Load models for operarion and planning of electricity distribution networks with metering data

Ni Ding

UNIVERSITÉ DE GRENOBLE

THÈSE

POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE GRENOBLE

Spécialité: Génie Électrique

Arrêté ministériel : 7 Aôut 2006

PRÉSENTÉE PAR

Ni DING

THÈSE DIRIGÉE PAR **Yvon BÉSANGER** ET

CODIRIGÉE PAR **Frédéric WURTZ**

PRÉPARÉE AU SEIN DU

Laboratoire G2ELAB

DANS L'**ÉCOLE DOCTORALE: EEATS**

Load models for operation and planning of electricity distribution networks with smart metering data

THÈSE SOUTENUE PUBLIQUEMENT LE **30 Novembre 2012**,

DEVANT LE JURY COMPOSÉ DE:

Pr. Nouredine Hadjsaid

Grenoble INP , Président

Pr. Carlo Alberto Nucci

Université de Bologne, Rapporteur

Pr. Corinne Alonso

Université de Toulouse, Rapporteur

Pr. Didier Mayer

Mine de Paris, Membre

Pr. Yvon Bésanger

Grenoble INP, Membre

Dr. Frédéric Wurtz

CNRS Grenoble, Membre

INVITÉS:

M. Olivier Devaux

EDF R&D

M. Alain Glatigny

Schneider Electric

Contents

List of Figures

List of Tables

Acknowledgments

Achieving my PhD degree has set a milestone in my life. The thesis defense has drawn an end to my best and worst moments during the three years that I have spent in G2elab. People that helped me will always remain dear to me for my future life journey.

Above all, I would like to thank the reviewers and committee members of my thesis. A very big thank to Pr. Carlo Alberto Nucci and Pr. Corinne Alonso for their time and energy to examine my dissertation and for their valuable opinions. Their encouragements and appreciations give me strength and confidence in my future work. Thanks also go to Pr. Didier Mayer for his interests and insightful comments about my work. It was a great honor for me to have Pr. Nouredine Hadjsaid as the president of the committee, and I acknowledge him for that.

I'm also grateful to the representatives of the industrial partners of the company that I work with: Mr. Oliver Devaux from EDF and Mr. Alain Glatigny from Schneider Electric. I thank them for this interesting subject that they set up and pertinent comments regarding the industrial applications of my models.

I owe my sincere gratitude to my principal advisor, Pr. Yvon Bésanger. I would like to thank him for his open-mind regarding collaborations, for his support and trust, and for his patience in guidance. I thank him for always being there for me as a teammate at difficult times that we encountered through publications, and administrations.

I would like to extend my gratitude to Dr. Frédéric Wurtz for his trust and appreciation for my work, for his warmly welcoming me into the laboratory.

I would not forget to grant my gratitude to Pr. Gérard Dreyfus, Pr. Jean-Louis Lacoume and Pr. Daniel Baudois for their scientific guidance. I highly respect their passion and rigorousness for the research. I thank them to accept to offer me technical advices without reserve.

I want to express my gratitude to Mr. Christophe Keiny, Mr. Guillaume Antoine and Miss Leticia De-Alvaro from EDF for their energy devoting to my thesis project. They have been supportive industrial advisers to keep me on track with the industrial needs, and in the meantime give me great freedom to develop independent solutions.

I also would like to thank Mr. Frédéric Gorgette from ERDF, my supervisor of the traineeship, for offering me such a great opportunity to pursue my PhD.

My deepest gratitude reserves to my family and my friends. Even though none of my family member could attend my thesis defense, their love and care are always around me. I want to thank my friends in G2elab for their tolerance and unconditional support to me during those years. I have never met so many great people during so short period of my life. I will treasure our friendships for a lifelong time.

You are responsible for what you have tamed. - "The little prince"
To them, I dedicate this work.

Acronyms

ADA Advanced Distribution Automation

ANN Artificial Neuronal Network

AI Artificial Intelligence

AR AutoRegressive

ARMA AutoRegressive Moving Average

ARIMA AutoRegressive Integrated Moving Average. Equation 2.5

ARMAX AutoRegressive Moving Average with eXogenous inputs

ARIMAX AutoRegressive Integrated Moving Average with eXogenous inputs

ACF AutoCorrelation Function. Equation 2.6

AFSA Artificial Fish Swarm Algorithm

ANOVA ANalyse Of VAriance

ADF Augmented Dickey-Fuller test

AMR Automatic Meter Reading

CV Cross-Validation. Equation 6.10

CDF Cumulated Distribution Function. Equation 3.13

CRLP Class Representative Load Pattern

CUSUM CUmulative SUM. Equation 6.3

DSO Distribution System Operator

DMS Distribution Management System

DG Distributed Generators

DWT Discrete Wavelet Transform. Equation 2.17

DFT Discrete Fourier Transform. Equation 3.6

DLE Distribution Load Estimation

ERDF Electricité Réseau Distribution France

EMS Energy Management System

FL Fuzzy Logic

FCM Fuzzy C-Means

GDP Gross Domestic Product

GA Genetic Algorithm

GEV Generalized Extreme Value

GPD Generalized Pareto Distribution

HV/MV High Voltage/Medium Voltage

HV High Voltage

HW Holt-Winters

IA Immune Algorithm

ISODATA Iterative Self-Organizing DATA-analysis technique algorithm

KNN K-Nearest Neighbor(s)

KPSS Kwiatkowski-Phillips-Schmidt-Shin tests

KDE Kernel Density Estimation. Equation 6.1

LV Low Voltage

LFC Load Frequency Control

LTLF Long-Term Load Forecast

LOO Leave-One-Out. Equation 4.6

LL Local Linear. Equation 6.9

LL2 Adapted Local Linear

MV Medium Voltage

MV/LV Medium Voltage/ Low Voltage

MTLF Medium-Term Load Forecast

MAPE Mean Absolute Percentage Error. Equation 2.32

MA Moving Average

MLE Maximum Likelihood Estimation

MLP Multi Layer Perceptron

MAE Mean Absolute Error. Equation 2.33

MSE Mean Square Error

NN Neuronal Network

NARMA Nonlinear AutoRegressive Moving Average

NW Nadaraya-Watson. Equation 6.6

OLS Ordinary Least Square

PDF Probability Density Function

PARMA Periodic AutoRegressive Moving Average

PACF Partial AutoCorrelation Function

PSO Particle Swarm Optimization

PRESS Predicted REsidual Sum of Squares. Equation 4.6

PNN Probability Neural Network

RNN Recurrent Neural Network

RBF Radial Basis Function

RBFN Radial Basis Function Networks

RLP Representative Load Pattern

SE State Estimator

STLF Short-Term Load Forecast

SOM Self-Organizing Maps

SLP Single Layer Perceptron

SVM Support Vector Machine. Equation 2.27

SVR Support Vector Regression. Equation 2.27

SCADA Supervisory Control And Data Acquisition

SSR Sum of Square Residuals

SSE Sum Square Error

TLP Typical Load Profile

THI Temperature-Humidity Index

TMB Minimum Temperature Base

VVC Volt VAR Control

VLOO Virtual Leave-One-Out

VSTLF Very Short-Term Load Forecast

WCI Wind Chill Index

WNN Wavelet Neuronal Network. Equation 2.18

WLSE Weighted Least Squares Estimation

i.i.d. independent and identically distributed

Mathematical Notations

x Influence variable vector or variable to be determined

X_i Sampled values, measurements

X Observaton matrix, whose element x_{ij} is the measured value of variable j in example i

y_t Load (power) measured at time t

y Load vector

y_i Sampled load value

ϵ_t Model's noise at time t

e Difference between output of the model and measured value

$\gamma_y(t, t-\tau)$ Autocovariance function of y process at the time t and $t-\tau$

$E(\cdot)$ Expected value operator

$\{P_i, y_i\}$ Historical data inputs/outputs pair, learning set or training set, $i = 1, \cdots, N$

f_t Trend model value at time t

S_t Cyclic model value at time t

$D_\alpha, \alpha = 1, \cdots, \kappa - 1$ Dummy variables, where κ is the number of different categories

$\gamma_\alpha, \alpha = 1, \cdots, \kappa - 1$ Dummy regression coefficients

T_t Temperature at time t

W_t Detrended series

$p(\epsilon)$ Probability density function of the random variable ϵ

$F_\epsilon(x)$ Cumulative distribution function of the random variable ϵ

P Vector variables $\{p_j, j = 0, \cdots, R\}$ of neural networks, where R is the total number of input variables

Ω_i Vector of the parameters (or weights) $\{\omega_{ij}, j = 0, \cdots, R\}$ of hidden neuron i

Ω Set of the parameters of the neural network model

ω Vector of weights of the linear combination, between the hidden layer and output layer of the neural network model

C Vector $\{c_i(P,\Omega_i), i = 1,\cdots,M\}$ of the outputs of hidden neurons, where M is the number of hidden neurons

r_0 The threshold rank of the orthogonal forward regression

r_{probe} The rank of a probe variable

ξ_i The i-th candidate variable vector

$f(P_i,\Omega)$ Output of the neural network with respect to the variable vector P_i and the parameters Ω

$f^{-i}(P_i,\Omega)$ Output of the neural network model when example i is withdrawn from the training set

n_p Number of realizations of the probe variable

n_{rp} Number of realizations of the random probe whose rank is smaller than or equal to rank r

δ Risk chosen by the designer to control the number of inputs

h_{ii} Leverage, i-th diagonal element of the hat matrix H

p Number of set of parameters of the neural network model, which is equal to $(R+1)M + (M+1)$

E_{LOO} Leave-one-out score

E_p Approximation of the leave-one-out score

Z Jacobian matrix of the neural network model

E_{yr} Yearly energy consumption

E_0 Non-heating daily energy consumption

s Temperature sensibility, indicating the amount of energy consumed (kWh) by decreasing 1 ^{o}C temperature

E_n Annual energy consumption adjusted to the normal climatic condition

$P(t)$ Estimated mean power at time “t”

$\sigma(t)$ Estimated standard deviation at hour “t”

$\nu(t)$ Estimated margin at time “t”

$a(t)$ Common group coefficient for the BAGHEERA model converting non heating daily energy into non heating power at time “t”

$b(t)$ Common group coefficient for the BAGHEERA model converting heating daily energy into heating power at time "t"

E_d Daily energy consumption

E_i Meter recording energy consumption during n_i days

Dd_i Degree days during a period of n_i days

$Dd_{\mathbf{365}}$ Yearly degree days in the normal climatic condition

T_d Daily average temperature

T_{Nh} Non heating temperature, a temperature threshold below which the consumption rises due to the electrical heaters

$\hat{g}_h(x)$ Kernel density estimator of variable x, with smoothing parameter h

h Smoothing parameter

$K(\mu)$ Normal kernel function of variable μ

γ Excess probability of the load estimation model

h_{cv} Optimal smoothing parameter defined by CV technique

y_{TMB_i} Estimated power at TMB condition

$\hat{f}_{h'}(\cdot)$ Kernel-type estimator with its optimal smoothing parameter h'

$P_{threshold}$ Estimated threshold power

Chapter 1

General introduction: the new problematic of load models in the smart grid context

Contents

Abstract

Ground-breaking evolutions have been brought to traditional electrical distribution grids by the concept of "smart grids". The smart meter system, as one of the most important infrastructures in the smart grids, gives us detailed information on electricity consumption of an individual customer. In this context, we aim at designing forecasting models and estimation models based on these information for needs in distribution network operation and network planning. The contributions, a quick overview of the scope, and the organization of this dissertation are also presented in this chapter.

1.1 Background: smart grid and smart meters for load modeling

The smart-grid concept combines advanced communication technologies with traditional electrical distribution grids in order to improve the transparency and the controllability of distribution grids. Facing several ground-breaking evolutions in the electricity systems, such as the large penetrations of the renewable power generation, the rapid load growth due to plug-in electric vehicles, to name a couple, numerous advanced algorithms appear in this circumstance to enhance the stability and the efficiency of the system. These Advanced Distribution Automation (ADA) functions include *Volt VAR Control (VVC)* [1], *self healing*, and *direct load control* [2] (to name a few). The ADA functions are calculated in real-time or in ahead of time in order to help making decisions. Generally, the monitoring and the control process of distribution networks are performed at the Medium Voltage (MV) level.

One of the smart-grid goals is to make distribution systems economically efficient with reliable energy supplies and less costs. Distribution network planning involves developing a schedule of future additions that ensure the quality of energy delivery as well as the lowest possible cost. On the one hand, the electricity infrastructure must meet the needs of peak loads. On the other hand, over-dimensioned systems can be very expensive. Thus, reliable load estimation models are required to tighten distribution margins and optimize the planning investment by performing distribution network calculations, i.e., carrying out the power flow calculation in critical situations so as to identify poor electricity supply zones. Nevertheless, the complexity in the problem is related to the uncertainty and randomness in the clients' electricity consumptions.

In the current state, the scarcity of measurements on the distribution system introduces bottlenecks in carrying out the ADA functions as well as the network optimization calculations. The available measurements in distribution networks are mainly on the secondary of source substations. It is economically non-feasible to implement electric meters in all 738 000 Medium Voltage/ Low Voltage (MV/LV) substations. Today, for the operation need, applying very approximate probabilistic models with 50% of precision seriously affects the efficiency of the ADA functions, resulting dubious analysis results. In order to comply with the planning need, the actual model applied by the French electricity company, termed BAGHEERA, depends mainly on the client's individual information, which becomes less and less available. Thus, a new model must be designed at the request of replacing the BAGHEERA model.

Starting from 2010, the Electricité Réseau Distribution France (ERDF) (French Distribution System Operator (DSO)) launched the "Linky" (baptized name for the smart meter in France) project, which aims at installing 35 000 000 smart meters in France. On the one hand, end users will pay electricity bills based on their real consumptions rather than on the estimated ones as in the today's case in France. On the other hand, thanks to these measurements, distribution network operators can have a better vision of the current situation on networks. Actually, there are no available measurements on MV/LV

substations on the French distribution networks. In the experimental phase of the "Linky" project, the consumption information of each individual is sampled on a 30-minute basis and transferred once a day to the correspondent data center. However, as data are gathered in packages and sent with a certain frequency [3], some delay is found in the measurements.

Therefore, using the accurate information provided by the smart meters to develop load models is the silver bullet that makes key smart-grid applications feasible.

1.2 Motivation and objectives

The supervision of the power and voltage dispatching of the networks is a critical task in distribution exploitation. It guarantees an economical optimum and a dynamic stability of the networks. Unlike transmission networks, on which abundant measurements exist, distribution networks have much less measurements. As a matter of fact, because of the complex structure and a great number of nodes (MV/LV substations) in distribution networks, it is economically impossible to install meters in a great quantity on these substations in distribution networks. Thus, the distribution system is considered as "blind" or "non observable". One solution to improve the " observability " of distribution networks is to introduce load models in order to replace the measurements.

In terms of loads in distribution networks, we distinguish two types:

- MV clients directly connected to MV networks
- Numerous Low Voltage (LV) clients connected to MV networks through the public MV/LV substations

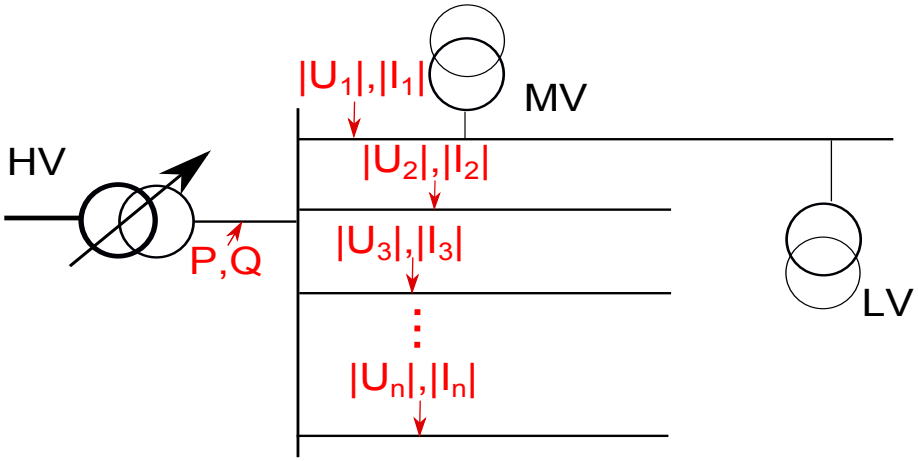


Figure 1.1: Available measurements (marked in red) [4] in the French distribution networks. $|\cdot|$ represents the norm notation, equivalent to the magnitude.

Currently, the measurements in the French distribution networks are (figure 1.1):

- The active and the reactive power on the secondary High Voltage/Medium Voltage (HV/MV) substations
- The mean voltage value on head of every MV feeder sampled every 10 minutes

- The magnitude of the current on head of every MV feeder
- The active and the reactive power of some MV clients

On the LV clients' side, the only available data are limited to the subscribed power in the supply contract and billing information of the clients connected to the public MV/LV substations.

With the new available individual consumption data collected by smart meters, the objective of the research program presented in this thesis is to build new load models for the need in operation and in planning in distribution networks. This context makes possible the design of accurate models for the distribution network planning, monitoring and control, in absence of the costly measuring equipments in distribution networks.

1.2.a For network operation need

For the sake of control and configuration in distribution systems, the evolution of the MV/LV substation load needs to be known. Mainly, we can point out three different reasons described as follows:

- During a failure: in order to efficiently restore electricity in regions where a fault occurs, loads in the affected regions should be known in the following three minutes.
- During network maintenance: the variation of the consumption needs to be known to restore the power supply. Generally, a two-day period is considered by the French electricity distributor ERDF as a normal repairing time. In this case, a two-day load forecast with its standard error is needed.
- As inputs for the SE: the SE [5] is the core function of any energy management system. It aims at estimating the network variables, such as the voltage magnitudes and angles. Figure 1.2 shows the schematic of the relationship among forecasting models, SE, and ADA functions. The forecasting models as well as the network data are considered as inputs for the SE. The network data [5] includes the information about the network topology, line resistance, reactance, tap setting, and line charging, etc. The output of the SE will lead the Distribution Management System (DMS) control scheduling block to perform concerned ADA functions for operational decisions. These decisions enable the monitoring and control of various devices in the networks such as capacitor banks, Distributed Generators (DG), on-load tap changing transformers, and switches/breakers, etc.

The idea is then to design forecasting models for MV/LV substations relying on the aggregated smart metering data for the next two days.

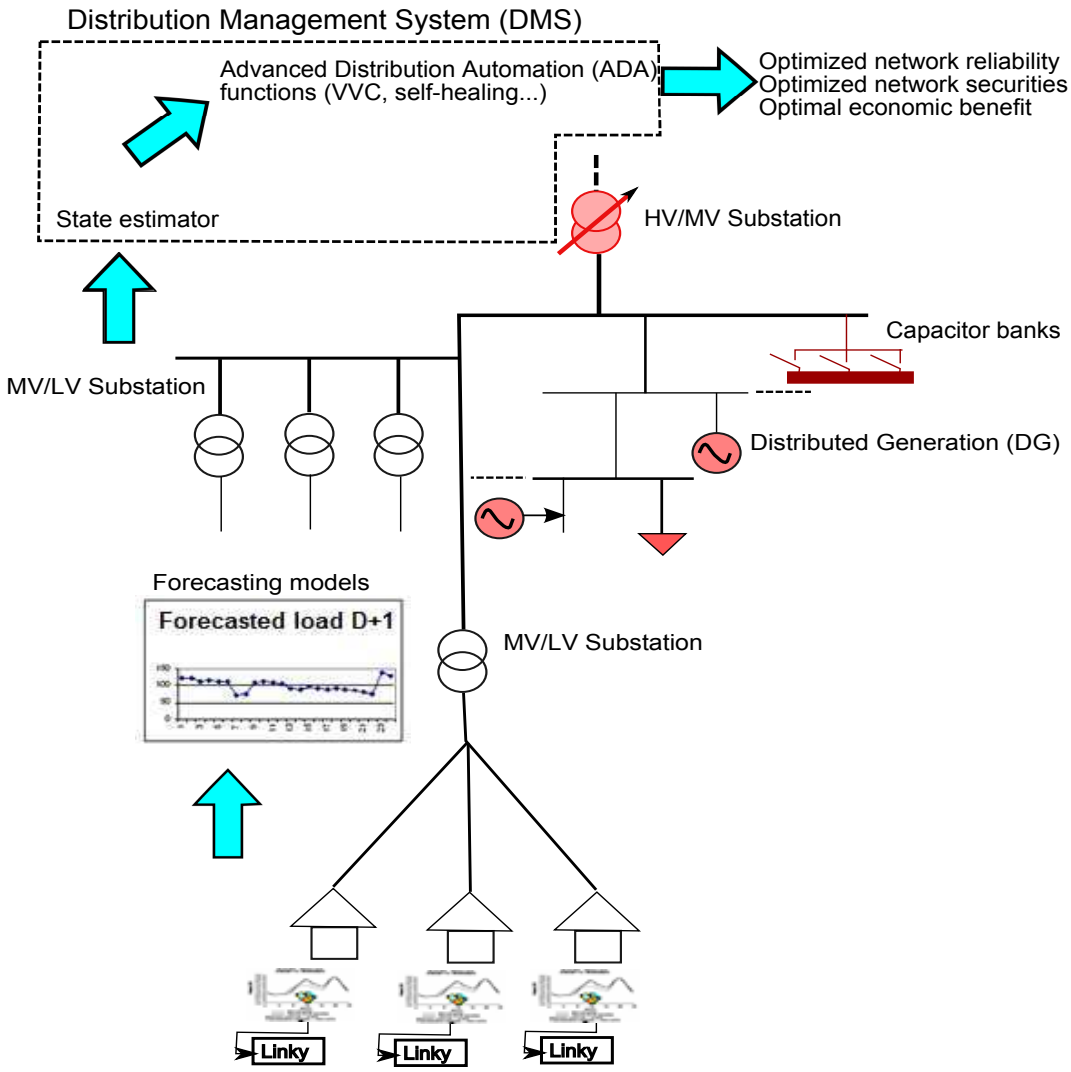


Figure 1.2: Relationship among forecasting models, SE, and ADA functions

1.2.b For network planning need

Designing a reliable distribution network is challenging since it needs to guarantee a stable and continuous power supply to the customers. As a matter of fact: *a utility must maintain the voltage delivered to each customer within a narrow range centered within the voltages that the electric equipment is designed to tolerate* [6]. In the European electricity regulation, for the LV networks, a 10% out-of-range voltage is acceptable. Beyond this range, the customer is defined as a "poorly supplied customer".

For the sake of planning, network calculations are performed under extreme situations in order to handle worst case scenarios [7, 8]. More specifically, these calculations are carried out when either of these two cases occurs: maximum demand with minimum generation, and maximum generation with minimum demand [9]. With the arrival of a considerable portion of DGs into the networks, the later case can be expected. Because of the customers' various behaviors, in a same geographic zone, peak demands of different customers seldom happen at the same moment. Therefore, estimation of customer daily

load pattern at every hour is required. Uncertainty of the load estimation also needs to be considered [10]. Generally, for the voltage-drop calculation, the excess probability, which defines the threshold power bounds, is fixed to 10% [11]. Consequently, in the network planning, a client's total supply demand includes the client's daily load pattern and his 10% excess probability uncertainty.

Moreover, in the distribution network planning process, a list of standards and criteria for equipments, such as distribution lines and MV/LV transformers, to name a couple, is also defined. Reliable load models are also required in the following cases in order to carry out distribution network calculations:

- Distribution line losses
- Currents in line segments
- Network voltage-drops
- MV/LV transformers: two-hour equivalent power[1], voltage-drops, and electrical losses

Therefore, the objective is to define an individual customer's maximum and minimum load limits of a year with 10% excess probability.

1.3 Contributions of the thesis

The major purposes of this dissertation are twofold: designing new load forecasting models for the network operation and load estimation models for the network planning in distribution networks. The contributions of the thesis can be summarized as follows:

- Load forecast is an extensively investigated subject on the transmission level [12, 13, 14, 15, 16, 17, 18, 17, 19, 20]. However on the distribution level, with the same characteristic consumption data (several dozen kW), to the best of our knowledge, few works have been done.

  We can think of three plausible reasons explaining this fact: first, more attention has been paid to the transmission level, as the transmission grid extents on longer distances and covers larger territories. The transmission grid, involving huge costs, is the backbone of the power systems. Second, the higher voltage level consumption has a more regular load curve pattern, which makes it easier to forecast. Third, there were no available measurements on the MV/LV substations relying on which the forecasting models can be designed and validated.

  In this research project, two models based on time series and neural networks have been proposed, for the network operation need.

- The different load types (residential, commercial, and industrial) are examined in this research, and their different properties are pointed out.

[1] Define a transformer's capacity

The two methods are designed and evaluated on *real measurements* collected from the French distribution grids in the framework of the "Linky" project. Reference case is established in order to be compared to the two proposed models. Advantages and drawbacks of the two methods are drawn through the comparison. The two methods are alternative and, at the same time, compliment to each other to set the precision limit due to the intrinsic characteristics of the substation load data.

- Time series method presented in this dissertation is an original work, where numerous statistical tools are integrated in for the pursuing of precision. Residual of the model is looked through in details to ensure the model's well being.

- Neural network method, on the other hand, is inspired by the selection procedure proposed by Gérard Dreyfus [21], a renowned expert in the neural network modeling. We focus on the Neuronal Network (NN) model design, which has been fully exploited for the first time in the short-term load forecast.

- Until the implementation of the smart meters, usually there were no available historical load data besides the demand survey data on a limited number of clients. In the load modeling field, most of the works concerning the distribution network planning aim at estimating the peak demand for a group of customers during the peak demand of the system, namely the coincident peak demand [6]. For this purpose, some researches based on end-use method [9, 22] decompose the load model of the residential customer into appliance elementary units. Others focus on the classification method, sorting customers into different categories, and on the representation each category with a Typical Load Profile (TLP). In the smart metering context, we are the first to propose the concept of individual data-driven load model for customers, for the network planning need.

- To build an individual estimation model, the relationship between the electricity consumption and the temperature is deduced by nonparametric estimators. The method is applied to real consumption data of individual customers in France. Performance is compared to the current load model (termed BAGHEERA) of the electricity company EDF through different validation studies.

1.4 Scope and organization of the dissertation

The dissertation is organized as follows:

- Chapter 1 states that the smart grid and the smart meters are in rapid development to improve the efficiency and the controllability of distribution networks. The revolutionary changes in distribution network systems urge the accurate load forecasting models. Recording individual consumption information, the smart meters enable the building of these models. The thesis is encouraged by the "Linky" project launched by the French electricity distributor ERDF, aiming at installing 35 000 000 smart meters in France. Two main objectives are evoked in the smart metering context: short-term load forecasting models for the network operation and the estimation models for the

network planning. Reasons for the development of these two objectives are declared. Contributions of the thesis are highlighted.

The rest of the dissertation is divided into two parts, dealing with two distinct objectives. Part A tackles the forecasting load models for the operation need, and includes chapters 2,3 and 4.

- Chapter 2 gives a review of the load forecasting methods in literatures and set up basic framework for the performance evaluation. Load forecasts are stratified into different objectives according to their lead time scale. Different objectives are devoted to different applications. More input information is needed for a longer lead time forecast. Forecasting methods are divided into two categories: the classical approach and the artificial intelligent approach. Hybrid models that possess the advantages of both categories gain more and more popularity in the applications. Data used in our study are thoroughly analyzed, suggesting influence factors to our short-term load forecasting models. We argue for our choices of time series and neural network methods. Performance criteria and a reference model (termed naive model) are also established in this chapter, building a solid framework base for the presentation of the applied methods in the next two chapters.

- Chapter 3 presents the short-term load forecasting model based on the time series method. The forecasting procedure is detailed. The additive time series model contains three components: a trend, a cyclic and a random error. The first two components are deterministic, and are designed into models respectively. The trend model is temperature-dependent, linear, and dummy variables integrated, indicating day types. Cyclic model is composed by the Fourier components, whose frequencies are found by a smoothed periodogram. Numerous statistical tools are applied pursuing a better precision: sliding window strategy is adopted so that the model is updated during each forecasting period; ANalyse Of VAriance (ANOVA) nullity test is applied to estimate the significance of variables. Available data are divided into a learning set and a test set. Important parameters of the model are defined thanks to the learning set. Whereas the test set is reserved for the performance evaluation. Residual is examined, making sure the well fit of the models. Weather uncertainty impact on the precision of the forecasting model is discussed in the end of the chapter. We concluded that even with the weather uncertainty, our proposed time series model still outperforms the naive model.

- Chapter 4 introduces the neural network model from the artificial intelligent approach family. In our study, we focus on the design of the neural network model, choosing the optimal model that has the best achievable predictive ability. A general concept of the machine learning technique is stated, and difficulties such as finding the relevant variables and the bias-variance dilemma, are explained. The orthogonal forward regression and the Virtual Leave-One-Out (VLOO) technique are proposed as solutions to these difficulties. Experiments on the same data show that the proposed methodology behaves better than the time series model in term of accuracy. A comparison in divers properties is made between the two models in the end of this chapter.

Part B introduces the load estimation problem for the planning need, and proposes solutions. It contains chapters 5 and 6.

- Chapter 5 focuses on the load research projects in distribution networks. Load research projects aim at providing hourly load estimation models for individual client. Three steps, i.e., technical analysis, economical analysis, and final decision, for the decision makings in distribution network planning are presented. The outputs of the load research projects contribute to the technical analysis, devoting to finding solutions in network planning. The mechanism of the aggregation of loads, the coincident load, is explained. Common methods of finding TLP, which represents the daily load pattern of a certain group of clients, are presented. Load models used by electrical unities in Finland, Denmark, Norway, and Taiwan are described. Finally, the components of the BAGHEERA model applied by the French DSO are detailed and the method is demonstrated with real measurements.

- Chapter 6 proposes a novel approach for the individual load estimation in the context of smart meters. With the abundant individual consumption information, in our opinion, the load model is ready to be individualized rather than estimated through the TLPs. Thus, in this chapter, the individual load estimation model based on the nonparametric estimators is put forward. Numerous statistical tools, such as binary hypothesis tests, kernel density estimation, CUmulative SUM (CUSUM) algorithm, and Cross-Validation (CV) technique, are integrated in the proposed method. Three kernel regressors, i.e., Nadaraya-Watson (NW), Local Linear (LL), and Adapted Local Linear (LL2) are applied to deduce the relationship between the load and the temperature variations. Different application cases according to the quality and quantity of the data are suggested. The method is illustrated with real measurements and compared with the BAGHEERA model. The validity of the method is examined with extensive examples. In the end of the chapter, a discussion on the definition of the uncertainty bound of the estimation model is carried out.

- Chapter 7 concludes the dissertation and proposes perspectives for the future research interests.

Part A

Short-term load forecasting models for monitoring and state estimator

Chapter 2

Load forecasting techniques and short-term model framework

Contents

Abstract

Load forecast plays an important role in decision makings in power systems. This chapter begins with the review of load forecasting models in literatures. The second part of the chapter contributes to a framework consisting of data description, method selection, performance criteria, and reference case introductions. In the reviewing part, we classify a wide range of approaches of load forecast into two categories: classical approach, and artificial intelligent approach. Methodologies in each category are briefly presented. Their advantages, disadvantages, applications and pertinent research works are also developed. Popular hybrid models combining two or more different approaches are also involved. In the framework part, data used for the design and the evaluation of our methodologies are analyzed. Certain behavioral "components" in the data are pointed out. The choices of the methodologies based on the time series and the neural networks are argued. These two methodologies are going to be detailed in the following two chapters. Performance criteria and reference case are stated so as to lay a good foundation for the presentation of the models in the next two chapters.

2.1 Literature review

The quality of the decision making in electric power systems strongly depends on the accuracy of the power load predictions. Various decisions require reliable and accurate load forecasting models with different time-scales as well as on different hierarchical levels in network systems [23].

A wide range of approaches have been proposed to the load forecasting problems. In this section, we aim at presenting briefly the different approaches found in literatures, their specificities, applications and techniques applied to load forecast.

The organization of the section is as follows: first, we start by introducing definitions, applications and influence factors of different lead time load forecasts. Then, a two dimensional digest in lead time and in voltage hierarchy scales summarizes load forecasting methods, followed by the descriptive presentation of every method. Related works are also depicted. Finally, we conclude the literature review in a table.

Note that for the sake of clarity and ease of understanding, mathematical notations in the referenced works and internal reports have been adapted in order to keep coherence through the entire dissertation.

2.1.a Forecasting lead times and influence factors

Different forecasting lead times result into different forecasting models as well as their input variables. Numerous factors, such as weather conditions, seasonal effects, and social, economic, demographic factors explain the variations in the load [23]. Table 2.1 summarizes the applications and influence factors for different time horizon forecasting models [23, 24, 18].

Notice that more input variables are included when the time horizon becomes longer. For a VSTLF, univariate (only the historical power samples are considered as inputs) models can offer satisfactory results. These VSTLFs often participate to improve the efficiency and reliability of the real-time electrical systems. For longer lead time forecast, multivariate models with exogenous variables are favored. The STLF needs mainly three categories of inputs: weather, calendar, and historical variables [25]. Due to some measurement delays, or the computational time for the execution of the ADA functions, the STLFs often replace VSTLFs to fulfill needs in network operations. STLF forecasts also help reducing equipment failures and system blackouts by indicating the operational margins in power systems. The MTLF models help making financial decisions, such as evaluation of the price of energy products and investment interests. In such cases, the forecasting models need additional inputs as social and economical factors. The LTLF, concerning energy system capital expenditures and more important economic investments, needs to take into account more socio-economic factors, and sometimes even their future evolutions.

The herein description is in a general way, as in section 2.2.a, we will talk about an industrial MV/LV substation load that is independent to weather conditions. Thus, whatever the forecasting lead time is, for this load example, weather variable is not considered as an influence factor.

As described in the table 2.1, for our application in network operations, especially to cooperate with the ADA functions, we focus on the STLF. Weather, calendar and historical data inputs are the most concerned influence factors.

Table 2.1: Different time horizon load forecasts

Time horizon	Applications	Influence factors
Very Short-Term Load Forecast (VSTLF) (1Min ~ 1h)	ADA functions in DMS, Load Frequency Control (LFC) in Energy Management System (EMS)	Historical consumptions
Short-Term Load Forecast (STLF) (1h ~ 1 week)	Operation (ADA functions), estimation of load flows, representation of saving potential for economic and secure operation of power systems	Historical consumptions, calendar factors (day type and hour of the day), weather conditions (∗)
Medium-Term Load Forecast (MTLF) (1 week ~ 1 year)	Negotiation of electricity contracts, scheduling of fuel supplies and maintenance operation	(∗) + population, economic factors, etc (⋄)
Long-Term Load Forecast (LTLF) (1 year ~ several years)	Capital expenditures and planning operations	(⋄) + more information such as: population growth, Gross Domestic Product (GDP)

(∗) and (⋄) represent respectively influence factors for short-term and medium-term forecasts.

For a weather sensitive load, including the credible forecasting weather information as input is recommended as it can improve the performance of the forecasting model. Temperature and humidity are the most frequently used load predictors. Composite weather variables such as Temperature-Humidity Index (THI), Wind Chill Index (WCI) [24], and smoothed weather variables [26] are often adopted. THI and WCI indicate respectively the discomfort caused by summer heat and winter wind chill. The smoothed weather variable represents the effects of changes in weather accumulated over the time. In practice, regarding STLF, weather forecasting data are applied to calculate the performance of the model [27]. However, for the most of the time, in the construction phase of the model, the predicted weather forecast is not available. In this case, most authors in the load forecasting field run simulations with the realized weather data [28]. One should bear in mind that using forecasting weather information will surely decrease the model's overall precision [26, 29]. Therefore, some authors [30] proposed omitting imprecise weather information as a conservative solution, since it would bring large variance to the model. A promising way to handle the uncertainty in weather variables is the weather ensemble predictions that generate the load forecasts in a probabilistic form [29]. In chapter 3, we will re-discuss and show empirical evidence addressing to this issue.

The calendar inputs include the time of the year, the day of the week, the hour of the day as well as day types (working days, weekends or national holidays). There are important differences in load between weekdays and weekends. Weekends and holidays are

often more difficult to forecast than working days due to their relative infrequent occurrence and the clients' irregular behaviors. Some authors work on the classification methods [19] in order to find or even create similar days [31] for these "anomalous" day forecasts.

Historical data inputs are also very important to the STLF, from which the seasonality information can be extracted [14].

2.1.b Forecasting methods

This subsection gives an overview of various approaches for load forecasts. Many of them are developed for STLF on the High Voltage (HV) level, although MV and LV levels began to attract more attention with the expansion of the smart grids during the past years. Figure 2.1 gives a two dimensional digest on the methods and the models in the load forecasting field both in the lead time and the voltage hierarchical scales.

Mainly, two classes of approaches can be distinguished [14, 23]: classical approach and Artificial Intelligence (AI) approach. Classical approach requires an explicit mathematical model which interprets the relationship between load and its influence factors. This family includes regression model, time series method, similar day approach, end-use method and econometric approach. AI approach, on the other hand, extracting non linear relationships between input factors and load has become very popular nowadays. This family of algorithms includes Artificial Neuronal Network (ANN), fuzzy logic, and expert systems.

In the sequel, we introduce these different approaches.

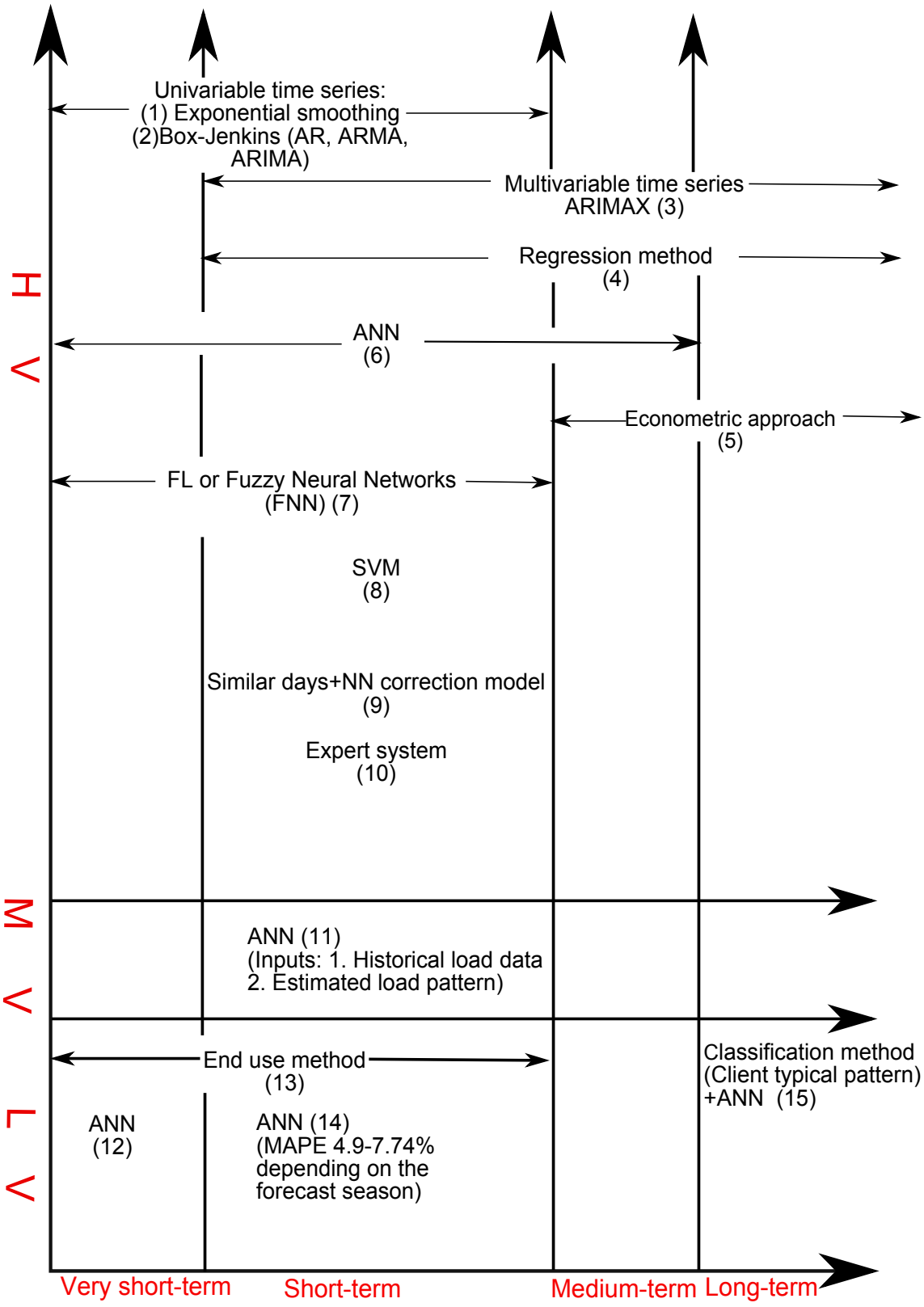


Figure 2.1: Summary of load forecasting methods in two dimensions: time horizon and voltage hierarchy. HV: High Voltage, MV: Medium Voltage and LV: Low Voltage. Numbers appear in the figure correspond to the related works.[a]

[a] (1):[13, 14] (2):[13, 15, 14] (3):[16] (4):[17, 18, 17] (5):[24] (6):[19, 20] (7):[32, 33, 34] (8):[35, 36] (9):[31] (10):[37, 38] (11):[39] (12):[22] (13):[40] (14):[41]

[b] Univariate time series model refers to the model with only one observation series, i.e., load data.

[c] Multivariate time series model corresponds to the model containing both the exogenous variables and the historical load data.

2.1.b-i Classical approach

Regression model. Regression is one of the most widely used statistical techniques. Electric load forecasting regression methods are usually used to express the relationship between load consumption and external factors [24, 42]:

$$y_i = a_i x_i + e_i \tag{2.1}$$

where y_i is the i-th load sample, x_i is the influence variable vector correspondent to the i-th load sample, a_i is the transposed regression coefficient vector, and e_i is a Gaussian error.

The advantages of regression methods are relatively easy implementation and interpretation for the relationship between input and output variables. Another advantage of the method is that it is easy to compute the prediction interval through estimation error of the model. The disadvantage of regression methods is the need to identify a correct form including effective inputs and output. This is hard due to the complex non-linear relationship [23].

C.L. Hor et al. [43] developed several multiple regression models incorporating weather-related and socio-economic variables on the load demand for England and Wales. Monthly data from 1989 to 1995 are used for the coefficient estimation of the model and monthly data from 1996 to 2003 are used to evaluate the accuracy of the forecasting model. The non-linear relationship between weather-related factors (mean monthly temperature value) and load suggests introducing other weather composite variables, such as Heating Degree Days (HDD), Cooling Degree Days (CDD), and Enthalpy Latent Days (ELD). The socio-economic variable GDP has evidently an impact on the trend. One of their regression model is set up as: $\hat{E}_1 = (\hat{E}_A + \alpha_7 \text{GDP}) F_{adj}(y)$, where $\hat{E}_1$ is the predicted electricity demand, $\hat{E}_A = \alpha_0 + \alpha_1 \text{CDD} + \alpha_2 \text{HDD} + \alpha_3 \text{ELD} + \alpha_4 V_w + \alpha_5 M_s + \alpha_6 M_r$, which represents the weather-related model. V_w stands for the mean monthly wind speed, M_s stands for the mean monthly sunshine hours, and M_r stands for the monthly rainfall. $\alpha_n, n = 0, \cdots, 7$ are constant coefficients. $F_{adj}(y)$ is the adjustment factor for each year. The Mean Absolute Percentage Error (MAPE) of the model, which represents the average portion of the absolute forecasting errors to the real forecasting values, was around 2%.

A. Bruhns et al. [17] designed a non-linear regression model for MTLF. This hourly load prediction model termed "Eventail" has been applied by the French electricity company EDF since 2001. They decomposite the load P_i into three components: $P_i = Phc_i + Pc_i + e_i$, where Pc_i and Phc_i are respectively the weather-dependent and the weather-independent parts, e_i is a Gaussian error. The weather-dependent part is fitted by a non-linear model of the observed temperature, the exponential smoothing temperature, and the cloud cover. Two thresholds for heating and cooling temperatures are also adopted in order to cope with the non-linearity. The exponential smoothing on temperature reflects the inertia of temperature inside buildings to the outside temperature variation. The weather-independent part integrates trends, day, week, year periods, and day type information. The dummy variables are used to indicate day types and four terms of Fourier series are used to model the seasonality pattern. Despite of the computational difficulty in estimation due to the temperature smoothing parameters, thresholds, and strong non-linearities, they declared that with the known weather data, a MAPE of around 2% for one-year-ahead forecast and a MAPE of 1.5% for one-day-ahead forecast have been achieved.

W. Charytoniuk et al. [26] introduced a nonparametric regression model for STLF. They considered load as a weighted average of past loads. The specific weights are defined by a multivariate kernel and its smoothing parameters. The optimal values of these smoothing parameters are calculated using CV technique. A conditional expectation of the load is built based on the local neighborhood loads P_i: $\hat{P}(x) = \frac{\sum_{i=1}^{n}\{P_i \prod_{j=1}^{r} K(\frac{x_j - x_{ji}}{h_j})\}}{\sum_{i=1}^{n}\{\prod_{j=1}^{r} K(\frac{x_j - x_{ji}}{h_j})\}}$, where $h_1, \cdots, h_r$ are smoothing parameters and $K(\mu)$ is a normal kernel. j denotes the j-th influence variable in the exogenous variable vector x and i denotes the i-th observation sample. They have obtained with accurate temperature values, up to one-week-ahead, a MAPE of 2.78% compared to an ANN model of 2.64%. They argued that even the errors are slightly higher than the ANN, the errors obtained by the regression model have better properties compared to those of the ANN model.

Time series method. Time series method is a linear model based on the assumption that there exists an internal structure within data. This structure consists of autocorrelation, trend and periodic variations. Time series methods detect and explore these relationships between the current load value, historical data and sometimes exogenous factors. A distinctive property of time series models is to consider time as one of the explanatory variables. As a linear model, time series method is suitable for VSTLF and STLF and it is able to provide the prediction interval. Whereas for MTLF and LTLF, these internal relationships are no longer linear and the model becomes complicated. Thus, it is no longer efficient to use a linear time series model for the MTLF or the LTLF application. In particular, Box-Jenkins models and exponential smoothing models are the most widely used time series methods.

1. Box-Jenkins models [44] include AutoRegressive (AR), Moving Average (MA), AutoRegressive Moving Average (ARMA), AutoRegressive Integrated Moving Average (ARIMA), Periodic AutoRegressive Moving Average (PARMA), and AutoRegressive Moving Average with eXogenous inputs (ARMAX). They are adapted for STLF, since they allow for the explicit modeling of time dependence. The basic element of these models are AR and MA models. AR model aims at estimating the current load by the moving average of historical load. MA model, on the other hand, explains the current load with the past errors committed by the model.

   In some cases, where the stationary condition[1] is not met, the trend and the periodic effects can be removed by carrying out transformations as follows:

   Let y_t be a stochastic process. With a first difference, the linear trend is removed:

   $$Z_t = y_t - y_{t-1} \tag{2.2}$$

   With a second difference, quadratic trend is removed:

   $$Z_t = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) \tag{2.3}$$

   For periodic series, it is possible to adjust observation seasonal effect by carrying out

[1] Mean and variance values stay the same over the time

the transformation:

$$Z_t = \frac{y_t - \overline{y}_t}{\sigma_t} \tag{2.4}$$

where $\overline{y}_t$ is the mean value, and σ_t is the standard deviation of y_t.

Being a rather complete model, ARIMA(p,d,q) is chosen herein to represent the Box-Jenkins family. It denotes the relationship between the load y_t at the time "t", and its historical data $y_{t-i}, i = 1, \cdots, p$, historical errors $e_{t-j}, j = 1, \cdots, q$ committed by the model and the error e_t at time "t", [45] :

$$\Delta y_t = \sum_{i=1}^{p} \phi_i \Delta y_{t-i} + \sum_{j=1}^{q} \theta_j e_{t-j} + e_t \tag{2.5}$$

The above model is decomposed of the following parts: AR(p)I(d)MA(q)

- AR(p): p is the order of autocorrelation (indicate using weighed moving average over the p latest observations).
- I(d): d is the order of integration (differencing) (indicate a linear trend or d order polynomial trend).
- MA(q): q is the order of moving averaging(indicate using weighted moving average over the p latest errors of the model).
- Δ refers to the transformations (equations 2.2 and 2.3) in order to achieve stationary.

Thus, an AR model equals to ARIMA(p,0,0) and a MA model equals to ARIMA(0,0,q). When the introduction of some explicative variables is necessary, transfer function becomes AutoRegressive Integrated Moving Average with eXogenous inputs (ARIMAX). "X" signifies the integration of exogenous variables.

A practical way of estimating the orders of an ARMA model is to calculate its AutoCorrelation Function (ACF) and Partial AutoCorrelation Function (PACF) [44]. Applying on the load data, the ACF expresses the linear predictability of the load data y_t at time "t", by using only the value τ instants before $y_{t-\tau}$ [44]:

$$\rho_y(t, t-\tau) = \frac{\gamma_y(t, t-\tau)}{\sqrt{\gamma_y(t,t)\gamma_y(t-\tau, t-\tau)}} \tag{2.6}$$

where $\gamma_y(t, t-\tau) = E[(y_t - E(y_t))(y_{t-\tau} - E(y_{t-\tau}))]$ is the autocovariance function, $\gamma_y(t,t)$ and $\gamma_y(t-\tau, t-\tau)$ are respectively the variance function of y at time t and $t-\tau$. $E(\cdot)$ refers to the expected value. If the y_t process is stationary, $\rho_y(t, t-\tau) = \rho_y(\tau)$,$\gamma_y(t, t-\tau) = \gamma_y(\tau)$ and $\gamma_y(t,t) = \gamma_y(t-\tau, t-\tau) = \gamma_y(0)$. Equation 2.6 becomes:

$$\rho_y(\tau) = \frac{\gamma_y(\tau)}{\gamma_y(0)} \tag{2.7}$$

The PACF, on the other hand, measures the additional correlation between y_t and $y_{t-\tau}$ after adjustments being made to exclude the dependence between the intermediate observations $y_{t-1}, \cdots, y_{t-\tau-1}$. Let y_t be a stationary process, the PACF is [44]:

$$\begin{aligned} & \psi_y(\tau) = \rho_{y'}(t, t-\tau) \\ \text{where} \quad & y'_t = y_t - E(y_t \mid y_{t-1}, \cdots, y_{t-\tau-1}) \\ & y'_{t-\tau} = y_{t-\tau} - E(y_{t-\tau} \mid y_{t-1}, \cdots, y_{t-\tau-1}) \end{aligned} \tag{2.8}$$

Both ACF and PACF are within $\{-1, 1\}$. The PACF of the AR(p) process is near zero from the p-th lagged value. The ACF of the MA(q) process is near zero from the q-th lagged value. A good fitness of model demands a Gaussian noise. Let e_t be a Gaussian noise, the coefficients of the ARMA model are calculated relying on Maximum Likelihood Estimation (MLE) [44].

J. Nowicka-Zagrajek et al. [14] applied a two-step procedure to address the load modeling and forecasting issue. First, weekly and annual seasonalities are removed by the moving average technique. Then, the deseasonalized data are fitted with an ARMA model. The performed residual analysis suggests that the residual follows a hyperbolic distribution. By comparing to a 1.7% error committed by the official forecast of the CAlifornia Independent System Operator (CAISO), the proposed model yields a $1.2 - 1.25\%$ error. However, this ARMA model cannot capture the holiday structure.

M. Zhou et al. [46] proposed an ARIMA approach to electricity price forecast with accuracy improvement by predicting errors. Besides a conventional ARIMA model for the electricity price forecast, several ARIMA models are also established for the residual error forecast. Results show that the method requires some easy-implemented low-order models instead of one complex model and the accuracy of the forecast is improved significantly. They argued that choosing the same or different kinds of approaches for the model forecast and the error forecast does not potentially influence the forecasting accuracy but the time of modeling.

2. Exponential smoothing [47, 12] is a univariate time series method that assigns exponentially decreasing weights to the historical loads. With a forgetting parameter, it corrects the previous smoothed load value by the new load measurement. According to the number of smoothing components, three models are the mostly used: simple, double, and triple exponential smoothing models.

   Simple smoothing model assumes that load data vary around a stable mean with no trend [48]: $r_t = \alpha y_t + (1-\alpha) r_{t-1}$, where the smoothed value r_t is the weighted average of the current load y_t and the previous smoothed value r_{t-1}. The forgetting parameter $\alpha \in [0, 1]$ controls the exponentially decreasing weights. The greater this parameter is, the greater influence the current load measurement has on the current smoothing value. The initial smoothing value r_1 is often chosen as y_1 or the average of the first four or five load values.

   Double exponential smoothing works with the load data with trend. The smoothing is done on the "level" and "trend" components of the time series data [48]:

$$\begin{aligned} \text{Level:} \quad & r_t = \alpha y_t + (1-\alpha)(r_{t-1} + b_{t-1}) \\ \text{Trend:} \quad & b_t = \beta(r_t - r_{t-1}) + (1-\beta) b_{t-1} \end{aligned} \tag{2.9}$$

   where the forgetting parameters $\{\alpha, \beta \in [0, 1]\}$ respectively control the level and the trend components' decreasing weights. The initial values for these two components are: $r_1 = y_1$ and $b_1 = y_2 - y_1$ (or $b_1 = \frac{y_n - y_1}{n-1}$).

   Triple exponential smoothing (also known as Holt-Winters (HW) method) works with "trend", "seasonal", and "level" components of the data set. Mainly, two kinds of mod-

els exist: multiplicative seasonal model and additive seasonal model. Multiplicative seasonal model assumes that the original data value in the data set is the product of the seasonal pattern and the average level. While additive seasonal model suggests that the seasonal pattern is independent to the average level of the data set and their sum equals to the original data set.

The multiple model can be expressed as follows [48]:

$$\begin{aligned} & y_t = (r_t + b_t t) l_t + \epsilon_t \\ \text{Level:} \quad & r_t = \alpha \frac{y_t}{l_{t-L}} + (1-\alpha)(r_{t-1} + b_{t-1}) \\ \text{Trend:} \quad & b_t = \beta(r_t - r_{t-1}) + (1-\beta) b_{t-1} \\ \text{Seasonal:} \quad & l_t = \gamma \frac{y_t}{r_t} + (1-\gamma) l_{t-L} \end{aligned} \tag{2.10}$$

where $\{\alpha, \beta, \gamma \in [0,1]\}$ are three forgetting parameters and L is the length of a period. Dividing the new load measurement y_t by the estimated seasonal component value of the last period l_{t-L} gives us the deseasonalized level, which participates of updating its old value estimated in the last round $r_{t-1} + b_{t-1}$. The trend variable is corrected by the difference of the level and its old value. The seasonal factor is updated by the most recently observed seasonal component given by y_t divided by the smoothed level estimate. For the initialization of these three components, a minimum of two full seasons data (2L-period data) are required.

The multiple-step-ahead load forecast for T periods is given by:

$$\hat{y}_{t+T} = (r_t + T b_t) l_{t+T-L} \tag{2.11}$$

The additive model can be expressed as follows:

$$\begin{aligned} & y_t = r_t + b_t t + l_t + \epsilon_t \\ \text{Level:} \quad & r_t = \alpha(y_t - l_{t-L}) + (1-\alpha)(r_{t-1} + b_{t-1}) \\ \text{Trend:} \quad & b_t = \beta(r_t - r_{t-1}) + (1-\beta) b_{t-1} \\ \text{Seasonal:} \quad & l_t = \gamma(y_t - r_t) + (1-\gamma) l_{t-L} \end{aligned} \tag{2.12}$$

Compared to the multiplicative model, the additive model simply changes the seasonal component multiple relationship by the additional relationship. The multiple-step-ahead load forecast for T periods is given by:

$$\hat{y}_{t+T} = r_t + T b_t + l_{t+T-L} \tag{2.13}$$

According to the different ways of assigning values to its parameters $\{\alpha, \beta, \gamma\}$, there exist two techniques: non-adaptive and adaptive. Non-adaptive technique fixes the parameter values, once they are initialized. The advantages of the non-adaptive technique are twofold:

- one, it is economic in terms of computational time. Once the parameters have been established, the forecast can proceed without any delay in re-computation of the parameters.

- Two, that no historical data needs to be recorded saves the storage space.

On the other hand, the adaptive technique keeps adapting the parameters to the changes in the undergoing process. The newly computed parameters may be computed using: all available data till that time or only k most recent data values. Thus, a certain amount of data need to be memorized.

J.W. Taylor [13] proposed a new triple seasonal exponential smoothing approach that captures intraday, intraweek, and intrayear seasonal cycles for a 24-hour-ahead load forecast. Single, double, triple seasonal ARMA, single, double, triple seasonal HW exponential smoothing and double, triple Intraday Cycle (IC) exponential smoothing methods were compared. The IC exponential smoothing method omits the occurrence of the intraweek seasonal cycle, but introduces some dummy variables in order to distinct different intraday cycles during a week (Monday, Saturday, and Sunday, etc.). The approaches were tested on the British and the French national data. He showed that for prediction up to a day-ahead, the triple seasonal methods outperform the single and double seasonal methods, and also a univariate ANN approach. Moreover, he concluded that there was little difference among the same seasonal versions of ARMA, HW, and IC exponential smoothing. He concluded that a simple average combination of these methods led to a better accuracy than any of the single method.

Similar day approach (also referred as K-Nearest Neighbor(s) (KNN) method). This is a data-based nonparametric approach that simply consists in searching past observations of the process for k events, which are most likely to be similar to the forecasting situation on some characteristics, such as the weather, the day of the week, and the day type [24]. A forecast is considered as a linear combination or a regression procedure of these k events with unequal weights. These specific weights are defined by the distance in characteristics between the similar days and the forecasting day. The task of calibrating these specific weights is delicate but crucial to the precision of the method. It mainly relies on the choices of the comparison characteristics and the distance metrics. In order to get a satisfactory result, this approach requires a large database and a stationary process. However, the required computational time, to calibrate and to get the output forecasts, increases rapidly with the size of the database [42].

T. Senjyu et al. [49] designed a Recurrent Neural Network (RNN) model for the next day load curve forecast by correcting the output of the similar day approach. Euclidean norm D with weighted factors is applied to evaluate the similarity of similar days:

$$\begin{aligned} D &= \sqrt{\hat{w}_1(\Delta L_t)^2 + \hat{w}_2(\Delta L_{t-1})^2 + \hat{w}_3(\Delta L_{t-2})^2} \\ \Delta L_{t-k} &= L_{t-k} - \widetilde{L}^p_{t-k} \end{aligned} \tag{2.14}$$

where ΔL_{t-k} is the deviation between the load on the forecasting day L_{t-k} and the load on a similar day $\widetilde{L}^p_{t-k}$, and $\hat{w}_i, i = 1, \cdots, 3$ are the weighted factors. The weighted factors are estimated in a regression model of historical temperature and load data by the least-square method.

A load correction method by correcting the data of historical days with the load correction rates was proposed to generate new similar days in case of the shortage in the training data. The load correction rate $\hat{L}_t^r$ is determined by the following least-square method based on the regression model.

$$\begin{aligned} L_t^r &= L_t^{sl}/L_t^{st} \\ T_t^r &= T_t^{sl}/T_t^{st} \\ \hat{L}_t^r &= \hat{w}_0 + \hat{w}_1 W_t H_t + \hat{w}_2 T_t^r + \hat{w}_3 Day \\ L_t^{pnew} &= L_t^{pold} \hat{L}_t^r \end{aligned} \tag{2.15}$$

where L_t^{sl} and T_t^{sl} are respectively the load and temperature data of similar days selected by using equation 2.14, L_t^{st} and T_t^{st} are respectively load and temperature data of past days selected by using a forecasting maximum temperature, L_t^r is the deviation rate between L_t^{sl} and L_t^{st}, and T_t^r is the deviation rate between T_t^{sl} and T_t^{st}. $H_t, t = 1, \cdots, 24$ is time, Day indicates the day of the week. L_t^{pnew} and L_t^{pold} are respectively corrected load data and original load data of past days. $\hat{w}_i, i = 0, \cdots, 3$ is the weighted factor and W_t is the weighted factor on each hour. These new load data L_t^{pnew} were then used to train the neural network.

This method gained an improvement on forecasting accuracy for special days such as weekends and national holidays. A MAPE of 2.78% was reported being obtained during a year.

End-use method. The assumptions underlying end-use methods consider that every electrical housing device (light, cooling, heating, refrigeration, cooking, and so on) has a specific electrical signature. Such signature consists of a unique temporal and spectral pattern that can be separated, classified and detected. Every time a given electrical device is switched on, specific sensor can detect its activity. Prior knowledge of the expected energy consumption of the housing electrical device enables an easy estimation of the user's electrical consumption [24]. Finally, a specific profile for every end-user, consisting of his age, income, the size of the house as well as the usage frequency of the different considered devices, is designed. It would be possible to accurately forecast the power consumption of the whole network.

Gathering all these necessary information is usually a cumbersome yet necessary step to build end-use models. Several steps in the process can be identified: first, relying on the historical data, energy consumption of every device needs to be identified (i.e., find out its power consumption as well as its expected usage time for every house). Ideally speaking, the forecaster estimates the number of houses having these devices in the region of interest as well as the probability of the usage of the devices at a specific time. Finally, summing up all the estimated energy consumption for all the considered houses, the estimation of the electricity in the region can be made. If all these above influence factors are projected into the future, the forecaster can provide an estimation for the future energy consumption. Of course, such approach is generalizable to residential, commercial and industrial sectors with no restrictions [50].

Designing such forecasting methods can be time consuming however the benefits of such accuracy could be worth the effort. As a matter of fact, the bottom-up method can

provide means to optimize the grid's energy demand, such as improving certain appliance efficiencies, and improving insulation levels of some apartments. Thus, energy can be saved, installations can be preserved and developed depending on the accurate knowledge regarding the local demand, and so on.

Due to the amount of data and prior information on the end-user as well as on all electrical device usages in the considered sectors, these methods are proved to be difficult to implement. The shortage of such information usually discards end-use methods from the set of possible choices. As a matter of fact, the reliability of the forecast highly depends on the amount of data collected as well as their quality.

Econometric approach. Econometric approach usually refers to a mathematical formulation that incorporates statistical tools in order to explicit the complex relationships that exist between electricity demand with economic, demographic and other influence factors mentioned previously [24, 51]. For instance, econometric tools can provide a piece of explanation regarding the electricity consumption behavior of a target population on the electricity price fluctuations.

Besides the economic theories, an econometric forecasting model is established relying on statistical assumptions about the joint distribution of explanatory variables and a set of unobservable variables [51]. The economic theory is used to specify the explanatory variables and the dependent variable, and to clarify how institutional and economic conditions can affect their relationships. Joint distribution density of the load and explanatory variables can then be estimated by the functional forms, such as parametric, nonparametric or semiparametric statistical models. The estimation of the models largely depends on the assumptions made about the unobservable, which impacts the consistency of the estimation criterion [51].

The main advantage of such approaches is explicative, such that they can explain how the evolution of the influence factors working on the load demand [50]. Namely, they fit the behavior of targeted populations relying on past observations, and thus can adapt to various populations. In our case, it can help designing a load forecasting model that adapts to different sectors, e.g., residential, commercial, and industrial sectors.

Unfortunately, what represents its strength, i.e., relying solely on past observations, can also be seen as its weakness. As a matter of fact, usually, econometric based forecasters assume that the relationship between the electricity load and influence factors is stationary and thus remains the same during the forecasting process [50]. Such assumption is usually justified for short periods, not for a long-term forecasting processes. Similarly, such designed models are not efficient when the real observations follow a highly fluctuating dynamism such as in small regions. It behaves better on a national scale where electricity load fluctuations are steadier. Consequently, the econometric model needs to be reevaluated on regular bases in order to maintain its reliability.

2.1.b-ii Artificial intelligent approach

ANN (also known as NN and RNN). Over the past two decades, the ANN in the AI technique group has been applied in various fields, e.g., pattern recognition, robotics, and load forecast, to name a few. Thanks to its non-linearity and great learning ability, ANN

quickly gained its popularity in load forecast. It is widely used in all forecasting lead times and all electrical voltage hierarchies. Compared to the classical approaches, it is often considered as a benchmark of performance [13, 39], or more often as a complimentary to enhance their abilities facing dynamic varying environments.

The nonlinear model can interpret complex relationships between multiple inputs and outputs. Its ease of implementation, automatic mapping ability, and generalization ability are often commented as principal advantages. Generalization ability allows the network to provide satisfactory responses to new inputs after its learning phase [52]. The use of the ANN is often divided into two phases: a learning phase and a prediction phase. In the learning phase (also known as training phase), ANN adapts its parameters, namely the synaptic weights and biases, to learn the knowledge within inputs and outputs containing in the historical data set (also known as learning data). In the prediction phase, the parametered ANN is used as a transfer function to produce the output(s). ANN is very flexible as when the operating environment changes, with the same structure, it is capable of rapidly adapting its interconnection weights to respond to the changing situation after the training process. ANN is also said to have great skills coping with the noisy data set [39].

The disadvantages are related to its large-scale choices, such as for the architecture (feedforward or recurrent, etc.), for the structure (number of hidden layers and number of hidden neurons on each layer), for the learning set, and the number of iterations for the training (to avoid overflow). The justification of the choices is not systematically argued or commented in a clarified way [39].

The ANN is considered as a class of algorithms capable of approximating any transfer function, provided that the activation functions of neurons in the intermediate layer are not polynomials, are bounded and piecewise continuous [53].

There are a great variety of ANNs found in the literature. Here, we present some frequently encountered structures: Multi Layer Perceptron (MLP), RNN, Wavelet Neuronal Network (WNN), Radial Basis Function Networks (RBFN) and Self-Organizing Maps (SOM).

1. As a universal approximator and having the simplest structure, the Multi Layer Perceptron (MLP) is the most widely used ANN model. It has one input layer, where the influence factors are introduced and one output layer for the estimated value, the load data. Between the input and the output layers situate the hidden layers with neurons represented by nonlinear activation functions. Figure 2.2 shows a single layer perceptron's structure, which is the simplest of the feedforward layered structures.

   Single Layer Perceptron (SLP) consists a layer of input neurons $\{p_j, j = 1, \cdots, R\}$, associated weights $\{\omega_{1,j}, j = 1, \cdots, R\}$, a weight ω_{10} related to a constant p_0 termed "bias", which equals to 1, an activation function g, and an output c_1. The single perceptron compares the weighted sum of the inputs with the weighted bias, and the comparison result n goes through one activation function g in order to get the output. Lots of activation functions exist, such as step functions, linear functions, and sigmoid functions. For the approximation of a non linear model, the sigmoid function $g(n) = \frac{1}{1+e^{-n}}$ is the most commonly used for the hidden neurons. The

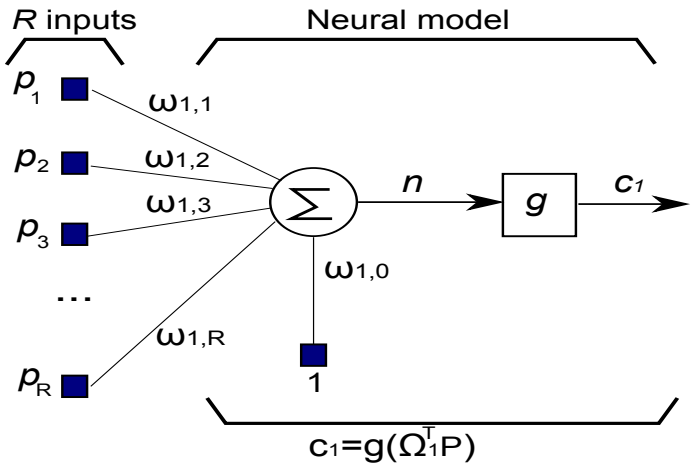


Figure 2.2: Single perceptron structure [54]

weights $\Omega_1, \{\omega_{1,j},\, j = 0, \cdots, R\}$ of the neuron represent the efficiency of the synaptic connection. A negative weight inhibits its correspondent input, while a positive weight accentuates its correspondent input. The bigger the absolute value of the synaptic weight is, the more important the input is to the result. The extension to the SLP is the MLP (figure 2.3).

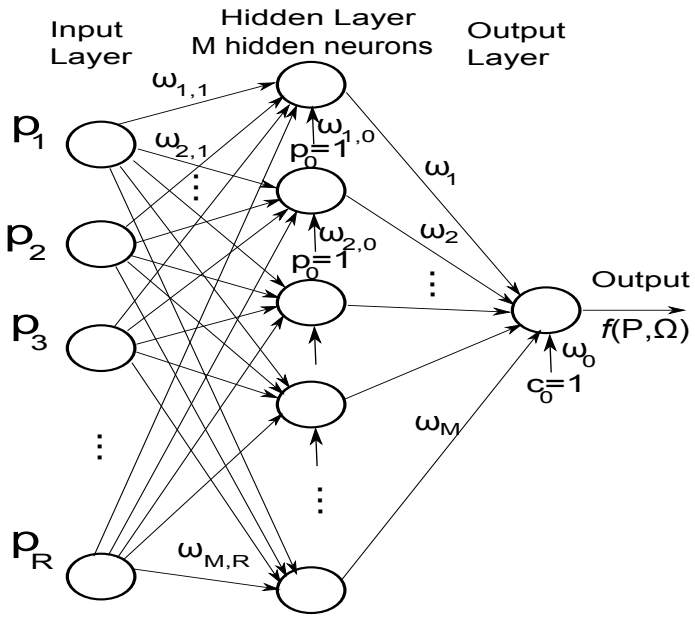


Figure 2.3: One-hidden-layer network structure, where P is the vector of variables $\{p_j, j = 0, \cdots, R\}$, and Ω is the vector of parameters $\{\omega_{i,j}$ and $\omega_i, i = 1, \cdots, M, j = 0, \cdots, R\}$. Therefore, the total number of variables is $(R+1)M + (M+1)$.

The most popular artificial neural network architecture for electric load forecast is the back propagation [55, 24]. Back propagation neural networks use continuously valued functions and supervised learning. Under the supervised learning, the actual numerical weights and bias assigned to element inputs are determined by matching N historical data inputs/outputs pairs $\{P_i, y_i\}$ (regarded as "learning set" or "training

set") in the training process.

The objective of the learning is to match the output of the MLP to the desired output, given the same inputs. This objective is reached by tuning the parameters Ω of the neural network, i.e., interconnection weights and bias. The learning procedures are grouped into three different categories: supervised learning, unsupervised learning, and hybrid learning. The essential element that differs supervised learning from other categories of learning is the presence of a "teacher". In the supervised procedure (figure 2.4), the teacher has some prior knowledge about the environment. Thus, when an input P_i in the learning set is presented, the teacher is capable of providing the perfect output y_i. The "supervised neural network" presents the student who tries to learn from the teacher by providing a close answer to the same input. The knowledge of the student is gained by the adjustment of his parameters under the influence of the deviation from the teacher's answer, e_i. The learning process stops when the difference between two responses is under a certain threshold, or the maximum number of iteration reached.

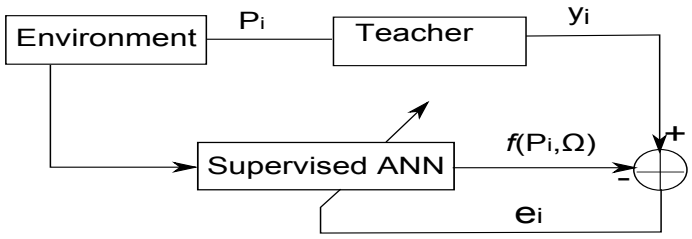


Figure 2.4: Supervised learning procedure [54]

When to our load forecasting issue, let a set of N past historical data $\{P_i, i = t-1, \cdots, t-N\}$, such as time and weather, be the training set inputs, historical consumption samples $\{y_i, i = t-1, \cdots, t-N\}$ that are collected with a regular time step be the desired outputs, and Ω be the set of ANN's parameters. Prediction is solved by minimizing errors in the training procedure presented in figure 2.4:

$$\min_{\Omega}\{\frac{1}{2}\sum_{i=t-N}^{t-1}(y_i - f(P_i,\Omega))^2\} \tag{2.16}$$

A. Kusiak et al. [40] have developed a building steam load predicting model based on MLP. This load is used to represent heating and cooling loads of buildings. First, boosting tree algorithm and correlation coefficients between input variables are used to deduce the most important input variables. Then, among several data-mining algorithms [56], namely, CART, CHAID, exhaustive CHAID, boosting tree, MARSplines, random forest, Support Vector Machine (SVM), MLP, MLP Ensemble and KNN, that map input variables to output steam load, MLP Ensemble performs the best. This latter involves five MLPs. Five different activation functions are selected, namely, logistic, identity, hyperbolic tangent, exponential, and sine functions. The number of hidden units is set between 5 and 18, and the bias for both the hidden and output layers varies from 0.0001 to 0.001. Many ANN has been applied to the load forecasting field, this example is chosen to state here since it is one of the rare

research working on the individual load forecast. It reveals the great ability of ANN coping with all kinds of forecasting issues.

2. Recurrent Neural Network (RNN) are equivalent to Nonlinear AutoRegressive Moving Average (NARMA) models. There are a variety of different RNN structures. We present here the most commonly used structure in the RNN family: complete RNN [57]. In this structure, the intermediate outputs or the final outputs are introduced as inputs for the model. This brings in unity delay factors to the model.

   T. Senjyu et al. [49] employed the RNN coordinating with similar day's data for the next day 24-hourly load forecasts. Figure 2.5 shows the given structure. The inputs of the RNN include the deviation of load ΔL, deviation of temperature ΔT between forecasting day and similar day, and previous output of the RNN values ΔC. The output of the RNN is the correction value ΔC_{t+1} applying to correct the similar days' value. Each block represents one unit discrete delay. They are used to memorize the previous activations and re-injected as the inputs for the RNN.

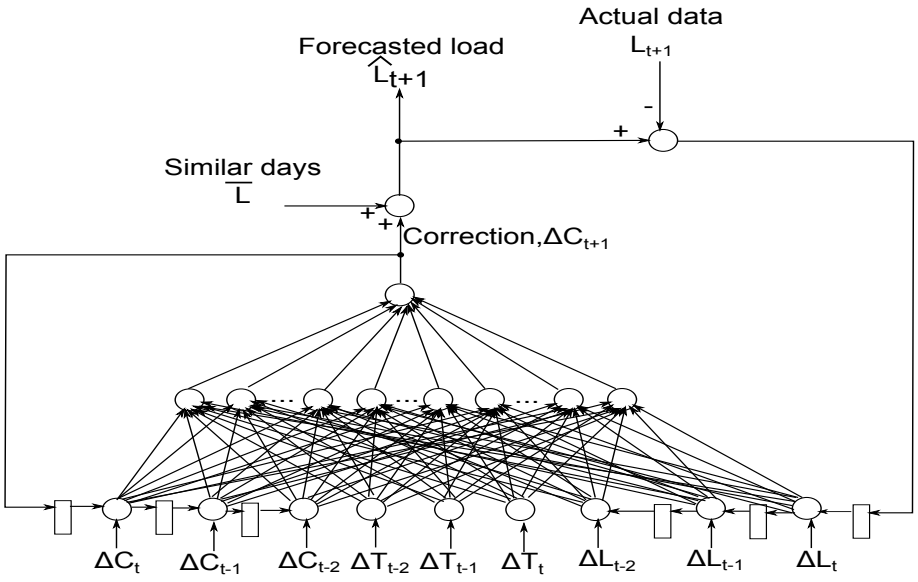


Figure 2.5: Recurrent neural network structure [49]

3. Wavelet transformation is frequently used in the signal preprocessing by decomposing the original signal into a set of better-behaved constitutive series. This transformation can isolate the outliers, and convert series into stationary, etc. The Discrete Wavelet Transform (DWT) can capture high and low frequency information by producing fine and broad scale coefficients [58]. For a given signal $y(t)$, a DWT can be expressed as [59]:

$$y(t) = \sum_k c_{j0,k}\Phi_{j0,k}(t) + \sum_{j>j0}\sum_k \omega_{j,k}2^{\frac{j}{2}}\psi(2^j t - k) \tag{2.17}$$

   where ψ is a mother wavelet function, j is the dilation or level index, k is the translation or scaling index, $\Phi_{j0,k}$ is a scaling function of broad scale coefficients, $c_{j0,k}$, $\omega_{j,k}$ are the scaling function of fine scale coefficients, and all functions of $\psi(2^j t - k)$ are orthonormal.

Wavelet based signal processing relies on a decomposition step and a reconstruction step. On the one hand, the decomposition step filters the data through a high pass/low pass filter. This operation is usually performed by computing a convolution between the data and the designed filter. On the other hand, the construction step performs the inverse of the decomposition step [58].

Wavelet processing serves as an efficient filter that can be integrated with any other method such as Box-Jenkins [60], ANN [58, 61, 62], and Kalman filtering [63]. ANN based on wavelets inherits this advantage.

Ajay Shekhar Pandey et al. [59] applied the wavelet decomposition to, and compared with, classical and artificial intelligent approaches. Different from other works, which have used the wavelet coefficients of the data set as input data, the proposed approach used the wavelet pre-processed data set after removing the highest frequencies (fast changing) components. They showed the superiority of the proposed wavelet based approach over the non-wavelet methods for the same set of data.

Wavelet Neuronal Network (WNN) combines the wavelet theory and NN into one. A WNN generally consists of a feedforward NN, with one hidden layer, whose activation functions are drawn from an orthonormal wavelet family. The outputs of WNN can be expressed as [64]:

$$y(u) = \sum_{i=1}^{M} \omega_i \psi_{\lambda_i, t_i}(u) + \overline{y} \tag{2.18}$$

where $\overline{y}$ value is to deal with functions whose mean is nonzero. The $\overline{y}$ value is a substitution for the scaling function $\phi(u)$ at the largest scale. Parameters $\{\omega_i, \lambda_i, t_i, \overline{y}\}$ are adjustable by learning procedure.

The WNN is believed to behave better than the MLP in nonlinear systems and variant environment such as electricity load forecasting.

4. Radial Basis Function Networks (RBFN) structures often have one radial hidden layer and one linear output layer. The hidden neurons compute the similarity between any input pattern and the neurons assigned point by means of a distance measure. The underlying idea is to make each hidden neuron represents a given region of the input space. When a new input signal is received, the neuron representing the closest region of input space activates a decisional path inside the network leading to the final result. Radial functions are applied as activation functions on the hidden layer. There are several radial functions, among which the most frequently used is the multi-variant Gaussian type function [54]:

$$g(x) = exp(-\frac{(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)}{2}) \tag{2.19}$$

where μ_i and Σ_i are respectively the centre and the covariance matrix of the radial neuron. These two parameters are defined before the training process.

Architecture of the RBFN can be found in [65]. The RBFN owns the same universal approximation capacity as the MLP and the calibration of its parameters is less time consuming [65]. The RBFNs are flexible, simple in structure and tolerant to Gaussian noise of the inputs.

Despite of its simple structure, RBFN suffers several drawbacks related to the learning process. On the one hand, since the learning step deals with two different activation functions (radial and linear), the convergence rates of the different layers are different. More specifically, the convergence on the parameters of the radial functions is much slower than that on the parameters of the linear functions. Moreover, the number of hidden neurons depends exponentially on the input space. For large input spaces, the structure of the neural network might become intractable.

5. Self-Organizing Maps (SOM) is an unsupervised ANN [66]. The learning process aims at exhibiting neighborhood relationships among vectors of an unlabeled data set. The neurons in the SOM are organized in one, two or three dimensional arrays. Each neuron i has a weight vector ω_i with the same dimension as the input vector x. Moreover, this weight vector stores the information about the inputs and outputs of the mapping being studied. The network weights are trained according to a competitive cooperative scheme, in which the weight vectors of a wining neuron i^* and its neighbors in the output array are updated after the presentation of a new input vector. The learning procedure of the SOM consists of a set of training steps. During the learning stage, the wining neuron i^* at step n is determined based on the input of the training sample x [66]:

$$i^*(n) = \arg(\min_{i \in A}\{\|x - \omega_i(n)\|\}) \tag{2.20}$$

The learning rules update the weights of the wining neuron i^* that has the minimal distance among all neurons on the output layer A between the weight vector ω_i and the input vector x:

$$\omega_i(n+1) = \omega_i(n) + \alpha(n)k_{ii*}(n)(x - \omega_i(n)) \tag{2.21}$$

where $\alpha(n)$ is the learning rate, and $k_{ii*}(\cdot)$ is a kernel function that spreads the updates to the neighborhood region. The neighborhood kernel defines the influence region that the input sample has on the SOM. Finally, the neurons on the grid become ordered: neighboring neurons have similar weight vectors [67].

Thus, SOM is a classification method that is able to decompose the input training data into several subsets with specific characteristics in an unsupervised manner [66].

Support Vector Machine (SVM). Introduced in 1995 [68], SVM (or Support Vector Regression (SVR)) is a nonlinear kernel based machine learning algorithm applied to data classification and regression. Its underlying mathematical concept is based on the transformation of the input data into a much higher dimensional feature space. In this new feature space, the original problem appears as a linear optimization problem with constraints:

$$f(x) = \omega m(x) + b \tag{2.22}$$

where ω is the weight vector, b is the threshold value, $m(\cdot)$ is the mapping function. The

SVM solves an optimization problem [23, 30, 66]:

$$\min_{\omega,b,\xi,\xi^*} \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$$

subject to

$$\begin{aligned} & y_i - (\omega^T m(x_i) + b) \le \epsilon + \xi_i^* \\ & (\omega^T m(x_i) + b) - y_i \le \epsilon + \xi_i \\ & \xi_i, \xi_i^* \ge 0,\ i = 1, \cdots, N \end{aligned} \tag{2.23}$$

The optimization equation is composed of two terms: a regularization term $\frac{1}{2}\omega^T\omega$ and a term added to control the error margin of the algorithm $C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$ in order to avoid over-fitting or under-fitting phenomena, where all input data x_i are mapped to a feature space with a much higher dimension than the original one. ξ_i^* and ξ_i are optimization variables introduced to take into account the optimization constraints on the error margin. Such margin is further modeled through the variable ϵ, such that $\|(\omega^T m(x_i) + b) - y_i\| \le \epsilon$. Finally, the cost of the classification or regression errors are quantified through the variable $C > 0$. Note that the overall accuracy of the training is controlled by $m(\cdot)$, ϵ and C.

Hence, relying on Lagrange multipliers, the problem is transformed into an unconstrained Lagrange equation [69]:

$$\begin{aligned} R = & \frac{1}{2}\omega^T\omega + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N} a_i(\epsilon + \xi_i^* - y_i + (\omega^T m(x_i) + b)) \\ & - \sum_{i=1}^{N} a_i^*(\epsilon + \xi_i + y_i - (\omega^T m(x_i) + b)) - \sum_{i=1}^{N}(\eta_i\xi_i + \eta_i^*\xi_i^*) \end{aligned} \tag{2.24}$$

where $\{a_i, a_i^*, \eta_i, \eta_i^*\}$ are Lagrange coefficients.

By assigning the partial differential equation of each parameter $\{b, \omega, \xi_i, \xi_i^*\}$ zero, we obtain:

$$\begin{aligned} \frac{\partial R}{\partial b} &= \sum_{i=1}^{N}(a_i^* - a_i) = 0 \\ \frac{\partial R}{\partial \omega} &= \omega - \sum_{i=1}^{N}(a_i^* - a_i)x_i = 0 \\ \frac{\partial R}{\partial \xi_i} &= C - a_i^* - \eta_i = 0 \\ \frac{\partial R}{\partial \xi_i^*} &= C - a_i - \eta_i^* = 0 \end{aligned} \tag{2.25}$$

Bring the above equations into equation 2.24, this later is equivalent to maximize [69]:

$$\frac{1}{2}\omega^T\omega - \epsilon\sum_{i=1}^{N}(a_i^* + a_i) + \sum_{i=1}^{N} y_i(a_i - a_i^*) \tag{2.26}$$

where a_i and a_i^* are Lagrange multipliers with $0 \le a_i, a_i^* \le C$ and $\sum_{i=1}^{N}(a_i) = \sum_{i=1}^{N}(a_i^*)$.

Substitute $\omega = \sum_{i=1}^{N}(a_i^* - a_i)x_i$ into equation 2.22, the regression function becomes:

$$f(x) = \sum_{i=1}^{N}(a_i - a_i^*)k(x_i, x) + b \tag{2.27}$$

where $k(x_i, x) = m(x_i)m(x)$ is a kernel function, which is the inner product of two vectors in feature space $m(x_i)$ and $m(x)$.

One kernel function should respect that after the transformation, the function must have the same result than the product of the projection. Common kernels include linear kernels $m(x_i)^T m(x_j) = x_i^T x_j$, polynomial kernels $m(x_i)^T m(x_j) = (\gamma x_i^T x_j + b)^d$, and Radial Basis Function (RBF) kernels $m(x_i)^T m(x_j) = exp(-\|x_i - x_j\|^2/2\sigma^2)$. In the load forecasting context, the RBF kernels are the most commonly used.

In addition to its strong and reliable mathematical foundations, the SVM is found to behave well in practice. The ANN structure minimizes the deviation between every output of the ANN and the desired value in the learning data, whereas the SVM searches to minimize the upper error bound of the generalization error [70]. Moreover, the SVM is commented having the ability reaching the global optimum. Therefore, it is indeed able to avoid over-fitting and thus offers interesting classification or regression generalization properties [66]. Compared to the numerous choices to be made before applying the ANN structure, the choices for the SVM restrain to the kernel function selection and its associated parameters [71]. These later are often solved by CV technique applying to the learning data set.

In 2001, Bo-Juen Chen et al. [30] obtained the first price at load forecasting competition organized by EUNITE [2] relying on a SVM based forecaster. The participants aim at predicting daily maximum load demand of the next month. Bo-Juen Chen et al. [30] chose the calendar dates as well as historical load data as inputs of the SVM optimizer. The model was implemented with the software LIBSVM, a library for SVM.

Fuzzy Logic (FL). Fuzzy logic based systems rely on approximated or fuzzy information rather than precise values or descriptions [32]. Figure 2.6 depicts the different steps of a fuzzy logic based system. Mainly, we observe five steps: first, the model begins with defining the set of input(s)/output(s), the parameters as well as their relationships. Then, the input values are mapped to a set of fuzzy variables that represent a qualitative range rather than a quantitative data. For instance, a classic example when dealing with temperature data is to associate their values to a set of fuzzy variables: “Low”, “Medium” and “High”. Each measurement is equivalent to a degree of truth (also called degree of membership) ranging from 0 to 1. The degree of truth represents the percentage that the measurement belongs to each fuzzy variable, i.e., a given temperature can be 0.2 “High” and 0.8 “Medium”. This step is usually called fuzzification. The next step consists of building a set of rules, depending on which the fuzzy variables are affected to the fuzzy outputs. The number and complexity of these rules mainly rely on the number of input parameters as well as the number of fuzzy variables associated to each input parameter. This is followed by an inference step that aims at evaluating the degree of fulfillment (firing strength) of the different rules. Finally, the defuzzification step transforms the fuzzy output of a control order or forecasting process into a crisp value.

FL based forecasters have many advantages. There is no need for a mathematical model to explicit input/output functional relationship [24] and it can extract similarity information from a large set of data as long as such data exist. Thus, K.Liu and al. [32]

[2] European Network on Intelligent Technologies for Smart Adaptive Systems

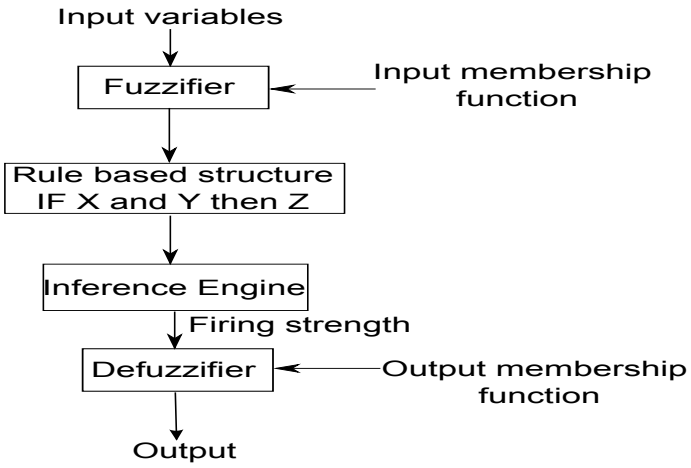


Figure 2.6: Fuzzy Logic process [72]

identified the similarities present in their data relying on two descriptions of the differences defined as *first-order difference* (V_k) and *second-order difference* (A_k):

$$V_t = \frac{y_t - y_{t-1}}{T}, \quad A_t = \frac{V_t - V_{t-1}}{T} \tag{2.28}$$

where y_t is the load sample at time "t" and T is the size of a period. $(m-1)$ former samples before t-th sample and n samples after t-th sample are respectively used as fuzzy inputs and outputs.

FL has also been used to design load correction models (along with a RNN correction model) [31]. T. Senjyu et al. [31] chose, in their study, special day, precipitation and discomfort index values as fuzzy data. The input and output membership functions are further depicted in figures 2.7 and 2.8. The membership function parameters ($a_i, i = 1, \cdots, 7$ and $b_j, j = 1, \cdots, 8$) of the fuzzy sets are also defined. Fuzzy rules are formulated based on the gained experience [31]. The firing strength is calculated base on the Mamdani method [72]. The output of the FL is the correction rate that applies to correct the selected similar days' data, which have already been corrected by the RNN correction model. This method by adding the FL correction is reported having an improvement from 1.63% to 1.43% on a whole year data.

Fuzzy linear regression [42] can be seen as an alternative approach to the statistic regression, when this latter fails due to unfulfilled requirements. The general model assumes a linear model such that:

$$Y = A_0 + \sum_{i=1}^{n} A_i(x_i - \overline{x}_i) \tag{2.29}$$

where A_0 and $A_i, i = 1, \cdots, n$ are the fuzzy coefficients and $\overline{x}_i, i = 1, \cdots, n$ is a vector in the space chosen at first that we consider as the most accurate sample. The main advantage of the fuzzy regression model resides in its less restrictive application condition compared to the classic linear regression model. However, it suffers many disadvantages: it is very sensitive to extreme values and the estimation of the coefficients is usually a complex task. As a matter of fact, 2^n inequalities need to be solved with $2(n+1)$ unknown coefficients and $2N$ constraints where N is the number of samples.

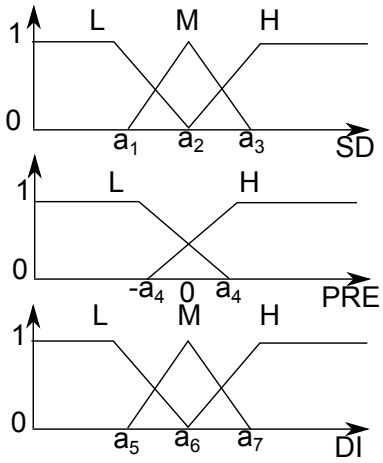


Figure 2.7: Fuzzy logic: input variables membership function. SD: Special days, PRE: Precipitation and DI: Discomfort index [31]

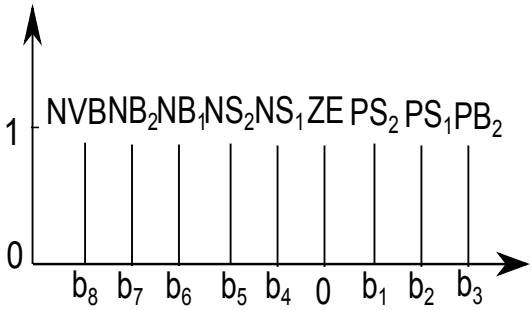


Figure 2.8: Fuzzy logic: output variables membership function. P: Positive, N: Negative, V: Very, B: Big, S: Small, ZE: Zero [31]

Expert systems. An expert system is a *computer program that has the ability to reason, explain and has its knowledge base expanded as new information becomes available* [73]. Building such systems relies on extensive knowledge gathered by human experts and transformed into simple IF-THEN rules. Its basic form does not allow the evolution of the rules while acquiring new knowledge. Since such systems show poor performances when dealing with new environments, new rules need to be coded into the systems in order to keep a good reliability on their decision making. Expert systems can handle thousands of rules [24] in order to take into account a large set of possible input scenarios. Forecasters based on such methods can be as reliable as the human experts designing them. Consequently, they perform very well when an extensive knowledge is available about the functional relationship between input and outputs. However, when dealing with very complex problems with insufficient knowledge about the model to be built, it appears as a poor choice.

2.1.b-iii Hybrid models

Hybrid models aim at taking advantage of the assets of various tools by combining them. Thus, they can avoid some of the drawbacks of the original methods. We can classify the hybrid models into three classes:

- Classical methods combined with classical methods: James W. Taylor mentioned in his work [13] that combining has particular appeal in cases when the methods are based on different information. He combined a weather-based method with those univariate methods [15] for one-hour-ahead load forecasts, and a simple average of the forecasts from the triple season ARMA and the Holt-winter methods [13] for a-day-ahead load forecasts.

- Artificial methods combined with artificial methods: [19] proposed combining the SOM and the ANN models for the STLF during the anomalous situations (holidays

and long weekends). S. Fan et al. have successfully applied a two-stage hybrid network of the SOM and the SVM for the next day's electricity price [74] and the electricity load forecasts [66]. In their works, the SOM clusters the input data set into several subsets and then a group of SVMs is used to fit the training data of each subset. I.Drezga et al. [75] have presented a new ANN based STLF. The ensemble of local ANN predictors was used to produce the final forecast, whereby the iterative forecasting procedure used a simple average of ensemble ANNs. A.A. Kusiak et al. [40] also suggested a five MLP ensemble model for the steam load forecast of buildings. I.Drezga et al. argued that by averaging individual forecasts that are generated by the identical modules over one and the same set of input data, the generalization of an ANN predictor can be improved [75].

- Classical methods combined with artificial methods: [76] described an approach to identify the best similar day parameters for ANN based electricity price forecast. T. Senjyu et al. presented several works by combing the RNN with the similar days approach. The forecasting load is obtained by adding a correction generated by RNN to the selected mean similar days' data (cf. figure 2.5). The load correction method was proposed to generate new similar days. The main purpose is, on the one hand, to train the neural network, and on the other hand, to deal with cases of shortage of similar day's data. They have successfully applied the method for one- [77], six-hour-ahead [78] and one-day-ahead [49] electricity load forecasts. In one of their studies, besides the RNN correction on the similar day's base, authors have added the FL approach that integrates other input parameters such as special days, precipitation and discomfort index values in order to refine the forecasting precision [31].

Apart from these methods, there exist also several artificial intelligence optimization methods that determine model's coefficients, such as gradient descent, Genetic Algorithm (GA), Particle Swarm Optimization (PSO) (also known as swarm-based optimization method), Artificial Fish Swarm Algorithm (AFSA), and Immune Algorithm (IA). The first one is a deterministic method, while the others are stochastic ones. According to the literatures, these stochastic methods adapt to extreme nonlinear situations and are particularly efficient in the dynamic environment, which the gradient descent cannot follow.

These stochastic algorithms begin with a random population and a fitness function, with which the efficiency of each solution can be evaluated and after each generation, the fittest survive. Their random, yet structured, search and parallel evaluation of the points are their main advantages. Therefore, they are capable of exploring and exploiting a given operating complex search space. Their operation on an encoded parameter string, but not directly on the parameters, enables the user to treat any optimization problem [79], such as regression models [80, 79], time series models [16, 81], and SVM [35]. Xuejun Chen et al. [82] have pointed out that GA methods have several disadvantages such as: slow convergence speed, sensitive to initial values, easily trapping into local optimum, premature convergence, and parameter selection problems. With the increasing length of number of individual parameters in the GA algorithm, the speed for optimal solution becomes unacceptable.

Compared to the similar optimization method GA, PSO is appreciated by its advantage, which is the rapid convergence to the global optimum, without being trapped in local

optima. However, the disadvantages of PSO are about its low precision, slow convergence in the later stage of the evolution, and parameter selection problems. They proposed AFSA for the SVM parameter estimation. According to them, AFSA can efficiently avoid local optimal operating points and thus can reach global optimality.

The usage of AFSA is flexible and shows a rather quick convergence rate. Nevertheless, none of them guarantees a convergence to the optimal solution at the end of the procedure. For further information we invite the reader to refer to paper [83], where the basic principles of GA and [36] of IA are described at length.

To conclude this subsection, it is worth mentioning that a comparison of some of these approaches is proposed in the literature. Indeed, paper [36] argues that SVR models combined with IA outperforms the SVR models with GA (proposed in [35]).

2.1.c Literature review conclusions and perspectives

The conclusions of this survey are threefold: firstly, researches seem to focus their energy towards techniques borrowed from the Artificial Intelligence community. As a matter of fact, we noticed during this last decade, the rise of new techniques in the field of power systems such as FL, SVR as well as variant forms of ANN models. Secondly, some authors still try to improve the traditional classical methods, although some of them seem to lose some interests because of their limited success [23], e.g., state space and Kalman filter modeling [27], and cubic splines [84, 85]. These latter gave hope to possible improvements, which could lead to promising results capable of competing with Artificial Intelligence based algorithms. Last but not least, hybrid methods become more and more popular as they combine the strength of several techniques (e.g., different models are cooperated in order to model the linear/non-linear parts, the rapidly fluctuating/slow fluctuating; to remove outliers, and to adapt coefficients of the models). These approaches lead to very interesting and promising results that may be confirmed in the next years.

In this chapter, we mainly discussed the classical approaches and artificial approaches applied to the load forecasting field. For the sake of clarity, we summarized the models and their features in table 2.2.

The literature review provides some interesting clues that could possibly improve the load forecasting result:

- Decomposing the original load data into a set of better-behaved constitutive series, the wavelet transformation can be used as a pre-treatment to filter noise data or outliers in the learning set for whatever load forecasting method.

- The clustering method, such as SOM, K-Means, and hierarchical clustering (the last two techniques haven't been discussed here, as they do not belong to load forecasting methods. They are very popular approaches for cluster analysis) can be applied to historical data to divide them into different categories, relying on which specific model to each category can be built.

- Fuzzy logic is based on the fuzzification and the defuzzification. The fuzzification converts the accurate information into fuzzy information and the defuzzification quantifies the fuzzy result into crisp values. This technique can be applied to uncertain

intrinsic variables such as the weather variables, which then can be used as inputs for various methods (for example, times series or neural networks, etc.).

Table 2.2: Summary of load forecasting approaches and their features. "√" signifies that the property of the model corresponds to the attribute. "×" signifies that the property of the model does not correspond to the attribute.

Method	**very STLF**	**STLF**	**MTLF**	**LTLF**	**Linearity**	**Long historical**	**Features**
Regression model	×	√	√	√	×	√	Simple, explicit relationship with exogenous variables
Time series	√	√	√	×	√	×	Extract periodocities, historical data and errors, including: Box-Jenkins, exponential smoothing, etc
Similar day approach	×	√	×	×	√	√	Nonparametric, delicate coefficient adjustment, provide learning set for AI methods
End-use method	×	×	√	√	×	×	Bottom-up, each housing device, require for extensive information
Econometric approach	×	×	√	√	×	√	Including social and economical variables, complex equations
ANN	√	√	√	×	×	√	Universal approximator, great learning ability, mapping complex relationship, delicate in structure selection, heavy computation
SVM	√	√	√	×	×	√	Non-linear function mapping input(s) to a higher dimensional space
FL	√	√	×	×	×	×	Fuzzification, defuzzification, absence of mathematical model, similarity, fuzzy linear regression, complicated parameter estimation
Expert system	√	√	×	×	×	√	Computer based, IF-THEN rules, require frequent update

2.2 Data description

For the sake of building short-term load forecasting models on the MV level, we have in possession the measurements on seven MV/LV substations and on four MV feeders. Before selecting the forecasting approach, we start by observing these data. The purpose of the observing is to give some evidence on the nature of the data: what certain behavioral "components" present in the data, what influencing factors can possibly explain the variation of the data, etc. These specifications in the data can help us in selecting the models with potentials to produce the best forecasts.

2.2.a MV/LV substation

The data used to validate our methodology are the real measurements collected from the French distribution network in the experimental phase of the "Linky" project. Mainly, there are seven load curves, and each one represents the consumption of a specific MV/LV substation, sampled every thirty minutes, in a same region, during the same period from September 9, 2009 to October 27, 2010. As an important factor affecting the consumption level, the temperature of the region is also provided on an hourly basis during the corresponding period (figure 2.9). Hourly temperature data are linearly interpolated to obtain the thirty minute temperature data. Each substation load curve represents the sum of all connected clients' consumption.

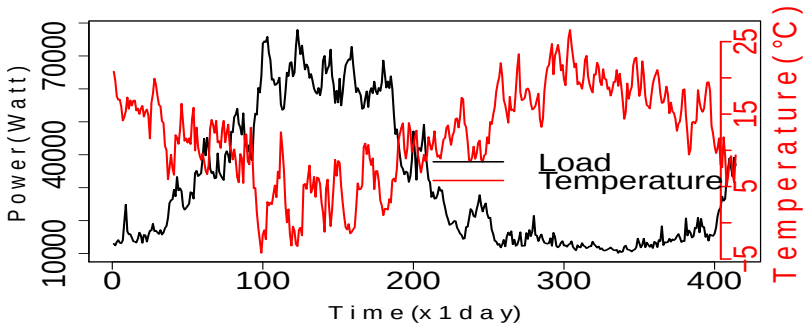


Figure 2.9: Daily average load and temperature data through 414 days (from Sept. 9, 2009 to Oct. 27, 2010) of *substation CE_MOU* (mainly residential)

Figures 2.9 , 2.10 and 2.11 are examples of daily average substation loads and the temperature variation through 414 days. We can observe that the load variation can significantly differ from substations. We aim, in this subsection at explaining these different behaviors in order to identify the impact factors for the forecasting model design. The mainly residential substation CE_MOU (figure 2.9), including 61% domestic , 23% service sector, and 16% industrial clients, where the percentage refers to the total power supply provided for the clients in each category, follows the variation of the temperature. The substation VI_LOG (figure 2.10), composing of one third service sector clients and two thirds industrial clients, follows a weekly cycle and varies with the temperature. The substation CE_FRO (figure 2.11) suppling electricity to a particular industrial client stays stable all year long and follows an evident weekly cycle.

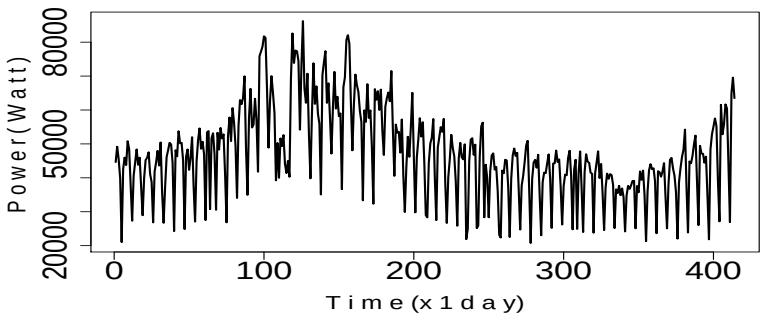


Figure 2.10: Daily average load through 414 days (from Sept. 9, 2009 to Oct. 27, 2010) of *substation VI_LOG* (mixed service sector and industrial)

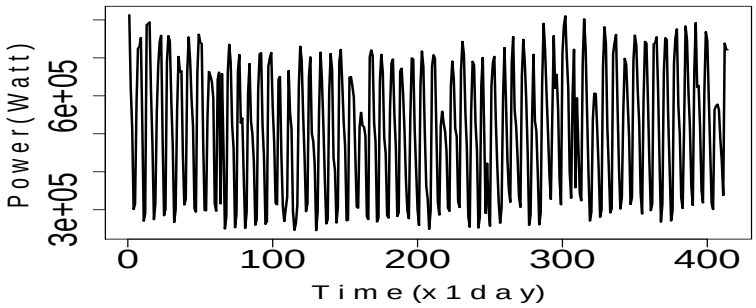


Figure 2.11: Daily average load through 414 days (from Sept. 9, 2009 to Oct. 27, 2010) of *substation CE_FRO* (an industrial client)

2.2.a-i Temperature influence

A correlation coefficient is computed to illustrate the substation's sensitivity regarding the temperature:

$$\rho_{X,Y} = corr(X,Y) = \frac{cov(X,Y)}{\sigma_X \sigma_Y} = \frac{E(X - E(X))E(Y - E(Y))}{\sigma_X \sigma_Y} \tag{2.30}$$

Where $E(\cdot)$ is the expected value operator, $cov(\cdot,\cdot)$ represents the mean covariance, $E(X)$ and $E(Y)$ are the expected values of the random variables X and Y, σ_X and σ_Y are the correspondent standard deviations. The value of the correlation coefficient is between -1 and 1, and it indicates the linear dependence between the two variables X and Y. The closer is to the values -1 or 1, the stronger the correlation is.

Thus, relying on equation 2.30, the correlation coefficient computed between the CE_MOU's load curve and the temperature equals -0.78, where the minus sign indicates that the energy consumption and the temperature evolves in an opposite direction. Indeed, when the temperature drops, it becomes colder and people tend to turn on their electric heaters, which increases electricity consumption.

The clients' compositions as well as the correlation coefficients with temperatures of the seven substations are summarized in table 2.3.

Table 2.3: Seven substation clients' compositions and correlation coefficients with temperatures

Substation	Correlation coefficient	Residential	Service sector	Industrial
		Number of clients/ Total power (Percentage)		
CE_MOU	-0.78	40/351 KVA (93.6%)	2/24 KVA (6.4%)	0/0 KVA (0%)
CE_CHA	-0.78	26/229 KVA (97.5%)	1/6 KVA (2.5%)	0/0 KVA (0%)
CE_FRO	0.07	0/0 KVA (0%)	0/0 KVA (0%)	1/1600 KVA (100%)
CE_CER	-0.78	73/629 KVA (88.6%)	6/81 KVA (11.4%)	0/0 KVA (0%)
VI_LOG	-0.24	1/3 KVA (1%)	5/108 KVA (36.4%)	3/186 KVA (62.6%)
VI_PRI	-0.81	35/285 KVA (69.3%)	2/18 KVA (4.4%)	1/108 KVA (26.3%)
VI_PAU	-0.83	62/545 KVA (61.2%)	16/201 KVA (22.6%)	3/ 144 KVA (16.2%)

We can see from the table 2.3 that according to the correlation coefficient's magnitude, three categories can be divided: one, the industrial substation CE_FRO is independent to the temperature variations. Two, the mixed substation VI_LOG is relatively influenced by the temperature. Three, all other mainly residential substations are sensitive to the temperature variations. This preliminary analysis suggests that we should establish different forecasting models (with or without taking into account the temperature variations) for these three categories of substations. We will explain in chapter 3 that time series method has a limit on the weather non-sensitive substations.

2.2.a-ii Day type influence

Another type of information that can be exploited is the calendar information. As a matter of fact, different patterns depending on the type of day, e.g., a working day or a public holiday, can be easily distinguished in the "industrial" substation's case.

Notice (figure 2.12) that the level of power consumption during a public holiday is more similar to weekends than normal working days. However, for the weather sensitive substations, it is hard to draw a conclusion by a simple observation on the load curves, since the load level varies with the temperature.

2.2.a-iii Time influence

The last impact factor that we intend to examine is the "time". In order to find out whether day of the week has some impact on the load curve, a *similarity index* is designed. This later indicates a similarity degree between the current day with the previous days. Here, we aim at figuring out the cyclic patterns in the load data. Therefore, the temperature influence is exempt by removing the daily average from the data. The similarity index is based on the Euclidean distance between two zero-centered days.

Let $x = (x_1, x_2, \cdots, x_n)$, $y = (y_1, y_2, \cdots, y_n)$ be two vectors. Their Euclidean distance is defined as:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2} \tag{2.31}$$

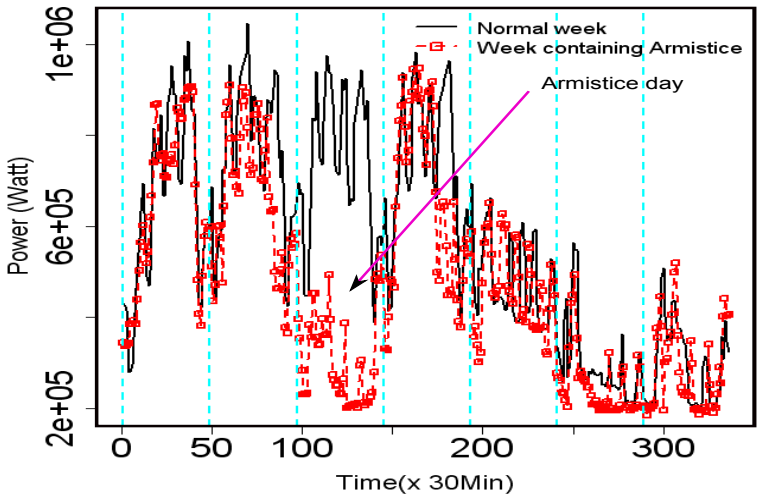


Figure 2.12: Normal week compared to the "Armistice" week (from Nov. 9, 2009 to Nov. 15, 2009) of *Substation CE_FRO* (an industrial client). Each pair of dotted lines separates a day.

In our case, x and y represent the vectors of 30-minute-step measurements collected in the days to compare. Thus, the size of the vector is $n = 48$, which is the number of electricity consumption samples recorded during a day. The smaller the Euclidean distance is, the more similar the two vectors (days) are. The similarity index is calculated as follows:

1. Remove the daily average from each day samples: $x'_i = x_i - \overline{x}_A, \{i \in A\}$, where $\overline{x}_A$ is the correspondent daily average of day A.

2. Calculate an Euclidean distance matrix with all the zero-centered sample days. e.g., "N" is the total number of days, the Euclidean matrix is of "N×N" dimension.

3. For each row in the Euclidean distance matrix, choose the 5 smallest Euclidean distances, which represent the 5 most similar days to the chosen day, and record their lagged-day numbers relative to this day.

4. The similarity index equals the number of lagged day divided by its total possible number. e.g., for the number of lagged day equal to 5, the similarity index is the number of selected similar pairs divided by its total number "N-5".

Since the daily average power is removed from the data, the distances in the Euclidean matrix only reflect the similarity in the intraday variation of the load pattern, independent to their consumption levels.

Figure 2.13 shows the similarity index for all lagged days. Furthermore, we withdraw the influence of the day type from the data as well. Figure 2.14 shows the similarity index for lagged days without weekends and holidays. The high similarity indexes in both figures on the one lagged day and on the multiplicative weekly lagged days indicate daily and weekly cyclic patterns. The cyclic patterns can also be observed in the frequency domain by plotting a sample analog of spectral density, i.e., the periodogram [14]. We consider

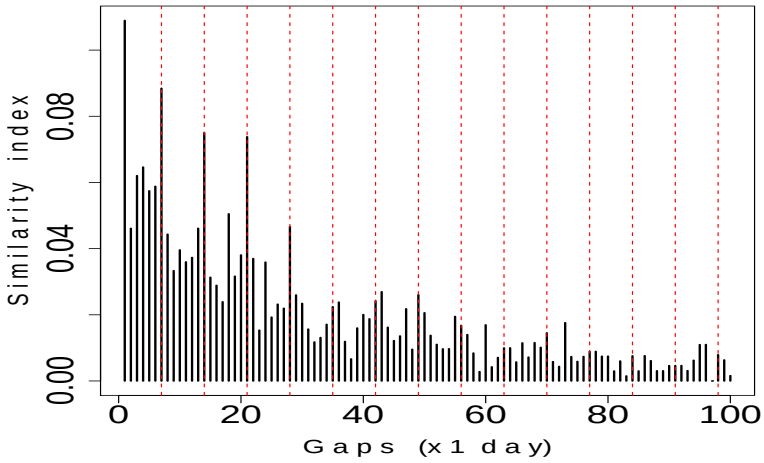


Figure 2.13: Similarity index calculated based on all days of *substation CE_MOU*. The dashed lines represent the multiplicative values of 7 lagged days. They stand for the similarity degrees with the same day of the previous week, the same day of the previous two weeks, etc.

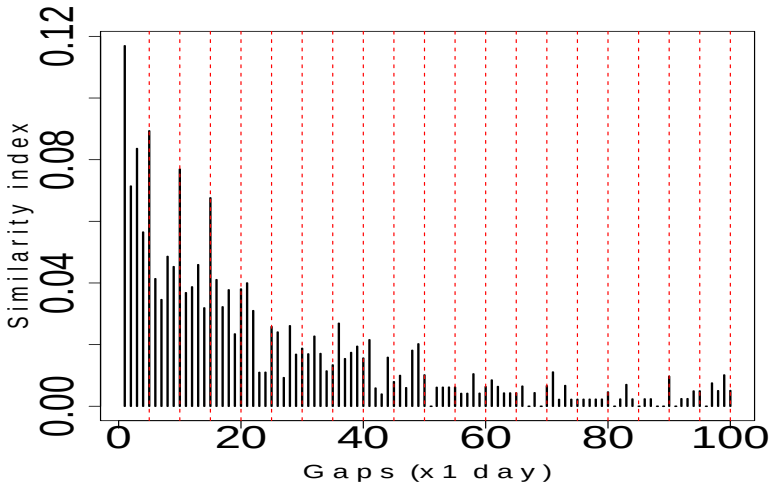


Figure 2.14: Similarity index without weekends and public holidays of *substation CE_MOU*. The dashed lines represent the multiplicative value of 5 lagged days.

that as the first step, examining periodicities in the time domain is more comprehensive than in the frequency domain. More details concerning the periodogram will be presented in the section 3.2.c.

We can conclude that the load pattern of a substation depends largely on the composition of the clients connected. The residential, service sector and industrial clients have their own specific load curve pattern:

- Residential load curve varies with the temperature: the electricity consumption level increases as the *temperature* drops because the use of the electrical heating device.
- The industrial client is not influenced by the change of seasons or the temperature variation, but has an evident *weekly period* and is sensible to the *national holidays*.
- The service sector client is between the two types: being influenced by the temperature as well as having a *weekly pattern*.

Thus, being the sum of a mixed categories of clients, the substation's electricity load mainly depends on the *temperature*, the *day type* ,and the *time*.

2.2.b MV feeder

A more aggregated level, load curves of MV feeders are also used for the validation of our methods. Four MV feeders' loads are considered in the study. Figure 2.15 shows the composition of two MV feeders as examples. Each represents the electricity demand of several MV/LV substations presented in section 2.2.a. The observations are collected on a half-hourly basis from September 9, 2009 to September 22, 2010. The other two MV feeders not included in the figure are also in the same region. Being a mixture of residential, service sector and industrial clients, patterns of these MV feeders' loads are more similar to the VI_LOG substation's case (figure 2.10), namely, having weekly patterns and being influenced by the temperature.

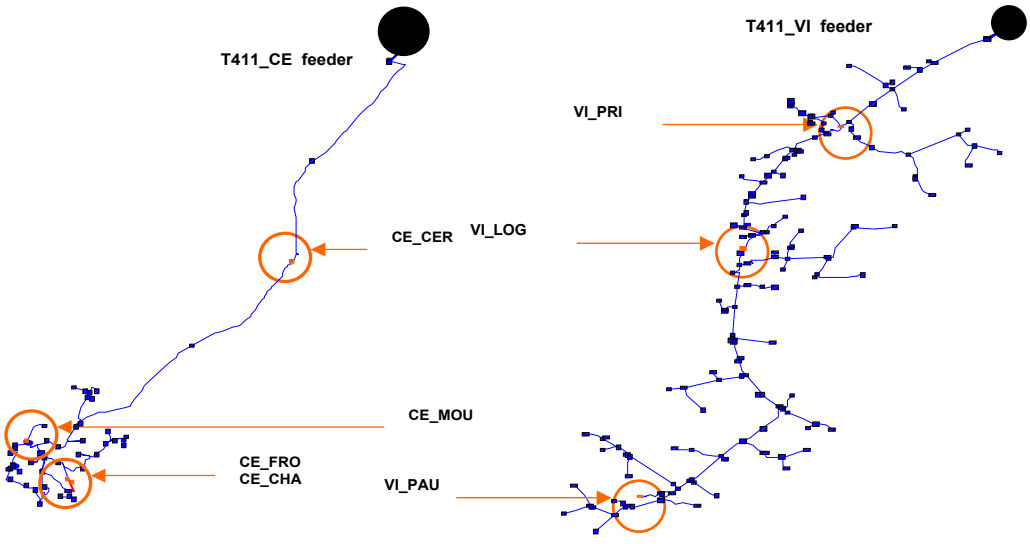


Figure 2.15: MV feeders and position of connected MV/LV substations

2.3 Choices of Time series and NN methods

Most forecasting models and methods presented previously have already been tried out on load forecast, with varying degrees of success. According to our objectives, our focus is on the short-term MV level. We can see in figure 2.1 that most of the models exist on the

HV level. The idea is to borrow algorithms on the HV level and adapt to the MV data. Compared to the HV level data, MV level data have more irregular patterns.

The choice of the model depends mainly on the nature of the data [86]. From the section 2.2, we concluded that our load examples are influenced by the nature of the customers connected, the temperature, the time (of the day and of the week), as well as the calendar information. Thus, from the presented two categories of methods, we have selected the time series method and the ANN model to address to our issue. Different from the Box-Jenkins method, we formulate our time series model in a regressional way, in which the physical interpretation can be attached to the components and explanation can be easily reached. Advanced signal processing tools are integrated in the time series method in order to get the best accuracy. Another method chosen from the artificial intelligent method family is the ANN method, more precisely, the Multi Layer Perceptron (MLP) structure ANN. It is the feed-forward ANN, the simplest to implement and the less computationally expensive to run. Moreover, A. Khotanzad et al. have declared in [20] that according to their investigation of the use of different structures of ANNs, no major advantage over the MLP of adopting a more complicated structure has been discovered. Different from most of the implementations in the literatures, the efforts are concentrated on the optimal structure selection. As explained, ANNs are often accused for its large architecture possibility and lack validity of its model [87]. Furthermore, both selected methods have sound basis for the confidential interval computation. This later gives indication to the uncertainty of the forecasting model.

2.4 Performance criteria and reference case

2.4.a Performance criteria: MAPE and MAE

The comparison of the model's performance relies on the computation of two quantities: MAPE and Mean Absolute Error (MAE). They are presented throughout this paper as performance indexes.

On the one hand, MAPE represents the average ratio between the absolute errors and real observations. The result is given in percentage:

$$MAPE(\%) = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| * 100 = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{e_t}{y_t}\right| * 100 \tag{2.32}$$

where N is the total number of forecasting samples, $\hat{y}_t$ is the forecasting value, y_t is the real measured value and e_t is the error of the model which is the difference between forecasting values and real values.

MAPE represents the accuracy of the forecasting models in the percentage of the real observations. Thus, the value of the MAPE is dependent to the real observations. More specifically, committing the same magnitude of errors, during the winter, when the electricity consumption level is high, the MAPE is low; during the summer, when the electricity consumption level is low, the MAPE would have an important value. In order to avoid this drawback mainly caused by the forecasting magnitude, another performance indicator is introduced.

The MAE indicates the average absolute error result. The result given in our example is in Watt:

$$MAE(W) = \frac{1}{N}\sum_{t=1}^{N}|y_t - \hat{y}_t| = \frac{1}{N}\sum_{t=1}^{N}|e_t| \tag{2.33}$$

The MAPE and MAE are the most widely used performance quantities in the load forecasting field.

2.4.b Reference case: the naive model

In order to realize a first benchmark, a naive model [12, 60] is designed to do the load forecast. The naive model does not exploit explanatory information such as weather, neither does it rely on sophisticated statistical or machine learning techniques.

The naive model replaces the forecasting result (for day D) by its most similar historical consumption pattern (day D-ℓ), where "ℓ" is the general delay between the most similar days which are defined by computing the euclidean distances among all days.

The naive model behaves in two different ways depending on the clients' composition. For the substation principally connected to domestic clients, as the consumption's level is mainly due to the variation of the temperature, its load pattern is most similar to the day before. As a result, in this case $\ell = 1$. While for an industrial substation, because of its weekly pattern, the most similar day is more likely to be the same day of a week before. Therefore, $\ell = 7$.

2.5 Conclusion

This chapter is served as an introduction to our short-term load forecasting objective. Firstly, the various approaches in the literatures are presented. Their applications in the related works are briefly introduced. Next, the data of the MV/LV substations and the MV feeders that are used in our study are examined. We pointed out that the main influence factors for the variations in those load curves are the temperature, the day type, and the time. Consequently, the choices of load forecasting methods being the time series and the NN methods are argued. Performance criteria, based on which the accuracy of different methods is compared, and the reference case are also set up.

In the next two chapters, the chosen methodologies are thoroughly stated.

Chapter 3

Time series model

Contents

Abstract

After a thorough scrutiny into the different load forecasting methods, we have chosen the time series and the neural network methods to answer the request of the short-term load forecast. We tackle in this chapter an additive time series method, dividing our analysis into three parts; namely, a trend component, a cyclic component, and a random error component. The first two parts are deterministic and are estimated separately with dummy variable regression models. The random error part is looked through in detail to ensure the well being of the designed model. Advanced signal processing tools are integrated in the proposed method such as ANOVA nullity test and smoothed periodogram. The result of the suggested method is encouraging, compared with the naive model. Weather uncertainty as an important aspect to the load forecast is also discussed at the end of the chapter.

In this chapter, we address the short-term load forecasting issue with an additive time series method. Being a linear model, the additive time series model is widely used for modeling for its ease of implementation, understanding, high precision and computational efficiency. In section 3.1, we introduce the additive time series model and the overall procedure of our designed methodology. Next, two sections are developed respectively describing theories and techniques including in the proposed procedure, and showing every step in the application of the model on a substation load example. In section 3.4, we confirm the conclusion obtained in some works that the weather uncertainty decreases the model's forecast precision and argue that our proposed methodology still largely outperforms the naive model.

3.1 Additive time series model and procedure overview

Time series [88] represents the evolution of a set of observations sampled at a regular interval along time. The specificity of time series models, compared with other statistic methods, is the introduction of "time" as one of its explicative variables.

The considered time series model contains three parts: a trend component, a cyclic component and a random error. Let y_t denote the measured load at time t, the additive model then can be written as:

$$y_t = f_t + S_t + \epsilon_t \tag{3.1}$$

where f_t denotes the trend component at time t, S_t denotes the cyclic component at time t and ϵ_t denotes the random error at time t. The first two components are deterministic parts, while the random error part can be examined after the model estimation to guarantee its good performance. We aim at building two parametric models representing the two deterministic parts, which are termed respectively the trend and the cyclic models.

Figure 3.1 describes the process to build the time series forecaster. The historical consumption loads, the correspondent historical temperatures, and the historical day types are used for the conception and parameter estimation of the models. The forecasting temperature data and the forecasting day types are applied as inputs for the parameterized model in order to get the forecasting results. The categorical variables, i.e., the dummy variables, are integrated to distinguish the different day types in the trend model. The cyclic model is composed of Fourier components, whose main frequencies are found by the smoothed periodogram. Additive time series model considers the sum of the forecasting results of the two models as the final load forecast (equation 3.1). At the end of the forecasting step, the estimation window, indicating the length of the historical data, with which the parameters are estimated, is shifted to the next period, ready to get the next forecasting process started.

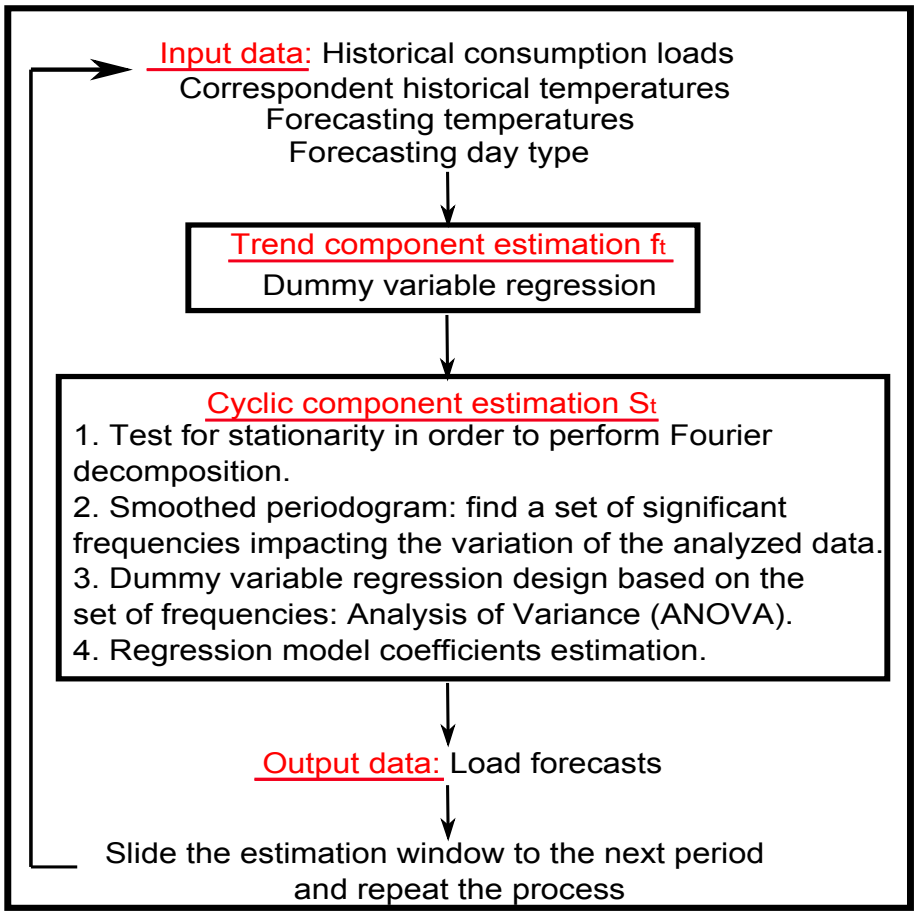


Figure 3.1: Steps of the designed time series forecasting method

3.2 Statistical tools

3.2.a Dummy Variable Regression

We have concluded in the chapter 2 that one of the most important influence factors that explain the variations in the load curves is the day type. Here, in our specific example, the load curves can be classified into three day-type categories: the working days, the Saturdays, and the Sundays and national holidays. The explanation of this category division is later carried on in subsection 3.3.a. Categorical variables are often employed for the injection of categorical information to models.

Dummy variable regression is a method for incorporating categorical variables into a regression model. Dummy variable takes the value of either 0 or 1, and indicates the presence of its coefficient in the equation. Let $\Gamma(\cdot)$ denote a regression equation, the dummy variable regression can be written as:

$$y_t = \Gamma(x_t) + \sum_{\alpha=1}^{\kappa-1} D_\alpha \Gamma_\alpha(x_t) \tag{3.2}$$

where $D_\alpha, \alpha = 1, \cdots, \kappa - 1$ are dummy variables, $\Gamma_\alpha(\cdot)$ denotes the regression equation associated to the dummy variable D_α, κ is the number of different categories and x_t, y_t represent respectively the independent variable vector and the studied variable. The number of the

dummy variables in the equation is $\kappa - 1$, as the category that all dummy variables equal to zero is considered as the other categories' reference. For a given category, at most one dummy variable is equal to 1.

The dummy variable regression is applied both for the trend component and for the cyclic component.

3.2.b Trend Component Estimation

The trend function f_t usually represents a slow variation of the analyzed quantity y_t. It is usually modeled by a linear function, a polynomial function or an exponential function. In section 2.2.a-i, the load variation (except the industrial substations) is mainly proved to be very linearly dependent to the temperature data. As a first approach, the simplest form, the linear trend model is conceived. Thus, here we suggest both a time and temperature correlated linear trend function.

$$f_t = at + b + cT_t + \sum_{\alpha=1}^{\kappa-1} D_\alpha \gamma_\alpha \tag{3.3}$$

where t refers to the time index whose value is from 1 to the estimation window size, T_t refers to the temperature of the region where the measurements took place, and $\sum_{\alpha=1}^{\kappa-1} D_\alpha \gamma_\alpha$ represents the dummy regression part. The parameters a, b, c and $\gamma_\alpha, \alpha = 1, \cdots, \kappa - 1$, are the coefficients of the regression model.

Thus, in this section, we aim at evaluating the $\kappa + 2$ regression coefficients $\{a, b, c, \gamma_\alpha\}$ relying on the Ordinary Least Square (OLS) approach. To that purpose as a first approximation, the sum of the cyclic part during one period is assumed to be zero. In order to adapt to our short-term forecasting problem, a sliding window is performed, within which we estimate the coefficients. With the computed coefficients, the day type of the forecast day as well as its temperature prevision, the trend component can be predicted. Then the sliding window is shifted for one day to the next period, within which the coefficients are recalculated. This procedure is repeated until the end of the data. Since we do not possess the temperature forecasts, we use the real temperature measurements instead. This is a common practice [26, 27, 28]. We analyze the impact of the weather uncertainty on forecasting model's accuracy later in section 3.4.

3.2.c Cyclic Component Estimation

Here, we complete our forecasting model by taking into account the periodicity of the observations. The cyclic part captures the cyclic behaviors of the load y_t. Namely, let p denote the periodicity of the function without trend, then we can write for all samples t: $S_{t+p} = S_t$, where S_t represents the cyclic component at t. The sum of the cyclic components over an entire period equals zero: $\sum_{t=1}^{p} S_t = 0$.

Once the trend model has been determined, it can be removed from the raw data y_t. Let W_t denote the detrended series:

$$W_t = y_t - f_t = S_t + \epsilon_t \tag{3.4}$$

Sum of the sine and cosine functions, the Fourrier series can fully describe the dynamism of the stationary periodic signals. Thus, we suggest performing the cyclic estimation using

a Fourier component regression. To do so, four steps are needed (figure 3.1). First, we test the detrended series W_t to see whether they are stationary. Then, a smoothed periodogram is performed in order to find out the harmonic frequencies that explain the variations in W_t. Next, with these harmonic frequencies, a regression model of the Fourier components can be constructed. Last, an ANOVA nullity test [89] integrated OLS approach is applied to the coefficient estimation in the sliding window for the cyclical part. The ANOVA nullity test is used for the discrimination of the significant regression coefficients. This technique will be presented in subsection 3.2.g.

3.2.d Tests of Stationarity

The purpose of performing tests of stationarity is to find out whether the detrended series W_t is stationary in order to ensure the primary condition of carrying out the smoothed periodogram presented in the next subsection. Several approaches exist, among which we chose two: the Kwiatkowski-Phillips-Schmidt-Shin tests (KPSS) test [90] and the Augmented Dickey-Fuller test (ADF) test [91]. The KPSS tests the null hypothesis of stationary time series against non-stationarity, while the ADF tests in the time series samples the existence of a unit root, which signifies the non-stationarity against stationarity. These two tests are jointly used to ensure a reliable result. For more information, readers can refer to appendix B.

Both tests confirm that the analyzed load data set is well detrended and is stationary with any size of the sliding window.

3.2.e Smoothed Periodogram

The purpose of using the smoothed periodogram is to find the main harmonic frequencies of the cyclic component in order to constitute the cyclic regression model. For a stationary data set, its periodogram can be defined as [88]:

$$I_n(\nu_j) = \frac{1}{n}\left|\sum_{t=1}^{n} W_t e^{-2i\pi\nu_j t}\right|^2 = d_c^2(\nu_j) + d_s^2(\nu_j) \tag{3.5}$$

where n is the sample size, $\nu_j = \frac{j}{n}$, $j = \{0, 1, \cdots, n-1\}$ are the fundamental frequencies of the Discrete Fourier Transform (DFT) of the data set W_t. $d_c(\nu_j)$ and $d_s(\nu_j)$ are respectively the normalized real and imaginary component of the performed DFT.

$$\begin{aligned} d_c(\nu_j) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} W_t \cos(2\pi t\nu_j) \\ d_s(\nu_j) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} W_t \sin(2\pi t\nu_j) \end{aligned} \tag{3.6}$$

We have:

$$\sum_{t=1}^{n}(W_t - \overline{W})^2 = 2\sum_{j=1}^{m}[d_c^2(\nu_j) + d_s^2(\nu_j)] = 2\sum_{j=1}^{m} I(\nu_j) \tag{3.7}$$

where $m = \frac{n-1}{2}$, and $\overline{W}$ stands for the average of the data set W_t. Equation 3.7 indicates that the sum of squares can be partitioned into harmonic components represented by the periodogram's amplitude at frequency ν_j. In other words, if there exist peaks in the periodogram, these frequencies explain the variation of the data.

However, the raw periodogram usually has a large variance at a given frequency. Consequently, its raw form is not directly used as an estimator of the spectral density function. One solution to reduce the estimator's variance is to use the smoothed periodogram.

It is assumed that the spectral density is fairly constant in the considered frequency band, and the adjacent frequencies have asymptotically independent values. We can set up a symmetric smoothing filter B of size $2l+1 << n$, which is centered around a frequency ν_j such that:

$$B = \left\{ \nu : \nu_j - \frac{l}{n} \leq \nu \leq \nu_j + \frac{l}{n} \right\} \tag{3.8}$$

The Daniell kernel [92] is a set of symmetric positive weights that center in the estimated frequencies. The sum of all weights is 1 :

$$h_k = h_{-k} > 0 \; and \sum_{k=-l}^{l} h_k = 1 \tag{3.9}$$

Thus, the smoothed periodogram becomes

$$\widetilde{I}(\nu_j) = \sum_{k=-l}^{l} h_k I_n(\nu_j + k/n) \tag{3.10}$$

Through the smoothed periodogram plot, frequencies $\mathcal{F}$ with the most significant amplitudes can be identified.

3.2.f Regression Model with Fourier Components

We build the cyclic component based on the set of frequencies $\mathcal{F} = \{\nu_1, \nu_2, \ldots, \nu_{N_f}\}$ such that:

$$\begin{aligned} S_t = & \sum_{i=1}^{N_f} c_i \cos(2\pi\nu_i t) + \sum_{i=1}^{N_f} s_i \sin(2\pi\nu_i t) + \\ & \sum_{\alpha=1}^{\kappa-1} D_\alpha \left(\sum_{i=1}^{N_f} c_{i,\alpha} \cos\left(2\pi\nu_i t\right) + \sum_{i=1}^{N_f} s_{i,\alpha} \sin\left(2\pi\nu_i t\right) \right) \end{aligned} \tag{3.11}$$

where N_f is the total number of frequencies in $\mathcal{F}$, and $\sum_{\alpha=1}^{\kappa-1} D_\alpha \times (\cdot)$ stands for the dummy variable regression part. As described in Equation 3.11, every frequency of the set $\mathcal{F}$ provides two contributions: a sine component and a cosine component. The $2\kappa N_f$ unknown coefficients in the equation are to be determined in a sliding window using the OLS approach.

3.2.g ANOVA Nullity Test

In order to determine the relevancy of the coefficients, and improve the efficiency of coefficient estimation, an ANOVA nullity test is performed.

Mainly the ANOVA test aims at partitioning the observed variance into several variances due to each explicative variable and the residual. The significance of every coefficient is identified by its ratio of within-group variance to the global variance. As the value of the ratio is larger, the coefficient is more significant (cf. appendix C).

Practically speaking, we start with an ANOVA nullity test with all parts of the regression model. Then, after the removal of non significant parts according to the test's result, the rest of the coefficients are re-estimated.

3.2.h Complete Forecasting Model

As described in the beginning of the section, the forecasting result of the time series model is the sum of the trend part forecast and the cyclic part forecast.

Thus, the complete forecaster can be expressed as:

$$\begin{aligned}\hat{y}_t &= f_t + S_t \\ &= at + b + cT_t + \sum_{i=1}^{N_f} c_i \cos(2\pi\nu_i t) + \sum_{i=1}^{N_f} s_i \sin(2\pi\nu_i t) + \\ &\sum_{\alpha=1}^{\kappa-1} D_\alpha \left(\gamma_\alpha + \sum_{i=1}^{N_f} c_{i,\alpha} \cos(2\pi\nu_i t) \sum_{i=1}^{N_f} s_{i,\alpha} \sin(2\pi\nu_i t) \right)\end{aligned} \tag{3.12}$$

The error part: $\epsilon_t = y_t - f_t - S_t$ is examined in Section 3.3.c.

3.3 Application example results

The results has been realized with the open source software "R" [93], which is dedicated to statistical computing and graphics. We chose "R", since it has a variety of up-to-date packages [94] that support all our functional statistic techniques. The sample period of consumption and temperature data is from September 9, 2009 to October 27, 2010. The available data is divided into two parts: a training set and a test set (figure 3.2). A training set is used to define the value of some important variables such as number of categories, the length of the sliding window, and the important frequencies. A test set is the data based on which the performance of method is evaluated and compared with the naive model. The choice of the size of these two sets is arbitrary: about one third of data (117 days) for the training set and two thirds (297 days) for the test set.

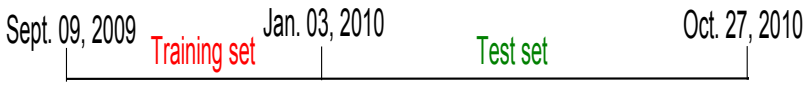


Figure 3.2: Training set and test set periods of the available data.

3.3.a Training set

Figure 3.3 shows the load variation from October 5, 2009 to October 11, 2009 (a whole week from Monday to Sunday) as an example and suggests three apparent categories: weekdays, Saturdays, and Sundays. Notice that in France, activities related to service sectors are closed on both Sundays and holidays, whereas usually, most of them are open on Saturdays. Consequently, national holidays are included in the same category as Sundays.

As presented in the section 3.2.b, a sliding window is performed to estimate coefficients of both the trend and cyclic parts. The size of the sliding window obviously has an impact

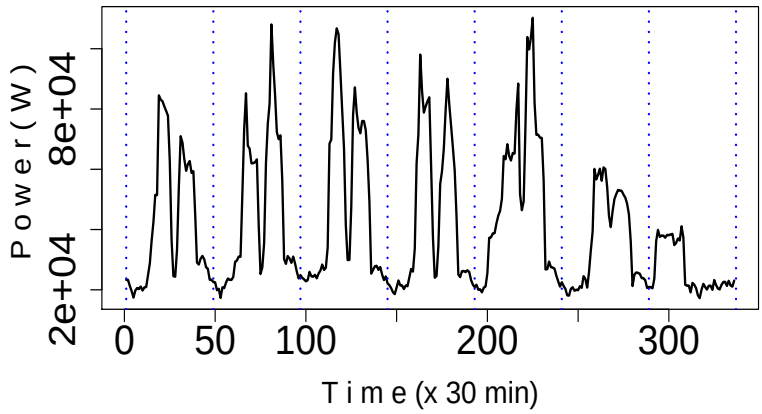


Figure 3.3: A weekly consumption pattern (October 5, 2009 to October 11, 2009) of a mixed industrial and service sector *substation VI_LOG*. Each pair of dotted lines defines one day.

on the forecast precision. With a larger window size, the trend model takes into account the slow variation of the time and temperature, which leads to a higher accuracy.

The sliding window size is selected by the MAE criteria computed on the intermediate performance of the trend forecasts with different window sizes. The smallest unit of the sliding window length is a week within which at least one day of each category can be found. In this chapter, the mixed industrial and service sector *Substation VI_LOG* is used as an illustrative example. Figure 3.4 presents the MAE criteria for the different sliding window size results.

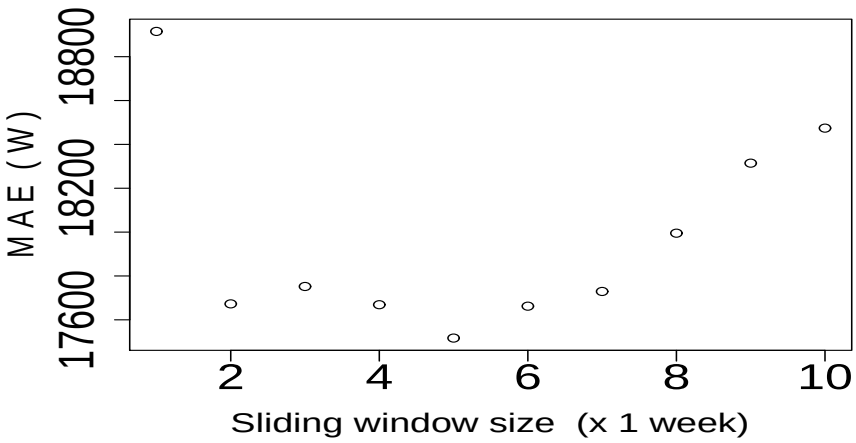


Figure 3.4: *Substation VI_LOG*, MAE criteria calculated on the training set (117 days) for different sliding window sizes (weeks)

Based on the above results, the sliding window length is fixed up to five weeks.

The estimation of the trend component is removed once determined. The detrended series W_t is proved to be stationary by both KPSS test and ADF test.

A smoothed periodogram is performed on the detrended series, with a Daniell kernel (7,7) filter [95, 96], which is a Daniell filter of size 7 (length 15) applied twice. The length of the filter is selected by trial and error.

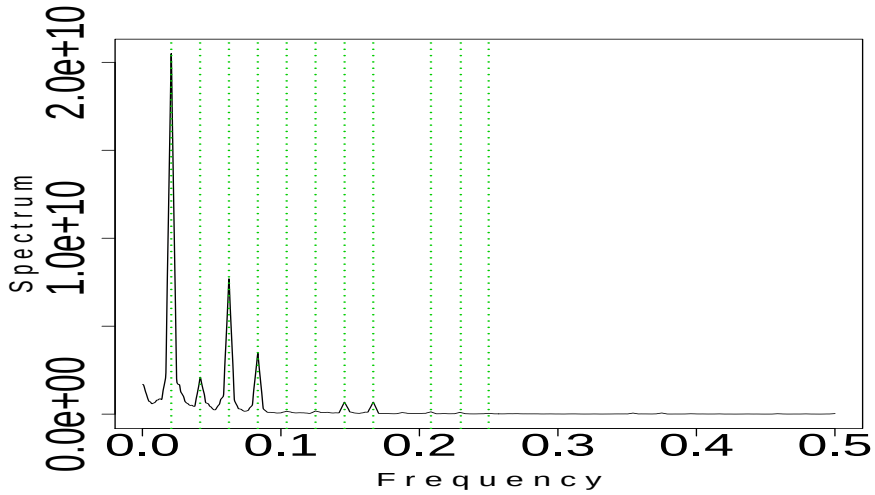


Figure 3.5: Periodogram of the detrended training data set smoothed by the Daniell kernel (7,7). The dotted lines indicate the considered harmonic frequencies, which are $\{1/48 = 0.021, 1/24 = 0.042, 1/16 = 0.063, 1/12 = 0.083, 1/9.6 = 0.104, 1/8 = 0.125, 1/6.85 = 0.146, \cdots\}$.

Based on equation 3.10, the smoothed periodogram of the detrended training data set is shown in figure 3.5. The dotted lines point out the set of the most significant frequencies $\mathcal{F} = \{\nu_1 = 1/48, \nu_2 = 1/24, \nu_3 = 1/16, \nu_4 = 1/12, \nu_5 = 1/9.6, \nu_6 = 1/8, \nu_7 = 1/6.85, \nu_8 = 1/6, \nu_9 = 1/5.35, \nu_{10} = 1/4.8, \nu_{11} = 1/4.35\}$. They represent the periodicities of one day, half day, 8 hours, and 6 hours, etc. These frequencies are then assigned to build the cyclic part according to the equation 3.11.

3.3.b Test set

With the parameters defined by the training data set, in this subsection, we present the performance of the designed forecasting model compared with the naive model on the test data set. The trend estimation model and the cyclic estimation model are independently set up according to the equations 3.3, 3.4 and 3.11. The coefficients are estimated with the past five weeks sliding window for the different three day type categories. The ANOVA nullity test is integrated into both the trend and the cyclic models to remove the irrelevant parts. Combining the trend and the cyclic forecasts, the final forecasting results are computed with the correspondent coefficients depending on the forecasting day type.

Figure 3.6 shows the overall results of the average daily forecasting results compared with the real measurements. Figures 3.7 and 3.8 show some more detailed results on a 30-minute basis of the forecasting model on the weekdays and weekends compared with the real measurements.

Composing of one third service sector clients and two thirds industrial clients, the

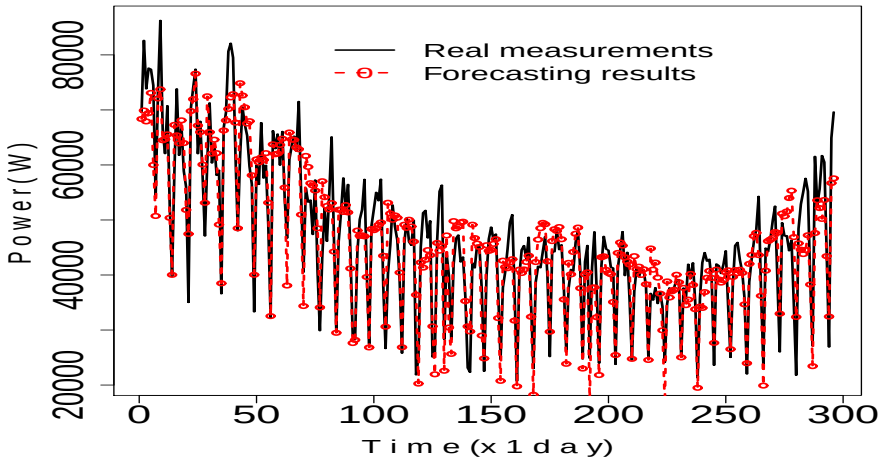


Figure 3.6: *Substation VI_LOG*, comparison of the forecasting results with the real measurements on the test set period (297 days).

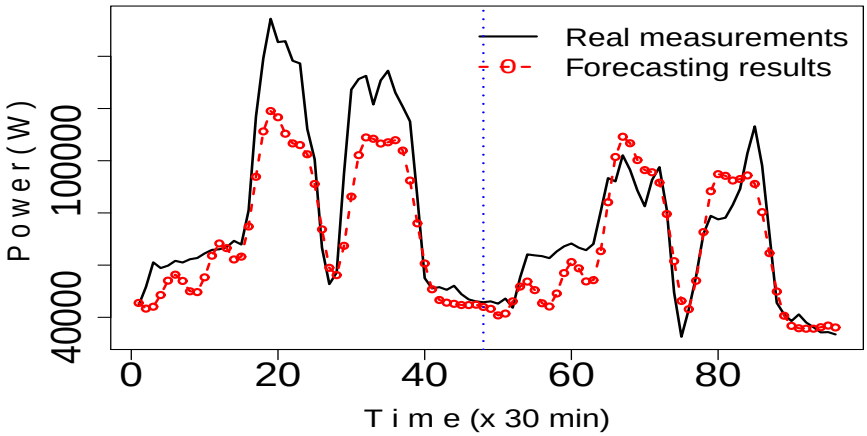


Figure 3.7: *Substation VI_LOG*, two-day-ahead load forecasting results on weekdays (Tuesday and Wednesday: Jan. 12, 2010- Jan. 13, 2010). The vertical dotted line separates two days.

illustration example *substation VI_LOG* is more strongly influenced by the weekly pattern than the temperature variation. Thus, the loads of the same day of a week before are used to replace the forecasts in the naive model. In this case, as shown in table 3.1, the one-day-ahead and two-day-ahead forecasting performances of the naive model are the same. Table 3.1 demonstrates that the presented method based on time series improves load forecast on a MV/LV substation level. Similar results have been found with the other mixed substation examples with an improvement of 2−3% compared with the naive model (Complete results in Appendix A). Considering the 738 000 MV/LV substations in France [97], a marginal improvement on one MV/LV substation in precision can have a significant gain in both efficiency and economy of the whole electrical system. The method is also successfully

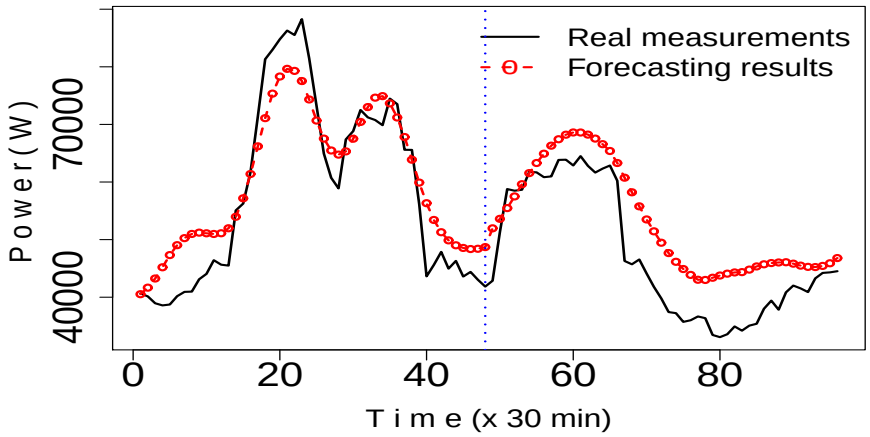


Figure 3.8: *Substation VI_ LOG*, two-day-ahead load forecasting results on weekends (Saturday and Sunday: Jan. 30, 2010- Jan. 31, 2010). The vertical dotted line separates two days.

applied to a more aggregated level, specifically to four Medium Voltage (MV) feeders' load forecast. For one-day-ahead forecast, the results have an improvement of $2-5\%$ compared with the naive model (Complete restults in appendix A). Moreover, bigger improvement of $3-11\%$ is found on the two-day-ahead forecasting scenarios (table 3.2, complete restults in appendix A). Representing a mix of all types of clients, the load curve of a MV feeder is influenced by both day type and temperature variation. In other words, neither last day's load ("D-1") nor the same day of the last week's load ("D-7") can properly replace the forecasting results. It is the reason that with the forecasting period becoming longer, the naive model deteriorates more or less rapidly on the feeders depending on the clients' composition. The designed time series model on the other hand, taking these influencing factors into account, is more robust compared with the naive model for a longer period load forecast.

Table 3.1: Forecasting results: comparison between the naive model and the complete time series model on the test set period (297 days) of *Substation VI_ LOG*.

		Naive model	**Time series model**
One-day-ahead	**MAPE (%)**	18.9	16.4
	MAE (kW)	8.82	7.13
Two-day-ahead	**MAPE (%)**	18.9	16.6
	MAE (kW)	8.82	7.23

Table 3.2: Forecasting results: comparison between the naive model and the complete time series model during the period: Mar. 9, 2010~ Sept. 21, 2010 (197 days) of *MV feeder CL*

		Naive model	**Time series model**
One-day-ahead	**MAPE (%)**	16.7	12.2
	MAE (kW)	166.48	117.60
Two-day-ahead	**MAPE (%)**	25.8	14.2
	MAE (kW)	248.50	135.51

3.3.c Residual Analysis

Residual contains the variation that has not been explained by the fitted model. It is mainly due to the random irregularities of the sampled observations.

In this section, we examine the residual in two aspects:

- Normality is examined by the Probability Density Function (PDF) and the Cumulated Distribution Function (CDF) plot. A normal distributed error guarantees the goodness fit of the model.
- Independence is examined by the ACF. An independent residual argues that there remains no useful information in the historical data for the forecasting model.

3.3.c-i Normality

We usually expect the residual to follow a normal distribution. When the errors are normally distributed, OLS regression is known to be optimal for coefficient estimations [93]. Moreover, if the residual does not follow a normal distribution, wrong intervals could be set up; for example, the prediction interval.

The PDF of the residual $p(\epsilon)$ at a specific value is the probability that the residual ϵ will have that value.

The CDF $F_\epsilon(x)$ represents the probability that the residual variable takes a value less than or equal to x, such that:

$$F_\epsilon(x) = \int_{-\infty}^{x} p(\epsilon)\,\mathrm{d}\epsilon \tag{3.13}$$

The PDF plot as well as the CDF plot is used to compare the residual distribution with other known distribution functions. In our case, we compared the residual distribution with the standard normal distribution. These plots are built by plotting the specific values of the residuals versus the correspondent theoretical values from the standard normal distribution. This standard normal distribution is built by a large number of random samples with same mean and standard deviation as the residual. If the residual is normally distributed, its curves must be close to the normal distribution curve. Instead, if the residual density function has some significant deviations from the standard normal distribution curve, it suggests that the residual series are probably not normally distributed.

Both, the probability density function plot and the empirical cumulative distribution function plot of the residual (figure 3.9) indicate that it is reasonable to believe the residual of the regression process to be approximately a normal distribution.

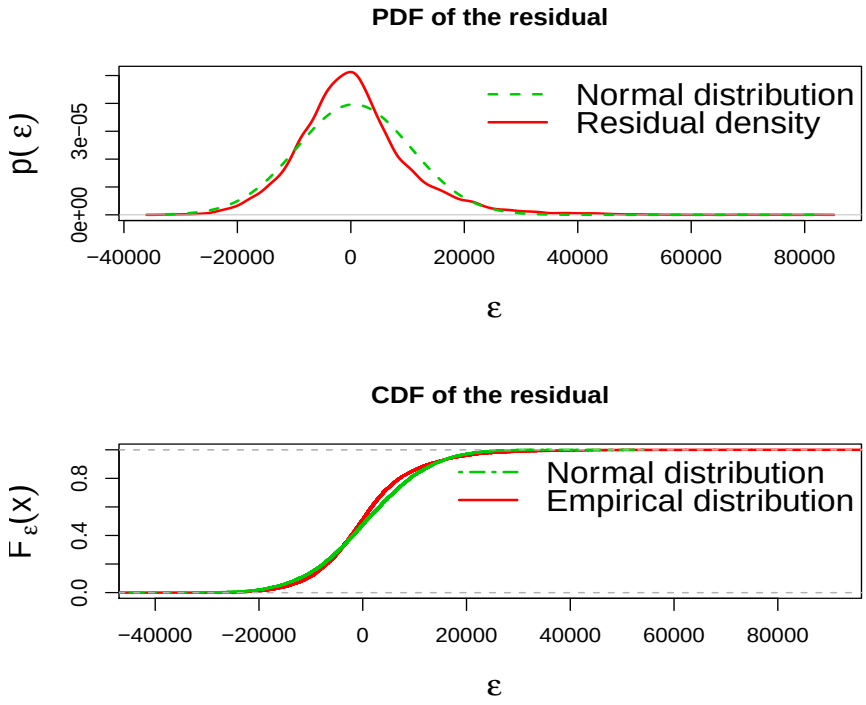


Figure 3.9: *Substation VI_LOG*, density function plot and cumulative density function plot of the residual.

3.3.c-ii Independence

The independence of the data set is measured by the AutoCorrelation Function (ACF) [88]. ACF of the residual measures the linear predictability of ϵ_t, the value of the residual at time t, by using the previous residual values. The computed ACF takes values between −1 and 1. The closer the value is to one of the two limits (−1 and 1), the greater linear predictability we can get. If the residual set is random, its autocorrelation should be near zero for all time-lags. If there are significant non-zero values, the residual set probably has no randomness.

Our plots of residual autocorrelation function in figure 3.10 show the independence evolution through the designed process. At the top, the ACF plot of the original data indicates that the raw data should be differenced at least once. In the middle, the ACF plot of the detrended data shows the strong evidence of periodicities. At the bottom, the ACF plot of residual is much closer to an independent data compared with the former ones. It indicates that the residual of the complete model is not random, and there still exists some moderate degrees of autocorrelation between residual at time t and the residuals at time $(t-1),(t-2),\cdots,(t-18)$. However, since the load consumption data is available once per day, when forecast at time t is computed, the real values at time $(t-1),(t-2),\cdots,(t-48)$

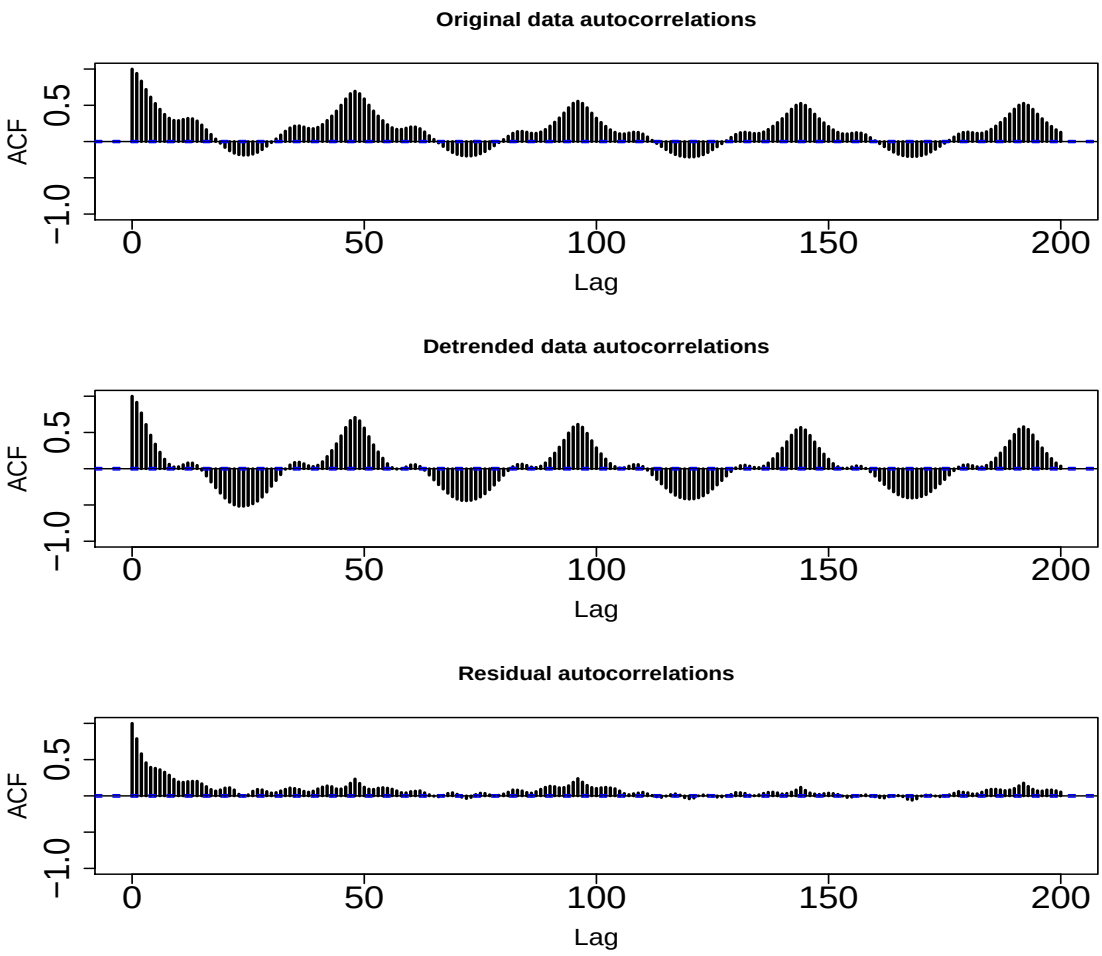


Figure 3.10: *Substation VI_ LOG*, evolution of autocorrelation functions of each step. Regarding equation 2.6, lag is the variable τ.

are unknown. As a result, it is difficult to obtain any improvement from these remaining correlations.

3.4 Weather uncertainty

Since the forecasting temperature data are not in our possession, the above results are calculated based on the actual realized temperature data. W. Charytoniuk et al. [26] have clearly pointed out that their results are also obtained by using the actual temperature values instead of forecasting ones. At the same time, they explained that using temperature forecasts would decrease the forecasting precision. Henrique Steinherz Hippert et al. [28] have explained in their short-term load forecasting review that using observed weather values instead of forecasting values which we don't usually have for the research use, is a common practice. They also concluded that in the real industrial use, when using forecasting weather values, the forecasting result is less accurate. V. Dordonnat et al. [27] have made experiments both with the actual and one-day-ahead forecasting temperature data. They found that temperature uncertainty influences results in respect to the methods and

to the seasons. Some methods are more resistant to the temperature errors than others during certain periods of the year. In the actual French electricity system, temperature forecasting errors have smaller impacts on summer than other seasons due to the rarely air-conditioners' use [27].

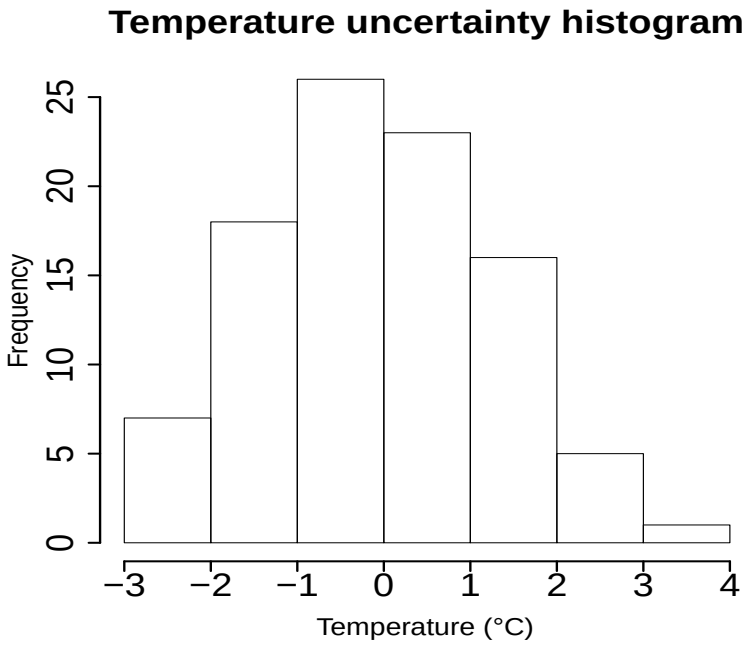


Figure 3.11: Histogram of the Gaussian distributed temperature uncertainty adding to the actual temperature. Frequency: occurrence during two days (96 points in total)

In this section, we aim at analyzing the temperature errors' impact on the proposed time series method. According to the French weather forecast bureau, the one-day-ahead temperature forecasts have a precision of 1.5^oC on an hourly basis [98]. Therefore, in order to create the uncertainty effect, a Gaussian noise zero centered with standard deviation of 1.5^oC is added to the actual temperature values. The same procedure of the model design (figure 3.1) has been carried out. For the coherence, the example borrowed here is still based on *substation VI_ LOG*. Figure 3.11 illustrates the histogram of the one day's (48 points) temperature uncertainty. Figure 3.12 demonstrates the forecasting temperatures and actual temperatures on three successive days. Table 3.3 shows the overall performance comparison among time series model with forecasting temperatures, time series model with actual temperatures and the naive model.

The results in table 3.3 shows that with the "forecasting" temperature, the time series model still largely outperforms the naive model. The obtained results also confirm the conclusion drawn by other researchers [26, 28, 27] that the precision suffers from the imprecision in the weather data.

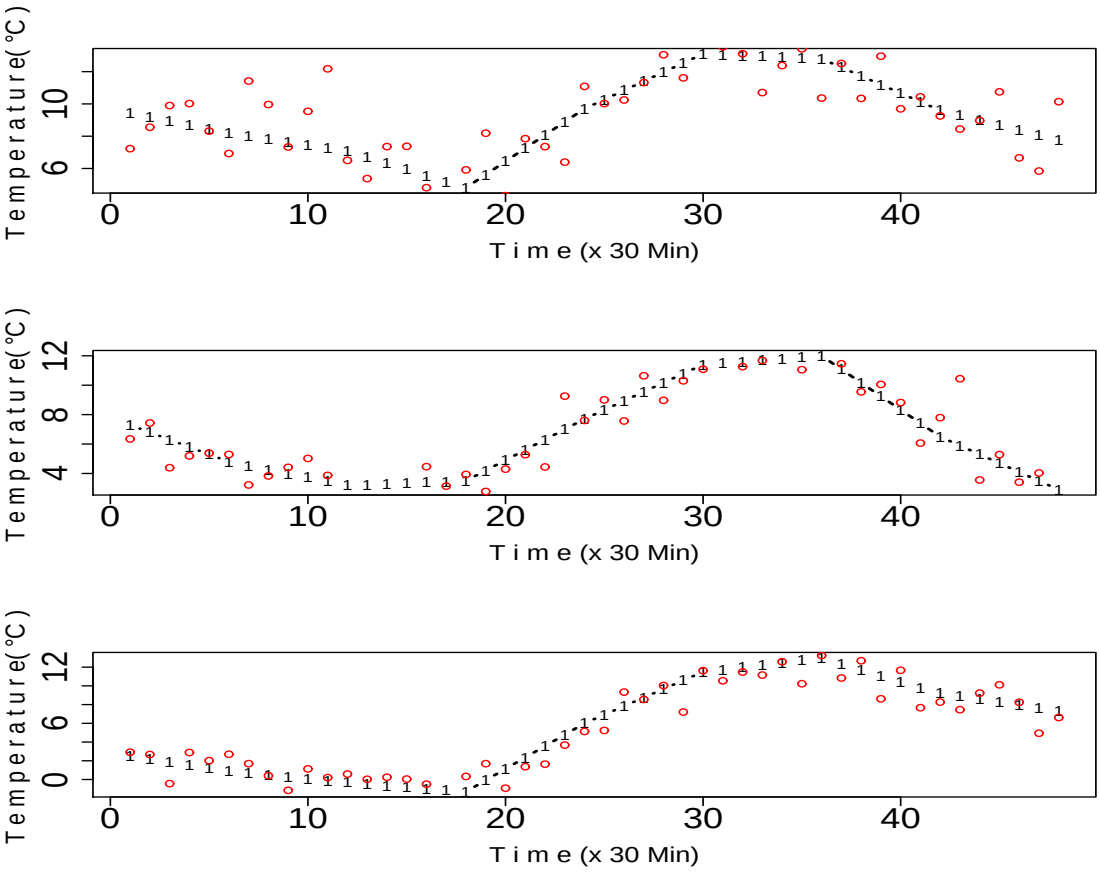


Figure 3.12: Three-day forecasting temperatures compared to the actual temperatures. The "1" represents the real temperature, while the red circle represents the forecasting temperature.

Table 3.3: Performance comparison among Time Series (TS) models with forecasting temperature, actual temperature, and naive model.

	MAPE (%)	**MAE (kW)**
TS model with forecasting temperature	16.9	7.32
TS model with actual temperature	16.4	7.13
Naive model	18.9	8.82

3.5 Conclusion

This chapter proposes a short-term forecasting model based on time series method for the MV/LV substations. The time series model divides the load curve into 3 separate parts: trend, cyclic and random errors. The first two deterministic parts are built with dummy variable regression models. The sliding window and ANOVA nullity test are combined with the regression models to give a better accuracy to the coefficient estimation. The model can be used to all types of mixed substations as it takes into account both the day type

and temperature information. The time series model is also successfully applied to the MV feeders' load forecasts.

Residual is examined by its independence in order to ensure the model with a reliable behavior. The results show that the residual is greatly uncorrelated compared with the original data despite of the small remained correlation. The residual analysis shows that there is hardly any possible amelioration on the forecasting results with statistical tools.

The weather uncertainty is also examined in this chapter. A Gaussian noise is attached to the actual measured temperature with an official published standard deviation to simulate the forecasting weather data. Experiments show that even with the forecasting weather data, our time series method still largely outperforms the naive model.

The proposed time series model has taken into account numerous explanatory variables, such as day type, temperature and periods. It is simple to implement, and easy to interpret. It shows great efficiency in computation that with little computational effort, the obtained forecasting results are satisfactory. Supplying electricity to a single industrial client, *substation CE_FRO* (figure 2.11) is independent to the temperature variation. Moreover, except the week periodicity, which has been taken into account by the naive model, its periodicities are not evident, relying on the productivity plan of the client. Therefore, the time series model, which exploit these information, cannot deal with this single industrial client's case. In the next chapter, we present the model based on neural networks in the artificial intelligent family. It is a promising candidate to answer the load forecasting request, since it gives better results than classical approach according to literatures [99] and it is capable of dealing with all types of load forecasts. Our work concentrates on the neural network model's structure selection strategies.

Chapter 4

Neural network model

Contents

Abstract

The chapter describes the design of a class of machine learning models, namely neural networks, for the load forecasts of electrical substations. Real measurements collected in French distribution systems are used to validate our study. We focus on the methodology of model design, in order to obtain a model that has the best achievable predictive ability given the available data. Principled methods are applied for variable selection and model selection. The results show that the neural network based models are more accurate than the time series models for load forecast, based on the same data. Comparison between the neural network and the time series models in various aspects is made in the end of the chapter.

4.1 Machine learning technique

We address the load forecasting problem with neural networks, a popular class of machine learning methods. One of the purposes of machine learning is the design of predictive models from data whenever prior knowledge on the process to be modeled is unavailable, or not accurate enough. Such situations being very frequent, machine learning pervaded essentially all scientific and technical fields during the past few years. In particular, neural networks have become very popular as nonlinear regression methods, for reasons that will be developed in section 4.2. Naturally enough, they have been frequently advocated as load forecasting tools [28, 99].

The simplest machine learning models for load forecast are AR models, whereby the future load is modeled as a parameterized function of selected past values, either linearly as a linear combination of its past values (usually termed AR model), or nonlinearly. More elaborate models make use, in addition, of past values of exogenous variables such as ARX and NARX models (cf. subsection 2.1.b-i). In the linear case, the estimation of the parameters of the linear models is performed by least squares fitting or variants thereof. In the nonlinear case, a family of parameterized nonlinear functions must be postulated, and the estimation of the parameters requires more elaborate optimization procedures than standard least squares fitting.

In the present work, we postulate that the family of functions known as neural networks can be used appropriately for power consumption prediction on the basis of both past consumptions and exogenous variables. In other words, we expect neural networks to provide nonlinear functions that can map past consumption values and exogenous variables at a given time to a future value of the consumption. The mapping is expected not only to explain the data from which the model is designed (termed training set), but also to generalize to "fresh" data (termed test set), i.e., to data that have not been used for designing the model. Therefore, in order to build an efficient model by machine learning, one must address two problems:

- Find the endogenous and exogenous variables that are relevant for predicting the power consumption. This is known as the *variable selection* problem.
- Find the appropriate model complexity given the available data. This is known as the *model selection* problem.

Note that the variable selection problem is not specific to nonlinear models, and must be addressed similarly for traditional linear models. By contrast, the model selection problem is ubiquitous for nonlinear models, be they simple (such as polynomial models), or more elaborate (such as neural networks [21], SVM [100], hidden markov models [101]). The following universal property is observed experimentally and explained theoretically: a model that is not complex enough is unable to explain the available data and to generalize to previously unseen data; conversely, a model that is too complex fits the available data very well, but is unable to generalize because it is "specialized" on the training data and fails to capture the deterministic behavior of the process. This is known as the bias-variance dilemma. Complexity selection aims at solving the dilemma, i.e., finding the optimal model complexity, given the available data, in order to build an optimal predictive model. The complexity of a model is defined rigorously by its Vapnik-Cervonenkis dimension [68]; for

practical purposes, it can be roughly described by the number of parameters of the model: the more parsimonious the model (i.e., less number of parameters of the model), the lower its complexity. The ratio of the complexity to the number of training examples is also crucial: it must be kept as low as possible.

According to the literatures and the number of devoted researches, the NN is the most popular model in the load forecasting applications [39]. Other methods such as SVM and hidden markov models could also accommodate to the load forecasting issue. However, as a first approach, they go beyond the scope of this dissertation. In the present work, we address the problems of variable selection and model selection by statistically principled methods that proved successful in the past for many applications in a wide variety of fields [21]. Variable selection is performed by the random probe method, which will be outlined in section 4.2. Model selection is based on the Virtual Leave-One-Out (VLOO) score, an empirical estimator of the generalization error and the state of the neural network's Jacobian matrix. The VLOO score requires the computation of the leverages of the training data. The leverage reflects the influence of each training example on the model. The Jacobian matrix examines the effectiveness of the complexity of models. These methods will be described in section 4.3.

In addition to making use of a full model design methodology, the main contributions of our work are the following :

- The load curve is decomposed into the daily average power and the intraday power variation parts, each part being separately forecast by a dedicated model.
- The procedure is validated with real data collected from the French distribution network. Results are compared to the time series models [102] on the same data sets.

The rest of the chapter is organized as follows: section 4.2 presents briefly neural networks. Section 4.3 details the statistical tools of the design procedure. Section 4.4 presents the applications of the designed models on the prediction of the load data of French MV/LV substations. Section 4.5 compares the time series and the neural network models in various aspects. Research perspectives in the application of neural networks are proposed. Section 4.6 concludes the chapter.

Note that this work is in close collaboration with G. Dreyfus, author of the book [21] where the principled methods of neural network are thoroughly presented. Thanks are also due to doctor engineer G. Foggia and research engineer C. Benoit for their great contributions respectively in the development of theoretical and application results.

4.2 Multi Layer Perceptrons and training process

The purpose of the present study is to perform a 24 hour-ahead load prediction in a black-box fashion, i.e., without incorporating prior knowledge into the model. Therefore, one must postulate a family of parameterized functions that have sufficient flexibility for implementing the unknown mapping between variables and the load to be predicted. One-hidden layer Multi Layer Perceptron (MLP)s in the neural network family are nonlinear functions that were proved to be universal approximators, i.e., to be able to approximate a sufficiently regular function to an arbitrary degree of accuracy [103]; in addition, they were

proved to be parsimonious. Therefore, MLPs are good candidates for solving our problem. In this section, we first recall the definition of these functions.

A one-hidden layer MLP with M "hidden neurons" is a linear combination of M parameterized nonlinear functions called neurons. A neuron is a nonlinear, bounded function of a linear combination of its variables, usually an s-shaped (sigmoid) function such as the hyperbolic tangent (tanh). Therefore, the i-th "hidden neuron" of an MLP can be conveniently written as:

$$c_i(P,\Omega_i) = \tanh(\sum_{j=0}^{R} \omega_{i,j} p_j) = \tanh(\Omega_i^T P) \tag{4.1}$$

where P denotes the $(R+1)$ vector of the variables $\{p_j, j = 0,\cdots,R\}$ of the model, and Ω_i denotes the vector of the parameters (or weights) $\{\omega_{i,j}, j = 0,\cdots,R\}$ of hidden neuron $i, i = 1,\cdots,M$. Note that variable p_0 is a constant equal to 1, termed "bias" of the hidden neurons.

Thus, the MLP function can be written as:

$$f(P,\Omega) = \sum_{i=1}^{M} \omega_i c_i(P,\Omega_i) = \omega^T C \tag{4.2}$$

where Ω denotes the set of the $(R+1)M + (M+1)$ parameters (cf. figure 2.3) of the model, ω denotes the $(M+1)$ vector of weights of the linear combination, and C denotes the vector of the outputs of hidden neurons $\{c_i(P,\Omega_i), i = 1,\cdots,M\}$ with an additional component $c_0 = 1$, named the output bias.

Note that $f(P,\Omega)$ is a linear function of the parameters of vector ω and a nonlinear function $c_i(\cdot)$ of the parameters of vectors $\{\Omega_i, i = 1,\cdots,M\}$. The nonlinearity with respect to some of the parameters is the origin of the parsimony of the models, the price to be paid being an increased complexity of training algorithms.

In the present work, $f(P,\Omega)$ is the 24-hour ahead load prediction, P is the set of endogenous and exogenous variables, selected as described in subsection 4.3.a, and Ω is the set of parameters of the model, which are estimated by training the neural network from a set of past measurements, the training set. No attempt is made at designing a model that provides simultaneously predictions of all 24 hours of the next day, i.e., a neural model with 24 outputs [104]. It is a very difficult task that cannot conceivably be performed by a parsimonious model, leading to an overparameterized model with dubious generalization capability.

The training set consists of N pairs $\{P_i, y_i\}$, where P_i denotes the set of variables related to example i, and y_i is the corresponding load measurement. The training process consists of estimating the parameters of the model Ω by minimizing the distance between the output of the neural networks $f(P_i,\Omega)$ and load values y_i (figure 2.4). For nonlinear regression problems such as investigated here, the usual distance is the least squares cost function:

$$J(\Omega) = \sum_{i=1}^{N} (y_i - f(P_i,\Omega))^2 \tag{4.3}$$

Since the MLP function is nonlinear with respect to its parameters, the cost function is not quadratic with respect to the parameters of the model, so that the usual least squares

fitting algorithm cannot be applied: one has to resort to nonlinear, iterative optimization algorithms, based on the computation of the gradient of the cost function. At each iteration, the gradient of the cost function is computed, and a parameter-update procedure that makes use of the value of the gradient of the cost function is applied, until convergence to a minimum of the cost function. This procedure can be accomplished efficiently by the so-called "backpropagation" algorithm. In the present study, the gradient of the cost function was computed by backpropagation, and the parameter update procedure was the Levenberg-Marquardt algorithm, a well-documented [105] second-order nonlinear optimization algorithm [1].

4.3 Model design

A model design procedure aims at providing the model that has the best generalization capability, given the available data. As outlined in section 4.1, this requires a search for the optimal model complexity. Since the number of parameters is linear with respect to the number of model variables and to the number of hidden neurons, the design procedure requires discarding all candidate variables that are not relevant for the prediction task at hand, and finding the appropriate number of hidden neurons.

Several related works exist: V.H. Ferreira et al. developed method based on Bayesian approach that has solved the problems of Neural Network (NN) complexity and variable selection in a coupled way [99]. N. Amjady et al. proposed a feature selection technique based on mutual information and a cascaded neuro-evolutionary algorithm (CNEA) [104]. The number of variables and the number of hidden neurons of the 24 NNs are determined by an iterative search procedure. I. Drezga et al. set up the pilot simulation, where the nearest neighbor target days are used instead of the original target days, to determine the number of hidden neurons [75]. In their works, the input variable selection for NN is based on the phase-space embedding method (Integral Local Deformation (ILD)) and the final forecast is the average of the forecasts provided by two identically structured networks [106].

In this section, we describe the statistical tools used for variable selection and model selection in the present study.

4.3.a Variable selection

There are two categories of variable selection methods: *filters* and *wrappers*. *Filters* aim at finding relevant variables irrespective of the model. By contrast, *wrappers* take into account the model, so that variable selection is "wrapped around" the model training procedure. In this manner, variable selection and model selection are performed simultaneously. An in-depth discussion of variable selection procedures can be found in [107]. *Wrappers* tend to be more computationally costly than *filters*, but they are of more general applicability. In the present study, as a first approach, a filter was implemented.

The principle of the method is the following [108] : in order to discriminate between irrelevant and relevant variables, a set of realizations of a random, hence irrelevant, vari-

[1] Unfortunately, the term backpropagation is frequently used for the computation of the gradient of the cost function followed by simple gradient descent, a training method that is less efficient.

able, called "probe variable", are generated by randomly shuffling the components of the vectors of candidate variables. The probe variables, together with the candidate variables, are ranked in order of decreasing relevance by orthogonal forward regression [109]. The cumulative probability distribution of the rank of the probe variable is estimated, and the threshold rank r_0 is determined such that the probability for an irrelevant variable to be selected is smaller than a threshold δ chosen by the model designer:

$$p(r_{probe} < r_0) < \delta \tag{4.4}$$

where r_{probe} is the rank of a realization of the probe variable. Therefore, δ is the risk of selecting a variable although it is irrelevant. Since the ratio of the complexity to the number of examples is a crucial figure, a low value of δ must be chosen if data are sparse in order to maintain a low complexity, while one may afford to choose a higher value of δ if data are abundant.

We denote by ξ_i the vector whose components are the N measured values of the i-th candidate variable ($i = 1, \cdots, n$), and by y the vector whose components are the N measured values of the quantity to be predicted. If the variables have zero mean, the square of the correlation coefficient between the i-th candidate variable and the quantity of interest is given by [21]

$$\cos^2 \varphi_i = \frac{(\xi_i^T y)^2}{(\xi_i^T \xi_i)(y^T y)} \tag{4.5}$$

where φ_i is the angle between vectors ξ_i and y in the observation space. The smaller φ_i, the larger the correlation between the candidate variable and the quantity to be predicted.

Ranking is performed by orthogonal forward regression, based on Gram-Schmidt orthogonalization [110, 21]: as illustrated in figure 4.1, the candidate variable whose vector ξ has the smallest angle with vector y is first selected; the vectors of all other candidate variables, and the output vector y are projected onto the null space [2] of the variable ranked first, thereby eliminating the contribution of the first variable. The same computation is then carried out in that space, providing the second variable in the ranking. The process is iterated until all variables are ranked, or until another stopping criterion is satisfied as described below. In order to take nonlinearities into account, the candidate variables are the primary variables (i.e., the variables that are considered by the process experts to be probably relevant) and their pairwise cross-products.

The next step is the definition of the threshold rank such that all variables that are ranked below the threshold are discarded. To that end, n_p vectors of random variable ("realizations of the probe variable") are generated by randomly shuffling the components of each vector of candidate variables, so that they have the same probability distributions as the original candidate variables. Those n_p realizations of the probe variables are appended to the set of candidate variables and ranked with the latter. The cumulative distribution function of the rank of the probe is estimated as follows: the estimated probability that the probe rank is smaller than or equal to a given rank r is the ratio n_{rp}/n_p, where n_{rp} is the number of realizations of the random probe whose rank is smaller than or equal to r. During the ranking process, when a rank r is reached such that $n_{rp}/n_p > \delta$ (where δ is

[2] Orthogonal subspace

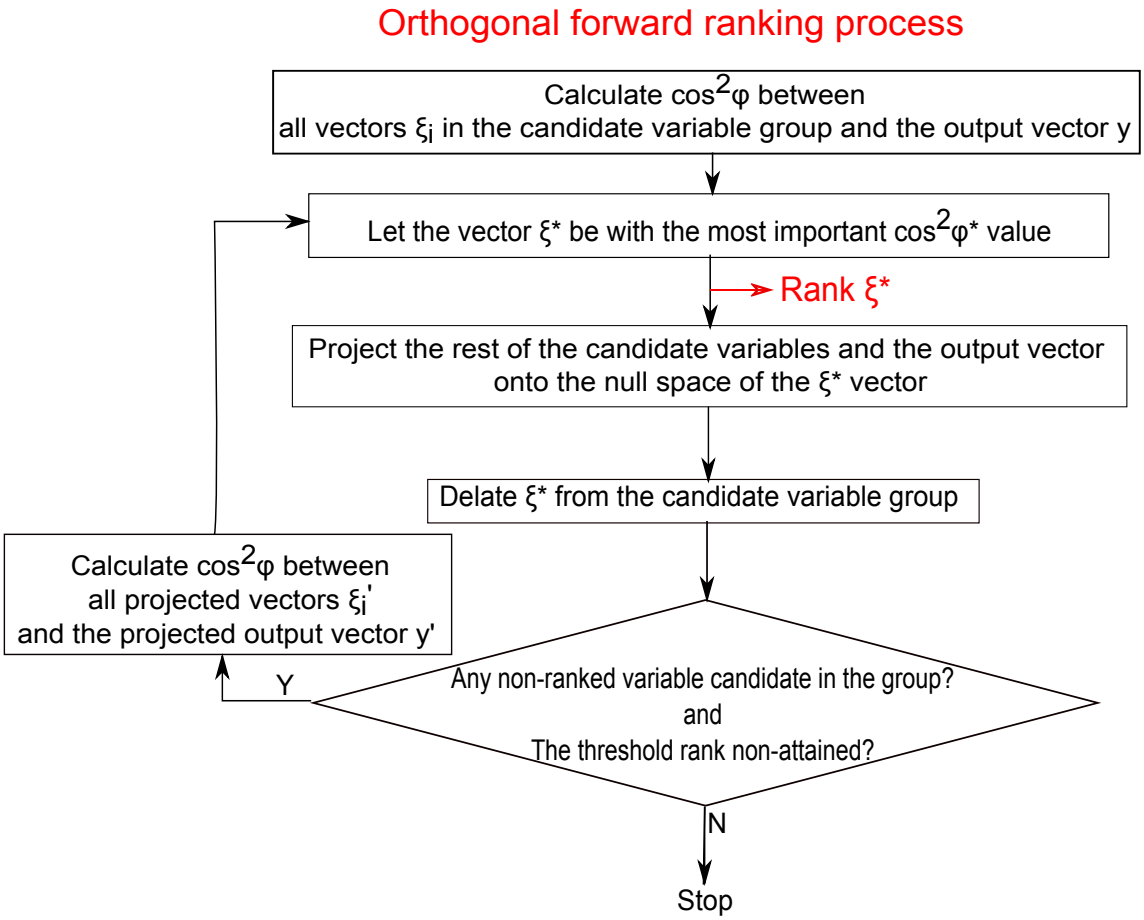


Figure 4.1: Orthogonal forward ranking process

the risk chosen by the designer as described above), the procedure is terminated and the threshold rank r_0 is set equal to $r-1$.

A general presentation of the above method, put into the perspective of alternative variable selection methods, is provided in [111].

4.3.b Model selection

As mentioned in subsection 4.3.a, variable selection by filter methods separates variable selection from model selection. In this section, we describe the model selection method that was used in the present study.

4.3.b-i Model selection methodology

Since the cost function used for training neural networks is not quadratic with respect to the parameters of the model, it has local minima. The optimization algorithms being iterative, initial values of the parameters are required: they are generally drawn randomly from a probability distribution with zero mean and variance $1/R$ [21]. The minimum reached by the optimization algorithm depends on the initial value of the parameters. Therefore, for a given number of hidden neurons, a variety of models can be obtained, each of these corresponding to a minimum of the cost function. Due to the bias-variance

dilemma mentioned in section 4.1, there is no reason why the model corresponding to the global minimum of the cost function should generalize better than a model pertaining to a local minimum. Therefore, a specific methodology must be used, as illustrated graphically in figure 4.2: the number of hidden neurons is increased from zero (linear model) to a maximum value (typically smaller than 10 in real applications); for each number of hidden neurons, many models are trained with different initial parameter values, and the model with the best generalization ability is selected as described below; if, at the end of step i (i.e., when models with i hidden neurons have been trained), it is found that the generalization ability of the best model with i hidden neurons is significantly worse than that of the best model with i-1 hidden neurons, the procedure is stopped, and i-1 hidden neurons is considered to be the optimum complexity given the available data.

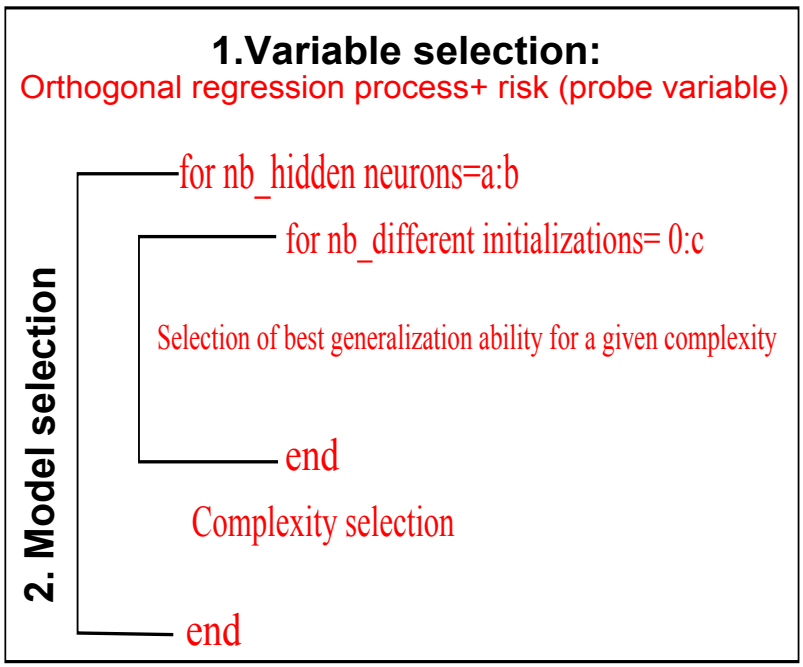


Figure 4.2: Neural network selection procedure({a,b}:{min, max} number of hidden neurons; c: max number of initializations)

4.3.b-ii Assessment of the generalization ability of the models

The methodology described in the previous section relies on the assessment of the generalization ability of the models. The present section describes the assessment methods used in the present study.

Leave-One-Out (LOO) (also termed *Jackknife*) is a procedure that provides an unbiased estimation of the generalization error of a model [68]. It consists of withdrawing an example from the available dataset, training a model from the $N-1$ remaining examples, and computing the prediction error on the left-out example. The procedure is iterated by withdrawing each example in turn from the training set. The leave-one-out score is defined

as :

$$E_{LOO} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - f^{-i}(P_i, \Omega))^2} \tag{4.6}$$

$f^{-i}(P_i, \Omega)$ is the model obtained when example i is withdrawn from the training set. If the number of examples N is large, this is a computationally expensive procedure, since the number of models is equal to the number of examples. However, if the model is linear with respect to its parameters, the exact value of the leave-one-out score is obtained by training a model from all training examples (withdrawing none) and using the Predicted REsidual Sum of Squares (PRESS) statistics [112] defined as:

$$E_p = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\frac{y_i - f(P_i, \Omega)}{1 - h_{ii}})^2} \tag{4.7}$$

where $f(P_i, \Omega)$ is the model trained from all examples and h_{ii} is the i-th diagonal element of the "hat matrix" $H = X(X^T X)^{-1}X^T$, where X is the observation matrix, i.e., the (N, p) matrix whose general element $x_{i,j}$ is the measured value of variable j in example i and p is the number of set of parameters, which equals $(R+1)M + (M+1)$ (cf. figure 2.3). h_{ii} is termed the leverage of example i. If matrix X has full rank, H is the orthogonal projection matrix onto the subspace defined by the columns of matrix X. As a result, the leverages have the following properties: $0 \leq h_{ii} \leq 1$; $\forall i \ \sum_{i=1}^{N} h_{ii} = p$.

Therefore, the leverage of example i can be viewed as the proportion of the parameters of the model that is devoted to fitting the model to example i.

VLOO [113, 114] is a generalization of the PRESS statistic. It differs from the PRESS statistic in two respects:

- E_p is an approximation of the leave-one-out score E_{LOO}
- The leverages are the diagonal elements of $H = Z(Z^T Z)^{-1}Z^T$, where Z is the Jacobian matrix of the model.

Therefore, the model selection procedure is the following (figure 4.2): starting from a linear model (zero hidden neuron), the number of hidden neurons is gradually increased. For each number of hidden neurons, different models are trained with different initial values of the parameters; the VLOO score of each model is computed, and the model with the smallest VLOO score is selected. When, upon addition of a number of hidden neurons, the VLOO score starts increasing significantly, the procedure is stopped and the number of hidden neurons giving the minimum value of VLOO is retained.

It may happen, however, that the VLOO does not vary significantly around the minimum value, in a certain range of number of hidden neurons. If such is the case, an additional selection criterion is taken into account. As mentioned above, the leverage of example i represents the proportion of the parameters of the model that is devoted to fitting the model to example i. Therefore, a model that has one or more examples with leverages close to 1 is certainly very dependent on the specific measurement errors on these examples; thus, it will generalize poorly. On the contrary, a model whose leverages are close to their mean value p/N will be influenced equally by all the examples, hence will have a good generalization capacity. Therefore, as a final selection criterion, the quantity

$\mu = \frac{1}{N}\sum_{i=1}^{N}\sqrt{\frac{N}{p}h_{ii}}$ is computed. μ is always smaller than 1, and it is equal to 1 if and only if all leverages are equal to p/N. Among the models with similar VLOO scores, if any, the model with the highest value of μ is favored.

4.4 Numerical illustration

4.4.a Framework

In this section, we illustrate the design procedure of section 4.3 by applications to power consumption prediction on the MV/LV substations. We compare in the end of the section the performance of the NN based model to the time series method presented in chapter 3. We also explain the limitations of the time series method on the industrial case, which NN based model can properly handle.

The data set used to test and validate our methodology is the measurements collected from the French distribution network. Each measurement represents an MV/LV substation load sampled every 30 minutes during the period from 09 Sept., 2009 to 02 March, 2011 (540 days). This period is longer than the period used in chapter 3 for the time series method study, since new data are collected as time goes by [3]. The period is divided into two parts: a training set and a test set. All reliable statistical predictions are made on the basis of a data set that samples appropriately the space of variables of the model; therefore, the training set must contain at least one whole period in order to guarantee a good performance on the test set. We assume that the consumption scenario has a one-year cycle: therefore, at least one year of data is assigned to the training set. The rest of the data set is the test set, on which the MAE and the MAPE are computed as performance indicators in the same way that for time series in the previous chapter. Since temperature is known to be an influential variable, temperature data and normal temperature data (i.e., average temperature of the same moment during the past 30 years) are also provided on a one-hour basis. The objective is to forecast the loads 24h ahead (sampled every 30 minutes).

Two MV/LV substation load curves are chosen as the study examples. *Substation CE_MOU* is a "mixed" substation with 61% domestic, 23% service sector and 16% industrial customers. The percentage indicates the total subscribed power supply in each customer category. *Substation CE_FRO* is a substation dedicated to an industrial customer. As a result, the electricity consumption of *substation CE_MOU* varies with the temperature while the consumption of *substation CE_FRO* stays stable all along the year [102].

For substation load modeling, we propose to split the power consumption into two parts: the daily average power and the intraday power variation (figure 4.3). These two parts are modeled independently for two reasons:

- We separate the temperature-dependent part (daily average power) from the part mainly influenced by calendar information, thereby reducing the input complexity of the neural networks.

[3] Between the design of the time series model and the design of the neural network model, new data are collected and provided for the study.

- For the intraday variation part, the same number of samples is observed in a smaller scale, thereby providing a better accuracy.

The price to pay is the fact that the design procedure must be applied twice. In addition, for the daily average power part, the amount of training data decreases to 1/48 (2%) of the original data size. This limits the acceptable model complexity.

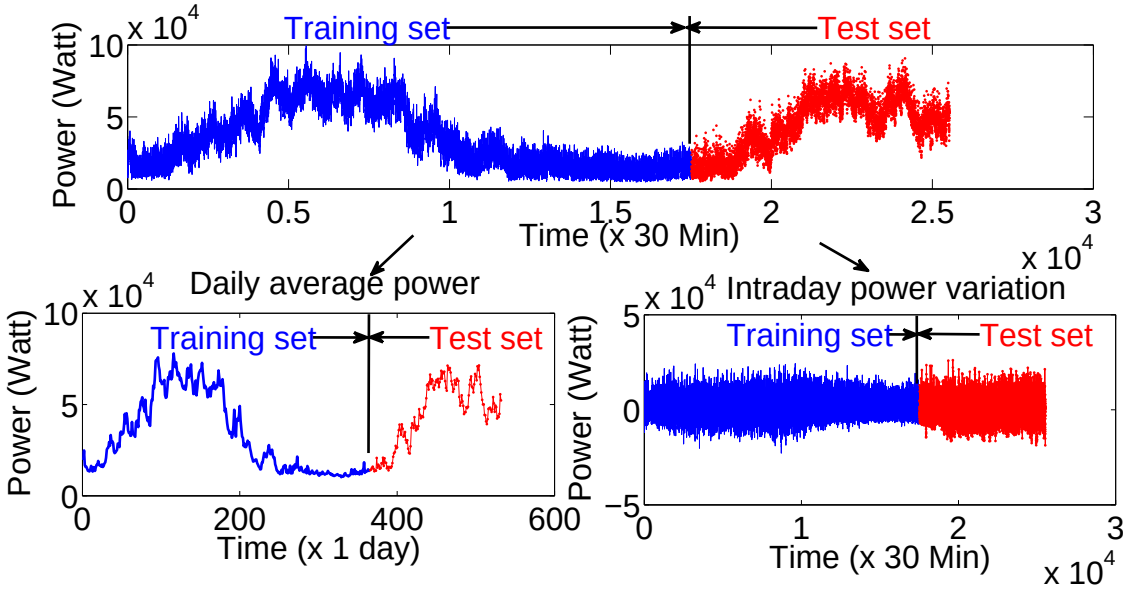


Figure 4.3: Separation of the load curve into the daily average power and the intraday power variation

During training, the cost function is optimized by the Levenberg-Marquardt algorithm, the gradient of the cost function being computed by backpropagation. These algorithms are implemented in the Matlab neural network toolbox.

4.4.b Model design: an illustrative example

In this subsection the various steps of the model design procedure, described in section 4.3 are illustrated on the *substation CE_MOU* example.

4.4.b-i Variable selection example

As described in subsection 4.3.a, the set of candidate variables is the set of the primary variables and of their pairwise cross-products. Let $X_i, (i = 1, \cdots, l)$ be the primary variables: the secondary variables are $\{X_i X_j\}, (i \neq j, i, j = 1, \cdots, l)$. Figure 4.4 illustrates the process of generating the secondary and the probe variables.

In our case, three types of primary candidate variables are proposed: load variables, temperature variables and calendar variables. The first two types are numerical variables, while the calendar variables are categorical, but are transformed to numerical variables.

According to the cyclic behavior of the data described in subsection 2.2.a-iii, we have chosen six load variables as the primary candidate variables: load at time $t - 1day$, $t - 1day - 30Min$, $t - 1week$, maximum load of the previous day, mean load of the previous

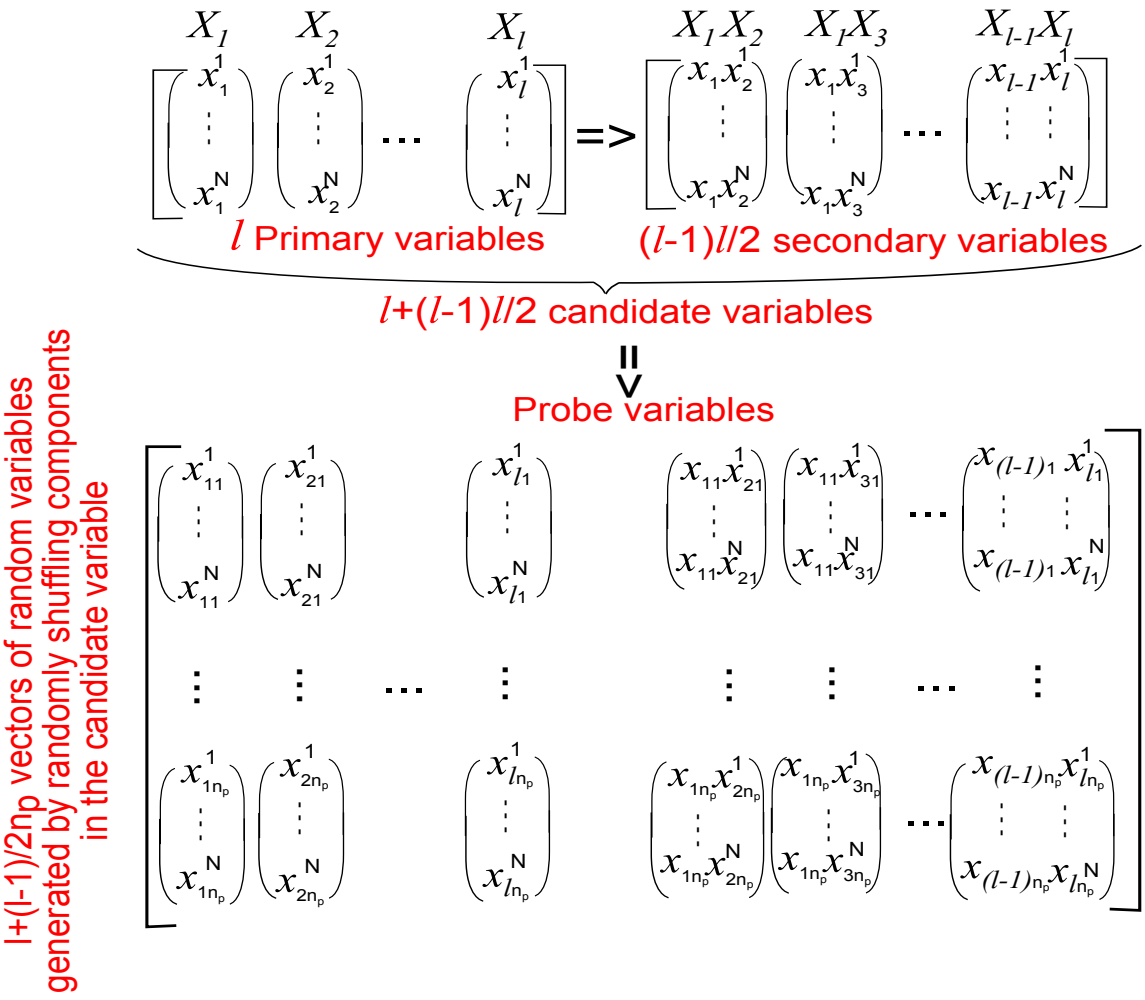


Figure 4.4: Generation of secondary variables and probe variables. Random vector $\{x_{ij}^1, \cdots, x_{ij}^N\}, i = 1, \cdots, l, j = 1, \cdots, n_p$ is generated by randomly shuffling the candidate variable $\{x_i^1, \cdots, x_i^N\}$.

day, and power consumption measured on the previous day at $11:30pm$ (last sample collected before the forecast, assuming that the prediction for the next 24h is carried out at 0am).

Temperatures as well as normal temperatures are included in the set of primary candidate variables. In fact, because of the thermal inertia of the buildings, the response of the load demand to a temperature change may have a time constant that is much larger than the sampling period. Therefore, the average temperature over a sliding window of m hours was considered as an additional primary candidate variable [26]:

$$T_m(t) = \frac{1}{m} \sum_{j=0}^{m-1} T(t-j) \tag{4.8}$$

As stated above, the temperature is sampled every hour; therefore, in order to provide data every 30 minutes, both constant and linear interpolations were used. As a result, fourteen temperature variables are included: temperature at time t, $t-1day$, past $3, 6, 24$-hour average temperatures (T_m, $m = 3, 6, 24$), normal temperature at time t, maximum

and minimum temperatures of the previous day. Sequences of the first six variables are provided both with constant and linear interpolations, thereby yielding fourteen candidate variables.

As usual, the variables are centered and normalized, in order to avoid the problems due to the fact that different variables may have very different magnitudes:

$$x' = \frac{x - \overline{x}}{\sigma_x} \tag{4.9}$$

where $\overline{x}$ and σ_x are the mean and the standard deviation of x.

The calendar variables contain cyclic variables and day type variables. The cyclic variables are designed in order to facilitate capturing cycles [106, 115]. The dominant frequencies of the load data have been determined with spectrum analysis [102]. Half-day, day, week, and year have been identified to be the dominant periods. A pair of variables is represented for every frequency [106]:

$$\begin{aligned} c_1(t) &= \sin(2\pi t/T) \\ c_2(t) &= \cos(2\pi t/T) \end{aligned} \tag{4.10}$$

where t is the time index of the measurement sampled every 30 min: it ranges from 1 to 17 520 for one year; T is the cycle period (17 520 for a year cycle, 336 for a week cycle, 48 for a day cycle, and 24 for a half-day cycle).

Working days and non-working days (weekends and national holidays) are also discriminated and represented by 1 and 0 in the day type variable.

As the daily average power model has one sample per day and the intraday power variation model has forty eight samples per day, their primary candidate variables are different. We suggested different candidate variables for the average power and the intraday power variation NNs: for the daily average power model, all the load variables and the temperature variables are averaged over 24 hours ($\frac{1}{48}\sum_{t=1}^{48} x'(t)$). By contrast, for the intraday power variation model, we have taken the previously presented candidate variables $x'(t)$, and the "daily average removed" candidate variables such that: $x'(t) - \frac{1}{48}\sum_{t=1}^{48} x'(t)$.

For the variable selection, ten realizations of the probe variable are generated from each primary and secondary candidate variable (cf. figure 4.4). In our case, for the average power model, 24 primary variables are designed. In consequences, 276 ($24\times23/2$) secondary varaibles and 3300 ($(24+276)\times10$) realizations of the probe variable are built. For the intraday power variation model, 43 primary variables, 903 ($43\times42/2$) secondary variables and 9460 ($(43+903)\times10$) realizations of the probe variable are generated. Figure 4.5 shows the cumulative probability for a probe variable to have a better rank than a candidate variable for the daily average power model. We observe in the zoom window that from the 10th rank onward, the risk of retaining a variable although it is not relevant is larger than $1/n_p$. More specifically, it signifies that in the orthogonal forward ranking, a probe variable is classed on the 10th position. In this example, 5 primary variables and 4 secondary variables are selected as the result of the ranking process. As explained, only the primary variables are used as the variables in the NN model. Therefore, the set of selected variables is the set of primary variables selected as such, and the set of primary variables that appear in the secondary variables, without duplication. Table 4.1 shows the selected variables with

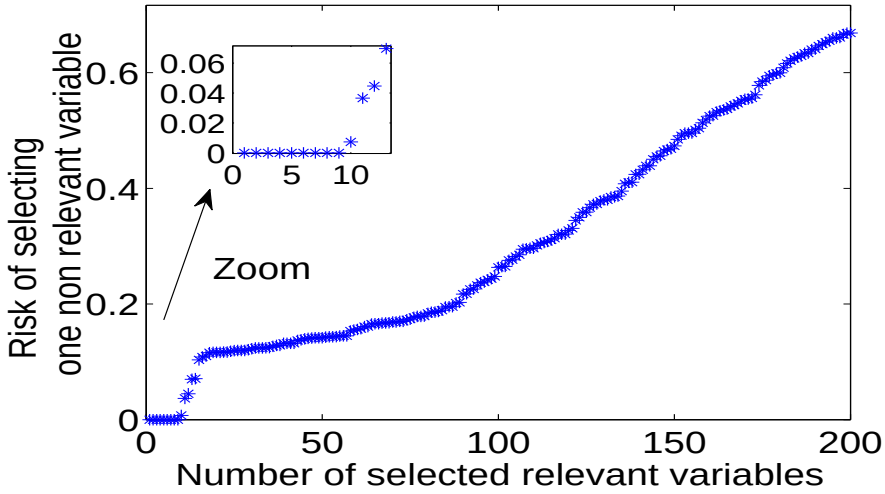


Figure 4.5: Cumulative probability for a probe variable to have a better rank than a candidate variable.

Table 4.1: 9 variables for the average power model. "mean" indicates that the candidate variables are averaged over 24 hours (48 values). "ci" indicates that the temperature variable has been constantly interpolated and "li" indicates the linear interpolation operation. hist_T stands for historical temperature variables.

Category	**Variable**
Load variables	1. P_{mean}(t-1day)
	2. P_{mean}(t-1day-30Min)
	3. P_{mean}(23:30pm)
Temperature variables	4. $\mathrm{hist}_{mean}\_\mathrm{T}^{ci}$(t)
	5. T^{li}_{mean6h}(t)
	6. T_{mean}(max)
	7. T_{mean}(min)
Cycle variables	8. $\sin(2\pi t/7)$
	9. $\cos(2\pi t/365)$

a risk smaller than $\rho_{average}$ = 1/3300 for the daily average model. In the later of the chapter, we'll see that because of the random shuffling effect in creating the probe variables, the outcome of the ranking result could be different (i.e., different numbers of variables are selected, taken a same risk). For the intraday power variation model, over 30 variables are selected with a risk smaller than $\rho_{intraday}$ = 1/9460. Here, for clarity, we only show the top nineteen variables in the ranked list in table 4.2. In subsection 4.4.c, forecast performances with different risks in both models are shown and discussed.

Table 4.2: 19 variables for intraday power variation model. $P_{variation}$ and $T_{variation}$ are daily average removed load and temperature variables.

Category	**Variable**
Load variables	1. P(t-1day) 2. P(t-1day-30Min) 3. P_{max}(t-1day) 4. P_{mean}(t-1day) 5. $P_{variation}$(t-1day) 6. $P_{variation}$(t-1day-30min) 7. $P_{variation}$(t-1week)
Temperature variables	8. hist_T^{ci}(t) 9. T^{li}_{6h}(t) 10. T^{li}(t-1day) 11. hist_$T^{li}_{variation}$(t) 12. hist_$T^{ci}_{variation}$(t) 13. $T^{li}_{3h\_variation}$(t) 14. $T^{ci}_{3h\ \ variation}$(t)
Cycle variables	15. $\cos(2\pi t/336)$ 16. $\sin(2\pi t/48)$ 17. $\sin(2\pi t/24)$ 18. $\cos(2\pi t/24)$ 19. $\sin(2\pi t/17\,520)$

4.4.b-ii Selecting the best model for a given complexity

As explained in subsection 4.3.b-i, model selection requires finding the optimal model for a given complexity. For a given number of hidden neurons, ten models are trained, the models with rank-deficient Jacobian matrix are discarded, and the models with the smallest VLOO score are stored in memory.

Figure 4.6 illustrates the model selection result on the intraday variation power model. The MAE is computed on the test set. For a given number of hidden neurons, all trained models are applied to the test data, and their MAE is computed. Figure 4.6 shows the maximum, minimum and average values of the MAE, together with the MAE obtained by the model selected by model selection. With the proposed model selection strategy, the performance of the selected model follows well that of the minimum error model for all numbers of hidden neurons. Thus, the VLOO score and the rank of the Jacobian matrix are good criteria for assessing the generalization ability of models of a given complexity.

4.4.b-iii Complexity selection example

In subsection 4.4.b-ii, we showed how model selection was performed among models having the same complexity (corresponding to local minima of the cost function). In the present subsection, we show how to select the model that generalizes best, among the models selected in the previous section. As explained in subsection 4.3.b-ii, for models whose VLOO scores are roughly equal, models with the most peaked leverage distribution (values

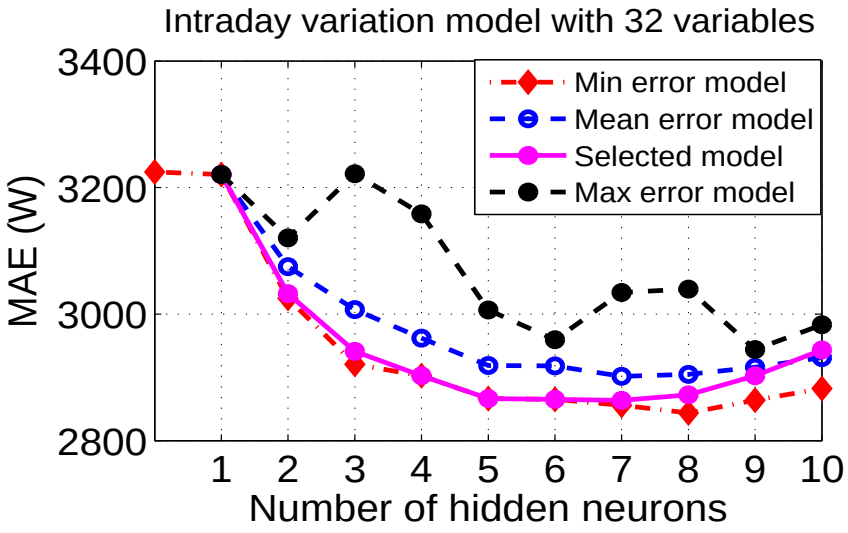


Figure 4.6: Model selection for the intraday power variation model

of μ closest to 1) should be favored. We describe the heuristic that was used to that end. Two different strategies are applied to the daily average power model and to the intraday power variation model.

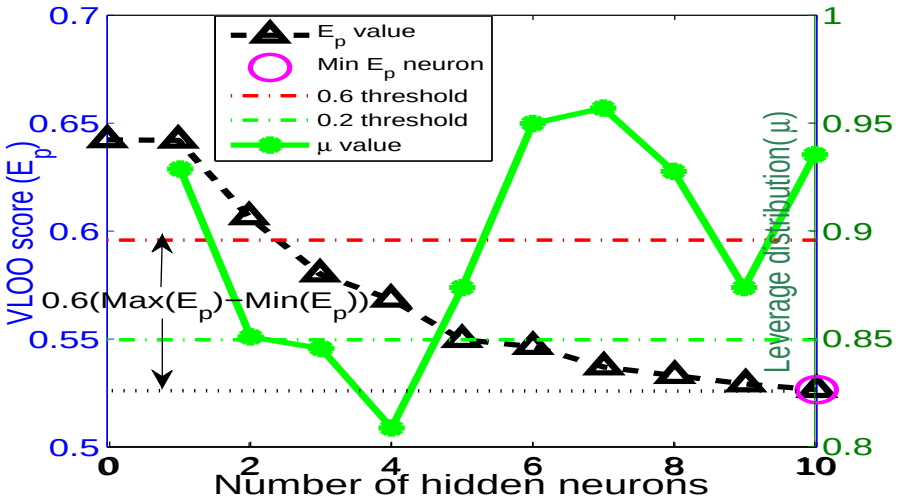


Figure 4.7: Neural network complexity selection strategies with VLOO score and leverage distribution

The daily average power model has a limited number of examples in the training set and very relevant input variables as temperatures, etc. Therefore, low complexity models must be favored. The following heuristic is used. Assume that a model with r hidden neurons has the smallest VLOO score E_{p0} with a leverage distribution of parameter μ_0. The most parsimonious model that satisfies the three conditions:

- Number of hidden neurons $\leq r$
- $E_{p0} \leq E_p \leq E_{p0} + 0.6(Max(E_p) - E_{p0})$

- $\mu \geq \mu_0$, is selected

In the example shown in figure 4.7, a neural network with 10 hidden neurons had the smallest VLOO score; a model with 6 hidden neurons had $\mu > \mu_0$ and its VLOO score was in the prescribed range. Therefore, it was considered that a network with 6 hidden neurons had the appropriate complexity for the prediction of the daily average power, given the available data.

By contrast, intraday power variation model, has a large training set so that one can afford a larger complexity. In this heuristic, the optimal model satisfies three conditions:

- Number of hidden neurons $\leq r$
- $E_{p0} \leq E_p \leq E_{p0} + 0.2(Max(E_p) - E_{p0})$
- The model that has the maximum μ is selected

In the example shown in figure 4.7, the model with 7 hidden neurons had the maximum μ in the prescribed range. Therefore, it was considered that a network with 7 hidden neurons had the appropriate complexity for the prediction of the intraday power variation (figure 4.6), given the available data. That the neural network with 7 hidden neurons gives one of the best performances among all the structures confirms the heuristic choice of the 0.2 threshold.

4.4.c Results

In this subsection, the final results, obtained by adding the prediction results of the daily average power model and the intraday power variation model, are presented and discussed. The MAPE and the MAE are chosen as the performance indicators. The results are compared with the naive model [102] as well as the time series model presented in chapter 3 [102]. The naive model replaces the daily prediction results by the most similar historical day's real consumption data. The similar historical day is often selected as the last day or the same day of the last week. Time series model is a regression model that combines a dummy variable integrated indicating day types, temperature-time dependent trend model and a Fourier component periodic model. The test period is from September 16, 2010 to March 01, 2011 (167 days). Table 4.3 shows the results.

The number of variables is varied both for the daily average power model and for the intraday power variation model. It is unnecessary to use too many hidden neurons for simple structure (low number of variables). As in such conditions, for most of the time, a model with a rank deficient Jacobian matrix is obtained. The selected complexity for every applied model is indicated after colon (table 4.3). In fact, since the daily average power model has a much larger mean value than the intraday power variation model, its performance has a much more important impact on the final result. Therefore, we focused mainly on the daily average power model.

In table 4.3, 5 different cases are presented. For the first case, with a risk smaller than $\rho = 1/n_p$, 6 variables are selected for the daily average power model, and 40 variables for the intraday power variation model. Allowing a 5% and a 10% risks of keeping an irrelevant variable, 10 variables are selected for the daily average power model in both the second and third cases. While with a risk smaller than $\rho_{intraday}$ for the intraday variation model, for

Table 4.3: *Substation CE_ MOU*, forecasting results: comparison among the naive model, time series model and NN models. "Hn": Hidden neuron(s). "NNa": Neural Network for average power, and "NNi": Neural Network for power variation. "Var": Variable(s).

	Naive model	**Time series model**	**NNa: 6 Var, 1 Hn + NNi: 40 Var 6 Hn**	**NNa: 10 Var, 1 Hn + NNi: 37 Var 4 Hn**	**NNa: 10 Var, 1 Hn + NNi: 37 Var 5 Hn**	**NNa: 24 Var, 2 Hn + NNi: 32 Var 7 Hn**	**NNa: 6 Var, 1 Hn + NNi: 19 Var 9 Hn**
MAE (kW)	4.80	4.16	3.62	3.58	3.53	3.68	3.68
MAPE (%)	12.9	11.0	10.2	10.2	10.1	10.3	10.5

both trials, 37 variables are selected. In the fourth case, a 50% risk is taken for the daily average power model and a risk smaller than $\rho_{intraday}$ for the intraday variation model. In the fifth case, we keep a risk smaller than $\rho_{average}$ for the daily average power model but a limited number of variables for the intraday variation model, less than the number allowed with a risk smaller than $\rho_{intraday}$ taking. Two conclusions can be drawn: first, in all cases, our model design methodology yields more accurate predictions than the time series model and the naive model. Second, our proposed neural network model design is a robust methodology. With the various risks taken, the precisions stay steady.

We have explained in chapter 3 that the time series method cannot be applied efficiently to the single industrial client's case. In order to show the great capacity of neural networks in dealing with all types of substations, we also chose the industrial *substation CE_ FRO* as an illustrative example. Table 4.4 shows the result on *substation CE_ FRO*. The number of variables for the daily average power model and the intraday power variation model are both defined by a risk smaller than ρ. For such an industrial substation, the time series method [102] is unable to extract further information other than the day type and the main periods. More information is needed by the time series method to provide a better result than the naive model. By contrast, the neural network models providing a non linear input-output mapping yields 4.7% improvement compared to the naive model. We are planning to adapt this neural network model design approach on reactive power forecasts as well.

Table 4.4: *Substation CE_ FRO*, forecasting results: comparison between the naive model and the neural network model. "Hn" stands for "Hidden neuron(s)".

	Naive model	**14 variables average model(1Hn)+28 variables variation model (10Hn)**
MAE (kW)	99.49	75.07
MAPE (%)	20.2	15.5

4.5 Overall comparison with the time series model

In this section, we aim at comparing the neural network method with the time series method for the short-term load forecast. Six aspects are compared and results are summarized in table 4.5.

Table 4.5: Summary of comparison aspects between neural network models and time series models for the short-term load forecasting application. ☺ indicates the model that has the better attribute.

	Neural network model	**Time series model**
Precision	☺	
Ease of interpretation		☺
Computational complexity		☺
Learning data quantity		☺
Update frequency	☺	
Adaptability	☺	

- **Precision**

  Results on *CE_MOU* and *CE_FRO substations* are detailed in appendix D. In general, the neural network models have a better precision compared to the time series models and to the naive models.

- **Ease of interpretation**

  Both the neural network models and the time series models are divided into two parts. One part representing the slow variations due to the exogenous factors, such as the temperature, often indicates the consumption level. The other part, on the other hand, involving the rapid variations, refines the final results.

  Being a parametric model, the time series method is easy to interpret. The relationship between power and other influence factors can be easily deduced. Following a black-box fashion, the neural network model, on the other hand, is difficult to draw an explicit equation between power and other influence factors [99].

- **Computational complexity**

  Both methods have two periods to output the forecasting results: the learning period and the test period. The learning period for the time series method aims at defining the values of several important variables, such as, the main periods, and the width of the sliding window. While during the test period, with the sliding window strategy, the time series model adapts the values of its parameters iteratively.

  Whereas, the neural network model, during the learning period, calibrates its parameters and, at the same time, selects the optimal model and the variables regarding the learning set. It's a very long process [99]. Independently, for the daily average and the intraday variation power models, the process is performed twice.

- **Learning data quantity**

  For the time series model, the learning data quantity is determined during the learning period. It is often of the size of one week to some weeks. The neural network, requiring an entire period as the learning set, needs to have one year as the learning set to perform correctly.

- **Update frequency**

  Update leads to the changes in the model's structure or parameters. For the time series model, the structure is fixed, but the parameters are re-estimated during each forecasting period, i.e., each day.

  One-year data are set to train the neural networks. Compared to the time series model results, the precision of the neural network forecasting results begins to decrease from May, 8 months after the model training. An appropriate update frequency needs to be chosen for the neural network models. Tables in appendix D show the detailed results of two substations on each month of the forecasting period.

- **Adaptability**

  Having a great adaptability refers to that the model is capable of being easily applied to other similar situations. Time series models, due to their invariant structure, can only extract the simple day type, the temperature, and the principal periods. The model cannot deal with the industrial substations because these substations are independent to the temperature variations and the most important period, the week pattern, has already been exploited by the naive model. Thus, the performance of the time series model can hardly compete with that of the naive model.

  Known for its great learning ability, the neural network model can handle all kinds of load forecasts. It can adapt its structure to the learning set by the variable and the structure selections. Therefore, it behaves better than the naive model in the industrial substation's case (appendix D). It is also a promising solution to the reactive power forecasts.

4.6 Conclusion and perspective

We present a new approach to the forecast of MV/LV substation loads by neural networks. By using separate predictive models for the daily average power and the intraday power variation, and by focusing on the methodology of model design, with principled statistical methods for variable and model selection, we improved the prediction accuracy with respect to those of the naive model and the time series model presented in chapter 3 [102], for an extensive database of actual measurements. In the end of the chapter, a comparison in various aspects is made between the time series and neural network methods.

The future work will focus on the update frequency of the neural network models. An adaptation of the method to the prediction of the reactive power can be envisaged.

In Part A, we have tackled the STLF for operation need in distribution networks respectively with two models based on the time series and the neural networks. Next,

in Part B, we are presenting the second objective of the dissertation, namely the design of load estimation models for the planning need in distribution networks.

Part B

Load estimation models for distribution network planning

Chapter 5

Load research projects in distribution networks: state of the art

Contents

Abstract

Network planning plays an important role in electricity distribution systems in both of economical and technical terms. It impacts the most important decisions in the investment as well as the quality of the electricity supply. The accurate load estimation model is the main key to reduce the network planning uncertainty so as to make the best strategy among various design alternatives. With the development of smart meters, in part B, we aim to build accurate power estimation models in order to improve the efficiency of the network planning. This chapter serves as a brief introduction in the part B. The chapter first talks about the decision making procedure and problems of measurement scarcity. Solutions adopted by most of DSOs, such as integrating TLP recognition and aggregation method using coincidence factor, are introduced. Secondly, the chapter presents the load research projects in four countries: Finland, Denmark, Norway and Taiwan. At the end, French load research project is presented in details. The load estimation model BAGHEERA is carefully scrutinized. For the sake of demonstration, survey client's data are borrowed as the illustrative examples to the method. Advantages and drawbacks of the model are also commented.

The design of an electrical network is a vital but a hard task presented in front of the planners. On the one hand, it defines where the most important investment in the energy sector goes, as well as the quality of the electricity supply to the end-users in the following future years. On the other hand, there are so many uncertain factors in the current distribution system, which makes it a dynamic changing system hard to take under control. These factors mainly include power consumptions, and distributed generations. Therefore, the essential core of the task is to reduce the uncertain factors and provide accurate information so as to make the best strategy among various design alternatives.

With the development of smart grid and the widely opening electricity market, great changes are taking place in the electrical networks. This brings new producers, services, thus introducing new flexible relationships among consumers, regulators, and operators. The smart meter enables collecting accurate "real time" information about individual client's power consumption. These information can draw a detailed energy behavior profile that interests both customers and planners. For customers, the profile can help them to reduce electricity bills by adjusting their behaviors and by choosing the best supply agreement. For planners, they can postpone the expansion of their network by minimizing the margin between the peak demands and the network's capacity and invest money more wisely. The arrival of the smart meter enables building reliable load estimation models, which solve the above problems.

In this chapter, we argue the importance of load estimation models for decision making in distribution network planning and show the state of the art for some models applied in different countries. Section 5.1 firstly presents the dilemma encountered in the decision making: on the one hand, the load information is in need for the technical analysis in order to work out solutions for network reinforcement. On the other hand, till occurrence of the smart meter, that only a limited number of survey clients' consumption data are collected raises difficulties in the design of the load models. A common way to solve the problem is also introduced in section 5.1 by construction TLPs, which represent the homogeneous clients' load patterns and aggregate the TLPs with coincident factors in an upper lever. The coincident factors reflect the fact that not all the individual peak occur at the same time. Load research projects are stated in section 5.2 and specificities of every model are commented. At last, French load research project is thoroughly presented in section 5.3. The BAGHEERA model in consistency with the general solution idea is detailed and shown with illustrative examples.

5.1 Decision making in distribution network planning

The distribution network planning is applied to give a vision of the optimal network in the future so as to help decision makers making investment decisions such as the feeder's and transformer's capacity, size, and location of substations. The decisions are mainly based on the technical and economical analysis. Figure 5.1 describes the three steps of the decision making procedure in the network planning. In the technical analysis, the load estimations, the topology of the existing network as well as the network data (transformers, transmission lines) are taken into consideration. Next, the economic analysis integrating the component's prices, the line losses and the maintenance costs evaluates the cost of each solution proposed by the former step. At last, based on these results and the network

yearly budgets, the plans for investment are made.

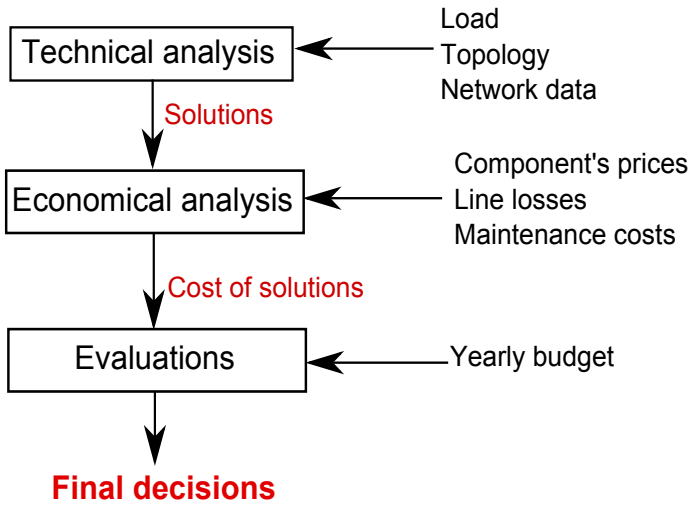


Figure 5.1: Network decision making procedure

As a matter of fact, the technical analysis can be considered as an optimization problem of the losses in the network in maintaining the voltage at each node of the network in an authorized range. In France, the accessible voltage is defined as ±10% of the nominal voltage value. The client, whose connection node is outside this scale, is regarded as *poorly supplied client*. To define the voltage at the connection node, the losses and the voltage calculation need to be carried out. A part from the formerly mentioned information, these calculations require other factors, i.e., the unbalance coefficient, the capability of new producers and consumers, to name a few. In this chapter, we focus on the techniques providing load estimation models to these technical calculations for network planning.

Till the occurrence of the smart meters, the only available measurements in the distribution systems are active and reactive power, voltage and current level of the HV/MV substations as well as several energy consumption readings per year of each customer for billing's use. The consumption of a MV/LV substation is estimated by the proportion of the maximum values between MV/LV substation and its head HV/MV substation. In order to estimate load models on a lower hierarchy for the planning's sake, most of the electricity companies also collect some survey data on a regular basis, i.e., 10 minute, 15 minute, half hourly or hourly, on a limited number of clients.

A common practice applied by most of the electricity companies is to form TLP that represents approximately an individual client's consumption and then aggregate these roughly estimated consumptions to an upper voltage level. For the ease of presentation, the planning method is organized into two steps and presented in an up-down aggregation direction. Firstly, we talk about the *coincidence effect* of clients so as to deduce the MV/LV substation's capacity. Secondly, attention is focused on the various classification methods in finding *TLPs* that represent the homogeneous clients' groups.

5.1.a Coincident load

The coincident load reveals the variety in the clients' load behaviors. It describes the fact that the peak demand of each client does not occur simultaneously. As a matter

of fact, because of their different housing appliances, as well as their different life styles, household load patterns are rarely the same. Moreover, industrial, commercial clients' load curves are different from that of residential clients. The industrial load depends mainly on the production schedule, and their load curve is often flat during a day because of the continuous manufacturing activity. The commercial clients are more influenced by the hour and the peak demand usually appears during business hours. The residential peak demand often happens in the evening time [116]. Therefore, if a substation provides energy supply to different categories and large number of clients, there is a coincident effect that the peak demand of the substation's load is inferior to the sum of all peak demands of clients connected.

In terms of equation, the coincident effect is expressed by a proportion named as "coincident factor" (also known as "reduction factor"). It divides the observed maximum power of the HV/MV substation by the sum of all the estimated maximum power of MV/LV substations connected. The capacity of every connected substation is adjusted by the product of the coincident factor and the estimated maximum power. Thus, in example of figure 5.2, the *coincidence factor*= $\frac{4000}{300+1500+1000+200+2000} = 0.8$. The adjusted estimated value for the peak demand of MV/LVs are: $300 \cdot 0.8 = 240$ kW, $1500 \cdot 0.8 = 1200$ kW, etc.

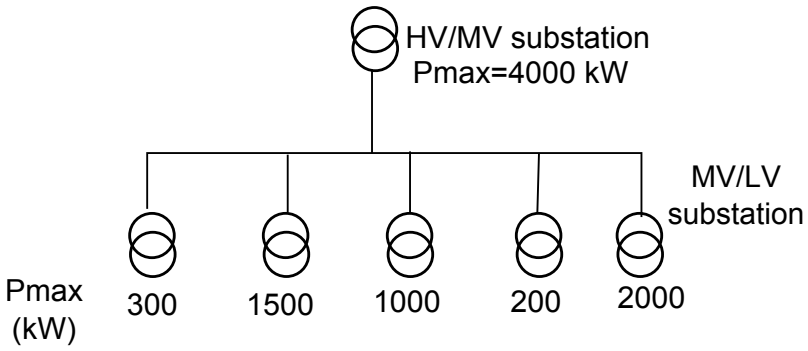


Figure 5.2: Example of coincidence factor calculation

In [9], J. Dickert et al. have summarized several variants of the coincident factors that exist in the literature by taking influence of variety of customers as well as degree of aggregation (number of customers) into account.

The estimated power of the MV/LV substations are calculated by aggregating individual load patterns. If there's no correlation among clients' consumptions, the mean power of the substation $P_{ag}(t)$ equals to the sum of all clients' load curves $P_i(t), i = 1, \cdots, n$ and its standard deviation σ_{ag} equals to the sum square of all clients' standard deviations $\sigma_i, i = 1, \cdots, n$ [117, 11, 118]:

$$\begin{aligned} P_{ag}(t) &= P_1(t) + P_2(t) + \cdots + P_n(t) \\ \sigma_{ag}(t) &= \sqrt{\sigma_1(t)^2 + \sigma_2(t)^2 + \cdots + \sigma_n(t)^2} \end{aligned} \tag{5.1}$$

An excess probability [117] representing the probability of the defined limit being exceeded is integrated into the estimation of maximum power [117, 11, 118]:

$$P_p(t) = P_{ag}(t) + z_p \sigma_{ag}(t) \tag{5.2}$$

where z_p is the standard normal deviate corresponding to the excess probability p.

Otherwise, the mean and standard deviation of substation's load are directly modeled. V. Neimane presented a probabilistic load modeling on the substation level in her Ph.D dissertation [117] which separates the three year's 110 kV and 33 kV substations data into three seasonal modes (winter, spring and autumn, and summer) and two day type modes (working day and weekend), resulting in six modes. The central moments are calculated in order to affect each mode into an appropriate distribution according to Pearson's chart.

5.1.b Typical Load Profile (TLP)

Typical Load Profile (TLP) refers to a daily load curve pattern that represents the load behavior of a group of coherent clients. As a matter of fact, all clients are not metered on a regular time interval basis due to the expenses, and load measurements are only collected on a limited number of clients as survey data. Because of the scarcity of the individual regular sampling load measurements, the only way that each customer can have a representative daily load pattern is to use classification methods. These classification methods assign every non survey client to a TLP group predetermined by the survey data at an early stage. According to the TLP of his group, a client's annual energy consumption can be disaggregated into time interval segments, i.e., hourly, half-hourly, quarter hourly and into day types, i.e., working day, weekend, public holiday. These individualized profiles will then participate to the network calculations.

The classification must follow the practical criteria such as [119]: each customer is affected to one unique class. In condition that the number of the classes is reasonable (not too much), the variance within each class must be as small as possible. Each class should be representative.

[120] explains the process from load survey data to the TLP (also called Class Representative Load Pattern (CRLP) in this article). The survey load data is first normalized with respect to a reference value in order to form the Representative Load Pattern (RLP). This later represents the pattern of the survey client and assumes the maximum power as the reference value. Thus, the RLP ranges in the $[0, 1]$ scale. The effectiveness of the classification methods such as hierarchical clustering, K-means, Fuzzy C-Means (FCM), modified follow-the-leader and SOMs are compared while assigning the RLP into different client's classes. As a result, the modified follow-the-leader and the hierarchical clustering outperform other clustering techniques. Then, the CRLP is computed as a weighted average of the original survey data of the clients affected in the group. The weights are defined by the reference values (the maximum power values of the survey client).

G. Chicco et al. [121] summarized the procedure of obtaining TLPs (referred as class representative load diagrams in the article) for electricity tariff structure. The procedure begins by collecting the sampling data, through bad data detection, and ends with clustering feature selection and clustering techniques. The clustering techniques are grouped into two systems: time domain approaches and frequency domain approaches. The clustering performance is indicated by the adequacy indicators.

Various classification methods exist. According to D. Gerbec et al. [122], methods of estimating TLP are derived into two systems: one is by predefining consumer's groups from the load survey information. The other one is by using the pattern recognition of the load curves. The first category is mainly based on the qualitative information to stratify clients,

whereas the second category depends mainly on the sampled data for the classification and the adjustment automatically according to the new data are often allowed. It is worth mentioning that end-use method can also be served for the TLP characterizing [123, 9]. Nevertheless, the end-use method requires extensive information about end-use appliances and end-users. Considering the great difficulty getting these information, this method is beyond scope in this chapter. For further information, readers can reference to subsection 2.1.b-i.

In the first system, customers are classed in different groups according to their specific characteristics, such as the subscribed power in the supply agreement, the electricity usage (customer's activity) [118], monthly consumption [118], to name a few. Each class has its TLP presented by different statistically calculated coefficients. Load curves are often considered following the Gaussian distribution that is simply represented by its mean and variance [118]. A given client's daily load profile can be obtained by scaling the average unit pattern of the group to its annual energy consumption. The results showed that the aggregated mean and standard deviation estimation of the power on the transformer level are close to the actual measurements [118]. Later in this section, we will present in details the actual load estimation method BAGHERRA applied by the French electricity company EDF, which belongs to this category of methods.

In the second system, numerous algorithms of clustering appear in recent years with the development of smart meters. These later provide detailed indexed consumption of individual client and enable the pattern recognition methods for clustering.

D. Gerbec et al. [122] proposed a clustering method integrating wavelet method for denoising, FCM method for creating the clusters from the client's survey data and Probability Neural Network (PNN) for attributing business code to the clusters deduced by the former FCM method. Finally, the cluster formed by PNN conducted the sub-TLPs representing several business codes attributing to the same cluster.

T. Zhang et al. [124] have proposed a stability index and priority index for choosing the most suitable clustering algorithms as well as the optimal number of clusters among K-means, fuzzy c-means and the SOM.

A. Mutanen et al. [11] proposed a customer pattern recognition classification method named Iterative Self-Organizing DATA-analysis technique algorithm (ISODATA). With at least 80% of the Automatic Meter Reading (AMR) covering by the end of 2013 in Finland, they suggested an iterative process to reclassify and update the customer classes. The clustering is based on the weighted Euclidean distance on temperature dependency parameters and three-month actual load data. The algorithm includes a variant of the K-mean procedure, a splitting and a merging procedure. The efficiency of the proposed clustering algorithm and adopting pattern vector instead of direct measurements are argued in the end of [11].

5.2 Load research projects in different countries

The load research converts customer's annual energy consumption into hourly load values through load models or patterns. In this section, we present the load research in four different countries: Finland, Denmark, Norway, and Taiwan. In Scandinavian countries, for example, the most usual method to estimate the annual peak of a client's load is the

"Velander's formula". The method is based on the correlation between the yearly energy consumption (kWh) E_{yr} and the maximum power (kW) P_{max}, such as [117, 9]:

$$P_{max} = K_1 E_{yr} + K_2 \sqrt{E_{yr}} \tag{5.3}$$

where K_1 and K_2 are two coefficients depending on the category of the client.

The "Velander's formula" (equation 5.3) is a very approximative estimation model that transforms energy consumption to maximum power with correspondent coefficients. It assumes that the connected clients are homogeneous, more specifically, have the similar behavior. Next, it simply sums up the maximum power of each client as the maximum power for the substation. Thus, the more the clients' behaviors are homogeneous, the more accuracy this method will get. As a result, it is only applicable to big scale systems (at least medium voltage network) and this method often overestimates the maximum power.

The TLPs are often obtained on a national wide range, which signifies that they may not adapt to the local circumstances. Thus, local DSO sometimes create new load models based on TLPs in a Distribution Load Estimation (DLE) process.

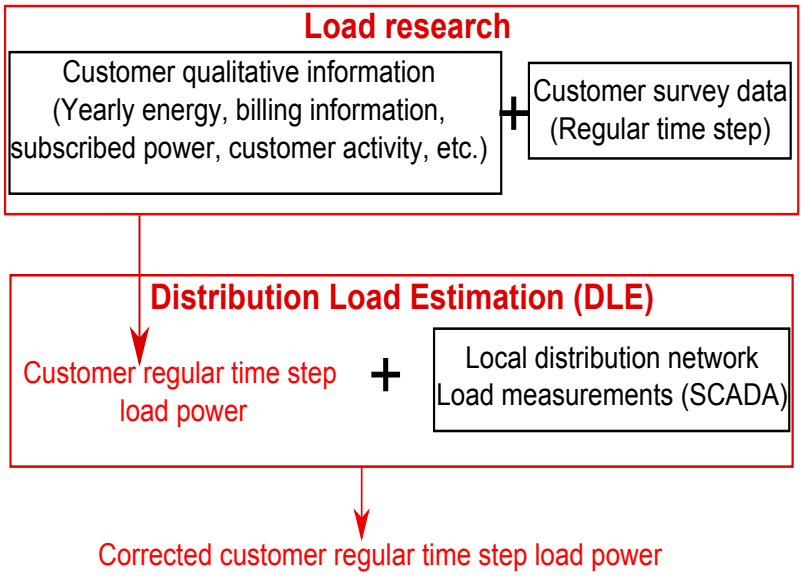


Figure 5.3: Distribution Load Estimation (DLE) process. Regular time step refers to hourly, half-hourly or quarter-hourly load power.

The load measurements in the DLE process are often referred to the current and active power metering at the primary substations [119]. Weighted Least Squares Estimation (WLSE) method is widely applied for the new customer class load estimations [119] in the DLE process. Errors are shared among models and measurements depending on their variances.

5.2.a Finland DSO model

The load research program in Finland started in 1983 [119]. The load power of over 1000 customers is hourly recorded as survey data. The annual energy consumption is used as criteria to class non survey customers into 46 groups. The output of the model can give the load power of any client at any hour of the year. In the actually used model by Finnish

DSOs, the individual client's hourly load $P(t)$ and the standard deviation $\sigma(t)$ are modeled as a linear function of the annual energy consumption E_{yr} in two ways: topography and index series, such that [11]:

$$\begin{aligned} \text{Topography}: \quad P(t) &= L_{topo}(t)E_{yr}/E_{base} \\ \sigma(t) &= s_{topo}(t)E_{yr}/E_{base} \end{aligned} \tag{5.4}$$

where $L_{topo}(t), \{t = 1, \cdots, 8760\}$ and $s_{topo}(t), \{t = 1, \cdots, 8760\}$ are respectively the coefficients of the expectation value and the standard deviation, depending on the geographic information for a base energy consumption E_{base} of 10 MWh/yr.

$$\begin{aligned} \text{Index series}: \quad P(t) &= \frac{E_{yr}}{8760}\frac{Q(t)}{100}\frac{q(t)}{100} \\ \sigma(t) &= P(t)\frac{s_{\%}(t)}{100} \end{aligned} \tag{5.5}$$

where $Q(t)$ is a two-week index of seasonal variation, $q(t)$ is an hourly power variation for three day types (working day, Saturday, and Sunday). Thus the overall power expectation of a year is modeled by 26 $Q(t)$ values and 1872 $(= 26 \times 3 \times 24)$ $q(t)$ values. The standard variation of a client's load is proportional to the average load by the percentage index $s_{\%}(t)$.

Both forms give the power expectation values and standard deviations for each hour of the year. They take special holidays into account, but in different manners. For topography model, they're independently defined, whereas in the index series, the eves and special holidays are respectively mixed with Saturday and Sunday indexes.

The temperature dependent part is the product of the individual client's hourly load $P(t)$, the difference between the average temperature of the previous day T_{ave} and normal temperature $hist\_T(t)$ (long-term historical temperature at a given time on a given day of the year), and a seasonal temperature-dependency parameter α:

$$\Delta P(t) = \alpha(T_{ave} - hist\_T(t))P(t) \tag{5.6}$$

In Finland, with the widespread of the smart meter implementation, the ancient customer classification based on their qualitative information becomes out of date. Methods of classification based on the accurate measurements collected by AMR systems are popular [11].

5.2.b Denmark Dong Energy

In 2008, Dong Energy, the Denmark leading company teamed with IBM and proposed a "SmartPIT" solution for distribution network planning [125]. Two stages are included in the solution. First, the energy sale of each category is converted to peak loads. Number of categories is increased to 27 from the traditional 6 categories. Velander's formula is adopted but the K_1 and K_2 parameters are independently calculated for each category and updated regularly. Second, the peak load is modeled by 22 different season patterns according to the category into hourly consumption estimations. The estimated load values on a MV feeder compared to the real measurement on 5 weeks from 7th December 2007 to 14th January 2008 gave an MAPE of 9%. It is reported that the business unit estimated to have a reduction in the network reinforcement of 80% with SmartPIT.

5.2.c Norway SINTEF Energy Research

SINTEF Energy Research has developed a program named "USELOAD" based on statistical analysis. Each end user i is modeled with a normal distribution, whose PDF is [126]:

$$p(P_i(t)) = \frac{1}{\sigma_{i,t}\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{P_i(t)-\mu_{i,t}}{\sigma_{i,t}})} \tag{5.7}$$

where $P_i(t)$ is the load of client i at time t. $\mu_{i,t}$ and $\sigma_{i,t}$ are independently the mean value and the standard deviation of client i's load at time t. Thus, the peak demand of client i, $P_{i,max}$ with a given excess probability is [126]:

$$P_{i,max} = \mu_{i,\tau} + k\sigma_{i,\tau} \tag{5.8}$$

where k corresponds to the standard normal cumulative function value for a given excess probability. τ is the expected moment when the peak demand occurs.

Let $\mu_{1,i,\tau'}$ and $\mu_{2,i,\tau'}$ be the expected loads at time τ' for category 1 and category 2, n_1 category 1 clients' and n_2 category 2 clients' aggregated load $P_\Sigma(\tau')$ follows a normal distribution such that:

$$\mathcal{N}(\mu_{\Sigma,\tau'}, \sigma_{\Sigma,\tau'}) \tag{5.9}$$

where the expected value $\mu_{\Sigma,\tau'} = \sum_{i=1}^{n_1} \mu_{1,i,\tau'} + \sum_{j=1}^{n_2} \mu_{2,j,\tau'}$ and the standard deviation $\sigma_{\Sigma,\tau'} = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{n} \rho_{i,j,\tau'}\sigma_{i,\tau'}\sigma_{j,\tau'}}$. The total number of aggregated clients is $n = n_1 + n_2$. $\rho_{i,j,\tau'} \in [-1, 1]$ is the correlation coefficient between load of client "i" and "j" at time τ'.

The peak demand of n_1 category 1 clients and n_2 category 2 clients is then obtained by equation 5.8 by changing $\mu_{i,\tau}$ and $\sigma_{i,\tau}$ with the correspondent expected value $\mu_{\Sigma,\tau'}$ and standard deviation $\sigma_{\Sigma,\tau'}$.

The advantage of this program is commented to have a great capacity dealing with a mixed of great variety of customer's loads. Apart from estimating the coincident demand of a given number of clients described above, the program can also segment metered energy consumptions collected in a large scale into small scaled end uses or different customers [127].

5.2.d Taipower system

Since 1993, Taipower system has built up TLPs by surveying 1500 customers on a quarter-hourly basis using smart meters. The active and reactive power of survey customers are recorded in the local meter's memory and transferred to the DSO's data center every three months [128]. The total period of survey is of four years. These survey data enable multiple functions such as system planning, operation, maintenance, marketing, rate tariff structure and load management [128]. An iterative stratified algorithm is adopted in order to determine the number and location of the survey clients. The stop criterion is defined as matching the system's power profile with the real time Supervisory Control And Data Acquisition (SCADA) system data. In this way, the derived TLPs represent effectively the actual customer's consumption [128]. In each class, a multiple regression analysis between load and temperature variation is carried out to design the load model, i.e., $P_n(t) = a_0(t) + a_1(t)T_n(t) + a_2(t)T_n(t)^2$, where $\{a_0(t), a_1(t), a_2(t)\}$ are coefficients at time t, $P_n(t)$ and $T_n(t)$ are respectively the normalized power and temperature data at time t [129, 116].

In [116], C.S. Chen et al. have analyzed the effect of the temperature on different customer's class. The temperature sensitivity of each class is found by first order derivative of the multiple regression equation of the correspondent class. The power loss on load buses under the temperature variation is solved by the temperature sensitivity of each client's class as well.

5.3 French load research project

The French load research program was set up for the technical analysis, which is an important step for the investment decision making. The good electricity supply is defined by the voltage scale. In the MV network, all nodes must be within ±5% of the nominal voltage value and in the LV network, the scale is defined as ±10% of the nominal value [130, 131]. With more and more clients connected to the networks and rising of the consumption, the voltage drops, especially for the clients at the end of the distribution line. Their voltage is most likely to drop off the admissible scale. As explained before, these clients that are outside the permitted scale are called "poorly supplied clients". In this circumstance, two devices run to rescue: HV/MV transformer's on-load tap changer (also known as under-load tap changers) and MV/LV transformer's no-load tap changer (also known as de-energized tap changers) [130]. Let U_0 denote the nominal value i.e., 20 kV (occasionally 15 kV) for the MV network, and 400V for the LV network. The HV/MV transformer's on-load tap changer varies $U_0+2\% \sim U_0+4\%$ [97]. The MV/LV transformer's no-load tap changer has five step switch values: ±5%, 0 and ±2.5% (new generations have 3 step values: +5%, 0 and +2.5%). The reference value for the HV/MV transformer's on-load tap changer is normally $U_0+4\%$. When a great number of productions connect to the MV network, the reference value varies and can reduce till to $U_0+2\%$.

For the reason of coherence and ease of computation, the voltage calculations in French DSO systems are carried out in terms of percentage to the nominal values. Figure 5.4 [97] shows two critical voltage-drop situations: no MV production and maximum MV production. In the first situation, the HV/MV transformer's tap changer is at the maximum value +4% in order to provide the maximum capacity to connect clients. The voltage-drop on the MV level is limited to −5%. Thus, in the most unfavorable situation, where we have the maximum voltage-drop on the MV level, the primary voltage of the MV/LV transformer is of −1%. The MV/LV's no-load tap changer is often set up at +2.5% in order to connect most client's load possible and at the same time guarantees a certain LV production's connection [97]. With a margin of 1.5% kept for the voltage-drop due to the client's in-house connection, the maximum voltage-drop on the LV level (including transformer's voltage-drop) is 10% $(4\% - 5\% + 2.5\% + 8.5\%)$ [97]. In the second situation, where there is a massive production penetrating into the MV network, the HV/MV transformer's tap changer is reduced to +2% to avoid the over voltage situation in the MV network [97]. In the mean time, the MV network is heavy loaded with a voltage-drop of 5%. The MV/LV transformer's tap changer is placed at +2.5% [97]. Therefore, the total voltage-drop is 8% $(2\% - 5\% + 2.5\% + 8.5\%)$.

Each regional DSO computes the percentage of the "poorly supplied clients" connected to his network. This percentage should not exceed 5% in any department [131]. If this standard is not respected, the regional DSO needs to reinforce his local network.

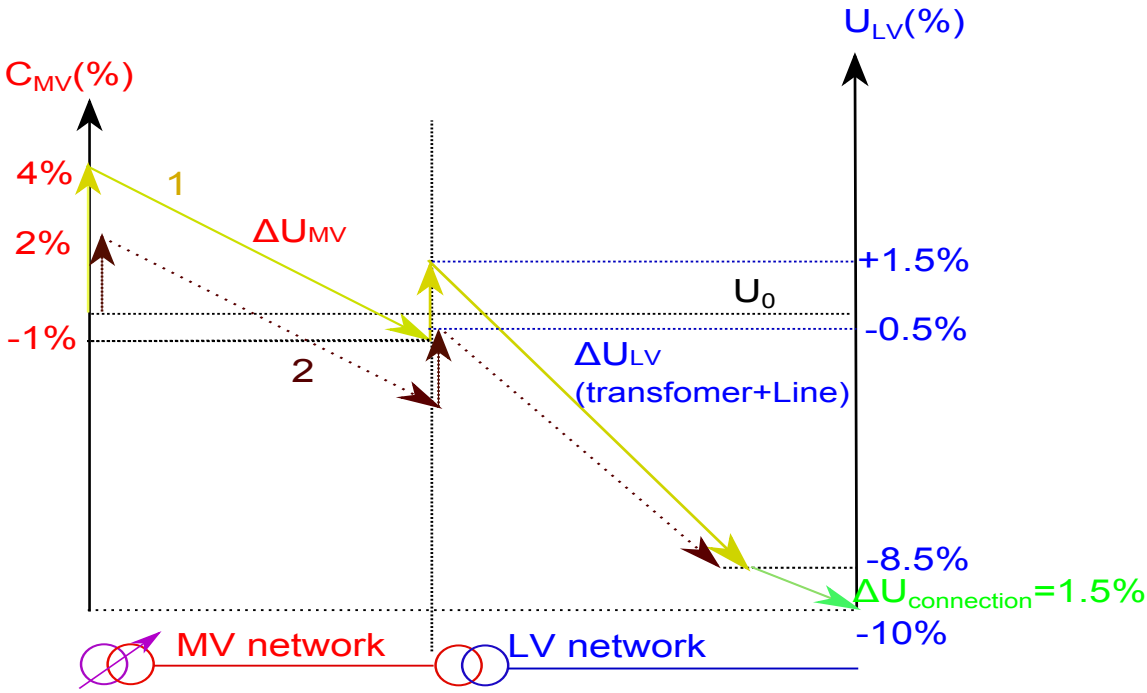


Figure 5.4: Voltage-drop and tap changer adjustment.

From 1997, EDF adopted an estimation model named BAGHEERA for clients connected to public MV/LV substations. Like most of the load research models, it clusters the clients into groups according to their billing information, such as client's social activity (residential, agriculture, commercial, industrial, etc.), and tariff option (base, off-peak/on-peak, Tempo [1]). It aims at estimating the peak demand and provides 48 hourly power estimations for working days and weekends of an individual LV client. The advantages of BAGHEERA model can be summarized as [132]:

- A universal model that adapts to all categories of clients
- The on-peak/off-peak period information of every client is taken into consideration
- The output of the model evolves 48 power values. Thus when carrying out the electrical computation, the coincident effect is already included.

As a matter of fact, in the voltage-drop calculation, the consumption MV client is often directly measured while LV client's measurements are replaced by the 48 hourly estimations of BAGHEERA model. Information such as network data (the conductor's standard, network topology), location of the clients, and connection mode (single, bi- or three phases) is provided by a data base called "GDO" (network infrastructure management). The output of the calculation gives 48 images of the voltage map of the network. In the absence of the information on which phase each client is connected, an imbalance coefficient is applied in the voltage-drop calculation. The imbalance coefficient is independently computed for each LV feeder. The line losses are calculated according to the voltage dispatching situation of the network.

[1] Three day tariffs combining with off-peak/on-peak periods, i.e. six different tariff rates

5.3.a Data description

Mainly, there are three contract types in France: basic option, off-peak/on-peak option, and Tempo option. Each of them associates with a standing charge and a particular tariff rate. The standing charge is a fixed part paid by clients to connect to the electricity network. While the tariff rate is the price paid for one unit energy (1 kWh) consumed. Clients are free to choose any kind of contract with the usage of their house appliances. Samples from the first two types are used for the analysis sake. The basic option is the simplest type with the cheapest standing charge and a uniform tariff rate for all the times of the year. It is more suitable for housings without electric heater or holiday homes with only occasional usage. The off-peak/on-peak option has a higher standing charge than basic option but offers an eight-hours-cheaper tariff rate during a day. This option is suitable for families having hot water tank which turns on automatically when switch to the off-peak rate. Generally the off peak hours are in the night and in the middle afternoon. The Tempo option is the most complicated charging system among all. The system has six rates of electricity pricing depending on EDF's load forecast on that day. This option is mainly for large households with electric heating and full time occupation.

In France, over 2000 clients' load consumptions are regularly collected as the survey data. These consumptions are sampled on a 10-minute basis. Other clients have three meter recordings a year for billing's use. For our analysis sake, 70 load curves from the survey data are collected from different regions in France during 2 years from July 01, 2004 to June 30, 2006. These load curves represent the power consumption of 35 basic option clients and 35 off-peak/ on-peak option clients. Clients are numbered from 1 to 35 in each group in the database. These numbers are simply to identify the clients in the group. The numbers have no specific meaning in terms of load consumption. Thanks to these identification numbers, each client is related to some useful information, such as geographical location, subscribed power, and activity category, to name a few. As a critical factor, the temperatures of the correspondent region are provided during the same period. Figures 5.5 and 5.6 illustrate representatively the daily average loads of clients subscribing to the two types of contract. We can easily notice the increase in the power consumption of *client no.5* (figure 5.5) during winters due to the use of electrical heating devices. On the other hand, because alternative heating devices powered by gas, oil, or wood are used, for instance, instead of electrical heating devices connected to the distribution grid, the energy consumption pattern of *client no.18* (figure 5.6) appears stable during the year. For the ease of the demonstration, hereinafter *client no.5* from the off-peak/on-peak option and *client no.18* from the basic option are the main examples for demonstration and comparison of the methods.

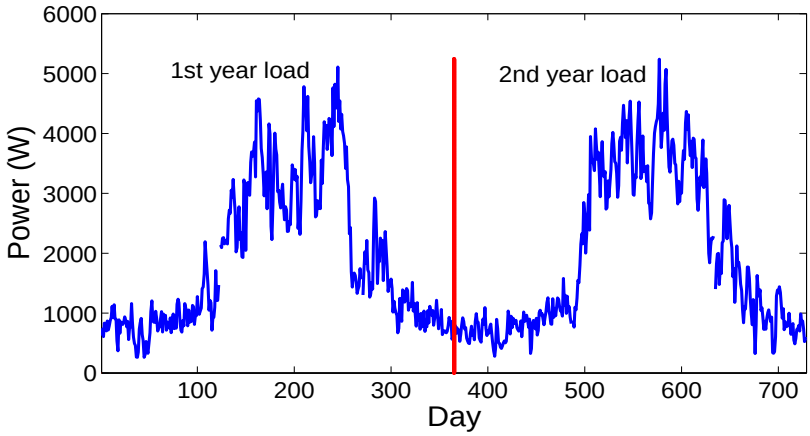


Figure 5.5: Two-year (July 01, 2004 ~ June 30, 2006) daily average loads of off-peak/ on-peak option *client no.5*. The vertical line separates the two-year period.

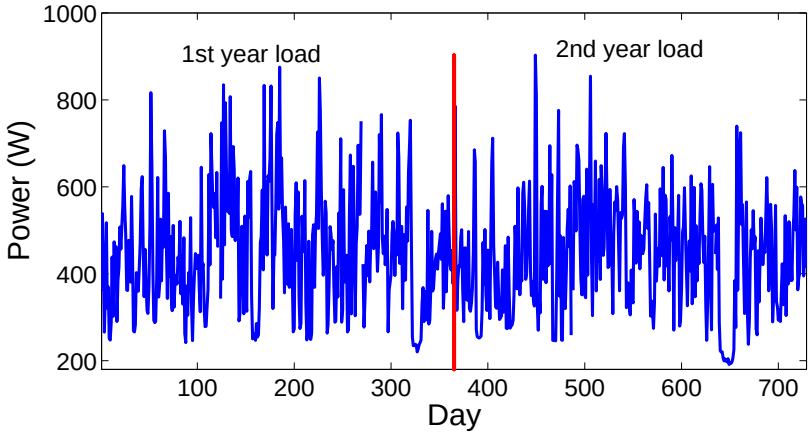


Figure 5.6: Two-year (July 01, 2004 ~ June 30, 2006) daily average loads of basic option *client no.18*. The vertical line separates the two-year period.

5.3.b EDF BAGHEERA model

BAGHEERA model is an estimation model collaborate with temperature. Like the most of the models presented previously, it works with the clustering groups and TLPs. There are four steps to build the model [133]. Firstly, cluster the clients into coherent groups by their qualitative information. Secondly, estimate 48 couples of values (mean and variance for weekdays and weekends) for every individual survey client. A model describing the load at hour "h" required by a single client in respect of the temperature and the chronological pattern is built. Thirdly, estimate the TLP of the group based on the mean and variance values of the survey client in the group. In the end, the mean and variance of each client's load pattern is scaled with his specific coefficients i.e. daily energy, temperature sensibility and yearly energy.

In this section, we aim at presenting the BAGHEERA model in details. The results are demonstrated with data examples introduced in the section 5.3.a. Firstly the Minimum

Temperature Base (TMB) temperature[2] and basic form of the model are revealed. As a matter of fact, the estimation values at TMB temperature are applied for the decision makings in network planning. The BAGHERRA model is composed by two parts: the mean power and the margin. Next, the common coefficient estimation based on which the TLP of the group is defined is explained. Then, the estimation of specific parameters of each client is depicted in distinctive situations: with sufficient historic data, with insufficient historic data and with no historic data. In the end, the output of the model is shown through demonstrative examples.

5.3.b-i TMB temperature and basic model

Cooperating with electrical calculation in a worst-case situation, the BAGHEERA model estimates an individual client's load at TMB temperature. The French electricity company EDF defines TMB as the temperature threshold for every region. This latter is defined such that there exists only one day per year with an equal or lower temperature[3]. Consequently, the probability of observing a day with a temperature equal to or lower than TMB is, in general, 1/365, i.e., 0.3%. The model provides hourly estimations of weekday and weekend load profile (48 points) for every individual client. There are two components in the basic model: the mean power $P(t,T_d)$ at temperature T_d and the margin $\nu(t)$. For each hour, we have:

$$P(t,T_d) = a(t)E_0 + b(t)s(T_d - T_{Nh})|_{T_d<T_{Nh}} \tag{5.10}$$

$$\nu(t) = \sigma(t)^2 E_n \tag{5.11}$$

where $\{a(t), b(t), \sigma(t)\}$ are coefficients statistically calculated and shared with clients in the same class. The parameters $\{E_0, s, E_n\}$ are specific to each client and respectively stand for daily non-heating energy use, gradient, and annual energy use adjusted to the normal climatic condition. T_d is the daily average temperature. Equation 5.10 assumes that above the "non-heating temperature" T_{Nh}, which differs from region to region, the consumed power $(a(t)E_0)$ is independent of the temperature and that below this temperature, the relationship between consumed power and temperature variation is linear. The linear relationship is indicated by the gradient s ($s<0$).

The mean power represents the expected hourly load estimation at temperature T_d, while the margin represents the uncertainty of the estimation. The confidence bound fixed for the distribution network planning study is 10%, which signifies that the upper bound has a 10% chance of being exceeded. The 10% upper bound is defined as:

$$P_{10\%}(t,T_d) = P(t,T_d) + c_{10\%}\nu(t) \tag{5.12}$$

where $c_{10\%}$ is called the risk coefficient at 10%. Therefore, theoretically, the planning confidence bound at TMB temperature is exceeded during $8760h/year * (10\% * 0.3\%) \approx 2.5$ hours per year.

[2] In French: Temperature Minimum de Base

[3] In practice, EDF defines the TMB based on a 30-year historical period. The probability of one day per year is an average value: in reality, during a warm year, we would probably find no daily temperature below this TMB value, while during a cold year, several daily temperatures would be found below this value.

5.3.b-ii Common coefficient estimation

The power measurement y_t can be divided into two independent parts: the thermosensitive power $\varphi(T_t)$, which is a power function depending on the temperature at the same moment T_t and the non thermosensitive power $P_{t,0}$, such that [133]:

$$y_t = P_{t,0} + \varphi(T_t) \tag{5.13}$$

Thus, the correspondent energy consumption is

$$E_d = E_0 + f(T_d) = E_0 + s(T_d - T_{Nh})|_{T_d<T_{Nh}} \tag{5.14}$$

where E_d is the daily energy consumption and $f(\cdot)$ is the function of the daily heating energy consumption influenced by daily temperature.

Bring equation 5.13 and 5.14 into equation 5.10, we have [133]:

$$\begin{aligned} a(t) &= \frac{P_{t,0}}{E_0} \\ b(t) &= \frac{\varphi(T_t)}{f(T_d)} \end{aligned} \tag{5.15}$$

where $a(t)$ and $b(t)$ can be respectively considered as the coefficient that converts non-heating daily energy and heating daily energy to the correspondent hourly power. In fact, coefficients $a(t)$ and $b(t)$ depict the reduced scale pattern of the heating and non-heating power pattern.

$\sigma(t)$ stands for the standard deviation of the error term. This later is the difference between the observed power y_t and estimated power $P(t, T_d)$.

5.3.b-iii Specific parameter estimation

The specific parameters include the temperature sensibility s, non-heating daily energy consumption E_0 and yearly energy consumption E_n. For clients that do not belong to the survey group, there are three categories according to their data's availability. Here, we present how these specific coefficients are estimated for different categories of clients.

1. Client with sufficient historical data

   This category represents the situation where at least six meter recordings during two years (three recordings per year) of an individual client are available. These recordings must be regularly taken and represents the heating and non-heating periods' consumption of the client. E_0 and s are defined based on the six consumption loads of every four months.

   Integrate equation 5.14 on n_i days (a four-month period, in this case) [133]:

   $$\sum_{d=1}^{n_i} E_d = n_i E_0 + s \sum_{d=1}^{n_i} (T_d - T_{Nh})|_{T_d<T_{Nh}} \tag{5.16}$$

   Divide equation 5.16 by n_i, we get:

   $$\frac{E_i}{n_i} = E_0 + s\frac{Dd_i}{n_i} \tag{5.17}$$

where $E_i = \sum_{d=1}^{n_i} E_d$ is the meter recording for the given period, and the degree days Dd_i, which is the sum of the temperature degrees inferior to the non-heating temperature T_{Nh} during the given period: $Dd_i = \sum_{d=1}^{n_i}(T_d - T_{Nh})|_{T_d<T_{Nh}}$. As a matter of fact, a fitter correlation is found with the exponentially smoothed temperature data than with the actually measured data regarding to the power [133]. In consequence, the exponentially smoothed temperature is used for the Dd_i calculation. E_0 and s are estimated by the ordinary least square on the six-couple data $(\frac{E_i}{n_i}, \frac{Dd_i}{n_i})$.

The annual energy consumption adjusted to the normal climatic condition E_n can be estimated by integrating equation 5.14 on a year period [134]:

$$E_n = 365E_0 + sDd_{365} \tag{5.18}$$

where Dd_{365} is the degree days of a year period on the normal climatic condition.

Since off-peak and on-peak periods' consumptions are measured with two electricity meters seperately, the off-peak/on-peak option clients often have twelve recordings instead of six. Therefore, with the equation 5.17, we obtain two couples of specific coefficients: $\{s_{HP}, E0_{HP}\}$[4] for the on-peak period, and $\{s_{HC}, E0_{HC}\}$[5] for the off-peak period. In this way, differences in temperature sensibility due to the heating device modes as well as due to the day/night modes can be figured out [133]. The off-peak/on-peak clients have also two sets of common coefficients $\{a(t), b(t), \sigma(t)\}$. They're respectively calculated following equations 5.13 ,5.14 and 5.15. When reconstruct the group's TLP, the off-peak period variation replaces the on-peak period variation on the off-peak periods. The clients that subscribe the off-peak/on-peak tariff option have an overall eight-hour off-peak period a day. The eight hours can be allocated to one, two or even three distinct periods in the afternoon and at night, when the network is not heavily loaded. The off-peak common coefficients are applied each time the off-peak periods occur.

2. Client with insufficient historical data

Sometimes, the six meter recordings are not complete due to the incorrect readings or data absence. In this case, the clients are considered not having sufficient historical data [135]. For this category of clients, equations (5.10, 5.11) estimating the power and the variance depend on the specific variable E_n, such that [135]:

$$\begin{aligned} P(t, T_d) &= a(t)'E_n + b(t)'E_n(T_s - T_d) \\ \nu(t) &= \sigma(t)'^2 E_n \end{aligned} \tag{5.19}$$

where the client's annual energy consumption adjusted to the normal climatic condition E_n is expressed as a function of the mean daily non-heating energy consumption values of the clients in the same category $\overline{E}_0$ and the heating period in proportion to estimated temperature sensibility of the client $(E_a - (n_a\overline{E}_0))/Dd_a$, such that [135]:

$$E_n = 365\overline{E}_0 + (E_i - (n_i\overline{E}_0))(Dd_{365}/Dd_i) \tag{5.20}$$

[4]The index "HP" means in French "Heure Pleine" (on-peak).
[5]The index "HC" means in French "Heure Creuse" (off-peak).

E_i is the meter recording during the given period. n_i is the correspondent number of days during the period and Dd_i is the number of degree days during this period. {$a(t)$', $b(t)$', $\sigma(t)$'} are correspondent common coefficients. T_s is the temperature threshold where the client's annual degree days equal to 50. More detailed information is described in [133].

3. Client with no historical data

   For this category of clients, the only possible estimated parameter E_n is calculated based on the subscribed power P_s, such that:

$$E_n = c + dP_s \tag{5.21}$$

   where $\{c, d\}$ are parameters depending on the client's activity and tariff option [135].

   Their mean power values and margins are also computed following the equations 5.19.

5.3.b-iv Illustrative example and model's output

The survey load data of the clients are recorded on a 10-minute basis. The temperature data are recorded on an hourly basis. These two data are converted to a daily basis in order to estimate the specific parameters. As previously stated, the specific parameters {E_0,s} are obtained by curve fitting the daily energy measurements with the temperature variations (equation 5.14). Given an equation with parameters and samples to be fitted, the curve fitting tools can calculate the values of the parameters by minimization of the Sum of Square Residuals (SSR). Being an open source Matlab toolbox, which is simple, quick and direct for the curve fitting, the ezyfit toolbox [136] is applied in our study. The above T_{Nh} part data are fitted with a constant E_0 function and the below T_{Nh} part data are fitted with a linear function $(s(T_d - T_{Nh}) + E)$ using least square error criteria. E_n is calculated according to equation 5.18, which equals the integration of the consumed energy on the normal condition basis.

As explained, the off-peak/on-peak option client has two electricity meters, which record separately the energy consumed during on-peak hours and that consumed during off-peak hours. Figures 5.7 and 5.8 show the curve fittings to daily energy measurements of off-peak hours and on-peak hours. The consumed energy is observed to vary with the temperature. In other words, the off-peak/on-peak option client often has an important gradient value. We can see from the figures that the off-peak/on-peak option client no.5 is more sensitive to the temperature during the on-peak periods than the off-peak periods. In fact, as the T_{Nh} reflects the region's mean non-heating temperature, this factor would vary among different clients in the same region. We can see in figure 5.7 that the average energy value during non-heating days (those whose temperature is superior to T_{Nh}) $E0_{HC}$ is different from the regression non-heating energy value E_{HC}. The same situation is found in figure 5.8, i.e., $E0_{HP} \neq E_{HP}$ for the on-peak period. Basic option client, using other means of heating rather than electricity stays stable with the temperature variation (figure 5.9). Thus a small s value is often found in this category of clients. For the same reason, the average non-heating energy $E0$ is different from the regression non-heating energy value E.

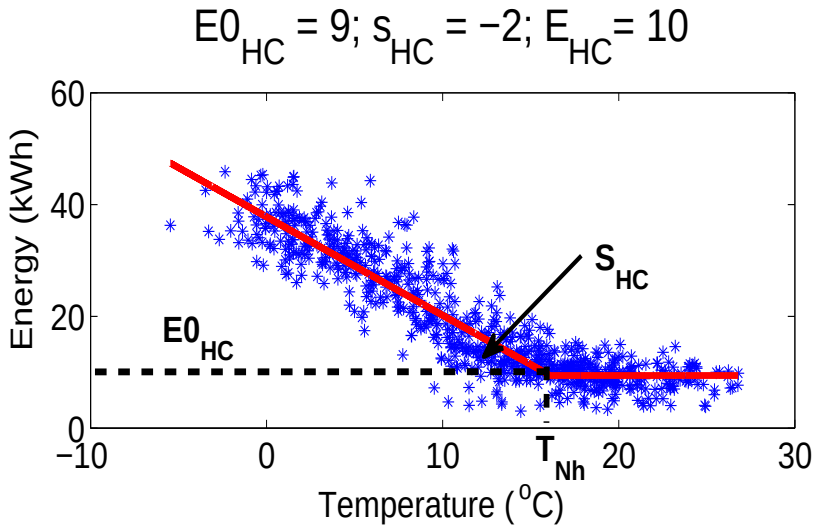


Figure 5.7: Off-peak/on-peak option *client no.5*: curve fitting on off-peak daily energy use

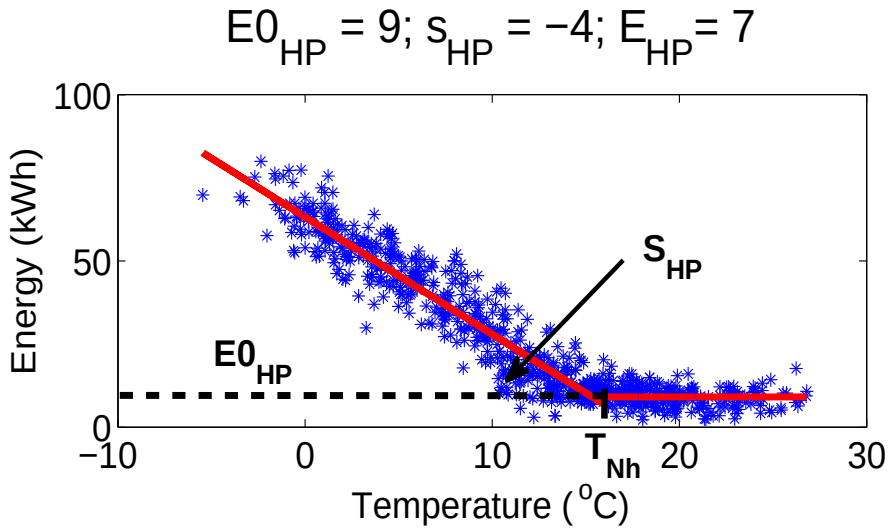


Figure 5.8: Off-peak/on-peak option *client no.5*: curve fitting on on-peak daily energy use

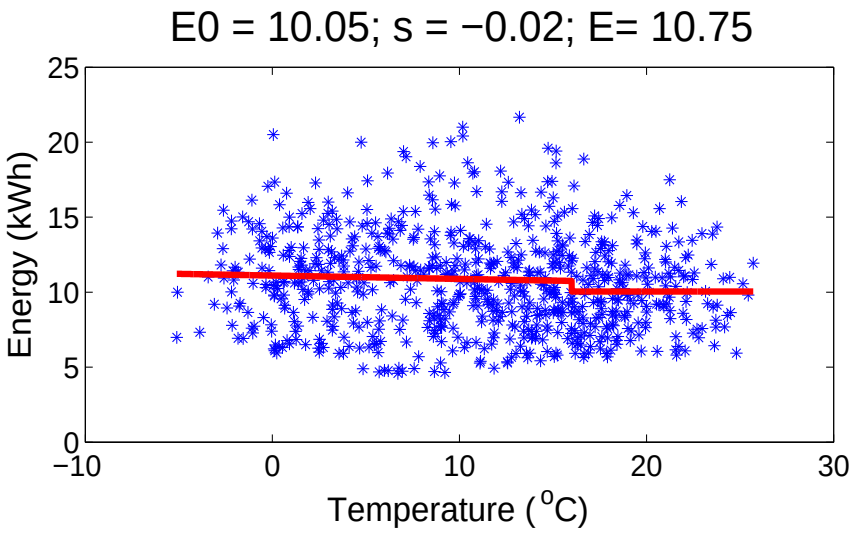


Figure 5.9: Basic option *client no.18*: curve fitting on daily energy use.

The common coefficients $\{a(t), b(t), \sigma(t)\}$, on the other hand, are known based on the category the client belongs to. Equations 5.10, 5.11, and 5.12 are carried out to calculate the 48 estimation values and their upper bounds. Notice that different groups of $\{a(t), b(t), \sigma(t)\}$ values exist for weekdays and weekends, and for on-peak and off-peak hours. To calculate the off-peak/on-peak option clients' load estimations, the on-peak estimations are replaced with the off-peak estimations during the off-peak periods. If there is more than one off-peak period, the estimated power is the same at the beginning of each period. For example, the *client no.5* has two off-peak periods: 01:00 to 07:00 and 12:00 to 14:00. We can see in figure 5.10 that two peaks at the beginning of each off-peak period have the same magnitude. The method also limits the 10% upper bound by the subscribed capacity of the client. In figure 5.10, the bound estimations higher than 9 kW are replaced by 9 kW, the subscribed capacity of the *client no.5*. Figure 5.11 shows the differences in the shape of mean powers between weekday and weekend day type. These differences are depicted by their different common coefficients of the day type.

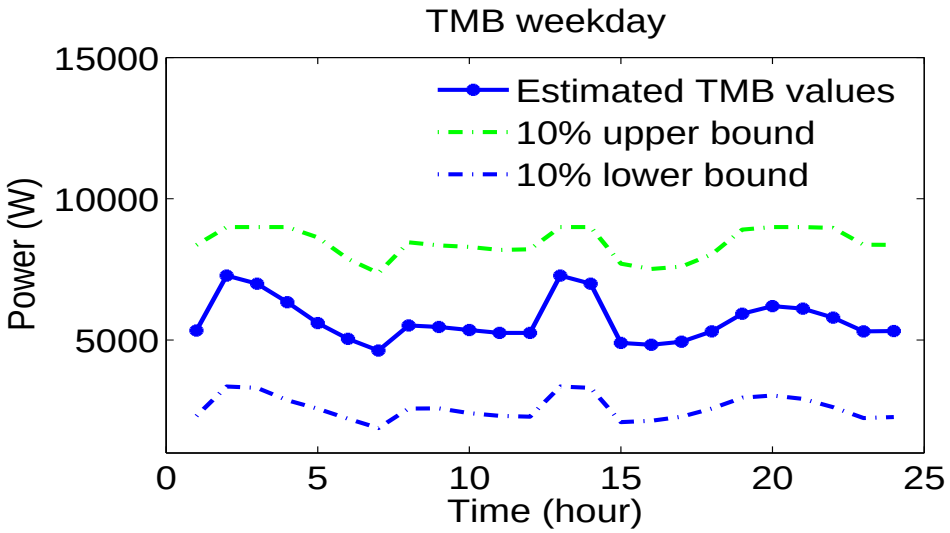


Figure 5.10: Off-peak/on-peak option *client no.5*: outputs of the BAGHEERA model, TMB load estimations on weekdays

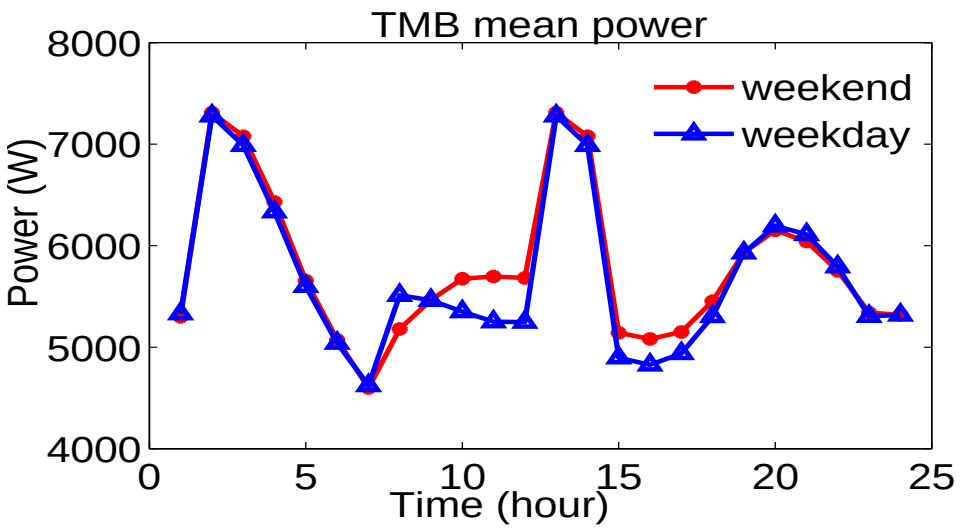


Figure 5.11: Off-peak/ on-peak option *client no.5* : comparison of TMB weekend's and weekday's load estimation

5.4 Conclusion

In this chapter, we briefly described the decision making process and explained that the load estimation model is vital to the first step of the process: the technical analysis. We discussed the solutions to the measurement scarcity in the actual distribution network. Coincident loads are used to estimate the MV/LV substation's capacity and TLPs are used to replace the individual measurements.

Load research projects in different countries were introduced. Widely used "Velander's formula" for the individual's maximum power estimation is depicted. DLE process is often applied to adapt national TLPs to local TLPs according to the SCADA measurements. Finland DSO models the mean load power and the standard deviation in both topography and index means. Temperature dependence is also associated to the model. Denmark Dong Energy company proposed a "SmartPIT" solution that updated the "Velander's formula" and modeled the peak loads by 22 different season patterns. They reclaimed to have a reduction of 80% in the network reinforcement. The Norway SINTEF energy research modeled the end user's load with a normal distribution. The Taipower system built up the power as a regression function of the temperature, so that the temperature sensitivity can be easily analyzed.

The French load research project is carefully scrutinized. The "poorly supplied clients" are defined and the voltage-drop plan with the tap changers' adjustment is explained. The BAGHEERA model actually applied by the EDF is thoroughly presented, including the estimation model as well as the coefficients' estimations. The model's outputs are shown through illustrative examples.

In fact, the BAGHEERA model assumes that the influence of the consumption by the temperature is linear. However, this is not realistic due to the limited maximum power of heating devices. Moreover, the smart grid concept leads to a great dynamism in the client's behavioral consumption as well as flexible relationships among electricity producer, distributor and client. Thus, it is no longer reasonable to prefix a client to a certain group.

Furthermore, with the air-conditioner gaining its popularity, load consumption will grow when the weather gets hot. The assumption made in BAGHEERA that the load stays unchanged above a certain non-heating temperature is no longer real. Hence, our idea is to build an individual model for each customer without clustering step. This model must be completely data-driven and independent to the qualitative information of the client whose quality is decreasing. Especially with the development of smart meters, providing detailed individual consumption information, we are convinced that the clustering step is no longer necessary. As a result, in the next chapter, we propose a non parametric individual estimation model.

Chapter 6

Nonparametric model

Contents

Abstract

In this chapter, we tackle client load estimation in a smart grid network. For that purpose, we propose an individual model based on nonparametric estimators. The model is designed for voltage-drop calculations for use in network planning. Completely data-driven, the proposed methodology can be applied to both thermosensitive and non-thermosensitive clients. Real measurements collected in French distribution systems are used to validate our methodology. The proposed approach produces more reliable estimation results than the current model BAGHEERA of the French electricity company EDF does on the same data. A discussion on the definition of the threshold power estimations is carried out in the end of the chapter.

Distribution network planning involves developing a schedule of future additions that ensure the quality of energy delivery as well as the lowest possible cost. On the one hand, the electricity infrastructure must meet the needs of peak loads. On the other hand, over-dimensioned systems can be very expensive. Thus, reliable load models are required to perform distribution network calculations, such as power flow calculations in critical situations so as to identify poor electricity supply zones for investment planning.

Individual load is mostly influenced by two main factors: day types and weather conditions. The daily load pattern can be very different from weekdays to weekends. Temperature is a primary weather factor that affects load variation. The customer model is designed for worst-case calculations [7, 8], namely, the time of maximum demand and minimum supply as well as the time of minimum demand and maximum supply contrawise [9]. As the peak loads of several customers rarely occur at the same time, estimation of the customer consumption load at every hour is required. Uncertainty regarding load estimations also needs to be taken into account [10]. Generally, for the voltage-drop calculation the excess probability is fixed at 10% [11]. Therefore, the objective is to define an individual customer's maximum and minimum load limits at different day types of a year with 10% excess probability.

As stated in section 5.3, the BAGHEERA model [137] is actually applied by the French electricity company EDF for its distribution network planning. Over 2000 demand survey load curves sampled every 10 minutes all over France are collected to predefine 66 classes. According to some qualitative information, such as supply agreement and client activity category, to name a couple, every client in the French territory is associated with a predefined class. Clients in the same class share some statistically calculated coefficients defining a daily profile pattern for weekdays and for weekends. The load estimations of a specific client are computed by scaling the average shape of his particular group to his annual unit consumption. The output of the model is the one hour interval power consumption (48 points: 24 for weekdays and 24 for weekends) of each client at minimum temperature of the region, i.e., Minimum Temperature Base (TMB).

However, the smart grid paradigm introduces three ground-breaking changes to current electricity networks. Firstly, it enables new products, services, and markets, and thus establishes a more flexible relationship among operators, clients and regulators [138]. Consequently, the qualitative information of each client is less and less precise. Secondly, customers are encouraged to modify their behavior to interact with the real-time electricity market [139]. Therefore, it is difficult to assign a client to a predefined class. Thirdly, in 2009 the ERDF (French electricity distributor) launched the "Linky" project to install 35 million smart meters in France. These smart meters collect the power consumed by individual clients on a 30-minute basis and automatically transfer the electricity consumption information to the data center of the ERDF at the end of every day. The data archive system is designed to keep the two-year historical consumption data of every client [3]. The extensive consumption information collected by smart meters enables us to build more accurate individual estimation models. Within this context, this chapter aims to design a universal individual load estimation model for the interests of distribution network planning.

Before the implementation of the smart meter, historical load data were not usually available, apart from the demand survey data of a limited number of clients. Most of

the works concerning distribution network planning aimed to estimate the peak demand for a group of customers during periods of peak system demand, namely, coincident peak demand [6]. For this purpose, three categories can be identified. Some methods provide a relationship between the maximum demand of a group of clients and individual clients by some correction factors [9] , such as coincidence factor, Velander's formula, etc. Others concentrate on classification methods [11, 122, 140] to sort clients into TLPs. Still others apply the end-use method [9, 22] by composing the residential customer model into appliance elementary units. The drawback of the end-use method is the extensive demand for input data. Our work, inspired by [141], proposes the building of nonparametric regressors to express the relationship between load demand and temperature and the application of nonparametric probability density function to model the random nature of demand of a customer.

The main contributions of our work are threefold. First, the method can be adapted to any client's load, regardless of his load thermosensibility, and is entirely data driven. Second, nonparametric methods are applied so that the model is independent from the client's qualitative information as well as from assumptions made on load functions. Third, the outputs of the method include both maximum and minimum daily power consumption patterns.

The rest of the chapter is organized as follows. Section 6.1 details the nonparametric model procedure as well as the coordinate statistical tools. A computational example is presented and the performance of different models is discussed in section 6.2. Validation study in section 6.3 compares the precision of the models on the extensive examples. In section 6.4, we carried out a discussion on the definition of the threshold power estimations. Section 6.5 concludes the chapter.

6.1 Nonparametric model

Figure 6.1 illustrates the procedure of the nonparametric model. First, the load curve of a given client is associated with the local temperature. Second, *statistical tests* are performed to examine whether the client is thermosensitive. If the client's power consumption is not influenced by temperature, there will be no difference between minimum power and power at TMB temperature. Thus a *Kernel Density Estimation (KDE)* method is applied to determine the median values as well as the 10% upper boundary and the 10% lower boundary. If the client is thermosensitive, the *CUSUM chart algorithm* is employed to separate the heating season from the non heating season. Next, the non-heating season data are used to compute the minimum power, and the heating season data are used to compute the power at TMB temperature. Notice that the TMB temperature is the coldest temperature that happens on average one day per year; basic knowledge of electricity demand in this extreme weather condition is often insufficient. This lack of knowledge may cause a problem for the TMB load estimation. Our strategy is to include past years' consumption information if the compatibility of the data is proven by *statistical tests*. At the end of the tests, compatible data are brought in and the relationship between the temperature and the load consumption is defined with a nonparametric *kernel regression method*. The smoothing parameter required by the kernel regression method is computed by the *Cross-Validation (CV) technique*. Once the relationship between temperature and load

consumption is established, all the heating season data are brought to the TMB condition in maintaining their uncertainties. In the end, the median power and the excess probability boundaries can be computed at TMB temperature. These statistical tools applied in the procedure are thoroughly presented in the following subsections.

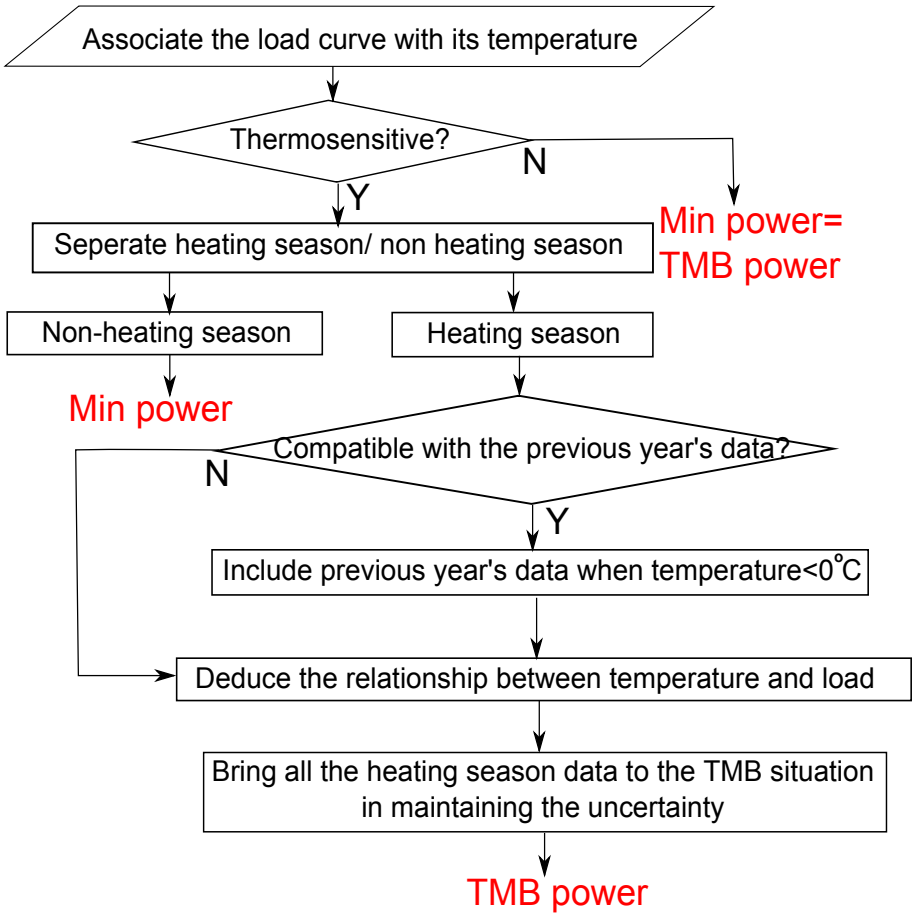


Figure 6.1: Overview of the nonparametric model

6.1.a Statistical tests

Statistical tests are often used for decision making by examining basic sample information. We aim to test if there is any significant difference between the mean of two different groups. In such a situation, *Student's t-test* for difference of means is applied. However, one assumption in carrying out *Student's t-test* is that the variance of the two populations is equal; if not, then *Welch's t-test* is used. An *F-test* can be used to test the hypothesis that the population variances are equal. Figure 6.2 states the relationships between the above statistical tests.

Practically speaking, at the beginning of these tests, two complementary hypotheses are set up, namely, the null hypothesis and the alternative hypothesis. The null hypothesis often indicates that no significant difference is found between two examined groups. The decision is made by comparing a p-value indicating the probability to accept the null hypothesis with a type I error percentage referring to the risk of wrongly rejecting the

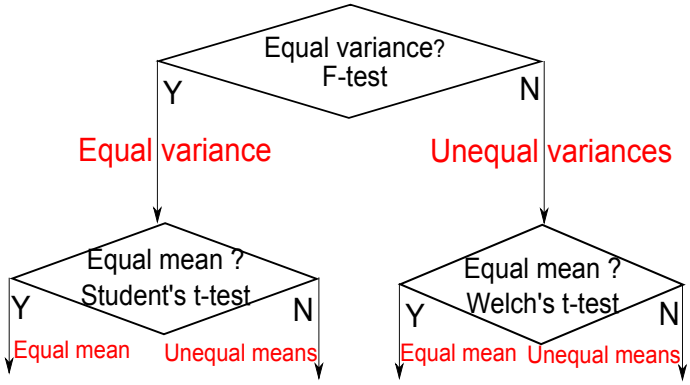


Figure 6.2: Statistical tests procedure

null hypothesis. This type I error is often set at 0.05. In other words, if the p-value of a statistical test is greater than 0.05, one may accept the null hypothesis. Otherwise, the alternative hypothesis is accepted.

6.1.b Kernel density estimation

KDE is a nonparametric method that estimates the density directly from the data without making any parametric assumption about the underlying distribution. In our applications, KDE is used to define the γ, $(0 < \gamma < 1)$ upper and lower bounds of the data set. Let $X_i, i = 1, \cdots, n$ be independent samples drawn from a distribution $g(x)$, the kernel density estimator $\hat{g}_h(x)$ is defined as:

$$\hat{g}_h(x) = \frac{1}{nh} \sum_{i=1}^{n} K(\frac{x - X_i}{h}) \tag{6.1}$$

where $K(\mu) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}\mu^2}$ is a normal kernel function. The smoothing parameter is h.

The adaptive KDE method [142] is applied, where the smoothing parameter is chosen automatically. The reliable kernel estimator based on the smoothing properties of linear diffusion processes deals well with multimodal densities. Readers who want to have more details on the adaptive KDE method can refer to [142].

Thus, the bound limits $\{x_{min}, x_{max}\}$ of $100(1-\gamma)\%$ probability are defined as:

$$\int_{-\infty}^{x_{min}} \hat{g}_h(x)\, \mathrm{d}x = \gamma \quad and \int_{-\infty}^{x_{max}} \hat{g}_h(x)\, \mathrm{d}x = 1 - \gamma \tag{6.2}$$

6.1.c CUSUM algorithm

The idea of the CUmulative SUM (CUSUM) algorithm is to automatically detect the breaking points in the data set that separate the heating season and the non-heating season. Afterwards, the non heating season data are used to estimate the minimum power, and the heating season data are used to estimate the power in TMB condition.

As the name suggests, CUSUM calculates the accumulated deviations of the sampled values $y_i, i = 1, \cdots, N$ from the mean value $\overline{y}$ [143]:

$$S_m = \sum_{i=1}^{N} (y_i - \overline{y}) \tag{6.3}$$

The CUSUM plot is centered on zero. During the heating period, consumption is supposed to be superior to the mean value of the year and the CUSUM points drift upwards; during the non-heating period, the consumption is likely to be inferior to the mean value of the year and the CUSUM points drift downwards. Thus, the breaking points of the two periods correspond to the maximum and the minimum values of the CUSUM plot.

6.1.d Kernel regression

Kernel regression is a nonparametric approach using historical data to define the relationship between load y $\{y_i, i = 1, \cdots, N\}$ and temperature variation X $\{X_i, i = 1, \cdots, N\}$. Three kernel-type regressors are introduced in this subsection: the NW, the LL, and the LL2 estimators. Being the zero- and the first-order nonparametric estimators, the NW and the LL estimators have gained large popularity in function estimations [144]. The LL2 estimator is an adapted estimator originating from the LL estimator in order to adapt the special situation, which will be presented later in this subsection.

Unlike the parametric estimation, the nonparametric approach imposes no assumption on the functional form, and thus, is much more flexible than parametric regressions. The method requires the smoothing parameter h, which can be calculated by applying the CV technique [26]. Estimating the conditional expectation of the random variable x, the kernel regression can be defined as:

$$E(y|X = x) = \hat{f}(x) \tag{6.4}$$

where the $E(\cdot|\cdot)$ is the conditional expectation operator.

By using local constant approximations, namely a local polynomial regression of degree 0, the estimator $\hat{f}(x) = \beta_0(x)$ minimizes the kernel weighted least-squares:

$$\sum_{i=1}^{N} \{y_i - \beta_0(x)\}^2 K(\frac{x - X_i}{h}) \tag{6.5}$$

We have the *Nadaraya-Watson (NW)* estimator:

$$\hat{f}_{NW}(x) = \beta_0(x) = \frac{\sum_{i=1}^{N} K(\frac{x-X_i}{h}) y_i}{\sum_{i=1}^{N} K(\frac{x-X_i}{h})} \tag{6.6}$$

Equation 6.6 shows that the load estimation $\hat{f}_{NW}(x)$ at temperature x is a local weighted average of the historical neighborhood samples. The size of the local neighborhood and specific weights are defined by its smoothing parameter h.

However, the NW estimator applied to define the relationship between load variation and temperature has certain drawbacks. It may have a large bias in the "boundary region" due to the non positive sample distribution function and in the interior points due to unequal spacing of the samples [145]. Therefore, we propose another kernel-type estimator, the LL estimator of degree 1. The idea is to fit the load temperature relationship with local model $\hat{f}(x) = \alpha(x) + \beta(x)(x - X_i)$ to minimize

$$\sum_{i=1}^{N} \{y_i - \beta(x)(x - X_i) - \alpha(x)\}^2 K(\frac{x - X_i}{h}) \tag{6.7}$$

We have the explicit expressions:

$$\alpha(x) = \frac{S_2 \sum_{i=1}^{N} y_i K(\frac{x-X_i}{h}) - S_1 \sum_{i=1}^{N} y_i K(\frac{x-X_i}{h})(x-X_i)}{N(S_2 S_0 - S_1^2)}$$
$$\beta(x) = \frac{S_0 \sum_{i=1}^{N} y_i K(\frac{x-X_i}{h})(x-X_i) - S_1 \sum_{i=1}^{N} y_i K(\frac{x-X_i}{h})}{N(S_2 S_0 - S_1^2)} \tag{6.8}$$

where $S_j = \frac{\sum_{i=1}^{N}(x-X_i)^j K(\frac{x-X_i}{h})}{N}$, $j = 0, 1, 2$.

Thus, the *Local Linear (LL)* estimator can be written as

$$\hat{f}_{LL}(x) = \frac{\sum_{i=1}^{N} K(\frac{x-X_i}{h})(S_2 - S_1(x-X_i))y_i}{N(S_0 S_2 - S_1^2)} \tag{6.9}$$

The LL estimator preserves the linear trend and stays invariant to the first derivative to $\hat{f}_{LL}(x)$. According to the literatures [146, 147, 144], the LL estimator also has much better properties at the boundary than the NW estimator has. High-order estimation functions have better performance on steeper and curvier regression functions. However, when the regression function is quite flat, the NW estimator behaves better.

The nonparametric estimators depend mainly on samples. If there is a declining trend shown by the data, the nonparametric estimators will follow this trend. Even though in the reality, it has no actual physical meaning. Lack of samples in the cold conditions, every sample under such condition has a big influence on the nonparametric regression. Due to the imprecision in measurements or special family occasions, the few samples can sometimes cause an unreasonable trend: i.e, when the temperature decreases, the load trend decreases. However, it is known that when the temperature drops, the electricity consumption rises because of the use of electrical heaters. Thus, we propose the LL2 estimator to correct the flaw in the trend caused by these "misleading" scarce samples. Two steps are followed to establish this LL2 estimator. First, we identify the cutoff point that separates uprising trends and declining trends of the LL estimator. Second, we filter the points on the left side of the cutoff point and conserve the local trend of the points on its right side.

Since the estimation is made by local averaging of historical load data in a correspondent neighborhood, it is obvious that its ability of extrapolation beyond the available historical data domain is rather limited. Hence, the estimations on the "boundary" are rather sensitive to the boundary samples. In our application, as TMB load needs to be estimated, but very few load samples are collected under this very cold weather condition, we propose to add previous years' data with cold weather conditions to the estimation samples if the data characteristics are similar to those of the current year, having the same mean and variance.

6.1.e Smoothing parameter selection: cross-validation technique

It is well known that the result of kernel regression depends crucially on the bandwidth h selection. A large h would give an oversmoothed estimation with a large bias, whereas a too small h would give an undersmoothed estimation with a large variance. The CV technique aims at finding the parameter h_{cv} that minimizes the CV function defined as

follows [144]:

$$CV(h) = \frac{1}{G}\sum_{i=1}^{G}\frac{1}{g_i}\sum_{j=1}^{g_i}(y_j - \hat{f}_h(x_j))^2 \tag{6.10}$$

where $\hat{f}_h(x_j)$ is the kernel-type estimator, G is the number of parts and g_i is the number of elements in the i-th part.

The idea is to randomly divide the data into T equal or quasi-equal parts and to estimate the regression function by using $G-1$ parts while calculating the square error on the left part called validation data. The procedure is repeated G times till every part has been used once for the validation. The CV function is a mean square error estimator.

6.2 Computational example

For ease of demonstration and in coherence with the previous sections, the off-peak/on-peak option client no.5 is selected to illustrate a more complex example. We aim to estimate the maximum and minimum loads of *client no.5* during the first year. In this section, we first present the process in subsection 6.2.a. Subsection 6.2.b illustrates and comments on the results as compared with the BAGHEERA model. All the presented results are run in the Matlab environment.

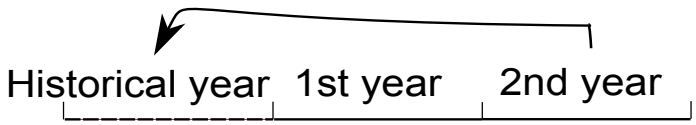


Figure 6.3: Data diagram: historical data, 1st-year data, and 2nd-year data.

As mentioned previously, the French legislation permits a period of maximum two-year consumption data being recorded [3]. The recording data used in our study are from July 1, 2004 to June 30, 2006. The overall two-year data are used both for the estimation and the validation of the model. Data diagram indicating the utilization of the available data is shown in figure 6.3. The 1st-year data are used for the model estimation. The 2nd-year data have two practical applications: as the historical year data and as the test data. In the proposed methodology, because of the scarce samples in cold conditions (temperature $< 0^oC$), we suggested joining coherent historical year data in such condition to the current year data for a more accurate estimation (cf. figure 6.1). We assume that the consumption data have a one-year period and that there is no difference if the year data are taken before or afterwards. Thus, the 2nd-year data are borrowed as the historical year data for the model estimation. In the future practice, we suggest the utilities to record historical data, so that this problem won't exist. In the validation phase, the 2nd data are used as the test data. Since the test data cannot be used during the learning process, the 2nd data are not joined to the 1st data to compute the kernel regression in the validation phase. Therefore, hereafter, when we mention historical (also termed "past") year data and test data, they actually mean the 2nd-year data.

6.2.a Illustrative example results

As previously described, after the local temperature is associated with the client's load curve, the statistical tests are carried out to check whether the examined client is thermosensitive. The statistical tests are performed by comparing the mean values between the load samples inferior to 0^oC and those superior to 20^oC. The choice of these two temperature degrees is arbitrary. We assign 0.05 to the type I error in the statistical test. As described in subsection 5.3.a, the off-peak/on-peak option *client no.5* has probably some electrical heating device at home. Figure 6.4 shows the tests result. The p-value of Welch's t-test rejects the equal mean hypothesis and confirms that this client is thermosensitive.

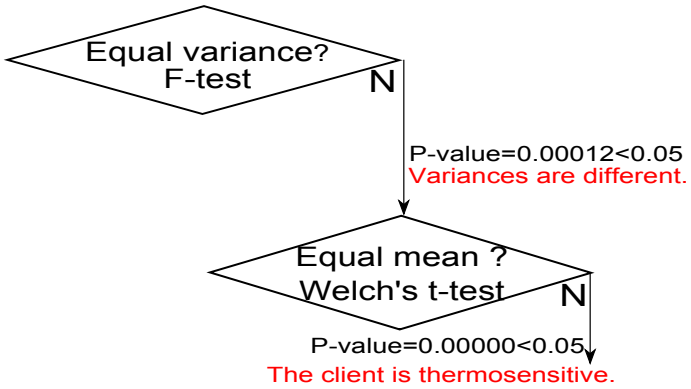


Figure 6.4: Off-peak/on-peak option *client no.5*: statistical tests result of thermosensitive check

Then, the thermosensitive client's load data (one entire year) are divided chronologically by the CUSUM algorithm. The pair of dotted lines indicates the maximum and minimum values in the CUSUM plot (Fig. 6.5). They are interpreted as the breaking points of the heating season and the non-heating season (Fig. 6.6).

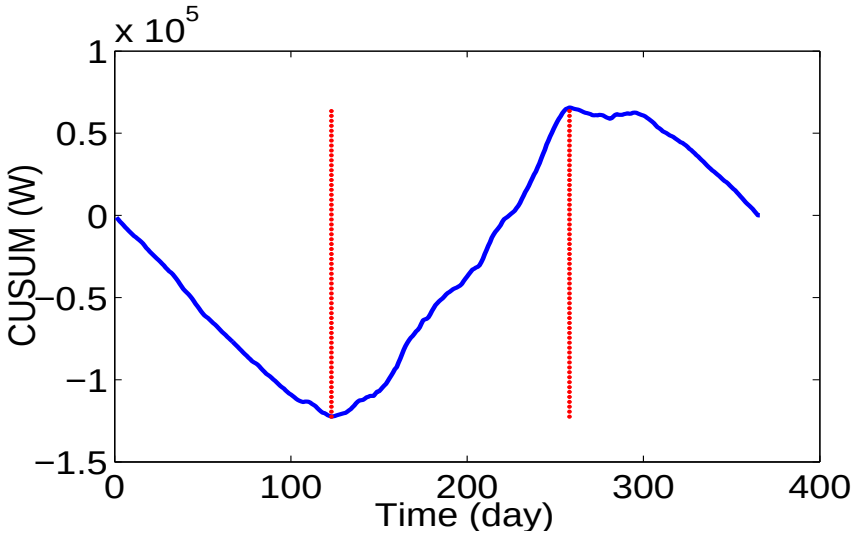


Figure 6.5: Off-peak/on-peak option *client no.5*: CUSUM chart of daily average power

On the one hand, the minimum power load pattern can be estimated once the heating

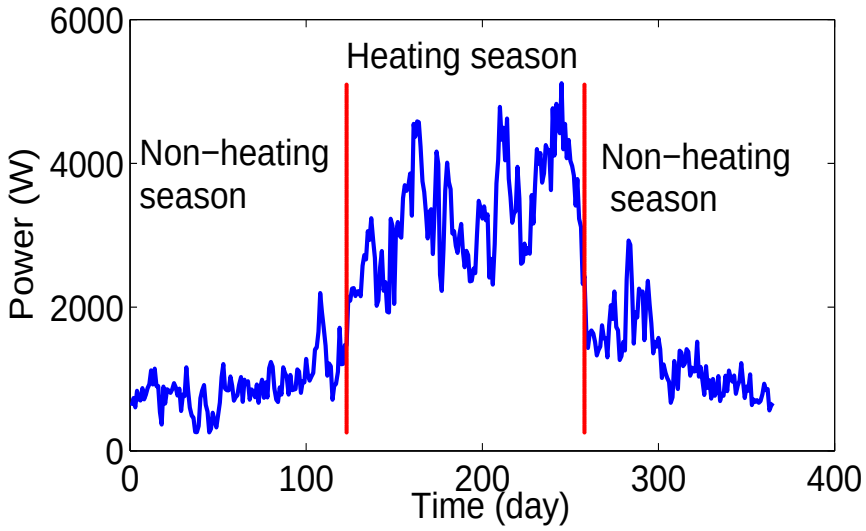


Figure 6.6: Off-peak/on-peak option *client no.5*: separation result of one year's power data by CUSUM algorithm

season and the non-heating season are separated. As the consumption during this period is considered independent of temperature, all days' loads are overlapped and the median [1] value of a specific hour and its 10% excess probability upper and lower bounds are estimated by KDE [142]. The median value rather than the mean value is calculated because of the former's statistical robustness to the large dispersion samples. Similar to the BAGHEERA model, three categories of different day types are made, namely, weekday, Saturday, and Sunday and holiday. Moreover, "Linky" in the future will collect load samples averaged every 30 minutes [3], and the estimation will also be made in 30-minute intervals. Figure 6.7 shows the result of the minimum power estimation on weekdays.

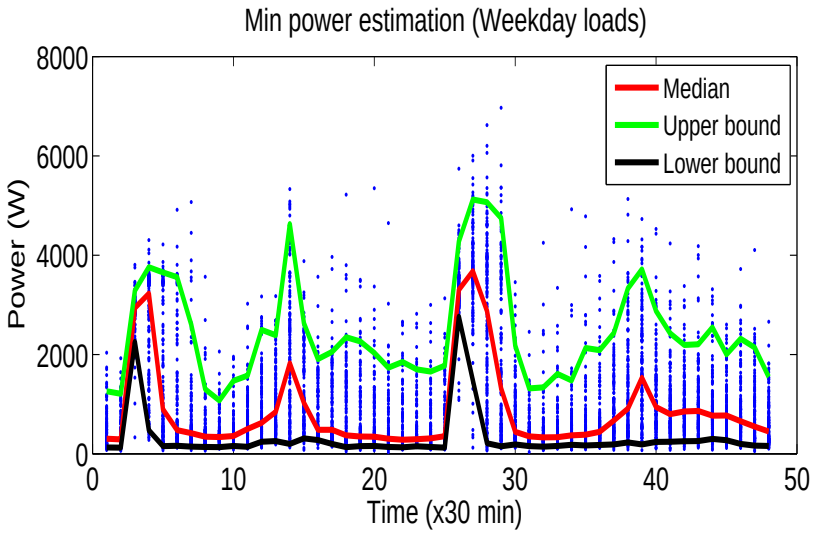


Figure 6.7: Off-peak/on-peak option *client no.5*: weekday minimum power estimations. The points represent the real measurements.

[1] Sorting data into increasing order, median is the middle entry or the mean of the two middle entries.

On the other hand, the estimation for TMB power or rather maximum power of the year involves three steps. First, statistical tests are performed between the group of load samples inferior to 0^oC of the examined year and those of the past years to check whether the consumption behavior of the client under cold weather conditions stays unchanged. In other words, whether the information collected during extreme conditions in past years can be used for the estimation of the current year. Figure 6.8 shows the statistical test results and suggests that the data of past year have the similar characteristics (mean and variance values) to those of the current year. As a result, the past year's inferior 0^oC data are joined.

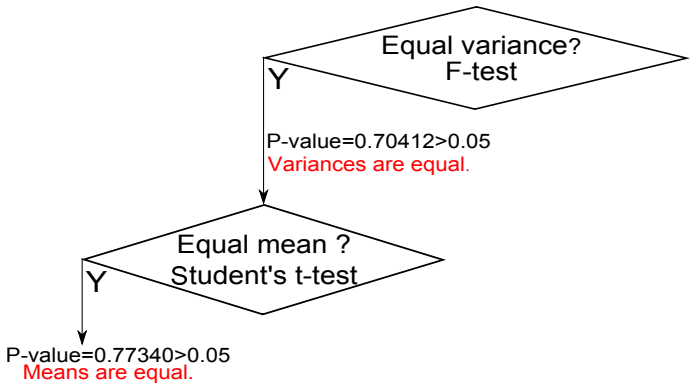


Figure 6.8: Off-peak/on-peak option *client no.5*: statistical tests result for the data coherence check

Second, the kernel regression method is applied in order to define the nonparametric relationship between power consumption and temperature. Before this is done, the smoothing parameter has to be selected. The CV function in equation 6.10 is applied with 10 parts. Figure 6.9 illustrates the Mean Square Error (MSE) variation of the different smoothing parameters. We choose $h_{cv} = 0.9$, which corresponds to the minimum MSE.

We propose herein three kernel-type estimators, referred to as: NW, LL, and LL2. These estimators are computed to deduce the relationship shown in figure 6.10.

Notice that the NW and the LL curves are rather close to each other except near the "boundary zone". The "boundary zone" is defined as the region where a very limited number of samples are taken. In fact, the nonparametric estimators are very sensitive to these samples.

Third, all the heating season data y_i at temperature X_i are brought to the TMB condition with their uncertainty level maintained. The estimated power at TMB condition y_{TMB_i} can be defined as

$$y_{TMB\_i} = \hat{f}_{h'}(TMB) - \hat{f}_{h'}(X_i) + y_i \tag{6.11}$$

where $\hat{f}_{h'}(\cdot)$ is the kernel-type estimator with its optimal smoothing parameter h'.

Figure 6.11 explains the uncertainty level, which represents the difference between the estimated power and the real measurement, given a NW estimator as an example. Figure 6.12 shows the maximum power in the TMB condition estimated by the kernel density method.

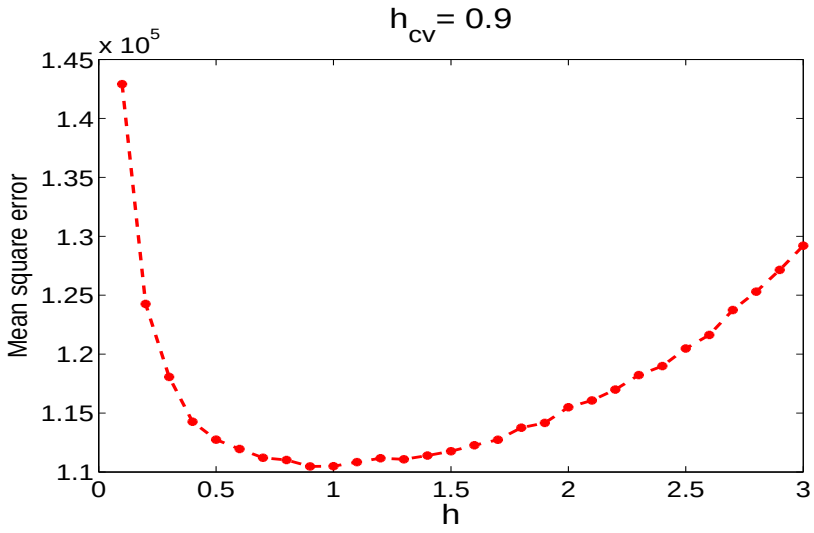


Figure 6.9: Off-peak/on-peak option *client no.5*: cross-validation result on the smoothing parameter selection of the kernel estimation

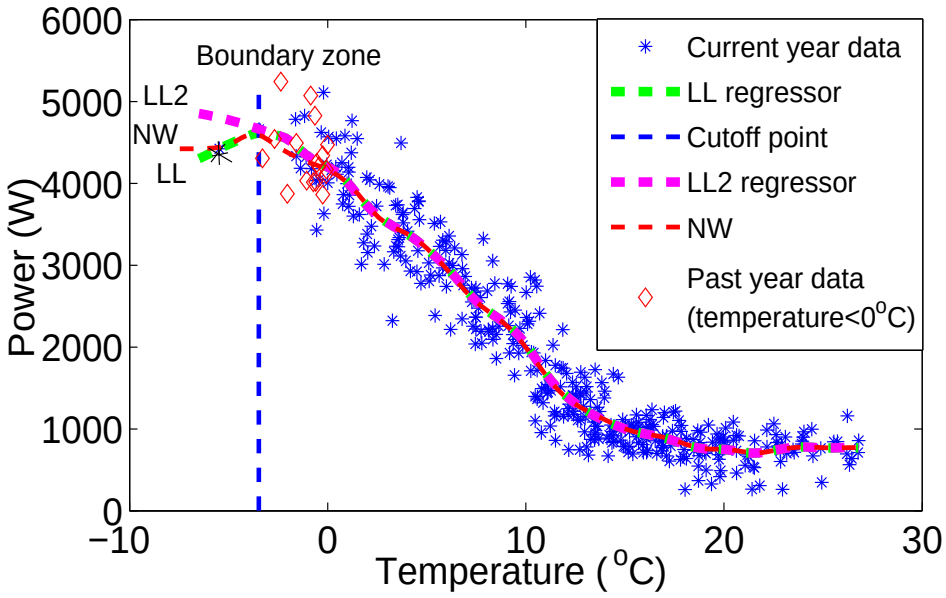


Figure 6.10: NW, LL, and LL2 regressors, indicating the relationship between the variation of temperature and the client's daily power consumption

6.2.b Comparison with the BAGHEERA model

In this subsection, we compare the proposed nonparametric models with the BAGHEERA model. Since the objective is to estimate the power consumed in the TMB condition, where in our example there is no measurement, the performance criteria are hard to establish. However, the accuracy of the model can be computed in cold weather conditions instead of the TMB condition. We believe that if the power is more accurately estimated in the neighborhood of the TMB temperature by one model, this model could possibly also provide a more credible estimation in the TMB condition.

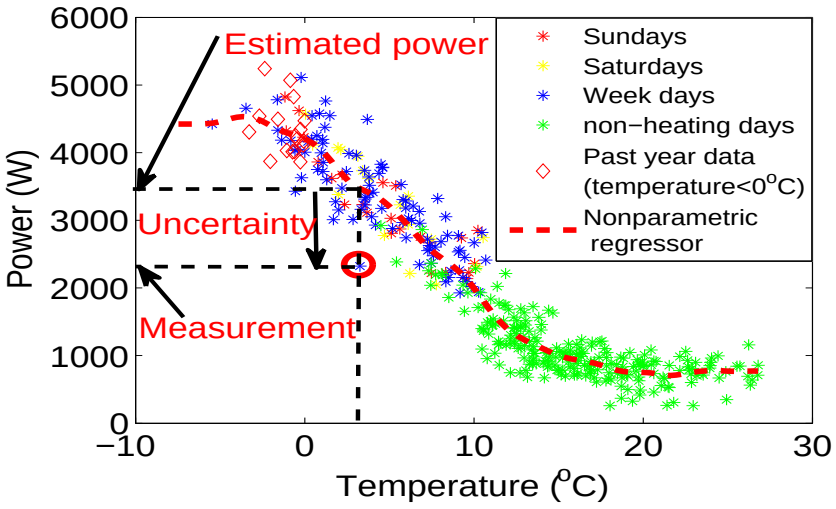


Figure 6.11: Off-peak/on-peak option *client no.5*: presentation of uncertainty of the red circled sample. The dotted red curve represents the kernel estimation result of the relationship between power consumption and temperature.

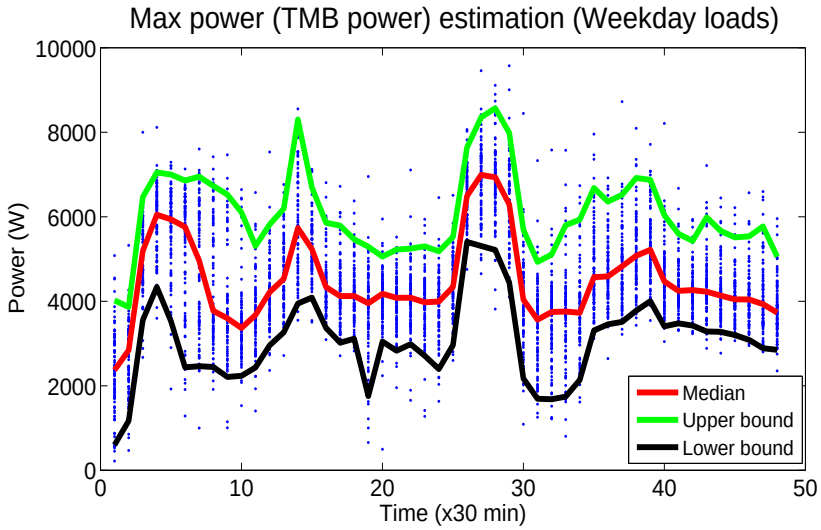


Figure 6.12: Off-peak/on-peak option *client no.5*: maximum power estimation of weekday loads. The points represent the heating season data brought to the TMB condition.

More specifically, the first-year data are used to establish the models respectively. The data of the second year, when the temperature is inferior to 0^oC, serve as the test data. This latter correspond to seventeen days, situated on the right side of the cutoff line as illustrated in figure 6.10. As previously explained, the second-year data are used as the test data but not as the past year data, thus they do not participate to compute the kernel regression.

The results of the Sum Square Error (SSE) shown in the figure 6.13 are calculated on a one-hour data base.

The results indicate that the three nonparametric estimators (NW, LL and LL2) give a better estimation than BAGHEERA does on the test data. There are three plausible

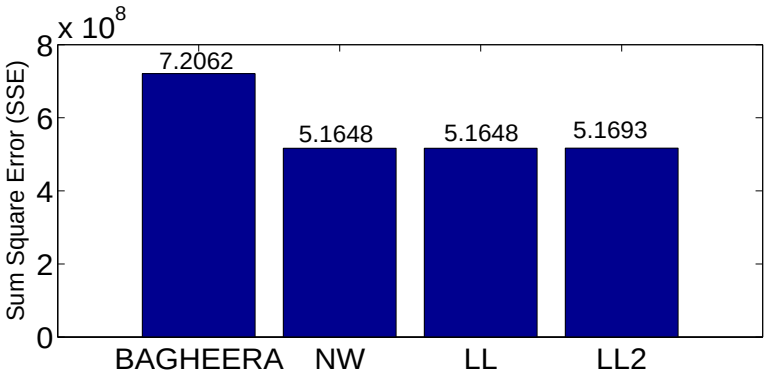


Figure 6.13: SSEs of the BAGHEERA estimator, NW, LL, and LL2 estimators on the test data

explanations:

- one, the hypothesis on which BAGHEERA is based, that the relationship between temperature and power consumption during cold weather is linear, is not verified. Our analysis according to figure 6.11 argues that below a certain temperature, the power consumption saturates, which is closer to the reality. Since very cold temperatures urge users to exploit their heating devices at maximum, assuming a linear relationship between the power and temperature variation does not seem very appropriate.
- Two, the BAGHEERA model calculates its coefficients with a group of clients. Hence, its estimation on one individual client may possibly be less accurate than a data-driven individual estimation model.
- Three, through our 70 examples, we find several clients misplaced in the groups with the qualitative information: the observations in the data base show that the qualitative information is not a satisfactory criterion for group discrimination. As a matter of fact, relying on qualitative information to assign clients to different groups can lead to assignment errors that impact the validity of the BAGHEERA model.

The difference in SSE between the NW and the LL estimators is fairly small with this example, as only one-year cold-temperature data were used as the test data. In other words, the LL estimator presents many advantages with respect to the NW estimator. Nevertheless, the improvements with LL fit do not necessarily produce a better fitness load function in our case. We believe that since near the TMB temperature the power function is rather flat, there is no necessity to try local polynomial regressions of higher degree.

An adapted estimator, LL2, is also proposed. It isolates the area where scarce samples are collected. As it continues the local trend drawn by the "credible data", in our example it gives a slightly higher estimation than the other two nonparametric estimators. We believe that with enough data, especially with the arrival of smart meters, the NW and the LL estimators can give accurate estimations in the TMB condition. On the other hand, if we do not have sufficient data, particularly near the TMB region, it is better to adapt the LL2 estimator, which gives more reliable estimation values for distribution network planning.

Since our model is individualized and completely data-driven, compared with the BAGHEERA model and most of the pre-defined group methodologies, the proposed method-

ology is more flexible in facing new types of consumption patterns. As a matter of fact, pre-processing steps such as discrimination of clients in groups are omitted, leading to a reduced computational complexity. Since the computational time is short, the model can be frequently updated. Thus, compared with the BAGHEERA model, whose coefficients are updated every one or two years, the accuracy of the proposed estimation model is guaranteed.

6.3 Validation study

In this section, we validate our methodology by exploiting all the clients' data. Two study cases are set up in order to check the accuracy of the nonparametric estimators for the "median" models. As explained, we have two-year recording load data of 35 off-peak/on peak clients, in these studies, the first-year data of the client are used to build up models, while the second-year data are used as the test data.

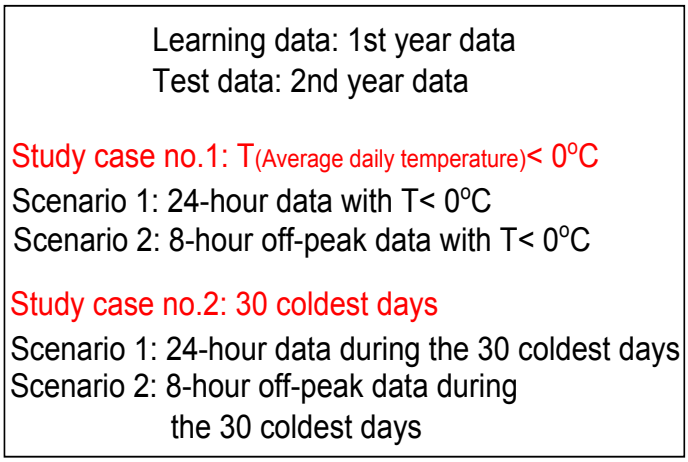


Figure 6.14: Study cases and scenarios in the validation study

As explained in the subsection 6.2.b, we aim at validating the models in the neighborhood conditions of the TMB temperature, for the reason of scarce samples collected in the TMB condition. Figure 6.14 summarizes the designed study cases and scenarios in the validation study. The first study case is an extension of the example introduced in subsection 6.2.b, i.e., the test data being the second-year data when the temperature is inferior to zero degree. Two scenarios are compared: during the whole test period and during the off-peak test period. The off-peak period is interesting for the validation study, since the peak loads mostly happen during this period.

Figures 6.15 and 6.16 show the comparison results of the two scenarios. Only the clients who have the coherent data during the two years are included in the study, i.e., the first year inferior to zero degree data and that of the second year having the same mean and standard deviation values proved by the statistical tests. In both scenarios, most of cases, the nonparametric estimators have a better precision than the BAGHEERA model, committing less SSE.

As a matter of fact, regarding the first study case, the number of days in the test set depends on the temperature variation of the test year. For instance, sometimes, only two

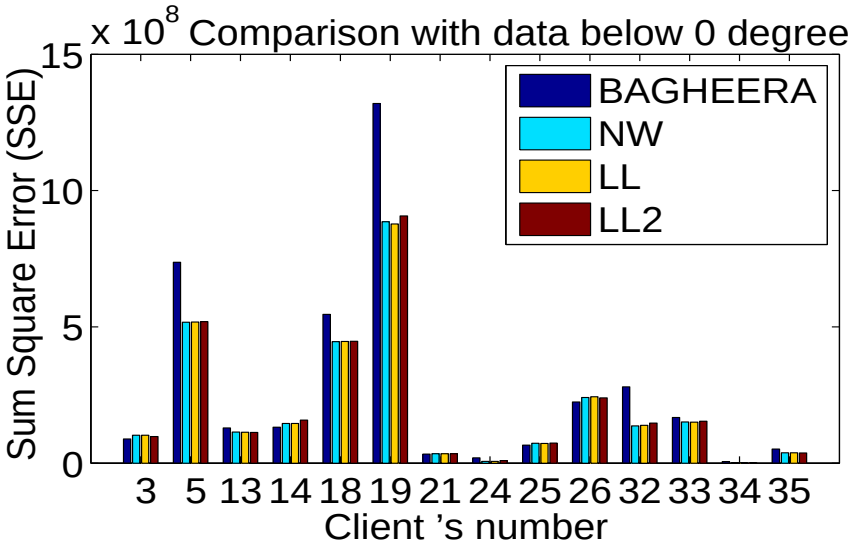


Figure 6.15: Study case no.1, scenario 1: off-peak/on-peak option clients, comparison of SSEs of BAGHEERA, NW, LL and LL2 estimators on the days below 0 degree during the second year

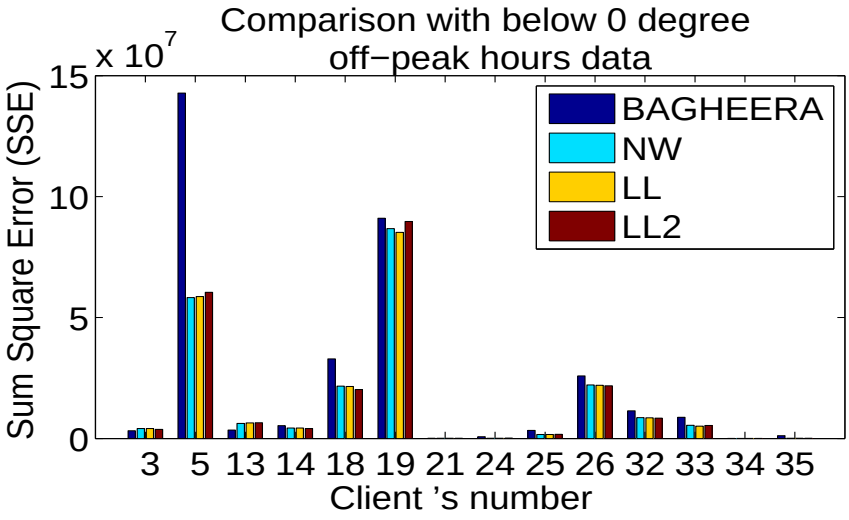


Figure 6.16: Study case no.1, scenario 2: off-peak/on-peak option clients, comparison of SSEs of BAGHEERA, NW, LL and LL2 estimators on the off-peak hours of the days below 0 degree during the second year

days in the second year are inferior to zero degree. As a result, in such cases, we compared the estimators based on very few data, and the compare results are far to be pertinent. Therefore, in the second study case, we compare the estimators on a 30 day database, where the coldest temperatures occurred. Same as the first study case, only the clients having the coherent data during the 30 coldest days are included. Figures 6.17 and 6.18 show the results on the whole test period and off-peak test period scenarios. The results indicate that the nonparametric estimators have a better precision.

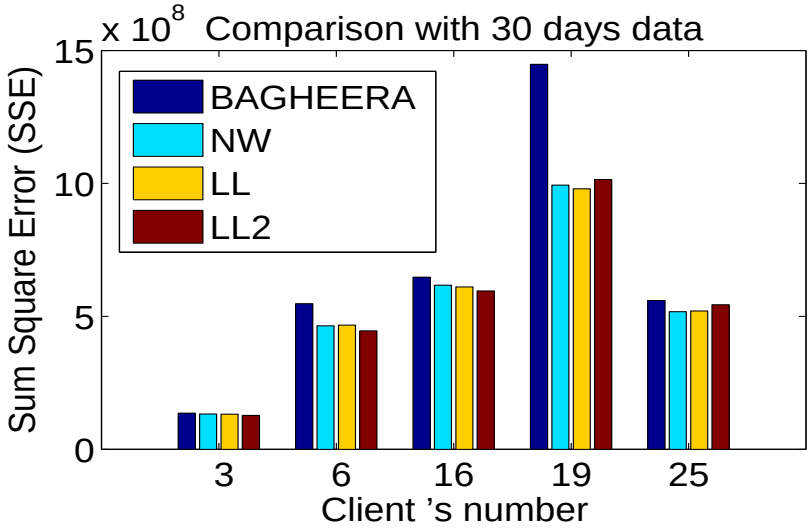


Figure 6.17: Study case no.2, scenario 1: off-peak/on-peak option clients, comparison of SSEs of BAGHEERA, NW, LL and LL2 estimators on the 30 coldest days of the second-year data

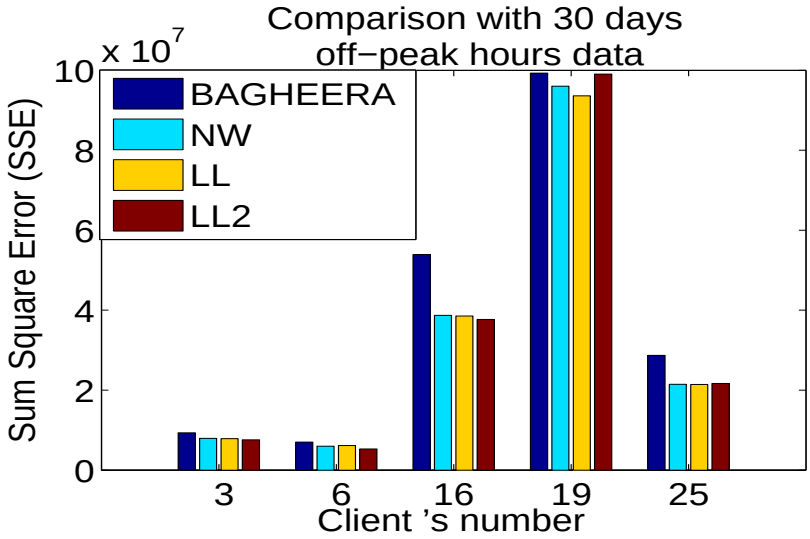


Figure 6.18: Study case no.2, scenario 2: off-peak/on-peak option clients, comparison of SSEs of BAGHEERA, NW, LL and LL2 estimators on the off-peak hours of the 30 coldest days of the second-year data

6.4 Discussion

In this section, we open a discussion on the definition of the upper bound for planning needs in distribution networks. First, we present some citations from the internal memo reports of the French electricity company EDF. Next, we focus on the calculation of the upper bound in our proposed method. In the end, we give some clues for the validation study for both our proposed methods and the BAGHEERA model. The first trial results are shown. We hope that this discussion can help readers getting a better understanding on the definition and the utilization of the upper-bound estimations. At the same time,

we hope to attract the attention of the load estimation model designers and the network planners to the validity of the models.

6.4.a Citations of the upper-bound definitions in EDF reports

We have found the following citations in the internal memo reports of the French electricity company EDF:

- [133]: "the new (BAGHEERA) model will provide the 10-minute maximum power having a risk of γ being exceeded, for a given temperature. Furthermore, it will give, for each network section and for a given temperature, one mean value estimation and one variance estimation of the 24 powers averaged from 10-minute data for every hour for the concerned group of clients." [2]

- [132]: "thus, the network planer accepts to take a certain risk γ, an arbitrary choice. The model results a power level P_γ, which has the γ probability being exceeded." "The load model produces 48 powers required by each load point, that the electric calculation takes into account for the calculation of 48 voltage-drops and deduce the maximum value. Thus, the assumption of the power synchronization can be canceled." [3]

- In [148], the risk (also called margin) calculation in the BAGHEERA model is stated in details.

  The objective of the BAGHEERA model indicated in [148] is "to estimate the risk of the actual power required by a group of clients exceeding the maximum power." "The quality of the network planning is measured by the maximum voltage-drop on the MV feeder among 48 voltage-drops." "The idea is to define a planning power or a power for a group of clients, which must not be exceeded with a risk inferior to or equal to 10%." [4]

  [148] shows the mathematical demonstration of the risk γ, and argues that:

$$\forall T \geq TMB, \quad \gamma = p(P > P_{threshold}) \leq \{\alpha + (1-\alpha)p(P > P_{threshold} | T = TMB)\} \quad (6.12)$$

  where $\alpha = p(T < TMB) = 0.003$. Since $\alpha \ll 1$, we have $\gamma = p(P > P_{threshold}) \approx p(P > P_{threshold} | T = TMB)$, which signifies that the probability of the powers P excess the

[2] Original sentences in French: "Le nouveau modèle fournira, pour une température donnée, la puissance 10 minutes maximale ayant un risque γ d'être dépassée. En outre, il donnera pour chaque tronçon de réseau et pour une température donnée, une estimation de la moyenne et de la variance des 24 puissances moyennes 10 minutes appelées heure par heure par l'ensemble des clients concernés."

[3] Original sentences in French: "Le planificateur acceptant donc de prendre un certain risque γ, résultat de son arbitrage, le modèle lui donne un niveau de puissance P_γ qui a cette probabilité γ d'être dépassée." "Le modèle de charge produit 48 puissances appelées par point de charge, que le calcul électrique prend en compte pour calculer 48 chutes de tension et en déduire la valeur maximale. Il n'y a plus d'hypothèses sur le synchronisme des puissances."

[4] Original sentences in French: "Contexte: estimer le rique de dépassement de la puissance maximale appelée par un groupe de clients." "la qualité est appréciée par la chute de tension maximale sur le départ, maximum des 48 chutes de tension horaires" "L'arbitrage se fait en définissant ≪ une puissance de planification ≫ ou puissance d'un groupe de clients qui ne peut être dépassée qu'avec un risque inférieur ou égal à 10%."

power threshold $P_{threshold}$ equals to the probability of the powers excess the power threshold under the TMB temperature, as long as attaining the TMB temperature being a rare occurrence. In other words, the probability of going beyond a certain threshold power in all temperature conditions is equivalent to the probability of going beyond the same threshold power in the TMB condition, as the probability of occurrence below the TMB temperature is near zero.

Next, in this report [148], the hourly model is built, they assume that for a given temperature, a given day type and a same tariff period (off-peak or on-peak period), the average 10-minute power require by the client in a same category follows a same distribution.

Since the distribution followed by one particular client's power is hard to be identified as a known distribution, a group of clients' power in a given category is written as a linear model:

$$P_c(t) = m_c(t) + k_c^{\gamma}\sigma_c(t) \tag{6.13}$$

where $m_c(t)$ is the mean value of the power of the group of clients at time "t", $\sigma_c(t)$ is the standard deviation of the power. k_c^{γ} is the coefficient different from category to category. Its value is estimated by bootstrapping method.

As a matter of fact, from the above citations, we can conclude that the required upper bound is actually composed by hourly values and the assumption that the power at the same hour of the clients in the same category, at the same temperature, and for the same day type, follows a same distribution is made in the BAGHEERA model. Thus, a 10% risk is actually taken for every time step. In order to comply with the voltage-drop calculation presented in the section 5.3, detailed estimated power thresholds at every time step are required, taking coincident effect into consideration (subsection 5.1.a). The definition of the upper bound in the reports of EDF, in our opinion, should highlight that the 10% (γ) risk is taken for every time step.

6.4.b Upper bound in the nonparametric models

The upper-bound estimation of our method is inspired by the BAGHEERA model. The 10% threshold power is calculated for every time step. After bring all the heating period samples to the TMB condition (figure 6.11), we assume that the hourly power consumption of a client on a given day type follows a same distribution. Figure 6.19 shows the samples, which are extrapolated to the TMB condition, being overlapped. The kernel density estimation (subsection 6.1.d) is applied in order to calculate the 10% border for each time step [5]. In figure 6.19, $P_{median}^{(m)}, m = 1, \cdots, 48$ stands for the estimated median value curve, and $P_{threshold}^{(m)}, m = 1, \cdots, 48$ stands for the estimated 10% upper-bound curve.

[5] In order to be coherent with the Linky's sampling time, the time step herein is equal to 30 minutes, whereas the output of the BAGHEERA model is on an hourly basis.

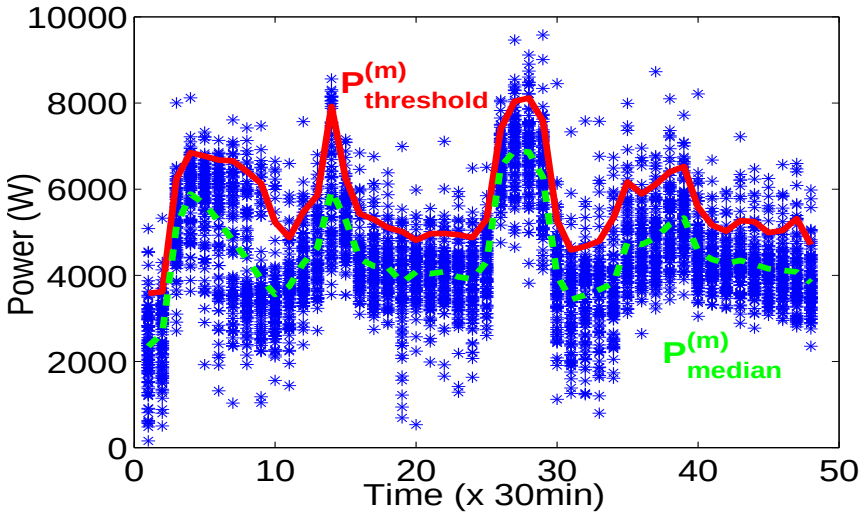


Figure 6.19: 10% hourly power excess probability threshold and median value for every time step (30 min/step) (e.g.: the Off-peak/on-peak option *client no.5*). Asterisks: estimated powers under TMB condition. Solid curve: estimated threshold powers, dotted curve: estimated median powers.

6.4.c Validation trial on the upper-bound estimation

Same to the "median" models (section 6.3), we try to analyze the validity of the upper-bound models by comparing some critical values in the measurements, with the upper bounds estimated by the BAGHEERA model and the nonparametric models. The first-year data is used as the learning data, building models, and the second-year data is used as the test set. The critical values we choose are the "2.5h data" and the "0.03% data".

As explained in subsection 5.3.b-i, theoretically, the power threshold that we define is in average exceeded during 2.5 hours, or 0.03% of a year's data. The test data, second-year 30-minute data are arranged in decreasing order, and the 5th largest is assigned as the "2.5h data". The "0.03%" data is correspondent to the value calculated by the kernel density estimation on the second-year data. In figure 6.20, the BAGHEERA, the NW, the LL, and the LL2, respectively correspond to the most important values in the upper-bound estimations with the first-year data.

The ideal case is that the four estimated upper bounds (obtained by the BAGHEERA, the NW, the LL, and the LL2) are close to the critical measurement values. This means that the estimations of the upper bound based on the first-year data are compatible with the second-year data. However, in figure 6.20, it is not often the case. We can find some plausible explanations in their plots of power data.

Figure 6.21 plots the average daily power consumption of client no.22 during two years. We can see from figure 6.20 that for client no.22, the measurement critical values of the second-year data are higher than output of any estimation model based on the first-year data. As stated in the first two subsections, the daily upper-bound estimation is composed by 24 (or 48 in models with 30-minute step time) mean average estimations and 24 (or 48 in models with 30-minute step time) risk (margin) estimations. The mean average estimations are mainly dependent on the daily average, and the risk (margin) estimations

Figure 6.20: Summary of the upper-bound comparison of the real measurements, the BAGHEERA model, and nonparametric models

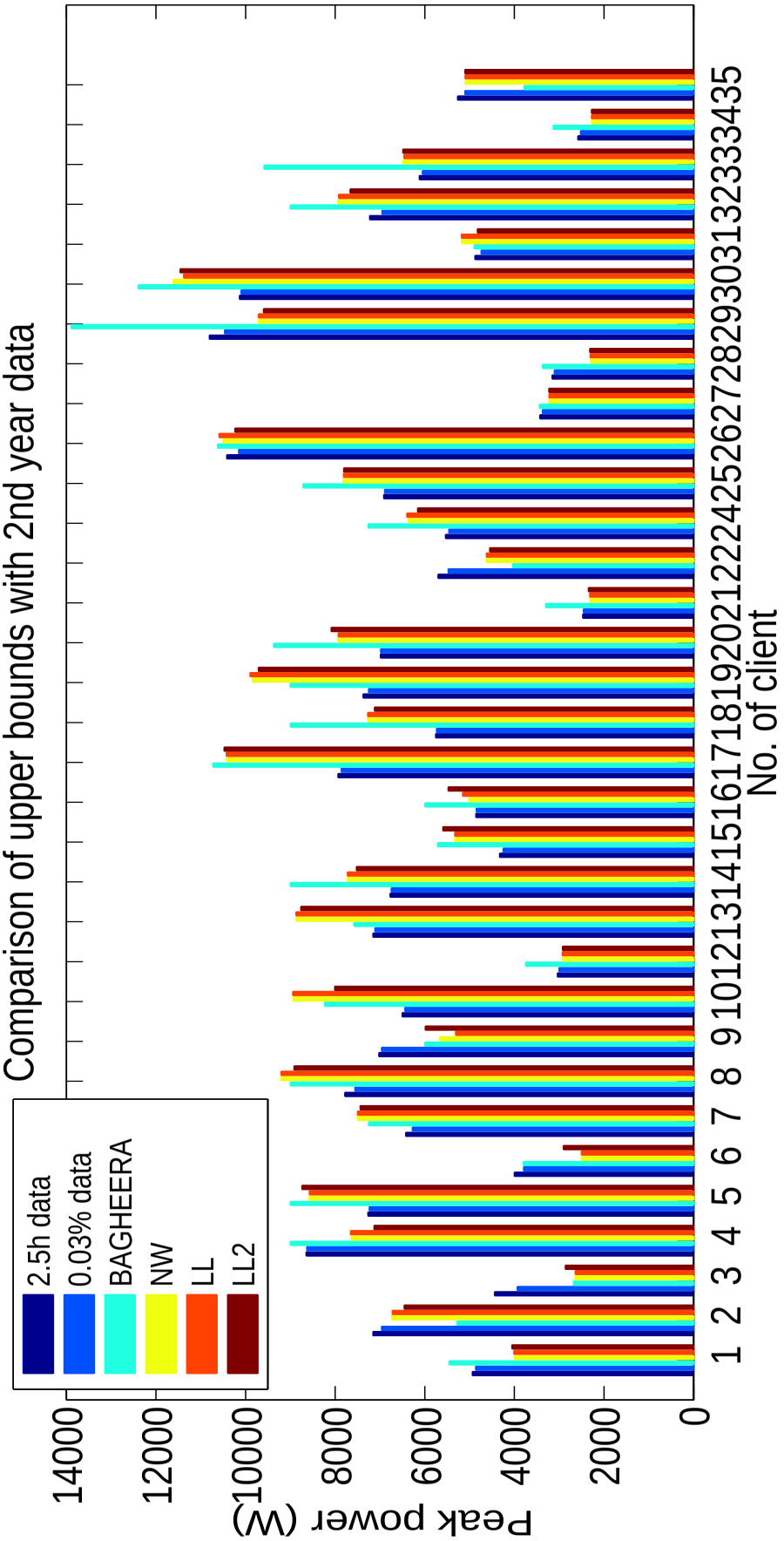
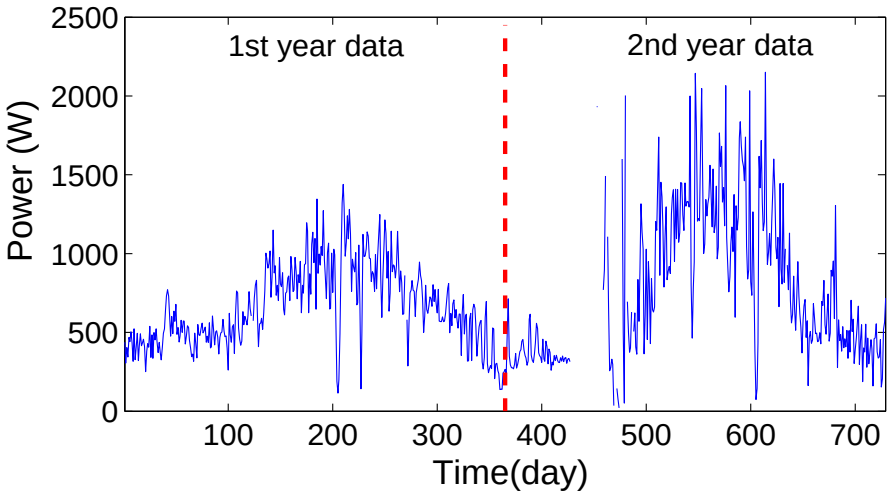


Figure 6.21: Power consumption of *client no.22* during two years (July 01, 2004 ~ June 30, 2006). The vertical line separates the two-year period.



are in close relationship with the dispersion (uncertainty) on every time-step. Thus, it is easy to understand that because of the incompatibility in the data, more specially, the consumption level of the first year being much lower than the second year, the estimation of the upper bound based on the first-year data cannot cover the measurement critical values collected in the second year.

Let the client no.17 be another example. We can see in figure 6.20 that the measurement critical values collected in the second year are smaller than the upper-bound estimations provided by the 4 estimators. Figure 6.22 shows daily average power plot of client no.17 during two successive years. The first-year consumption level is slightly higher than the second year. Figure 6.23 illustrates the standard deviation values on every 30-minute. The standard deviation value Sd_m on time step m is defined such that:

$$Sd_m = \sqrt{\frac{1}{K}\sum_{i=1}^{K}(X_{i,m} - \overline{X}_i)^2} \tag{6.14}$$

where $X_{i,m}$ is a measurement collected on the i-th day, m-th time step, K is the total number of days participating to the calculation, and $\overline{X}_i$ is the i-th daily average.

We can see in the figure 6.23 that the first-year standard deviations are also slightly higher than those of the second year. Thus, both the higher daily average values and the higher standard deviations result higher estimated values than the measurement critical values taken in the second year.

Of course, here we presented the simplified explanations for the examples. A more precise explanation takes into account numerous factors applied in the constitution of the upper-bound estimation, such as the correlation between the daily average power and the daily temperature, and the distribution followed by samples on each time step, etc. In this subsection, we aim at giving an idea for the validation of the upper-bound estimation. Even though that the BAGHEERA model has been put into service for more than 20 years, and that we can nearly say that the validity of the model is approved in practice,

Figure 6.22: Power consumption of *client no.17* during two years (July 01, 2004 ~ June 30, 2006). The vertical line separates the two-year period.

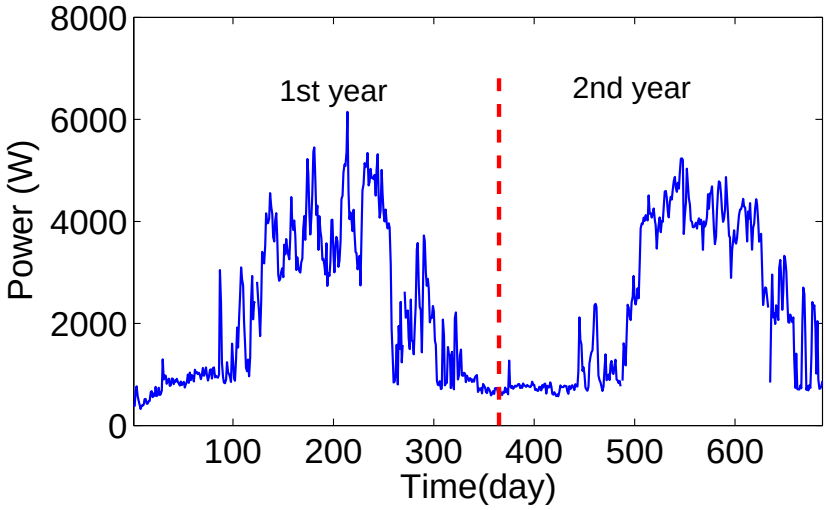
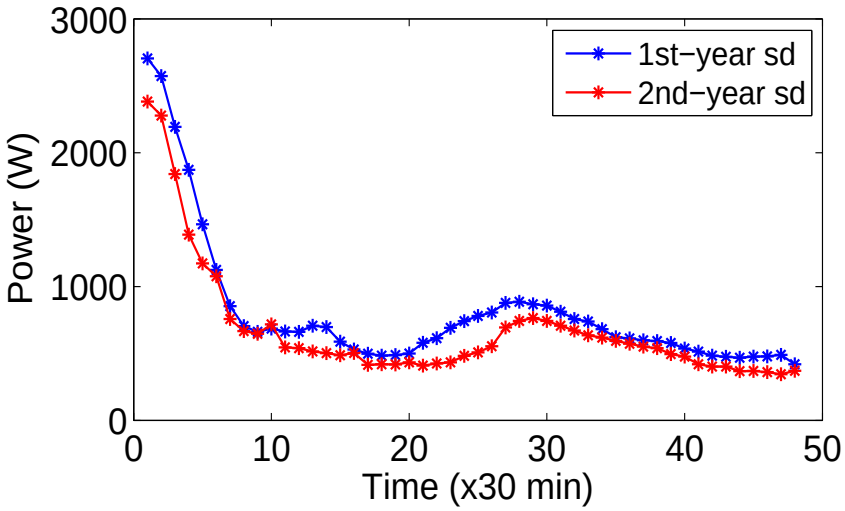


Figure 6.23: 30-minute time step standard deviation(sd) of *client No.17*



we believe that for the optimal balance of the network security and the economic benefice, the question of the validation of the estimation model output should be raised.

6.5 Conclusion and perspective

This chapter presents a novel individual load estimation model based on nonparametric methods for power distribution network planning. Making no hypothesis on the load function and being completely data-driven, the method is more objective in providing accurate maximum and minimum individual load estimations. Three different nonparametric estimators are proposed, and the choice of estimator under different data conditions is discussed. The nonparametric models are compared with the current industrial model BAGHEERA on the test data. Validation study cases are designed and carried out for the comparison of precisions of different estimators. We argue that our method is more adapted and gives more reliable estimations. Discussion dedicated to the upper-bound estimations is brought up.

As presented in the subsection 5.3.b-i, the threshold power has a 0.03% (0.3% attaining TMB temperature and taking a 10% excess probability at TMB temperature) probability being exceeded by the measurements. Therefore, without extrapolating data to the TMB temperature, the problem can be converted into the calculation of an extremal event, which includes finding extreme values whose probability of occurrence is more extreme than any sampled observation. *Extreme value distribution* can be determined based on sound mathematical principals, such as the Generalized Extreme Value (GEV) distribution [149, 150], Generalized Pareto Distribution (GPD) [151] and Quantile based methods (Hill's estimator [152], Pickands estimator [153], Moment estimator [154], to name a few). Comparison then is to be made among threshold powers obtained by different methods.

For further industrial implementation, however, more considerations need to be taken into account, such as the feasibility of 35 million individual models on French territory and the frequency to update these models, among others.

Chapter 7

General conclusion and perspective

7.1 Conclusion

In this report, we tackled two distinct load model problems in distribution networks for operation need and planning need. Recording detailed individual loads, the smart meters in smart-grid context enable the design of the accurate forecast models on the LV level, and the individual load estimation models.

Thus, after a first general introduction in chapter 1, we divided the report into two parts to answer respectively the two needs: chapters 2, 3, and 4 contribute to the short-term load forecasting issue in the need of network operation; chapters 5 and 6 deal with the load estimation model in the need of network planning. Three models are developed, and the validity of the models is tested with real measurements collected in the French distribution networks.

In the first part, we selected the time series method from the classical approach family and the neural network method from the AI approach family after a thorough review of the load forecasting literatures in chapter 2. These two methods are respectively depicted in chapters 3 and 4.

- The time series method divides the data into three parts: the trend, the cyclic components, and the random errors. The first two deterministic parts are designed into parametric models. The trend model takes into account the temperature, the time, and the day types. The cyclic model is composed of Fourrier components, whose frequencies are found by the smoothed periodogram technique. The residual analysis shows that there is hardly any possible amelioration on the forecasting results with other classical approach.

  The weather uncertainty results a decrease in precision in the load forecasting model. We showed that despite of the weather uncertainty, our proposed time series model still largely outperforms the reference model: the naive model.

- Having a universal approximation capacity and a great mapping capacity in dealing with complex nonlinear relationship, neural network can provide superior forecasting performance to classical methods [99]. In most of works dedicating in applying neural networks to the short-term load forecast, it is declared that there is no consistent methodology for variable selection for neural network model [106] and it is often observed that choices of the model are not systematically justified and the results are not well commented or not presented in a clarified way [39, 28].

  We focus on the methodology of model design, i.e., variable and model selections, for the neural network model. The variable selection is based on the orthogonal forward regression, ranking the variable candidates in the order of the pertinence

to the predicted load. The model selection is based on the VLOO process, which estimates the generalization capacity of the model. The load curve is decomposed into the daily average power and the intraday power variation parts, in order to reduce the complexity of the load model. The efficiency of the selection strategy is proven with the real MV/LV substation loads.

In the second part, we tackled the individual load estimation model. After a review of the load models applied in some countries, and a detailed presentation of the load estimation model of the French DSO in chapter 5. We introduced the idea of building the individualized load estimation model.

- A completely data-driven nonparametric model, making no prior hypothesis to the relationship between load and variation of the temperature, is proposed in chapter 6. Three different nonparametric regressors are designed, for the different data conditions. Validation study by comparing the estimation precision of the BAGHEERA and the nonparametric models on the same data shows that the nonparametric estimators have a better precision than the BAGHEERA model. A discussion of the definition of the uncertainty threshold for the voltage-drop calculation is developed.

These three load models answer the total requests of the objectives described in chapter 1.

7.2 Perspective

- The proposed short-term load forecasting models have a better precision than the naive model. However, in terms of efficiency and benefit brought to the ADA functions, they are not yet quantified. Thus, further study of integrating the designed load forecasting models to the ADA functions can be conceived.

- The proposed neural network design methodology seems a very promising way for the short-term load forecast, providing a more accurate load forecast than the time series method. It can be extended to all types of load forecast, for instance, the reactive power. This first study is quite satisfactory. However, some research efforts are still required for the following reasons:

  In the dissertation, we have presented using the orthogonal forward regression for the variable selection of neural network model. I. Drezga and S. Rahman commented in their paper [106] that *there is no consistent methodology for determining relevant variables for ANN based STLF*. They explained that the neural network models are designed for modeling nonlinear complex functions, whereas the input variables are for most of the time selected based on the linear or the local-linear correlation coefficients.

  In our methodology, the secondary candidate variables are conceived in order to take into account the nonlinearity of the model. The secondary candidate variable represents only the product effect between two primary variables. Nevertheless, other nonlinear combination forms are ignored.

Moreover, the risk of selecting one non effective variable is estimated by the generation of the probe variables. Due to the randomly shuffling effect, the number of retained variables with the same risk is varied in different trials.

Other input variable design methodologies, such as the entropy analysis [115] and phase-space embedding technique [106] could open some possible ways to improve the efficiency of the model.

Wrapper presented in the subsection 4.3.a, on the other hand, being very computational costly, does not seem to perform well in our first trials.

We have observed a declining precision in the forecasting results in Appendix D over time. As a non-stationary process, the forecasting model is suffered from slow changes, because of the gradual load growths [20], etc. Therefore, the update frequency for the model needs to be figured out in order to get the optimal performance.

- In subsection 6.3, we have shown that the nonparametric model has a better precision compared to the BAGHEERA model on the mean value estimations. A discussion was held, aiming to clarify the definition of the upper bound, whose value(s) directly participate in the voltage-drop calculation. This discussion should be continued with engineers in the DSO. The validation of the methodology for the upper bound estimation is also a difficult task. Since few samples are recorded under the TMB condition during the two-year period [3], the 10% excess probability is hard to validate with the real measurements. The proposed methodology needs to gain the validation from the industrial implementation.

Today, large amount of renewable energy integrate to the electrical network. Their occurrences change the transits of the electrical lines. They can be considered as negative electricity consumptions to the network.

Different from the traditional power generations (nuclear, fuel, and hydrogenation, etc.), these distributed renewable energies (mainly from wind, photovoltaic, and biomass) so far connected to the electricity distribution systems are intermittent, very dependent to the weather conditions and can change instantly [155]. The accuracy of the forecast for these power generations depend more on the quality of the forecasted influence factors than on the efficiency of the predictive algorithm.

The influence factors for the photovoltaic energy for instance depend directly on the cloud covering, the solar radiation, and the temperature [156]. Other parameters such as the wind, the rain and the snow indirectly impact the photovoltaic energy generation [155]. The great challenge is the accurate prediction on the cloud covering based on some complex physical process, since it is much more difficult to predict than the mean daily temperature. The wind power generation on the other hand, depend mainly the wind force, the direction, geography situation and the type of the wind turbine. The prediction of the influence parameters is beyond scope of this dissertation.

In order to correctly carry out the voltage-drop calculation for the operation need and for the planning need, the renewable power generations need to be predicted. Two ways are possible:

- a predictive model dedicated to the power generation

- a model that predicts the aggregated value of the generation and consumption

Most of researches concentrate on the first track. There are mainly three categories of models for the power generation predictive model: based on a physical model, on a statistical model or on a hybrid model that combines the advantages of the two previous ones [155]. The methods are similar to that of the load forecasting [155, 157, 158].

We suggest two models developed in this dissertation, the neural network model and the nonparametric model to reply the demand in the second way. Because of the good learning ability of the neural network model, if the influence factors are properly handled, it is capable to map the complex nonlinear relationship between inputs and outputs. In this case, the output of the model would be the forecast of the difference between power production and electricity consumption. The variable and the model selection strategies (sections 4.3.a and 4.3.b) can be easily implemented. Regarding the nonparametric model, making no priori hypothesis on the load function and being completely data driven, can be easily applied to the situation with the integration of the renewable power generations.

For future work, real consumption and power generation data should be tested on these two models.

Bibliography

[1] M. Biserica, B. Berseneff, Y. Bésanger, and C. Kieny. Upgraded coordinated voltage control for distribution systems. In *PowerTech Conference Proceedings, IEEE Trondheim, Norway*, 2011.

[2] R. Podmore and M.R. Robinson. The role of simulators for smart grid development. *IEEE Transactions on Industry Applications*, Volume 1, September 2010.

[3] French Energy Regulatory Commission. Deliberation on linky report (in french). http://www.cre.fr/content/download/4627/38116/version/1/file/100211experimentationLinky.pdf, February 2011.

[4] EDF internal memo report No. HR 21/99/029A, J-B. Jouve, M-L. Fourel, and D. Cortinas. Nouveau modèle de charge HTA: état de l'art sur la modélisation, December 1999.

[5] R. Singh. *State Estimation in Power Distribution Network Operation*. PhD thesis, Imperial College London, April 2009.

[6] H.L. Willis. *Power distribution planning reference book*. MARCEL DEKKER, INC, 1997.

[7] S.M.M. Agah and H.A. Abyaneh. Effect of load models on probabilistic characterization of aggregated load patterns. *IEEE Transactions on Power Systems*, 26:811 – 819, May 2011.

[8] P.E. McSharry, S. Bouwman, and G. Bloemhof. Probabilistic forecasts of the magnitude and timing of peak electricity demand. *IEEE Transactions on Power Systems*, 20:1166–1172, May 2005.

[9] J. Dickert and P. Schegner. Residential load models for network planning purposes. In *Modern Electric Power Systems 2010, Wroclaw, Poland*, 2010.

[10] D.H.O. McQueen, P.R. Hyland, and S.J. Watson. Monte Carlo simulation of residential electricity demand for forecasting maximum demand on distribution networks. *IEEE Transactions on Power Systems*, 19:1685 – 1689, August 2004.

[11] A. Mutanen, M. Ruska, S. Repo, and P. Järventausta. Customer classification and load profiling method for distribution systems. *IEEE Transactions on Power Delivery*, 26:1755 – 1763, July 2011.

[12] J.W. Taylor and P.E. McSharry. Short-term load forecasting methods: An evaluation based on european data. *Power Systems, IEEE Transactions*, 22(4), November 2007.

[13] J.W. Taylor. Triple seasonal methods for short-term electricity demand forecasting. *European Journal of Operational Research*, 204:139–152, 2010.

[14] J. Nowicka-Zagrajek and R. Weron. Modeling electricity loads in california: Arma models with hyperbolic noise. *Signal Processing*, 82:1903–1915, 2002.

[15] J.W. Taylor. An evaluation of methods for very short-term load forecasting using minute-by-minute british data. *International Journal of Forecasting*, 24:645–658, 2008.

[16] B. Wang, N. Tai, H. Zhai, J. Ye, J. Zhu, and L. Qi. A new armax model based on evolutionary algorithm and particle swarm optimization for short-term load forecasting. *Electric Power Systems Research*, 78:1679–1685, 2008.

[17] A. Bruhns, G. Deurveilher, and J.S. Roy. A non-linear regression model for mid term load forecasting and improvements in seasonality. In *15th Power Systems Computation Conference*, 2005.

[18] A. Ghanbari, A. Naghavi, S.F. Ghaderi, and M. Sabaghian. Artificial neural networks and regression approaches comparison for forecasting iranŠs annual electricity load. In *Power Engineering, Energy and Electrical Drives, 2009. POWERENG '09. International Conference on*, 18-20 March 2009.

[19] R. Lamedica, A. Prudenzi, M. Sforna, M. Caciotta, and V.O. Cencelli. A neural network based technique for short-term forecasting of anomalous load periods. *IEEE Transactions on Power Systems*, 11:1749–1756, November 1996.

[20] A. Khotanzad, R. Afkhami-Rohani, and D. Maratukulam. Annstlf-artificial neural network short-term load forecaster-generation three. *IEEE Transactions on Power Systems*, 13:1413–1422, 1998.

[21] G. Dreyfus, J.M. Martinez, M. Samuelides, M.B. Gordon, F. Badran, and S. Thiria. *Apprentissage statistique*. EYROLLES, 2008, 3rd edition.

[22] A. Capasso, W. Grattieri, R. Lamedica, and A. Prudenz. A bottom-up approach to residential load modeling. *IEEE Transactions on Power Systems*, 9:957–964, May 1994.

[23] H. Hahn, S. Meyer-Nieberg, and S. Pickl. Electric load forecasting methods: Tools for decision making. *European Journal of Operational Research*, 199:902–907, December 2009.

[24] E.A. Feinberg. *Short Term Electric Load Forecasting: A Tutorial. in Applied Mathematics for Restructured Electric Power Systems*. Springer, 2005.

[25] E. Kyriakides and M. Polycarpou. *Short Term Electric Load Forecasting: A Tutorial. In: Chen, K., Wang,L.(Eds.), Trends in Neural Computation, Studies in Computational Intelligence.* Springer, 2007.

[26] W. Charytoniuk, M.S. Chen, and P.V. Olinda. Nonparametric regression based short-term load forecasting. *IEEE Transactions on Power Systems*, 13:725–730, August 1998.

[27] V. Dordonnat, S.J. Koopman, M. Ooms, A. Dessertaine, and J. Collet. An hourly periodic state space model for modelling french national electricity load. *International Journal of Forecasting*, 24:566–587, 2008.

[28] H.S. Hippert, C.E. Pedreira, and R.C. Souza. Neural networks for short-term load forecasting: A review and evaluation. *IEEE Transactions on Power Systems*, 16:44–55, 2001.

[29] J.W. Taylor and R. Buizza. Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, 19:57–70, 2003.

[30] B. Chen, M. Chang, and C. Lin. Load forecasting using support vector machines: a study on eunite competition 2001. *Power Systems, IEEE Transactions on*, 19:1821–1830, 2004.

[31] T. Senjyu, P. Mandal, K. Uezato, and T. Funabashi. Next day load curve forecasting using hybrid correction method. *IEEE Transactions on Power Systems*, 20:102–109, 2005.

[32] K. Liu, S. Subbarayan, R.R. Shoults, M.T. Manry, C. Kwan, F.L. Lewis, and J. Naccarino. Comparison of very short-term load forecasting techniques. *IEEE Transactions on Power Systems*, 11:877–882, 1996.

[33] H.M. AI-Hamadi and S.A. Soliman. Fuzzy short-term electric load forecasting using kalman filter. *Generation, Transmission and Distribution, IEE Proceedings*, 153:217–227, 2006.

[34] A.M. Sharaf and T.T. Lie. A novel neuro-fuzzy based self-correcting online electric load forecasting model. *Electric Power Systems Research*, 34:121–125, 1995.

[35] P. Pai and W. Hong. Forecasting regional electricity load based on recurrent support vector machines with genetic algorithms. *Electric Power Systems Research*, 74:417–425, 2005.

[36] W. Hong. Electric load forecasting by support vector model. *Applied Mathematical Modelling*, 33:2444–2454, 2009.

[37] S. Rahman and O. Hazim. Load forecasting for multiple sites development of an expert system-based technique. *Electric power systems research*, 39:161–169, 1996.

[38] S. Rahman and O. Hazim. A generalized knowledge-based short term load forecasting technique. *IEEE Transactions on Power Systems*, 8:509–514, 1993.

[39] J.M.C. Sousa, L.M.P. Neves, and H.M.M. Jorge. Short-term load forecasting using information obtained from low voltage load profiles. *Power Engineering, Energy and Electrical Drives, 2009. POWERENG '09. International Conference on.*

[40] A. Kusiak, M. Li, and Z. Zhang. A data-driven approach for steam load prediction in buildings. *Applied Energy*, 87:925–933, 2010.

[41] J. Nazarko and Z.A. Styczynski. Application of statistical and neural approaches to the daily load pirofiles modelling in power distribution systems. In *Transmission and Distribution Conference, 1999 IEEE*, April 1999.

[42] V. Fortin, T.B.M.J. Ouarda, P.F. Rasmussen, and B. Bobée. Revue bibliographique des méthodes de prévision des débits. a review of streamflow forecasting methods. *Revue des sciences de l'eau*, 4:461–487, 1997.

[43] C.L. Hor, S.J. Watson, and S. Majithia. Analyzing the impact of weather variables on monthly electricity demands. *IEEE Transactions on Power Systems*, 20:2078–2085, 2005.

[44] A. Charpentier. Cours de séries temporelles théorie et applications. VOLUME 1: Introduction à la théorie des processus en temps discret modèles arima et méthode box & jenkins. http://perso.univ-rennes1.fr/arthur.charpentier/TS1.pdf.

[45] B. Hung. Univariate arima models, econometric modeling and analysis, lecture note5. www.ist.yildiz.edu.tr/dersler/dersnotu/arima1.doc, 2004.

[46] M. Zhou, Z. Yan, Y. Ni, and G. Li. An arima approach to forecasting electricity price with accuracy improvement by predicted errors. In *IEEE Power Engineering Society General Meeting*, June 2004.

[47] J.G. De Gooijer and R.J. Hyndman. 25 years of time series forecasting. *International Journal of Forecasting*, 22:443–473, 2006.

[48] P.S. Kalekar. Time series forecasting using holt-winters exponential smoothing. http://www.it.iitb.ac.in/~praj/acads/seminar/04329008_ExponentialSmoothing.pdf, December 2004.

[49] T. Senjyu, P. Mandal, K. Uezato, and T. Funabashi. Next day load curve forecasting using recurrent neural network structure. *Generation, Transmission and Distribution, IEE Proceedings*, 151:388 – 394, 2004.

[50] Appendix B: load forecasting methods. http://140.194.76.129/publications/eng-manuals/EM_1110-2-1701_pfl_noE/a-b.pdf, December 1985.

[51] P.C. Reiss and F.A. Wolak. Structural econometric modeling: Rationales and examples from industrial organization. http://www.stanford.edu/group/fwolak/cgi-bin/sites/default/files/files/Structural%20Econometric%20Modeling_Rationales%20and%20Examples%20From%20Industrial%20Organization_Reiss,%20Wolak.pdf, 2007.

[52] C. Christodoulou and M. Georgiopoulos. *Applications of Neural Networks in Electromagnetics*. ARTECH HOUSE, INC, Boston, London, 2001.

[53] M. Leshno, V. Lin, A. Pinkus, and S. Shochen. Multilayer feedforward networks with a polynomial activation function can approximate any function. *Neural Networks*, 6:861–867, 1993.

[54] M. Parizeau. Réseaux de neurones (notes de cours), université laval. http://wcours.gel.ulaval.ca/2009/a/GIF4101/default/8fichiers/reseauxdeneurones.pdf, 2006.

[55] K. Kalaitzakis, G.S. Stavrakakis, and E.M. Anagnostakis. Short-term load forecasting based on artificial neural networks parallel implementation. *Electric Power Systems Research*, 63:185–196, 2002.

[56] P.N. Tan, M. Steinbach, and V. Kumar. *Introduction to data mining*. New York: Addison Wesley, 2005.

[57] J.T. Connor, R.D. Martin, and L.E. Atlas. Recurrent neural networks and robust time series prediction. *Neural Networks*, 5:240–254, 1994.

[58] P. Bunnoon, K. Chalermyanont, and C. Limsakul. Mid-term load forecasting: Level suitably of wavelet and neural network based on factor selection. In *2nd International Conference on Advances in Energy Engineering (ICAEE)*, 2012.

[59] A.S. Pandey, D. Singh, and S.K. Sinha. Intelligent hybrid wavelet models for short-term load forecasting. *IEEE Transactions on Power Systems*, 25:1266–1273, 2010.

[60] A.J. Conejo, M.A. Plazas, R. Espínola, and A.B. Molina. Day-ahead electricity price forecasting using the wavelet transform and arima models. *IEEE Transactions on Power Systems*, 20:1035–1042, 2005.

[61] Y. Chen, P.B. Luh, C. Guan, Y. Zhao, L.D. Michel, M.A. Coolbeth, P.B. Friedland, and S.J. Rourke. Day-ahead electricity price forecasting using the wavelet transform and arima models. *IEEE Transactions on Power Systems*, 25:322–330, 2010.

[62] D. Benaouda and F. Murtagh. Electricity load forecast using neural network trained from wavelet-transformed data. In *Engineering of Intelligent Systems, 2006 IEEE International Conference on*, September 2006.

[63] T. Zheng, A.A. Girgis, and E.B. Makram. A hybrid wavelet-kalman filter method for load forecasting. *Electric Power Systems Research*, 54:11–17, 2000.

[64] D. Veitch. Wavelet neural networks and their application in the study of dynamical systems. Master's thesis, University of York, UK, August 2005.

[65] C-M. Lee and C-N. ko. Time series prediction using rbf neural networks with a nonlinear time-varying evolution PSO algorithm. *Neurocomputing*, 73:449–460, 2009.

[66] S. Fan and L. Chen. Short-term load forecasting based on an adaptive hybrid method. *IEEE Transactions on Power Systems*, 21:392–401, 2006.

[67] S.V. Verdú, M.O. García, C. Senabre, A. Gabaldón Marín, and F.J. García Franco. Classification, filtering, and identification of electrical customer load patterns through the use of self-organizing maps. *IEEE Transactions on Power Systems*, 21:1672–1682, 2006.

[68] V.N. Vapnik. *The Nature of Statistical Learning Theory*. Springer, 2000.

[69] C.C Cai and M. Wu. Support vector machines with similar day's training sample application in short-term load forecasting. In *Electric Utility Deregulation and Restructuring and Power Technologies, 2008. DRPT 2008. Third International Conference on*, 6-9 April 2008.

[70] K.J. Kim. Financial time series forecasting using support vector machines. *Neurocomputing*, 55:307–319, 2003.

[71] M. Hasan and F. Boris. Svm : Machines à vecteurs de support ou séparateurs à vastes marges. georges.gardarin.free.fr/Surveys_DM/Survey_SVM.pdf, Janvier 2006.

[72] M. Lescieux. Introduction à la logique floue. application à la commande floue. http://193.52.220.131/automatique/AUA/ressources/Introduction_logique_floue.ppt.

[73] H.K. Alfares and M. Nazeeruddin. Electric load forecasting: literature survey and classification of methods. *International Journal of Systems Science*, 33:23–34, 2002.

[74] S. Fan, C. Mao, and L. Chen. Next-day electricity-price forecasting using a hybrid network. *Generation, Transmission & Distribution, IET*, 1:176–182, 2007.

[75] I. Drezga and S. Rahman. Short-term load forecasting with local ann predictors. *IEEE Transactions on Power Systems*, 14:844–850, 1999.

[76] P. Mandal, A.K. Srivastava, and J. Park. An effort to optimize similar days parameters for ann-based electricity price forecasting. *IEEE Transactions on Industry Applications*, 45:1888–1896, 2009.

[77] T. Senjyu, H. Takara, K. Uezato, and T. Funabashi. One-hour-ahead load forecasting using neural network. *IEEE Transactions on Power Systems*, 17:113–118, 2002.

[78] P. Mandal, T. Senjyu, N. Urasaki, and T. Funabashi. A neural network based several-hour-ahead electric load forecasting using similar days approach. *Electrical Power and Energy Systems*, 28:367–373, 2006.

[79] M.EL-N. Khaled and A.AL-R. Khaled. Electric load forecasting using genetic based algorithm, optimal filter estimator and least error squares technique: Comparative study. In *World Academy of Science, Engineering and Technology*, 2005.

[80] S.T. Ng, M. Skitmore, and K.F. Wong. Using genetic algorithms and linear regression analysis for private housing demand forecast. *Building and Environment*, 43:1171–1184, 2008.

[81] H. Yang, C. Huang, and C. Huang. Identification of armax model for short term load forecasting: An evolutionary programming approach. *IEEE Transactions on Power SYstems*, 11:403–408, 1996.

[82] X. Chen and J. Wang. Time series forecasting based on novel support vector machine using artificial fish swarm algorithm. In *Natural Computation, 2008. ICNC '08. Fourth International Conference on*, 2008.

[83] D. Thilakawardana and K. Moessner. Traffic modelling and forecasting using genetic algorithms for next-generation cognitive radio applications. *Annals of Telecommunications*, 64:535–543, 2009.

[84] A. Harvey and S.J. Koopman. Forecasting hourly electricity demand using time-varying splines. *Journal of the American Statistical Association*, 88:1228–1236, 1993.

[85] J.W. Taylor. Short-term load forecasting with exponentially weighted methods. *IEEE Transactions on Power Systems*, 27:458–464, 2012.

[86] G.A. Darbellay and M. Slama. Forecasting the short-term demand for electricity. do neural networks stand a better chance? *International Journal of Forecasting*, 16:71–83, 2000.

[87] M. Adya and F. Collopy. How effective are neural networks at forecasting and prediction? a review and evaluation. *Journal of Forecasting*, 17:481–495, 1998.

[88] R.H. Shumway and D.S. Stoffer. *Time Series Analysis and Its Applications With R Examples*. Springer, 2006.

[89] The significance test in ANOVA. http://davidmlane.com/hyperstat/B87993.html.

[90] S.E. Said and D.A. Dickey. Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika*, 71:599–607, 1984.

[91] D. Kwiatkowski, P.C.B. Phillips, and P. Schmidt. Testing the null hypothesis of stationarity against the alternative of a unit root: How sure are we that economic time series have a unit root? Cowles Foundation Discussion Papers 979, Cowles Foundation, Yale University, May 1991.

[92] P.J. Brockwell and R.A. Davis. *Time Series: Theory and Methods*. Springer, 1991.

[93] J.J. Faraway. *Practical Regression and Anova using R*. CRC Press, 2002.

[94] R archive network: CRAN for download and packages. http://www.r-project.org/.

[95] Spectral analysis – smoothed periodogram method. http://www.ltrr.arizona.edu/~dmeko/notes_6.pdf, 2011.

[96] Smoothing kernel objects. http://stat.ethz.ch/R-manual/R-patched/library/stats/html/kernel.html.

[97] ERDF. Principes d'étude et de développement du réseau pour le raccordement des clients consommateurs et producteurs BT. http://www.erdfdistribution.fr/medias/DTR_Racc_Conso/ERDF-PRO-RES_43E.pdf, March 2011.

[98] French Weather Forecast Bureau. Les performances des modèles de prévision. http://entreprise.meteofrance.com/nous_connaitre2/activites_moyens/prevision/performances_modeles?page_id=14163.

[99] V.H. Ferreira and A.P. Alves da Silva. Toward estimating autonomous neural network-based electric load forecasters. *Power Systems, IEEE Transactions*, 22(4), November 2007.

[100] N. Cristianini and J. Shawe-Taylor. *An Introduction to Support Vector Machines and other kernel-based learning methods*. Cambridge University Press, 2000.

[101] C.M. Bishop. *Pattern Recognition and Machine Learning, chapter 8*. Springer, 2004.

[102] N. Ding, Y. Bésanger, F. Wurtz, G. Antoine, and P. Deschamps. Time series method for short-term load forecasting using smart metering in distribution systems. In *Power Tech Conference Proceedings, IEEE Trondheim, Norway*, 2011.

[103] K. Hornik, M. Stinchcombe, and H. White. Multilayer feedforward networks are universal approximators. *Neural Networks*, 2:359–366, 1989.

[104] N. Amjady and F. Keynia. Day-ahead price forecasting of electricity markets by mutual information technique and cascaded neuro-evolutionary algorithm. *IEEE Transactions on Power Systems*, 24:306–318, February 2009.

[105] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *Quarterly Journal of Applied Mathmatics*, II(2):164–168, 1944.

[106] I. Drezga and S. Rahman. Input variable selection for ann-based short-term load forecasting. *Power Systems, IEEE Transactions on*, 13:1238 – 1244, November 1998.

[107] I. Guyon, S. Gunn, M. Nikravesh, and eds L. Zadeh. *Feature Extraction, Foundations and Applications*. Springer, 2006.

[108] H. Stoppiglia, G. Dreyfus, R. Dubois, and Y. Oussar. Ranking a random feature for variable and feature selection. *Journal of Machine Learning Research*, 3:1399Ũ1414, March 2003.

[109] S. Chen, S.A. Billings, and W. Luo. Orthogonal least squares method and their application to non-linear system identification. *Int. J. Control*, 50:1873–1896, 1989.

[110] Gram-Schmidt orthogonalization process. http://www.math.dartmouth.edu/archive/m24w07/public_html/Lecture25.pdf.

[111] G. Dreyfus and I. Guyon. *Assessment methods, in Feature extraction, foundations and applications. Edited by I. Guyon and S. Gunn and M. Nikravesh and L. Zadeh.* Springer-Verlag, 2005.

[112] N.R. Draper and H. Smith. *Applied regression analysis*. Wiley, 1998.

[113] G. Monari and G. Dreyfus. Withdrawing an example from the training set : an analytic estimation of its effect on a non-linear parameterised model. *Neurocomputing*, 35:195–201, 2000.

[114] G. Monari and G. Dreyfus. Local overfitting control via leverages. *Neural Computation*, 14:1481–1506, 2002.

[115] P.J. Santos, A.G. Martins, and A.J. Pires. Designing the input vector to ann-based models for short-term load forecast in electricity distribution systems. *International Journal of Electrical Power & Energy Systems*, 29:338Ű347, May 2007.

[116] C.S. Chen, M.S. Kang, J.C. Hwang, and C.W. Huang. Temperature effect to distribution system load profiles and feeder losses. *IEEE Transactions on Power Systems*, 16:916–921, 2001.

[117] V. Neimane. *On Development Planning of Electricity Distribution Networks*. PhD thesis, Royal Institute of Technology, 2001.

[118] J.A. jardini, C.M.V. Tahan, M.R. Gouvea, S.U. Ahn, and F.M. Figueriredo. Daily load profiles for residential, commercial and industrial low voltage consumers. *IEEE Transactions on Power Delivery*, 15:375–380, January 2000.

[119] A. Seppälä. Vtt publication: Load research and load estimation in electricity distribution. www.vtt.fi/inf/pdf/publications/1996/P289.pdf, November 1996.

[120] G. Chicco, R. Napoli, and F. Piglione. Comparisons among clustering techniques for electricity customer classification. *IEEE Transactions on Power Systems*, 21:933–940, 2006.

[121] G. Chicco, R. Napoli, F. Piglione, P. Postolache, M. Scutariu, and C. Toader. A review of concepts and techniques for emergent customer categorisation.

[122] D. Gerbec, S. Gašperič, I. Šmon, and F. Gubina. Allocation of the load profiles to consumers using probabilistic neural networks. *IEEE Transactions on Power Systems*, 20:548 – 555, May 2005.

[123] R. Yao and K. Steemers. A method of formulating energy load profile for domestic buildings in the UK. *Energy and Buildings*, 37:663–671, 2005.

[124] T. Zhang, G. Zhang, J. Lu, X. Feng, and W. Yang. A new index and classification approach for load pattern analysis of large electricity customers. *IEEE Transactions on Power Systems*, 27:153–160, 2012.

[125] P. Vinter. Dong energy-towards the intelligent utility network. In *CIRED Seminar 2008: Smart Grids for Distribution, Frankfurt*, 23-24 June 2008.

[126] I. Wangensteen (NTNU). Tet4185: Power markets. http://www.elkraft.ntnu.no/elkraft3/fag/TET4185%20Kompendium%2006.pdf, 2006.

[127] SINTEF: N. Feilberg. Description of useload package. http://www.sintef.no/home/SINTEF-Energy-Research/Project-work/USELOAD-/, March 2012.

[128] C.S. Chen, M.S. Kang, J.C. Hwang, and C.W. Huang. Implementation of the load survey system in taipower. In *Transmission and Distribution IEEE Conference*, 1999.

[129] C.S. Chen, J.C. Hwang, Y.M. Tzeng, C.W. Huang, and M.Y. Cho. Determination of customer load characteristics by load survey system at taipower. *IEEE Transactions on Power Delivery*, 11:1430–1436, July 1996.

[130] P. Sarrouy-ARAR. ERDF internal memo: Planification des réseaux de distribution basse tension, 2009.

[131] décret 2007-1826 Texts qualité. Principales étapes de la méthode de détermination des Clients Mal Alimentés (Tenue de la tension), December 2007.

[132] EDF internal memo report No. HR 21/96. BAGHEERA modèle basse tension des charges et de calcul de l'état électrique des réseaux arborescents implantés sur poste PANTER, August 1996.

[133] T. Boulassier and M. Patin. EDF internal memo report No. HR/24-2345: Spécifications générales d'un nouveau modèle de charge BT, 1989.

[134] ERDF Direction Réseau. Présentation : La méthode de calcul des clients mal alimentés: Le modèle de calcul BAGHEERA, March 2008.

[135] F. Boulanger. Guide technique de la distribution délectricité A65.12: Calcul des réseaux BT: Le logiciel BAGHEERA, 2001.

[136] F. Moisy. A free curve fitting toolbox for Matlab. http://www.fast.u-psud.fr/ezyfit/, July 2010.

[137] E. Julien, J.P. Cosperec, and P. Feron. Lv network planning tools at EDF. In *Electricity Distribution, 1993. CIRED. 12th International Conference on, Birmingham, UK*, May 1993.

[138] T.D. Jin and M. Mechehoul. Ordering electricity via internet and its potentials for smart grid systems. *IEEE Transactions on Smart Grid*, 1:302 – 310, December 2010.

[139] A. Abdollahi, M.P. Moghaddam, M. Rashidinejad, and M.K. Sheikh-El-Eslami. Investigation of economic and environmental-driven demand response measures incorporating uc. *IEEE Transactions on Smart Grid*, 3:12 – 25, March 2012.

[140] G. Chicco, R. Napoli, P. Postolache, M. Scutariu, and C. Toader. Customer characterization options for improving the tariff offer. *IEEE Transactions on Power Systems*, 18:381–387, FEBRUARY 2003.

[141] W. Charytoniuk, M.S. Chen, P. Kotas, and P. Van Olinda. Demand forecasting in power distribution systems using nonparametric probability density estimation. *IEEE Transactions on Power Systems*, 14:1200 – 1206, November 1999.

[142] Z.I. Botev, J.F Grotowski, and D.P. Kroese. Kernel density estimation via diffusion. *Annals of Statistics*, 38:916–2957, 2010.

[143] M. Basseville and I.V. Nikiforov. *Detection of Abrupt Changes: Theory and Application*. Prentice Hall PTR, 1 May 1993.

[144] J.M. Vilar-Fernández and R. Cao. Nonparametric forecasting in time series. a comparative study. *Communications in statistics. Simulation and computation*, 36:311–334, 2007.

[145] D. Ruppert. Local polynomial regression and its applications in environmental statistics, 1996.

[146] J. Fan. Design-adaptive nonparametric regression. *Journal of American Statistical Association*, 87:998–1004, 1992.

[147] J. Fan and I. Gijbels. Variable bandwidth and local linear regression smoothers. *The Annals of Statistics*, 20:2008–2036, 1992.

[148] C. Degand. EDF internal memo report No. HR/21/95/029/A: BAGHEERA calcul de risque dans le modèle de charge BT, January 1996.

[149] M. Guida and M. Longo. Estimation of probability tails based on generalized extreme value distributions. *Reliability Engineering and System Safety*, 20:219–242, 1988.

[150] P. Embrechts, C. Kl'uppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer-Verlag, December 1996.

[151] J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3:119–131, 1975.

[152] B.M. Hill. A simple general approach to inference about the tail of a distribution. *The Annals of Statistics*, 3:1163–1174, 1975.

[153] J. Segers. Generalized pickands estimators for the extreme value index. http://alexandria.tue.nl/repository/books/560862.pdf, October 2002.

[154] A.L.M. Dekkers, J.H.J. Einmahl, and L. De haan. A moment estimator for the index of an extreme-value distribution. *The Annals of Statistics*, 17:1833–1855, 1989.

[155] S. Morel and E. Neau. Photovoltaïque : des prévisions météo aux prévisions de production. http://ewh.ieee.org/r8/france/pes/activities.html, November 2012.

[156] A. UI haque J. Meng P. Mandal, S.T.S Madhira and R.L. Pineda. Forecasting power output of solar photovoltaic system using wavelet transform and artificial intelligence techniques. *Procedia Computer Science*, 12:332–337, 2012.

[157] Y. Yang W-J. Lee, Y. Liu and P. Wang. Forecasting power output of photovoltaic system based on weather classification and support vector machine. In *Industry applications society annual meeting (IAS)*, 2011.

[158] T. Senjyu A.Yona and T. Funabashi. Application of recurrent neural network to shortŰterm ahead generating power forecasting for photovoltaic system. In *IEEE power engineering society general meeting*, 2007.

[159] Hypothesis testing. http://people.csail.mit.edu/siracusa/doc/hyptest-faq.pdf.

[160] G. Schwarz. Estimating the dimension of a model. *The Annals of Statistics*, 6:461–464, 1978.

[161] H. Akaike. On entropy maximization principle. *Krishnaiah, P.R. (Editor). Applications of Statistics, North-Holland, Amsterdam*, pages 27–41, 1977.

Appendix A

Time series model's result summary

The forecasting results of the time series method on the different MV/LV substations are shown (table A.1, A.2) during the period from September 16, 2009 to October 27, 2010.

Table A.1: *MV/LV substations*, forecasting results: comparison between the naive model and the complete Time Series (TS) model of **one-day-ahead** forecasts.

Substation	**Naive model** MAPE(%)/MAE (kW)	**TS model** MAPE(%)/MAE (kW)
VI_PAU(to June 29, 2010)	13.8/10.92	11.5/8.91
CE_MOU	17.1/4.12	15.1/3.65
CE_CHA	16.0/3.77	14.5/3.34
CE_CER	12.7/8.28	11.7/ 7.29
VI_PRI	13.3/8.66	11.5/7.18
VI_LOG (January 3, 2010 to October 26, 2010)	18.9/8.82	16.4/7.12

Table A.2: *MV/LV substations*, forecasting results: comparison between the naive model and the complete Time Series (TS) model of **two-day-ahead** forecasts.

Substation	**Naive model** MAPE(%)/MAE (kW)	**TS model** MAPE(%)/MAE (kW)
VI_PAU(to June 30, 2010)	16.2/13.17	12.8/10.25
CE_MOU	18.9/4.70	17.09/4.29
CE_CHA	18.7/4.57	16.3/3.86
CE_CER	14.9/10.05	13.2/8.53
VI_PRI	15.8/10.82	12.9/8.48
VI_LOG (January 4, 2010 to October 27, 2010)	18.9/8.82	16.6/7.23

The forecasting results of the time series method on the different MV feeders are shown (table A.3, A.4) during the period from September 16, 2009 to September 22, 2010. Due to the data missing in the MV feeder's data set, the results are given in discontinued periods.

Table A.3: *MV feeders*, forecasting results: comparison between the naive model and the complete Time Series (TS) model of **one-day-ahead** forecasts.

Feeder	**period**	**Naive model** MAPE(%)/MAE (kW)	**TS model** MAPE(%)/MAE (kW)
AL	Sept.16, 2009~ Sept.21, 2010	9.7/157.26	9.3/145.84
VI	Sept.16, 2009~ Feb. 26, 2010	10.8/164.97	7.8/124.14
	Apr.1, 2010~ Jun.30, 2010	12.2/125.98	10.7/101.93
	Jul.14, 2010~ Sept.21, 2010	10.5/88.70	8.5/70.70
CL	Sept.16, 2009~ Feb. 23, 2010	13.4/179.39	10.8/145.38
	Mar.9, 2010~ Sept.21, 2010	16.7/166.48	12.3/119.94
CE	Sept.16, 2009~ Feb. 26, 2010	13.2/386.80	9.9/299.83
	Apr.1, 2010~ Jun.30, 2010	15.3/312.88	13.6/265.56
	Jul.24, 2010~ Sept.21, 2010	15.4/270.57	11.6/197.94

Table A.4: *MV feeders*, forecasting results: comparison between the naive model and the complete Time Series (TS) model of **two-day-ahead** forecasts.

Feeder	**period**	**Naive model** MAPE(%)/MAE (kW)	**TS model** MAPE(%)/MAE (kW)
AL	Sept.17, 2009~ Sept.22, 2010	13.3/214.61	10.7/171.48
VI	Sept.17, 2009~ Feb. 27, 2010	16.0/243.88	9.2/149.25
	Apr.2, 2010~ Jul.1, 2010	18.8/188.02	12.3/116.14
	Jul.14, 2010~ Sept.22, 2010	15.0/125.80	9.7/79.90
CL	Sept.17, 2009~ Feb. 24, 2010	20.0/264.68	12.7/173.69
	Mar.10, 2010~ Sept.22, 2010	25.8/248.50	14.6/140.92
CE	Sept.17, 2009~ Feb. 27, 2010	19.9/570.99	12.0/365.20
	Apr.2, 2010~ Jul.1, 2010	24.6/490.67	15.0/289.31
	Jul.25, 2010~ Sept.22, 2010	23.9/419.01	13.5/225.45

Appendix B

Binary hypothesis test

Hypothesis test is one of the most important tools in the statistic applications. Decisions are often required to make after carrying out binary hypothesis tests on the sample based information.

In a general case, people formulate two complementary hypothesis on the studied population: null hypothesis and alternative hypothesis. Null hypothesis denoted by H_0 often states that there's no difference between the procedures. The alternative hypothesis denoted by H_A on the other hand, supposes that there are differences between the examined procedures.

- KPSS test

  In the KPSS tests, the null hypothesis states that an observable time series is stationary around a deterministic trend. Let $X_t, t = 1, 2, \cdots, N$ be the observed series, it is decomposed into the sum of a deterministic trend βt, a random walk γ_t and a stationary error ϵ_t, such that:

  $$X_t = \gamma_t + \beta t + \epsilon_t$$

  The random walk $\gamma_t = \gamma_{t-1} + \mu_t$, where μ_t is an independent and identically distributed (i.i.d.) Gaussian distribution $\mathcal{N}(0, \sigma_\mu^2)$. To test the stationarity of X_t around the deterministic trend βt, the null hypothesis H_0 is set as $\sigma_\mu^2 = 0$. It means that the random walk γ_t is stationary against the alternative hypothesis H_A: $\sigma_\mu^2 > 0$.

  The decision of whether or not we accept H_0 is made by comparing a calculated p-value with a predefined type I error. The bigger the p-value is, the greater chance that we accept H_0. The consistency is measured by calculating the probability of getting the test statistic value greater than the observed value from our sample data, assuming the H_0 is true [159]. The type I error symbolizes the false rejection of the null hypothesis H_0. Conventionally, 5% is chosen as the type I error (also known as the significance level), which corresponds to 1 in 20 chances being wrong. Based upon the sample distribution of the estimator for the parameter, the critical value is obtained by using the probability tables. The acceptance region and the rejection region are defined by the critical value. Results of the KPSS test on the trend removal series is presented as follows:

KPSS Test for Level Stationarity data: trend removal consumption data
KPSS Level = 0.147, Truncation lag parameter = 32, p-value = 0.1
Message d'avis :
In kpss.test (W_t) : p-value greater than printed p-value

"p-value greater than printed p-value", which certificates the stationary of the series.

- ADF test

  ADF test is a test for a unit root in a time series sample. The tests are based on the following three regression forms:

  1. Without constant and trend $\Delta X_t = \delta X_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta X_{t-i} + \epsilon_t$
  2. With constant $\Delta X_t = \alpha + \delta X_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta X_{t-i} + \epsilon_t$
  3. With constant and trend $\Delta X_t = \alpha + \beta t + \delta X_{t-1} + \epsilon_t$

  where X_t is the studied series, and Δ represents the difference operator. The number of maximum lags p is determined by minimizing the Bayesian information criterion (BIC) [160] or Akaike information criterion (AIC) [161]. Parameters $\{\alpha, \beta, \delta, \alpha_i\}$ are computed with OLS. ϵ_t corresponds to the random error.

  The null hypothesis of the ADF test is $\delta = 0$ (The data needs to be differenced to be stationary) versus the alternative hypothesis $\delta < 0$ (The data is stationary). An example of ADF test is given below:

Augmented Dickey-Fuller Test
data: trend removal consumption data
Dickey-Fuller = -22.1512, Lag order = 26, p-value = 0.01
alternative hypothesis: stationary
Message d'avis :
In adf.test (W_t) : p-value smaller than printed p-value

"p-value smaller than printed p-value", which confirms that the data is stationary.

Appendix C

Example of ANOVA nullity test

Suppose that we have the following equation:

$$W_t = c_1\cos(2\pi\omega_1 t) + s_1\sin(2\pi\omega_1 t) + c_2\cos(2\pi\omega_2 t) + s_2\sin(2\pi\omega_2 t) + c_3\cos(2\pi\omega_3 t) + s_3\sin(2\pi\omega_3 t) + c_4\cos(2\pi\omega_4 t) + s_4\sin(2\pi\omega_4 t) \quad \text{(C.1)}$$

The ANOVA nullity test can be applied on the estimated parameters $c_i, s_i (i = 1, \cdots, 4)$ to check if each regression part is significant for the response variable. The following shows the output of the test in "R" environment:

Analysis of Variance Table

Response: Trend removal data

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
c_1	1	4.8271e+10	4.8271e+10	392.9564	<2.2e-16	***
s_1	1	3.1526e+07	3.1526e+07	0.2566	0.612450	
c_2	1	1.3067e+10	1.3067e+10	106.3697	<2.2e-16	***
s_2	1	1.1836e+09	1.1836e+09	9.6355	0.001913	**
c_3	1	1.8030e+11	1.8030e+11	1467.7810	<2.2e-16	***
s_3	1	8.7868e+09	8.7868e+09	71.5296	<2.2e-16	***
c_4	1	1.1978e+11	1.1978e+11	975.0600	<2.2e-16	***
s_4	1	6.6406e+09	6.6406e+09	54.0584	2.072e-13	***
Residuals	12087	1.4848e+12	1.2284e+08			

Signif.codes: 0 '*' 0.001'**' 0.01'*' 0.05'.' 0.1' ' 1**

The analysis of variance table is organized in lines, each of which denotes the variation assigned to every regression part. In our example, there are 8 regression parts and one residual part. The "Df" in the table stands for "degree of freedom". Every independent input/output vector that participates for the model estimation is counted as one Df. In the regression model, each regression part occupies one degree of freedom, and the residuals keep all the rest of the Df of the data set. In this example, we have 12087 Df on the residuals, plus 8 Df, and we must also count 1 for the intercept part. Thus the number of input/output examples for the estimation use is 12096 (12087+8+1). "Sum Sq" denotes the sum of square, which represents the sum of squared deviations. "Mean Sq" denotes the mean of square, which is an estimator of average squared error between the estimated value and the real value. These two terms indicate how where the correspondent regression part explain the variability of the final observations. The greater value they are, the bigger variability the correspondent regression part explains.

The F value is the ratio between Mean sq of regression part and Mean sq of the residual. As a matter of fact, for each line, we compare the variability explained by the regression part with the one explained by the residuals. H_0 is set up as: There's no difference between Mean sq of regression part and that of the residuals. The alternative hypothesis H_A on the other hand defines that there are differences in means, so the regression part is significant to the response variable. The critical value can be found in the F distribution table with a give α level, the Df of the regression part and the residuals'. In our example, the critical value is $F_{0.05;8,12087} = 1.94$.

The p-value represents the probability that we accept H_0. In a word, the greater the F value is, the smaller the p-value will be. This later signifies that H_0 has a greater possibility to be rejected. In the software "R", the significance is both expressed by the p-value and asterisks. The result is flagged with"$*$", when p-value is less than 0.05, with "$**$", when p-value is less than 0.01, and with three "$***$", when p-value is less than 0.001.

The ANOVA table in our example indicates that the "s_1" regression part is not significant for the variability of the studied variable. As a result, the regression equation C.1 to the ANOVA nullity test is adjusted as:

$$\begin{aligned} W_t = {} & c_1 \cos(2\pi\omega_1 t) + c_2 \cos(2\pi\omega_2 t) + s_2 \sin(2\pi\omega_2 t) \\ & + c_3 \cos(2\pi\omega_3 t) + s_3 \sin(2\pi\omega_3 t) + c_4 \cos(2\pi\omega_4 t) + s_4 \sin(2\pi\omega_4 t) \end{aligned} \tag{C.2}$$

Appendix D

Comparison results of naive model, time series model and neural network model

In this appendix, we compare the detailed results of three models. Results on *CE_MOU* and *CE_FRO* substations are shown. The first one is a substation constituted mainly by residential clients. The other one is a substation that only supplies to an industrial client. Lack of information, time series model is not suitable for the industrial substation short-term forecast. Therefore, for *substation CE_FRO*, the comparison is made only between naive model and neural network model. The entire available data for this study is from September 9, 2009 to April 5, 2012 (940 days). We applied the maximum data in this study since that we wanted to show the performance of the models over time and to support the discussion on the update frequency of the neural network model in chapter 4. The testing period can only begin one year after the beginning of the available data for considering that the neural network model takes one-year period as the learning data. The testing period is from September 16, 2010 to April 5, 2012 (568 days).

As explained in chapter IV that with different risks taken in selecting one non effective variable can lead to different number of variables to both the daily average power model and the intraday power variation model. We have also explained in chapter 4 that even with the same risk taken, because of the random shuffling effect in creating the probe variables, the variable numbers of different trials can be varying. We present the result on the *substation CE_MOU* in the following:

1. 0% risk for the average power model and the 20 top ranking variables for the power variation model

Table D.1: 6 variables for the daily average power model and 19 variables for the intraday power variation model. "NM", "TS", and "NN" respectively stand for Naive model, Time Series model, and Neural Network model. "PH" stands for Public Holiday(s).

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Sept. 2010 (15 days)	3.33	2.73	2.89	22.9	18.0	21.2
Oct. 2010	4.09	3.41	3.16	18.0	15.2	14.7
Nov. 2010	5.13	4.79	3.77	13.3	12.3	10.1
Nov. 2010 (without PH)	5.14	4.64	3.61	13.2	11.7	9.7
Dec. 2010	5.27	4.48	4.09	8.3	7.1	6.6
Dec. 2010 (without PH)	5.19	4.44	4.13	8.3	7.1	6.7
Jan. 2011	5.12	4.57	3.97	9.1	8.1	7.2
Jan. 2011 (without PH)	5.10	4.55	3.99	9.1	8.1	7.2

Table D.1 – continued from previous page

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Feb. 2011	5.17	4.27	3.79	10.9	8.8	8.2
Mar. 2011	4.79	4.13	3.73	13.3	11.3	10.3
Apr. 2011	3.38	3.19	2.92	18.7	17.5	17.5
Apr. 2011 (without PH)	3.39	3.23	2.94	18.4	17.4	17.3
May 2011	2.66	2.25	2.66	19.8	16.7	22.1
May 2011 (without PH)	2.57	2.23	2.60	19.1	16.3	21.6
June 2011	3.02	2.69	2.61	22.8	20.3	19.8
June 2011 (without PH)	2.98	2.68	2.64	22.5	20.4	20.1
July 2011	2.52	2.20	2.02	20.0	17.6	16.0
July 2011 (without PH)	2.38	2.19	2.15	19.4	17.8	17.4
Aug. 2011	3.06	2.72	2.46	25.4	23.1	20.5
Aug. 2011 (without PH)	3.05	2.72	2.48	25.1	22.9	20.6
Sept. 2011	2.75	2.36	2.28	21.7	18.8	18.6
Oct. 2011	3.61	2.93	3.17	19.2	16.5	19.0
Nov. 2011	4.40	3.53	3.81	15.8	12.9	14.2
Nov. 2011 (without PH)	4.40	3.49	3.82	15.7	12.7	14.2
Dec. 2011	4.59	3.76	4.27	10.7	8.7	10.4
Dec. 2011 (without PH)	4.71	3.79	4.21	11.0	8.8	10.3
Jan. 2012	4.84	4.05	4.29	10.2	8.6	9.6
Jan. 2012 (without PH)	4.85	4.04	4.32	10.2	8.6	9.6
Feb. 2012	4.73	4.89	4.44	9.2	9.7	9.1
Mar. 2012	4.80	4.29	3.80	15.1	13.8	12.4
Sept. 16, 2010 ~ Mar. 01, 2011 (Total 167 days)	4.80	4.16	3.68	12.9	11.0	10.5
Sept. 16, 2010 ~ Apr. 05, 2012 (Total 568 days)	4.09	3.56	3.55	15.9	13.8	13.9

2. 0% risk for the average power model and the 30 top ranking variables for the power variation model

Table D.2: 6 variables for the daily average power model and 23 variables for the intraday power variation model

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Sept. 2010 (15 days)	3.33	2.73	2.93	22.9	18.0	21.5
Oct. 2010	4.09	3.41	3.15	18.0	15.2	14.7
Nov. 2010	5.13	4.79	3.77	13.3	12.3	10.1
Nov. 2010 (without PH)	5.14	4.64	3.63	13.2	11.7	9.6
Dec. 2010	5.27	4.48	4.06	8.3	7.1	6.6
Dec. 2010 (without PH)	5.19	4.44	4.10	8.3	7.1	6.7

Table D.2 – continued from previous page

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Jan. 2011	5.12	4.57	3.95	9.1	8.1	7.2
Jan. 2011 (without PH)	5.10	4.55	3.96	9.1	8.1	7.2
Feb. 2011	5.17	4.27	3.72	10.9	8.8	8.5
Mar. 2011	4.79	4.13	3.79	13.3	11.3	10.5
Apr. 2011	3.38	3.19	2.79	18.7	17.5	16.5
Apr. 2011 (without PH)	3.39	3.23	2.81	18.4	17.4	16.4
May 2011	2.66	2.25	2.60	19.8	16.7	21.6
May 2011 (without PH)	2.57	2.23	2.54	19.1	16.3	21.1
June 2011	3.02	2.69	2.70	22.8	20.3	20.7
June 2011 (without PH)	2.98	2.68	2.73	22.5	20.4	21.0
July 2011	2.52	2.20	2.09	20.0	17.6	16.8
July 2011 (without PH)	2.38	2.19	2.22	19.4	17.8	18.2
Aug. 2011	3.06	2.72	2.52	25.4	23.1	20.9
Aug. 2011 (without PH)	3.05	2.72	2.53	25.1	22.9	20.8
Sept. 2011	2.75	2.36	2.23	21.7	18.8	18.0
Oct. 2011	3.61	2.93	3.12	19.2	16.5	18.7
Nov. 2011	4.40	3.53	3.83	15.8	12.9	14.2
Nov. 2011 (without PH)	4.40	3.49	3.85	15.7	12.7	14.3
Dec. 2011	4.59	3.76	4.32	10.7	8.7	10.5
Dec. 2011 (without PH)	4.71	3.79	4.27	11.0	8.8	10.5
Jan. 2012	4.84	4.05	4.53	10.2	8.6	10.2
Jan. 2012 (without PH)	4.85	4.04	4.56	10.2	8.6	10.3
Feb. 2012	4.73	4.89	4.32	9.2	9.7	9.0
Mar. 2012	4.80	4.29	3.86	15.1	13.8	12.6
Sept. 16, 2010 ~ Mar. 01, 2011 (Total 167 days)	4.80	4.16	3.69	12.9	11.0	10.5
Sept. 16, 2010 ~ Apr. 05, 2012 (Total 568 days)	4.09	3.56	3.41	15.9	13.8	14.0

3. 0% risk both for the average power model and for the power variation model

Table D.3: 6 variables for the daily average power model and 40 variables for the intraday power variation model

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Sept. 2010 (15 days)	3.33	2.73	2.84	22.9	18.0	20.9
Oct. 2010	4.09	3.41	2.99	18.0	15.2	13.8
Nov. 2010	5.13	4.79	3.56	13.3	12.3	9.5
Nov. 2010 (without PH)	5.14	4.64	3.43	13.2	11.7	9.1
Dec. 2010	5.27	4.48	3.92	8.3	7.1	6.3

Table D.3 – continued from previous page

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Dec. 2010 (without PH)	5.19	4.44	3.98	8.3	7.1	6.4
Jan. 2011	5.12	4.57	3.88	9.1	8.1	7.1
Jan. 2011 (without PH)	5.10	4.55	3.89	9.1	8.1	7.1
Feb., 2011	5.17	4.27	4.15	10.9	8.8	9.0
Mar. 2011	4.79	4.13	3.98	13.3	11.3	11.2
Apr. 2011	3.38	3.19	2.91	18.7	17.5	17.4
Apr. 2011 (without PH)	3.39	3.23	2.91	18.4	17.4	17.1
May 2011	2.66	2.25	2.70	19.8	16.7	22.5
May 2011 (without PH)	2.57	2.23	2.60	19.1	16.3	21.7
June 2011	3.02	2.69	2.72	22.8	20.3	20.7
June 2011 (without PH)	2.98	2.68	2.74	22.5	20.4	21.0
July 2011	2.52	2.20	2.06	20.0	17.6	16.4
July 2011 (without PH)	2.38	2.19	2.25	19.4	17.8	18.2
Aug. 2011	3.06	2.72	2.58	25.4	23.1	21.2
Aug. 2011 (without PH)	3.05	2.72	2.60	25.1	22.9	21.3
Sept. 2011	2.75	2.36	2.15	21.7	18.8	17.7
Oct. 2011	3.61	2.93	3.09	19.2	16.5	18.9
Nov. 2011	4.40	3.53	3.84	15.8	12.9	14.3
Nov. 2011 (without PH)	4.40	3.49	3.87	15.7	12.7	14.4
Dec. 2011	4.59	3.76	4.55	10.7	8.7	11.1
Dec. 2011 (without PH)	4.71	3.79	4.52	11.0	8.8	11.1
Jan. 2012	4.84	4.05	4.83	10.2	8.6	10.8
Jan. 2012 (without PH)	4.85	4.04	4.84	10.2	8.6	10.9
Feb. 2012	4.73	4.89	4.32	9.2	9.7	9.0
Mar. 2012	4.80	4.29	4.07	15.1	13.8	13.3
Sept. 16, 2010 ~ Mar. 01, 2011 (Total 167 days)	4.80	4.16	3.62	12.9	11.0	10.2
Sept. 16, 2010~ Apr. 05, 2012 (Total 568 days)	4.09	3.56	3.45	15.9	13.8	14.2

4. 5% risk for the average power model and 0% risk for the power variation model

Table D.4: 10 variables for the daily average power model and 37 variables for the intraday power variation model

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Sept. 2010 (15 days)	3.33	2.73	2.81	22.9	18.0	20.7
Oct. 2010	4.09	3.41	3.23	18.0	15.2	14.7
Nov. 2010	5.13	4.79	3.56	13.3	12.3	9.6
Nov. 2010 (without PH)	5.14	4.64	3.42	13.2	11.7	9.1

Table D.4 – continued from previous page

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Dec. 2010	5.27	4.48	3.88	8.3	7.1	6.2
Dec. 2010 (without PH)	5.19	4.44	3.90	8.3	7.1	6.3
Jan. 2011	5.12	4.57	3.88	9.1	8.1	7.1
Jan. 2011 (without PH)	5.10	4.55	3.91	9.1	8.1	7.1
Feb. 2011	5.17	4.27	3.72	10.9	8.8	8.0
Mar. 2011	4.79	4.13	3.83	13.3	11.3	10.8
Apr. 2011	3.38	3.19	2.98	18.7	17.5	17.8
Apr. 2011 (without PH)	3.39	3.23	2.98	18.4	17.4	17.5
May 2011	2.66	2.25	2.64	19.8	16.7	21.9
May 2011 (without PH)	2.57	2.23	2.55	19.1	16.3	21.1
June 2011	3.02	2.69	2.62	22.8	20.3	20.2
June 2011 (without PH)	2.98	2.68	2.63	22.5	20.4	20.3
July 2011	2.52	2.20	2.11	20.0	17.6	17.0
July 2011 (without PH)	2.38	2.19	2.19	19.4	17.8	18.0
Aug. 2011	3.06	2.72	2.60	25.4	23.1	22.3
Aug. 2011 (without PH)	3.05	2.72	2.60	25.1	22.9	22.2
Sept. 2011	2.75	2.36	2.33	21.7	18.8	19.6
Oct. 2011	3.61	2.93	3.25	19.2	16.5	20.3
Nov. 2011	4.40	3.53	4.11	15.8	12.9	15.6
Nov. 2011 (without PH)	4.40	3.49	4.17	15.7	12.7	15.7
Dec. 2011	4.59	3.76	4.73	10.7	8.7	11.7
Dec. 2011 (without PH)	4.71	3.79	4.68	11.0	8.8	11.6
Jan. 2012	4.84	4.05	5.33	10.2	8.6	12.1
Jan. 2012 (without PH)	4.85	4.04	5.35	10.2	8.6	12.2
Feb. 2012	4.73	4.89	4.43	9.2	9.7	9.4
Mar. 2012	4.80	4.29	4.20	15.1	13.8	14.1
Sept. 16, 2010 ~ Mar. 01, 2011 (Total 167 days)	4.80	4.16	3.58	12.9	11.0	10.2
Sept. 16, 2010 ~ Apr. 05, 2012 (Total 568 days)	4.09	3.56	3.50	15.9	13.8	14.6

5. 10% risk for the average power model and 0% risk for the power variation model

Table D.5: 10 variables for the daily average power model and 37 variables for the intraday power variation model

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Sept. 2010 (15 days)	3.33	2.73	2.78	22.9	18.0	20.6
Oct. 2010	4.09	3.41	3.20	18.0	15.2	14.8
Nov. 2010	5.13	4.79	3.53	13.3	12.3	9.5

Table D.5 – continued from previous page

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Nov. 2010 (without PH)	5.14	4.64	3.37	13.2	11.7	8.9
Dec. 2010	5.27	4.48	3.81	8.3	7.1	6.2
Dec. 2010 (without PH)	5.19	4.44	3.85	8.3	7.1	6.2
Jan. 2011	5.12	4.57	3.81	9.1	8.1	6.9
Jan. 2011 (without PH)	5.10	4.55	3.82	9.1	8.1	7.0
Feb. 2011	5.17	4.27	3.70	10.9	8.8	8.0
Mar. 2011	4.79	4.13	3.83	13.3	11.3	10.7
Apr. 2011	3.38	3.19	2.94	18.7	17.5	17.6
Apr. 2011 (without PH)	3.39	3.23	2.93	18.4	17.4	17.2
May 2011	2.66	2.25	2.62	19.8	16.7	21.5
May 2011 (without PH)	2.57	2.23	2.52	19.1	16.3	20.6
June 2011	3.02	2.69	2.67	22.8	20.3	20.6
June 2011 (without PH)	2.98	2.68	2.69	22.5	20.4	20.9
July 2011	2.52	2.20	2.14	20.0	17.6	16.8
July 2011 (without PH)	2.38	2.19	2.25	19.4	17.8	18.3
Aug. 2011	3.06	2.72	2.52	25.4	23.1	21.1
Aug. 2011 (without PH)	3.05	2.72	2.53	25.1	22.9	21.1
Sept. 2011	2.75	2.36	2.32	21.7	18.8	19.3
Oct. 2011	3.61	2.93	3.14	19.2	16.5	29.5
Nov. 2011	4.40	3.53	3.94	15.8	12.9	14.8
Nov. 2011 (without PH)	4.40	3.49	4.01	15.7	12.7	14.8
Dec. 2011	4.59	3.76	4.56	10.7	8.7	11.1
Dec. 2011 (without PH)	4.71	3.79	4.48	11.0	8.8	11.0
Jan. 2012	4.84	4.05	5.13	10.2	8.6	11.7
Jan. 2012 (without PH)	4.85	4.04	5.14	10.2	8.6	11.7
Feb. 2012	4.73	4.89	4.56	9.2	9.7	9.7
Mar. 2012	4.80	4.29	4.19	15.1	13.8	14.1
Sept. 16, 2010 ~ Mar. 01, 2011 (Total 167 days)	4.80	4.16	3.53	12.9	11.0	10.1
Sept. 16, 2010 ~ Apr. 05, 2012 (Total 568 days)	4.09	3.56	3.46	15.9	13.8	14.3

6. 50% risk for the average power model and 0% risk for the power variation model

Table D.6: 24 variables for the daily average power model and 32 variables for the intraday power variation model.

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Sept. 2010 (15 days)	3.33	2.73	2.80	22.9	18.0	20.3
Oct. 2010	4.09	3.41	3.01	18.0	15.2	13.8

Table D.6 – continued from previous page

	MAE (kW)			MAPE (%)		
	NM	TS	NN	NM	TS	NN
Nov. 2010	5.13	4.79	3.31	13.3	12.3	9.1
Nov. 2010 (without PH)	5.14	4.64	3.21	13.2	11.7	8.7
Dec. 2010	5.27	4.48	3.81	8.3	7.1	6.0
Dec. 2010 (without PH)	5.19	4.44	3.86	8.3	7.1	6.1
Jan. 2011	5.12	4.57	4.16	9.1	8.1	7.7
Jan. 2011 (without PH)	5.10	4.55	4.15	9.1	8.1	7.7
Feb. 2011	5.17	4.27	4.63	10.9	8.8	10.2
Mar. 2011	4.79	4.13	3.97	13.3	11.3	11.5
Apr. 2011	3.38	3.19	2.61	18.7	17.5	15.1
Apr. 2011 (without PH)	3.39	3.23	2.60	18.4	17.4	14.8
May 2011	2.66	2.25	2.35	19.8	16.7	18.4
May 2011 (without PH)	2.57	2.23	2.24	19.1	16.3	17.4
June 2011	3.02	2.69	2.55	22.8	20.3	19.5
June 2011 (without PH)	2.98	2.68	2.56	22.5	20.4	19.8
July 2011	2.52	2.20	2.05	20.0	17.6	16.6
July 2011 (without PH)	2.38	2.19	2.21	19.4	17.8	18.3
Aug. 2011	3.06	2.72	2.50	25.4	23.1	20.6
Aug. 2011 (without PH)	3.05	2.72	2.51	25.1	22.9	20.7
Sept. 2011	2.75	2.36	2.21	21.7	18.8	17.8
Oct. 2011	3.61	2.93	3.02	19.2	16.5	18.1
Nov. 2011	4.40	3.53	4.61	15.8	12.9	17.5
Nov. 2011 (without PH)	4.40	3.49	4.67	15.7	12.7	17.7
Dec. 2011	4.59	3.76	5.56	10.7	8.7	13.7
Dec. 2011 (without PH)	4.71	3.79	4.48	11.0	8.8	13.6
Jan. 2012	4.84	4.05	6.84	10.2	8.6	15.5
Jan. 2012 (without PH)	4.85	4.04	6.87	10.2	8.6	15.6
Feb. 2012	4.73	4.89	5.89	9.2	9.7	12.3
Mar. 2012	4.80	4.29	5.13	15.1	13.8	17.0
Sept. 16, 2010 ~ Mar. 01, 2011 (Total 167 days)	4.80	4.16	3.68	12.9	11.0	10.3
Sept. 16, 2010 ~ Apr. 05, 2012 (Total 568 days)	4.09	3.56	3.75	15.9	13.8	14.7

Table D.7 shows the comparison results in months of the industrial *substation CE_FRO* during the testing period. Notice that the NN model has a much better precision than the naive model.

Table D.7: *Substation CE_FRO*: 14 variables for the daily average power model and 28 variables for the intraday power variation model.

	MAE (kW)		MAPE (%)	
	NM	NN	NM	NN
Sept. 2010 (15 days)	88.57	74.81	18.3	15.8
Oct. 2010	100.38	74.31	20.2	15.5
Nov. 2010	104.30	70.39	21.8	15.1
Nov. 2010 (without PH)	94.40	70.10	18.6	14.7
Dec. 2010	112.23	85.90	23.1	17.3
Dec.2010 (without PH)	113.81	88.01	22.1	17.3
Jan. 2011	93.31	72.81	18.2	14.4
Jan. 2011 (without PH)	94.85	72.82	18.3	14.2
Feb. 2011	91.92	71.58	18.4	14.9
Mar. 2011	63.28	68.47	12.2	13.2
Apr. 2011	106.71	83.88	24.9	21.0
Apr. 2011 (without PH)	103.21	82.31	22.2	19.2
May 2011	116.65	87.78	23.7	18.8
May 2011 (without PH)	118.36	89.14	23.5	18.6
June 2011	143.22	94.54	36.5	25.6
June 2011 (without PH)	129.98	89.26	28.0	21.5
July 2011	129.87	93.34	31.2	23.3
July 2011 (without PH)	108.85	84.79	21.7	19.0
Aug. 2011	131.44	95.40	28.8	21.7
Aug. 2011 (without PH)	123.39	93.20	24.7	20.2
Sept. 2011	90.44	78.83	19.1	18.4
Oct. 2011	99.07	82.08	23.0	21.1
Nov. 2011	138.16	84.07	32.3	21.2
Nov. 2011 (without PH)	123.05	77.91	23.2	17.1
Dec. 2011	94.41	80.97	22.5	19.91
Dec. 2011 (without PH)	98.48	83.97	22.5	19.8
Jan. 2012	89.35	78.22	94.3	18.5
Jan. 2012 (without PH)	91.01	78.17	19.4	17.9
Feb. 2012	94.63	72.73	19.8	15.7
Mar. 2012	104.21	94.83	21.3	21.0
Sept. 16, 2010 ~ Mar. 01, 2011 (Total 167 days)	99.49	75.07	20.2	15.5
Sept. 16, 2010 ~ Apr. 05, 2012 (Total 568 days)	105.20	81.79	23.0	18.6

Appendix E

Résumé français

E.1 Introduction générale: la nouvelle problématique du modèle de charge dans le contexte du réseau intelligent

E.1.a Réseau intelligent et compteurs intelligents pour les modèles de charge

La conception du réseau intelligent (smart grid en anglais) combine les nouvelles technologies de communication et l'ancien réseau électrique de distribution. Le but du développement de ce réseau intelligent est d'améliorer l'observabilité et le moyen de contrôle dans le réseau électrique de distribution. Face aux évolutions révolutionnaires dans les systèmes électriques, telles que la quantité importante d'énergie renouvelable connectée au réseau, l'augmentation de la demande d'énergie due aux véhicules électriques, etc., de nombreux algorithmes avancés sont apparus pour renforcer la stabilité et l'efficacité du système. Ces Fonctions Avancées du Réseau (FAR) incluent notamment le réglage de tension [1], l'autocicatrisation, et le contrôle directe de la consommation électrique [2]. Les FARs peuvent être calculées en temps réel ou à l'avance pour aider la prise des décisions. Généralement, le contrôle et la conduite des réseaux de distribution sont effectués sur le réseau moyenne tension.

Un des objectifs du réseau intelligent est de rendre les réseaux de distribution efficaces du point de vue économique, et d'obtenir une alimentation d'énergie à faible coût. La planification du réseau de distribution consiste en le développement d'un programme des investissements futurs qui garantissent la distribution de l'énergie ainsi que les coûts les plus faibles possibles. D'un côté, les infrastructures électriques doivent répondre aux exigences du réseau pendant les pointes de consommation. D'autre part, un réseau surdimensionné coûte très cher. Par conséquent, des modèles d'estimation de charge les plus fiables possibles sont nécessaires pour minimiser la marge et optimiser les investissements dans le réseau de distribution en mettant en place des calculs électriques. L'exécution du calcul de calcul de répartition de charge sous des conditions critiques afin d'identifier des zones mal-alimentées est un exemple concret. La complexité de ce problème est aussi liée aux incertitudes sur la consommation électrique des clients.

Dans l'état actuel, un nombre très limité de mesures dans le réseau de distribution introduit des difficultés pour la mise en place des FARs ainsi que des calculs d'optimisation du réseau. Les mesures disponibles dans les réseaux de distribution sont généralement dans les postes sources. Il est économiquement infaisable d'installer des capteurs de mesure sur les 738 000 postes HTA/BT. Aujourd'hui, dans la conduite, on remplace ces mesures par des modèles probabilistes avec 50% de précision. Cet acte influence l'efficacité des FARs et génère des analyses peu crédibles. Pour répondre aux besoins de la planification, un

modèle intitulé BAGHEERA est actuellement appliqué par l'Electricité de France (EDF). Ce modèle dépend principalement des informations qualitatives du client individuel, qui deviennent de moins en moins disponibles et précises. En conséquence, un nouveau modèle pour remplacer le modèle BAGHEERA est demandé.

En 2010, ERDF a démarré un projet nommé « Linky », qui envisage d'installer 35 000 000 compteurs intelligents en France. D'un côté, les clients vont payer leurs factures d'électricité selon leurs consommations réelles au lieu des consommations estimées comme cela est le cas actuellement. D'autre part, grâce à ces mesures, les opérateurs de réseau de distribution peuvent surveiller de près la situation en temps réel des réseaux. Aujourd'hui, il n'y a pas de mesures disponibles dans les postes HTA/BT en France. Dans la phase expérimentale du projet « Linky », les consommations individuelles des clients sont échantillonnées toutes les 30 minutes et transférées une fois par jour aux concentrateurs correspondants. Néanmoins, puisque les données sont assemblées en paquets et transférées à une certaine fréquence [3], il y a des retards dans ces données de mesure.

E.1.b Objectifs et plan du résumé français

Avec l'arrivée des données détaillées de la consommation individuelle du client fournies par les compteurs intelligents, l'objectif des travaux de recherche présentés ici est la conception de nouveaux modèles de charge pour la conduite et la planification des réseaux de distribution. Ce contexte permet d'envisager des modèles plus précis pour la planification, le contrôle et la conduite du réseau, en l'absence de capteurs des mesures couteux pour le réseau de distribution.

Pour la conduite, nous avons besoin de la prévision de charge sur une ou deux journées (« J+1 » et « J+2 ») pour les postes HTA/BT, basée sur les données agrégées des compteurs intelligents. Nous considérons trois raisons pour cet objectif:

- En cas de défaut: pour la réalimentation, le département de conduite aimerait avoir le total de charge à reprendre pour les 3 minutes suivantes.

- En cas de travaux sur le réseau: l'objectif est similaire à la prévision en cas de défaut, il faut réaliser une estimation de la courbe de charge à reprendre durant les travaux. Le programme de reprise s'établit en général à « J+2 » (sur le lendemain). Donc il faut faire une prévision avec dispersion d'erreur pour « J+2 ».

- Pour le réglage de la tension de consigne (étant une entrée pour l'estimateur d'état [5]): l'estimateur d'état est une fonction clé pour tous les systèmes de gestion de l'énergie. Il estime les grandeurs du réseau, telles que l'amplitude et l'angle de la tension. La Figure E.1 montre la relation entre les modèles de charge, l'estimateur d'état et les fonctions avancées du réseau. Les modèles de charge prédictifs et les données du réseau sont considérés comme des entrées pour l'estimateur du réseau. Les données du réseau [5] incluent les informations sur la topologie du réseau, les impédances (résistances et réactances) des lignes, les réglages des prises, et les charges des lignes, etc. La sortie de l'estimateur d'état active le module de contrôle de la gestion avancée des systèmes de distribution afin de pouvoir prendre des décisions pour la conduite. Ces décisions concernent les contrôles commandes des dispositifs dans le

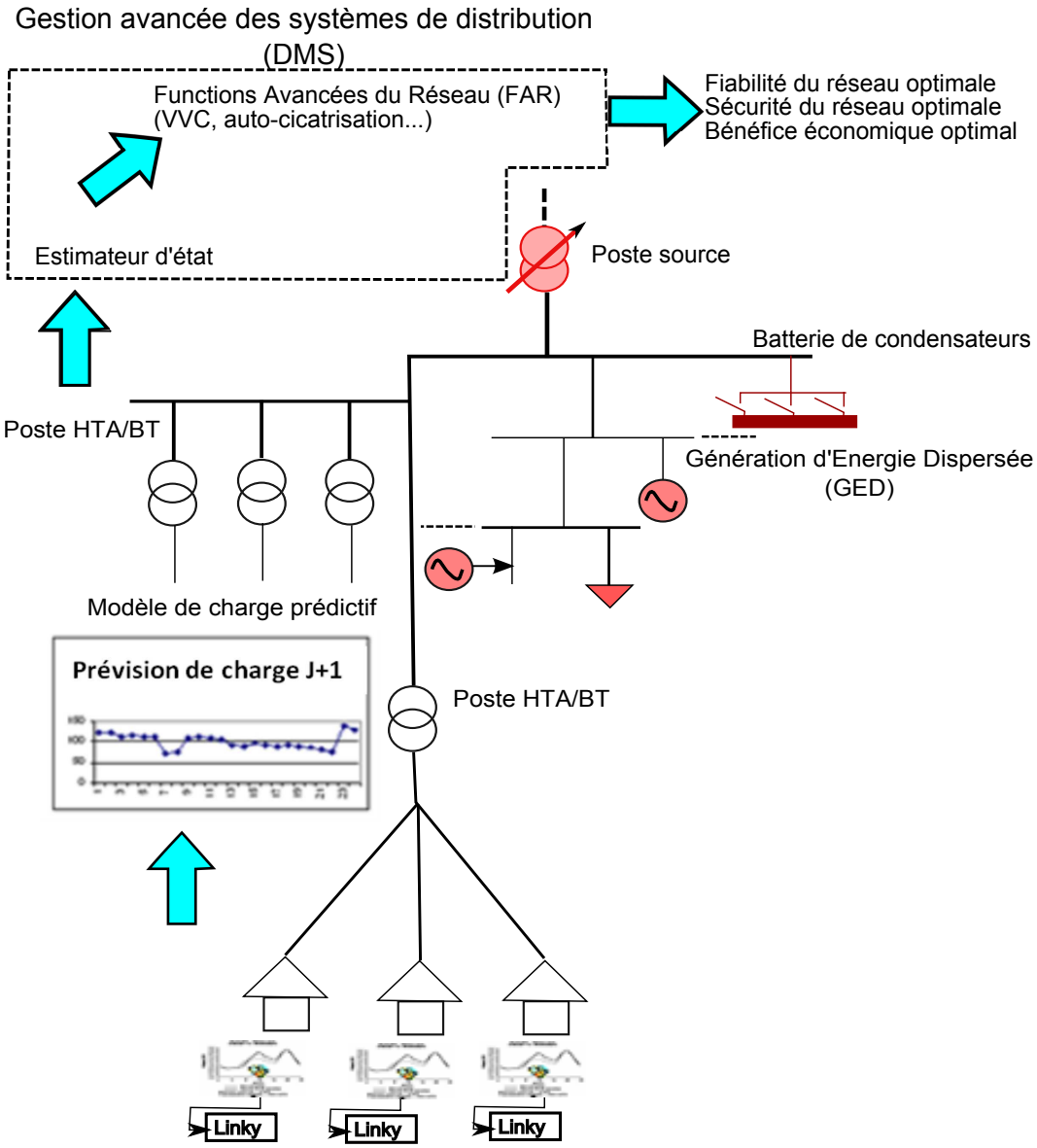


Figure E.1: Relation entre les modèles de charge prédictifs, l'estimateur d'état, et les fonctions avancées du réseau

réseau, tels que les batteries de condensateurs, les Générateurs Electriques Décentralisés (GEDs), les réglages en charge des transformateurs, les interrupteurs/sélecteurs, etc.

La planification d'un réseau fiable est un défi puisque cela signifie que ses clients alimentés doivent avoir une bonne fourniture d'électricité stable et continue. Les opérateurs des réseaux doivent garantir la tension de chaque client dans une échelle admissible. En Europe, pour les réseaux Basse Tension (BT), la tension admissible est définie à ±10% de la tension nominale. En dehors de ces limites, les clients sont considérés comme des « Clients Mals Alimentés » (CMAs).

Pour la planification, les calculs électriques sont effectués dans les conditions les plus critiques [7, 8]. Plus précisément, ces calculs sont réalisés pour deux situations : une

demande maximale de consommation électrique avec une production minimale et une production maximale avec une consommation minimale [9]. Avec une grande pénétration de GEDs, on peut envisager le deuxième cas. Etant donné les comportements variés des clients, dans une même zone géographique, les pics de demande ont rarement lieu au même moment. En tenant compte de ce fait, l'estimation de la charge du client heure par heure est nécessaire. Il faut aussi tenir compte de l'incertitude de l'estimation de charge [10]. Généralement, pour le calcul de la chute de tension, il faut prendre un risque de 10% [11]. C'est ce risque qui définit les seuils des puissances. Ces seuils sont utilisés ensuite pour identifier les CMAs. En conséquence, pour la planification du réseau, la puissance demandée par un client est équivalente à la somme de la journée type moyenne de ce client et de 10% de son risque de puissance.

Pour conclure, l'objectif est de définir les puissances limites maximales et minimales de l'année avec 10% de risque de dépassement.

Ce résumé est découpé en quatre parties. Tout d'abord, le contexte du réseau intelligent et des compteurs intelligents est mis en avant comme une opportunité pour développer des modèles de charge plus performants. Les objectifs ainsi que les contributions de la thèse sont soulignés. Ensuite, deux sections sont consacrées pour répondre aux deux objectifs de la thèse : la conception des modèles de charge pour la conduite et pour la planification du réseau de distribution. Les théories de chaque méthodologie sont détaillées et les résultats comparatifs sont commentés. A la fin, ce résumé est terminé par des conclusions et perspectives générales.

E.1.c Contribution de thèse

Les contributions de la thèse peuvent être résumées ci-dessous :

1. La prévision de charge est un sujet très étudié au niveau des réseaux de transport [12, 13, 14, 15, 16, 17, 18, 17, 19, 20]. Cependant, dans les réseaux de distribution et avec les caractéristiques de données considérées dans notre étude (quelques dizaines kW), à notre connaissance, il existe peu de travaux.

De notre point de vue, trois raisons peuvent être utilisées pour expliquer ce fait : Premièrement, une attention particulière a été apportée au réseau de transport, puisque celui-ci s'étend sur une distance plus grande, et couvre un territoire plus large. Le réseau de transport est la « colonne vertébrale » du système électrique. Deuxièmement, le volume de charge considéré dans le transport rend la forme de la courbe régulière, par un effet de foisonnement très important. Cette caractéristique de la courbe de charge la rend plus facile à prévoir à ce niveau. Troisièmement, il n'y a encore aujourd'hui pas ou peu de mesures disponibles sur les postes HTA/BT, sur lesquels les modèles de charge pour la distribution peuvent être conçus.

Dans ce projet de recherche, deux modèles sont proposés pour la prévision de charge des postes HTA/BT. L'un est basé sur les séries chronologiques et l'autre est basé sur les réseaux de neurones.

2. Les différents types de charge (résidentielle, commerciale, et industrielle) sont étudiés dans le cadre de la thèse et leurs différentes propriétés sont illustrées.

Les deux méthodes sont conçues et évaluées grâce aux données provenant du réseau de distribution français dans le cadre du projet « Linky ». Un modèle référentiel est établi pour la comparaison des modèles. Les avantages et les inconvénients des deux modèles sont soulignés dans la comparaison. Les deux méthodes sont alternatives et en même temps, complémentaires. Elles permettent de définir avec une précision limitée les caractéristiques intrinsèques des données de charge du poste.

3. La méthode basée sur la série chronologique est un travail original. Elle intègre de nombreux outils statistiques pour atteindre une meilleure précision. Le résidu du modèle série chronologique est examiné en détail pour assurer un bon comportement du modèle.

4. La méthode basée sur le réseau de neurones est inspirée par la procédure de sélection proposée par le Professeur Gérard Dreyfus, un spécialiste reconnu dans le domaine des réseaux de neurones. Nous nous focalisons sur la méthodologie de conception du modèle, qui est pour la première fois exploitée entièrement dans la prévision de charge à court terme.

5. Avant la mise en œuvre des compteurs intelligents, il n'avait pas de données historiques à part pour un nombre limité de clients. Dans le domaine de la prévision de charge, la plupart des travaux se concentre sur l'estimation de la demande de pointe pour un groupe de clients pendant la pointe du système, i.e., la demande coïncidente. Quelques travaux ont aussi été effectués sur la technique « end-use », méthode décomposant la courbe de charge d'un client résidentiel en unité d'appareil électrique. Les autres se concentrent sur la méthode de classification pour diviser les clients en différents groupes, et sur la présentation de chaque groupe avec un profil de charge typique. Dans le contexte des compteurs intelligents, nous sommes les premiers à proposer la conception du modèle de charge individuel dirigé par les données pour le besoin de la planification.

6. La relation entre la consommation électrique et la température est définie par les estimateurs non-paramétriques dans notre modèle d'estimation de charge individuel. La méthode est appliquée aux vraies courbes de charge individuelle en France. La performance comparée avec le modèle actuel d'électricité de France (EDF) s'intitule BAGHEERA dans différents cas d'étude.

E.2 Modèle de charge prédictif court terme pour la conduite et l'estimateur d'état

La prévision de charge joue un rôle important dans la prise de décision dans le système électrique. Cette partie est consacrée au modèle de charge prédictif. Dans un premier temps, nous présentons rapidement des méthodes appliquées à la prévision de charge, leurs spécificités ainsi que leurs applications. Ensuite, les données utilisées pour la conception et l'évaluation de nos méthodologies sont étudiées. Certains composants importants dans la courbe de charge pour sa modélisation sont précisés. Nos choix de méthodes pour la série chronologique et le réseau de neurones sont argumentés. Les critères de performance et un modèle de référence sont aussi établis. Puis, les deux modèles de charge prédictifs sont détaillés. Les applications sur les vraies données de charge sont illustrées, et une comparaison de ces deux méthodes est faite à la fin de cette partie.

E.2.a Méthodes de la prévision de charge dans la littérature

Différentes techniques et différentes entrées sont appliquées à la prévision de charge avec un horizon de temps varié. De nombreux facteurs tels que les conditions météo, les variables saisonnières, et les facteurs sociaux, économiques et démographiques font partis des entrées pour les techniques de prévision de charge [23]. Tableau E.1 résumé des applications et des facteurs d'influence pour les modèles avec différents horizons de temps [23, 24, 18].

Table E.1: Différents horizons de temps pour la prévision de charge

Horizon de temps	Applications	Facteurs d'influence
Très court terme (1 Min ~ 1h)	Fonctions avancées du réseau, le réglage fréquence-puissance	Charges historiques
Court terme (1h ~ 1 semaine)	Conduite (fonctions avancées du réseau), calculs de répartition de charge pour l'estimation, de l'estimation d'épargne pour l'économie et la sécurité opérationnelle du système électrique	Charges historiques, informations calendaires (journée type et heure), conditions météo ($*$)
Moyen terme (1 semaine ~ 1 an)	Négociation des contrats, planification de la fourniture de d'énergie primaire et travaux de maintenance	($*$) + population, facteurs économiques, etc. ($\diamond$)
Long terme (1 an ~ plusieurs ans)	Dépenses en capital et planification des investissements	($\diamond$) + plus d'informations tels que : augmentation de la population, produit intérieur brut

($*$) et ($\diamond$) représentent respectivement les facteurs d'influence pour la prévision à court terme et à moyen terme.

Pour notre application de la conduite du réseau, surtout pour coopérer avec les fonctions avancées du réseau, nous nous focalisons sur la prévision à court terme. Selon le tableau E.1, les conditions météo, le calendrier et les charges historiques sont des entrées essentielles.

La figure E.2 donne une vue globale en deux dimensions (horizon de temps et hiérarchie de la tension) sur les méthodes appliquées pour la prévision de charge. La plupart d'entre elles est développée pour le court terme et la haute tension. Cependant, ces dernières années, l'expansion des réseaux intelligents attire de plus en plus l'attention sur les réseaux moyenne tension et basse tension.

Généralement, les méthodes prédictives peuvent être classées en deux catégories : les approches classiques et les approches « intelligence artificielle ». La première catégorie exige un modèle avec des équations mathématiques qui interprètent la relation entre la charge et les autres facteurs d'influence. Cette famille inclue le modèle de régression, la série temporelle, les jours similaires, la méthode end-use et l'approche économétrique. Alors que la famille des méthodes d'intelligence artificielle exploite la relation non linaire entre

la charge et les entrées. Cette catégorie comprend les réseaux de neurones (RN), la logique floue et les systèmes experts.

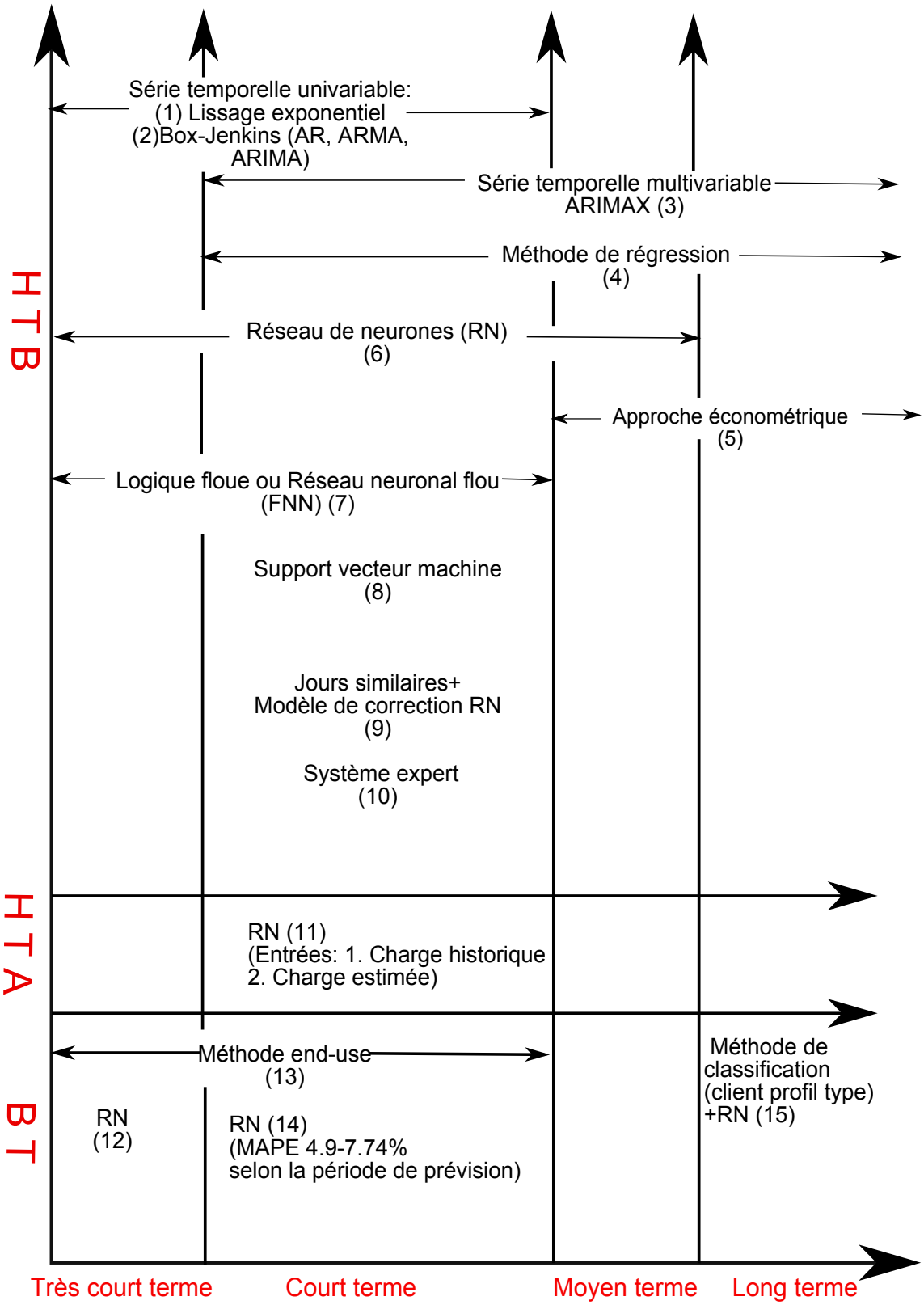


Figure E.2: Résumé des méthodes de charge prédictives en deux dimensions: horizon de temps et hiérarchie de la tension. HTB: haute tension et très haute tension, HTA: moyenne tension et BT: basse tension. Les numéros qui apparaissent dans la figure correspondent aux travaux liés.[a]

[a] (1):[13, 14] (2):[13, 15, 14] (3):[16] (4):[17, 18, 17] (5):[24] (6):[19, 20] (7):[32, 33, 34] (8):[35, 36] (9):[31] (10):[37, 38] (11):[39] (12):[22] (13):[40] (14):[41]

[b] Le modèle série temporelle univariable n'exploite que les informations de la courbe de charge historique.

[c] Le modèle série temporelle multivariable exploite non seulement les informations de la courbe de charge

Nous résumons trois points sur l'état de l'art des modèles de charge prédictifs:

1. Les méthodes d'intelligence artificielle deviennent de plus en plus populaires.
2. Des chercheurs essayent d'améliorer les méthodes classiques, telles que l'espace d'état et le filtre de Kalman, et l'algorithme de spline cubique, malgré les succès limités de ces méthodes. Ces méthodes améliorées peuvent parfois atteindre la même précision que les méthodes d'intelligence artificielle.
3. Les méthodes hybrides deviennent de plus en plus populaires également puisqu'elles combinent les avantages de plusieurs techniques. Par exemple, appliquer différents modèles pour modéliser séparément des parties linéaires et non linéaires, des fluctuations rapides et lentes; supprimer les données aberrantes ; et calculer les coefficients des modèles.

Les caractéristiques des modèles sont résumées dans le tableau E.2.

Table E.2: Résumé des modèles de charge prédictifs et leurs caractéristiques. « √ » signifie que la propriété correspond à l'attribut; « × » signifie que la propriété ne correspond pas à l'attribut.

Méthode	**Très court terme**	**Court terme**	**Moyen terme**	**Long terme**	**Linéaire**	**Historiques importantes**	**Caractéristiques**
Modèle de régression	×	√	√	√	×	√	Simple, relation explicite avec variables exogènes
Série temporelle	√	√	√	×	√	×	Extraire périodicités, données historiques et erreurs, inclure : Box-Jenkins, lissage exponentiel, etc.
Similar day approach	×	√	×	×	√	√	Non paramétrique, ajustement difficile des coefficients, fournir données d'apprentissage pour les méthodes d'intelligence artificielle
Méthode end-use	×	×	√	√	×	×	De bas en haut, chaque équipement électrique, exige des informations extensives

Suite à la page suivante

Table E.2 – Suite du tableau présenté page précédente

Méthode	**Très court terme**	**Court terme**	**Moyen terme**	**Long terme**	**Linéaire**	**Historiques importantes**	**Caractéristiques**
Approche économétrique	×	×	√	√	×	√	inclut les variables sociales et économiques, équations complexes
RN	√	√	√	×	×	√	Approximateur universel, bonne capacité d'apprendre, établit des relations complexes, nombreux choix pour la structure, temps de calcul important
SVM	√	√	√	×	×	√	Fonction non linaire transforme les entrées en un espace d'une plus grande dimension
Logic Floue (LF)	√	√	×	×	×	×	Fuzzification, défuzzification, dispensé du modèle mathématique, similarité, régression linéaire floue, estimation compliquée des paramètres
Système expert	√	√	×	×	×	√	Ordinateur assisté, règle Si-Alors, mises à jour fréquentes

E.2.b Description de données

Dans cette sous-section, nous examinons les courbes de charge des postes HTA/BT, et décrivons les caractéristiques de ces données. Ces spécificités vont nous aider à faire des choix parmi les nombreuses méthodes de prévision présentées précédemment.

Les données utilisées pour nos études sont des mesures des postes HTA/BT dans le réseau de distribution français. Plus concrètement, il s'agit de 7 courbes de charge de postes HTA/BT d'une même zone géographique échantillonnées par points 30 minutes. La période de mesure est du 9/9/2009 au 27/10/2010. Etant un des facteurs d'influence, les données de température sont aussi fournies heure par heure. Ces données de températures sont interpolées linéairement pour être sur la même fréquence que les données de la charge (points 30 minutes). Figure E.3 montre la variation journalière de la charge et de la température pendant la période de mesure. Chaque courbe de charge d'un poste HTA/BT

représente la somme des consommations des clients connectés en aval.

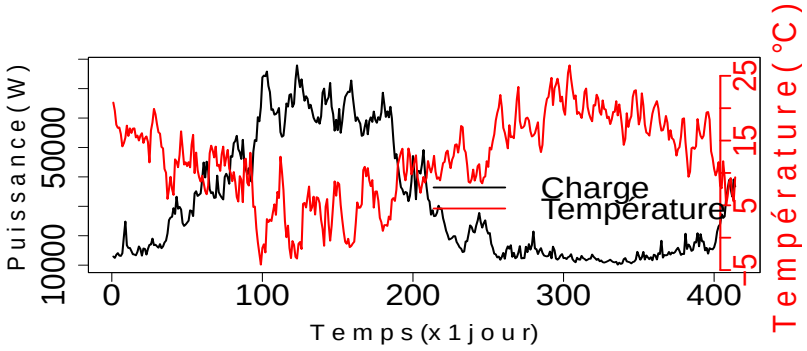


Figure E.3: Courbe de charge et température journalière pendant 414 jours (du 9/9/2009 au 27/10/2010) du poste HTA/BT *CE_MOU* (connecté principalement à des clients résidentiels)

Les figures E.4 et E.5 montrent deux autres exemples de courbes de charge journalières de postes HTA/BT. Le poste *CE_MOU* (figure E.3) alimente 61% de clients résidentiels, 23% de clients tertiaires, et 16% de clients industriels. Le pourcentage représente la puissance souscrite totale des clients dans cette catégorie par rapport à la puissance totale souscrite en aval de ce poste HTA/BT. La courbe de charge du poste *CE_MOU* suit l'évolution de la température. Le poste *VI_LOG* (figure E.4) contient un tiers de clients tertiaires et deux tiers de clients industriels. Sa courbe a une périodicité hebdomadaire et suit également la variation de la température. La courbe de charge du poste *CE_FRO* (figure E.5), qui n'alimente qu'un client industriel, reste stable toute année et possède une périodicité hebdomadaire.

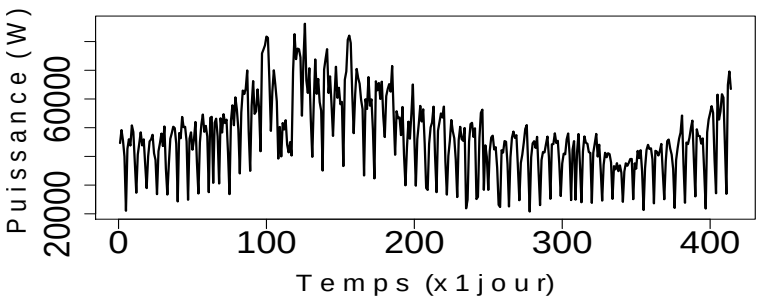


Figure E.4: Courbe de charge journalière pendant 414 jours (du 9/9/2009 au 27/10/2010) du poste HTA/BT *VI_LOG* (connecté aux clients mixtes tertiaires et industriels)

En analysant ces données, nous concluons que la courbe de charge d'un poste HTA/BT dépend principalement de la composition des clients connectés. Les courbes de charge des clients résidentiels, tertiaires, et industriels ont leurs propres spécificités:

- Une courbe de charge de type résidentiel varie avec la température : le niveau de consommation d'électricité augmente quand *la température* chute. Cela est dû à

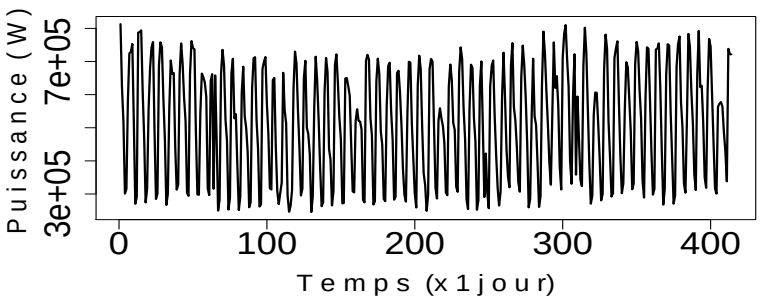


Figure E.5: Courbe de charge journalière pendant 414 jours (du 9/9/2009 au 27/10/2010) du poste HTA/BT *CE_FRO* (connecté à un seul client industriel)

l'utilisation du chauffage électrique.

- Une courbe de charege de type industriel ne dépend pas de la saison ni de la température. Cependant, elle a une *périodicité hebdomadaire* claire et est sensible aux *jours fériés*.

- Une courbe de charge de type tertiaire se situe entre les deux premiers types : sa courbe de charge varie avec *la température* et elle a une *périodicité hebdomadaire*.

Un poste HTA/BT est souvent connecté à un mix de catégories de clients. Il dépend donc la plupart du temps de facteurs tels que *la température*, *le calendrier* (le type de jour), et *le temps* (les périodicités, l'heure, etc.).

Afin de pouvoir tester la pertinence des modèles de charge prédictifs, 4 courbes de charge mesurées en tête de départs HTA (un niveau d'agrégation plus élevé, au niveau du poste source HTB/HTA) sont aussi utilisées pour l'étude des méthodes développées. La période de mesure est du 9/9/2009 au 22/9/2010.

E.2.c Choix des méthodes: série chronologique et réseau de neurones

Les méthodes présentées dans la sous-section E.2.a ont été appliquées aux différents cas de prévision de charge avec plus ou moins de succès. Selon l'objectif, notre intérêt se trouve dans la prévision de charge à court terme au niveau du poste HTA/BT. Nous percevons dans la figure E.2 que la plupart des modèles peuvent être appliqués au niveau HTB. L'objectif est d'emprunter ces méthodes appliquées au niveau HTB et de les adapter aux niveaux HTA et BT. Comparées aux données HTB, les données HTA et BT sont plus irrégulières car l'effet de foisonnement est moindre.

Le choix des méthodes est issu du résultat des analyses de la nature des données [86]. Dans la section E.2.b, nous avons conclu que la courbe de charge d'un poste HTA/BT dépend de la nature des clients connectés, de la température, de la météo et du calendrier. Ainsi, des deux catégories présentées, nous avons choisi la série chronologique et le réseau de neurones pour modéliser nos charges. Nous formulons le modèle série chronologique (différent de la méthode Box-Jenkins) avec une régression, dans laquelle on trouve une interprétation physique et une explication attachée à chaque composant. Les outils avancés

du traitement du signal sont intégrés dans cette méthode pour avoir la meilleure précision possible. L'autre méthode choisie, venant de la famille « intelligence artificielle » est le réseau de neurones, plus concrètement, la structure Perceptron MultiCouche (PMC) [21]. C'est un réseau type « feed forward », le plus simple à mettre en œuvre et dont le calcul est le moins couteux. De plus, A. Khotanzad et al. arrivent à la conclusion [20] que selon leurs investigations sur les différentes structures des réseaux de neurones, il y n'a pas d'avantage évident dans l'utilisation d'une structure plus compliquée. Contrairement aux applications trouvées dans la littérature, nos efforts se sont concentrés sur la sélection de la structure optimale d'un PMC, puisque le réseau de neurones est souvent critiqué pour sa complexité et le manque de la validité de son modèle [87]. De plus, les deux méthodes choisies peuvent facilement estimer l'intervalle de la prévision. Cette dernière donne une indication sur l'incertitude du résultat.

E.2.d Critères de performance et modèle de référence

MAPE et MAE sont choisis comme les critères de performance, puisqu'ils sont les plus utilisés dans le domaine de la prévision de charge. Ils sont les bases de comparaison pour les modèles conçus.

MAPE représente l'erreur prévisionnelle absolue relative à la grandeur mesurée. Le résultat est donné en pourcentage :

$$MAPE(\%) = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{y_t - \hat{y}_t}{y_t}\right| * 100 = \frac{1}{N}\sum_{t=1}^{N}\left|\frac{e_t}{y_t}\right| * 100 \tag{E.1}$$

Où N est le nombre totale des échantillons de la prévision, $\hat{y}_t$ est la puissance prévisionnelle, y_t est la mesure, et e_t est l'erreur du modèle, qui représente la différence entre la puissance prévisionnelle et la mesure.

MAPE représente une erreur relative, ce qui signifie que sa valeur dépend de la grandeur de la mesure. Autrement dit, ayant les mêmes valeurs d'erreur, pendant l'hiver, quand la consommation électrique est élevée, l'indice MAPE est petit ; et pendant l'été, quand la consommation électrique est basse, le MAPE est grand. Pour corriger ce biais, nous avons introduit un autre indice, MAE.

MAE correspond à une erreur absolue en unité Watt :

$$MAE(W) = \frac{1}{N}\sum_{t=1}^{N}|y_t - \hat{y}_t| = \frac{1}{N}\sum_{t=1}^{N}|e_t| \tag{E.2}$$

Pour avoir un premier repère de performance, nous avons établi un modèle de référence intitulé modèle naïf [12, 60]. Ce modèle n'exploite pas les informations exogènes telles que la température et le type de jour, Il n'utilise pas non plus d'outils mathématiques complexes. Il remplace simplement le résultat de la prévision (Jour J) par son historique (Jour J- ℓ) le plus similaire du jour de la prévision, où ℓ est le décalage général entre les jours les plus similaires.

D'après nos exemples, le modèle naïf d'un poste HTA/BT se comporte de deux manières différentes selon la composition des clients. Pour un poste principalement constitué de clients résidentiels, sa courbe de charge varie avec la température, et son modèle naïf se

base donc sur la veille pour la prévision (ℓ=1). Etant donné qu'un poste industriel a une évidente cyclicité hebdomadaire, le même jour de la semaine dernière (ℓ=7) est utilisé comme prévision dans le modèle naïf.

Dans cette première partie, nous avons choisi les méthodes, les critères de performance ainsi que le modèle de référence. Dans les prochaines deux sous-sections, nous allons détailler ces modèles de prévision de charge basés indépendamment sur la série chronologique et sur le réseau de neurones. Différents aspects des modèles sont aussi comparés et développés dans les sous-sections.

E.2.e Modèle série chronologique

La série chronologique représente l'évolution d'un set d'observations échantillonnées dans un intervalle régulier de temps. La spécificité de la série chronologique par rapport aux autres méthodes statistiques est qu'elle introduit le temps comme une de ses variables explicatives.

Nous avons choisi la série chronologique additive qui contient trois parties : une tendance, une périodicité et une erreur aléatoire. Supposons que la mesure de la puissance à temps t est y_t, le modèle est :

$$y_t = f_t + S_t + \epsilon_t \tag{E.3}$$

Où f_t représente la composante de tendance, S_t représente la composante cyclique et ϵ représente l'erreur aléatoire au temps t. Les deux premières parties sont déterministes et à partir d'elles, deux modèles paramétriques sont conçus.

Figure E.6 donne un aperçu de la procédure de la prévision de charge avec le modèle série chronologique. Nous allons ensuite décrire les outils mathématiques qui ont servis dans cette procédure.

1. Régression des variables catégoriques

   Nous avons conclu dans la section E.2.b que l'un des facteurs d'influence les plus importants qui expliquent la variation dans la courbe de charge est le type du jour. Dans nos courbes exemples, nous avons distingué trois types : « jour ouvrable », « samedi », et « dimanche et jour férié ». Des variables catégorielles sont utilisées pour intégrer ce type d'information dans le modèle. Ces variables prennent des valeurs 0 ou 1, et indiquent la présence de la partie correspondante dans l'équation. La régression des variables catégorielles s'écrit :

   $$y_t = \Gamma(x_t) + \sum_{\alpha=1}^{\kappa-1} D_\alpha \Gamma_\alpha(x_t) \tag{E.4}$$

   Où $D_\alpha, \alpha = 1, \cdots, \kappa - 1$ sont les variables catégorielles, $\Gamma(\cdot)$ est une régression, $\Gamma_\alpha(\cdot)$ est la régression associée aux variables catégorielles D_α, κ est le nombre de catégories et x_t, y_t représentent indépendamment le vecteur des variables indépendantes et la variable dépendante.

   Le nombre de variables catégorielles dans l'équation est $\kappa - 1$, puisque la catégorie où toutes les variables catégorielles sont égales à zéro est considérée comme la référence des autres catégories. Pour une catégorie donnée, il y a au maximum une variable

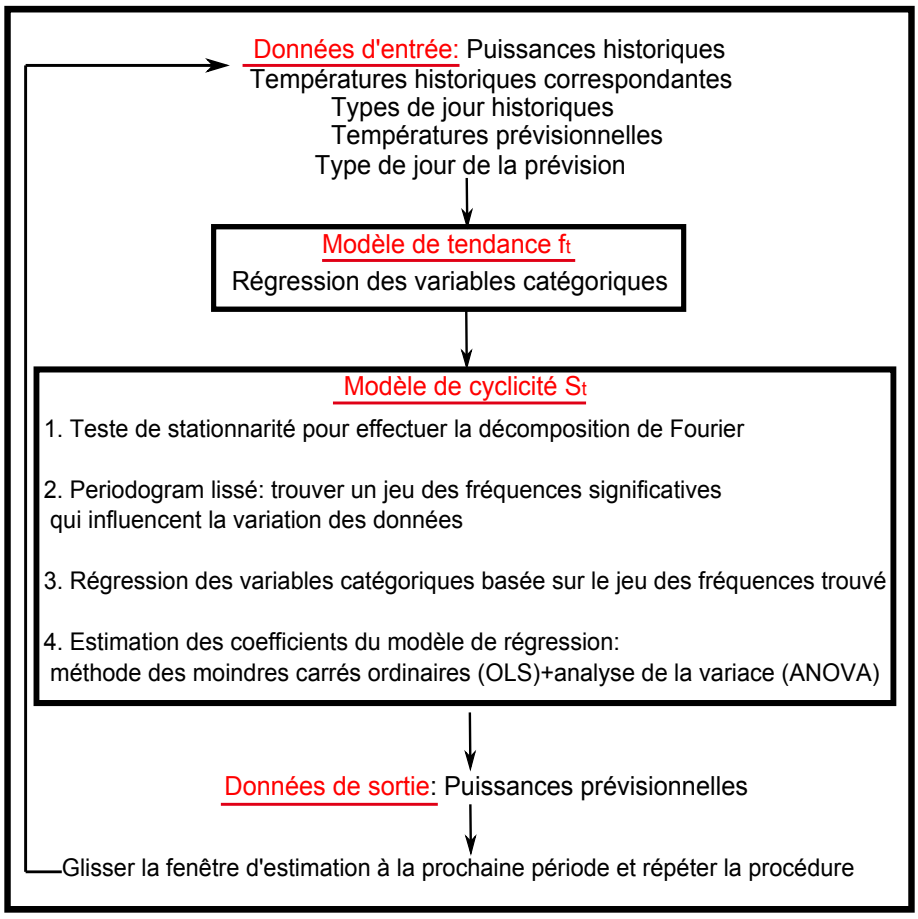


Figure E.6: Etapes pour construire le modèle série chronologique pour la prévision de charge

catégorielle égale à 1. La régression des variables catégorielles est appliquée à la composante de tendance ainsi qu'à la composante cyclique.

2. Estimation de la composante de tendance

La fonction de tendance f_t représente une variation lente de la variable étudiée y_t. Elle est souvent représentée par une fonction linéaire, polynomiale ou exponentielle. Dans nos exemples, la courbe de charge est linéairement dépendante de la température (tableau 2.3). En tant que première approche, nous avons donc pris une fonction de tendance linéaire. Nous suggérons une fonction linéaire corrélée au temps et à la température :

$$f_t = at + b + cT_t + \sum_{\alpha=1}^{\kappa-1} D_\alpha \gamma_\alpha \tag{E.5}$$

Où t correspond au temps, dont la valeur est de 1 à la taille de la fenêtre d'estimation. T_t correspond à la température de la région où les mesures sont faites. $\sum_{\alpha=1}^{\kappa-1} D_\alpha \gamma_\alpha$ correspond à la partie de la régression catégorielle.

Les $\kappa+2$ coefficients de la régression $\{a, b, c, \gamma_\alpha\}$ sont estimés avec le critère des moindres carrés. La somme de la partie cyclique pendant une période est supposée être 0. Nous avons adopté la stratégie de la fenêtre glissante pour notre problématique de prévision court terme. Les données dans la fenêtre glissante sont utilisées pour estimer les coefficients. Etant donné que les coefficients sont estimés, et qu'on connait la journée type du jour prévisionnel, ainsi que la température prévisionnelle, nous pouvons calculer la prévision de la partie tendance. Ensuite, la fenêtre va être glissée d'une journée pour prévoir en « J+1 », et les coefficients vont être recalculés dans cette fenêtre. Puisque nous n'avons pas les températures prévisionnelles des postes HTA/BT, nous appliquons les températures réalisées pour cette étude. Cette approximation est souvent utilisé [26, 27, 28]. La section 3.4 est consacrée à l'analyse de l'impact de cette incertitude sur la température prévisionnelle pour la précision de la prévision.

3. Estimation de la composante cyclique

Dans cette section, nous complétons le modèle de prévision en ajoutant la partie cyclique. La partie cyclique décrit les comportements cycliques de la charge y_t. Supposons que p est la période des données sans tendance, pour chaque observation, nous avons $S_{t+p} = S_t$, où S_t représente la composante cyclique à t. La somme des composantes cycliques pendant une période complète est 0 : $\sum_{t=1}^{p} S_t = 0$.

Après avoir déterminé le modèle de tendance, nous le retirons des données originales y_t :

$$W_t = y_t - f_t = S_t + \epsilon_t \tag{E.6}$$

Où W_t représente la série après avoir retiré la tendance.

La somme des fonctions sinus et cosinus peut très bien décrire le dynamisme des signaux périodiques stationnaires. Nous suggérons de modéliser la partie cyclique avec une régression des composantes de Fourier. Quatre étapes sont nécessaires pour construire le modèle cyclique (figure E.6). Premièrement, nous testons la stationnarité de la série après avoir retiré la tendance W_t. Ensuite, un périodogramme est effectué pour identifier les fréquences harmoniques qui expliquent les variations dans la série W_t. Puis, avec ces fréquences, nous établissons une régression avec les composantes de Fourier. Enfin, Nous intégrons le test de nullité ANOVA dans l'estimation des coefficients de la partie cyclique avec le critère des moindres carrés. Le test de nullité ANOVA permet de discriminer des coefficients de régression significatifs.

4. Tests de stationnarité

Le but d'effectuer les tests de stationnarité est de vérifier si la série, après avoir retiré la tendance W_t, est vraiment stationnaire puis d'appliquer un périodogramme lissé. Parmi les approches existantes, nous en avons choisi deux : le test KPSS [90] et le test ADF [91]. Le KPSS teste une hypothèse nulle de la stationnarité de la série contre la non stationnarité. D'un autre côté, le ADF teste la présence d'une racine unitaire, ce qui signifie la non stationnarité. Ces deux tests sont utilisés conjointement pour

assurer un résultat fiable. Plus de détails sur cette méthode sont fournis dans annexe B.

Les résultats des deux tests montrent que notre série (après avoir retiré la tendance W_t) est stationnaire avec n'importe quelle taille de fenêtre d'estimation.

5. Périodogramme lissé

Le périodogramme lissé est appliqué pour identifier les fréquences harmoniques principales dans la composante cyclique. Ces fréquences sont ensuite utilisées pour construire la régression de la composante cyclique. Pour une série stationnaire, son périodogramme est défini [88]:

$$I_n(\nu_j) = \frac{1}{n}\left|\sum_{t=1}^{n} W_t e^{-2i\pi\nu_j t}\right|^2 = d_c^2(\nu_j) + d_s^2(\nu_j) \tag{E.7}$$

Où n est la taille d'échantillons, $\nu_j = \frac{j}{n}$, $j = \{0, 1, \cdots, n-1\}$ sont des fréquences harmoniques de Transformée de Fourier Discrète (TFD) de la série W_t . $d_c(\nu_j)$ et $d_s(\nu_j)$ sont des composantes normalisées réelles et imaginaires de la transformée :

$$\begin{aligned} d_c(\nu_j) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} W_t \cos(2\pi t\nu_j) \\ d_s(\nu_j) &= \frac{1}{\sqrt{n}}\sum_{t=1}^{n} W_t \sin(2\pi t\nu_j) \end{aligned} \tag{E.8}$$

Nous avons :

$$\sum_{t=1}^{n}(W_t - \overline{W})^2 = 2\sum_{j=1}^{m}[d_c^2(\nu_j) + d_s^2(\nu_j)] = 2\sum_{j=1}^{m} I(\nu_j) \tag{E.9}$$

Où $m = \frac{n-1}{2}$, et $\overline{W}$ correspond à la valeur moyenne de la série W_t. Equation E.9 indique que la somme des carrés peut être partitionnée en composants harmoniques représentés par l'amplitude du périodogramme à fréquence ν_j. Autrement dit, s'il existe des pics dans le périodogramme, ces fréquences expliquent la variation des données.

Néanmoins, le périodogramme brut a souvent une grande variance à une fréquence donnée. Cette forme est rarement utilisée directement comme un estimateur dans la fonction de la densité spectrale. La solution pour réduire la variance de l'estimateur est d'employer un périodogramme lissé.

La densité spectrale est supposée être constante dans une bande de fréquence, et les fréquences adjacentes sont indépendantes asymptotiquement. Nous établissons un filtre glissant symétrique B de taille $2l+1 << n$, qui est centré autour d'une fréquence ν_j :

$$B = \left\{\nu : \nu_j - \frac{l}{n} \le \nu \le \nu_j + \frac{l}{n}\right\} \tag{E.10}$$

Le noyau Daniell [92] est un jeu de poids positifs symétriques centrés autour des fréquences estimées. La somme des poids est 1 :

$$h_k = h_{-k} > 0 \; et \sum_{k=-l}^{l} h_k = 1 \tag{E.11}$$

En conséquence, le périodogramme lissé devient :

$$\widetilde{I}(\nu_j) = \sum_{k=-l}^{l} h_k I_n(\nu_j + k/n) \tag{E.12}$$

Au travers de la présentation graphique du périodogramme lissé, les fréquences $\mathcal{F}$ avec les amplitudes les plus significatives sont identifiées.

6. Modèle de régression avec composantes de Fourier

   Nous construisons la composante cyclique basée sur le jeu de fréquences $\mathcal{F} = \{\nu_1, \nu_2, \ldots, \nu_{N_f}\}$ tel que:

$$\begin{aligned} S_t = &\sum_{i=1}^{N_f} c_i \cos(2\pi\nu_i t) + \sum_{i=1}^{N_f} s_i \sin(2\pi\nu_i t) + \\ &\sum_{\alpha=1}^{\kappa-1} D_\alpha \left(\sum_{i=1}^{N_f} c_{i,\alpha} \cos(2\pi\nu_i t) + \sum_{i=1}^{N_f} s_{i,\alpha} \sin(2\pi\nu_i t) \right) \end{aligned} \tag{E.13}$$

   Où N_f est le nombre total de fréquences dans le jeu $\mathcal{F}$, et $\sum_{\alpha=1}^{\kappa-1} D_\alpha \times (\cdot)$ correspond à la partie régression des variables catégorielles.

   Comme décrit dans l'équation E.13, chaque fréquence dans le jeu $\mathcal{F}$ a deux contributions : une composante sinus et une composante cosinus. Les $2\kappa N_f$ coefficients inconnus dans l'équation sont déterminés dans une fenêtre glissante avec les critères des moindres carrés.

7. Test de nullité ANOVA

   Le test de nullité ANOVA est effectué pour déterminer la pertinence des coefficients, et améliorer l'efficacité de l'estimation des coefficients.

   Le test ANOVA partitionne la variance observée en plusieurs variances des variables explicatives et celle du résidu. L'importance de chaque coefficient est identifiée par son rapport entre l'intra-groupe variance et la variance globale. Plus ce rapport est grand, plus le coefficient concerné est important. Pour plus de détails, les lecteurs peuvent consulter l'annexe C.

   En pratique, nous commençons le test de nullité ANOVA avec toutes les parties du modèle de régression. Ensuite, selon le résultat du test, nous retirons de l'équation des parties non significatives, les coefficients restants sont réestimés.

8. Modèle de prévision complet Le résultat de la prévision finale est la somme des prévisions de la partie tendance et de la partie cyclique. Le modèle complet est

exprimé de la façon suivante :

$$\begin{aligned}\hat{y}_t &= f_t + S_t \\ &= at + b + cT_t + \sum_{i=1}^{N_f} c_i \cos(2\pi\nu_i t) + \sum_{i=1}^{N_f} s_i \sin(2\pi\nu_i t) + \\ &\sum_{\alpha=1}^{\kappa-1} D_\alpha \left(\gamma_\alpha + \sum_{i=1}^{N_f} c_{i,\alpha} \cos\left(2\pi\nu_i t\right) \sum_{i=1}^{N_f} s_{i,\alpha} \sin\left(2\pi\nu_i t\right) \right)\end{aligned} \tag{E.14}$$

La partie erreur est: $\epsilon_t = y_t - f_t - S_t$.

Pour les résultats des applications, les lecteurs sont invités à lire les chapitres correspondants en anglais.

E.2.f Modèle réseau de neurones

Nous proposons dans cette partie de résoudre le problème de la prévision de charge avec les réseaux de neurones. Un avantage d'utiliser cet outil est qu'il est efficace même si les connaissances à priori sur le processus ne sont pas requises ou pas précisées.

Dans ces travaux, nous proposons d'exploiter nos connaissances dans les consommations historiques et les variables exogènes pour prévoir la consommation. Autrement-dit, nous utilisons les réseaux de neurones pour construire une relation non-linéaire entre les valeurs de la consommation historique, les variables exogènes et la future valeur de la consommation. Cette relation doit être vraie non seulement pour les données avec lesquelles elle est construite, mais aussi peut être généralisée pour les données fraiches (« données de test », qui ne sont jamais utilisées pour construire la relation). Ce problème peut être découpé en deux parties :

1. Trouver les variables endogènes et exogènes qui sont pertinentes pour la prévision de la consommation. C'est ce qu'on appelle *la sélection des variables.*

2. Trouver la complexité appropriée pour le modèle compte tenu des données disponibles. C'est ce qu'on appelle *la sélection du modèle.*

Dans nos travaux, la sélection des variables est effectuée par la méthode de la variable sonde et la sélection du modèle est basée sur le score de Leave-One-Out Virtuel (LOOV), un estimateur empirique de la généralisation d'erreur et l'état de la matrice jacobienne du réseau [21]. Le score LOOV exige le calcul du levier des données d'apprentissage. Le levier représente l'influence de chaque exemple d'apprentissage pour le modèle. La matrice jacobienne examine l'efficacité des modèles de différentes complexités. Ces méthodes vont être détaillées dans la partie suivante.

Ici, nous n'allons pas présenter la structure d'un Perceptron MultiCouche (PMC) et son processus d'apprentissage. Pour plus d'information, nous inviterons les lecteurs à se référencer à la section 4.2.

E.2.f-i Conception du modèle

L'objectif de la conception est de concevoir un modèle qui a une meilleur capacité de généralisation. Cette tâche contient une recherche de la complexité optimale du modèle, puisque le nombre de paramètres est linéaire par rapport au nombre de variables du modèle et au nombre de neurones cachés. La procédure de la conception doit rejeter les variables candidates non pertinentes à la prévision et définir le nombre de neurones sur la couche cachée. Cette procédure est présentée en deux étapes dans la partie qui suit: la sélection des variables et la sélection du modèle.

(1) Sélection des variables

La méthode pour la sélection des variables est la suivante : afin de distinguer les variables pertinentes et non pertinentes, une série de variables aléatoires nommées « variables sondes » sont créées à partir du vecteur des variables candidates. Elles sont générées en mélangeant aléatoirement des composantes dans le vecteur de la variable candidate. Les variables sondes vont être classées avec les variables candidates dans un ordre décroissant de pertinence par une régression orthogonale. La probabilité cumulative du rang des variables sondes est estimée et un seuil de rang r_0 est déterminé de façon à ce que la probabilité pour une variable non pertinente sélectionnée soit plus petite que le seuil r_0 choisit par le concepteur :

$$p(r_{probe} < r_0) < \delta \tag{E.15}$$

Où r_{probe} est le rang d'une variable sonde. En conséquence, δ est le risque de choisir une variable même si elle est non pertinente.

Notons ξ_i le vecteur dont les composants sont N valeurs mesurées du ième variable $(i = 1, \cdots, n)$, et y le vecteur dont les composants sont N valeurs mesurées de la variable à prévoir. Si les variables ont une moyenne égale à zéro, le carré du coefficient de la corrélation entre la ième variable candidate et la variable à prévoir est [21]:

$$\cos^2 \varphi_i = \frac{(\xi_i^T y)^2}{(\xi_i^T \xi_i)(y^T y)} \tag{E.16}$$

Où φ_i est l'angle entre vecteurs ξ_i et y dans l'espace d'observation. Plus φ_i est petit, plus grande est la corrélation entre la variable candidate et la variable à prévoir.

La procédure de classement basée sur la projection orthogonale de Gram-Schmidt est décrite dans la figure E.7. La variable candidate dont le vecteur ξ a le plus petit angle avec le vecteur y est classée en premier. Les vecteurs des autres variables candidates et le vecteur étudié y sont projetés sur l'espace nulle de la variable classée pour éliminer sa contribution. Le même calcul continue dans cet espace. Cette procédure est répétée jusqu'à ce que toutes les variables candidates soient classées ou qu'un critère d'arrêt soit atteint. Afin de tenir compte de la non linéarité, les variables candidates contiennent les variables primaires et leurs produits vectoriels.

La prochaine étape est la définition du seuil du rang tel que toutes les variables classées derrière soient rejetées. n_p vecteurs des variables sondes sont générés et la distribution cumulative du rang est estimée de la manière suivante : la probabilité estimée que le rang de la variable sonde soit supérieur ou égal au rang r est le rapport n_{rp}/n_p, où n_{rp} est le nombre de variables sondes dont le rang est inférieur ou égal à r. Pendant la procédure de

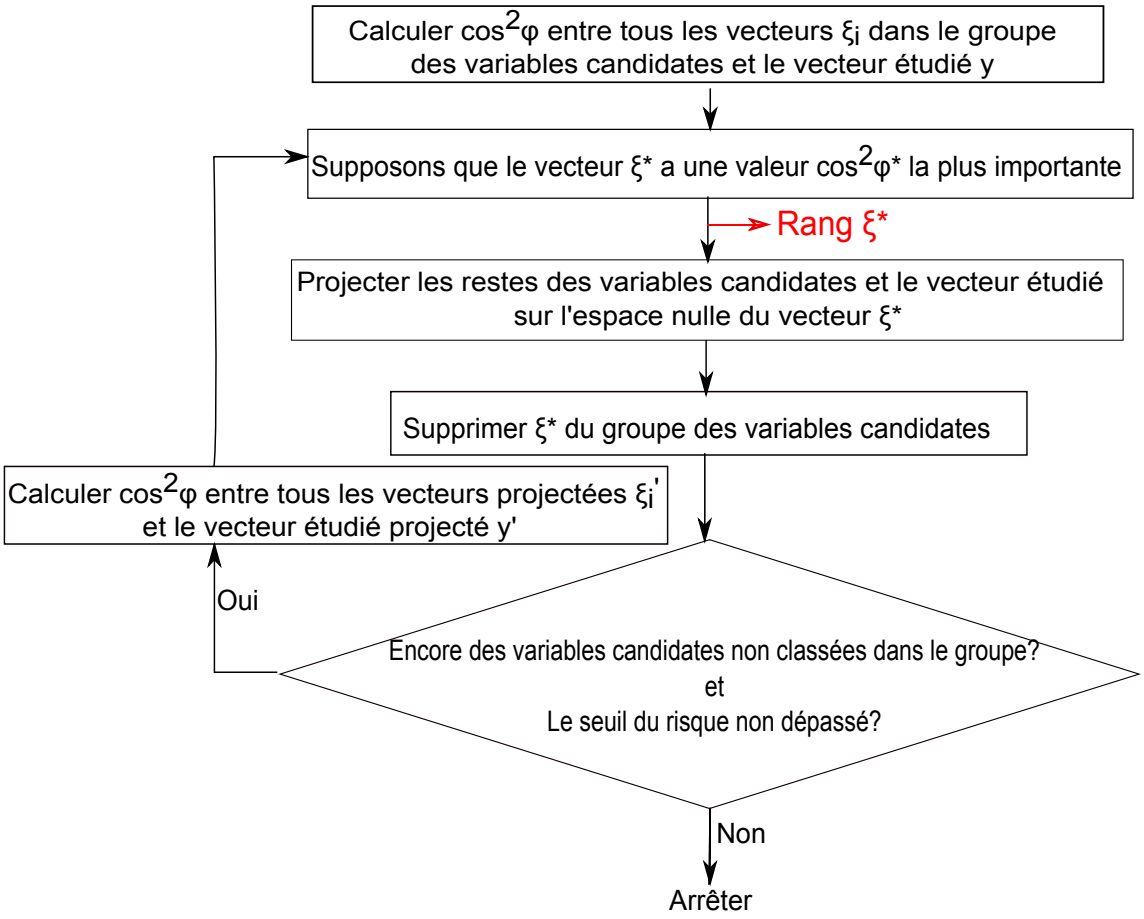


Figure E.7: Procédure du classement par la projection orthogonale de Gram-Schmidt

classement, quand le rang r est atteint tel que $n_{rp}/n_p > \delta$, la procédure s'arrête et le rang r_0 est fixé à $r-1$.

(2) Sélection du modèle

Puisque la fonction objectif appliquée pour l'apprentissage des réseaux de neurones n'est pas quadratique par rapport aux paramètres du modèle, elle a des minimums locaux. Les algorithmes d'optimisation sont itératifs et les initialisations de leurs paramètres sont nécessaires. Ces valeurs sont souvent aléatoirement générées par une distribution de probabilité avec une moyenne nulle et une variance $1/R$ [21]. La valeur minimale des algorithmes d'optimisation dépend de la valeur initiale des paramètres. En conséquence, pour un nombre fixe de neurones cachés, nous pouvons obtenir différents modèles. Chaque modèle correspond à un minimum local de la fonction objectif. Il n'y a pas de théorie qui montre que le modèle obtenu par le minimum global de la fonction objectif généralise mieux que les modèles obtenus par des minimums locaux. Par conséquent, nous proposons la stratégie décrite dans la figure E.8 : le nombre de neurones cachés augmente à partir de zéro (modèle linéaire) jusqu'à un nombre maximal (typiquement plus petit que 10 dans la plupart des applications); pour chaque complexité donnée, plusieurs modèles sont entrainés avec des initialisations des paramètres différentes. Le modèle avec la meilleure capacité de généralisation est sélectionné. En augmentant le nombre de neurones cachés, si la capacité de généralisation du meilleur modèle avec i neurones cachés se dégrade de

façon importante, la procédure s'arrête. La complexité avec i-1 neurones cachés est donc considérée optimale avec les données disponibles.

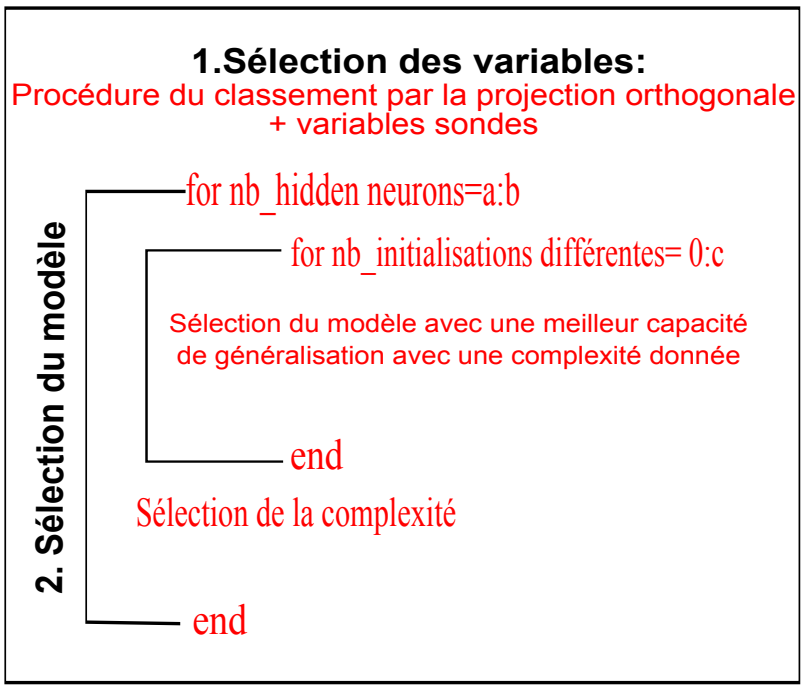


Figure E.8: Procédure de sélection du réseau de neurones ({a,b}:{min, max} nombre de neurones cachés; c: nombre d'initialisations maximal)

Leave-One-Out (LOO) (connu aussi sous le nom « Jackknife ») est un estimateur non biaisé de l'erreur de la généralisation du modèle [68]. Le concept consiste à retirer un exemple des données disponibles, faire l'apprentissage sur le reste des $N-1$ exemples, et ensuite calculer l'erreur de la prévision sur l'exemple retenu. Cette procédure est répétée jusqu'à ce que tous les exemples aient été retirés une fois des données d'apprentissage. Le score de LOO est :

$$E_{LOO} = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(y_i - f^{-i}(P_i,\Omega))^2} \tag{E.17}$$

$f^{-i}(P_i,\Omega)$ est le modèle obtenu quand l'exemple i est retiré des données d'apprentissage. Si le nombre d'exemple N est grand, cette procédure devient très lourde, puisque le nombre de modèles est égal au nombre d'exemples. Néanmoins, si le modèle est linéaire par rapport à ses paramètres, le score de LOO est obtenu par l'apprentissage d'un modèle de toutes les données d'apprentissage et estimé par PRESS [112] :

$$E_p = \sqrt{\frac{1}{N}\sum_{i=1}^{N}(\frac{y_i - f(P_i,\Omega)}{1-h_{ii}})^2} \tag{E.18}$$

où $f(P_i,\Omega)$ est le modèle obtenu avec tous les exemples et h_{ii} est le i-ième élément de la matrice $H = X(X^TX)^{-1}X^T$, où X est la matrice d'observation, i.e., la matrice de taille (N,p), dont les éléments $x_{i,j}$ sont les valeurs mesurées de la variable j dans l'exemple i et p est le nombre des paramètres, $p = (R+1)M + (M+1)$ (cf. figure 2.3). h_{ii} est le levier

de l'exemple i. Si la matrice X est de rang plein, H est la projection orthogonale sur le sous espace défini par les colonnes de la matrice X. En conséquence, les leviers ont les propriétés suivantes : $0 \leq h_{ii} \leq 1$; $\forall i \ \sum_{i=1}^{N} h_{ii} = p$.

Le levier d'exemple i peut être considéré comme la proportion des paramètres du modèle qui est dédiée pour calibrer le modèle avec l'exemple i.

Virtual Leave-One-Out (VLOO) [113, 114] est une généralisation du PRESS. Il est différent de PRESS en deux aspects :

- E_p est une approximation de score LOO E_{LOO}
- les leviers sont des éléments diagonaux de la matrice $H = Z(Z^T Z)^{-1} Z^T$, où Z est la matrice jacobienne du modèle

La procédure de sélection du modèle est décrite dans figure E.8 : on commence par un modèle linéaire (avec zéro neurone caché), le nombre de neurones cachés augmente un par un. Pour chaque complexité (nombre de neurones cachés) donnée, différents modèles sont calculés avec différentes valeurs initiales des paramètres ; le score de VLOO de chaque modèle est évalué et le modèle avec le plus petit score VLOO est choisi. Quand on atteint un certain nombre de neurones cachés, le score VLOO commence à s'accroître considérablement, la procédure s'arrête et la complexité du modèle qui donne la plus petite valeur du score VLOO est retenue.

Il peut arriver que le score de VLOO ne varie pas considérablement autour d'une valeur minimum, dans une certaine plage de nombres de neurones cachés. Si c'est le cas, un critère supplémentaire est pris en compte. Comme décrit auparavant, le levier de l'exemple i représente la proportion des paramètres du modèle qui contribue à caler le modèle avec l'exemple i. Le modèle qui a un ou plusieurs exemples avec levier proche de 1 dépend des erreurs de mesures de ces exemples, et doit donc être éliminé. Au contraire, le modèle dont les leviers sont proches de leurs valeurs moyennes p/N est dépendant également de tous ces exemples, il a donc une bonne capacité de généralisation. En conséquence, la quantité $\mu = \frac{1}{N} \sum_{i=1}^{N} \sqrt{\frac{N}{p} h_{ii}}$ est calculée, où μ est toujours plus petit que 1, et il est égal à 1 si et seulement si tous les leviers sont égaux à p/N. Parmi tous les modèles avec un petit score de VLOO, le modèle avec la plus grande valeur de μ est préféré.

Les détails des exemples numériques et les performances des modèles ne sont pas donnés dans ce résumé. Pour les lecteurs intéressés, veuillez-vous référencer à la section 4.4.

E.2.f-ii Comparaison globale avec le modèle de série chronologique

A l'issu de la présentation des deux modèles pour la prévision de charge à court terme, nous allons comparer les performances de la série chronologique et du réseau de neurones selon les six critères décrits dans le tableau E.3.

- **Précision de la prévision**

  Les résultats de la prévision sur les postes *CE_MOU* et *CE_FRO* sont listés dans annexe D. Dans le cas général, les réseaux de neurones ont une meilleure précision de la prévision que les modèles de série chronologique et les modèles naïfs.

- **Facilité d'interprétation**

Table E.3: Résumé de la comparaison entre le modèle du réseau de neurones et le modèle de la série chronologique pour la prévision de charge court terme. ☺ signifie que le modèle a une meilleure propriété.

	Réseau de neurones	**Modèle série chronologique**
Précision de la prévision	☺	
Facilité d'interprétation		☺
Complexité du calcul		☺
Quantité des données d'apprentissage		☺
Fréquence de mise à jour	☺	
Adaptabilité	☺	

Les modèles basés sur le réseau de neurones et la série chronologiques sont découpés en deux parties. Une partie représente les variations lentes de la consommation issue des facteurs exogènes, tels que, la température, la journée type, etc. et indique le niveau de la consommation. L'autre partie concerne les variations rapides et raffine les résultats finaux.

Etant paramétrique, le modèle série chronologique est facile à interpréter. La relation entre la puissance et les autres facteurs d'influence peut être déduite sans effort. Etant une boîte noire et non linéaire, le modèle réseau de neurones a des difficultés pour formuler une équation explicite entre la puissance et les autres facteurs d'influence [99].

- **Complexité du calcul**

  Les méthodes ont toutes deux périodes : la période d'apprentissage et la période de test. La période d'apprentissage pour la méthode série chronologique est consacrée à définir les valeurs de plusieurs variables importantes pour le modèle, telles que, les fréquences principales, la largeur de la fenêtre glissante, etc. Ensuite, pendant la période de test, le modèle ajuste les valeurs de ses paramètres itérativement grâce à la technique de la fenêtre glissante.

  De l'autre côté, pour le modèle réseau de neurones, les données d'apprentissage sont en même temps utilisées pour calculer les valeurs de ses paramètres, pour sélectionner le modèle optimal, et pour choisir les variables d'influence. Néanmoins, c'est un long processus. Il faut entrainer deux réseaux indépendamment : un pour prévoir la moyenne journalière, et l'autre pour prévoir la variation intra-journalière.

- **Quantité de données d'apprentissage**

  Ici, la quantité de données correspond au nombre de données consacrées aux calculs des paramètres dans un modèle. Pour le modèle chronologique, cette quantité est définie pendant la période d'apprentissage. Celle-ci peut durer d'une à plusieurs semaines. Le modèle de réseau de neurones a besoin d'une période complète pour apprendre correctement. Une année de données historiques est utilisée pour l'apprentissage de ce modèle.

- **Fréquence de mise à jour**

  La mise à jour d'un modèle change sa structure ou ses paramètres. Pour le modèle série chronologique, la structure est figée, mais les paramètres sont réestimés pendant chaque période de la prévision, i.e., chaque jour.

  Une année de données est utilisée pour l'apprentissage des modèles du réseau de neurones. Comparé au modèle de série chronologique, la précision de la prévision du modèle de réseau de neurones se dégrade à partir du mois de Mai, c'est-à-dire 8 mois après l'apprentissage. La fréquence optimale de mise à jour doit donc être choisie pour les modèles de réseau de neurones. Le tableau en annexe D montre les résultats détaillés de deux postes pour chaque mois de la période de prévision.

- **Adaptabilité**

  Avoir une bonne adaptabilité signifie que le modèle a la capacité d'être facilement appliqué à d'autres situations similaires. Le modèle de série chronologique, lié à la structure fixée, peut seulement extraire des informations simples telles que la journée type, la température, et les périodes principales. Le modèle ne peut pas prévoir la charge des postes industriels puisque ces postes ne dépendent pas de la température, et la période principale, hebdomadaire, a déjà été exploitée par le modèle naïf. La performance d'un modèle de série chronologique est donc un peu meilleure que celle du modèle naïf.

  Reconnu pour sa bonne capacité d'apprentissage, le modèle du réseau de neurones peut traiter toutes les sortes de prévision de charge. Le modèle adapte sa structure aux données d'apprentissage par la sélection de ses variables et de sa structure. Ce modèle est donc mieux adapté dans le cas industriel (annexe D). C'est aussi une solution prometteuse pour la prévision de la puissance réactive.

La section E.2 a présenté deux méthodes appliquées à la prévision de charge court terme pour la conduite et l'estimateur d'état. Différentes applications et différentes méthodes de la prévision de charge dans la littérature sont brièvement introduites au début de cette section. Ensuite, les différents types de courbe de charge sont illustrés, et leurs caractéristiques sont commentées afin de choisir des méthodes adaptées à notre besoin de la prévision court terme. Deux méthodes ont été choisies parmi de nombreux choix : série chronologique et réseau de neurones. A l'issue de la présentation théorique détaillée de ces deux méthodes, une comparaison sur différents aspects est effectuée. Dans la prochaine section, nous allons traiter le deuxième objectif de la thèse, la conception des modèles d'estimation pour la planification dans le réseau de distribution.

E.3 Modèle d'estimation de charge pour la planification du réseau de distribution

La planification joue un rôle important dans les réseaux de distribution en termes économique et technique. Elle impacte les décisions d'nvestissements ainsi que la qualité de la fourniture d'électricité. La précision du modèle d'estimation de charge est essentielle pour réduire l'incertitude afin de choisir la meilleure stratégie parmi toutes les alternatives. Dans

cette section, nous argumentons l'importance des modèles d'estimation de charge dans la procédure de prise de décision pour la planification des réseaux de distribution. Le modèle nommé BAGHEERA appliqué actuellement par le distributeur d'électricité français ERDF est détaillé. A partir de celui-ci, notre modèle est proposé, commenté et comparé.

Les décisions pour la planification des réseaux de distribution sont basées sur les analyses technique et économique. La Figure E.9 décrit des trois étapes de la procédure de prise décision dans la planification des réseaux. Dans l'analyse technique, les estimations de la charge, la topologie du réseau existant et des données du réseau (des transformateurs, des lignes et des câbles, etc.) sont prises en compte. Ensuite, l'analyse économique intègre les prix des composants, les pertes des lignes, le coût de maintenance et estime le coût pour chaque solution issue de la dernière étape. Celle-ci combine ces résultats avec le budget annuel pour faire des plans d'investissement.

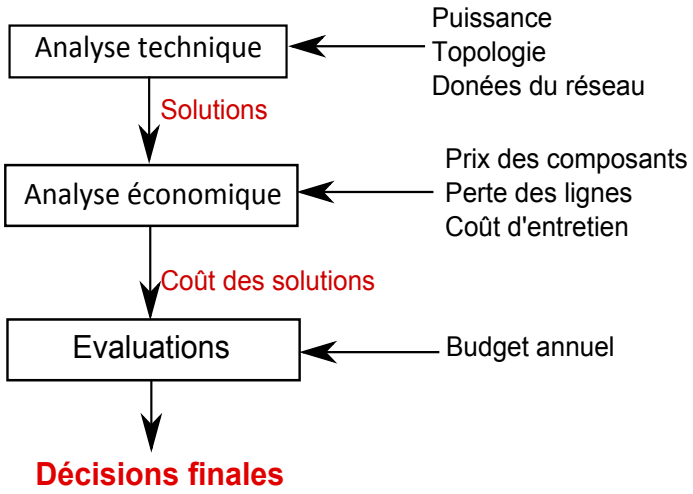


Figure E.9: La prise des décisions dans le réseau de distribution

Les modèles d'estimation de charge sont appliqués dans l'étape d'analyse technique. Celle-ci est considérée comme un problème d'optimisation des pertes électriques dans les réseaux de façon à ce que la tension sur chaque nIJud des réseaux reste dans une bande normalisée. Les clients dont la tension du point de raccordement est en dehors de cette bande sont considérés comme des CMAs (Client Mal Alimenté). Pour définir la tension du point de raccordement, les calculs des pertes et du plan de tension doivent être effectués. Avant l'apparition des compteurs intelligents, les seules mesures disponibles dans les réseaux de distribution étaient les puissances actives et réactives, la tension et le transit dans les postes sources et quelques relevés par an et par client pour la facturation. La consommation d'un poste HTA/BT est estimée par la proportion des valeurs maximales entre les postes HTA/BT et le poste source en amont. Afin d'estimer des modèles de charge sur un plus bas niveau de tension pour la planification des réseaux, la plupart des entreprises d'électricité collectent aussi des données de sondage sur une base régulière, i.e., 10 minutes, 15 minute, une demi-heure ou une heure, pour un nombre limité de clients.

La plupart des entreprises d'électricité établissent les profils de charge typiques qui représentent à peu près la consommation d'un client individuel. Ces profils sont ensuite agrégés à un niveau de tension plus élevée. Pour plus d'informations, les lecteurs peuvent

se référencer aux sous-sections 5.1.a et 5.1.b.

E.3.a Modèle BAGHEERA

Depuis 1997, Electricité de France (EDF) emploie un modèle nommé BAGHEERA pour ses clients connectés aux postes publiques HTA/BT. Il classifie les clients dans des groupes selon leurs informations qualitatives, telles que son code d'activité (résidentiel, agriculture, commercial, industriel, etc.) et son option tarifaire (base, heure creuse/pleine, Tempo). Ce modèle est utilisé pour l'estimation des pics de consommation et fournit 48 estimations de puissance (deux par heure) pour les jours ouvrables et les weekends pour chaque client individuellement. Ses avantages peuvent être résumés comme :

- Un modèle universel qui adapte aux clients de toutes catégories
- L'information de la période heure creuse/pleine de chaque client est prise en compte
- La sortie du modèle inclut 48 puissances. Quand nous faisons le calcul électrique, l'effet coïncidence (sous-section 5.1.a), ou effet de foisonnement, est pris en compte.

En collaboration avec le calcul électrique dans les cas les plus défavorables, le modèle BAGHEERA estime la charge d'un client individuel sous la condition TMB [1]. Il y a deux composants dans ce modèle : la puissance moyenne $P(t,T_d)$ sous la température T_d et la marge $\nu(t)$. Pour chaque heure, nous avons :

$$P(t,T_d) = a(t)E_0 + b(t)s(T_d - T_{Nh})|_{T_d<T_{Nh}} \tag{E.19}$$

$$\nu(t) = \sigma(t)^2 E_n \tag{E.20}$$

Où $\{a(t), b(t), \sigma(t)\}$ sont les coefficients calculés statistiquement et partagés pour tous les clients dans le même groupe. Les paramètres $\{E_0, s, E_n\}$ sont spécifiques à chaque client et correspondent respectivement à l'énergie journalière consommée hors chauffage, le gradient, l'énergie annuelle ajustée à la condition climatique normale. T_d est la température moyenne journalière. Equation E.19 assume que au-delà de la température de non chauffage T_{Nh}, qui est différente d'une région à l'autre, la puissance consommée ($a(t)E_0$) est indépendante de la température et au-dessous de cette température de non chauffage, la relation entre la puissance consommée et la variation de température est linéaire. Cette relation linéaire est indiquée par le gradient s ($s< 0$).

La puissance moyenne représente l'espérance de la puissance de chaque heure sous la température T_d, et la marge représente l'incertitude de l'estimation. Le risque fixé pour la planification des réseaux de distribution est de 10%, ce qui signifie que la borne supérieure a 10% de chance d'être dépassée. La borne « 10% supérieure » est donc définie comme :

$$P_{10\%}(t,T_d) = P(t,T_d) + c_{10\%}\nu(t) \tag{E.21}$$

[1] Température Minimum de Base : en pratique, EDF définie le TMB à partir de 30 ans de données historiques. La probabilité d'un jour par an est une valeur moyenne : en réalité, pendant une année chaude, nous pouvons probablement trouver aucun jour dont la température journalière est inférieure à cette TMB valeur, alors que pendant une année froide, plusieurs jours peuvent être inférieurs à cette TMB valeur.

Où $c_{10\%}$ est le coefficient du risque à 10%. En théorie, la borne supérieure de la température TMB est dépassée pendant $8760h/an * (10\% * 0.3\%) \approx 2.5$ heures par an.

Les détails du calcul des coefficients $\{a(t), b(t), \sigma(t)\}$ sont présentés dans sous-section 5.3.b-ii et celui des coefficients $\{s, E_0, E_n\}$ sont développés dans sous-section 5.3.b-iii.

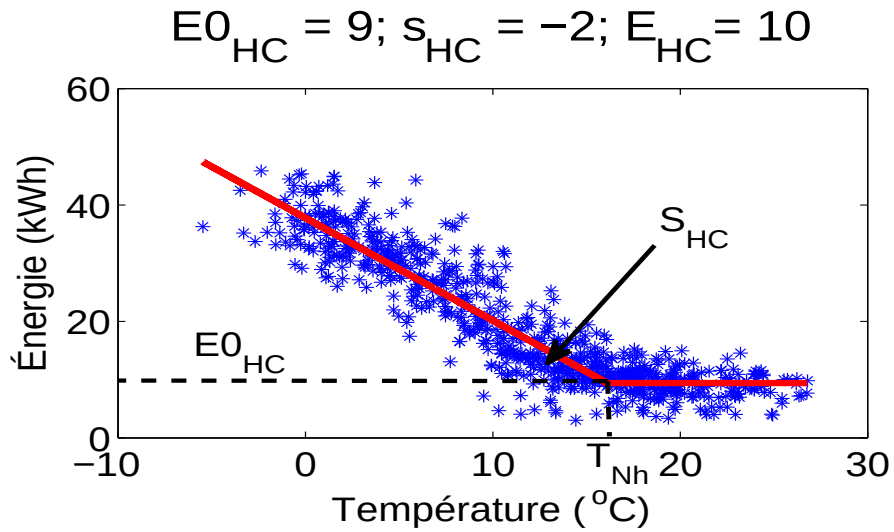


Figure E.10: *Client no.5* option heure creuse/pleine: ajustement de courbe sur l'énergie journalière pendant les heures pleines. L'indice « HP » signifie Heure Pleine et l'indice « HC » signifie Heure Creuse.

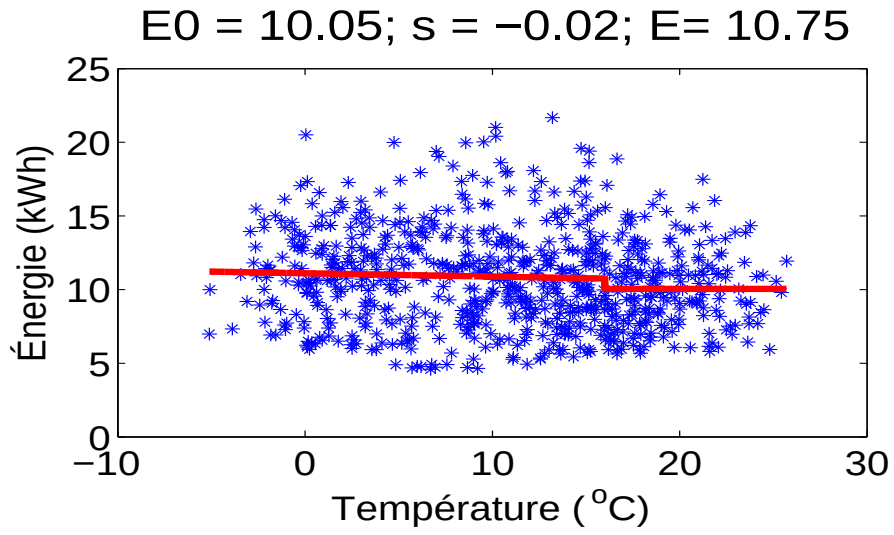


Figure E.11: *Client no.18* option de base: ajustement de courbe sur l'énergie journalière.

L'intervalle d'échantillon des clients sondés est de 10 minutes. Les données de la température sont échantillonnées par heure. Ces deux données sont moyennées sur une base journalière afin d'estimer les paramètres $\{s, E_0, E_n\}$ par l'ajustement de courbe. Les données au-dessus de la température T_{Nh} sont ajustées avec une constante E_0 et les données au-dessous de la température T_{Nh} sont ajustées avec une fonction linéaire $(s(T_d - T_{Nh}) + E)$

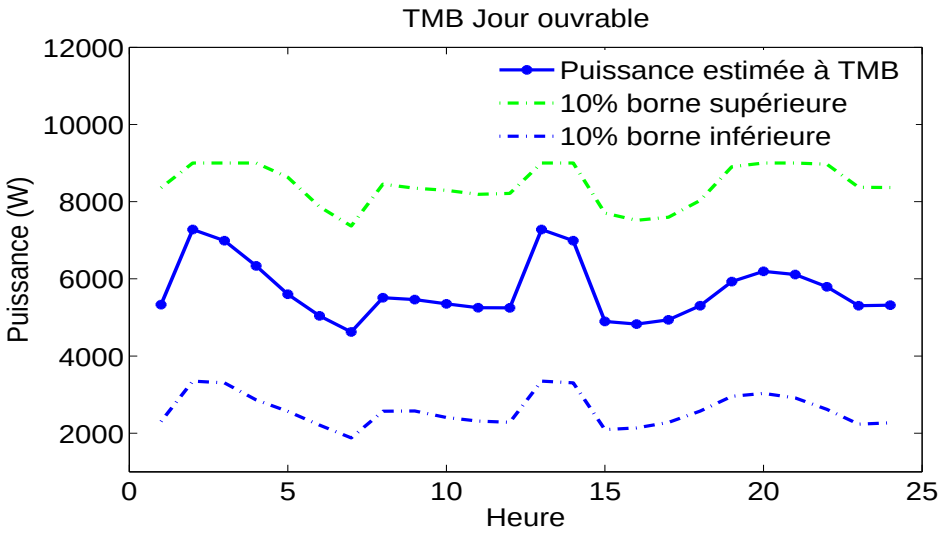


Figure E.12: *Client no.5* option heure creuse/pleine: TMB estimations de la charge pendant les jours ouvrables

qui minimise les moindres carrés.

Les clients heure creuse/pleine ont deux compteurs électriques, qui enregistrent séparément les énergies consommées pendant les heures pleines et les heures creuses. La figure E.10 illustre l'ajustement de courbe pendant les heures pleines. Une figure similaire peut être tracée pour les heures creuses. L'énergie journalière varie avec la température. De cette façon, les clients heure creuse/pleine ont souvent un grand gradient. En fait, T_{Nh} est une grandeur qui indique la température moyenne de non chauffage pour une région. Ce facteur peut varier d'un client à l'autre de la région. Dans la figure E.10, nous voyons que l'énergie moyenne pendant les périodes de non chauffage ($E0_{HC}$) est différente de l'énergie journalière hors chauffage déduite par la régression (E_{HC}). Les clients d'option de base utilisent d'autre moyens de chauffage, notamment des moyens non électriques. Leur consommation électrique reste stable par rapport à la variation de température (figure E.11). Ils ont donc un petit s. Pour la même raison, l'énergie journalière hors chauffage $E0$ est différente de celle de la régression E.

Les coefficients communs $\{a(t), b(t), \sigma(t)\}$ dépendent de la catégorie des clients. 48 estimations de la valeur moyenne ainsi que des bornes supérieures sont calculées avec les équations E.19, E.20 et E.21. Notons que différents ensembles de valeurs $\{a(t), b(t), \sigma(t)\}$ existent pour les jours ouvrables et les week-ends, ainsi que pour les périodes heures creuses et heures pleines. S'il existe plus d'une période d'heures creuses pendant la journée, la puissance estimée est considérée comme identique pour chaque début de période. Dans la figure E.11, le *client no.5* il y a deux périodes d'heures disjointes : 01 :00 à 07 :00 et 12 :00 à 14 :00. Nous constatons que deux pics pendant le début de chaque période ont la même valeur. Les valeurs des bornes supérieures sont écrêtées par la puissance souscrite. Dans la figure E.11, les estimations des bornes supérieures dépassant 9kW sont remplacées par 9kW, la puissance souscrite du *client no.5*.

Le modèle BAGHEERA suppose que l'influence de la température sur la consommation est linéaire. Néanmoins, cette hypothèse n'est pas réaliste. Cela est dû à la limite des

puissances maximales des chauffages. De plus, le concept du réseau intelligent conduit à une dynamique importante et des relations flexibles entre le producteur d'électricité, le distributeur et les clients. Il n'est donc plus raisonnable d'attribuer un client à un groupe fixe. Avec la climatisation, de plus en plus répandue en France, la consommation va aussi augmenter pendant les périodes chaudes. L'hypothèse du modèle BAGHEERA selon laquelle la charge reste stable quand la température dépasse une température de non chauffage n'est plus vraie. Notre idée est de construire un modèle individuel pour chaque client sans faire la classification. Ce modèle doit être complètement dirigé par les données et indépendant par rapport aux informations qualitatives des clients dont la qualité se dégrade. En tenant compte du développement des compteurs intelligents, qui fournissent des consommations détaillées de chaque individu, nous proposons un modèle d'estimation de charge non paramétrique individualisé pour chaque client.

E.3.b Modèle non paramétrique

Dans cette sous-section, nous abordons le modèle d'estimation de charge non paramétrique. Les avantages du modèle peuvent être résumés en trois points :

- La méthode est complétement dirigée par les données et peut être adaptée à tous les types de charge, pour les clients thermosensibles ainsi que pour les clients non thermosensibles.
- Les méthodes non paramétriques sont appliquées de sorte que le modèle est indépendant des informations qualitatives d'un client ainsi que de l'hypothèse faite sur les fonctions de charge.
- La sortie du modèle inclus les estimations de charge maximales et minimales.

La figure E.13 présente la procédure du modèle non paramétrique. Dans un premier temps, la courbe de charge d'un client est associée à la température de la région où les mesures sont prises. Ensuite, les tests statistiques sont effectués pour examiner si ce client est thermosensible. Si ce client n'est pas sensible par rapport à la température, il n'y aura pas de différence entre la courbe de la puissance maximale et celle de la température TMB. L'estimation de la densité de la probabilité par noyau est utilisée pour déterminer les valeurs médianes, ainsi que la borne supérieure à 10% et la borne inférieure à 10%. Si ce client est thermosensible, l'algorithme CUSUM (somme cumulée) est appliqué pour séparer la période de chauffage de la période de non chauffage. Ensuite, les données de la période de non chauffage sont utilisées pour calculer les puissances minimales et les données de la période de chauffage sont utilisées pour estimer les puissances sous température TMB. Notons que la température TMB est une de plus basses températures de l'année, qui a lieu en moyenne une fois par an. Les connaissances de la consommation électrique sous cette température sont souvent insuffisantes. Ce manque de connaissances peut engendrer des difficultés pour l'estimation des puissances à température TMB. Nous proposons donc d'inclure les mesures de la charge des années précédentes si ces données sont compatibles avec les données de l'année estimée. Cette compatibilité est vérifiée par des tests statistiques. Donc, éprouvées par les tests, les données des années précédentes compatibles avec les données de la période de chauffage de l'année estimée sont utilisées

pour estimer la relation entre la température et la charge par une méthode de régression par noyau. Le paramètre de lissage demandé par la méthode de régression par noyau est calculé par la technique de validation croisée. Une fois la relation entre la température et la charge établie, les données de la période de chauffage sont ramenées à la condition TMB selon cette relation en maintenant leurs incertitudes d'observation. Finalement, les puissances médianes et les bornes sont calculées à la température TMB. Les outils statistiques montrés dans la figure E.13 ne sont pas présentés ici. Pour plus de détails, nous invitons les lecteurs à se reporter à la section 6.1.

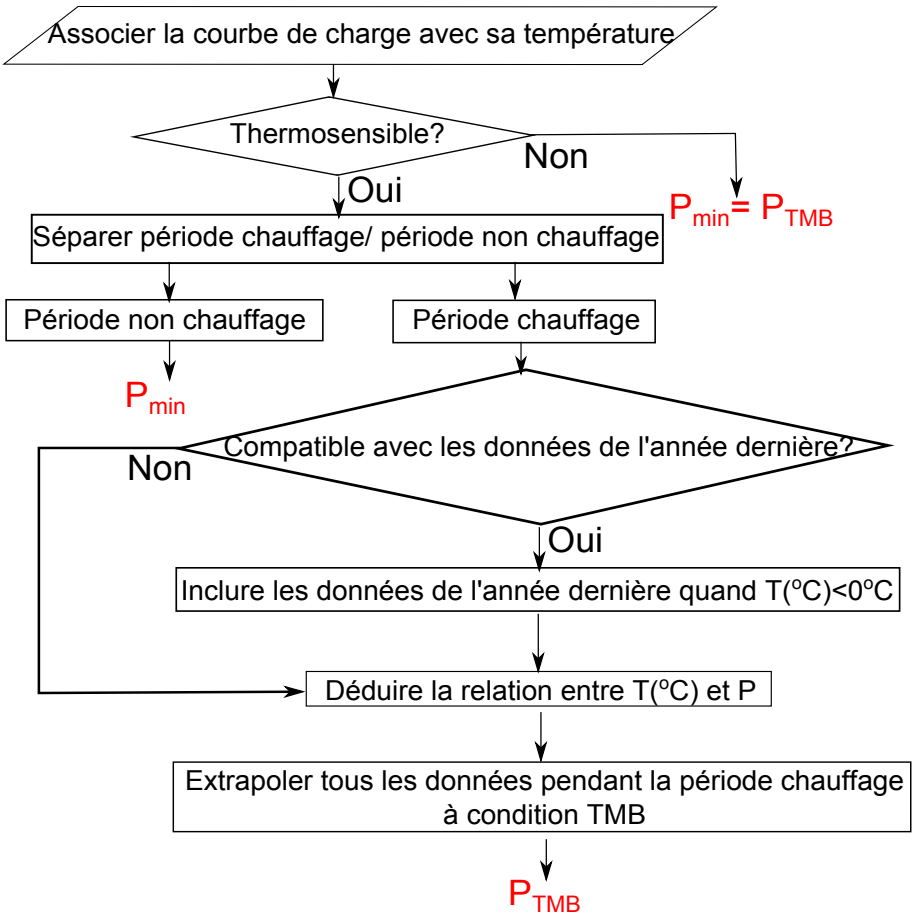


Figure E.13: La procédure du modèle non paramétrique

Les modèles non paramétriques sont comparés avec le modèle industriel BAGHEERA avec des données réelles (sous-section 6.2.a). Nous soulignons que les estimateurs proposés sont plus adaptés et donnent des résultats plus fiables (sous-section 6.2.b).

E.4 Conclusions et perspectives générales

Dans ce mémoire, nous traitons deux problèmes distincts sur la charge électrique dans les réseaux de distribution, respectivement pour la conduite et la planification. L'apparition des compteurs électriques intelligents dans un contexte de réseaux intelligents donne la possibilité de construire les modèles prédictifs de charge plus précis sur la basse tension, ainsi que des modèles d'estimation de charge individualisés.

Trois modèles ont été traités, les deux premiers pour l'estimation de charge en J+1 et J+2, le troisième pour l'estimation de charge à l'échelle de la planification :

- Le modèle de série chronologique découpe la courbe de charge en trois parties : la tendance, la cyclicité et les erreurs aléatoires. Les deux premières parties sont déterministes et sont présentées séparément par deux modèles paramétriques. Le modèle de tendance prend en compte la température, le temps et les journées types. Le modèle cyclique est composé des composantes de Fourrier, dont les fréquences principales sont identifiées par le périodogramme lissé. L'analyse du résidu montre qu'il n'y a plus de possibilité d'améliorer les résultats de la prévision avec les autres approches classiques.

  Nous avons aussi étudié l'influence de l'incertitude dans la prévision de la température vis-à-vis de la précision pour la prévision de charge. Nous avons montré que même avec l'incertitude dans la prévision de la température, notre modèle de série chronologique est beaucoup plus précis par rapport au modèle de référence : le modèle naïf.

- Le réseau de neurones a une capacité d'approximation universelle et une bonne capacité à traiter les relations non linéaires et complexes. Le réseau de neurone est souvent considéré comme ayant une prévision plus précise que les méthodes classiques [99]. Dans la plupart des travaux qui traitent de la problématique de la prévision de charge à court terme, le choix des variables ainsi que celui de la structure pour le réseau de neurones n'étaient pas systématiquement justifiés. Les résultats finaux n'étaient pas bien commentés ou présentés clairement [39, 28].

  Nous focalisons sur la méthodologie de la conception du modèle, i.e., les sélections des variables et du modèle. La sélection des variables est basée sur la régression orthogonale, qui range les variables candidates dans un ordre de pertinence décroissante. La sélection du modèle est basée sur le processus VLOO, qui estime la capacité de la généralisation pour le modèle. La courbe de charge est décomposée en la puissance moyenne journalière et les variations de la puissance intra journalière. Cette décomposition réduit la complexité du modèle. L'efficacité de la stratégie de sélection est prouvée en utilisant de vraies mesures de postes HTA/BT.

- Pour l'objectif du modèle d'estimation de charge, nous introduisons le concept consistant à construire le modèle individualisé pour chaque client sans faire de classification. Ce modèle non paramétrique est complètement dirigé par les données. Il ne possède pas d'hypothèses à priori sur la relation entre la charge et la variation de la température. Dans ce rapport, nous avons proposé trois estimateurs pour différents cas de disponibilités des données. Les études de validation comparent la précision des estimations du modèle BAGHEERA et les modèles non paramétriques sur la base des mêmes données. Ces études montrent que les estimateurs non paramétriques ont une meilleure précision que celui du modèle BAGHEERA. Une discussion pour la définition du seuil de l'incertitude pour le calcul de la chute de tension est développée.

Les objectifs de la thèse sont atteints.

Les modèles prédictifs de charge ont une précision meilleure que le modèle naïf. Néanmoins, en termes d'efficacité ainsi que de bénéfice économique sur les fonctions avancées du réseau, le rapport n'a pas pu être quantifié. Les prochains travaux devraient se focaliser sur l'intégration de ces modèles dans les fonctions avancées du réseau.

La méthodologie de la conception du modèle réseau de neurones est très prometteuse pour la prévision de charge. Elle fournit une prévision plus précise que la méthode série chronologique. Cette solution peut être étendue à tous les types de la prévision de charge, y compris la puissance réactive. Les résultats de ces premières études sont encourageants. Néanmoins, des études plus poussées devraient être effectuées pour les raisons suivantes :

- Dans notre méthodologie, les variables candidates secondaires sont conçues pour prendre en compte la non linéarité du modèle. Cependant, une variable candidate secondaire représente seulement le produit de deux variables primaires. Il y a peut-être d'autres formes de combinaison à explorer.

- Le risque de sélectionner une variable non pertinente est estimé par les variables sondes. A cause d'un effet de mélange aléatoire, le nombre de variables retenues malgré le même risque peut être différent. D'autres méthodologies de sélection de variables peuvent être envisagées, tels que l'analyse d'entropie [115], l'espace des phases associées [106].

- La performance se dégrade quand le temps de prévision augmente. Voir Annexe 4.3.a. Puisque les modèles de charge suivent un processus non linéaire, dû à l'augmentation progressive, ils doivent donc être mis à jours régulièrement. En conséquence, la fréquence de mise à jour du modèle doit être déduite pour atteindre une performance optimale.

Nous avons montré que le modèle non paramétrique a une meilleure précision que le modèle BAGHEERA sur l'estimation des valeurs moyennes. Une discussion est consacrée pour clarifier la définition de la borne supérieure, dont les valeurs participent directement dans le calcul de chute de tension. Cette discussion doit être prolongée avec les distributeurs. La validation de la méthodologie pour cette estimation est aussi une tâche difficile. Puisque peu d'échantillons sont pris à température TMB pendant une période de deux ans [3], la probabilité d'être dépassée pendant 10% de temps est donc difficile à vérifier. Nous proposons de tester cette méthodologie dans l'implémentation industrielle.

Aujourd'hui, un grand nombre des énergies renouvelables sont intégrés dans les systèmes électriques. Leurs occurrences changent les transits dans les lignes électriques. Elles peuvent être considérées comme des consommations électriques négatives.

Différentes des moyens de productions classiques (nucléaire, fioul, et hydraulique, etc.), ces générations d'énergies distribuées renouvelables (éolien, photovoltaïque, et biomasse) connectées aux réseaux de distributions sont intermittentes, et très dépendantes des conditions météo. La qualité de la prévision de ces productions dépend plus de la qualité des prévisions des facteurs d'influence que de l'efficacité de l'algorithme de la prévision.

Par exemple, pour le photovoltaïque, les facteurs d'influence directs sont la nébulosité, la radiation solaire, et la température [156]. Les autres paramètres comme le vent, la pluie et la neige sont des facteurs indirects [155]. La difficulté est d'obtenir une prévision

précise sur la nébulosité, puisque cela est basé sur un phénomène physique très complexe et cela représente plus de difficulté que la prévision de la température moyenne journalière. L'éolien, d'un autre côté, dépend de la force et de la direction du vent, de la géographie de la ferme, et du type de générateur. La prédiction des facteurs d'influence ne fait pas partie du périmètre de cette thèse.

Afin de calculer correctement la chute de tension pour la conduite et la planification du réseau de distribution, nous devons tenir compte de ces productions. Deux façons sont envisageables :

- un modèle prédictif pour la production
- un modèle qui prévoit la valeur agrégée de la génération et la consommation

La plupart des travaux de recherche se concentrent sur la première voie. Il y a trois catégories de modèles pour prévoir ces productions : basé sur un modèle physique, sur un modèle statistique ou sur un modèle hybride qui combine les avantages des deux premiers [155]. Les méthodes sont similaires à celles de la prévision de charge [155, 157, 158].

Nous proposons de déployer les deux modèles présentés dans ce rapport, le modèle du réseau de neurones et le modèle non paramétrique pour résoudre le problème dans la seconde voie. Grâce à la bonne capacité d'apprentissage du réseau de neurones, si les prévisions des facteurs d'influence sont assez précises, le modèle est capable de déduire la relation non linéaire complexe entre entrées et sorties. Dans ce cas, la sortie du modèle est la prévision de l'écart entre la production et la consommation. Les stratégies pour les sélections des variables et du modèle (parties *Sélection des variables* et *Sélection du modèle*) peuvent être aisément appliquées. En ce qui concerne le modèle non paramétrique, sans hypothèse à priori, et complétement dirigé par les données, il peut être facilement mis en œuvre pour la situation où ces énergies renouvelables sont présents.

Pour les futurs travaux, il faudrait tester ces deux modèles avec les vraies données des consommations électriques et des productions des énergies renouvelables.