industry, is due to the slightly different ways the charts were plotted: The curve in Fig. 28 is an average of work days but the monthly curve in Appendix 4.1 is an average of the whole month, holidays included.

Another important observation is that the results of electric heating are similar with the previous analysis presented in Appendix 4.1 although the temperature standardisation was applied to the analysis in Appendix 4.1 but not here. This is rather surprising as some reduction of variance due to the temperature standardisation is generally expected. The reason why the temperature standardisation did not reduce the variance should be a target for further study.

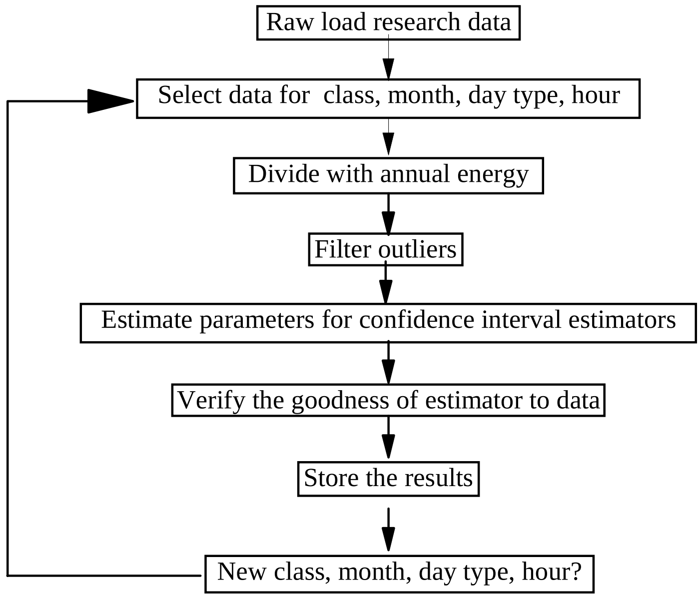


Fig. 27. Flow of the estimation and verification process.

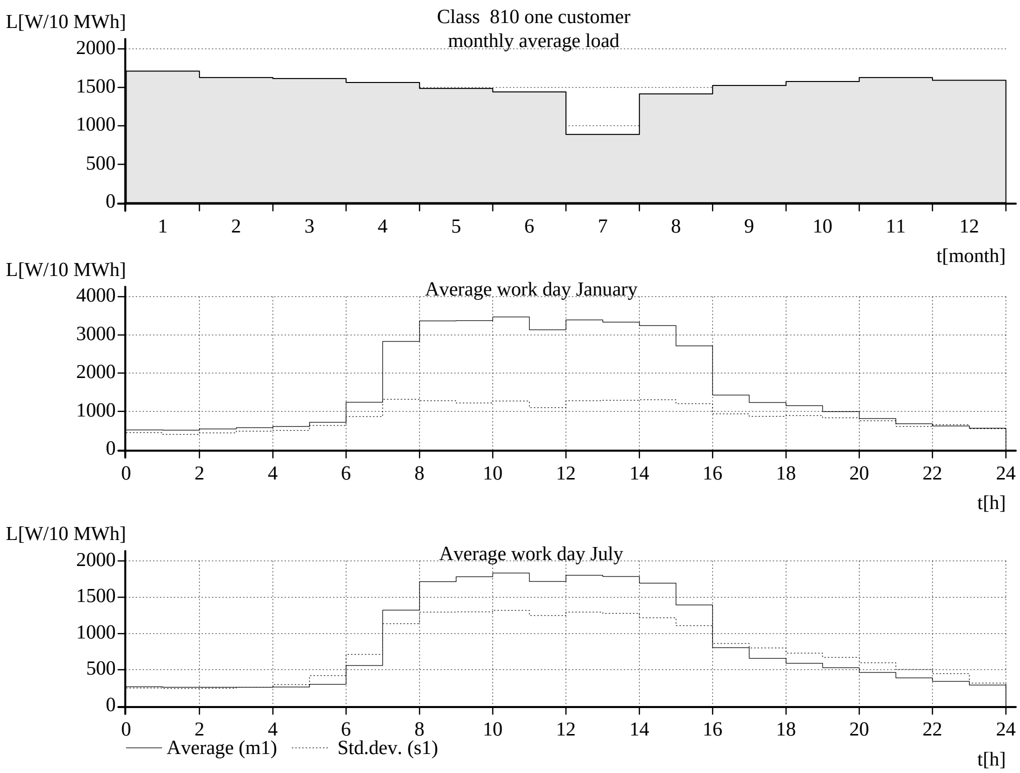


Fig. 28. Results of estimation of parameters m_1 and s_1. Class 810 Industry 1-*shift.*

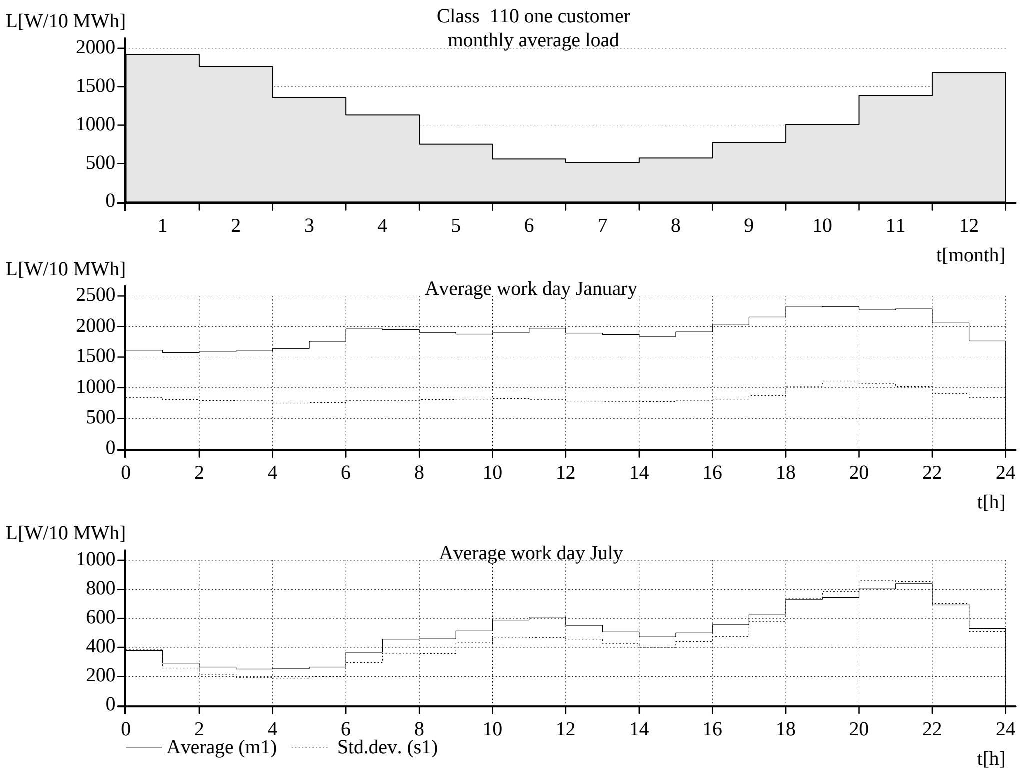


Fig. 29. Results of estimation of parameters m_1 and s_1. Class 110 direct electric heating, one family house. No temperature standardisation.

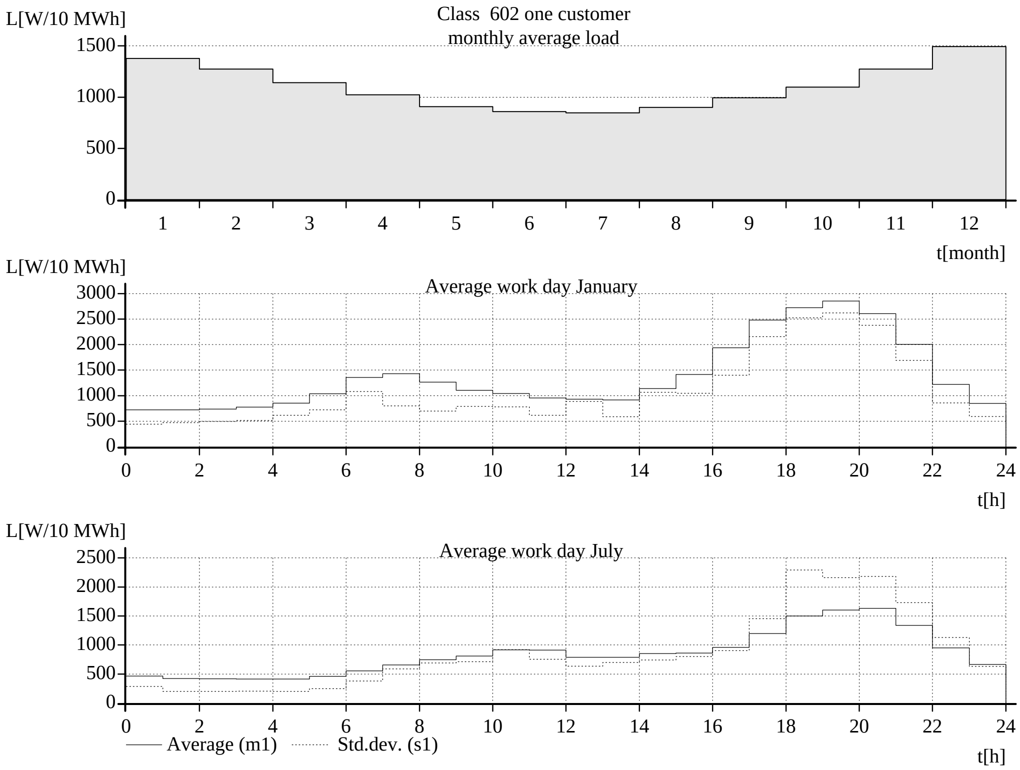


Fig. 30. Results of estimation of parameters m_1 and s_1. Class 602 residential, one family house.

5.4 VERIFICATION OF THE ESTIMATORS WITH THE LOAD RESEARCH DATA

5.4.1 General

The selected confidence intervals to be calculated will be 84.13 %, 95 %, 99 % and 99.5 %. The problem with this representation is to summarise and analyse the large number of different cases here.

The computational analysis was done for working days for
- 12 months
- 9 customer classes
- 24 hours
- 5 estimation methods
- 4 confidence intervals
- 2 estimation error values: q_1 and q_2

which make altogether $12 \cdot 9 \cdot 24 \cdot 5 \cdot 4 \cdot 2 = 103\,680$ values.

First we will study in chapter 5.4.2 the distribution functions with some selected cases which are assumed to be the most interesting. The selected customer classes are: industry 1-shift (810), residential electric heating (110) and residential (602). From these classes the hours 00.00 - 01.00 and 09.00 - 10.00 in January are selected to represent day and night. In ch 5.4.3 the error values for the selected hours are presented.

In ch. 5.4.4 we further limit the analysis to 99.5 % confidence interval and NE and SLNE estimation methods and finally in ch. 5.4.5 we study the estimation methods only with the maximum hour in each customer class.

5.4.2 Observed load distributions and estimated distribution functions

Figures 31 - 36 present distribution functions with the data from the load research. The selected data is from January work days, hours 00.00 - 01.00 and 09.00 - 10.00 of class 810 Industry 1-shift, class 110 residential, direct electric heating and class 602 residential without electric heating. The upper graph covers all data and the graph below is focused on the tail area to show the distribution of extreme values better. See also the graphs of the original data in Appendix 4.2.

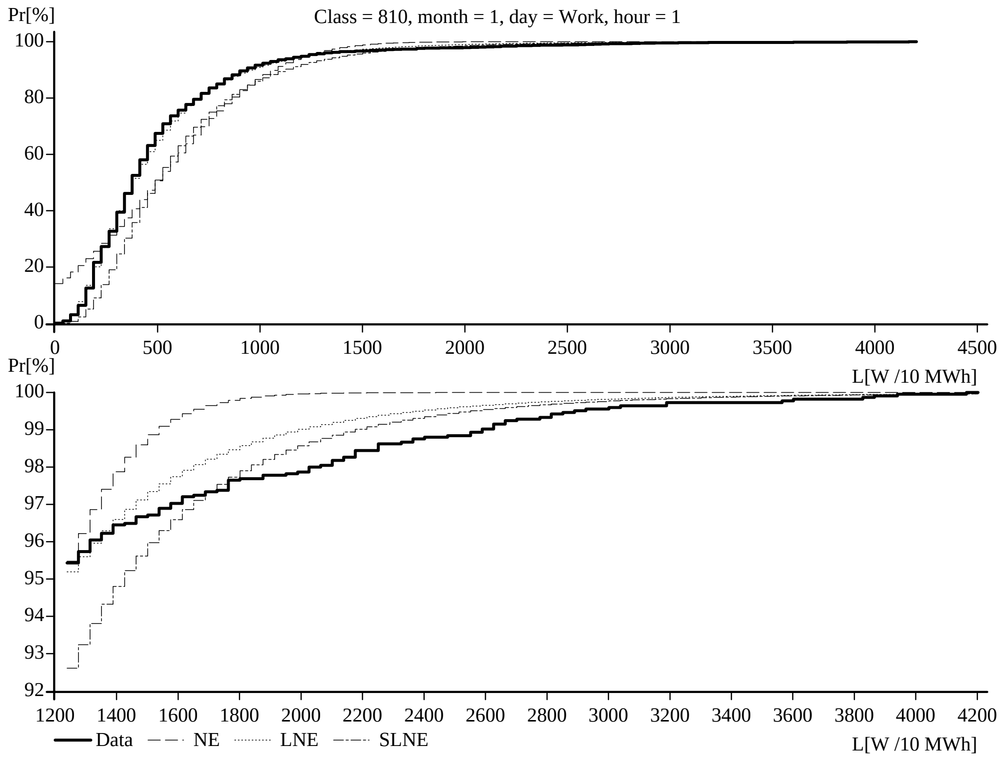


Fig. 31. Estimated distribution functions and distribution of the data. Class 810 Industry 1*-shift, January, working day, hour 00.00-01.00. Sample size is 2253.*

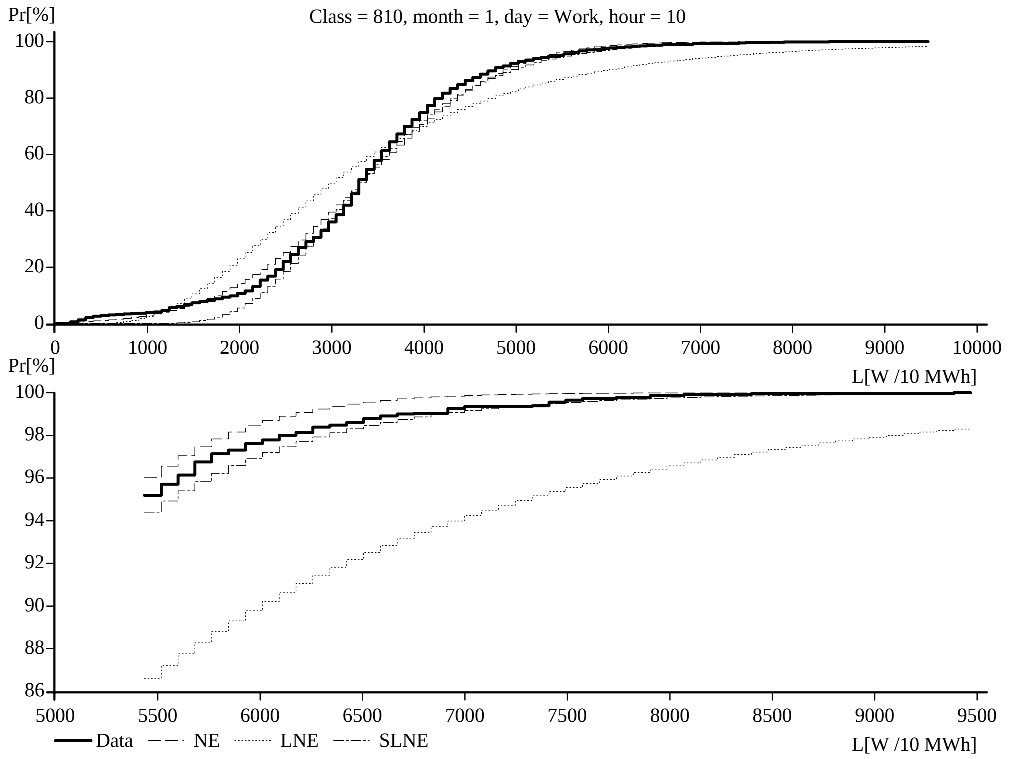


Fig. 32. Estimated distribution functions and distribution of the data. Class 810 Industry 1-shift, January, working day, hour 09.00-10.00. Sample size is 2311.

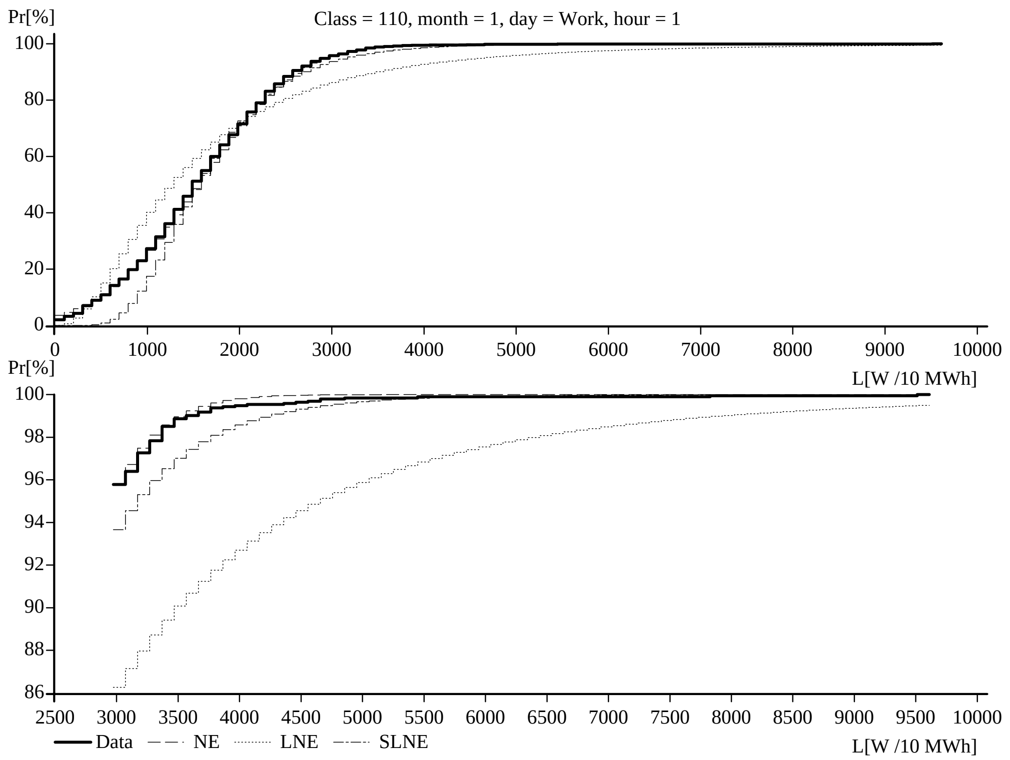


Fig. 33. Estimated distribution functions and distribution of the data. Class 110 direct electric heating, one family house, January, working day, hour 00.00-01.00. Sample size is 1944.

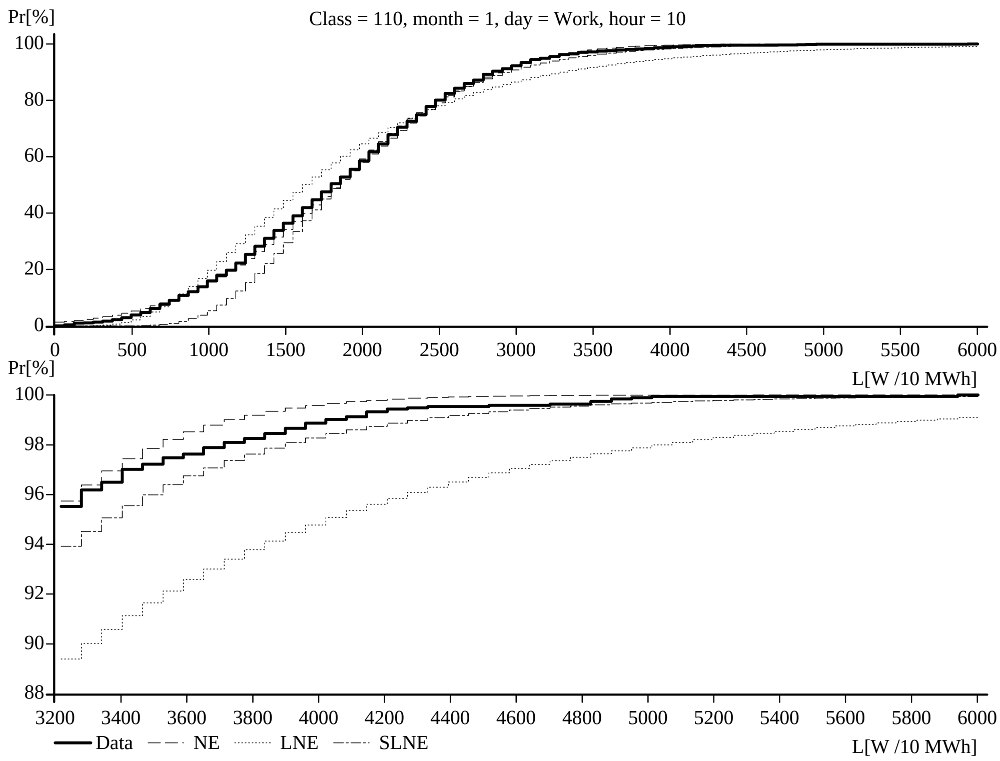


Fig. 34. Estimated distribution functions and distribution of the data. Class 110 direct electric heating, one family house, January, working day, hour 09.00-10.00. Sample size is 1943.

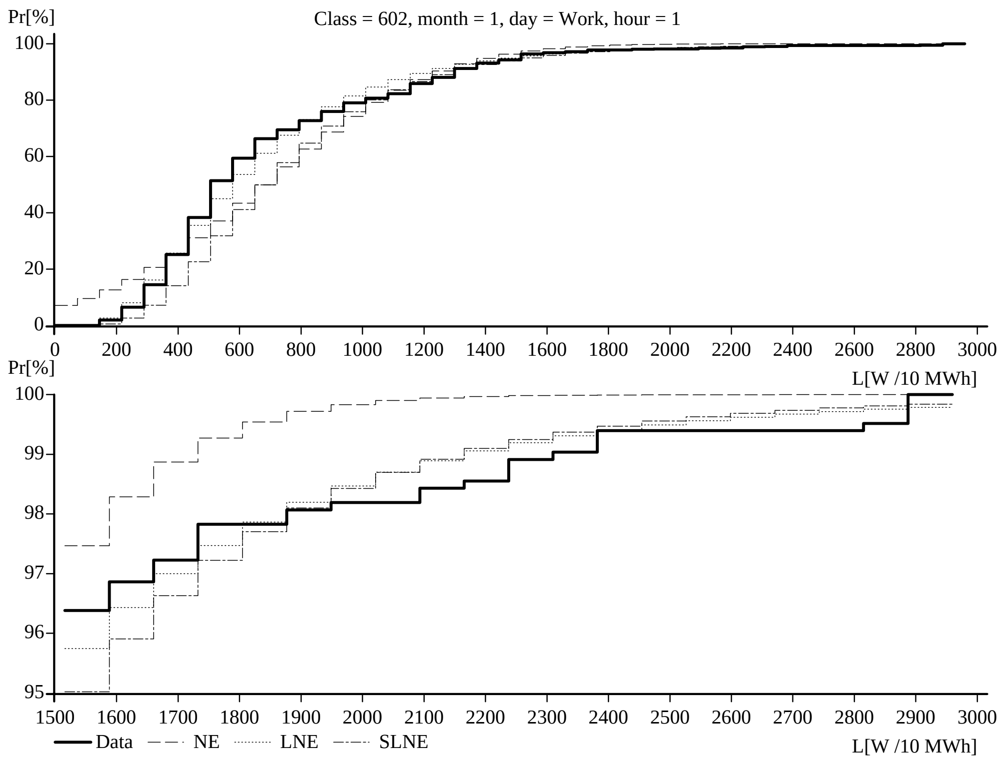


Fig. 35. Estimated distribution functions and distribution of the data. Class 602 residential, one family house, January, working day, hour 00.00-01.00. Sample size is 829.

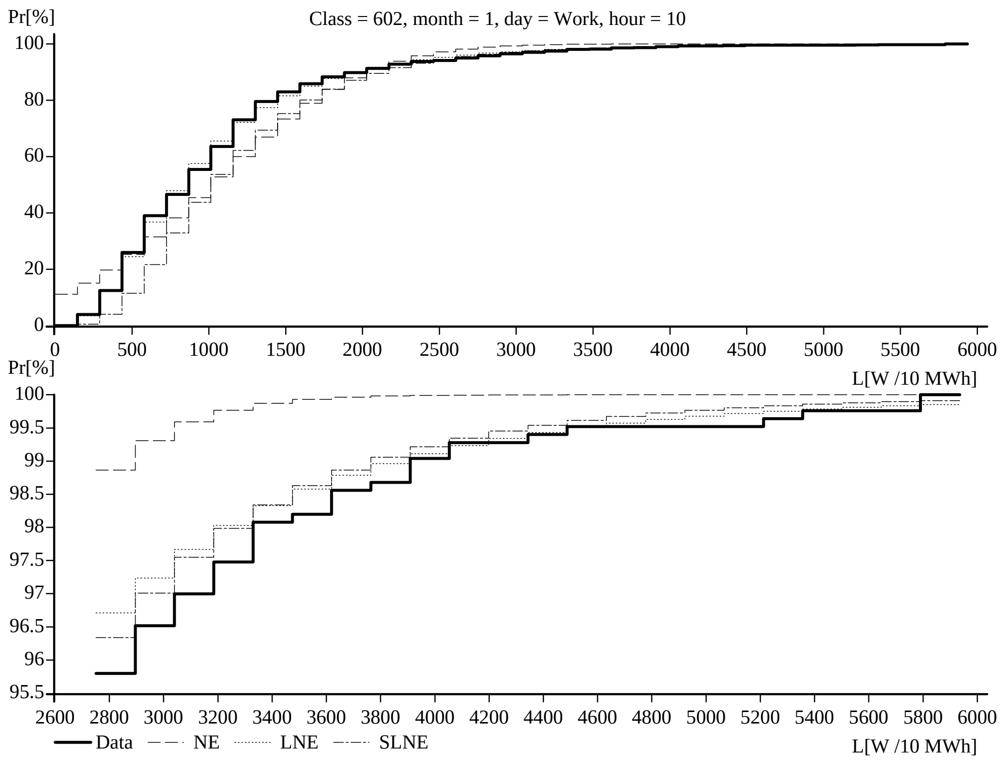


Fig. 36. Estimated distribution functions and distribution of the data. Class 602 residential, one family house, January, working day, hour 09.00-10.00. Sample size is 832.

5.4.3 Verification of confidence interval estimation

Tables 14 -19 present the estimation errors q_1 and q_2 of selected classes 810, 110 and 602, January workdays, hours 00.00 - 01.00 and 09.00 - 10.00. The estimation methods are described in chapter 5.3.

Table 14. The errors q_1 and q_2 of the confidence interval estimation. Class 810 Industry 1-shift, January, workday, hour 00.00 - 01.00.

Method	q_1[%]				q_2[%]			
α=	84.13%	95%	99%	99.5%	84.13%	95%	99%	99.5%
NE	6.28	0.21	-2.24	-2.30	20.12	0.90	-40.54	-42.79
LNE	0.20	0.30	-1.13	-0.74	0.14	1.46	-22.88	-18.00
LNEA	(0)	0.34	-1.00	-0.74	(0)	1.92	-22.05	-16.93
LNEB	-0.38	(0)	-1.13	-0.88	-1.14	(0)	-24.11	-19.35
SLNE	6.28	1.49	-0.55	-0.61	20.12	16.24	-15.78	-11.96

Table 15. The errors q_1 and q_2 of the confidence interval estimation. Class 810 Industry 1-shift, January, workday, hour 09.00 - 10.00.

Method	q_1[%]				q_2[%]			
α=	84.13%	95%	99%	99.5%	84.13%	95%	99%	99.5%
NE	3.06	-0.41	-0.90	-0.88	4.11	-1.42	-7.36	-12.53
LNE	9.51	4.39	1.00	0.50	17.97	34.68	57.60	61.52
LNEA	(0)	-0.97	-0.64	-0.50	(0)	-3.71	-5.71	-9.25
LNEB	1.94	(0)	-0.08	-0.15	2.31	(0)	-0.54	-3.75
SLNE	3.06	1.02	0.13	0.02	4.11	2.77	3.27	0.30

Table 16. The errors q_1 and q_2 of the confidence interval estimation. Class 110 direct electric heating, one family house, January, workday, hour 00.00 - 01.00.

Method	q_1[%]				q_2[%]			
α=	84.13%	95%	99%	99.5%	84.13%	95%	99%	99.5%
NE	1.26	0.47	-0.08	-0.32	2.55	0.92	-0.55	-7.11
LNE	9.49	4.79	0.95	0.50	19.19	58.43	121.67	135.87
LNEA	(0)	1.40	0.54	0.24	(0)	6.31	17.73	15.36
LNEB	-3.93	(0)	0.38	0.04	-3.64	(0)	7.99	4.87
SLNE	1.26	2.02	0.54	0.29	2.55	8.45	19.46	16.82

Table 17. The errors q_1 and q_2 of the confidence interval estimation. Class 110 direct electric heating, one family house, January, workday, hour 09.00 - 10.00.

Method	q_1[%]				q_2[%]			
α=	84.13%	95%	99%	99.5%	84.13%	95%	99%	99.5%
NE	1.00	-0.04	-0.90	-0.84	1.37	-0.26	-7.08	-8.38
LNE	5.32	4.02	0.95	0.50	7.99	26.30	45.14	54.96
LNEA	(0)	1.35	0.43	0.09	(0)	4.22	6.16	8.62
LNEB	-2.04	(0)	0.02	0.04	-2.47	(0)	0.14	1.85
SLNE	1.00	1.50	0.54	0.14	1.37	5.35	6.99	9.35

Table 18. The errors q_1 and q_2 of the confidence interval estimation. Class 602 residential, one family house, January, workday, hour 00.00 - 01.00.

Method	q_1[%]				q_2[%]			
α=	84.13%	95%	99%	99.5%	84.13%	95%	99%	99.5%
NE	8.97	0.40	-1.61	-1.49	33.31	10.08	-20.11	-21.87
LNE	2.22	-0.37	-1.45	-1.19	3.87	-5.38	-19.29	-15.58
LNEA	(0)	-0.98	-2.07	-1.19	(0)	-9.98	-24.16	-21.03
LNEB	2.99	(0)	-0.69	-0.42	6.58	(0)	-12.02	-6.98
SLNE	8.97	1.47	0.69	0.19	33.31	26.90	13.34	20.47

Table 19. The errors q_1 and q_2 of the confidence interval estimation. Class 602 residential, one family house, January, workday, hour 09.00 - 10.00.

Method	q_1[%]				q_2[%]			
α=	84.13%	95%	99%	99.5%	84.13%	95%	99%	99.5%
NE	3.70	-1.78	-1.62	-1.66	6.37	-4.50	-17.90	-26.14
LNE	3.39	1.15	0.23	-0.27	4.39	3.57	1.75	-3.52
LNEA	(0)	-1.93	-1.00	-0.89	(0)	-5.51	-11.78	-17.84
LNEB	2.77	(0)	-0.39	-0.27	3.49	(0)	-4.43	-10.26
SLNE	3.70	0.53	-0.39	-0.27	6.37	0.82	-5.56	-11.96

From Tables 14 - 19 we see how the error of estimation varies from hour to hour and in general the simplified lognormal estimation method SLNE results in less errors than the normal distribution method NE. The LNE method, which is based directly on lognormal distribution, results in very high errors as seen in Tables 15 -17.

Table 20 summarises the analysis where the different methods of estimating confidence intervals were compared to load research data from the Finnish load research project. The results in Table 20 are the average values from an analysis of 9 customer classes over 12 months. For each hour there are, on average, 800 observations.

Table 20. Average errors of confidence interval estimation. Average result of 9 customer categories, 12 months, working days and 24 hours.

Method	q_1[%]				q_2[%]			
α=	84.13%	95%	99%	99.5%	84.13%	95%	99%	99.5%
NE	3.15	-1.08	-1.94	-1.82	10.91	-4.82	-20.97	-25.63
LNE	2.76	1.44	0.21	0.07	5.55	13.38	26.53	33.87
LNEA	(0)	-0.05	-0.37	-0.35	(0)	-0.08	2.73	5.86
LNEB	-0.02	(0)	-0.04	-0.10	0.87	(0)	1.92	4.54
SLNE	3.15	0.63	-0.03	-0.07	10.91	4.92	0.11	-0.16

The conclusion is that applying normal distribution approximation when estimating 99.5 % confidence interval for one customer leads, on an average, to -25.63 % error. If we use the simplified lognormal approximation instead we get an average error of -0.16 %. According to this result the simplified lognormal approximation results in more accurate estimates than the normal approximation in the case of one customer.

In further analysis we concentrate on the NE and SLNE methods estimating the 99.5 % confidence interval.

5.4.4 Verification of 99.5 % confidence interval estimation

Now we concentrate on the 99.5 % confidence level which is the highest resonable level with the sample size around 1000. With a sample size of 1000 the 99.5 % interval means that 5 observations are expected to be above the confidence interval. The possible few errors in data will then not affect the result too much.

In Figs. 37 - 39 the estimation error q_2 for the 99.5 % percentile is plotted for January and July, workdays varying the hour of day, and for hours 00.00 - 01.00 and 09.00 - 10.00 varying the month. From these figures we see how the error of SLNE is in general smaller and the NE method in general estimates the confidence interval too low (q_2 is negative).

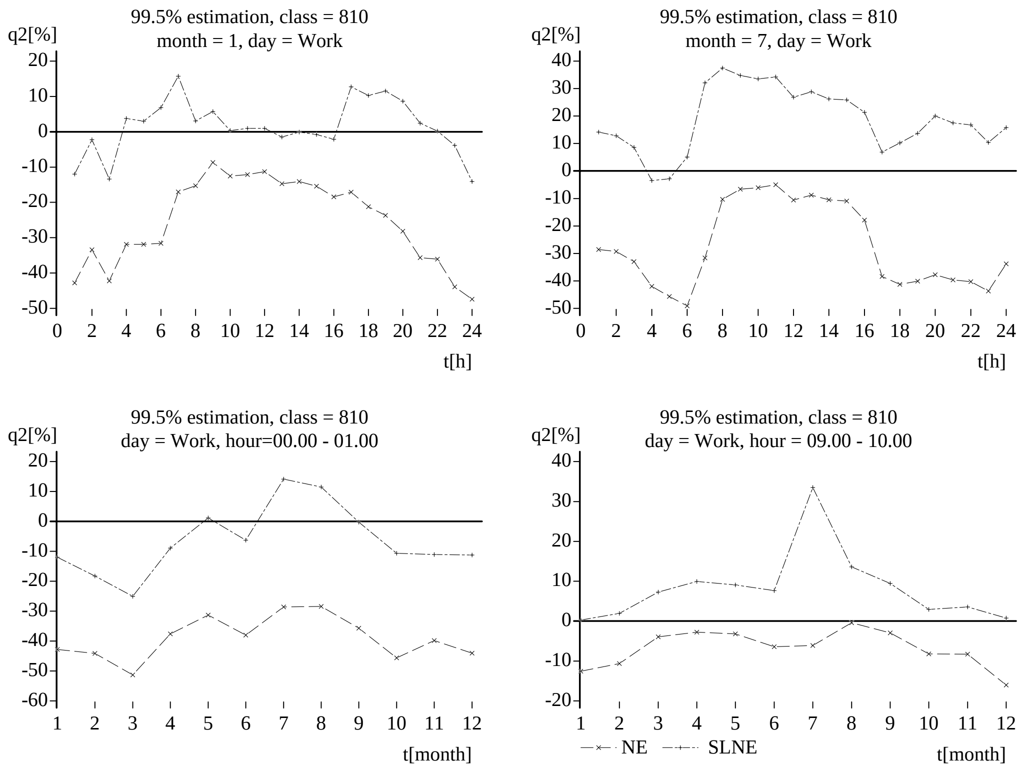


Fig. 37. Error q_2 of confidence interval estimation. Class 810 Industry 1-*shift. Above: the error for 24 hours of a workday in January and July. Below: the error for 12 months at hours 0.00 - 01.00 and 09.00 - 10.00.*

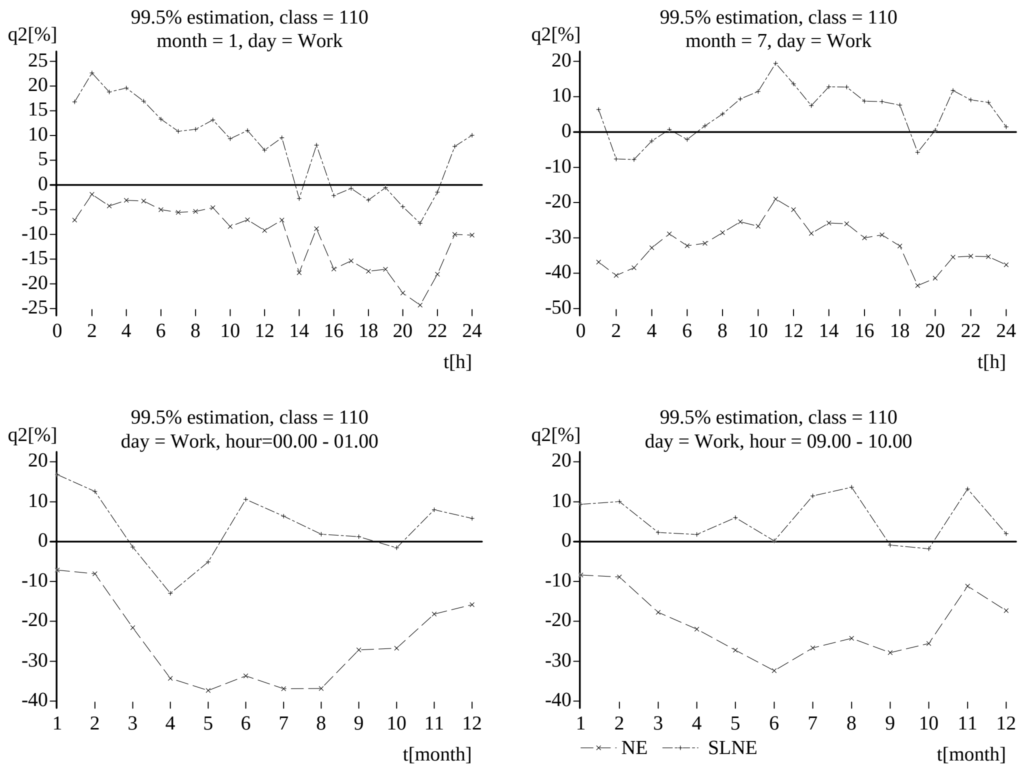


Fig. 38. Error q_2 of confidence interval estimation. Class 110 direct electric heating, one family house. Above: the error for 24 hours of a workday in January and July. Below: the error for 12 months at hours 0.00 - 01.00 and 09.00 - 10.00.

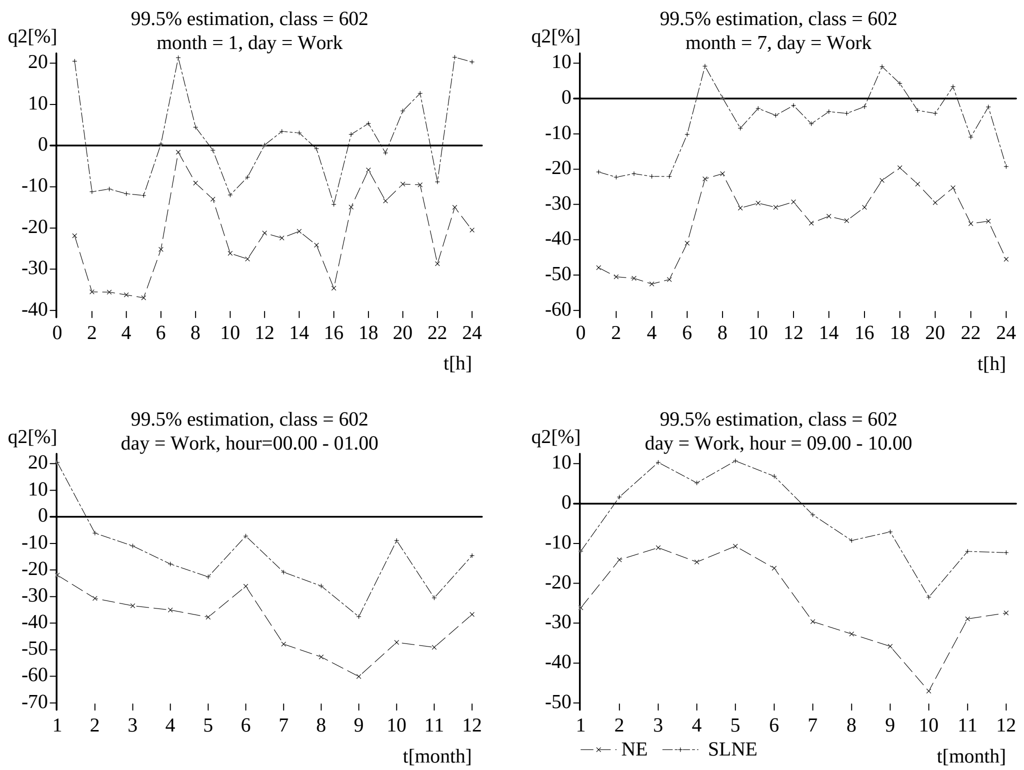


Fig. 39. Error q_2 of confidence interval estimation. Class 602 residential, one family house. Above: the error for 24 hours of a workday in January and July. Below: the error for 12 months at hours 0.00 - 01.00 and 09.00 - 10.00.

In Fig. 40 the average estimation error q_2 of the 99.5 % confidence interval is presented for each class for January and July, hours 00.00 - 01.00 and 09.00 - 10.00.

In Fig. 41 the average error q_2 over 24 hours of a day in January and July and average error over 12 months at hours 00.00 - 01.00 and 09.00 - 10.00 is presented.

From Fig. 40 and Fig. 41 we see the error of SLNE is smaller than the error when using NE method. Also from the figures it can be noted that the NE method gives too low estimates while the error q_2 is systematically negative.

In Fig. 42 the average errors for winter work days (1.11. - 31.3., hours 07.00 - 22.00), winter work days night (1.11. - 31.3., hours 22.00 - 07.00), summer work days (1.4. - 31.10, hours 07.00 - 22.00) and summer workdays night (1.4. - 31.10., hours 22.00 - 07.00). Again the error of SLNE estimation is smaller.

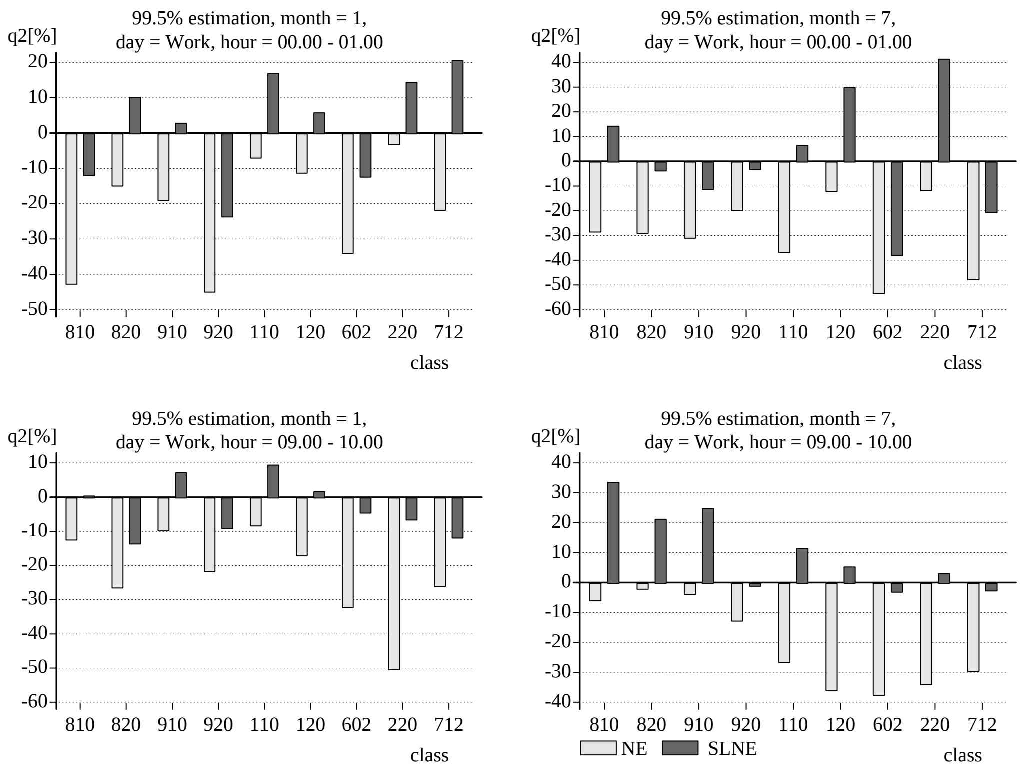


Fig. 40. Error q_2 of confidence interval estimation comparing different classes months and hours.

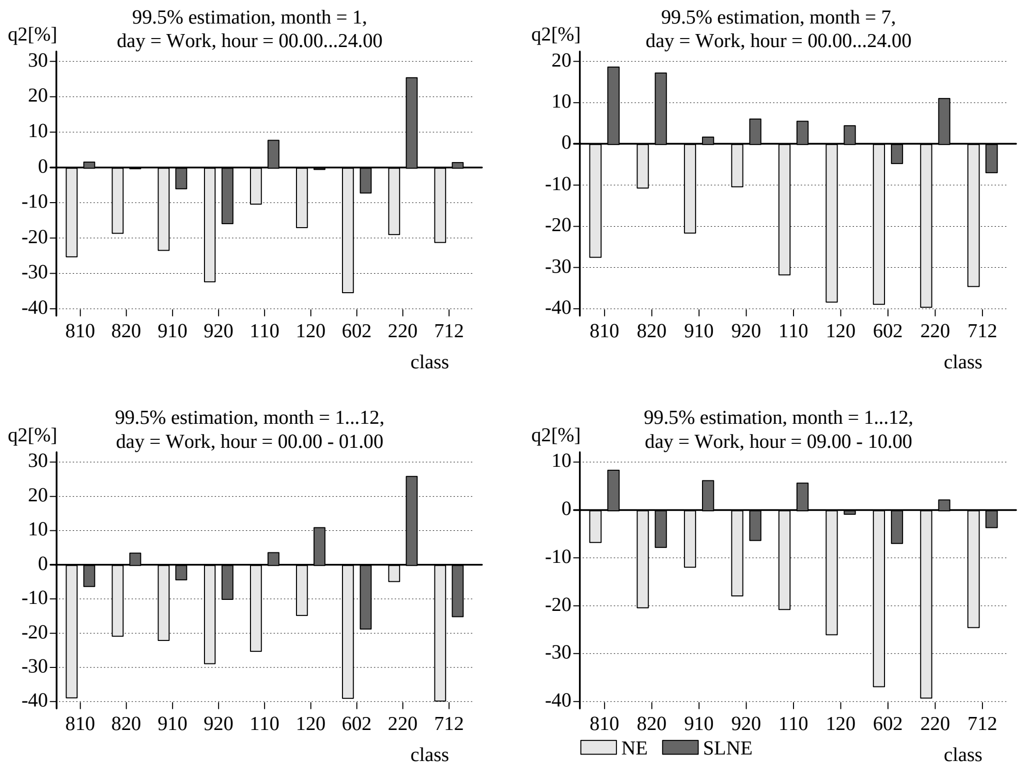


Fig. 41. Error q_2 of confidence interval estimation. Average over 24 hours in selected months and average over 12 months of selected hours.

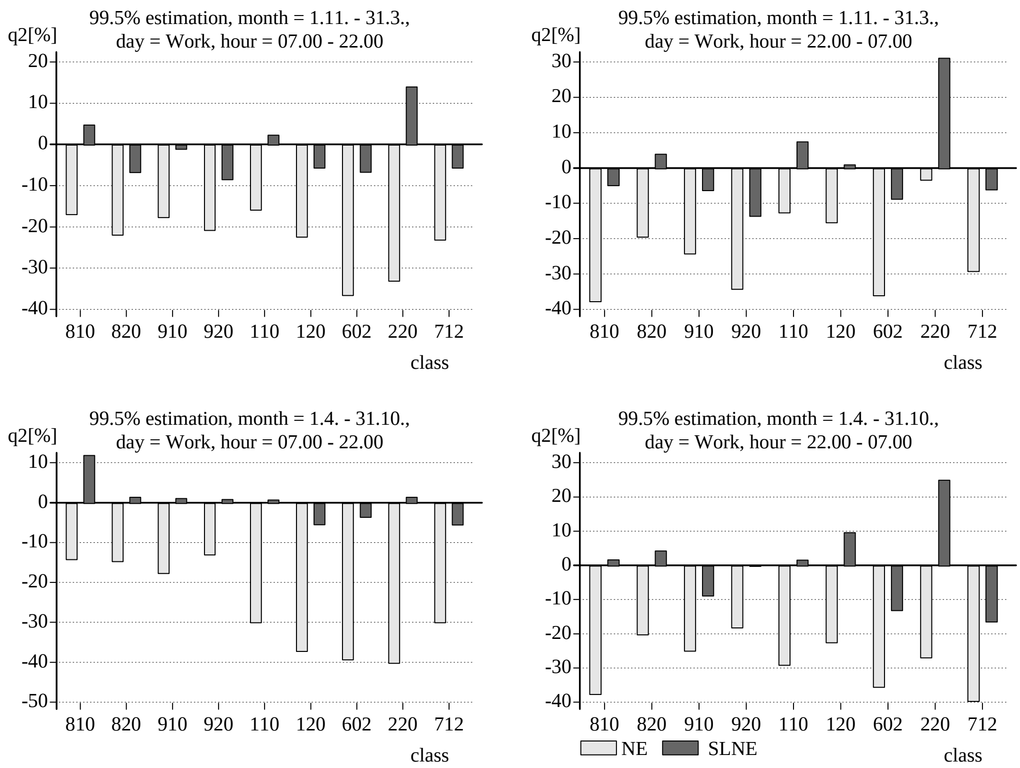


Fig. 42. Error q_2 of confidence interval estimation. Average over winter and summer day and night.

In Fig. 43 the average error over a year is presented. Except for class 220 the error of SLNE is small compared to NE estimation. Here we can see that the result in Table 20, SLNE error -0.16 % for α = 99.5 % percentile gives a bit wrong impression of the accuracy of the SLNE because the error in class 220 has a positive sign compensating for the overall other negative errors of other classes. Still this result proves the smaller error of the SLNE method when applied to all classes and all times.

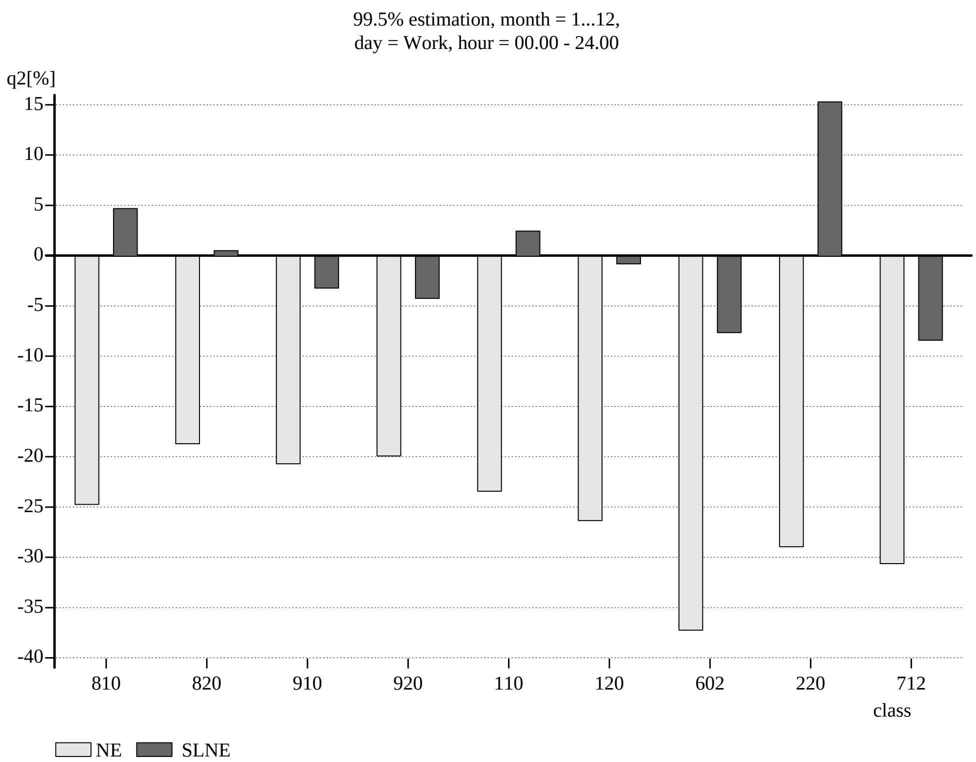


Fig. 43. Error q_2 of confidence interval estimation. Average over year and all hours of day.

5.4.5 Verification of confidence interval estimation of customer's maximum load

The final task is to study how the overall maximum of a customer load could be estimated. This is the most common engineering problem which occurs when the lines and other equipment near the customer are considered. The customer's maximum load is an important factor which determines how much transmission capacity should be available to gain a desired performance.

The results are shown in Table 21 where the confidence of interval estimation is shown for the month and hour of the highest 99.5% value for every class. In addition to the estimation results, the values of the parameters m_1 (mean) and s_1 (standard deviation) are shown.

From the table we also observe the Finnish peculiarity, the electric sauna, in class 602. The electric sauna is a high power appliance in many Finnish homes, usually 6 kW. While the figures in Table 21 are given for an annual energy of 10 MWh and the usual annual energy for a one family house without electric heating is 3 - 5 MWh per year, the power estimated using $L_{99.5\%}$ is reduced to 4 - 7 kW.

Table 21. Performance of NE and SLNE estimators when the 99.5 % confidence interval is highest. Unit of m_1 , s_1, $L_{99.5\%}$, NE and SLNE is [W/10 MWh/a].

Class	month	hour	m_1	s_1	$L_{99.5\%}$	NE	q_2[%]	SLNE	q_2[%]
810	12	9	3110	1449	8578	6835	-21	8313	-4
820	2	10	2124	939	6264	4538	-28	5444	-14
910	1	17	1813	905	7216	4140	-43	5134	-29
920	1	17	1994	692	5294	3774	-29	4289	-19
110	12	20	2137	1096	6688	4954	-26	6193	-8
120	12	23	2732	1108	7921	5581	-30	6554	-18
602	3	20	2803	2839	14873	10099	-33	16919	13
220	1	1	4094	1705	8763	8478	-4	10020	14
712	10	20	1767	1187	8859	4817	-46	6619	-26

5.5 ESTIMATING CONFIDENCE INTERVALS OF THE DATA FROM THE SIMULATION

An example of a customer's load variation simulation is presented here. The customer appliance data is the same as in the simulation presented in chapter 4.5.7.

Three simulations are presented in Fig. 44 - Fig.46. The estimated normal (NE), lognormal (LNE) and simplified lognormal (SLNE) distributions are also drawn. Also with this simulated data the simplified lognormal curve fits best.

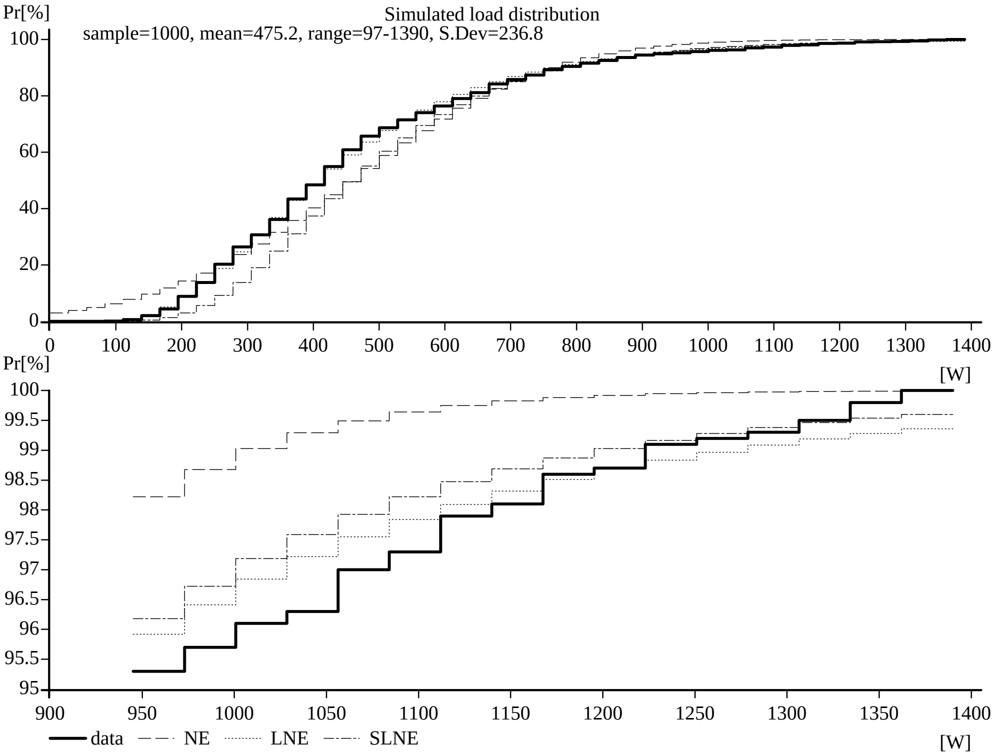


Fig. 44. Simulated distribution and normal, lognormal and simplified lognormal distribution functions. Expected load m_1 = 475 W and s_1 = 237 W.

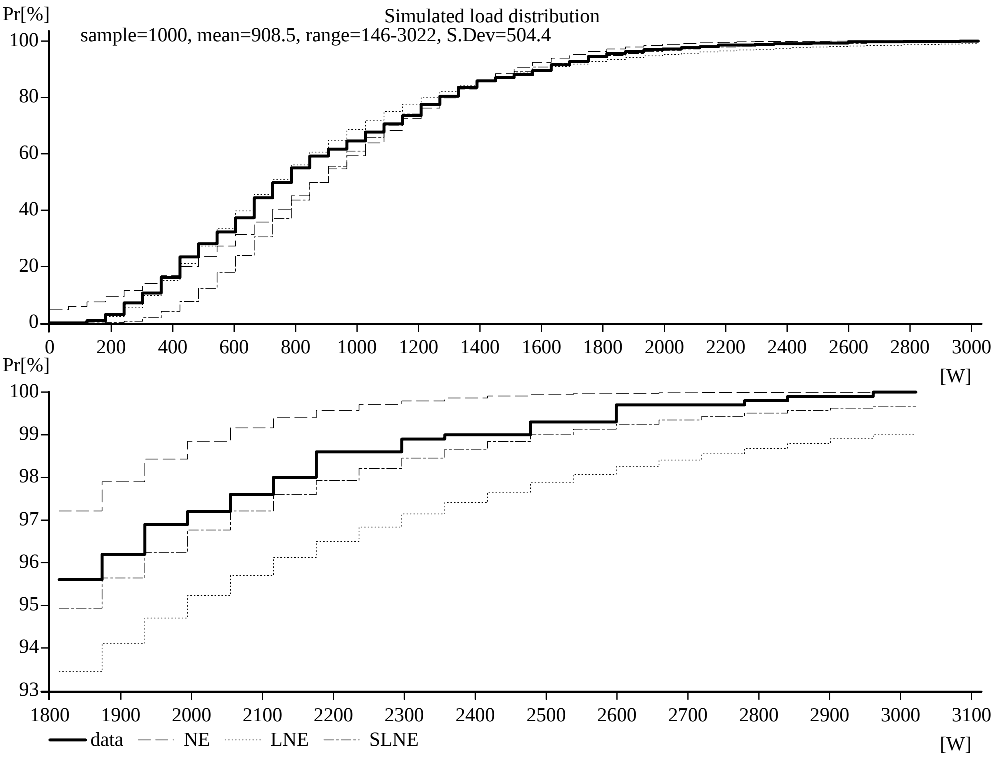


Fig. 45. Simulated distribution and normal, lognormal and simplified lognormal distribution functions. Expected load m_1 = 909 W and s_1 = 504 W.

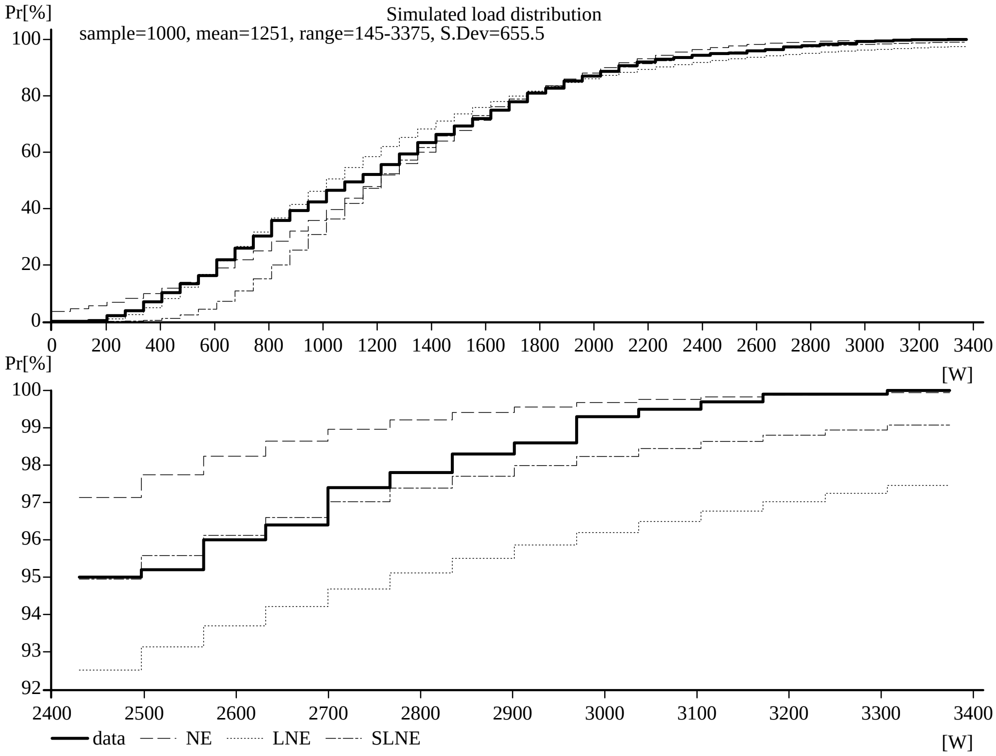


Fig. 46. Simulated distribution and normal, lognormal and simplified lognormal distribution functions. Expected load m_1 = 1251 W and s_1 = 656 W.

5.6 APPLICATION OF THE CONFIDENCE INTERVAL ESTIMATORS TO PRACTICAL DISTRIBUTION COMPUTATION

The previous results can be applied in practice to estimate the confidence intervals for one customer. Substituting the parameters of simplified lognormal approximation (45) to the equation of the estimator (46) we get

$$\hat{L}_{\alpha 4} = m_1 \left(1 + \frac{s_1}{m_1}\right)^{U_\alpha} \qquad (55)$$

recalling again from the definition of the model (35) that $L = m_1$ and $s_L = s_1$ we can now write the formula of the estimator using the parameters of the load model

$$\hat{L}_{\alpha 4} = L \left(1 + \frac{s_L}{L}\right)^{U_\alpha} \qquad (56)$$

Applying the load model for simplified lognormal approximation we get

$$\hat{P}_{\alpha 4} = \hat{L}_{\alpha 4} W_a = L\left(1+\frac{s_L}{L}\right)^{U_\alpha} W_a = \overline{P}\left(1+\frac{s_L}{L}\right)^{U_\alpha} \qquad (57)$$

and for normal distribution approximation we get

$$\hat{P}_{\alpha 1} = \hat{L}_{\alpha 1} W_a = L\left(1+U_\alpha \frac{s_L}{L}\right) W_a = \overline{P}\left(1+U_\alpha \frac{s_L}{L}\right) \qquad (58)$$

Example: A load model L_c gives, for a customer with a given annual energy W_a, one certain hour's load mean value $\overline{P}$ = 100 kW and the standard deviation s_P = 50 kW.

With the NE method we estimate the 95 % and 99.5 % confidence interval:

$$P_{95\%,1} = 100 \cdot (1 + 1.65 \cdot 0.5) = 182.50 \text{ kW}$$
$$P_{99.5\%,1} = 100 \cdot (1 + 2.58 \cdot 0.5) = 229.50 \text{ kW}$$

and with the SLNE method we get

$$P_{95\%,4} = 100 \cdot (1 + 0.5)^{1.65} = 195.23 \text{ kW}$$
$$P_{99.5\%,4} = 100 \cdot (1 + 0.5)^{2.58} = 284.65 \text{ kW}$$

If we have load values $\overline{P}$ = 100 kW and s_P = 75 kW with the NE method we estimate

$$P_{95\%,1} = 100 \cdot (1 + 1.65 \cdot 0.75) = 223.75 \text{ kW}$$
$$P_{99.5\%,1} = 100 \cdot (1 + 2.58 \cdot 0.75) = 293.50 \text{ kW}$$

and with SLNE we get

$$P_{95\%,4} = 100 \cdot (1 + 0.75)^{1.65} = 251.78 \text{ kW}$$
$$P_{99.5\%,4} = 100 \cdot (1 + 0.75)^{2.58} = 423.68 \text{ kW}$$

The SLNE method is important in the sense that there is no need for parameters other than the ordinary mean and standard deviation. These parameters are already available for utilities' applications and therefore their use will not require any activity other than a change in the computation algorithm.

6 ESTIMATION OF CONFIDENCE INTERVALS OF SEVERAL CUSTOMERS

6.1 GENERAL

The result of the previous chapter was a method to estimate confidence intervals for one customer load. When the load consists of more than one customer the estimators are not directly applicable.

According to the Law of Great Numbers, the distribution of the sum of the loads becomes normal when the number becomes large. A complete study of the problem requires analysis of the correlation between customer classes and will be left for further studies.

Here we assume the loads are from the same class and are independent. First we study how the estimators of the confidence intervals should be modified when the number of customers increases to 2, 3, 4 etc. Then we apply the method to load research data.

6.2 DEVELOPMENT OF THE ESTIMATION METHODS FOR SEVERAL CUSTOMERS

6.2.1 The parameters of the sum of random variables

According to probability theory, the mean of the sum of random variables is

$$\mu_s = \mu_1 + \mu_2 + \ldots + \mu_k \qquad (59)$$

Also if the variables are independent, the sum of the variances is

$$\sigma_s^2 = \sigma_1^2 + \sigma_2^2 + \ldots + \sigma_k^2 \qquad (60)$$

Assuming each variables' mean and standard deviation is equal we get

$$\mu_s = k\mu \qquad (61)$$

$$\sigma_s^2 = k\sigma^2 \Rightarrow \sigma_s = \sqrt{k}\sigma \qquad (62)$$

According to the simple load model in equation (4) the standard deviation s is a linear function of the customers annual energy consumption. Assuming that $W_{a,i}$ is the annual energy of customer i we calculate for k customers the one customer's standard deviation with the average annual energy

$$s_{\text{av. one customer}} = s_1 \frac{\sum_{i=1}^{k} W_{a,i}}{k} \tag{63}$$

We get the standard deviation of k customers by substituting (63) to the previous equation (62)

$$s_{k\ \text{customers}} = \sqrt{k}\, s_{\text{av. one customer}} = \sqrt{k}\, s_1 \frac{\sum_{i=1}^{k} W_{a,i}}{k} = \frac{s_1}{\sqrt{k}} \sum_{i=1}^{k} W_{a,i} \tag{64}$$

According to this result we shall now test the estimation of confidence interval of loads of more than one customer in a class by dividing the standard deviation by the square root of the number of customers.

6.2.2 Normal distribution confidence interval estimation NE for several customers

The parameters for normal distribution m_1 and s_1 are estimated from the data of one customer. The estimators for percentiles $L_{\alpha k\%}$ according to the result of chapter 6.2.1 are selected from respective percentiles of unit normal distributions U_α. See Table 13.

$$\begin{cases} m_{1k} = m_1 \\ s_{1k} = \dfrac{s_1}{\sqrt{k}} \end{cases} \tag{65}$$

$$\hat{L}_{\alpha 1k} = m_{1k} + U_\alpha s_{1k} = m_1 + U_\alpha \frac{s_1}{\sqrt{k}} \tag{66}$$

6.2.3 Simplified lognormal distribution confidence interval estimation SLNE for several customers

The parameters of lognormal distribution function are derived from m_{1k} and s_{1k} similarly:

$$\begin{cases} m_{4k} = \ln m_{1k} \\ s_{4k} = \ln(m_1 + s_{1k}) - \ln m_{1k} \end{cases} \qquad (67)$$

$$\hat{L}_{\alpha 4k} = m_1 \left(1 + \frac{s_1}{m_1 \sqrt{k}}\right)^{U_\alpha} \qquad (68)$$

The procedure for estimation computation is shown in Fig.47.

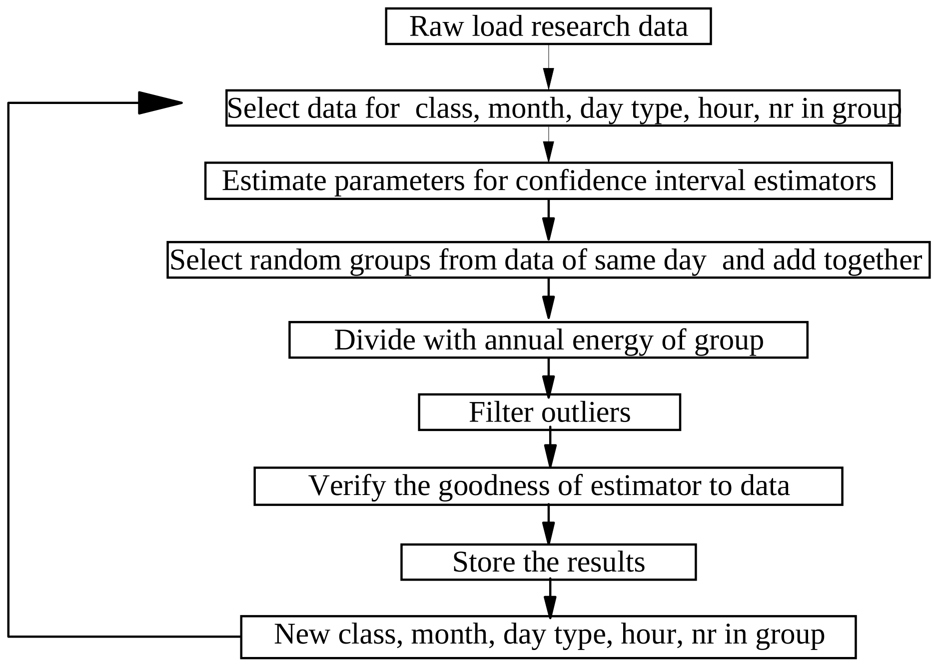


Fig. 47. Flow of the estimation and verification process for groups of customers.

6.3 VERIFICATION OF THE ESTIMATION OF SEVERAL CUSTOMER'S LOADS

6.3.1 Verification of 99.5 % confidence interval estimation

The following Figs. 48 - 52 present the results of estimation of 99.5 % confidence interval when the number of customers *k* is 1, 2, 3, 4, 6 and 8.

The data representing sums of several loads is calculated by selecting random combinations from the data

$$L_n = \frac{W_h(m, d, h)_1 + W_h(m, d, h)_2 + \dots}{W_{a1} + W_{a2} + \dots} \qquad (69)$$

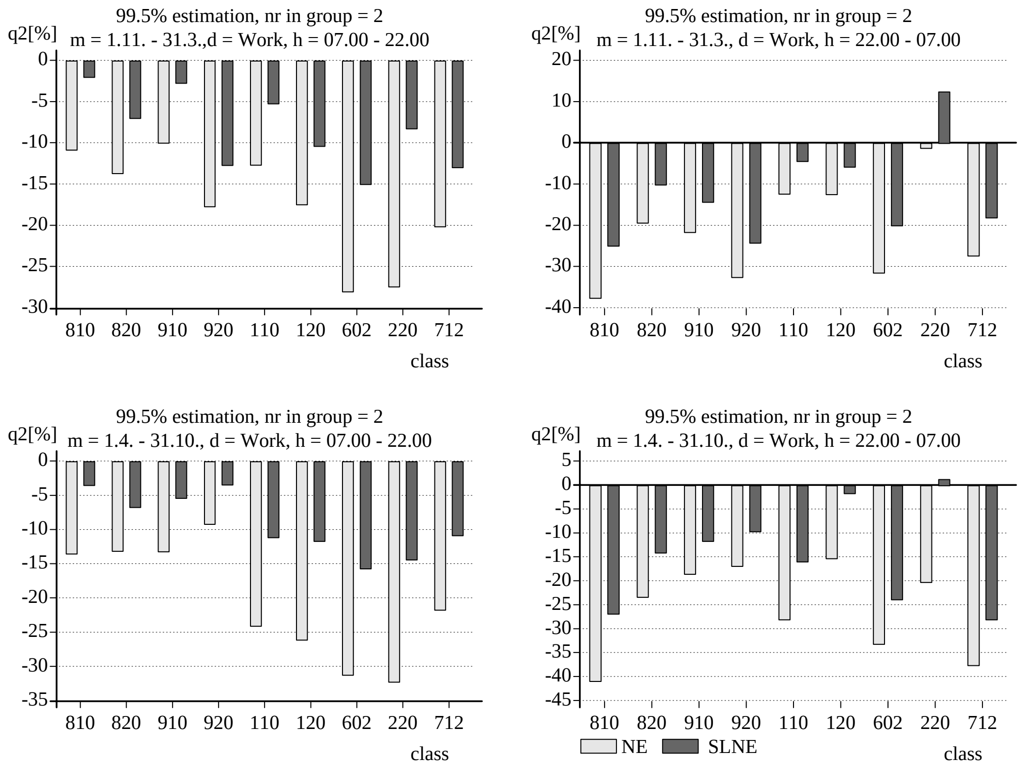


Fig. 48. Estimation of the 99.5 % confidence interval for 2 customers.

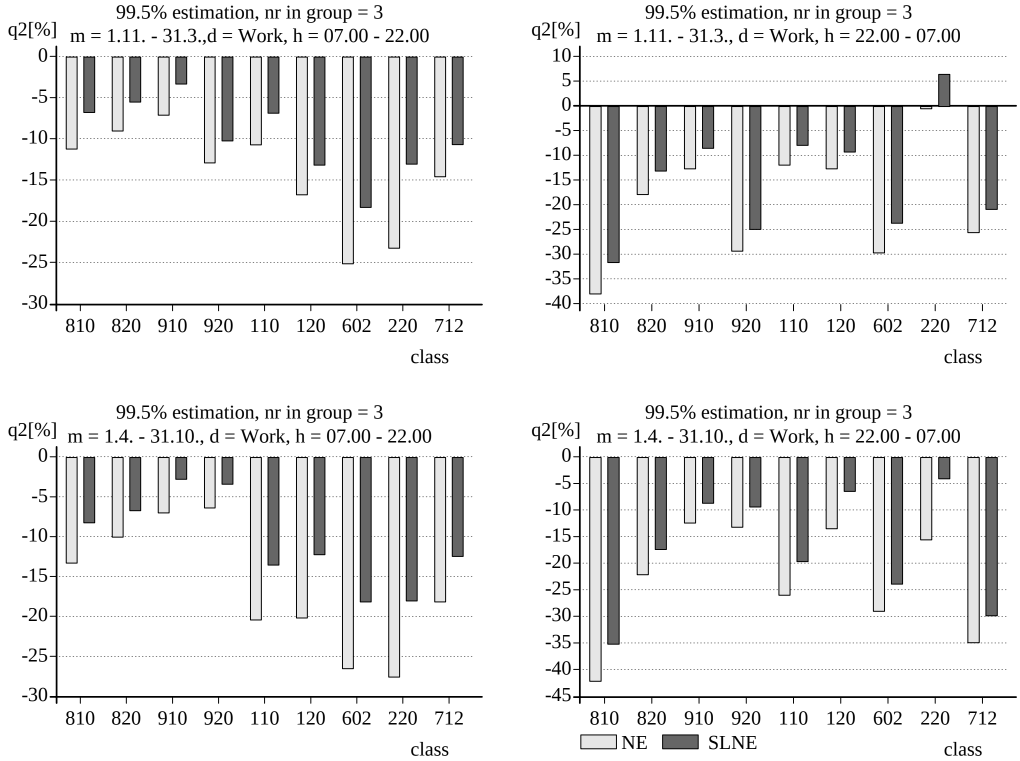


Fig. 49. Estimation of the 99.5 % confidence interval for 3 customers.

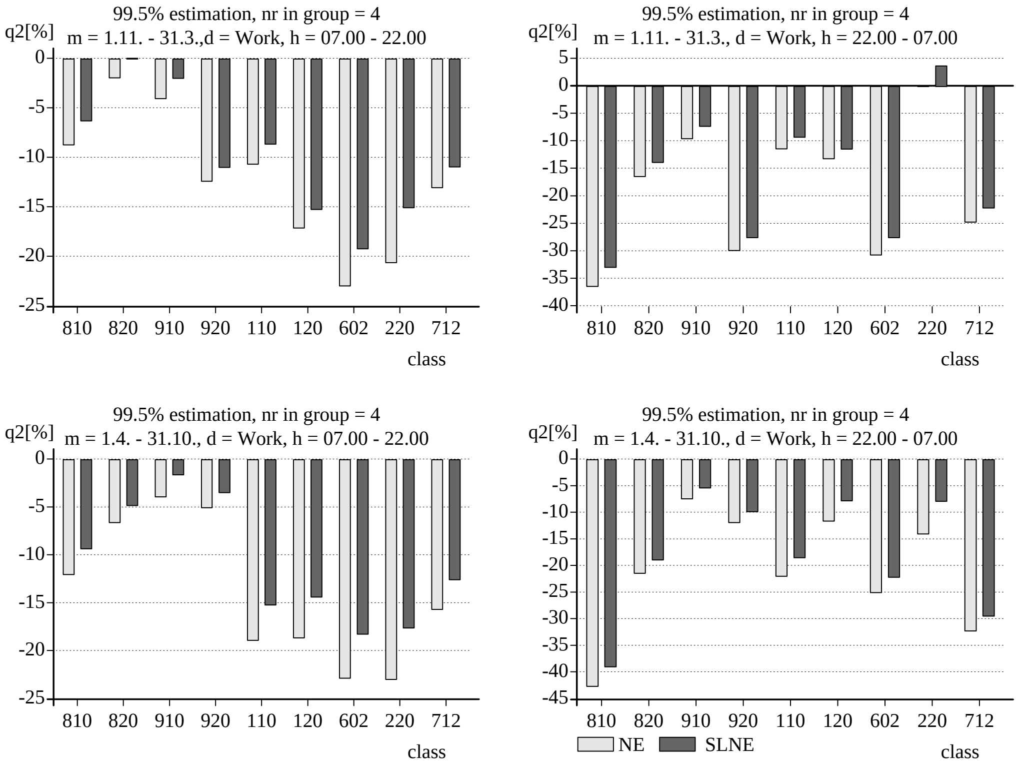


Fig. 50. Estimation of the 99.5 % confidence interval for 4 customers.

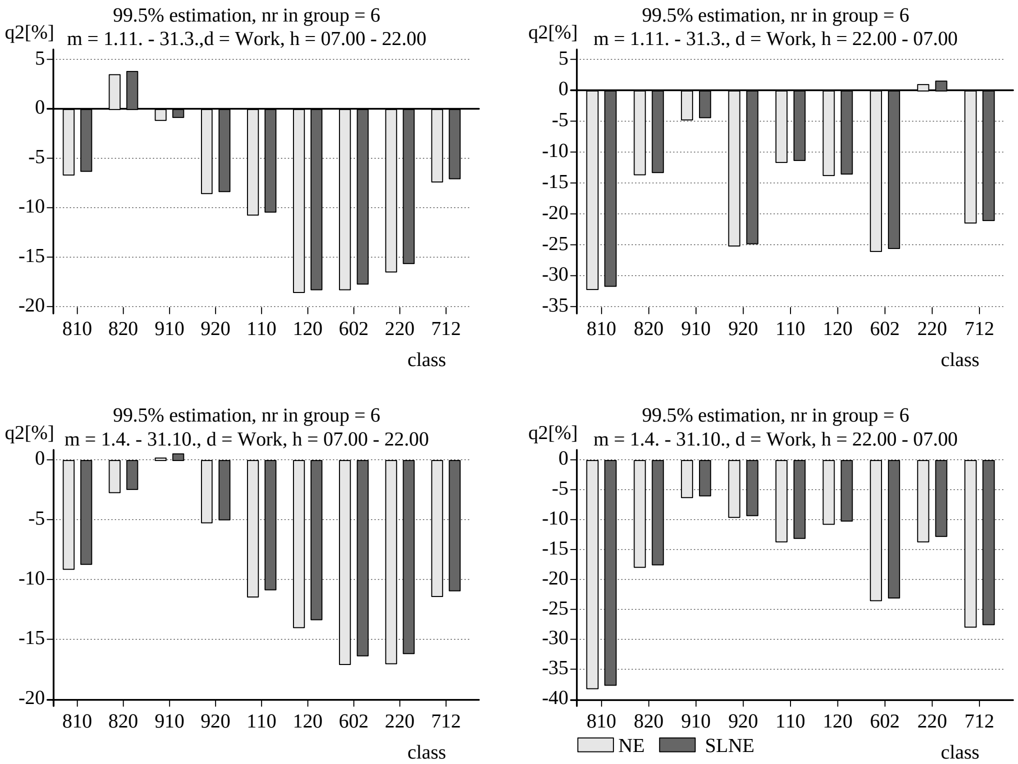


Fig. 51. Estimation of the 99.5 % confidence interval for 6 customers.

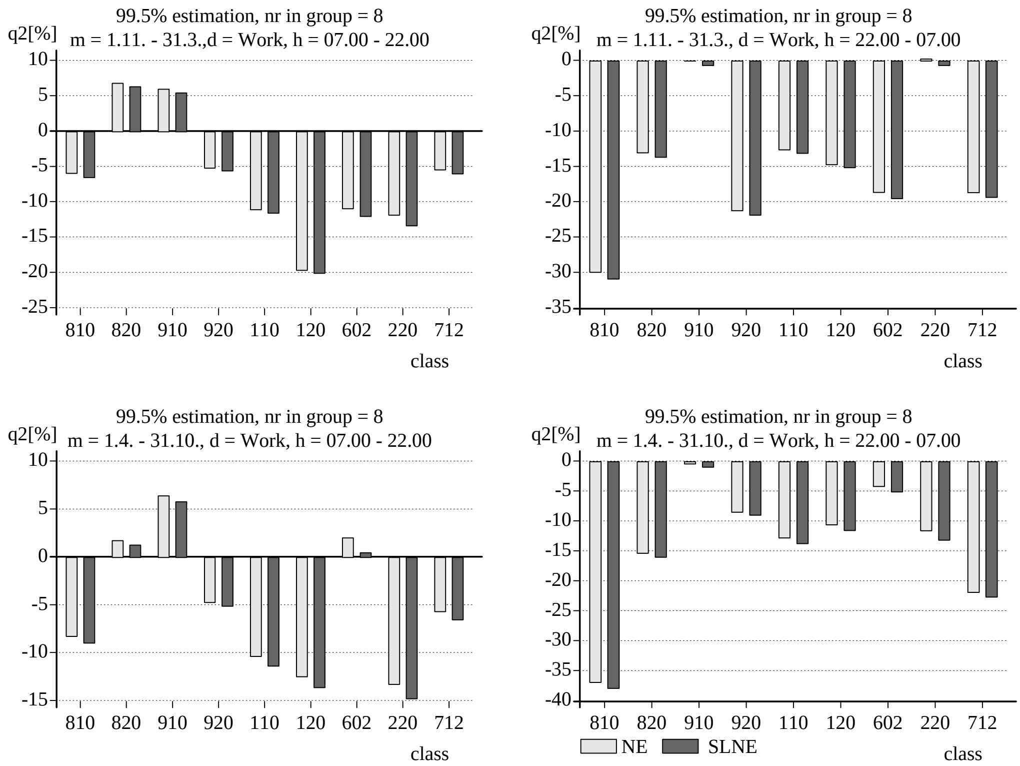


Fig. 52. Estimation of the 99.5 % confidence interval for 8 customers.

The figures clearly show that the assumption of independent loads is wrong. The estimates are, overall, below the observed confidence intervals. However the NE method for several customers gives almost the same results as the SLNE method. The overall error q_2 of the estimation becomes smaller as the number of customers k grows. When the number of customers is 8 (Fig. 52) the estimation error is about -10 % except class 120. Obviously the assumptions of the estimators are not fulfilled, but the accuracy is still in a range sufficient for most distribution applications.

The other conclusion is that only in the case of one or two customers is the SLNE estimation clearly better than NE. In the case of three and more customers the two estimation methods give similar results. The value of the error q_2 remains, in winter days, around -10 % when the number of customers grows. However, such an error (ranging from -10 % to +5 % estimating the 99.5 % confidence interval) is quite acceptable for most distribution applications.

One source for inaccuracy of this analysis is the random selection algorithm to form groups of 2, 3, 4,… customers. While the number of the available load data values varies, the random selection has different numbers of values to select. The work to improve this method will be left for further studies.

6.3.2 Verification of estimation of several customer's maximum loads

Here the result is presented for selected customer classes: class 810 Industry 1-shift, class 120 electric heating with storage water boilers and class 602 homes with electric sauna. The results of the other customer classes will be analysed in further studies.

Class 810 Industry is shown in Table 22. The estimation result is good and there is no notable difference between NE and SLNE when the number of customers is more than 1.

We also see some odd things. The time position of highest load changes as the number of customers grows, and the parameter values sometimes increase when the number of customers increases while a decrease in the parameter values would be expected. The main reason for these is obviously that the data values are the result of random selection forming the groups of 2, 3, etc. customers.

Table 22. The model parameters at the time of maximum load of Class 810 Industry 1-shift with different number of customers. Number of customers 1…8. Unit of m_1, s_1, $L_{99.5\%}$, NE and SLNE is [W/10 MWh/a].

Nr.	Month	hour	m_1	s_1	$L_{99.5\%}$	NE	q_2[%]	SLNE	q_2[%]
1	12	9	3110	1449	8578	6835	-21	8313	-4
2	12	15	2761	1124	6218	5590	-11	6114	-2
3	12	11	3021	928	5348	5295	-1	5501	2
4	1	9	3229	748	5018	5012	-1	5094	1
5	1	9	3226	694	4893	4838	-2	4876	-1
6	1	11	3330	568	4785	4800	0	4811	0
7	1	9	3202	629	4536	4610	1	4604	1
8	1	9	3202	597	4489	4530	0	4512	0

The following example is the class 120 Residential, direct electric heating shown in Table 23. In class 120 the electric water boilers are automatically switched on in the evening. Here we see how the $L_{99.5\%}$ does not decrease when the number of customers increases because the boilers are always switched on same time. The values when the number of customers is 3 and 8 are exceptional and likely because of errors in the random selection algorithm.

Table 23. The model parameters at the time of maximum load of Class 120 Residential, direct electric heat, one family house, water boiler at night. Number of customers 1…*8. Unit of* m_1 , s_1, $L_{99.5\%}$ *, NE and SLNE is [W/10 MWh/a].*

Nr.	Month	hour	m_1	s_1	$L_{99.5\%}$	NE	q_2[%]	SLNE	q_2[%]
1	12	23	2732	1108	7921	5581	-30	6554	-18
2	2	23	2795	837	5290	4888	-8	5213	-2
3	12	8	1605	579	5827	2532	-57	2619	-56
4	2	23	2766	638	4524	4289	-6	4364	-4
5	2	23	2758	583	4370	4136	-6	4170	-5
6	2	23	2762	558	4310	4023	-7	4033	-7
7	2	23	2749	521	4163	3935	-6	3930	-6
8	12	11	1469	415	4524	2023	-56	2013	-56

The last example again represents that Finnish peculiarity, the electric sauna. Many Finnish homes also have an electric sauna which is usually about 6 kW installed power. The other thing is that Finns have a habit of using their saunas at the same time which leads to a well known “sauna peak” for the electric utility. The “sauna peak” is usually on Saturdays, so the working days are not the best days to study it, but to give information on how the use of saunas is distributed over the hours of a week. The use of saunas however, is nowadays spread over the weekdays. In the table we see how the maximum load takes place in the evening of working days at hour 20. The value of the peak for one customer with a 99.5% confidence level is almost 15 kW but with eight customers, only 7 kW. Remember that the figures are given at 10 MWh annual energy and the usual household’s annual energy use is between 3.5 MWh and 5 MWh per year.

Table 24. The model parameters at the time of maximum load of Class 602 Residential, one family house, no electric heat, electric sauna. Number of customers 1…*8. Unit of* m_1 , s_1, $L_{99.5\%}$ *, NE and SLNE is [W/10 MWh/a].*

Nr.	Month	hour	m_1	s_1	$L_{99.5\%}$	NE	q_2[%]	SLNE	q_2[%]
1	3	20	2803	2839	14873	10099	-33	16919	13
2	3	21	2666	1962	11336	7750	-32	9741	-15
3	12	17	2156	1039	9749	4338	-56	4692	-52
4	12	21	2651	1278	8232	5812	-30	6143	-26
5	12	18	2411	939	7638	4944	-36	5072	-34
6	12	21	2701	1170	8038	5244	-35	5288	-35
7	2	19	2651	1087	7778	5096	-35	5070	-35
8	2	19	2755	1126	7070	4938	-31	4862	-32

The overall error q_2 in Table 24 is also systematically around -35 %, which again reminds us that the assumption of independence between loads is not valid.

When the number of customers increases the analysis of the confidence interval obtained from the sample distribution becomes very small. However this is only an applicable result when the sample is from exactly the same population and time. When we apply the load model to a larger number (10…1000…) of customers to estimate their total load, the error comes from differences between the target population and the sample population and the errors applying the model at different times.

In the following chapter the method of Distribution Load Estimation is introduced to apply load models to a large number of customers.

7 DISTRIBUTION LOAD ESTIMATION (DLE)

7.1 GENERAL

The load research has produced load models to convert the customers' annual energy consumption to hourly load values. The reliability of load models applied from a nation-wide sample is limited in any specific network because many local circumstances are different from utility to utility and time to time. Therefore there is a need to find improvements to the load models or, in general, improvements to the load estimates.

In Distribution Load Estimation (DLE) the measurements from the network are utilised to improve the customer class load models (See Fig. 53).

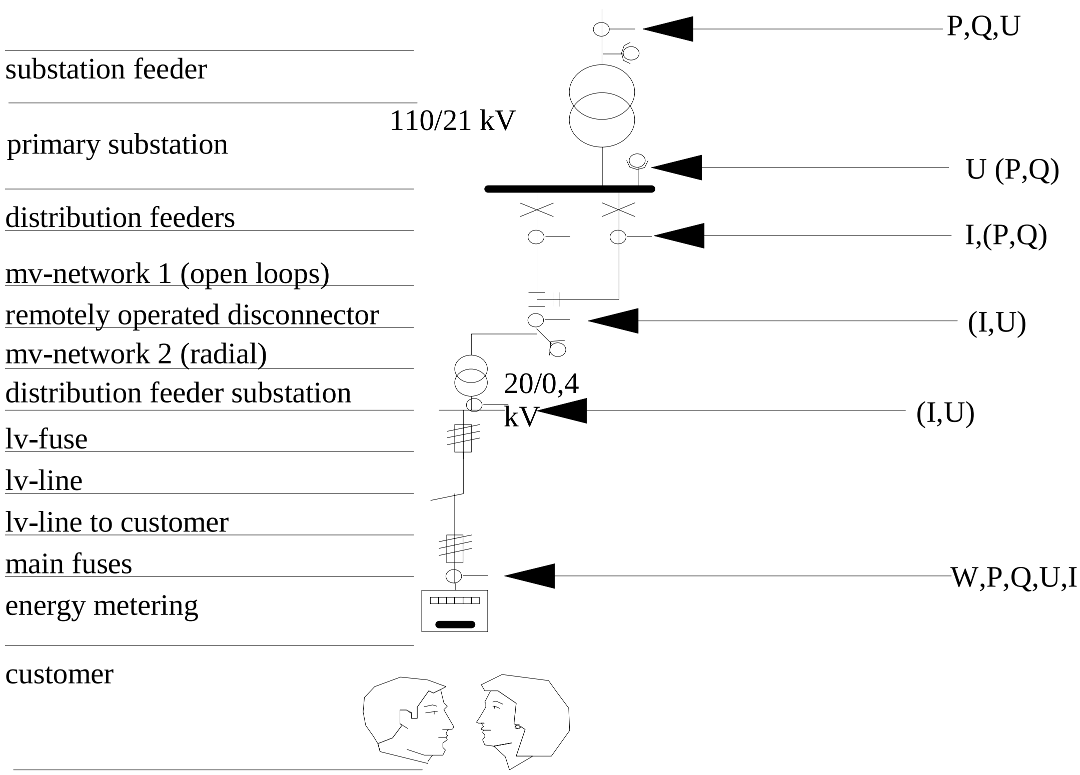


Fig. 53. Possible measurements in distribution network.

The results of DLE will be new load models that better correspond to the loading of the distribution network but are still close to the original load models obtained by load research. The principal data flow of DLE is presented in Fig. 54.

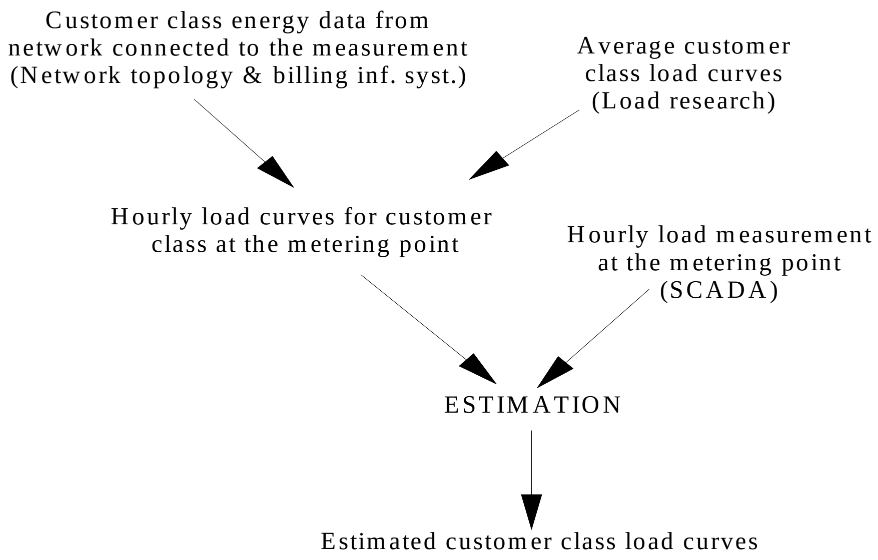


Fig. 54. The data flow of distribution load estimation (DLE) process.

7.2 BACKGROUND

Distribution system load estimation has been studied especially by Handschin and Dörnemann (1988), Dörnemann (1990) and Dörnemann et al. (1990). Their studies handle distribution network loads and customer class load models estimated from the busbar loads using the Bayesian estimation method.

In the USA, distribution network estimation is the subject of research activity, but the difference in the distribution system makes the problems there more complicated. Wu & Neyer (1990), Baran & Kelley (1995) and Ghosh et al. (1996a, 1996b) handle the total three phase system with state estimation techniques. The difference with conventional state estimation methods is that the load data is obtained from load models instead of direct measurements.

Utilising the distributed estimation algorithms, distributed load estimation was studied in (Seppälä 1991). This method could reduce the computation time by applying parallel algorithms when handling large distribution networks.

Kärenlampi et al. (1996) have developed estimation and monitoring systems for distribution networks. The remote meterings are used to adjust the load model data to better fit the measured loads of the feeders of the distribution network. The method is integrated into a network operator’s support system developed by Tampere University of Technology under the distribution automation research programme Edison.

7.3 THE ESTIMATION PROCEDURE

7.3.1 Definition of weighted least squares estimation

We assume that the distribution network is radial operated. Each meter measures the load for a specific part of the network. Usually the load metering points are at the primary substations, but also metering can exist deeper in the network (see Fig. 53). The medium voltage feeders are usually equipped with current metering and the substation primary feeder is also equipped with active power metering. See Fig. 55.

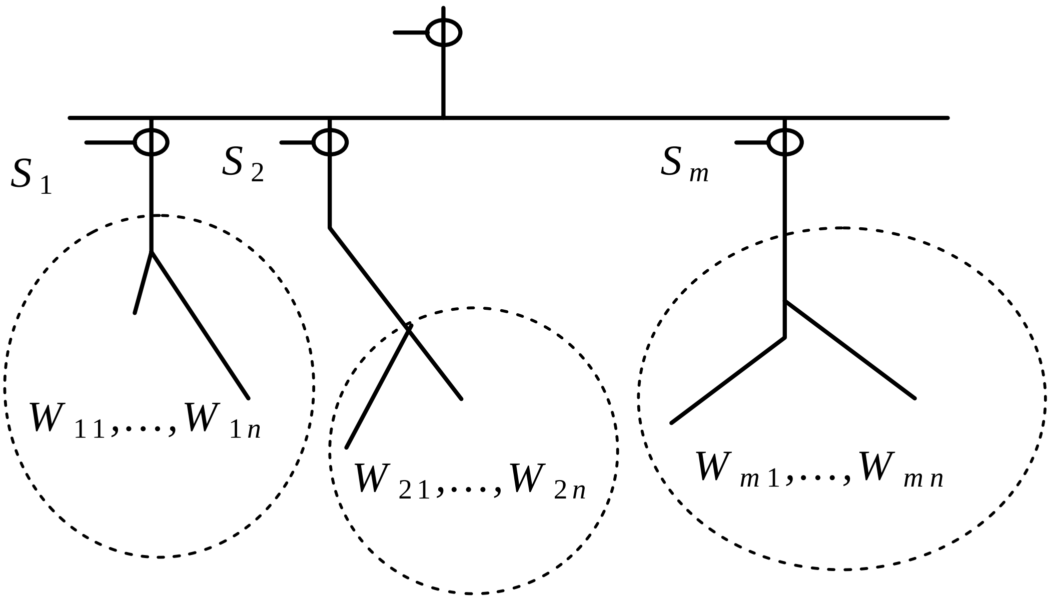


Fig. 55. Distribution station with feeder measurements $S_1 ... S_m$. Over an area i the total annual energy of a customer class j is W_{ij}.

To calculate the hourly load estimate for each customer class we use the linear load model (4):

$$\overline{P}(t) = L_c(m(t), d(t), h(t)) \cdot W_a \tag{70}$$

In the following mathematical manipulation, the load model is briefly presented as the equation

$$z = xW \tag{71}$$

where z denotes the customer class' one hour's expected load $\overline{P}$ derived from the model, W corresponds to annual energy W_a and x corresponds to the factor L_c, which represents the relationship of customer's hourly load to the annual energy consumption.

The annual energy in different areas of the network is represented by W_{ij}, where i defines the area and j defines the customer class.

In the distribution network, the total load is a sum of the loads of distinct customer classes 1...n. The total load of the customer class in the network is z_j. The total annual energy of each class j in the network is presented briefly as W_j

$$W_j = \sum_{i=1}^{m} W_{i,j} \tag{72}$$

The total annual energy consumption of customer classes 1...n over areas 1...m are (See Fig. 55)

$$\begin{cases} W_1 = W_{11} + \cdots + W_{m1} \\ \quad \vdots \\ W_n = W_{1n} + \cdots + W_{mn} \end{cases} \tag{73}$$

The equation between the actual customer class load and the load obtained from the model can be written adding an error term v_j:

$$\begin{cases} z_1 + v_1 = W_1 x_1 \\ \quad \vdots \\ z_\mathrm{n} + v_n = W_n x_n \end{cases} \tag{74}$$

The matrix form of the equations between customer class loads and customer class load models is

$$\begin{bmatrix} z_1 + v_1 \\ \vdots \\ z_n + v_n \end{bmatrix} = \begin{bmatrix} W_1 & \cdots 0 & 0 \\ \vdots & \ddots & 0 \\ 0 & \cdots 0 & W_n \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \Leftrightarrow \mathbf{z} + \mathbf{v} = \mathbf{W}_z \mathbf{x} \tag{75}$$

where $\mathbf{z}$ is one column matrix with the number of rows equal to the number of customer classes. $\mathbf{W}_z$ is a diagonal matrix of total annual energies of customer classes in the network. $\mathbf{x}$ is one column matrix of load model parameters.

For each load measurement S_i in the network we can write equations, where the value of the measured load is the total of the loads of customer classes in the corresponding area. The equations can be written adding an error term e_i

$$\begin{cases} S_1 + e_1 = W_{11}x_1 + \cdots + W_{1n}x_n \\ \vdots \\ S_m + e_m = W_{m1}x_1 + \cdots + W_{mn}x_n \end{cases} \tag{76}$$

The error term also includes the network losses unless the losses are defined as one customer class. The matrix form of the equations is

$$\begin{bmatrix} S_1 + e_1 \\ \vdots \\ S_n + e_n \end{bmatrix} = \begin{bmatrix} W_{11} & \cdots & W_{1n} \\ \vdots & \ddots & \vdots \\ W_{m1} & \cdots & W_{mn} \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \Leftrightarrow \mathbf{S} + \mathbf{e} = \mathbf{W}_S \mathbf{x} \tag{77}$$

S and **e** are one column matrixes where the number of rows is equal to the number of measurements. $\mathbf{W}_\mathrm{S}$ is a matrix where the number of rows equals the number of measurements and the number of columns equals the number of load classes. **x** is a one column matrix where the number of rows equals the number of customer classes.

The two matrix equations (75) and (77) presented together now describe the total system of customer class load models. The load models describe the customer class loads in the network $\mathbf{z} + \mathbf{v} = \mathbf{W}_z \mathbf{x}$ and the relation between the measurements and load models is described by the equation $\mathbf{S} + \mathbf{e} = \mathbf{W}_S \mathbf{x}$. Combining these equations we get an equation of partitioned matrixes

$$\begin{bmatrix} \mathbf{z} \\ \mathbf{S} \end{bmatrix} + \begin{bmatrix} \mathbf{v} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{W_z} \\ \mathbf{W_S} \end{bmatrix} \mathbf{x} \tag{78}$$

Now we want to find **x** that minimises the error $\begin{bmatrix} \mathbf{v} \\ \mathbf{e} \end{bmatrix}$.

For this kind of optimisation problem with several parameters, we will use the method of Weighted Least Squares Estimation (WLSE), which is widely utilised in the state estimation of transmission networks (Debs 1988 p. 291).

7.3.2 The formulation of WLSE

The general solution of the WLSE equation of the form

$$\mathbf{c} = \mathbf{Ab} \tag{79}$$

is solved by solving the minimum of the sum of squares weighted by $\mathbf{R}^{-1}$

$$\min\left[(\mathbf{c} - \mathbf{Ab})^{\mathbf{T}} \mathbf{R}^{-\mathbf{1}} (\mathbf{c} - \mathbf{Ab}\right] \text{ for } \mathbf{b}. \tag{80}$$

where the weights in $\mathbf{R}^{-1}$ are covariances of the variables and measurements. Assuming the variables are independent we get a diagonal matrix with variances.

$$\mathbf{R}^{-1} = \begin{bmatrix} \sigma_1^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \cdots & 0 \\ 0 & \cdots & \sigma_n^2 & 0 & \cdots & 0 \\ 0 & \cdots & 0 & \sigma_{S1}^2 & \cdots & 0 \\ \vdots & \cdots & \vdots & \vdots & \ddots & 0 \\ 0 & \cdots & 0 & 0 & \cdots & \sigma_{Sm}^2 \end{bmatrix}^{-1} \tag{81}$$

The solution of (80) is when the derivative of the equation is zero

$$-2\mathbf{A}^{\mathbf{T}} \mathbf{R}^{-\mathbf{1}} (\mathbf{c} - \mathbf{A}\hat{\mathbf{b}}) = 0 \tag{82}$$

and the solution and the estimate is $\hat{\mathbf{b}}$

$$\hat{\mathbf{b}} = [\mathbf{A}^{\mathbf{T}} \mathbf{R}^{-\mathbf{1}} \mathbf{A}]^{-\mathbf{1}} \mathbf{A}^{\mathbf{T}} \mathbf{R}^{-\mathbf{1}} \mathbf{c} \tag{83}$$

and applied to the DLE problem (78)

$$\hat{\mathbf{x}} = \left[\begin{bmatrix} \mathbf{W_z} \\ \mathbf{W_S} \end{bmatrix}^T \mathbf{R}^{-1} \begin{bmatrix} \mathbf{W_z} \\ \mathbf{W_S} \end{bmatrix} \right]^{-1} \begin{bmatrix} \mathbf{W_z} \\ \mathbf{W_S} \end{bmatrix}^T \mathbf{R}^{-1} \begin{bmatrix} \mathbf{z} \\ \mathbf{S} \end{bmatrix} \tag{84}$$

7.3.3 Definition of the weights

The weights in $\mathbf{R}^{-1}$ are inverses of the variances of models and measurements. This leads to a solution where the estimates of the load models and measurements of higher variance are subject to greater changes than those with lower variance.

Now the question is to select which model of variance should be used. One choice is to use the square of the standard deviation s_P of the load model and the other choice is to use the square of the standard deviation of the sum of k customers ($s/\sqrt{k}$). While the load research data is from different times and usually different population, the variability of the model error at the specific hour is not only a function of the number of customers. While we have no information from the other factors of error variance we rather apply the model (4) for standard deviation

$$s_P(t) = s_{Lc}(m(t), d(t), h(t)) \cdot W_a \tag{85}$$

to evaluate the variance $\sigma^2 = (s_P)^2$. The value of standard deviation of load models in general ranges between 30 % ... 100 % from the mean.

For the network measurements the standard deviation needs to be approximated. One practical method is first to approximate the maximum error. The distribution of the error is unknown, but assuming the error to be roughly normally distributed, the standard deviation is about 1/3 of the maximum error.

For example, when we have a current measurement, the evaluation of active power from current includes an error (the cos ϕ is usually unknown and needs to be estimated too) which may have a standard deviation of 10 % from the absolute value of the measurement. In the case of direct active power measurement the error could be 1 % or less. Thus the direct measurements have smaller error variance than the load models. Also the resolution of the SCADA communication between the remote terminal and central computer bring some error to the registered measurement values.

Only relative differences between weights in $\mathbf{R}^{-1}$ are important. When in practice the measurements get much smaller variances, the solution will more likely change the load models than the measurements.

7.3.4 Application of estimation

Finally with the help of the result $\hat{\mathbf{x}}$ we can solve the new customer class load estimates $\hat{\mathbf{z}}$ and the load at the points of measurements $\hat{\mathbf{S}}$:

$$\begin{aligned} \hat{\mathbf{z}} &= \mathbf{W_z}\hat{\mathbf{x}} \\ \hat{\mathbf{S}} &= \mathbf{W_S}\hat{\mathbf{x}} \end{aligned} \tag{86}$$

The network losses can be taken into account in two ways. The losses may be defined as one load model; otherwise the total losses will be included in the error of the models and measurements. The definition of a load loss model for estimation should be a subject for further studies.

7.4 A DLE EXPERIMENT WITH FOUR SUBSTATION FEEDER MEASUREMENTS

The previous DLE method will be dealt with here with some substation feeder data. Load current measurements from four 20 kV distribution feeders and the customer class information has been used to build the equations presented in the previous chapter. The measurements and load models of the feeders are presented in Fig. 56 and Fig. 57.

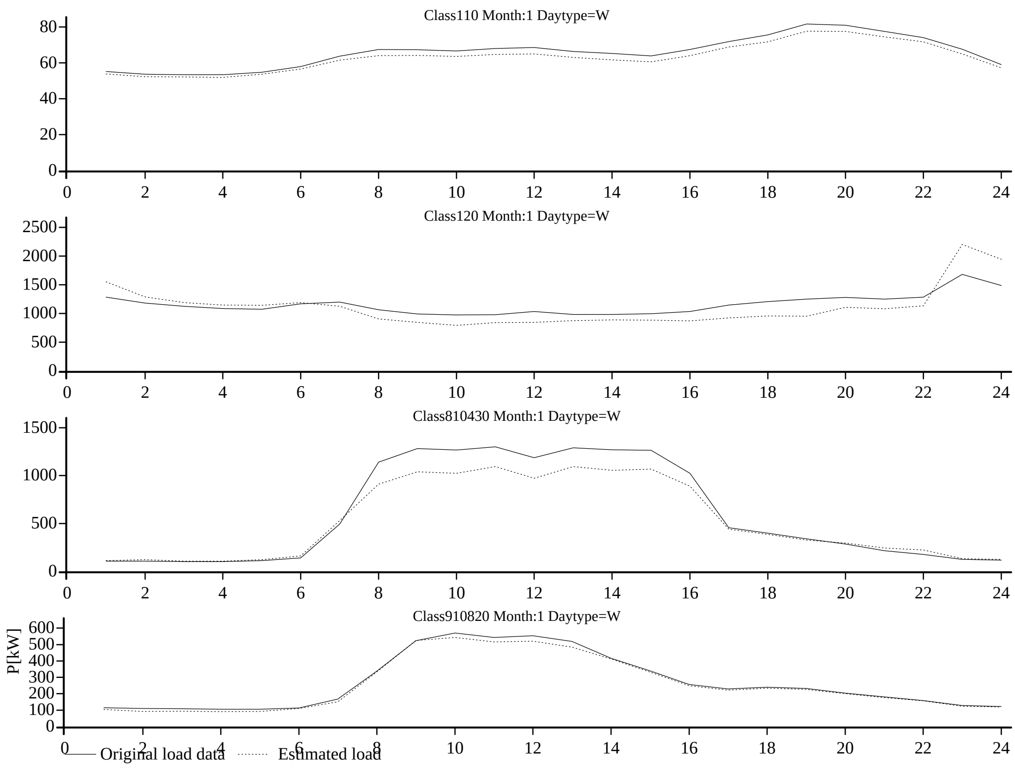


Fig. 56. Examples of how estimation affects customer class load models ($\mathbf{z}_j$). The result of estimation is represented by a dotted line and the original load model is represented as a solid line. 110 = direct electric heating, 120 = direct electric heating with storage water boilers, 810430 = 1 *shift industry (textile) and 910820 = service (private sector). Average values for working days over January.*

Fig. 56 shows an example of how the models change on an average in one month according to estimation with the four substation current measurements. The error is shared between models and measurements depending on how large a share of a load model is represented in one measurement and how high the model variance is. This also means that the small or negligible customer classes are not affected in estimation.

The curve of the customer class ”Electric heating with storage boilers” in Fig. 56 is interesting because it shows exactly the actual situation where the utility controls the boilers simultaneously and the peak is caused by the boilers which are switched on the same time (at 22.00).

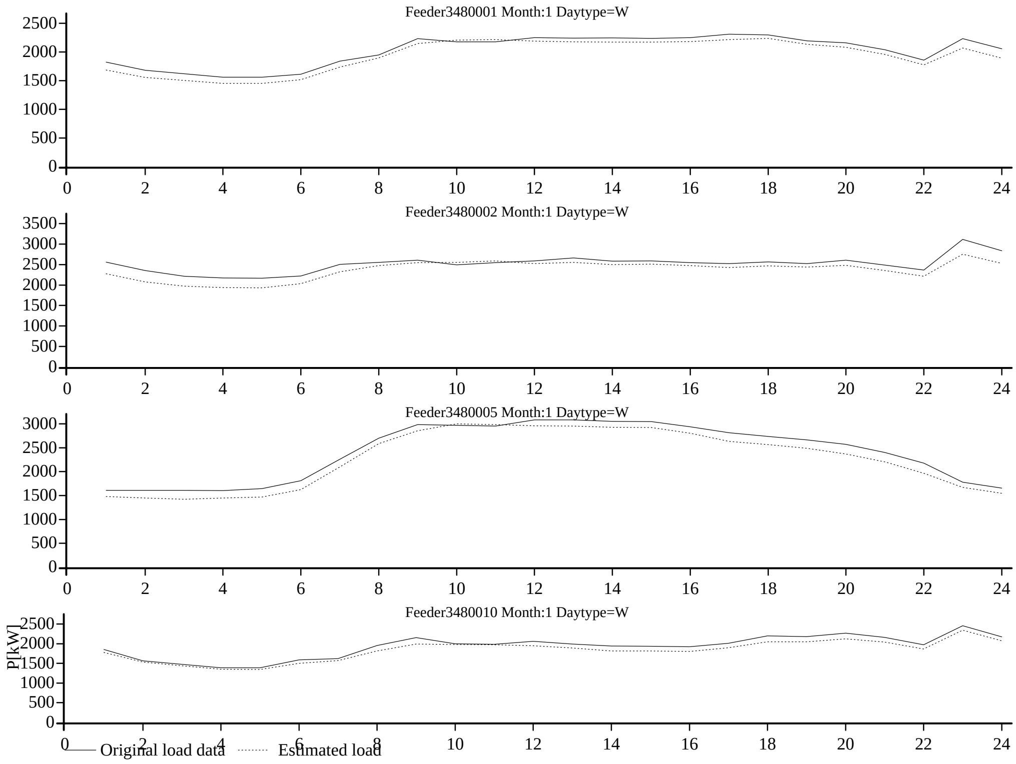


*Fig. 57. Feeders from Meriniitty and Perniö substations (nr. 3480001, 3480002, 3480005 and 3480010) are the measurements (**S**$_i$) . The result of estimation is represented by a dotted line. The feeder measurement, where the values are transformed from current to active power, is represented by a solid line. Average values for working days during January.*

7.5 LOAD ESTIMATION WITH ONE MEASUREMENT

The simplest case of DLE is one measurement in the feeder of a radial network . Here we analyse the special case more to see if it would be possible to formulate a simpler form of load estimation for the special case of one measurement in a radial distribution network. However this result can be generalised to any radial network split to areas of one meter in the feeding point.

The form of the state equations in DLE recalling (78) is

$$\begin{bmatrix} \mathbf{z} \\ \mathbf{S} \end{bmatrix} + \begin{bmatrix} \mathbf{v} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{W_z} \\ \mathbf{W_S} \end{bmatrix} \mathbf{x} \tag{87}$$

Now the **S, e, W**$_{\mathbf{s}}$ are one-row matrixes and the equation with one measurement can be written simply

$$\begin{bmatrix} z_1 + v_1 \\ \vdots \\ z_n + v_n \\ \sum_{j=1}^{n} \left(z_j + v_j \right) + e \end{bmatrix} = \begin{bmatrix} W_1 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & W_n \\ W_1 & \cdots & W_n \end{bmatrix} \cdot \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \tag{88}$$

We minimise the weighted sum of squares of errors in the following form by substituting

$$\begin{bmatrix} \mathbf{z} \\ \mathbf{S} \end{bmatrix} = \begin{bmatrix} z_1 \\ \vdots \\ z_n \\ \sum_{j=1}^{n} \left(z_j + v_j \right) \end{bmatrix} \quad \text{and} \quad \begin{bmatrix} \mathbf{v} \\ \mathbf{e} \end{bmatrix} = \begin{bmatrix} \mathbf{z} \\ \mathbf{S} \end{bmatrix} - \begin{bmatrix} \mathbf{W_z} \\ \mathbf{W_S} \end{bmatrix} [\mathbf{x}] = \begin{bmatrix} v_1 \\ \vdots \\ v_n \\ e \end{bmatrix} \tag{89}$$

to (80) and we get the function G representing the weighted sum which will be minimised

$$\min \left\{ G = \sum_{j=1}^{n} \left(\frac{1}{\sigma_j^2} v_j{}^2 \right) + \frac{1}{\sigma_S^2} e^2 \right\} \tag{90}$$

In practice the total error between the original models and the measurement is known as total value e'

$$e' = \sum_{j=1}^{n} v_j + e \Leftrightarrow e = e' - \sum_{j=1}^{n} v_j \tag{91}$$

Now we state the problem differently: What are the values of v_j to minimise the weighted sum of square error when the total difference e' is given? Or in other words: How should the error e' be divided among the customer classes to fill the WLS-criteria?

Substituting the equation (91) to equation (90) we get

$$\min\left\{G = \sum_{j=1}^{n} \frac{1}{\sigma_j^2} v_j^{\,2} + \frac{1}{\sigma_S^2}\left(e' - \sum_{j=1}^{n} v_j\right)^2\right\} \tag{92}$$

The minimum is found by solving the set of partial derivatives

$$\frac{\partial G}{\partial v_j} = 0 \quad j = 1 \ldots n \tag{93}$$

The result for v_j is (For derivation see Appendix 3)

$$v_j = \frac{\sigma_j^2}{\sigma_S^2 + \sum_{i=1}^{n} \sigma_i^2} e' \tag{94}$$

Substituting this to the formula

$$z_j + v_j = W_j x_j \tag{95}$$

we get the estimate in the form

$$\hat{x}_j = \frac{1}{W_j}\left(z_j + \frac{\sigma_j^2}{\sigma_S^2 + \sum_{i=1}^{n} \sigma_i^2} e'\right) \tag{96}$$

This result is useful for many practical distribution applications, where for example, the voltage drop of the radial distribution feeders is calculated separately. This result states that when the WLSE method is used the error between the loads and metering are divided in proportion to their variances.

EXAMPLE:
From a 20 kV feeder a current of 20 A was measured on Wednesday, the 15th March, 1995 14.00-15.00. The bus voltage was 20.7 kV. From the network information system we receive the information that the feeder was feeding five customer classes according to Table 25.

Table 25. The annual energies of the measured feeder

Customer class	[MWh/a]
1. 1-shift industry	1000
2. Agriculture	200
3. Residential with direct electric heating	1000
4. Storage heating	400
5. Residential	1800
Total	4400

From the load models from load research we obtain the relation between annual energy and the corresponding hour's load as shown in Table 26.

Table 26. Customer class load models and expected load P_j and standard deviation s_j.

Customer class	L_c [W/MWh]	s_{Lc} [W/MWh]	P_j [kW]	s_j [kW]
1. 1-shift industry	297	133	297	133
2. Agriculture	90	64	18	12.8
3. Residential, direct electric heating	135	81	135	81
4. Storage heating	24	15	10	6
5. Residential	101	70	182	126
Total			642	

This information was obtained from the published files of the Finnish load research project. The structure of the model is from the calendar year 1990. Thus we apply the day, March 14th, which was a Wednesday in 1990.

From the feeder current metering we get the active power by assuming first that the load factor cosϕ = 0.9. Thus $P_S = \sqrt{3} \cdot 0.9 \cdot 20 \cdot 20.7 = 710\text{kW}$. The standard error of the measurement σ_P will be approximated as 7 %, thus σ_P = 50 kW. The error between models and measurement *e* = 710 - 642 kW = 68 kW will be shared relative to the variances of the models and measurements. The result is shown in Table 27.

Table 27. The original model class loads, variances, error/correction v_j and new estimated value.

Customer class	Class load P_j [kW]	Variance σ_j^2 [kW^2]	Error v_j [kW]	Estimate $\hat{P}_j$ [kW]
1. 1-shift industry	297	17689	33	330
2. Agriculture	18	4096	8	26
3. Residential with direct electric heating	135	6561	12	147
4. Storage heating	10	225	0.5	10.5
5. Residential	182	4900	9	191
Total	642	33471	62.5	704.5

This shows how the difference between the total of models and the measurements can be quite easily shared between models. This method takes into account both the difference in the sizes of the customer classes and the uncertainty of the model and measurement expressed in the variances of the models and measurements.

7.6 UTILISATION OF DISTRIBUTION LOAD ESTIMATION

When integrated into the utility's information systems and SCADA the DLE does not require any additional investments. The DLE can be utilised in the distribution automation in several ways (Seppälä & Kärkkäinen 1995):

- The output of DLE is a selection of load curves for customer categories and load classes. These curves can be utilised in forecasting purposes.
- DLE brings the possibility of continuous load research where the need of customer level recordings is reduced compared to conventional load research.
- When the electricity markets are free from regulation the DLE brings online information from the loads when the final load values are not available due to the time consuming clearing between producers and sellers. With the help of DLE the utility can calculate their energy balance reliably on an on-line basis.

From the DSM point of view, the DLE can be utilised in several ways, for example:

- The accurate knowledge of feeder load gives the indication for the need of DSM at a specific time (load management, real-time pricing) and site,
- the better estimates of load curves of different customer classes can be utilised in the operation of the load management system (actual timing of the load control) and estimation of the effects of load management on the total load.

The benefits of DLE can be achieved from optimal utilisation of the distribution network capacity

- maximum utilisation of network components
- finding the most profitable targets for network investments and service.

The integration of DLE to the utility's information systems is a task which requires some further experiments. The basic problem is which form of information is needed in further applications. The suggested DLE method results in new load values for each hour. Such information is handled on line and requires applications capable of accepting on line information. Such applications are studied and presented in (Kärenlampi et al. 1996).

Another alternative is to collect the estimated load information to a database where the data will be retrieved for further study. When large differences to applied load models occur the reason for the difference should be analysed and the current load model changed. Some of the distribution network computation applications (at least Tekla, Tietosavo and Versoft) support a load model editor which can be used to update the load models according to the information retrieved from load estimation.

The estimation algorithm is very general and brings a lot of possibilities to develop applications. For example, the results of estimated loads could be used recursively in further estimations. Such variations and improvements should be targets for further analyses especially when there is a continuously running DLE installation available to test with live data.

8 DEVELOPMENT OF THE APPLICATIONS

The results of this thesis will now be reviewed together with an analysis of future applications for development. This work has been done as a part of the Finnish distribution automation research programme Edison, which is integrating several development projects to build a new scheme for a distribution automation system (Lehtonen 1996).

Today load research and utilisation of load models are on a high level in Finnish utilities. Similar high level integration of distribution applications is difficult to find in other countries. The direction of development from this point can be seen in two ways:

- Development of applications and improvement of load data in Finnish utilities
- Development of products for distribution and applications for domestic and international markets.

8.1 DEVELOPMENT OF UTILITIES' APPLICATIONS

The free electricity market in Finland has changed the prospects of load research. In the monopoly situation the local seller's developed local tariffs and pricing schemes which have also affected the load variation of customers. In the future such loacal features are expected to change. Also the number of distribution companies is speculated to shrink from the current 103.

The electricity market for small customers is being considered by the authorities. The requirement of hourly meterings for customers participating in the electricity market is too expensive. One alternative is that the small customers participate in the markets using type load models. In such a case the need for load research data will increase. Also the methods to estimate loads of customer classes using load estimation (DLE) techniques are needed to adjust the energy sold to the total load observed in the feeding substation of the distribution network (Lehtonen et al. 1996).

While the sales of energy will be changing in many ways the distribution function itself remains a monopoly. The authorities will be supervising the distributors and the main problem is to focus the network investments in an optimal way and keep operating costs under control. Load data and applications of load models will be needed again not only to help the functions of planning and operation but also to convince the authorities and public.

The application of confidence interval estimation presented in chapter 5 is basic. The new results are applicable at best in low voltage network calculations where there are only one or a few customers. While load estimation of one customer is not very important as a single case, it becomes important because the number of such cases becomes high in distribution networks. An analysis of correlation between customer classes is suggested for further study. The correlation between customer classes is needed to complete the calculations of the sums of loads of different customer classes in distribution networks.

The simplified confidence interval estimation (SLNE) could be added to the network computation and planning applications' toolbox (See example on page 85.).

The results of load distribution in chapter 4 and 5 are important tools when analysing the impact of Demand Side Management (DSM) functions and loads of end use appliances. The outcome of DSM operations are always random and require simulation tools. The analysis of the origins of distribution functions and confidence intervals is a contribution to development of DSM analysis evolving from end use appliances to customer load and further to total system load.

The new metering techniques and requirements of hourly metering in the electricity market bring large amounts of load data. The availability of the load data also brings a challenge to analyse and benefit from the information. The statistical modelling technique and simple load models bring a straightforward method to utilise the load data from remote meterings.

When the amount of load measurement data increases, data management becomes an essential factor when utilising the load information. This is an acute challenge to software and applications development.

8.2 DEVELOPMENT OF DISTRIBUTION AUTOMATION PRODUCTS

The Finnish distribution system and distribution applications are well advanced compared to corresponding systems in other countries. This is an advantage for the development of products for distribution applications. However, development of applications for other markets where the infrastructure is not similar to Finland, requires special attention.

For example, the concept of load research and load estimation, presented in this thesis, requires good background data from customers and their energy use. When such information is not available the applications should be capable of supporting the user to generate the needed information. The algorithms should be so robust that they accept rough approximations as well as completely defined models or measurements.

Therefore one target for further study should be the development of these methods to work with minimal data and also applications supporting a complete lack of background information.

9 CONCLUSIONS

The analysis of customer loads and load estimation is a traditional area of electricity distribution technology. Modern computers and load research data collecting techniques and analyses have led us to new sources of information. More accurate methods of analyses are required because the competition and market forces will increase the demands of productivity from the electric utilities. During the last decade the availability of load data has increased. This study has given some methods of how the information could be utilised.

In this thesis a model for customer electric load variation and a new method for estimating confidence intervals was developed. The method is simple and easy to use. It can be applied in network load flow and voltage analysis and customer pricing applications.

A method for Distribution Load Estimation (DLE) which combines the models and measurements was developed. The load estimation of a distribution network is reaching the level where it could be introduced as a practical application. The utilities are getting better estimates from the participation of different load categories to the system's total load.

The Finnish national load research project has proven the usefulness of customer load data analysis. In spite of the success of the load research project, the accuracy of the estimates is still limited. Local customer load research data is required before accurate load estimates from customer load research data are possible.

The electric utilities should systematise their load research, even on a smaller scale. The benefits of improved accuracy in the load estimates will no doubt pay back the costs of load research.

This thesis has pointed out several subjects for further study:

- New load research should be targeted to loads which don't have regular seasonal variations.
- The linking of the load models to utility databases should be analysed to get more accurate information on how the load models are in general applied.
- The customer load variation model should be analysed further to find out the complete theoretical basis determining the statistical distribution of customer loads.
- The applicability of the load variation model to DSM studies should be analysed.

- The temperature standardisation should be studied because the latest results show that the temperature standardisation of earlier studies did not have the expected results.
- The load correlation between and within classes should be studied to form models for cumulated loads for several customers from several classes.
- The random algorithm for simulation of the loads for several customers should be developed so that they better take into account the varying amounts of available load data.
- The transmission loss load model for DLE should be developed.
- The variation of DLE taking recursively the estimated values should be studied.
- The method of configuring DLE without preliminary load models should be developed.

REFERENCES

Aitchison, J., Brown, J. A. C. 1957. The lognormal distribution. Cambridge University Press. 176 p.

Allera, S. 1994. A letter sent 21.10.1994.

Baart de la Faille, C. J. 1915. On the logarithmic frequency curve and its biological importance. Recueil des Travaux Botaniques Néerlandais. Volume XII. Groningen: Société Botanique Néerlandaise. Pp. 349 - 368.

Baran, M. E., Kelley, A. W. 1995. A branch-current-based state estimation method for distribution systems. IEEE Transactions on Power Systems. Vol 10, No 1, pp. 483 - 489.

Debs, A. S. 1988. Modern power systems control and operation. Kluwer Academic Publishers. 368 p. ISBN 0-89838-265-3.

Dörnemann, C. 1990. Betriebsmittelbezogene Lastmodellierung für die Berechnung in Verteilungsnetzen. Dissertation. Universität Dortmund. 105 p.

Dörnemann, C., Freund, H, Handschin, E. 1990. Netzberechnung in Mittel und Niederspannungsnetzen mit Hilfe einer betriebmittelbezogenen Lastmodellierung. Elektrizitätswirtschaft, Vol 89, No 22, pp. 1248 - 1256.

EA Electricity Association 1994. Load research progress report July 1994. London: Electricity Association Services Limited. 5 p.

Erhiö, I. 1991. Sähkönjakeluverkon seurannan kehittäminen (The development of distribution network monitoring). MSc thesis. Helsinki University of Technology. 89 p. (In Finnish.)

Feilberg, N., Livik, K. 1993. Energy and load structure at various categories of end-users. Trondheim: EFI Norwegian Electric Power Research Institute. ISBN 82-594-0551-2.

Fikri, Z. 1975. Statistical load analysis for distribution network planning. Dissertation. Stockholm: Department of electric power systems engineering, The royal institute of technology. 376 p.

Gellings, W. C., Swift, M. A. 1988. The value of load research. Public Utilities Fortnightly, June 9, pp. 32 - 40.

Ghosh, A. K., Lubkeman, D. L., Downey, M. J., Jones, R. H. 1996a. Distribution circuit state estimation using probabilistic approach. In: Proceedings of IEEE Winter meeting, 96 WM 171-9-PWRS.

Ghosh, A. K., Lubkeman, D. L., Jones, R. H. 1996b. Load modelling for distribution circuit state estimation. In: Proceedings of IEEE Transmission & Distribution meeting, 1996.

Hald, A. 1967. Statistical theory with engineering applications. USA: John Wiley & Sons Inc. 7th printing. 783 p.

Handschin, E., Dörnemann, C. 1988. Bus load modelling and forecasting. IEEE Transactions on Power Systems, Vol 3, No 2, pp. 627 - 633.

Herman, R., Kritzinger, J. J. 1993. The statistical description of grouped domestic electric load currents. Electric Power Systems research, Vol 27. pp. 43 - 48.

Härkönen, H. 1987. Sähköverkon kuormitusten mallintaminen verkosto-suunnittelua varten. (Load modelling for distribution network planning). MSc thesis. Helsinki University of Technology. (In Finnish.)

Irwin, G. W., Monteith, W., Beattle, W. C. 1986. Statistical electricity demand modelling from consumer billing data. IEE proceedings, Vol 133, No 6, pp. 328 - 335.

Johnson, N. L., Kotz, S. 1970. Distributions in statistics: Continuous univariate distributions - 1. New York: John Wiley&Sons.

Juuti, P., Kärkkäinen, S., Kalevi, J. 1987. Load curve simulation experiences in network calculations and load management in Finland. In: Proceedings of 9th International Conference on Electricity Distribution CIRED. Pp. a.15.1 - a.15.6.

Kapteyn, J. C. 1916. Skew frequency curves in biology and statistics, Recueil des Travaux Botaniques Néerlandais. Volume XIII. Groningen: Société Botanique Néerlandaise. Pp. 105 - 151.

Karanta, I., Ruusunen, J. 1991. Short term load forecasting in communal electric utilities, Helsinki University of Technology, System Analysis Laboratory, Research Reports A40. 104 p.

Kofod, C., Lange-Hüsken, M., Paradinas, J. G., Voldhaug, L. 1988. Household load curve studies - Contribution to the system demand and load management. In: Proceedings of UNIPEDE Sorrento Congress 1988, Paper 88E 60/D.2. 33 p.

Kohtala, E., Koivuranta, K. 1991. A program operating in a workstation environment for LV and MV network design. In: Proceedings of the 11th International Conference on Electricity Distribution CIRED. Pp. 6.16.1 - 6.16.5.

Kärenlampi, M., Verho, P., Järventausta, P., Partanen, J. (1996). Forecasting the short-term loads of mv-feeders and distribution substations - a DMS function. In: Proceedings of the 12th Power Systems Computation Conference PSCC, Dresden, 19 - 23 August 1996. Pp. 757 - 763.

Lakervi, E., Holmes, E. J. 1995 Electricity distribution network design. 2nd ed. London: Peter Peregrinus Ltd, 321 p. ISBN 0-86341-309-9.

Lehtonen, M. (ed.) 1996. EDISON - research programme on electricity distribution automation 1993 - 1997. Interim report 1995. Espoo: Technical Research Centre of Finland, VTT Research Notes 1735. 127 p. + app. 7 p. ISBN 951-38-4890-6

Lehtonen, M. Seppälä, A., Takala, J. 1996. Pienten kuluttajien liittäminen vapaan sähkökaupan piiriin (Small cusotmer in a free electricity market). Helsinki: Sähkömarkkinakeskus, julkaisuja 1 / 1996. 63 p. (In Finnish.)

Leino, K. 1974. Kouvolan kaupungin sähkölaitoksen kuormituskäyrän analysointi pienkuluttajien tariffin muokkausta varten (Analysis of Kouvola town’s load curve for tariffs planning for small customers). Tilastotieteen laudaturtyö. Helsingin yliopisto, Tilastotieteen laitos. 41 p. (In Finnish.)

Lokki, O. 1980. Tutkimustulosten tilastollinen hallinta ja käyttö. Insinööritieto Oy. 591 p. (In Finnish.)

Meldorf, M. 1995. Steady-State Monitoring of Power System. Theses of Tallinn Technical University. 23 p.

Milton, J. S., Arnold, J. C. 1990. Introduction to probability and statistics: principles and application for engineering and the computing sciences. Singapore: McGraw-Hill Book Co. ISBN 0-07-100812-8

Ojala, V. 1992. An integrated PC-program for the tariff planning in electric utilities and for national price statistics on electricity. In: Proceedings of 7th International Conference on Metering Apparatus and Tariffs for Electricity Supply, 17 - 19 November 1992. London: IEE. Pp. 72 - 76 ISBN 0-85296-555-9.

Paananen, A. 1991. Kuormitusmittausten hyödyntäminen tariffien suunnittelussa (Utilising load recordings in tariff planning). MSc thesis. Lappeenranta University of Technology. (In Finnish.)

Pahkinen, E., Lehtonen, R. 1989. Otanta-asetelmat ja tilastollinen analyysi, Helsinki: Oy Gaudeamus Ab. 286 p. ISBN 951-662-470-7 (In Finnish.)

Papoulis A. 1965. Probability, random variables and stochastic processes. New York: McGraw-Hill. 583 p.

Partanen, J. 1991. A PC-based information and design system for electricity distribution networks. Dissertation. Tampere University of Technology, Publications 79. 104 p.

Puromäki, A. 1959. Sähkölaitoksen kuormituskäyräanalyysi moniregressio-menetelmällä (Analysis of electric utility's load curve with multiregression method). Voima ja Valo 1959, Nos 5 - 6 , pp. 116 - 123. (In Finnish.)

Rossinen, E. 1982. Kuormituskäyrätietoihin perustuva sähkönjakeluverkon laskentaohjelmisto (Distribution network software based on load curve data). MSc thesis. Tampere University of Technology. (In Finnish.)

Räsänen, M. 1995. Modelling processes in the design of electricity tariffs. Dissertation. Helsinki University of Technology, System Analysis Laboratory, Research Reports A60.

SEF (Association of Swedish Electric Utilities) 1991. Belastningsberäkning med typkurvor. Stockholm: Svenska Elverksföreningen. 217 p. (In Swedish.)

Seppälä, A., Paananen, A. 1992. Sähkölaitosten kuormitustutkimus 1992 (Electric utilities load research 1992). Helsinki: Suomen Sähkölaitosyhdistys julkaisuja 5/92. 172 p. (In Finnish.)

Seppälä, A. 1984. Sähkölaitosten kuormitustutkimus (Load research of electric utilities). MSc thesis. Helsinki University of Technology. 79 p. (In Finnish.)

Seppälä, A. 1991. Sähkönjakelun laskennan menetelmien ja tietotekniikan kehittäminen, (The development of methods and data processing for distribution network computation). Licentiate's thesis. Helsinki University of Technology. 58 p. (In Finnish.)

Seppälä, A. 1993. Load research and load modelling for distribution network planning and operation. In: Proceedings of Third International Symposium on electricity distribution and load management. Singapore, 27 - 29 October 1993, Singapore: IEEE Singapore section. Vol 1. Pp.284 - 288.

Seppälä, A. 1994. Load research and distribution load estimation. In: Proceedings of Nordic Distribution Automation Conference, 25 -26 May 1994, Copenhagen.

Seppälä, A. 1995. Load research and electrical distribution load modelling. In: Proceedings of the 13th International Conference on Electricity Distribution CIRED. Brussels, 8 - 11 May 1995. Pp. 5.22.1 - 5.22.5.

Seppälä, A., Kärkkäinen S. 1995. Load research and load estimation - tools for DA and DSM. In: Proceedings of DA/DSM'95 Europe conference. Vol II. Rome, 21 -23 November 1995. The Netherlands: Pennwell Conferences & Exehibitions. Pp. 349 - 366.

Seppälä, A 1996. The estimation and simulation of electricity customer hourly load distribution. In: Proceedings of the 12th Power Systems Computation Conference PSCC. Dresden, 19 -23 August 1996. Pp. 238 - 244.

Siirto, O. 1989. Sähkön tarpeen lämpötilariippuvuuden tutkiminen (Study of temperature dependency of electricity consumption). MSc thesis. Helsinki University of Technology. 73 p. (In Finnish.)

SRC (Synergic Resources Corporation) 1983. Selected statistical methods for analysis of load research data. SRC report no. 7054, draft May 20. Prepared for EPRI.

STYV (Co-operation organisation for Finnish power producers) 1985. Sähkön kulutuksen indeksisarjat (Index series for electricity consumption). Planning Committee Report, October 1985. (In Finnish.)

Särndal, C.-E., Wright R. L. 1984. Cosmetic form of estimators in survey sampling, Scandinavian Journal of Statistics. Vol 11, pp. 146 - 156.

UNIPEDE (International Union of Producers and Distributors of Electrical Energy) 1973. The study of load curves in electricity supply economics - Manual of theory and practical procedure. 130 p.

Wolf, M. 1959. Enzyklopädie der Energiewirtschaft, Zweiter Band: Belastungskurven und Dauerlinien in der elektrischen Energiewirtschaft. Berlin: Springer-Verlag. 563 p.

Wu, F. F., Neyer, A. F. 1990. Asynchronous distributed state estimation for power distribution systems. In Proceedings of 10th Power Systems Computation Conference PSCC, Graz, 19 - 24 August 1990. London: Butherworths. Pp. 439 - 446.

APPENDIX 1

List of publications on the results of the load research project.

The first results from the load research project were published in 1985 by the Co-operation organisation for Finnish power producers (STYV). The work was done jointly by the planning committee and published. The title of the publication was:

- *Sähkön kulutuksen indeksisarjat - selvitys tarkistustyöstä 1985. Suunnitteluvaliokunnan raportti . 3/85. 40 p. + app.*

The Association of Finnish Electric Utilities has published following reports presenting the results of the load research project:

1. *Sähkön käytön kuormitusmittaukset. SLY julkaisuja 1/1986. 14 p. + app.*

2. *Sähkölaitosten kuormitustutkimus. SLY julkaisuja 3/1988. 19 p. + app.*

3. *Sähkölaitosten kuormitustutkimus 1992. SLY julkaisuja 5/92. 172 p.*

4. *Data disks of the load curves and standard deviations in various formats.*

At least 3 and 4 are available from SLY-Palvelu Oy.

All publications are in Finnish.

APPENDIX 2

2.1 Customer classification in 1992 analysis.

Residential customers

Class	Class description	Number of recordings included
110	One family house, direct electric heat, water boiler <300l	54
120	One family house, direct electric heat, water boiler =300l	65
130	One family house, direct electric heat, floor heating > 2kW	18
210	One family house, partial storage electric heat, short disconnect periods	12
220	One family house, partial storage electric heat, long disconnect periods (7-22)	27
300	One family house, full storage electric heat, (7-22)	16
400	One family house, heat pump	34
510	One family house, dual heat, flat tariff	9
520	One family house, dual heat, night tariff	9
530	One family house, dual heat, seasonal tariff	17
601	One family house, no electric heat, no electric sauna	10
602	One family house, no electric heat, electric sauna	22
611	Flat, no elect. heat, no electric sauna.	24
612	Flat, no electric heat, electric sauna	4
1010	Block of flats, no flats included	6
1020	Block of flats	8
1030	Semi detached house, direct electric heating, whole building	18
1120	Summer cottages (sub-station level)	11
711	Agriculture, milk production, residence excluded	13
712	Agriculture, milk production, residence included	28
713	Agriculture, milk production, residence included, electric, sauna	13
714	Agriculture, milk production, residence included, electric sauna, electric heat.	15
721	Agriculture, meat production, residence excluded.	2
722	Agriculture, meat production, residence included	4
732	Agriculture, crop production, residence included	7
733	Agriculture, crop production, residence included, electric sauna	2

Industrial customers 1-shift

Class	Class description	Number of recordings included
810	1-shift industry all branches	all below
810430	Textile, clothing and leather industry, 1-shift	15
810440	Wood industry (mechanic), 1-shift	9
810460	Chemical, oil, gum and plastic industry, 1-shift	8
810480	Metal and machine works, 1-shift	17

Industrial customers 2-shift

Class	Class description	Number of recordings included
820	2-shift industry all branches	all below
820420	Food, drink & tobacco industry, 2-shifts	18
820430	Textile, clothing and leather industry, 2-shift	3
820452	Paper products manufacturing, graphical industry, 1-shift	6
820460	Chemical, oil, gum and plastic industry, 2-shift	9
820480	Metal and machine works, 2-shift	9

Service customers public

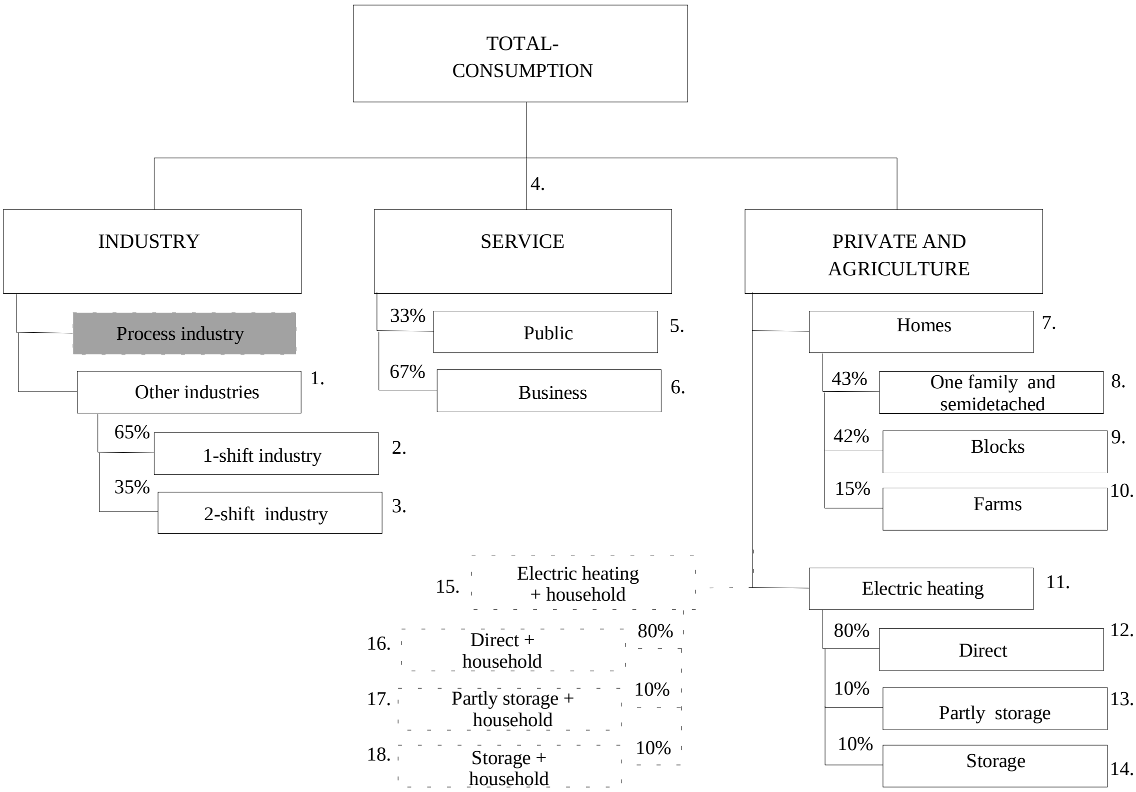
Class	Class description	Number of recordings included
910	All branches	all below
910810	Administration	8
910820	Education, schools	10
910830	Hospitals and health care	6

Service customers private

Class	Class description	Number of recordings included
920	All branches	all below
920610	Wholesale trade	5
920622	Department store	29
920622	Retail shops	8
920630	Car retail and service	6
920640	Hotels, accommodation service	5
920650	Restaurant and café	3
920660	Bank & Insurance	13
920670	Recreation and cultural service	4

2.2 Customer classification for total energy demand

This classification is traditional in Finnish production planning applications and is derived from the previous groups of Appendix. 2.1. for compatibility. The simplicity and general form of these load models keep them popular in applications which don't require more specific classification. The idea for this classification is presented in the following figure which describes the distribution of total (nation-wide) energy consumption to these different categories. The percentages describe how the larger groups are combined. For example, the category Other industries (1.) consists of 65% 1-shift industry (2) and 35% 2-shift industry.



APPENDIX 3

Derivation of the load estimation with one measurement

The problem is to solve the value for v_j so that the weighted sum of squares of errors is minimum in the equation

$$\min\left\{ G = \left(\sum_{j=1}^{n} \frac{1}{\sigma_j^2} v_j^{\,2} \right) + \frac{1}{\sigma_S^2} \left(e' - \sum_{j=1}^{n} v_j \right)^2 \right\} \qquad (1)$$

The minimum is found by solving the set of partial derivatives

$$\frac{\partial G}{\partial v_j} = 0 \qquad j = 1 \ldots n \qquad (2)$$

The set of equations is then

$$\begin{cases} \dfrac{2}{\sigma_1^2} v_1 - \dfrac{2}{\sigma_S^2} (e' - \sum\limits_{j=1}^{n} v_j) = 0 \\ \vdots \\ \dfrac{2}{\sigma_n^2} v_n - \dfrac{2}{\sigma_S^2} (e' - \sum\limits_{j=1}^{n} v_j) = 0 \end{cases} \qquad (3)$$

$\Leftrightarrow$

$$\begin{cases} \sigma_S^2 v_1 - \sigma_1^2 (e' - \sum\limits_{j=1}^{n} v_j) = 0 \\ \vdots \\ \sigma_S^2 v_n - \sigma_n^2 (e' - \sum\limits_{j=1}^{n} v_j) = 0 \end{cases} \qquad (4)$$

Now summing up these equations and solving $\sum\limits_{j=1}^{n} v_j$ we get

$$\sigma_S^2 v_1 + \ldots + \sigma_S^2 v_n - \left(\sigma_1^2 (e' - \sum_{j=1}^{n} v_j) + \ldots + \sigma_n^2 (e' - \sum_{j=1}^{n} v_j) \right) = 0 \qquad (5)$$

$\Leftrightarrow$

$$\sigma_S^2 (v_1 + \ldots + v_n) + \left((\sigma_1^2 + \ldots + \sigma_n^2) \sum_{j=1}^{n} v_j \right) - \left((\sigma_1^2 + \ldots + \sigma_n^2) e' \right) = 0 \qquad (6)$$

$\Leftrightarrow$

$$\sigma_S^2 \sum_{j=1}^{n} v_j + \left((\sigma_1^2 + \ldots + \sigma_n^2) \sum_{j=1}^{n} v_j \right) - \left((\sigma_1^2 + \ldots + \sigma_n^2) e' \right) = 0 \qquad (7)$$

$\Leftrightarrow$

$$\left(\sigma_S^2 + \sigma_1^2 + \ldots + \sigma_n^2 \right) \sum_{j=1}^{n} v_j - \left((\sigma_1^2 + \ldots + \sigma_n^2) e' \right) = 0 \qquad (8)$$

$\Leftrightarrow$

$$\sum_{j=1}^{n} v_j = \frac{\sigma_1^2 + \ldots + \sigma_n^2}{\sigma_S^2 + \sigma_1^2 + \ldots + \sigma_n^2} e' \qquad (9)$$

Substituting this to the set of equations (4)

$$\begin{cases} \sigma_S^2 v_1 - \sigma_1^2 (e' - \dfrac{\sigma_1^2 + \ldots + \sigma_n^2}{\sigma_S^2 + \sigma_1^2 + \ldots + \sigma_n^2} e') = 0 \\ \vdots \\ \sigma_S^2 v_n - \sigma_n^2 (e' - \dfrac{\sigma_1^2 + \ldots + \sigma_n^2}{\sigma_S^2 + \sigma_1^2 + \ldots + \sigma_n^2} e') = 0 \end{cases} \qquad (10)$$

$\Leftrightarrow$

$$\begin{cases} \sigma_S^2 v_1 - \sigma_1^2 (\dfrac{\sigma_S^2}{\sigma_S^2 + \sigma_1^2 + \ldots + \sigma_n^2} e') = 0 \\ \vdots \\ \sigma_S^2 v_n - \sigma_n^2 (\dfrac{\sigma_S^2}{\sigma_S^2 + \sigma_1^2 + \ldots + \sigma_n^2} e') = 0 \end{cases} \qquad (11)$$

$\Leftrightarrow$

$$\begin{cases} v_1 = \dfrac{\sigma_1^2}{\sigma_S^2 + \sigma_1^2 + \ldots + \sigma_n^2} e' \\ \vdots \\ v_n = \dfrac{\sigma_n^2}{\sigma_S^2 + \sigma_1^2 + \ldots + \sigma_n^2} e' \end{cases} \qquad (12)$$

$\Leftrightarrow$

$$v_j = \frac{\sigma_j^2}{\sigma_S^2 + \sum_{i=1}^{n} \sigma_i^2} e' \qquad (13)$$

APPENDIX 4

4.1 Simple load models for selected customer classes.

These figures are based on the load model data files published by the Association of Finnish Electric Utilities in 1992. The figures present simple form load model paramters, the average load and standard deviation in W for 10 MWh annual energy use W_a.

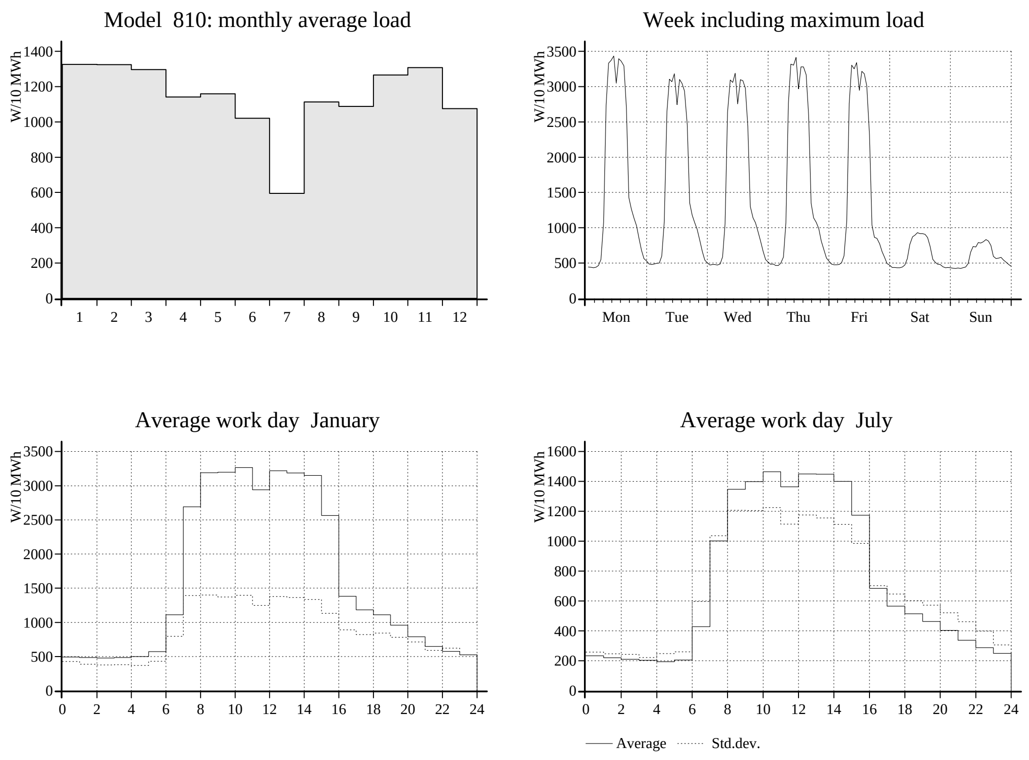


Fig. 1. Industry 1-shift. Model nr 810.

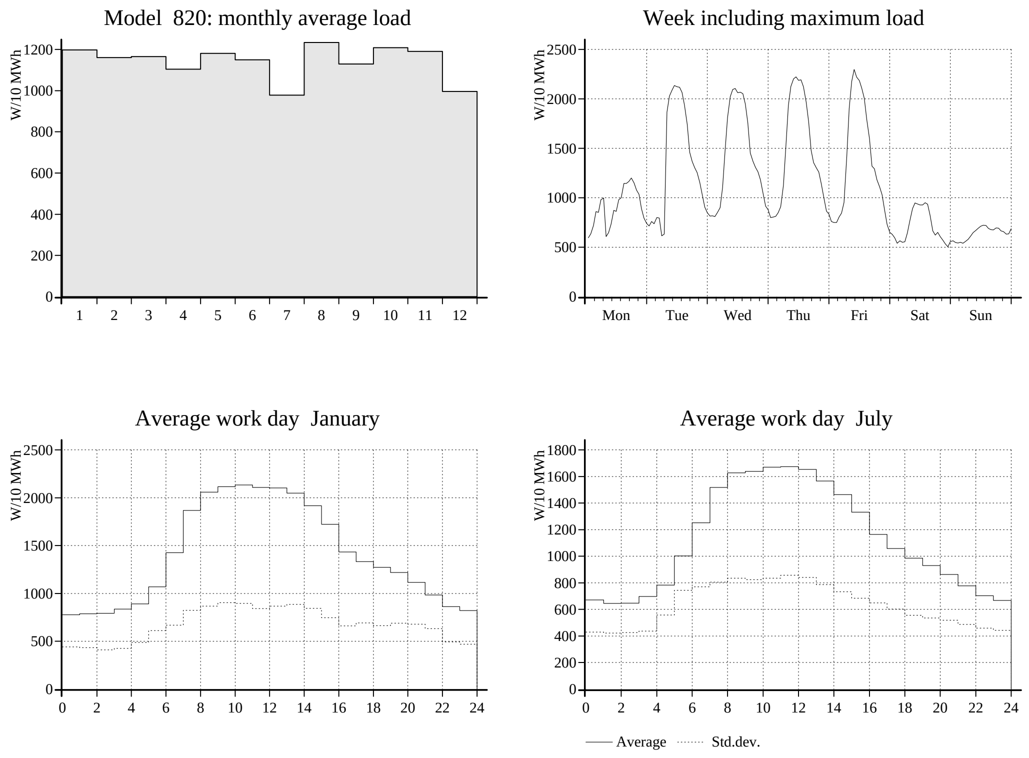


Fig. 2. Industry 2-shift. Model nr 820.

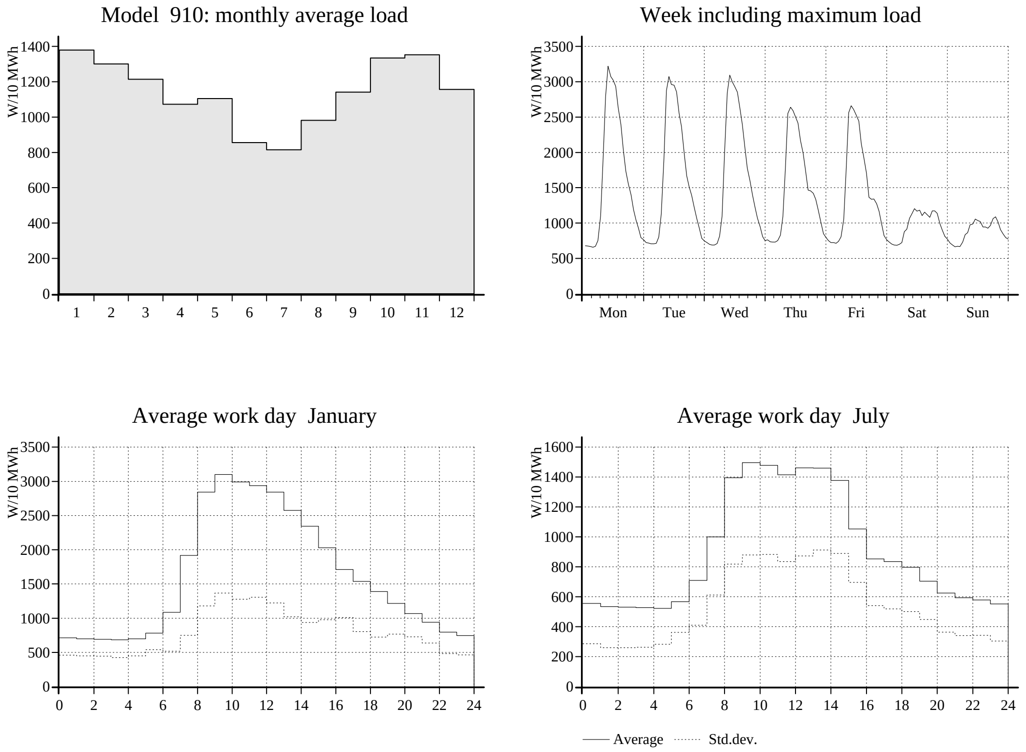


Fig. 3. Service, public. Model nr 910.

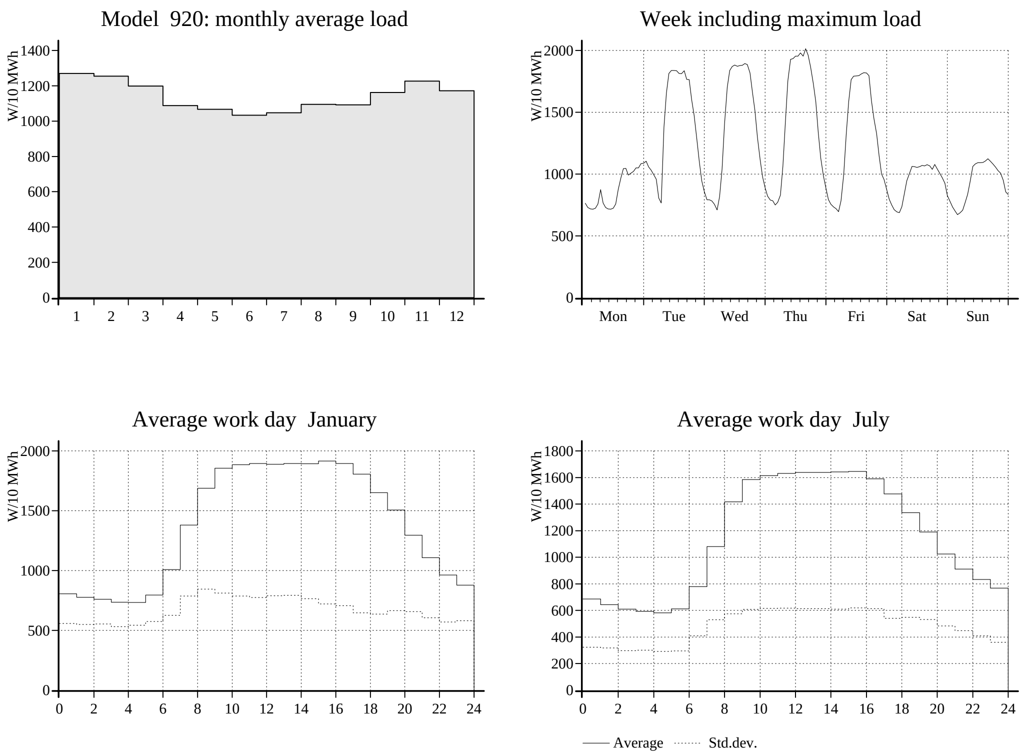


Fig. 4. Service, private. Model nr 920.

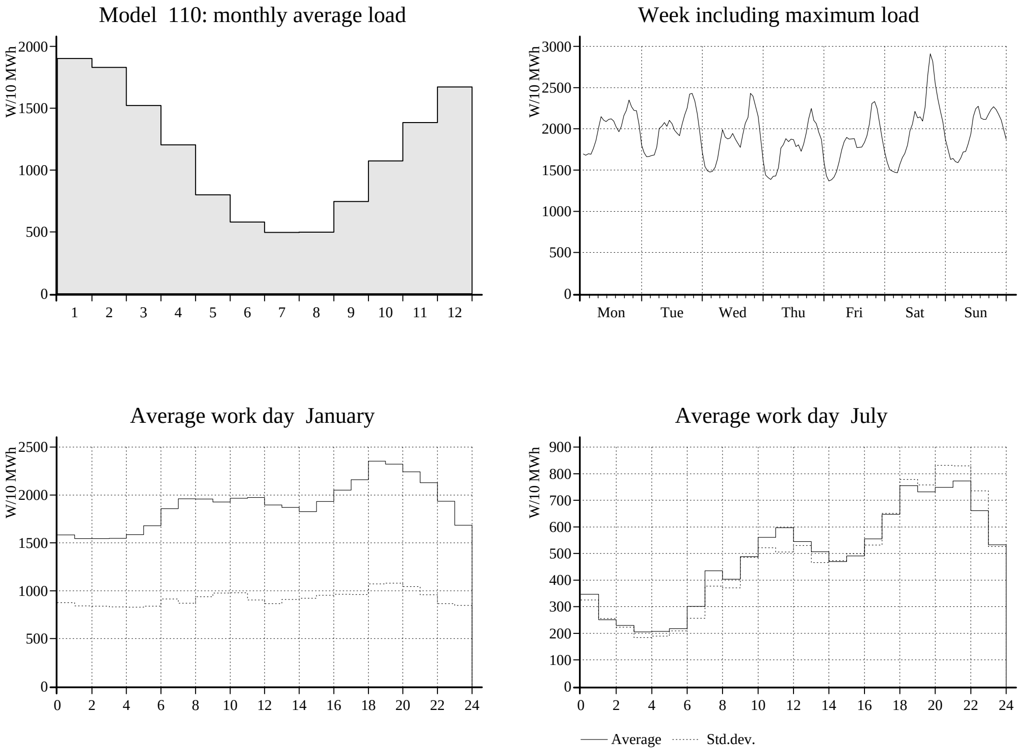


Fig. 5. Electric heat, one family house. Model nr 110. Standardised to long term average temperature.

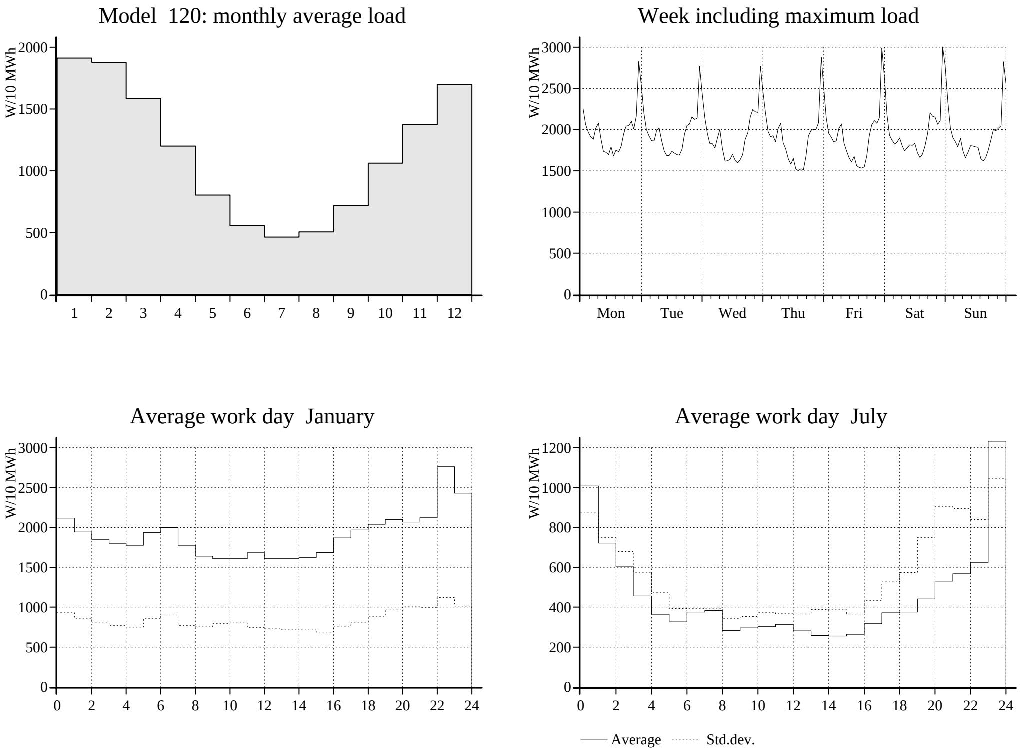


Fig. 6. Electric heat, one family house. Model nr 120. Storage water heating. Standardised to long term average temperature.

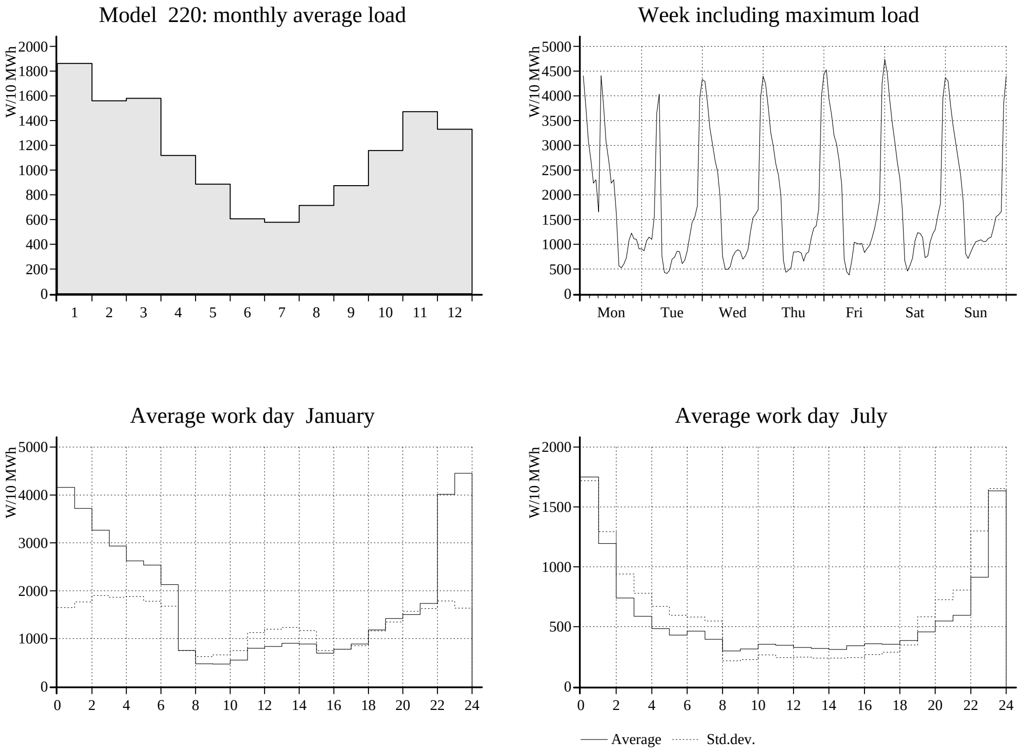


Fig. 7. Electric heat, partly storage, one family house. Model nr 220. Standardised to long term average temperature.

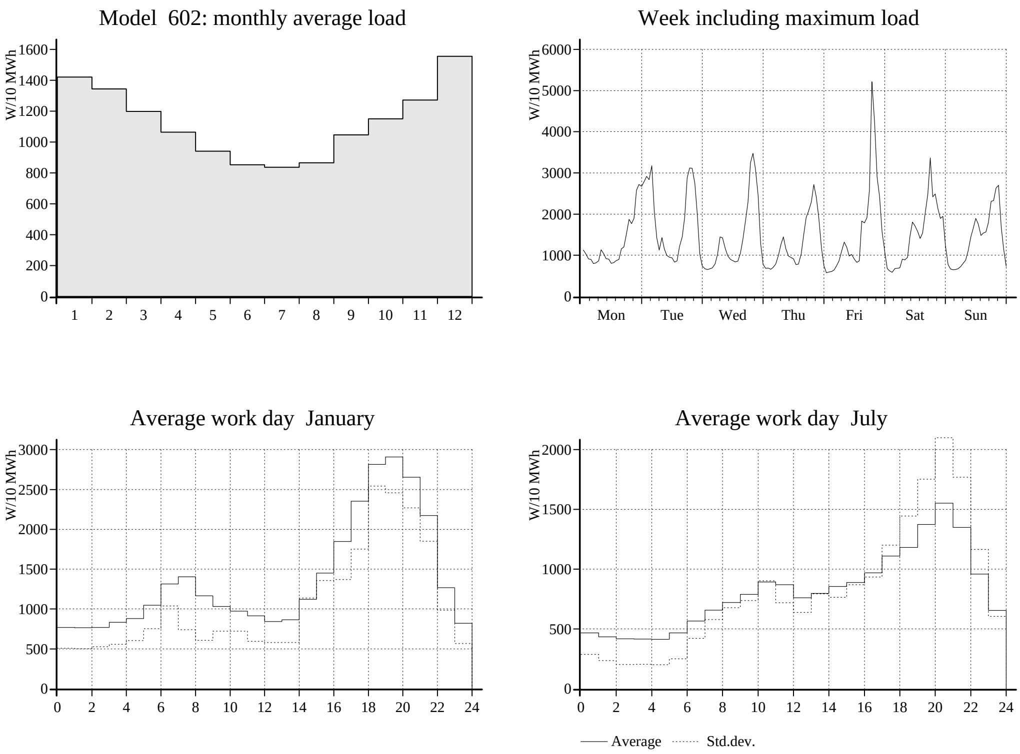


Fig. 8. Residential, one family house, electric sauna.

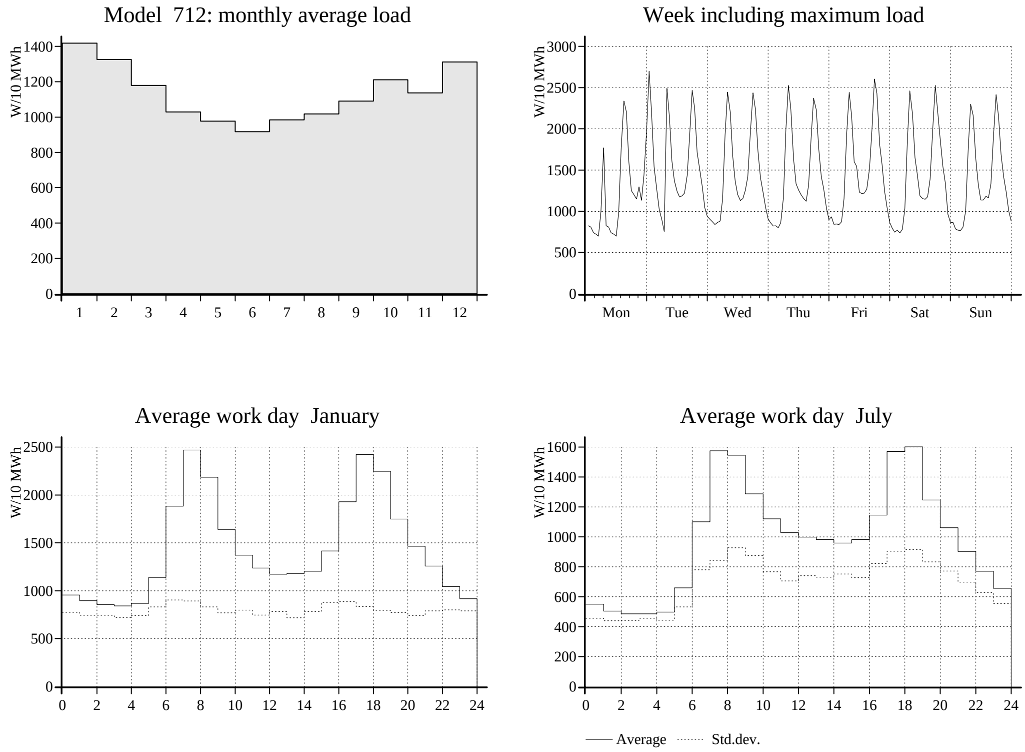


Fig. 9. Agriculture with milk production and residence consumption included.

4.2 Plots of load research sample data

The following figures present plots of load research sample data for a specific class, month, day-type and hour of day. Each value is plotted along the x-axis. These figures show the scatter of hourly loads in a sample and how the division with annual energy affects the distribution.

In each figure the above plot shows the original load data sample in watts. The plot below shows the same load data divided by customer's annual energy use and scaled as watts per 10 MWh/year.

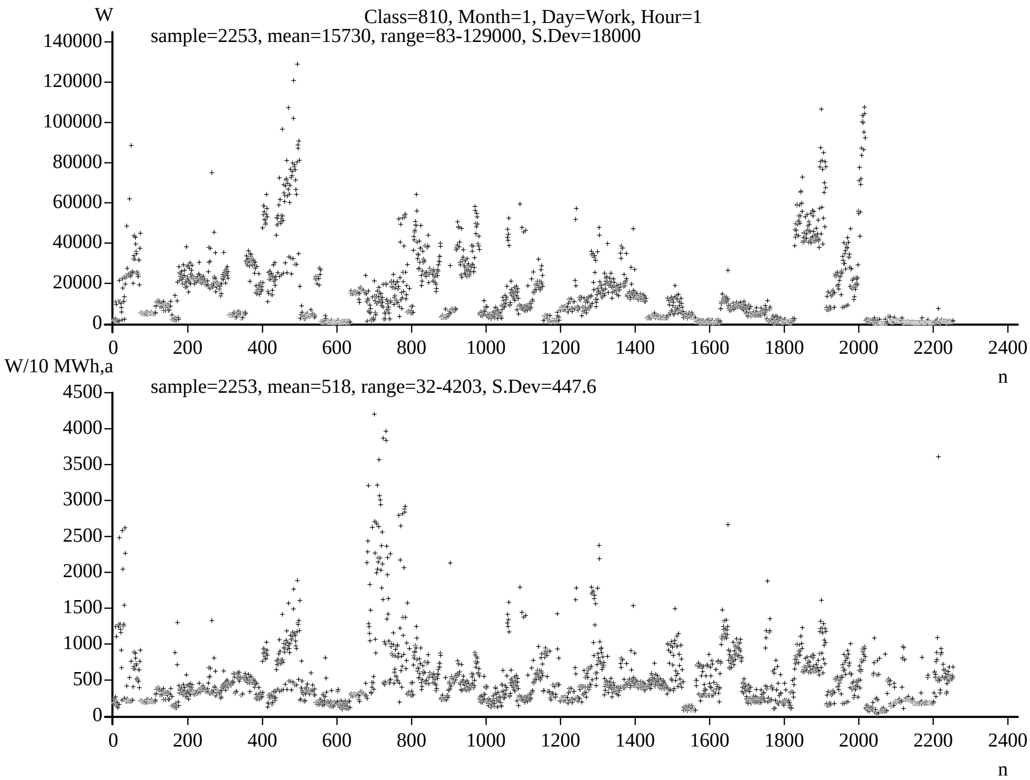


Fig. 10. Class 810 industry 1-shift, January, working day, hour 00.00-01.00.

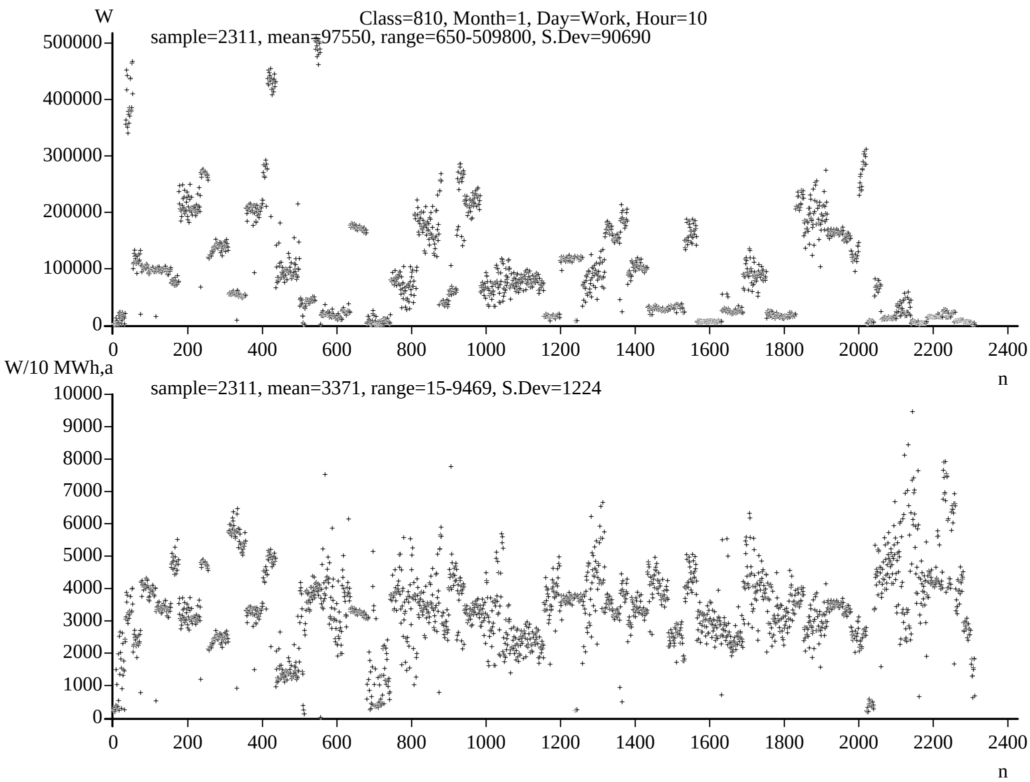


Fig. 11. Class 810 industry 1-shift, January, working day, hour 09.00-10.00.

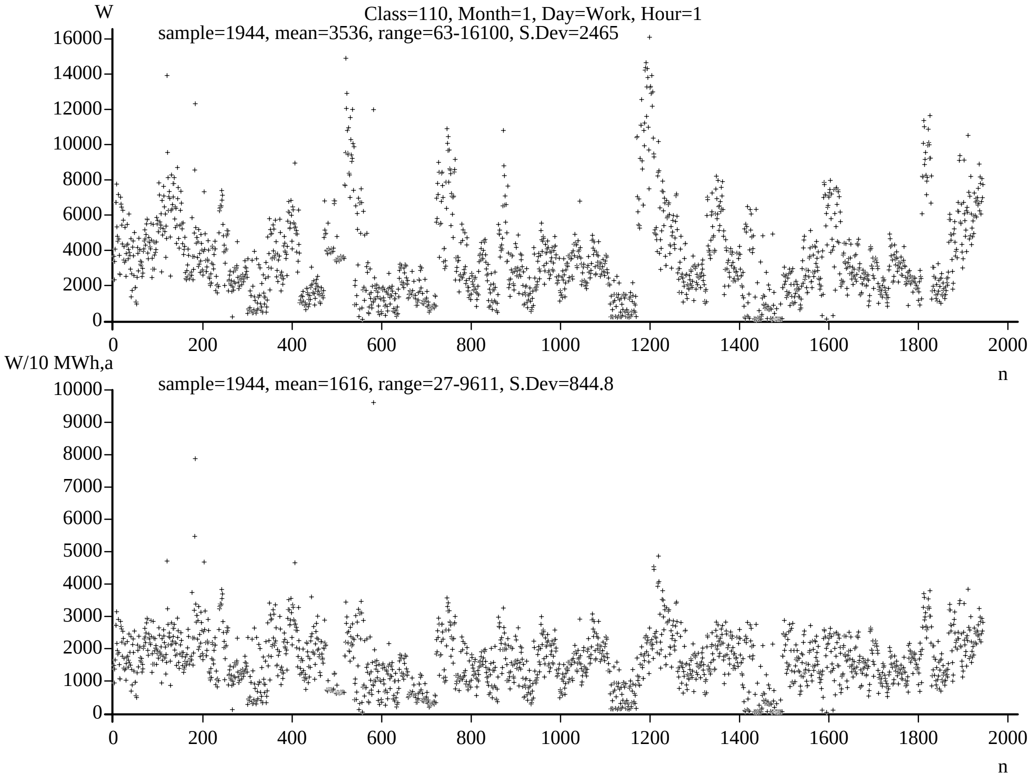


Fig. 12. Class 110 direct electric heating, one family house, January, working day, hour 00.00-01.00.

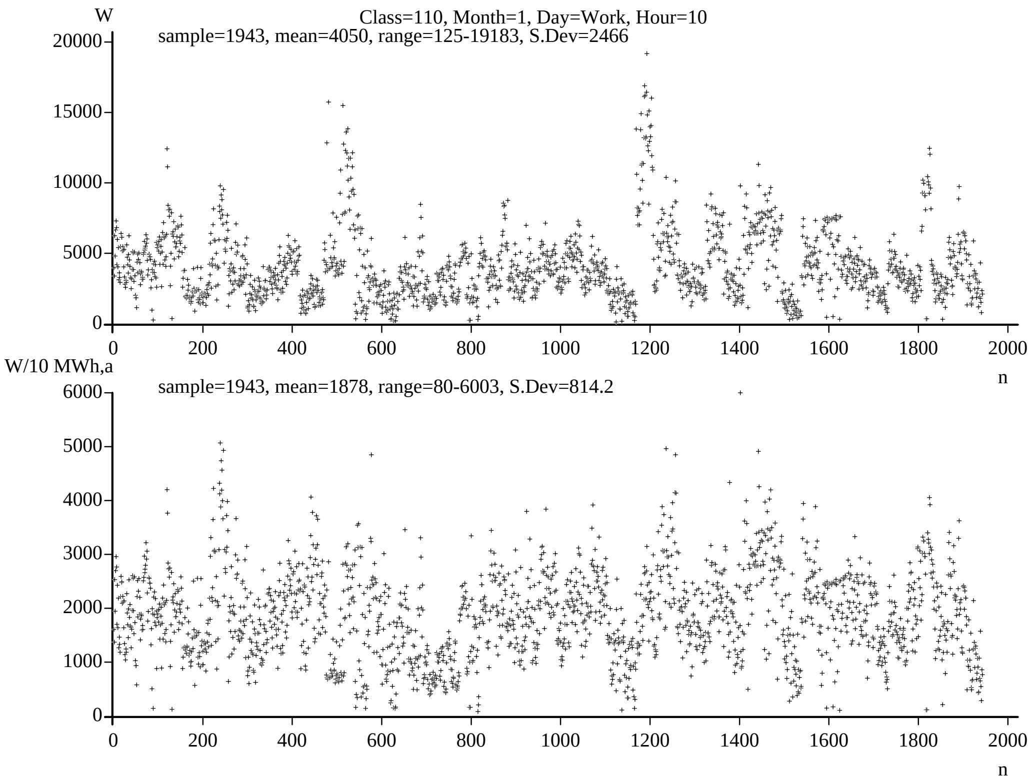


Fig. 13. Class 110 direct electric heating, one family house, January, working day, hour 09.00-10.00.

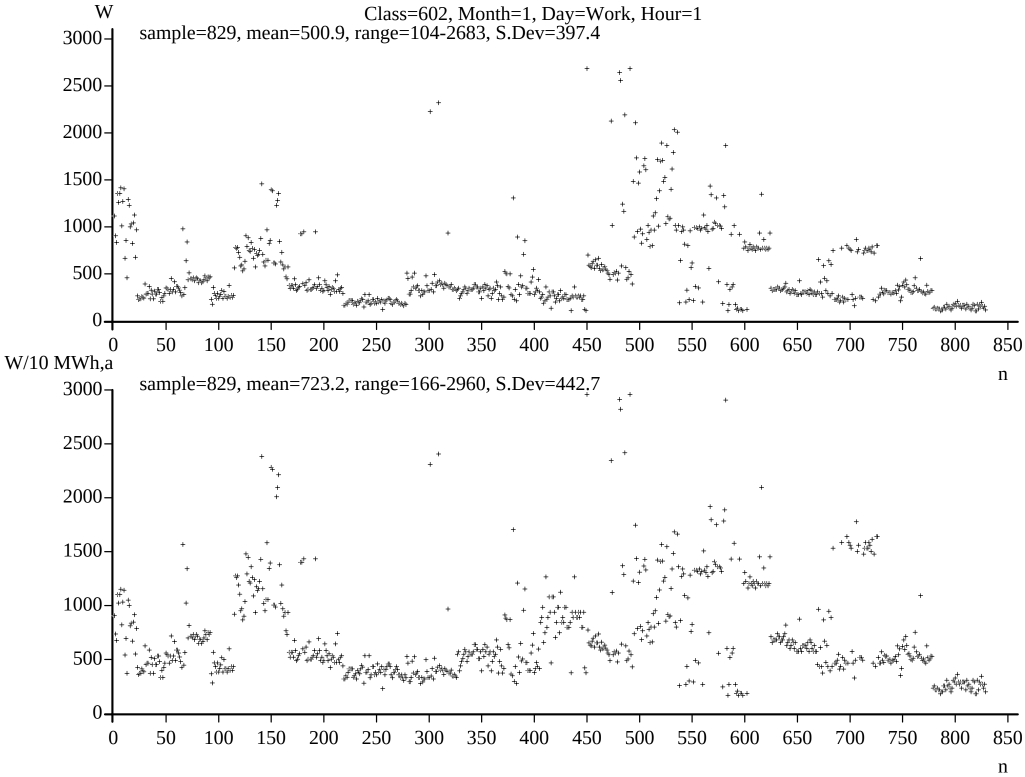


Fig. 14. Class 602 residential, one family house, January, working day, hour 00.00-01.00.

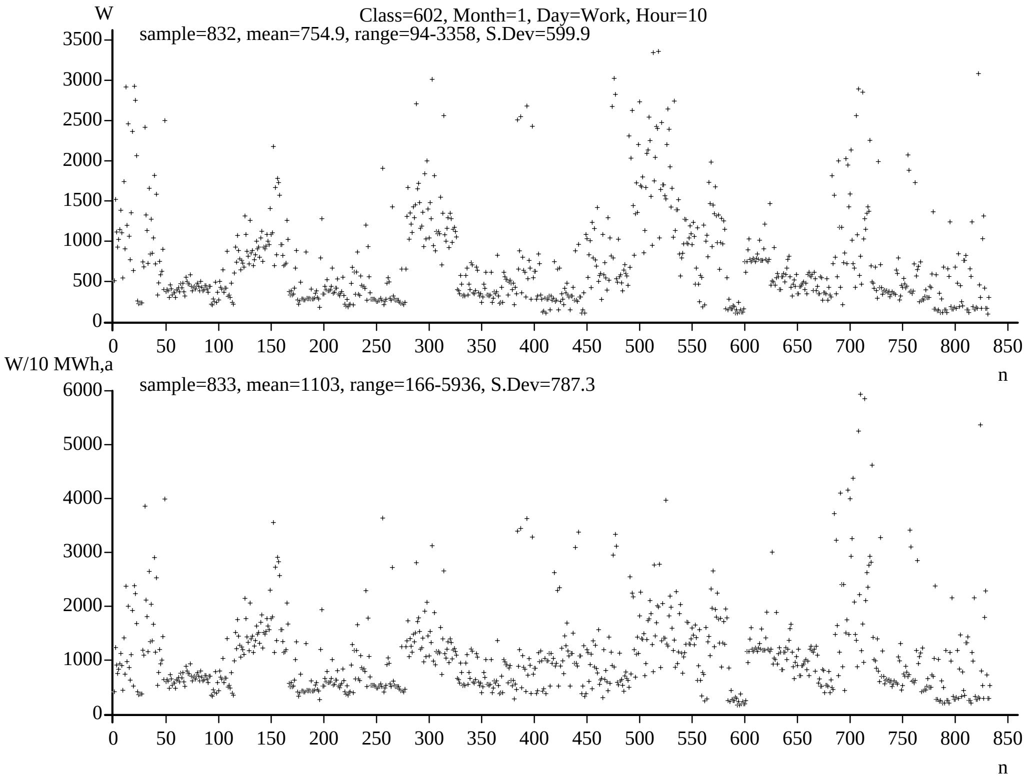
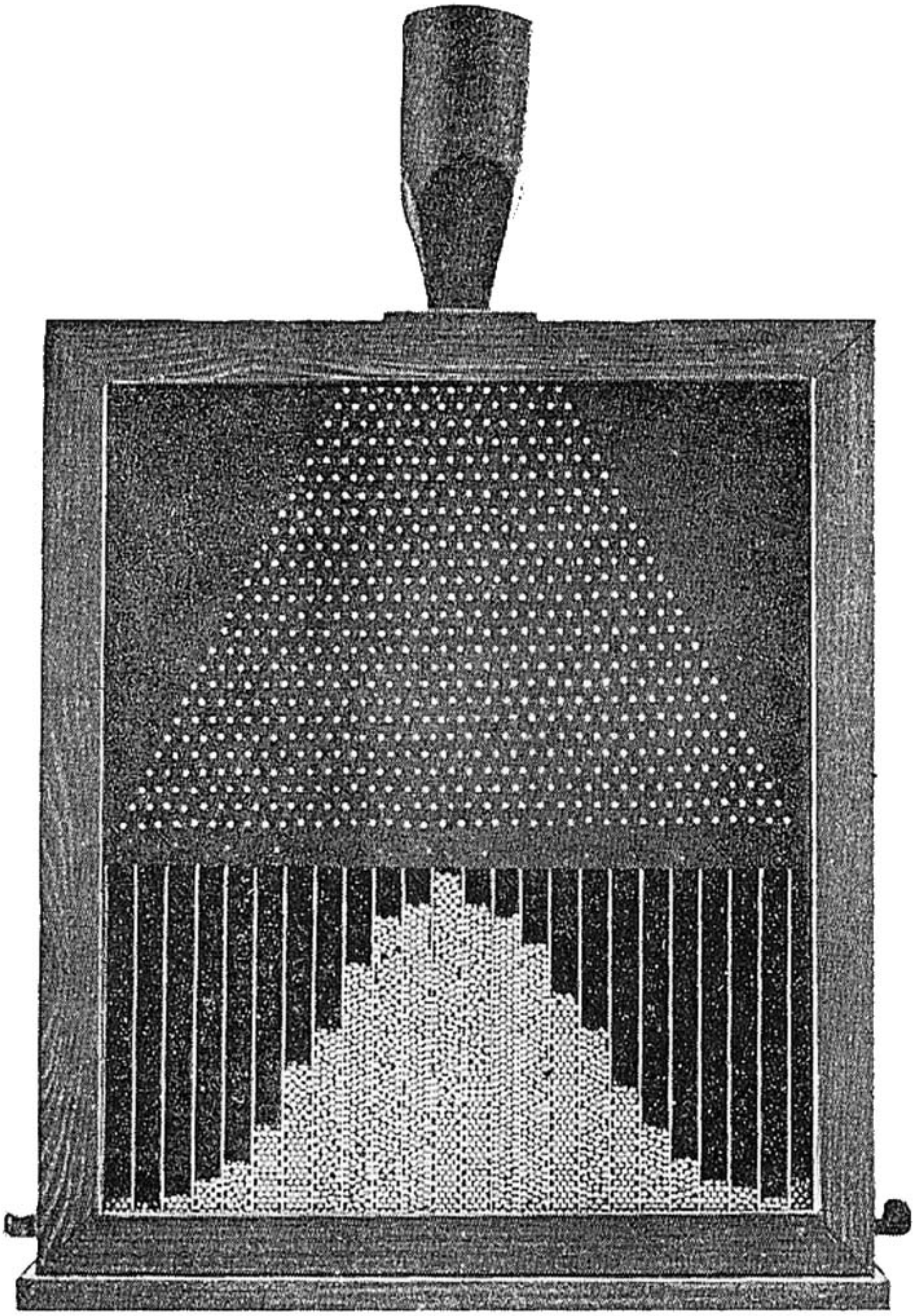


Fig. 15. Class 602 residential, one family house, January, working day, hour 09.00-10.00.

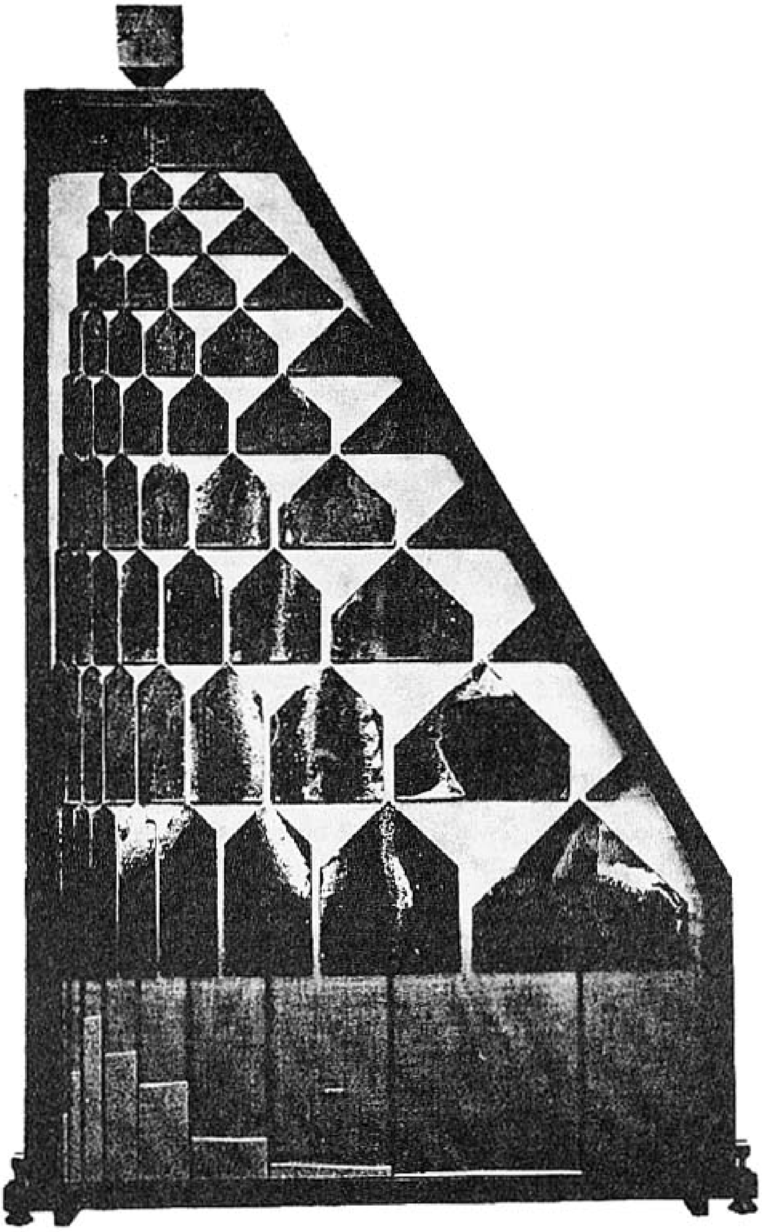
APPENDIX 5

Distribution machines

From the history of science we find methods which sometimes can give us another interesting view to a problem. These machines are nowadays replaced by computer programs. However the appearance of these machines gives better understanding to the physical origin of normal and lognormal distributions.



Galton’s normal distribution machine (Hald 1965, p. 32) from the book “Natural Inheritance” 1889. The apparatus consists of a board with nails of a given row being placed below the midpoints of the intervals in the row above. Steel balls are poured into the apparatus through a funnel, and the balls will then be “influenced” by the nails in such a manner that they take up positions deviating from the point vertically below the funnel. The distribution of the balls is of the same type as the theoretical distribution from a binomial process.



Kapteyn's skew distribution machine (Kapteyn 1916, fig. 7). The whole machine is 104 cm high. The pins of Galton's machine are replaced here with pentagon shaped deviators*, two sides perpendicular and the two upper ones inclined at a fixed angle (45 °) to the horizon. The deviators are of varying breadth. The breadth is proportional to the distance of the deviator from the left hand side of the machine. Sand is poured into a funnel situated at the top. The sand will arrive in the receptacles placed at the bottom of the machine and form a histogram approximating lognormal distribution.*