CONFIDENTIAL

Predictive Estimation of the Real-Time Electricity Market Price

I. de Hoogt

Master of Science Thesis

TU Delft Delft University of Technology

Delft Center for Systems and Control

MSCCONFIDENTIAL

Predictive Estimation of the Real-Time Electricity Market Price

Master of Science Thesis

For the degree of Master of Science in Systems and Control at Delft University of Technology

I. de Hoogt

March 1, 2017

Faculty of Mechanical, Maritime and Materials Engineering (3mE) · Delft University of Technology

The work in this thesis was supported by Peeeks. Their cooperation is hereby gratefully acknowledged.

Abstract

Electricity can not be stored efficiently in the grid, resulting in a need for demand and supply to be in balance. The Transmission System Operator operates a real-time electricity market to acquire extra supply or load. Traditionally, electricity usage is forecast as accurately as possible a day ahead and deviations from this electricity program are kept to a minimum as financial risk is involved with these deviations. Deviations from the submitted electricity program are settled at the real-time electricity market clearing price, which can be profitable to the parties involved.

The aim of this project is to achieve accurate, real-time and interpretable prediction of the electricity market price. To this end the Generalised Fuzzy Neural Network formulation of a Non-linear Auto-Regressive with eXogenous inputs model structure (GFNARX) model structure is proposed as contribution to existing fuzzy modelling literature. Accuracy of the proposed model is shown to be comparable to that of state-of-the-art fuzzy models on a popular literature benchmark, prediction of the Mackey-Glass chaotic time series, while using computationally cheaper means. As there is no substantial amount of literature on Real-time Electricity market Price (REP) forecasting which enables comparison of GFNARX predictions to any other, two other models from time series literature are used to generate comparison material: Seasonal Auto-Regressive Integrated Moving Average with eXogenous inputs and Generalized Auto-Regressive Conditional Heteroskedasticity (SARIMAX-GARCH) and Non-linear Auto-Regressive with eXogenous inputs (NARX).

To assess whether information about the real-time electricity price can be used to reduce electricity consumption costs for assets in a demand response portfolio, a benchmark method which simulates control of the cooling motor of a developed cold storage warehouse model according to the predicted state of the real-time electricity market, is proposed. Using the GFNARX model to predict the real-time price throughout the year 2015 based on historical data, a 25.5% reduction in cumulative electricity costs is obtained compared to the reference case where all electricity consumption is bought on the day ahead market. When comparing the GFNARX prediction result to using the most recent electricity price observation as naïve forecast which achieves 16.9% cost reduction, a relative 10.3% cost reduction is achieved.

Table of Contents

List of Figures

List of Tables

Acknowledgements

This thesis work was supervised by prof.dr.ir. M. Verhaegenof Delft Center for Systems and Control (DCSC) and ir. D. Beijerof Peeeks. I am grateful to their valuable feedback, without which the standard of this work would not have reached the current level. I would also like to thank my colleagues at Peeeks as it was a pleasure to share an office with them.

Delft, University of Technology I. de Hoogt
March 1, 2017

Chapter 1

Introduction

This chapter will introduce the topic of research in this thesis project and formulate a problem statement. Section 1-1 will introduce the problem within its context, before Sections 1-2 and 1-3 provide detail on the research scope and problem statement. Section 1-4 gives an overview of the thesis outline. A lot of material in this chapter is based on a literature survey performed during an earlier stage of this thesis assignment[14].

1-1 Research context and scope

In a lot of residential and industrial appliances, electricity is used as a power source. Electricity is produced at large scale by power stations, transmitted across high voltage lines and then distributed to consumers through local low voltage networks. Electricity, as opposed to most other commodities, cannot be stored in the grid in an efficient manner. This indicates a need for electricity supply and consumption to always be in equilibrium.

Traditionally, consumers subtract energy from the grid whenever there is a need to and generation is required to match the demand. Generators are thus faced with a predictive control problem. Conventional generation plants can regulate their output to match demand, but renewable energy sources are weather dependent and their generation output is hence not deterministic. The stochastic nature of weather conditions causes an increased volatility in the required control pattern of conventional generation plants, which are inherently limited in ramp rate and capacity.

To allow for a greater flexibility, the concept of demand response has been introduced. Demand response aims at matching supply and demand by altering the consumption pattern to match available supply levels[15]. During times when overall demand exceeds the capacity of generators, the consumption is lowered and this is compensated for in times of abundant supply. In a demand response framework, consumers are thus faced with a predictive control problem in which supply forecasts are required in order to control consumption levels.

If there is a disturbance to balance in the grid, both supply control and demand response can restore the equilibrium, combining the separate control problems into one balancing problem. To avoid unnecessary control action the Transmission System Operator (TSO) governs a real-time balancing energy market, in which it is the only broker. The marginal price of electricity bought or sold at the real-time market is applied to any unplanned electricity transaction occurring in a block of fifteen minutes, a Program Time Unit (PTU)[16]. This provides Balance Responsible Partys (BRPs) with a financial incentive to adjust their electricity program in a way which balances the grid.

The research for this thesis is performed at Peeeks, a company which investigates ways to increase grid tolerance to uncertain supply by renewable energy sources through deployment of flexible demand response services. In order to determine the grid state and thus the direction in which to steer the demand, the Real-time Electricity market Price (REP) is observed. REP is published by the TSO, TenneT, in the form of a time series[9] and reflects the price at which TenneT is buying or selling electricity to balance the grid. However, the measurements of this time series are not published in real-time, but with a delay of three minutes. This delay in availability of measured data inherently leads to a delay in demand response control action. Applying control based on delayed data can lead to undesirable effects like steering in the wrong direction which aggravates the problem instead of alleviating it. By forecasting REP, negative effects of this delay can potentially be reduced.

1-2 Scope

Based on the findings of the literature survey performed in an earlier stage of this thesis project, the scope of this project can be determined. This paragraph will shortly summarise a few key findings of this literature research. For more details the reader is referred to [14].

In electricity price forecasts, the time scale of interest is usually in terms of days with a time resolution of hours. These forecasts are done in order to optimise trading strategies on spot markets like the day-ahead Amsterdam Power Exchange (APX). Traditionally, BRPs try to forecast their electricity consumption and generation a day ahead as accurately as possible and avoid the highly volatile real-time market.

It was found that not much has been published about forecasting the real-time electricity price, which can be attributed to two reasons. The first of these reasons is the fact that in Europe, only the dutch TSO publishes information about REP within minutes of price formation. This started in September 2009[17], which is relatively recent. The second is that the real-time electricity market is structured to only have a single broker, the TSO[18]. Most TSOs choose to discourage BRPs from deliberately causing imbalance in their own portfolio by having the Settlement Price (SP) be the most penalising between REP and the day ahead spot market price. The dutch TSO however, allows the SPs to be fully governed by REP[19], creating opportunities for BRPs to profit from deliberate deviations from the forecast electricity program. The combination of these two reasons is an exclusive and recent feature of the dutch real-time electricity market, which explains why there has been so little research performed in this direction.

A recent paper aiming at providing a comprehensive overview of real-time electricity price forecasting methods ([20]) shows that prior work has mainly focused on forecasting SP a day ahead and it has been determined that existing methods fail to structurally outperform random guessing. In this research, the focus lies on real-time control of demand response assets instead of day-ahead optimisation. No published research has been found in the course of the literature survey which investigate usage of REP to control a demand response asset. This lead to the decision to focus on identifying models with which to forecast REP.

As REP is published by the TSO in the form of a times series, the literature survey investigated existing time series models with a focus on models used to forecast related variables like cumulative grid load and the APX day ahead price. Based on the lack of prior study, the choice has been made to follow a fuzzy data-driven modelling approach in which the influences of exogenous factors can be interpreted and assessed on a qualitative basis.

Research regarding model inputs will be limited to investigating publicly available data sources, as it is not yet known whether these data sources are sufficiently correlated to REP to use in REP forecasts.

In conclusion, the scope of this thesis research is the development of an interpretable model to forecast REP with and investigation of the use of REP forecasts to reduce negative effects of delay in the observations of the time series.

1-3 Problem statement

The real-time market prices for electricity are published in time series format with an approximate delay of three time steps. As these prices are highly volatile in nature, controlling assets in a demand response portfolio based on the time-delayed time series can lead to an aggravation of grid imbalance as well as monetary damages for the BRPs involved. It is thus of high interest to accurately forecast REP.

It has been established in [14] that forecasting REP has not received proper attention in literature. Therefore, although accurate predictions of REP are the main interest in controlling demand response assets, it is also desirable to have an interpretable model in order to gain enlightenment regarding the influential factors concerning the inner workings of the real-time market. To this end, fuzzy time series modelling techniques can be employed. To attain the dual goals of accuracy and interpretability, a novel self-organizing fuzzy neural network structural algorithm is proposed in which data-driven identification of a rulebase is performed.

The goal of this thesis is to compensate for the delay in REP measurement publishing. This is done by forecasting the REP time series. Three models are used to do so:

1. A popular model in time series literature: SARIMAX-GARCH
2. A recurrent neural network structure: NARX

3. A self-organising fuzzy neural network: GFNARX

The quality of the models is benchmarked against a naïve reference in which the latest REP observation serves as forecasted value. The benchmark method is a simulation in which the cooling motor of a Cold Storage Warehouse (CSW) is controlled based on the forecasts, with the aim of reducing costs of acquiring electricity.

There are four research aims:

- First of all it is investigated whether the proposed data-driven fuzzy modelling technique is capable of producing time series predictions which are comparable in accuracy to computationally more expensive models in literature
- Secondly, it is investigated which observable endogenous and exogenous factors can be used as influential inputs to a REP model identified through time series modelling techniques
- The third research challenge is to find which model structure of the ones selected in [14] (SARIMAX-GARCH, NARX, GFNARX) predicts REP most accurately
- The final research aim is to incorporate REP predictions in the demand response control strategy of Peeeks and investigate how much of the negative effects of the delay in publishing of REP can be mitigated when using REP forecasts to control the cooling motor of a CSW

The hypothesis here is that by using Generalised Fuzzy Neural Network formulation of a Non-linear Auto-Regressive with eXogenous inputs model structure (GFNARX), REP forecasts can be made which can compensate for the delay. Analysis of historical REP data of 2015 will verify the truth of this hypothesis.

1-4 Outline

Theoretical background of the models used throughout this work is given in Chapter 2. Chapter 3 identifies plausible inputs to the model and explains the experiments used to establish model performance, after which the results are provided in Chapter 4. The conclusions and recommendations for further research are found in Chapter 5

Chapter 2

Theory

The aim of this project is to achieve accurate, real-time and interpretable prediction of the electricity market price, which is published by the Transmission System Operator (TSO) in the form of a time series. To achieve this aim, time series models are identified and the value of their predictions are benchmarked on a cold storage warehouse temperature control simulation. In this chapter, a data-driven fuzzy time series modelling approach is proposed, which will be used to forecast the Real-time Electricity market Price (REP). This new method is coined Generalised Fuzzy Neural Network formulation of a Non-linear Auto-Regressive with eXogenous inputs model structure (GFNARX). As there is no benchmark available in literature, the new modelling approach is benchmarked against two well-known time series models. Seasonal Auto-Regressive Integrated Moving Average with eXogenous inputs and Generalized Auto-Regressive Conditional Heteroskedasticity (SARIMAX-GARCH) is chosen as benchmark model because it is a generalized form of the most popular model type used in econometry, ARIMA. As a neural network alternative benchmark, Non-linear Auto-Regressive with eXogenous inputs (NARX) is chosen. In Section 2-1, some concepts are introduced which serve as base to better understand the models introduced in subsequent sections. Section 2-2 discusses the SARIMAX-GARCH structure and the way model orders and parameters are estimated. The nonlinear alternative, NARX, is described in Section 2-3, before Section 2-4 elaborates on the proposed GFNARX model class.

2-1 Time series modelling concepts

In this section a few base concepts in time series modelling are introduced, starting with system identification concepts in Section 2-1-1. Section 2-1-2 describes wide sense stationarity and conditions under which a model can be identified from a single data sample series. Basic model prototypes which are used to describe time series are Moving Average (MA) and Auto-Regressive (AR) models. They are introduced in Sections 2-1-3 and 2-1-4. Finally, the Linear Least Squares (LLS) parameter estimation method is introduced in Section 2-1-5.

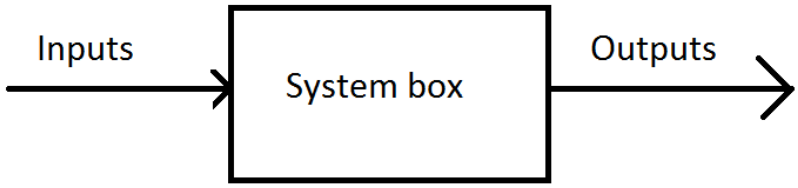


Figure 2-1: Abstract system identification structure where the input-output transfer is modelled as a box

2-1-1 System identification

When modelling systems, one has to derive a relation between system inputs and system outputs. In some cases, this relation can be derived from constitutional principles. An example of this is the simplified relation between driving force F as a single input and particle acceleration a as output: $F = ma$. In this example, the relationship can be derived from principles known from theoretical physics, and as such is called a first principles model.

Another way of viewing this relationship between inputs and outputs is shown in Figure 2-1 and can broadly be put into three categories:

White box modelling The inner workings of the system box can be inspected freely and the transfer function can be derived by first principles modelling

Grey box modelling The box is not fully transparent, but it is possible to derive a partial model or a sensible model structure from the parts which can be inspected. A common way of grey box modelling is to assume model structure a priori and subsequently estimate and optimise the model parameters

Black box modelling In case the box is fully opaque, the system identification procedure is called black box modelling. In black box modelling no prior assumptions about model structure are made. In many applications, the exact relations between inputs and outputs are irrelevant, as long as the output predicted by the model is accurate

2-1-2 Wide sense stationary series

When analysing time series, stationarity is an important concept. If a stochastic time series is stationary, its mean and variance do not change over time. Thus, the time at which an experiment is performed is of no consequence to the experiment results, allowing a time-invariant model representation. If a time series does have a variance which changes over time, but its autocovariance does remain constant, it is called Wide Sense Stationary (WSS)[21].

$$WSS \quad if \quad \begin{cases} \mathbb{E}[x_t] = \mu_x \quad \forall t \\ C_x(t_2, t_1) = \mathbb{E}[(x_{t_2} - \mu_x)(x_{t_1} - \mu_x)] = \mathbb{E}[(x_{t_2 - t_1} - \mu_x)(x_0 - \mu_x)] \end{cases} \tag{2-1}$$

with

$\mathbb{E}$ Expectancy operator

μ_x Mean of time series x

C_x Autocovariance of x

If the statistical properties of a process can be inferred from a single experiment, it is called ergodic. For example if one has a bag full of fair, six-sided dice, taking out one die and casting it a large amount of times will reveal the expected value and distribution of values for all of the dice in the bag. If part of the bag consists of four-sided dice, it is impossible to deduce the statistical properties of the entire bag in a single experiment.

In the context of forecasting a WSS time series, if a WSS process is ergodic in the mean, an unbiased predictor of the time series can be determined from a single long sample of its stochastic measurements:

$$\mathbb{E}[x_t] = \lim_{T\to\infty} \frac{1}{T} \sum_{t=0}^{T} x_t \tag{2-2}$$

A sufficient condition to establish ergodicity in the mean for a WSS time series, is based on the autocovariance being a convergent series [22]:

$$\sum_{j=0}^{\infty} C_x(t, t-j) < \infty \tag{2-3}$$

2-1-3 Moving average models

When analysing time series, Moving Average(MA) models are commonly used to smooth out noisy data and recover trends or cycles present within the series. As such, it can be viewed as being a low-pass filter. A pure MA process of order N_{MA} is described by Equation 2-4. The autocovariance then reduces to Equation 2-5. It cuts off after lag N_{MA}. The order of moving average terms can thus be inferred from inspection of the sample autocorrelation.

$$x_t = \mu + \epsilon_t + \sum_{i=1}^{N_{MA}} \theta_i \epsilon_{t-i} \tag{2-4}$$

$$\begin{aligned} C_x(t, t-j) &= \mathbb{E}[(x_t - \mu_x)(x_{t-j} - \mu_x)] \\ &= \mathbb{E}[(1 + \sum_{i=1}^{N_{MA}} \theta_i \epsilon_{t-i})(\sum_{i=1}^{N_{MA}} \theta_i \epsilon_{t-j-i})] \end{aligned} \tag{2-5}$$

with

ϵ_t Prediction error at time step t. It is often assumed to be white noise

θ Weighting parameter vector for the lagged error terms

2-1-4 Autoregressive models

If the value of a time series depends on its previous values, but is offset by a random noise disturbance at every measurable point, it is called an Auto-Regressive (AR) process. A pure AR process of order N_{AR} is described by Equation 2-6. The model order can now not be estimated from the sample autocovariance, as the autocorrelation at lag p is influenced by the propagation of autocovariance present in lagged terms of shorter order. To clarify, an example of an AR process of order 1 is given in Equation 2-7.

$$x_t = \epsilon_t + \sum_{i=1}^{N_{AR}} \phi_i x_{t-i} \tag{2-6}$$

$$x_t = \phi x_{t-1} + \epsilon_t \tag{2-7}$$

with

ϕ Weighting parameter vector for the lagged AR terms

ϵ_t Prediction error at time step t. It is often assumed to be white noise

It is easy to see that in Equation 2-7 x_t and x_{t-2} are not directly correlated. However, the propagation of correlations results in a non-zero correlation coefficient of ϕ^2.

To estimate the order N_{AR} the Partial Autocorrelation Function (PACF) can be inspected. The PACF estimates the direct correlation between x_t and its lagged values by removing linear dependencies on the lagged values. It has been proven by [23] that the PACF of the $j-th$ lagged term is equal to the coefficient α_j of the optimal linear prediction of x_t based on its lagged values up to x_{t-j}, as is described by Equation 2-8.

$$\hat{x_t} = \sum_{i=1}^{j} \alpha_i x_{t-i} \tag{2-8}$$

The order of an AR process can thus be inferred by inspection of the PACF. The coefficients α_i can be estimated through a LLS procedure. However, in system identification, these coefficients are constrained: to have an identified model describe a WSS series, the transfer function poles, which correspond to the solutions for r to Equation 2-9, have to lie within the unit circle. For AR models to generalise to out-of-sample predictions, WSS is required.

$$r^{N_A R} - \sum_{i=1}^{N_{AR}} \phi_i r^{N_{AR}-i} = 0 \tag{2-9}$$

2-1-5 Linear Least Squares estimation

When considering a quadratic objective function of the form $V(\beta) = ||y - x\beta||^2$, which represents a square absolute error for a model which is linear in the parameters β, the parameter

vector β which minimizes the cost function can analytically be derived by expanding and differentiating with respect to β:

$$\begin{aligned} V(\beta) &= (y - x\beta)^T (y - x\beta) \\ &= y^T y - \beta^T x^T y - y^T x\beta + \beta^T x^T x\beta \\ &= y^T y - 2\beta^T x^T y + \beta^T x^T x\beta \end{aligned} \tag{2-10}$$

$$\frac{dS(\beta)}{d\beta} = -x^T y + (x^T x)\beta = 0 \tag{2-11}$$

$$\beta = (x^T x)^{-1} x^T y \tag{2-12}$$

If x has full column rank, $x^T x$ is positive definite, which is sufficient to establish the solution as a minimum.

2-2 SARIMAX-GARCH

In the previous section, a few time series modelling concepts were introduced as necessary background knowledge. In this section a generalized version of Auto-Regressive Integrated Moving Average (ARIMA) models is described. ARIMA is a very popular grey box modelling technique in the field of econometry.

SARIMAX-GARCH has a few extensions to the ARIMA structure in order to produce more accurate forecasts:

Seasonality If there is seasonality present within a time series, regular periodic trends can be identified which can be subtracted from the time series to enhance stationarity

External inputs Adding external inputs as regressors will frequently provide higher model accuracy in any forced system

GARCH If a series has a variance which evolves over time, it is called heteroskedastic. The Generalized Auto-Regressive Conditional Heteroskedasticity (GARCH) structure proposed by [24] is a way to correct for time-varying volatility in order to more accurately describe the system dynamics in terms of SARIMAX parameters

In GARCH modelling, it is common practice to first fit the data with a pure AR model with least squares parameter estimation. Then GARCH terms are added based on heteroskedasticity effects in its residuals and all model parameters are re-estimated. In this thesis, the underlying prediction model will not be pure AR, but the more general linear regression model SARIMAX. The overall schematic overview of SARIMAX-GARCH forecasting is depicted in Figure 2-2.

The rest of this section is structured as follows: the Seasonal Auto-Regressive Integrated Moving Average with eXogenous inputs (SARIMAX) model structure will briefly be discussed in Section 2-2-1, before GARCH extension is described in Section 2-2-2. Finally, a parameter estimation technique for fitting a SARIMAX-GARCH model is given in Section 2-2-3.

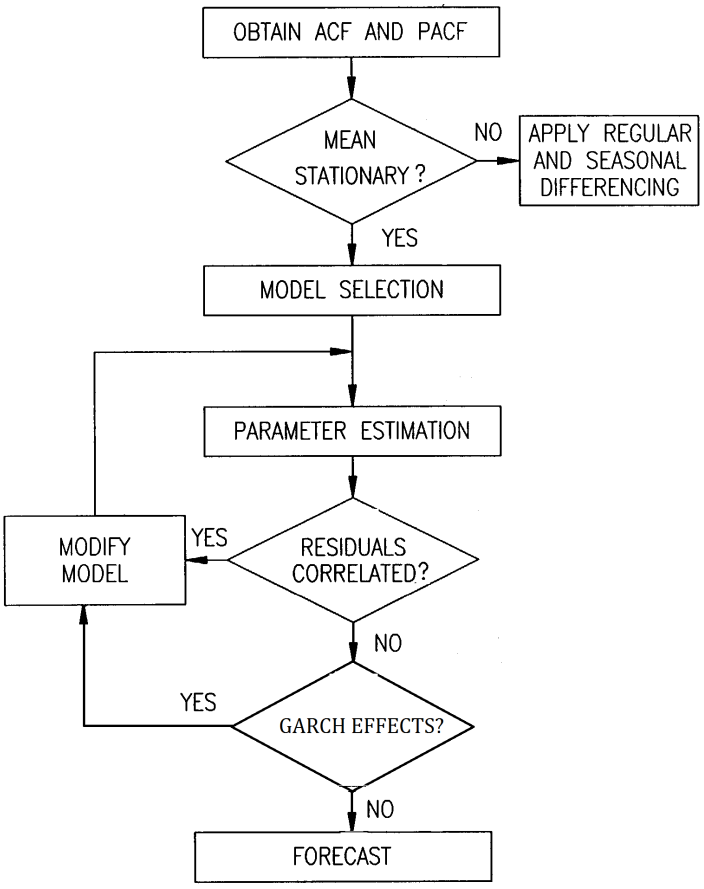


Figure 2-2: Schematic overview of SARIMAX-GARCH model determination (adapted from Fig. 10 in [1])

2-2-1 SARIMAX model

A popular regression model type for time series forecasting is the SARIMAX structure, which is linear in the parameters which enables linear regression techniques for estimating those parameters. SARIMAX type models fall within the category of multivariate regression models and have been used extensively in short term price predictions[25], which makes it viable to use SARIMAX as benchmark model when forecasting REP. A SARIMAX type model is an extension of the Auto-Regressive Moving Average (ARMA) class of models, which is a linear combination of the AR and MA basic model types described in Sections 2-1-4 and 2-1-3. The integrated part is introduced because of the requirement ARMA modelling imposes on a time series: it has to be WSS. A non-stationary series can often be rendered stationary by differencing the series to a sufficient order. A SARIMAX model is described by Equations 2-13 through 2-17.

$$\nabla^{N_d}\nabla_s^{N_{ds}} y_t = \frac{\theta(z)\Theta(z^s)}{\phi(z)\Phi(z^s)}\epsilon_t + \kappa(z)u_t \tag{2-13}$$

$$\phi(z) = 1 - \sum_{i=1}^{N_{AR}} \phi_i z^i \tag{2-14}$$

$$\Phi(z^s) = 1 - \sum_{i=1}^{N_{SAR}} \Phi_i z^{is} \tag{2-15}$$

$$\theta(z) = 1 + \sum_{i=1}^{N_{MA}} \theta_i z^i \tag{2-16}$$

$$\Theta(z^s) = 1 + \sum_{i=1}^{N_{SMA}} \Theta_i z^{is} \tag{2-17}$$

With the denotation s serving to distinguish between seasonal and non-seasonal components and:

z Lag operator: $z^n y_t = y_{t-n}$

$\phi, \Phi, \theta, \Theta$ Polynomials in z

∇_s Difference operator: $\nabla_s y_t = (1 - z^s)y_t$

y_t The predicted variable at time t

u_t Vector of external inputs

ϵ_t Prediction error

κ Input transfer function

N_{AR} Order of auto-regressive terms

N_{SAR} Order of seasonal auto-regressive terms

N_{MA} Order of moving average terms

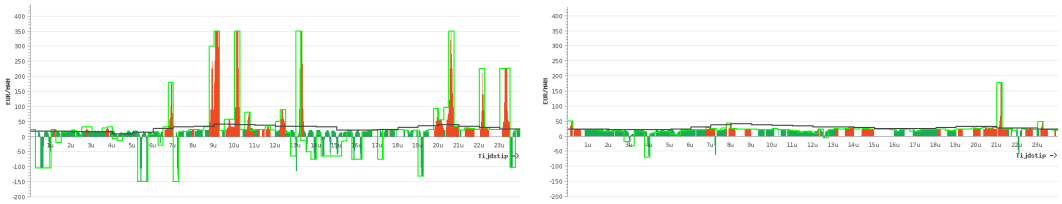
N_{SMA} Order of seasonal moving average terms

N_d Order of differencing required to render the series WSS

N_{D_s} Order of seasonal differencing

2-2-2 GARCH modelling

In time series where periods of high volatility alternate with periods of low volatility, stand-alone SARIMAX models which assume a constant variance will yield bad performance. In the real-time electricity market, volatility is time-dependent and price spikes tend to cluster. This is shown in Figures 2-3.



(a) 29-03-2016. The volatility is high **(b)** 31-03-2016. The volatility is low

Figure 2-3: Real-time electricity price for subtracting from the grid (REP_c) on two days exhibiting distinctly different volatility profiles

A way to correct for time-varying volatility is the GARCH structure proposed by [24]. GARCH updates the variance based on past squared values of the prediction error and past variances. As is shown in Equation 2-19, the prediction error is modeled as white noise with a time-varying variance. The variance σ_t^2 of the residual error term ϵ_t in a GARCH model is given by Equation 2-19. In order to allow for asymmetric effects of positive and negative prediction errors, [26] extended the determination of GARCH variance to Equation 2-20.

$$\epsilon_t = \sigma_t w_t \tag{2-18}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 \tag{2-19}$$

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{i=1}^{p} \beta_i \sigma_{t-i}^2 + \sum_{i=1}^{r} \gamma_i \epsilon_{t-i} \tag{2-20}$$

with

w_t Gaussian white noise $\mathcal{N}(0,1)$

ϵ_t Prediction error at time step t

ω Bias parameter

α Weighting vector of lagged prediction errors

β Weighting vector of previous variance values

γ Weighting vector of lagged prediction errors introduced to allow for asymmetric effects of positive and negative prediction errors

p, q, r Orders of the GARCH model

If GARCH is used as a model extension, the order of GARCH terms is iteratively determined in three steps.

- Estimation of underlying model parameters
- Computation of the residual autocorrelation to see if the underlying model is adequate
- Subjecting the squared residual sample autocorrelation to the Ljung-Box test to see if extra GARCH effects need to be added

The Ljung-box test which is used to test for GARCH effects in the residuals, checks for statistically significant lags in the squared residual autocorrelation. If there are such significant lags, GARCH terms are added to explain the correlation structure present in the variance[27].

In the previous section the SARIMAX model structure has been described, while this section introduced GARCH as an extension to incorporate dynamic time series variance in the model. The next section will describe the parameter estimation step of Figure 2-2.

2-2-3 Parameter estimation

When the model structure has been determined, the parameters still need to be estimated. For SARIMAX structures, the model is linear in the parameters, which enables LLS parameter optimisation. GARCH has been introduced as model extension to correct for a time-varying variance in the model. As the variance is now modelled in a nonlinear fashion, least squares optimisation is not applicable, but the principle of maximum likelihood can be exploited.

Maximum Likelihood estimation While LLS is suitable to estimate parameters of the initial SARIMAX model, model parameters have to be estimated in a different way after adding GARCH terms. As the prediction error is modelled to have a Gaussian distribution, the likelihood of a realization of ϵ_t is given by Equation 2-21. Realizations are assumed to be independent, so the joint realization is given by Equation 2-22.

$$L_t = \left(\frac{1}{\sqrt{2\pi\sigma_t^2}} \right) e^{-\epsilon_t^2/2\sigma_t^2} \tag{2-21}$$

$$L = \prod_{t=1}^{T} L_t \tag{2-22}$$

Taking the logarithm of the likelihood transforms the product into a sum, of which the derivative, and thus the argument of maximum likelihood, is more convenient to calculate.

The log-likelihood function is given in Equation 2-23. Once the prediction error ϵ_t of the SARIMAX model has been obtained and the GARCH model structure for σ_t has been selected (Equation 2-20), the quantities ϵ_t and σ_t^2 can be expanded, after which the parameters which maximize the log-likelihood function are determined numerically.

$$\ln L = -\frac{T}{2}\ln(2\pi) - 0.5\sum_{t=1}^{T}\ln\sigma_t^2 - 0.5\sum_{t=1}^{T}(\epsilon_t^2/\sigma_t^2) \tag{2-23}$$

In this section the model structure of SARIMAX-GARCH, which will be used as a benchmark model, has been described. A SARIMAX model, which is well known from econometry, is extended with GARCH effects to incorporate time-varying volatility in time series prediction. The next section will introduce another benchmark model, NARX.

2-3 Nonlinear Auto-Regression with eXogenous inputs network

The SARIMAX-GARCH model introduced in the previous Section assumes a linear correlation structure between time series values [28] with a nonlinear variance correction term. Moreover, it requires a model structure to be defined a priori. A popular nonlinear data-driven alternative is found in Artificial Neural Networks (ANNs). In order to process time series, feedback ANNs are used in which information about previous network states is fed to the network along with its external inputs in order to learn patterns. The Non-linear Auto-Regressive with eXogenous inputs (NARX) model class is reported by [29] to perform better than other feedback ANN structures in keeping track of long-term dependencies. This comes without loss of computational accuracy with respect to fully connected networks[30]. A fully connected network is a network in which all nodes are interconnected and the complete network state is fed back as additional input. In a NARX network, only the network output is used in a feedback loop, rather than the complete network state. The NARX principles are elaborated on in Section 2-3-1. Then, Section 2-3-2 elaborates on NARX structure determination before Section 2-3-3 briefly describes parameter estimation.

2-3-1 NARX modelling

In a NARX network, the estimated output is a function of past outputs and inputs of the time series, as shown in Equation 2-24.

$$\hat{y}(t+1) = \Gamma(y(t), y(t-1), ..., y(t-d_y), u(t+1), u(t), ..., u(t-d_u)) \tag{2-24}$$

A basic NARX type network is shown in Figure 2-4. External inputs at the current time step, as well as previous time steps are fed through the hidden layers of the network to produce a prediction which is looped back to serve as additional input.

The non-linear mapping which needs to be inferred in an identification experiment, Γ, consists of the connection weights between nodes in subsequent layers and the activation functions in the hidden and output layers. A commonly used activation function is the sigmoid activation function described by Equation 2-25. It is used as a continuously differentiable, monotonic

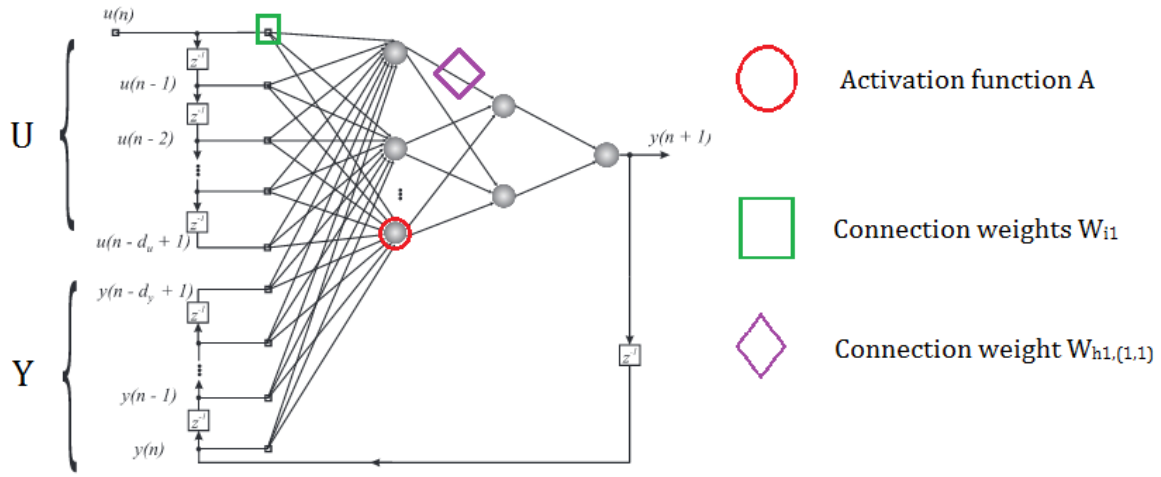


Figure 2-4: NARX type neural network with d_y output delays and d_u input delays and two hidden layers [2]. Each hidden layer transforms its inputs through activation function A (Equation 2-25). Each connection from input to hidden layer is weighted with an element from W_I. Connection weights from a hidden layer to the next layer are collected into the matrix W_H, of which a single weight $W_{h1,(1,1)}$, imposed on the connection of node 1 from the first hidden layer to node 1 of the second hidden layer, is highlighted

probability that a neuron is activated and the signal should thus be propagated through this node. The continuously differentiability guarantees gradient-based optimisation approaches are enabled, while it has been shown that nonlinearity in the activation function enables neural networks to approximate any continuous function to within arbitrary bounds[31]. A matrix notation for a single-layer NARX mapping is given in Equation 2-26.

$$A(\nu_i) = \frac{1}{1 + e^{-\nu_i}} \tag{2-25}$$

$$\Gamma = W_H A(W_I[Y \;\; U]) \tag{2-26}$$

With

A Sigmoid activation function

Γ Nonlinear mapping from network inputs to network output

W_H Connection weights on links between hidden layer nodes and outputs

W_I Connection weights on links between inputs to the network and hidden layer nodes

Y Input vector of past outputs

U Vector of current and past control inputs

The corresponding scalar version of Equation 2-24 with a single hidden layer is thus:

$$\hat{y}(t+1) = \sum_{k=1}^{N_h} \left[W_{hk} \left(1 + e^{-(\sum_{i=1}^{d_u} W_{Ii} u(t-i+1) + \sum_{j=1}^{d_y} W_{I,j+d_u} y(t-j+1))} \right)^{-1} \right] \tag{2-27}$$

with

N_h Amount of nodes in the hidden layer

W_{hk} Connection weight on the link between hidden layer node k and the output node

d_u Amount of external input delays

d_y Amount of output tap delays

The parameters to optimize in a NARX type neural network, are the connection weights W_H and W_I. Since the monotonicity of a sigmoid activation function renders the optimisation problem convex[32], least squares optimisation techniques are applicable.

2-3-2 Network structure determination

In a SARIMAX-GARCH model, the network structure is determined iteratively. For a NARX type neural network however, there are only two things to determine beforehand: the amount of hidden layers and model inputs. As its basis is still an AR model, selecting a relevant amount of feedback delays can be done based on inspection of the sample PACF.

Selecting the amount of hidden layers affects the severity of the transformation and can usually be done based on expert knowledge of the amount of transformations needed to go from input data to output data. If one takes the example of face recognition within a picture, it is nearly impossible to mark something as a face based on pixel intensities. However, transforming the raw pixel data into edge detectors and subsequently taking the detected edges as inputs to another layer piecing these edges together to recognize faces is more feasible. One can add in more intermediate steps to facilitate ease of detection, like grouping edges to features like mouths or eyes, then classifying part of a picture as a face if enough of these features are present within a bounded region.

If such expert knowledge about the amount of transformations from input to output is not available, there are still guidelines to select the amount of hidden layers in your network. [33] argues the following about the relationship between the amount of hidden layers and computational strength:

0 layers Only linearly separable functions or decisions can be represented

1 layer Any function which contains a continuous mapping from one finite space to another can be approximated

2 layers Arbitrary decision boundaries and smooth mappings can be approximated to arbitrary accuracy with rational activation functions

As the REP does not contain a continuous mapping, but rather an evolving discrete set of values, two hidden layers will be used.

2-3-3 Parameter estimation

After determination of the NARX network structure, connections weights between subsequent layers have to be optimised. Since the monotonicity of a sigmoid activation function renders the optimisation problem convex[32], least squares optimisation techniques are applicable. In a nonlinear least squares optimisation without constraints, the Levenberg-Marquardt algorithm is a standard technique to obtain a solution for the parameter vector[34]. It is based on a combination of a gradient descent method and the Gauss-Newton method. The generalized objective function OF to minimise in a least squares problem formulation, is the squared error criterion given by Equation 2-28.

$$OF(\beta) = (y - \hat{y}(\beta))^T W (y - \hat{y}(\beta)) \tag{2-28}$$

with

β Parameter vector

W Weighting matrix

The LLS problem is contained in this formulation as a special case: if the estimated function $\hat{y}(\beta)$ is linear in the parameters and the weighting matrix W is set to unity, the objective function reduces to Equation 2-10.

In the nonlinear case minimisation becomes an iterative process the details of which can be found in [34].

In this section the NARX neural network has been described. It will be used as a benchmark model alongside the SARIMAX-GARCH model which was topic of the previous section. The next section presents the time series prediction model which is proposed in this thesis in order to forecast REP.

2-4 Generalized Fuzzy Neural Network formulation of a NARX model structure

The previous sections described SARIMAX-GARCH and NARX which will be used as benchmark models. SARIMAX-GARCH assumes a linear correlation between subsequent time series values, which might not yield accurate predictions in case of the highly nonlinear REP time series. NARX provides a nonlinear model which is thought to yield more accurate REP predictions, but has as main drawback that the relation between inputs and output is not interpretable. It is yet unclear which observable factors contribute to the REP, so an interpretable model is key towards selecting appropriate inputs and learning something about

the underlying system. As a way to incorporate human reasoning expressed in natural language in modelling, fuzzy systems have been developed. Fuzzy systems thus naturally yield a linguistic interpretation.

In this Section the proposed GFNARX model class is presented. A brief introduction to fuzzy logic is given before the working principles of the Generalised Fuzzy Neural Network (G-FNN) algorithm, which serves as a basis for the proposed model, are described. Subsequently, improvements to this G-FNN model are discussed, leading to the proposed model structure.

2-4-1 Fuzzy logic

In this paragraph, a brief introduction to rule-based fuzzy inference systems will be given, based on [3]. While fuzzy systems are not limited to rule-based inference, in this thesis only rule-based fuzzy systems will be addressed.

When using natural language to express quantities which can be used in a model, one will use imprecise phrases like *The temperature is high* or *The temperature is low*. However, transition between the two is gradual instead of having a crisp boundary. As there is a gradual transition, there is overlap between the linguistic labels, making it possible for a temperature to partly belong to both categories. Hence the classification is called fuzzy.

In fuzzy inference systems, reasoning is typically of the general form **IF** *circumstance* **THEN** *consequence*, which is called a rule. The circumstances and consequences can be fuzzy propositions or numerical functions. Fuzzy propositions are characterized by linguistic labels, which represent a fuzzy set of a variable. An example of a variable with three fuzzy sets and its corresponding labels is found in Figure 2-5. The membership functions quantify to which extent a numerical value of the base variable belongs to a fuzzy set, while rules put a restriction on the simultaneous occurrences of variable values.

In fuzzy inference systems, the circumstance is always a fuzzy proposition, while the consequence structure can be broadly categorized into two:

- Linguistic reasoning, in which the consequences are also fuzzy propositions which are derived from the rules and the inputs. A well-known linguistic fuzzy inference model is the Mamdani model [35]. A visualized example of such a fuzzy inference model with two inputs and two rules is found in Figure 2-6. The reasoning in Figure 2-6 has as structure **IF** *circumstance* x_0 *AND circumstance* y_0 **THEN** *consequence* z. The consequence of each rule is triggered by the minimum fulfilment degree of the two circumstances, while the final output distribution is governed by the maximum consequence at each point.

  If a numerical value is required as output, this can be inferred from the output distribution. For example, when regarding the fuzzy partitioning of temperature in Figure 2-5, with corresponding singleton outputs:

  IF *temperature is low* **THEN** *comfort is 10*

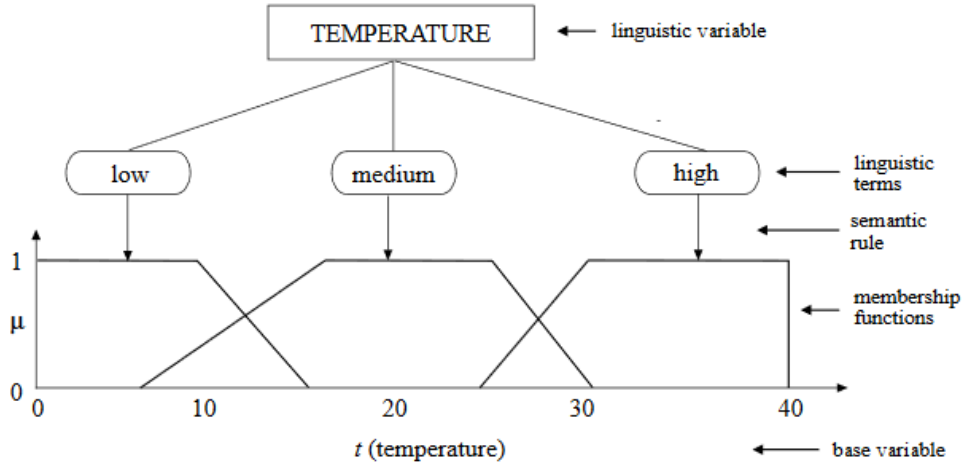


Figure 2-5: Example of a linguistic variable "temperature" with three linguistic terms[3]

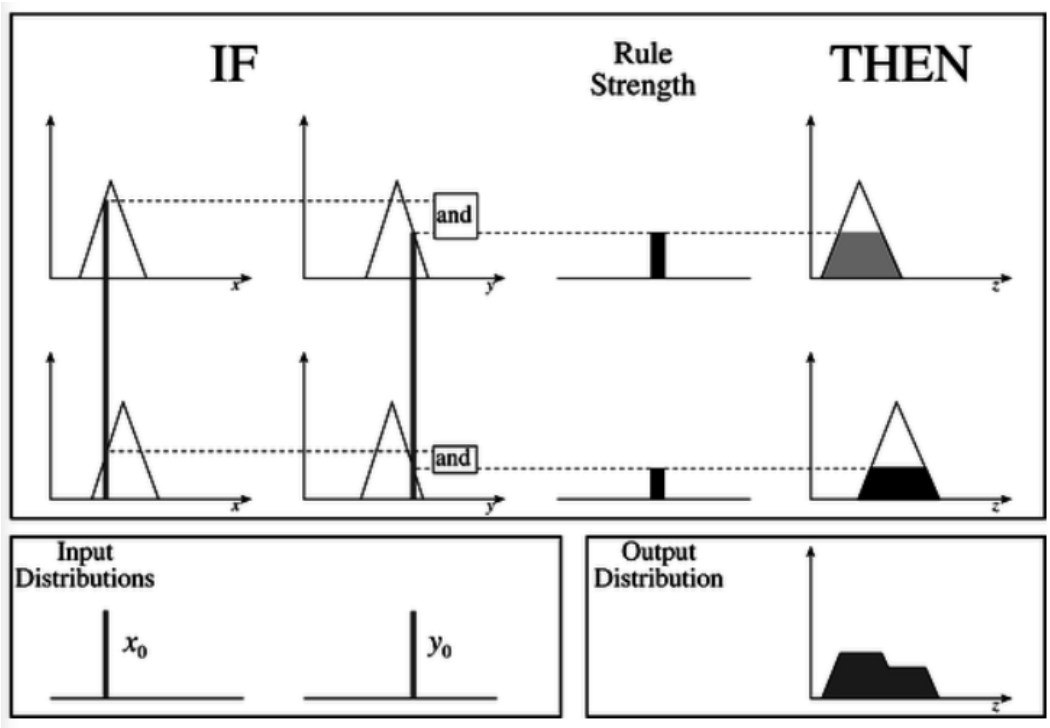


Figure 2-6: Example of Mamdani inference in a model with inputs $x0$ and $y0$ and output z [4]

IF *temperature is medium* **THEN** *comfort is 20*

IF *temperature is high* **THEN** *comfort is 10*

Then, if the temperature $t = 13$, $\mu_{low} = 0.2$, $\mu_{medium} = 0.7$, $\mu_{high} = 0$; resulting in $comfort = (0.2 * 10 + 0.7 * 20)/0.9) = 17.8$.

- Takagi-Sugeno-Kang (TSK) type fuzzy inference, in which the output is a function of the inputs instead of a fuzzy set. Rule firing strength here can be seen as the validity of using a local model to approximate the model output. The overall output $y(z)$ for input z is retrieved from the individual local models as:

$$y(z) = \frac{\sum_i \mu_i y_i(z)}{\sum_i \mu_i} \tag{2-29}$$

2-4-2 Rule base construction

In any fuzzy model, a rule base has to be constructed. The rule base is the agglomeration of all individual rules which govern the output distribution. Most fuzzy models known from literature take one of the following three approaches to constructing the rule base:

- Human expert knowledge is translated into linguistic if-then clauses, which are then described by fuzzy sets in premise and consequent parts[36]
- The input data space is partitioned into clusters. In this approach, either the amount of clusters is pre-defined[37], or a suitable amount of clusters is found by iteratively reducing the amount of clusters from an initial partitioning with a redundant amount of clusters[38]
- Region growing methods. In these kind of methods there is only a single cluster initially, and incoming data is either added to an existing cluster if it s similarity to this cluster is larger than a predefined threshold, or results in a new cluster being created around this data point

If a rule base is constructed based on human expert knowledge, the rules and membership function implementations are in general static. In an environment which is dynamic and not fully understood, such an approach is likely to fail. Moreover, as was argued by Zadeh 1973, *"as the complexity of a system increases, our ability to make precise and yet significant statements about its behavior diminishes until a threshold beyond which precision and significance (relevance) become almost mutually exclusive characteristics"*[39].

As a suitable amount of clusters is hard to determine a priori when it is unknown which inputs are relevant, region growing methods are well suited to identifying a model for the real-time electricity market while simultaneously assessing input significance.

2-4-3 Generalized Fuzzy Neural Network model building algorithm

In the previous section it was argued that region growing methods are well suited to rule base construction when identifying a fuzzy model of the real-time electricity market. A data-driven TSK type region growing method which automatically clusters the input data into

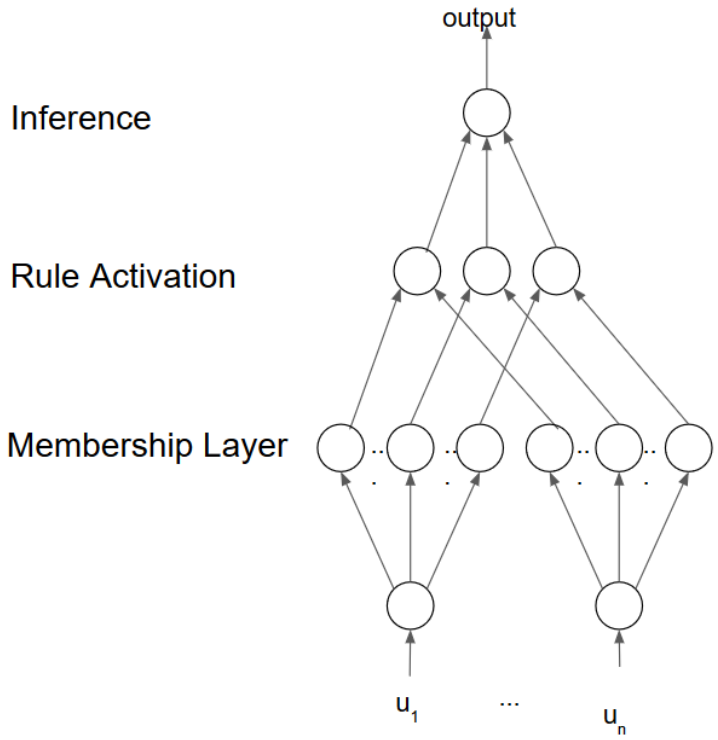


Figure 2-7: Structure of a fuzzy neural network

fuzzy sets and adapts the rule base online is the G-FNN learning algorithm proposed by [5]. This G-FNN uses fuzzy logic within a neural network framework. The general idea is depicted in Figure 2-7. Each input has its corresponding fuzzy sets of which a membership grade can be determined. These membership grades are fed through to the rule activation layer. Then the inference mechanism determines the output based on the activation of rules.

In a G-FNN, the mapping from inputs to outputs is defined as[5]:

$$\hat{y}(t) = \sum_{j=1}^{n_r} w_j e^{-(u(t)-c_j)^T V_j (u(t)-c_j)} \tag{2-30}$$

with

u Input vector of length N_u

w_j Connection weights from the inputs to rule j: $w_j = k_{0,j} + \sum_{i=1}^{N_u} k_{i,j} u_i$

c_j Center vector of the Gaussian membership functions

V_j Covariance matrix of the Gaussian membership functions of rule j: $V_j = \text{diag}\left(\frac{1}{\sigma_{1,j}^2}, ..., \frac{1}{\sigma_{N_u,j}^2}\right)$

During the training process, both the network structure and parameters are learned iteratively. A flowchart of the training process is shown in Figure 2-8.

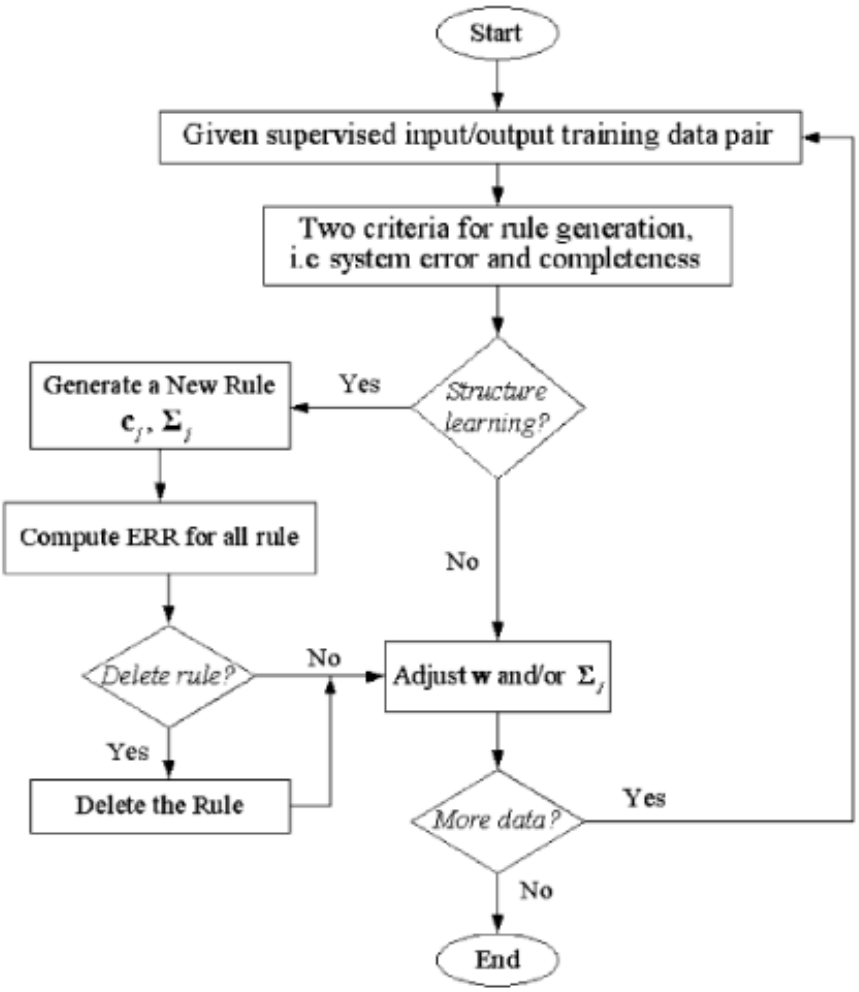


Figure 2-8: Flowchart of the G-FNN learning algorithm[5]

Structure learning constitutes determination of the proper amount of fuzzy logic rules and their corresponding membership function shapes. This is based on two criteria, the prediction error and completeness of the model. For every new input-output training data pair, there are three potential cases:

1. $\epsilon_t > K_e$ and $d_t > K_d$. A new rule is generated

2. $\epsilon_t > K_e$ and $d_t \leq K_d$. An existing rule is updated

3. $\epsilon_t \leq K_e$. No structural changes are applied

with

ϵ_t Prediction error at time step t: $\epsilon_t = ||y_t - \hat{y}_t||$

K_e Maximum permissible prediction error

K_d Maximum distance between an observation and the nearest cluster: $K_d = \sqrt{ln\frac{1}{\varepsilon_c}}$

ε_c Threshold for membership grade. For each input sample $u(t)$, $\max_j(\mu_j(z(t))) \geq \varepsilon_c$ should hold

d_t Normalised distance from the input observation to the nearest cluster. This distance, called the Mahalanobis distance, is described by Equation 2-31. A two-dimensional comparison to Euclidian distance is shown in Figure 2-9

$$d(t) = \min_j \sqrt{(u(t) - c_j)^T V_j (u(t) - c_j)} \tag{2-31}$$

If a new membership function is added, the center and width vectors are determined by Equations 2-32 through 2-35.

$$D_{i,j}(t) = ||u_i(t) - b_{i,j}|| \tag{2-32}$$

$$x = \underset{j}{\text{argmin}}\ D_{i,j}(t) \tag{2-33}$$

$$\text{if} \quad D_{i,x}(t) < K_{mf} : c_{i,n_r+1} = b_{i,x}, \quad \sigma_{i,n_r+1} = \sigma_{i,x} \tag{2-34}$$

$$\text{if} \quad D_{i,x}(t) \geq K_{mf} : c_{i,n_r+1} = u_i(t), \quad \sigma_{i,n_r+1} = (\frac{D_{i,x}}{K_d})^2 \tag{2-35}$$

with

$b_{i,j}$ Boundary point from the set $\{c_{i,1}, ..., c_{i,n_r}, u_{i,min}, u_{i,max}\}$

n_r Amount of fuzzy rules

K_{mf} Threshold of dissimilarity of neighbouring membership function, which regulates the maximum amount of partitions of the input space

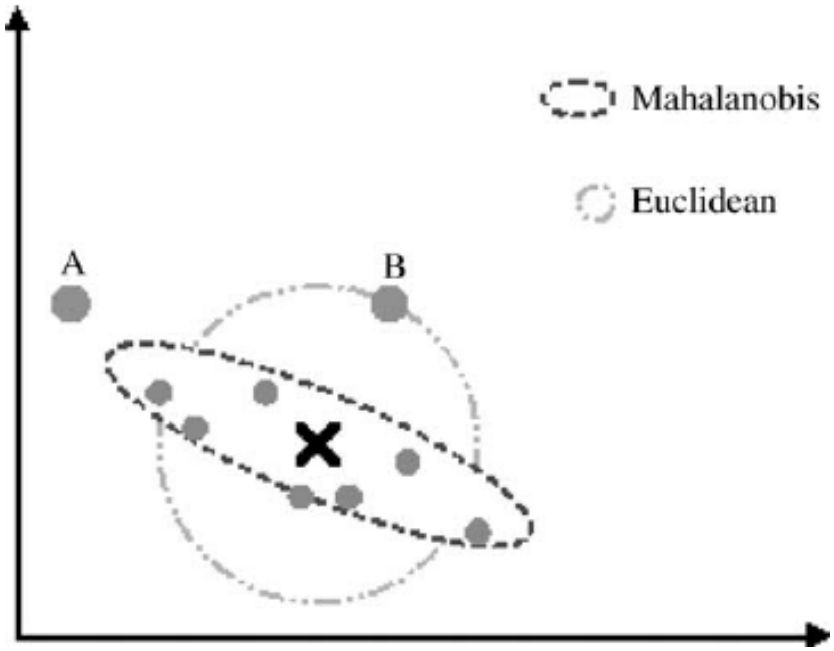


Figure 2-9: Comparison of Euclidean and normalised distance measures (Fig. 3 of [6]). In Euclidian distance, point B is closest to the cluster. Using the normalised Mahalanobis distance as measure however, point A is noticeably closer to the cluster

In case 2, the new data point falls within an existing cluster, yet the prediction error is too large. The width of the cluster is reduced according to Equation 2-36 in order to increase the chance a potential new cluster will emerge if more training data is added.

$$\sigma_{i,x} := K_s \sigma_{i,x} \tag{2-36}$$

with

K_s Reduction factor. It is recommended by [5] to choose $0.9 \leq K_s \leq 1$ based on the sensitivity of rule x to input dimension i

Redundant clusters are pruned by assessment of cluster influence on the estimated output. Equation 2-30 can be rewritten as a product of a regression vector Φ with weights w:

$$\hat{y}(t) = \Phi(t) w \tag{2-37}$$

with

$\Phi \in \mathcal{R}^{t\,x\,\nu}$

$\nu = n_r(n_u + 1)$

n_r Amount of fuzzy rules

n_u Amount of inputs to the network

$w \in \mathcal{R}^{\nu}$

Then, performing a QR decomposition, $\Phi = QR$, the Error Reduction Ratio (ERR) term in Figure 2-8 is calculated as[5]:

$$ERR_k = \frac{(q_k y)^2}{||q_k||^2||y||^2} \tag{2-38}$$

with

y output time series

q_k kth column of the Q matrix

Finally, all ERR_k terms corresponding to rule j are collected in a vector $ERR(j)$, and rule j is deleted if:

$$TERR(j) = \sqrt{\frac{||ERR(j)||^2}{n_u + 1}} < K_{err} \tag{2-39}$$

with

$TERR(j)$ Total ERR due to inclusion of rule j

K_{err} A designed threshold. According to [5], it is not tractable to design it analytically and $K_{err} = 0.005$ is a decent initial guess

Once the structure is in place, the weight vector w is estimated through a LLS procedure as described by Equation 2-40.

$$w = (\Theta^T \Theta)^{-1} \Theta^T y \tag{2-40}$$

Where $\Theta = [\Phi(1)\, \Phi(2) \ldots \Phi(t)]^T$. If the consequent TSK models depend linearly on their parameters, the least-squares solution is optimal.[3]

2-4-4 Issues with G-FNN

G-FNN as described in the previous section has a few shortcomings in its algorithm as well as theoretical flaws in its description, which leads to believe the authors did not succesfully implement the algorithm as described in [5]. The computationally inexpensive, dynamically self-organizing and data-driven nature, however, are attractive as basis to improve upon. In this section, a few weaknesses of the G-FNN are highlighted.

1. The G-FNN algorithm allows numerical instability. As there is no restriction on addition of rules, it can happen in G-FNN that the regression matrix becomes singular. Also, when clusters overlap, the output can potentially become unbounded, as there is no normalisation of rule activation

2. In the G-FNN algorithm, clusters are only updated in terms of their widths. The cluster centres are static after initialisation, unless the rule pertaining to these centres is pruned. This results in a loss of accuracy when compared to any method which optimises the cluster locations

3. In G-FNN, pruning of redundant rules is based on the ERR, which ranks rule importance according to its total contribution to model output in the training data. It has been proven by [7] that this method can produce an incorrect subset of fuzzy rules, because it neglects potential redundancy in input partitioning

4. When creating a new rule, minimum distance to existing clusters is defined in G-FNN by the Euclidian norm, while the need to create a new rule is governed by Mahalanobis distance. This is inconsistent. A normalised distance like the Mahalanobis distance is better suited to identify parts of the input space which are already covered, especially in the case where inputs do not have the same dimensions, because Euclidian distance will then underestimate deviations from inputs with relatively small values

5. Defining new cluster widths in terms of Euclidian distances suffers from the same limitation imposed by scale differences. The new cluster widths are governed by the most extreme input dimensions, which either causes the new cluster to cover a very small part of the input space (governed by smallest dimension) or causes large overlap with existing clusters (governed by largest dimension)

6. The local models in the consequent part of the G-FNN are TSK style affine models. This means that at every cluster centre, a locally linear approximation is made to the signal, between which is interpolated in a nonlinear manner. In case of at least twice differentiable target signals, Taylor's theorem suggests that an improvement to the accuracy of the model, as well as speed of convergence, is achieved by using local higher order polynomial models

With the weaknesses of G-FNN identified in this section, improvements to the G-FNN algorithm which address these weaknesses can be made. This will be done in the next section.

2-4-5 GFNARX modelling approach

In this section, improvements to the G-FNN algorithm proposed by [5] are suggested. These improvements solve issues described in Section 2-4-4. At the end of this section, the resulting GFNARX algorithm is given.

Bounding the amount of rules Issue 1 as described in Section 2-4-4 is that there is no restriction on the addition of rules. Prediction error and overall dissimilarity to existing rules are the only two criteria mentioned with regards to updating the rule base in the G-FNN description. However, there are two scenarios in which new rules should not be added:

- The addition of a new rule would render the regression matrix near singular. As the weight parameters are determined by the pseudo-inverse of the regression matrix, singularities in the regression matrix lead to unstable solutions
- The regression problem is underdetermined. Adding a new rule adds $n_u + 1$ columns to the regression matrix. If there are not enough observations, the amount of rows in the regression matrix is lower than the amount of columns, hence the matrix can never have full column rank, as is required for a stable solution to Equation 2-40

To allow the least squares parameter optimisation to assign meaningful weights to new rules, the regression matrix is rebuilt with the new fuzzy partitioning whenever a structural change has been applied. If one of the two cases above would happen, the new rule is not added.

Moving average cluster centring One of the weaknesses of the G-FNN is that it does not optimise its cluster parameters after creation. This is solved by adapting the cluster center of the nearest rule through averaging as in Equation 2-41 with every incoming training data pair. In this fashion, a cumulative moving average implementation of clustering is created which, combined with the rule base cluster segmentation method, optimises the cluster locations online.

$$c_{i,j} = \begin{cases} \dfrac{c_{i,j} n_{cl,j} + u_i}{n_{cl,j} + 1} & \text{if } j = \text{argmin}_j \sqrt{(u(t) - c_j)^T V_j (u(t) - c_j)} \\ c_{i,j} & \text{otherwise} \end{cases} \tag{2-41}$$

with

$c_{i,j}$ Cluster center of rule j at input i

$n_{cl,j}$ Amount of data points attributed to cluster j

$u(t)$ Input vector at time t

V_j Width matrix of the Gaussian membership functions of rule j: $V_j = \text{diag}\left(\frac{1}{\sigma_{1,j}^2}, ..., \frac{1}{\sigma_{N_u,j}^2}\right)$

Mahalanobis Distance as cluster separation criterion G-FNN uses Euclidian distance to determine whether parts of input space are already covered. Euclidian distance is unsuitable as distance measure when inputs have different scales. A unitless and scale-invariant distance measure, which is already used in another part of G-FNN is the Mahalanobis distance, which measures the distance between a data point and a distribution in terms of standard deviations from the distribution centroid.[40] Thus, when adding a new rule to the rulebase, a new cluster is only added when input $u_i(t)$ is separated from all existing clusters by more than distance threshold K_{md} standard deviations.

Instead of defining new cluster widths in terms of the geometrically closest neighbouring cluster center, new cluster widths are determined by calculating the permissible widths for each dimension based on maximum overlap with neighbouring clusters and selecting the minimum of these two widths. Since cluster widths can be reduced, but are not expanded in both G-FNN and GFNARX, it can be considered relatively safe to initialize the cluster width in such a way that the normalised distance of its maximal intersection with the nearest neighbour is exactly equal to the separation threshold K_{md}. By initializing the cluster in this manner, maximal coverage of the input space void between existing clusters is achieved while avoiding redundancy. This solves the problems with using an Euclidian distance threshold as separation measure.

The intersection of two membership functions of fuzzy sets A and B, based on the product t-norm which is the standard t-norm in fuzzy neural networks, is defined as:

$$\mu_A \cap \mu_B = \mu_A \mu_B = e^{-\left(\left(\frac{u-c_A}{\sigma_A}\right)^2 + \left(\frac{u-c_B}{\sigma_B}\right)^2\right)} \tag{2-42}$$

The product t-norm intersection of two Gaussian functions is a concave function, and hence there is a unique global maximum to the intersection which can be found by differentiating and equating the gradient to zero.

$$\frac{\partial \mu_A \mu_B}{\partial u} = \left(\frac{2u-2c_A}{\sigma_A^2} + \frac{2u-2c_B}{\sigma_B^2}\right) e^{-\left(\left(\frac{u-c_A}{\sigma_A}\right)^2 + \left(\frac{u-c_B}{\sigma_B}\right)^2\right)} = 0 \tag{2-43}$$

In a bounded input space, this means that equating the first bracketed term on the right-hand side of Equation 2-43 to zero will yield the solution. Solving for u, substituting the solution into Equation 2-42 and rearranging terms, the following maximum is found:

$$max(\mu_A \cap \mu_B) = e^{-\frac{(c_A - c_B)^2}{\sigma_A^2 + \sigma_B^2}} \tag{2-44}$$

As the cluster centers as well as one of the width matrices are known, the only unknown in Equation 2-44 is the width matrix of the new cluster. By similarity to the Mahalanobis distance (Equation 2-31), the maximum new width in input dimension i can then be solved for:

$$\sigma_{i,n_r+1} = \min_j \sqrt{|(c_{i,j} - c_{i,n_r+1})^2 - K_{md}/\sigma_{i,j}||/K_{md}} \tag{2-45}$$

Rule pruning by fraction of activation The ERR which is used in G-FNN to select a significant subset of rules is prone to overestimating the importance of redundant rules. A number of methods have been suggested by [7] to either extend the ERR or improve on it by, amongst others, pivotal QR decomposition, but in the end all of them rely on computationally relatively expensive orthogonal decompositions. In this work, a simple check, which hardly requires any memory or computational power, for rule significance is proposed in terms of fraction of activation: $f_{act,j} >= K_{act}$. An additional benefit of doing so, is that it also puts an interpretable upper limit on the amount of rules the model can have. For example, if $K_{act} = 0.025$, the absolute maximum amount of rules is $n_{r,max} = 40$. The model will avoid over-fitting on outliers by pruning rules which have less than K_{act} chance of activating, based on historical data. In order to render the model adaptive to changing circumstances, this activation-based pruning is coupled to a rolling window of historical training data which gradually forgets old data.

GFNARX algorithm On top of the improvements described in the previous paragraphs, a few minor improvements to G-FNN are implemented in GFNARX:

- Activation of rules is normalised in order to increase the consistency of predictions in output space. As the membership support in a fuzzy neural network is not unilaterally described by $\sum_j \mu_j = 1$, this is necessary and common practice in most fuzzy models

- The pseudo-inverse of the regression matrix is calculated by use of a QR-decomposition, instead of directly computing the inverse in Equation 2-40. This enhances numerical stability[41]

The overall proposed GFNARX algorithm can thus be described by:

1. Initialize first rule based on first training data pair
2. New incoming training data: calculate prediction error $e(t) = ||y(t) - \hat{y}(t)||$ and Mahalanobis distance (Equation 2-31)
3. If $e(t) \leq K_e$, no structural changes are applied; skip to 8. Otherwise continue to 4
4. If $d(t) \leq K_{md}$, reduce width matrices of the nearest rule
5. If $d(t) > K_{md}$, generate a new rule candidate
6. If the regression matrix becomes singular, reject candidate; skip to 8. Otherwise continue to 7
7. Redistribute available data over clusters
8. Attribute new data to clusters; update clusters according to Equation 2-41.
9. If $f_{act,j} < K_{act}$, delete rule j
10. Adjust regression matrix
11. Adjust weight matrix through least squares estimation
12. If regression matrix size over the historical data buffer limit: forget oldest data point
13. return to 2

In this section, issues with G-FNN identified in Section 2-4-4 are addressed. As a result, the improved GFNARX algorithm has been derived. The next section will describe how the identified fuzzy neural network model can be interpreted in linguistic terms.

2-4-6 Dimensionality reduction and linguistic interpretation

The aim of this project is accurate, real-time and interpretable prediction of the electricity market price. To achieve accurate, real-time prediction the GFNARX algorithm has been proposed. In order to retrieve an interpretable subset of the identified model, similarity-driven rule base simplification as proposed by [7] is employed. The pairwise similarity between fuzzy sets can be quantified by the fuzzy Jaccard index, defined as the intersection of two sets divided by the union, which is described in Equation 2-46.

$$JI(A_{li}, A_{mi}) = \frac{min(|\mu_{A_{li}}(u_i), \mu_{A_{mi}}(u_i)|)}{max(|\mu_{A_{li}}(u_i), \mu_{A_{mi}}(u_i)|)} \tag{2-46}$$

where

JI Jaccard index

μ Membership degree

u_i Input i

A_{ri} Fuzzy set of rule r at input i

l, m Rule indices

$|\cdot|$ Cardinality of a fuzzy set, $|\cdot| = \int_{u_i \epsilon Z_i} \cdot$

Sets which are similar to at least degree λ_s can be merged. If this simplification results in rules becoming equal, only one of the rules needs to be kept.

In this chapter, the GFNARX algorithm has been proposed as improvement to the G-FNN algorithm found in literature. As the goal is to predict REP, for which currently no benchmark method is known, theory has also been provided on models which will be used as benchmark models. SARIMAX-GARCH is a generalised version of the ARIMA model structure popular in econometry, while NARX is a neural network structure specifically catered towards time series prediction. In the next chapter, the experiments which will be used to assess the performance of GFNARX are described and the REP data is investigated in order to pre-process it into a time series suitable for identification purposes.

Chapter 3

Experiments

In this chapter the experiments which have been carried out to validate the performance of Generalised Fuzzy Neural Network formulation of a Non-linear Auto-Regressive with eXogenous inputs model structure (GFNARX) are described and plausible model inputs towards forecasting the Real-time Electricity market Price (REP) are found. GFNARX is proposed as an improvement to Generalised Fuzzy Neural Network (G-FNN) in Section 2-4-4. As there is no benchmark method found in literature which deals with REP prediction, GFNARX performance is first benchmarked on a time series prediction problem popular in fuzzy neural network literature. This time series, the Mackey-Glass series, and the benchmark experiment corresponding to it are described in Section 3-1.

GFNARX uses a method of selecting a significant subset of rules which does not rely on a computationally expensive orthogonal decomposition like G-FNN. To establish that this subset selection is able to filter redundant and inconsequential rules from the rule base, a literature problem which deals with identifying insignificant rules from an existing rule base is described in Section 3-2.

After testing GFNARX performance on literature problems, the real-time market data is examined in Section 3-3 in order to develop the REP model identification experiment. Finally, a novel benchmark method for assessing economic value of REP predictions is introduced in Section 3-4. REP predictions of the models described in Chapter 2 will be used to determine cooling motor behaviour of a Cold Storage Warehouse (CSW).

3-1 Mackey-Glass chaotic time series

This section presents the Mackey-Glass time series, which will be used to benchmark GFNARX model accuracy to other fuzzy neural network approaches. The Mackey-Glass chaotic time series is a well known method found in fuzzy neural network literature to benchmark the

performance of a time series prediction model. Since it is not known whether the GFNARX model is capable of predicting the highly volatile REP for which no benchmark is yet available, the contribution of GFNARX to time series forecasting literature will be supported by this Mackey-Glass prediction benchmark.

The Mackey-Glass time series is generated from a differential equation which has originally been used to characterize dynamics of physiological feedback systems [42], and is described by Equation 3-1. The most common form of this equation, when used for time series prediction benchmarking purposes, uses as parameters $\beta = 0.2$, $\gamma = 0.1$, $n = 10$ and $\tau = 17$.

$$y(k) = \frac{\beta y(k-\tau)}{1 + y^n(k-\tau)} + (1-\gamma)y(k-1) \tag{3-1}$$

with $y(0) = 1.2$.

The values of the time series are predicted by using four past values of the series itself:

$$\hat{y}(k) = f(y(k-24), y(k-18), y(k-12), y(k-6)) \tag{3-2}$$

The first 123 values generated by the difference equation are used to initialize the series, after which 500 training samples are seen and the obtained model is validated on 500 more samples. The model performance metric used is Root-Mean-Square Error (RMSE). RMSE of GFNARX prediction of the 500 test samples is compared to RMSE scores reported in literature.

In this section a time series model identification problem popular in fuzzy modelling literature has been described. The performance of GFNARX will be benchmarked against results reported in literature for this identification experiment, the results of which are discussed in Section 4-1. The next section introduces another experiment known from literature, which will test the ability of GFNARX to clean up inconsequential rules from the rule base.

3-2 Rule base pruning by fraction of activation

During training of the GFNARX model, rules are generated online. To avoid overparametrising the model, which will result in overfitting the training data and thus less prediction accuracy, it is necessary to also remove rules from the rule base if their significance is low. As part of GFNARX, a rule pruning method based on total fraction of activation is proposed in Section 2-4-4. As this method does not make use of orthogonal transforms to rank and reduce the amount of rules, it has to be validated that this method works well in practice. A problem found in fuzzy literature dealing with dimensionality reduction by identifying insignificant rules in an existing rulebase, is provided by [8]. This modelling problem has also been included in [7] in which it is shown that several existing methods to rank rules, among which the Error Reduction Ratio (ERR) used in G-FNN are inaccurate.

In this rule subset selection exercise, the modelled system is a nonlinear plant with two auto-regressive inputs and an unmeasured disturbance, described by Equations 3-3 and 3-4.

$$y(k) = f(y(k-1), y(k-2)) + u(k) \tag{3-3}$$

with

$$f(y(k-1), y(k-2)) = \frac{y(k-1)y(k-2)\left[y(k-1) - 0.5\right]}{1 + y^2(k-1) + y^2(k-2)} \tag{3-4}$$

with

$f(\cdot)$ Mapping of the unforced system. It has an equilibrium state at (0,0)

y Plant output

u Disturbance input

k Time step

Following the approach in [8], 1200 data points are generated, 1000 of which are obtained by using a random disturbance $u(k)$ uniformly distributed in the interval $[-1.5, 1.5]$, followed by 200 sinusoidal disturbances $u(k) = sin(2\pi k/25), k = 1001, ..., 1200$. This constitutes a signal-to-noise ratio of order -10dB. Starting from a given initial rule base, it is the task to identify the redundant rules and order the rules in terms of importance.

In this modelling exercise, a fuzzy rule is specified by its corresponding cluster centre and width parameters. The relevant parameters of the initial rule base are given in Table 3-1, while the input space partitioning is visualized in Figure 3-1. It can be seen that some rules are nearly duplicates of one another and there are two rules which are so narrow that they are essentially fuzzy singletons, which are bound not to contribute much in the presence of strong noise. It has to be noted that membership function widths σ_{ij} in Table 3-1 are $0.5\sqrt{2}\sigma_{ij}$ of GFNARX due to the specification of a fuzzy set A_{ji} of rule j at input i as:

$$A_{ji}(u_i(k)) = \exp - \frac{(u_i(k) - c_{ji})^2}{2\sigma_{ji}} \tag{3-5}$$

In [8] it has been shown that model performance is not degraded by removing one of the rules of each redundant pair and the two rules with overly narrow clusters. In this experiment it is the aim to obtain a significance ordering for the rules and verify that if a subset of 20 rules is selected based on this ordering, there are no more redundant or non-firing rules in the rule base.

Table 3-1: Parameters of the Gaussian membership functions in the experiment of [8]. It can be seen from the bold lines that there are insignificant membership functions due to redundancy and due to membership functions being too narrow

rule	center c_{i1}	center c_{i2}	width σ_{i1}	width σ_{i2}
1	**0.0930**	**-0.3630**	**0.7095**	**0.7095**
2	**0.0933**	**-0.3632**	**0.7095**	**0.7095**
3	1.3828	-0.6617	0.6271	0.6271
4	-1.0414	1.5397	0.7969	0.7969
5	**-1.8130**	**-1.6470**	**1.3205**	**1.3205**
6	**-1.8125**	**-1.6469**	**1.3205**	**1.3205**
7	0.7776	-1.1555	0.7800	0.7800
8	0.1898	1.0142	0.6141	0.6141
9	-0.4052	0.2798	0.8099	0.8099
10	**-0.6613**	**-0.4846**	**0.0100**	**0.0100**
11	-0.6613	-0.4846	0.7051	0.7051
12	0.9529	-0.3965	0.6313	0.6313
13	0.7860	0.7723	0.6177	0.6177
14	0.4329	0.1910	0.6652	0.6652
15	**1.2940**	**1.0740**	**0.6474**	**0.6474**
16	**1.2942**	**1.0738**	**0.6474**	**0.6474**
17	0.6801	1.4083	0.6370	0.6370
18	1.2656	0.2698	0.7156	0.7156
19	-0.3846	1.1827	0.6772	0.6722
20	**-1.2642**	**-0.1808**	**0.0100**	**0.0100**
21	-1.2642	-0.1808	0.7907	0.7907
22	-0.9099	-1.1750	0.7728	0.7728
23	-1.008	-1.1384	0.8046	0.8046
24	-1.1533	0.7037	0.8517	0.8517
25	1.7691	-1.2798	0.8746	0.8746

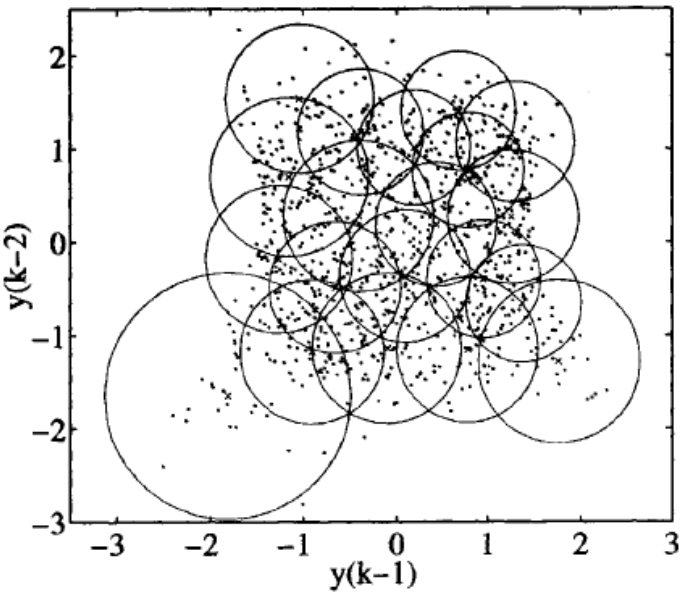


Figure 3-1: Clusters in the input space [7]

In this section an experiment, known from fuzzy modelling literature, has been introduced. This experiment checks whether a metric to assess the significance of rules is able to prune inconsequential rules from the rules base, the results of which are reported in Section 4-2. The next section will describe the actual REP forecast, which is the aim of this project.

3-3 Real-time Electricity market

The previous sections of this chapter described experiments known from fuzzy modelling literature in order to establish GFNARX performance in a known framework. These experiments are performed because there is no benchmark method available in literature for Real-time Electricity market Price (REP) forecasts, while the aim of this project is to achieve accurate and interpretable prediction of REP. In this section, the REP time series is investigated and pre-processed for model identification.

The remainder of this section is structured as follows. Section 3-3-1 briefly introduces the way the REP is formed, before Section 3-3-2 explains the relationship between REP and Settlement Price (SP). Pre-processing of the REP time series to a form suitable for identification and prediction is topic of Section 3-3-3. Section 3-3-4 investigates available data to select appropriate model inputs. Finally, the REP prediction experiment is described in Section 3-3-5.

3-3-1 Formation of the Real-time Electricity Market Price

Before delving into details of the REP time series and its pre-processing for model identification purposes, the way REP is formed is introduced in this section. This will provide fundamental insight into the REP signal, which is required to understand design choices made with regards to the model. As the REP time series is published on the website of the Transmission System Operator (TSO) with the aim of transparency regarding the potential settlement prices, this section will contain some allusion to Section 3-3-2.

To balance the electricity grid, the TSO operates a real-time electricity market in which it is the only broker. There are three products which are put up for auction on this market:

- Primary Control Reserve (PCR)
- Secondary Control Reserve (SCR)
- Tertiary Control Reserve (TCR)

In this thesis the focus will lie on the influence of these control products on the REP. For a full description, the reader is referred to [43].

PCR PCR is contracted by the TSO to counteract spontaneous grid disturbances. As PCR activation is an unpaid service, it does not influence the REP.

SCR SCR constitutes the largest part of the real-time electricity market. Balance Responsible Partys (BRPs) submit price bids (Eur/MWh) at which they are willing to supply or consume electricity. The minimum power which needs to be available in order to place a bid is $4MW$, but the TSO can choose to partially activate offered power in steps of $1MW$. The amount of MW activated is sent as setpoint signal to the BRP. Activation of bids is done by aggregating bids into a merit order dubbed as bidding ladder and sequentially activating up until the required power, thus minimising cost of activation for the TSO. Bid activation is updated every four seconds and the amount of MW activated per bid is sent out as setpoint to the BRPs.

TCR TCR is once more split into two categories, reserve power and emergency power. Emergency power is contracted in advance, so the price involved with activation of emergency power is known. Reserve power however can be put up for auction by means of bids. BRPs determine the price at which they are willing to supply or consume electricity as well as the corresponding volume of electricity. The TSO can opt to activate such a bid during a Program Time Unit (PTU).

The characteristics of PCR, SCR and TCR are summarized in Table 3-2. Following the activation of SCR and TCR bids in two possible directions, REP comprises two components, REP_c and REP_s, described by Equations 3-6 and 3-7.

Table 3-2: Characteristics of control products available on the real-time electricity market. Upwards regulation refers to the TSO buying electricity from BRPs, while downwards regulation implies the TSO sells energy to BRPs

Product	PCR	SCR	TCR
Price determination	-	Most extreme bid activated	Most extreme bid activated
Way of activation	Automatic, real-time response	Activation updated every 4s	Manual activation for next PTU
Response to activation	linear up to 100% within 30s	$\geq 7\%$/min ramp rate towards setpoint	$\leq 15min$
Direction	upwards and downwards	upwards and downwards (simultaneous possible)	reserve: upwards and downwards emergency: upwards only

$$REP_c = max([B_{up,SCR}, B_{up,TCR}]) \tag{3-6}$$

$$REP_s = min([B_{down,SCR}, B_{down,TCR}]) \tag{3-7}$$

with

REP_c Real-time electricity price for overconsumption. This occurs when the TSO needs to buy energy from BRPs

REP_s Real-time electricity price for oversupply. This occurs when the TSO needs to sell energy to BRPs

B_{up} Activated bids in the upwards direction. Upwards regulation refers to the TSO buying electricity

B_{down} Activated bids in the downwards direction. Downwards regulation refers to the TSO selling electricity

In this section, the formation of the REP has been described. The REP is a result of activated control energy bids on the real-time electricity market, with the most extreme bids activated determining the price. As the actual price BRPs have to pay or receive is determined by the SP rather than REP, the relation between these prices is described in the next section.

3-3-2 Imbalance settlement

In the previous section, the formation of REP has been described. This Section describes the way REP influences the Settlement Price (SP). In the Netherlands, following the terminology of [18], a dual pricing system is used in conjunction with an additive component. The latter is rarely employed and negligible in comparison to the imbalance price at its extremes, so will be omitted from the pricing scheme analysed in this study. The term dual pricing refers to there being separate SPs for upwards regulation and downwards regulation.

The imbalance price to be paid by a BRP when overconsuming, ISP_c, and the one paid to a BRP when oversupplying, ISP_s, are given in Table 3-3. The concept of the mid price, P_m, as a function of the bids submitted by the BRPs is given in Equation 3-8 with the subscripts once again denoting extra consumption or supply. The framework for determining the regulatory state S is captured by the following rules:

Table 3-3: The oversupply (ISP_s) and overconsumption (ISP_c) settlement prices as determined by the mid price P_m, regulation state S, the REP and the emergency power price P_e. The parameter $\alpha = 1$ if emergency power has been dispatched and $\alpha = 0$ otherwise, while the subscript i denotes individual values within the Program Time Unit

S	ISP_s	ISP_c
-1	$\min_i(REP_{s,i})$	$\min_i(REP_{s,i})$
0	P_m	P_m
1	$max(\max_i(REP_{c,i}), \alpha P_e)$	$max(\max_i(REP_{c,i}), \alpha P_e)$
2	$min(\min_i(REP_{s,i}), P_m)$	$max(\max_i(REP_{c,i}), \alpha P_e, P_m)$

- If no bids have been activated, $S = 0$
- If only downwards regulating bids have been activated, $S = -1$
- If only upwards regulating bids have been activated, $S = 1$
- If bids have been activated both ways, the balance delta $\delta_i = SCR_{c,i} - SCR_{s,i}$ with SCR_c denoting the extra upwards regulation power and SCR_s the extra downwards regulation power due to activated SCR bids, determines the actual regulation state:
  1. If it is continuously decreasing, $\delta_i - \delta_{i-1} \leq 0 \; \forall i \; \epsilon \, PTU$, $S = -1$
  2. If it is continuously increasing, $\delta_i - \delta_{i-1} \geq 0 \; \forall i \; \epsilon \, PTU$, $S = 1$
  3. $S = 2$ otherwise

$$P_m = \frac{max(B_{up}) + min(B_{down})}{2} \tag{3-8}$$

International Grid Control Cooperation The International Grid Control Cooperation (IGCC) is a collaboration between a number of TSOs in Central West European countries which is in effect since 2012[44]. Its basic product is a netting of imbalance, but the dutch TSO has some more cooperation schemes with TenneT DE and Elia BE, the specifics of which can be found in [43].

While the other potential points of cooperation hinge on reduction of required overall control capacity, the netting of imbalance leads to immediate reduction in activated SCR and thus to reduction in costs. Actual netting of imbalance can happen cross-border between any two adjacent control areas and the basic idea, as can be found in [44], is shown in Equations 3-9 and 3-10. The settlement price P_{IGCC} is multiplied by the net import volume, defined as the difference between the sum of imports I and sum of exports E to arrive at a cost J for participating TSO x. To achieve a fair distribution of the profits of the imbalance netting, the settlement price is determined as a volume weighted average of the imbalance settlement prices in each control area.

$$P_{IGCC} = \frac{\sum_x (I_x ISP_{s,x} + E_x ISP_{c,x})}{\sum_x (I_x + E_x)} \tag{3-9}$$

$$J_x = (I_x - E_x) P_{IGCC} \tag{3-10}$$

Since imbalance areas with a surplus receive more compensation for the energy they provide to the IGCC than selling through activating an additional BRP bid, and vice versa pay less for acquiring extra supply, this imbalance netting exchange is in general very profitable to the TSO. Information about the IGCC netting is published simultaneously with the REP time series and is an indication of potentially required SCR. If at any point the IGCC imbalance netting becomes unavailable, the power which was first netted cross-border needs to be acquired through activating SCR bids, resulting in more extreme REP.

In this section, the relation between REP and SP has been described. SP is the price at which electricity program deviations are settled, but it has been established that SP is a result of REP and SCR trend, which have a finer time resolution. This finer time resolution allows for improved SP predictions throughout a PTU and hence for better asset regulation. The REP is thus the time series to forecast and its pre-processing to yield a series suitable for black-box identification is described in the following section.

3-3-3 Time series representation of the Real-time Electricity Market Price

In the previous section, the relation between REP and SP has been described. It has been seen that the Dutch balancing market employs a dual pricing system in which SPs for subtracting electricity from the grid can differ from those for supplying to the grid if the regulation state $S = 2$. When forecasting REP it is thus necessary to forecast both elements of REP, REP_c and REP_s, separately in order to obtain an implicit intermediate forecast for SP which governs the actual financial transactions between the TSO and BRPs. The SP forecasts are not explicit, as this would constitute a many-to-one mapping, with infinite possibilities for intra-PTU profiles resulting in the same SP.

The resulting model identification scheme is shown in Figure 3-2. The planned supply and demand profiles are subject to disturbances d resulting in grid imbalance e. A PI controller is used to determine the desired SCR activation[45]. This desired amount of SCR is then acquired in a balancing auction. If it is possible, the IGCC is used to net imbalances cross-border, after which the SCR bids from the bid ladder are activated in sequence to satisfy the SCR demand (see Section 3-3-1). Information about REP is published once every minute on the website of the TSO[9]. A PTU comprises 15 minutes, so 15 published REP values determine the settlement price of a PTU. Yet this signal is not observable in real-time. The REP time series is published at a three minute delay. This three minute delayed signal is used as an endogenous input to the autoregressive model structures described in Chapter 2. The REP forecasts are fed through the imbalance settlement regulation mask in order to detect regulation state $S = 2$.

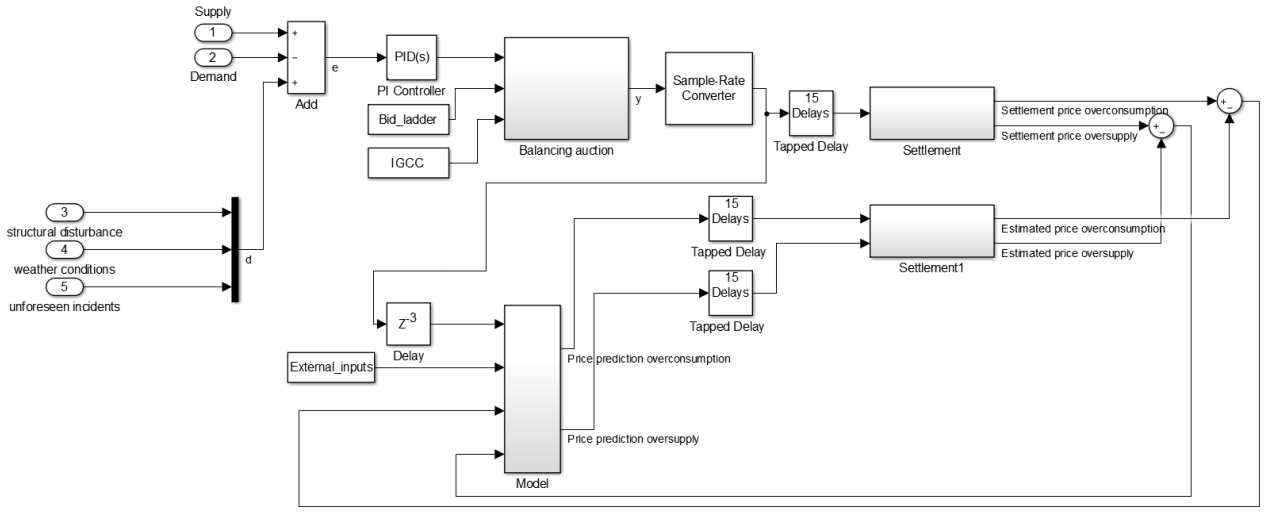


Figure 3-2: Schematic overview of the model identification problem

Pre-processing REP The endogenous signal published by the TSO consists of REP_c and REP_s. However, the TSO does not publish a value for both REP_c and REP_s at every time step. If no bid B_{up} is activated, there is no published value for REP_c and similarly if no bid B_{down} is activated, no value for REP_s is published. As all time series models described in Chapter 2 assume uniformly spaced data, the gaps in REP_c and REP_s data have to be filled. An intuïve way to fill the gaps can directly be inferred from Table 3-3: if there is no regulation during a PTU and hence no published values for either REP_c or REP_s, the settlement price $ISP_s = ISP_c = P_m$. Filling the gaps in the time series data with $REP_{gap} = P_m$ is thus representative of the actual REP state.

A large influence to the REP time series is the mid price P_m. The mid price is is a direct average of the first SCR bids to be activated on each side of the bid ladder, as described by Equation 3-8, and is thus a consequence of subjective BRP behaviour. The fact that the mid price is not fixed can be seen in Figure 3-3. This means that the REP time series exhibit a deterministic trend. The mid price is fixed throughout a PTU and as bids cannot be altered in the hour preceding the PTU of delivery[43], is known prior to the start of a PTU. It can then be subtracted from the REP time series to yield Real-time Electricity market Price Premium (REPP):

$$REPP_c = REP_c - P_m \tag{3-11}$$

$$REPP_s = REP_s - P_m \tag{3-12}$$

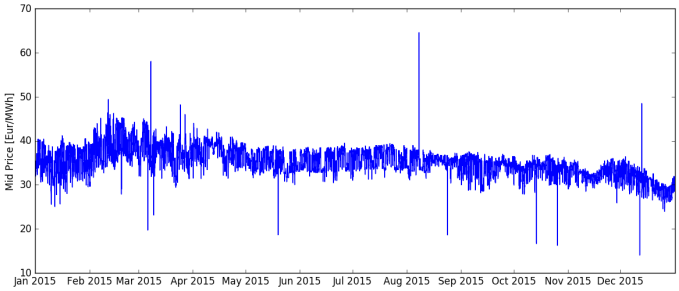


Figure 3-3: Mid price P_m throughout 2015. It is clear that P_m is not constant

The premiums described by Equations 3-11 and 3-12 are representative of the potential monetary value or risk for a BRP being in a state of imbalance. As the REP_c and REP_s series are forecasted separately, it is possible to specify $REPP_s$ in a way which represents the possible monetary gain if a BRP actively causes imbalance in its portfolio to counteract the overall grid imbalance. This is described by Equation 3-13.

$$REPP_s^* = P_m - REP_s \tag{3-13}$$

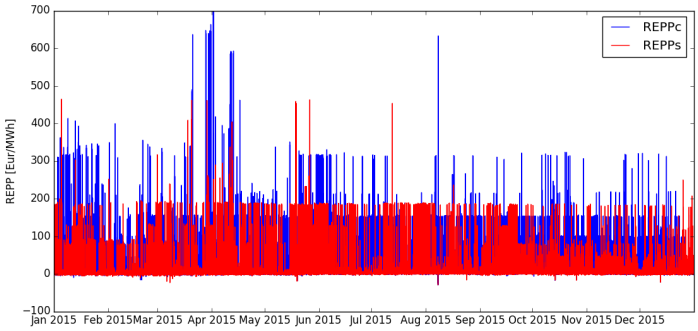


Figure 3-4: REPP throughout the year 2015. The series exhibit volatility clustering and price spikes, with $REPP_c$ spikes being more extreme

If one looks at the profiles of REPP over the year 2015, shown in Figure 3-4, it is visible that the overall REPP behaviour is mean-reverting, but there is a surplus of spikes. There is also a clear interval within which most of the REPP values reside, the interval $[0, 200]$. To avoid model overfitting on outliers, the REPP series is clipped. Some statistics about the REPP series are summarized in Table 3-4. Since only 1% of $REPP_c$ values are exceeding $200 Eur/MWh$, which is approximately the boundary between normal price ranges and price spikes from a visual inspection of Figure 3-4, $REPP_{c,max} = 200 Eur/MWh$ is chosen to clip the $REPP_c$ time series. As the threshold for clipping is based on visual inspection rather than calculation, optimisation might be possible for specific applications of REP forecasts, but this is not within the scope of this thesis.

Table 3-4: Statistics of REPP in 2015. The means are calculated over the values where a REP signal is published

	REPPc	REPPs
Percentage > 200 Eur/MWh	1%	0.1%
Percentage < 0 Eur/MWh	4%	4%
Mean	14.84 Eur/MWh	28.29 Eur/MWh
Mean (interval [0, 200])	14.68 Eur/MWh	26.17 Eur/MWh

The clipping threshold for $REPP_s$ is chosen at $REPP_{s,max} = 100Eur/MWh$ for the following reasons:

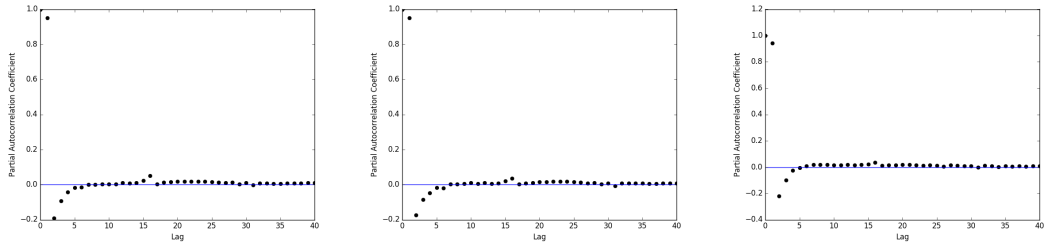
- The mean of $REPP_c$ is approximately twice as high as the mean of $REPP_s$ (see Table 3-4). This is not solely due to the difference in price spike intensity, as clipping both $REPP_c$ and $REPP_c$ at the interval $[0, 200]$, on average $\bar{REPP}_c \approx 1.8\bar{REPP}_s$
- From 31-03-2015 through 17-12-2015, the government provided state aid to renewable electricity sources through the Stimulering Duurzame Energieproductie (SDE). On average throughout 2015, this constituted $\bar{SDE} = 56.39Eur/MWh$[46]. The largest amount of state aid in terms of Eur/MWh was provided to wind turbines, with the turbines joining in from 31-03-2015 receiving a base amount of $SDE_{wind} = 70.00Eur/MWh$[47]. Adding the average mid price $\bar{P_m} = 35.23Eur/MWh$, this indicates that at $REPP_s = 91.52Eur/MWh$ it is not attractive for the average renewable energy source to deliver electricity to the grid. Similarly, at $REPP_s = 105.23Eur/MWh$ it is unattractive to wind parks on land. The clipping threshold of $REPP_{s,max} = 100Eur/MWh$ lies in between these two values with less than 1% of the values where $REPP_s > 100Eur/MWh$ falling within the range $[100, 105]$

At the lower end of the REPP spectrum, the time series are clipped at $REPP_{c,min} = REPP_{s,min} = 0Eur/MWh$. This is because it is assumed that BRPs with a demand response portfolio will not act voluntarily in a way which nets monetary losses. Summarising the properties of the pre-processed REPP:

$REPP_c$ $REPP_{c,min} = 0$; $REPP_{c,max} = 200$

$REPP_s$ $REPP_{s,min} = 0$; $REPP_{s,max} = 100$

In this section a closer look has been taken at the REP time series. As the Dutch TSO employs a pricing scheme in which settlement prices for overconsumption and oversupply can differ, REP_c and REP_s need to be forecast separately. Moreover, to facilitate model identification with the model structures described in Chapter 2 the REP time series are pre-processed: REP is detrended by the subtracting the deterministic mid price P_m and clipped to avoid overfitting on outliers. In the remainder of this thesis, the suffix 'premium' will be dropped, so the pre-processed series REPP is implicitly meant whenever actual values are mentioned with regards to REP forecasts. The next section will investigate available data on exogenous inputs as well as select auto-regressive model orders.



(a) PACF of TS **(b)** PACF of REP_c **(c)** PACF of REP_s

Figure 3-5: PACFs of the time series TS, REP_c and REP_s. They resemble each other closely

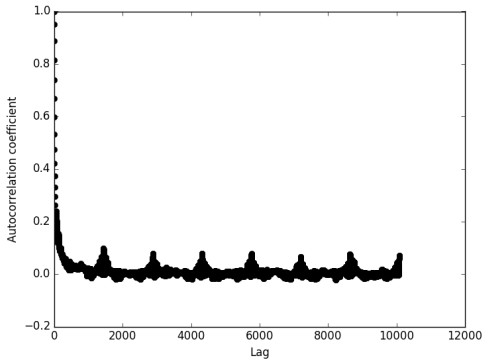
3-3-4 Model inputs

The previous section described how REP_c and REP_s have to be forecast separately and their pre-processing. The model structures introduced in Chapter 2 which will be used to forecast REP are not just auto-regressive in nature. They also allow for exogenous influences to be modelled. REP is a discrete result of the TSO activating bids in the real-time electricity market auction. Based on the properties of the SCR and TCR control products, described in Section 3-3-1, these bid activations typically last for a longer time than the three minute delay at which REP is published. In this section, potential model inputs are reviewed and a subset is selected to forecast the REP with. First, the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the REP time series are inspected to identify significant endogenous terms. Then, potential exogenous regressors are addressed.

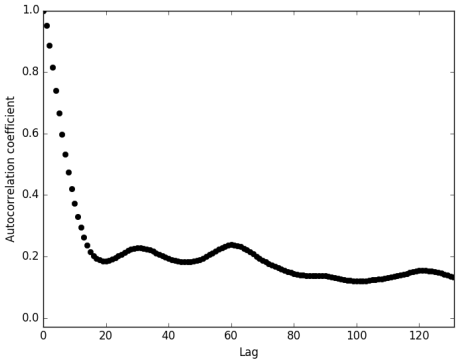
Endogenous inputs To select appropriate lagged REP values as inputs, the ACF and PACF of the REP time series are inspected. To avoid continuity issues due to data gaps in REP being filled with a constant value (and hence overestimation of autocorrelation between subsequent terms), the modified time series $TS = REP_c - REP_s$ is regarded. As can be seen in Figure 3-5, the PACF of TS resembles the PACF of REP_c and REP_s closely, so selection of Auto-Regressive (AR) terms is not influenced.

The ACF of this modified series is shown in Figure 3-6a. A zoomed version of this ACF is depicted in Figure 3-6b. Daily, hourly and half-hourly correlation peaks can definitely be established. The daily correlation peaks are the highest of these three possible seasonalities, but are low in comparison to shorter lags. While there is thus some evidence towards a daily seasonality, the significance is low. With some allusion to later sections: including the daily seasonality in the GFNARX model yielded lower accuracy on the training data set, after which this particular endogenous input was discarded. Other seasonalities are even less significant and have thus not been considered.

As was argued in Section 2-1-3, a pure Moving Average (MA) process would have its ACF cut off after a finite amount of lags, so there are AR influences and the PACF has to be regarded. Figure 3-5a shows the PACF of the modified time series TS. The PACF cuts off after around 6 lags, with a small extra spike at the 15 minute mark. This extra spike can be attributed to the mean-reverting step which frequently occurs at PTU borders (see Section 3-3-1). As



(a) Autocorrelation coefficients of the TS time series for a week in lags. There are periodic spikes at multiples of 1440 minutes, which constitute a day



(b) Autocorrelation of the TS time series for approximately 2 hours in lags. There are mild spikes at 30 minutes lag and 60 minutes lag

the time series is published by TenneT at a three minute delay, lags 3, 4, and 5 are most significant towards predicting the REP.

External inputs One property of all the time series models considered in this thesis, is the optional inclusion of exogenous inputs. In this section, publicly available exogenous inputs which could influence the REP are analysed.

SCR and TCR As the REP is formed by the TSO dispatching regulatory power to balance the grid, the activated SCR and TCR have a large influence on the REP. The regulatory state which ultimately determines the SP, however, is not influenced by the TCR. As described in Section 3-3-2, the regulatory state is solely governed by the activation of SCR throughout a PTU.

If the emergency TCR is activated in a PTU, but the SCR activated in the corresponding direction forms a monotone non-increasing series, while the activated SCR in the opposite direction has at least one increment in the duration of the PTU, the SP is not influenced by the activation of TCR at all. An example is found in PTU 61 of 03-03-2015 (15.00-15.15)[9], of which the relevant quantities as published by the TSO are given in Table 3-5. The emergency TCR is active and an additional $30MW$ of TCR is activated at the start of the PTU. However, the activated SCR is a monotonically non-increasing series throughout the PTU. Hence, the regulatory state $S = -1$ and $ISP_c = ISC_s = -35.60Eur/MWh$.

This means that including the TCR does not offer insight into the SP and hence has a deceptive influence on the REP. Moreover, as TCR is activated in blocks of full PTUs, its influence on the REP will be reflected in the published REP time series itself and there is no need to include TCR as explanatory variable.

The regulatory state is determined by the trajectory of SCR during a PTU. The published

Table 3-5: TenneT data for PTU 61 of 03-03-2015[9] for upwards(+) and downwards(-) regulation. The regulatory state $S = -1$ due to the non-increasing activation of SCR in the upwards regulation direction

Time	$IGCC_+$ [MW]	$IGCC_-$ [MW]	SCR_+ [MW]	SCR_- [MW]	TCR_+ [MW]	REP_c [Eur/MWh]	REP_s [Eur/MWh]	P_m [Eur/MWh]
15.00	0	0	229	0	30	366.40		41.96
15.01	0	97	213	0	30	366.40		41.96
15.02	0	132	180	0	30	366.40		41.96
15.03	0	194	166	0	30	366.40		41.96
15.04	0	228	149	0	30	366.40		41.96
15.05	0	50	119	0	30	366.40		41.96
15.06	0	134	104	1	30	366.40	40.39	41.96
15.07	0	119	88	6	30	366.40	38.39	41.96
15.08	0	132	56	14	30	366.40	36.39	41.96
15.09	0	96	45	25	30	366.40	27.39	41.96
15.10	0	55	32	43	30	366.40	27.39	41.96
15.11	0	95	3	64	30	366.40	-35.60	41.96
15.12	0	86	0	88	30	366.40	-35.60	41.96
15.13	0	105	0	125	30	366.40	-35.60	41.96
15.14	0	0	0	156	30	366.40	-35.60	41.96

values reflect the currently active amount of SCR rather than the total amount activated by the TSO, while the REP time series represent the most extreme bid activated, as described in Section 3-3-1. As such, the published SCR time series can show a ramp-up of SCR providers towards their setpoint provided by the TSO. It is also possible for bids lower than the price-setting bid to be deactivated while leaving the REP unaltered. In that case, the SCR shows a ramp down and it can be expected that the REP will soon return towards the mid price P_m as well. As such, the first difference of the SCR time series has been included as explanatory variable.

Bid price ladder REP is determined by the most extreme bid activated from the bid ladder. Therefore, it would be possible to create a deterministic mapping from SCR activation to REP if the bid ladder was completely observable. However, the only values published are:

- $min(B_{up,SCR})$
- $max(B_{down,SCR})$
- B_{up} when $SCR_+ = 100MW$
- B_{down} when $SCR_- = 100MW$
- B_{up} when $SCR_+ = 300MW$
- B_{down} when $SCR_- = 300MW$
- B_{up} when $SCR_+ = 600MW$
- B_{up} when $SCR_- = 600MW$
- $max(B_{up,SCR})$
- $min(B_{down,SCR})$

Bids are submitted per individual PTU and can vary in size as well as price. It is thus not trivial to reconstruct the entire bid ladder over time to obtain a mapping from SCR to REP. As most of the bid ladder is unobservable, it will not be used as explanatory variable in the REP prediction models.

IGCC The IGCC, which has been described in Section 3-3-1, was introduced in 2012. When regarding the statistics of regulatory state occurrence, a discrete step in occurrence of regulatory state $S = 0$, which corresponds to no SCR activated, is noticed in 2012, as shown in Table 3-6. This coincides with the introduction of the IGCC. Due to the netting of imbalance between countries, sometimes activation of regulatory reserves can be avoided as a whole, which results in regulatory state $S = 0$ even though there is local imbalance.

For the IGCC imbalance netting to have an effect, local imbalances between two adjacent countries have to have opposite signs. Price volatility will be a lot higher if such a netting can not take place. The IGCC contribution towards imbalance regulation is published on the website of the TSO and the availability of IGCC is clearly an important explanatory variable.

Table 3-6: Yearly relative occurrence of regulation states per PTU. Regulation states $S = -1$, $S = 1$ refer to the grid being in surplus or deficit; regulation state $S = 0$ means there was no regulation necessary and state $S = 2$ refers to both upwards and downwards regulation being required within a single PTU

year\S	-1	0	1	2
2008	43.94%	0.03%	38.23%	17.79%
2009	51.77%	0.14%	38.25%	9.85%
2010	50.11%	0.23%	40.85%	8.81%
2011	47.33%	0.23%	39.53%	12.91%
2012	45.04%	6.99%	36.44%	11.53%
2013	41.97%	3.39%	42.11%	12.53%
2014	46.04%	5.29%	41.80%	6.88%
2015	43.87%	5.12%	37.27%	13.74%

Weather conditions As renewable energy sources cannot reliably plan their power generation a day ahead due to dependence on weather conditions, wind farms and solar plants tend to frequently cause grid imbalance. To counteract the imbalance arising as a result, SCR has to be activated. It is then plausible to assume a correlation between weather conditions and the REP.

Available historical data on weather conditions provided by the Koninklijk Nederlands Meteorologisch Instituut (KNMI) has a time resolution of an hour. This is a low sample rate compared to the REP time series. As the REP is highly volatile and the grid balance state frequently changes sign multiple times within an hour, the correlation between the weather data provided in hourly intervals and REP published in one minute intervals is thought to not be significant due to undersampling of the weather data. Table 3-7 shows the correlation coefficients between weather variables and the REP time series. There is no strong

correlation, with only the correlation between REP_s and average wind speed during an hour $Corr(REP_s, \bar{V}_{wind}) > 0.05$.

Table 3-7: Cross-correlation coefficients between REP and weather conditions. The weather conditions investigated are the temperature T and average wind speed $\bar{V}_{wind}$

	REP_s	REP_c
Temperature	0.009	0.035
Average wind speed	0.054	0.010

To further investigate this correlation, inclusion of this wind speed in prediction of REP_s is tested within the framework of the prediction experiment described in Section 3-3-5. The result is that on the training data set, the RMSE of GFNARX prediction is increased from $RMSE(REP_s)_{without,training} = 9.78 Eur/MWh$ to $RMSE(REP_s)_{wind,training} = 10.72 Eur/MWh$. As including the statistically most significantly correlated weather variable, wind speed, as a model input results in a lower model accuracy, weather conditions are not used as exogenous inputs. In future research, the influence of weather conditions sampled at a higher frequency can still be investigated, but acquisition of such data by measurement is not within the scope of this thesis.

Grid frequency The TSO is tasked with balancing the grid through matching electricity consumption and supply. If these are not in balance, the grid frequency deviates from the nominal frequency $f_{nom} = 50Hz$. As control energy is dispatched by the TSO to maintain the grid frequency at approximately f_{nom}, it is investigated whether the grid frequency is correlated to REP.

As all of continental Europe is connected to the same electricity grid, the grid frequency is synchronous throughout Europe. The Mains frequency[48] is measured at an interval of 10 seconds, which is a fine enough time resolution to investigate its influence on REP. Contrary to initial expectation however, there is little to no correlation between REP and the Mains frequency deviation from f_{nom}. Figure 3-7 shows the frequency deviation and REP time series for a couple of hours in 2015. The lack of visible correlation between frequency deviation and REP trend is observed throughout the rest of the year 2015 as well. The cross-correlation coefficients for the first 60 lags, corresponding to an hour in lags, are shown in Figure 3-8. There is no evidence that the series are significantly correlated.

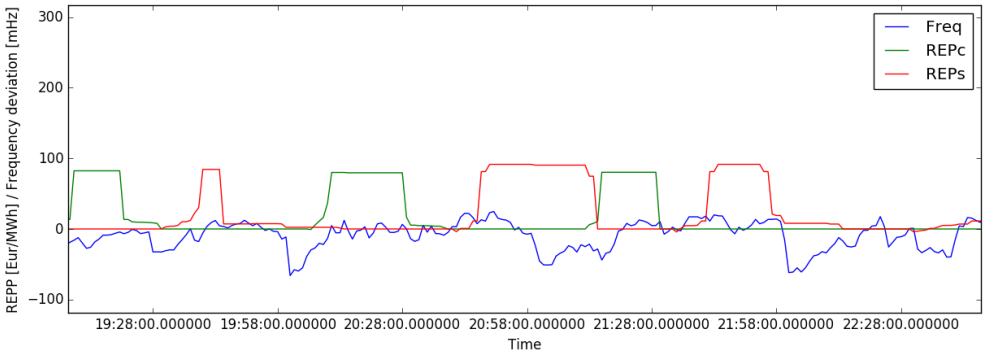


Figure 3-7: Frequency deviation from the nominal frequency $f_{nom} = 50Hz$ and REP for a couple of hours in 2015. There seems to be no correlation

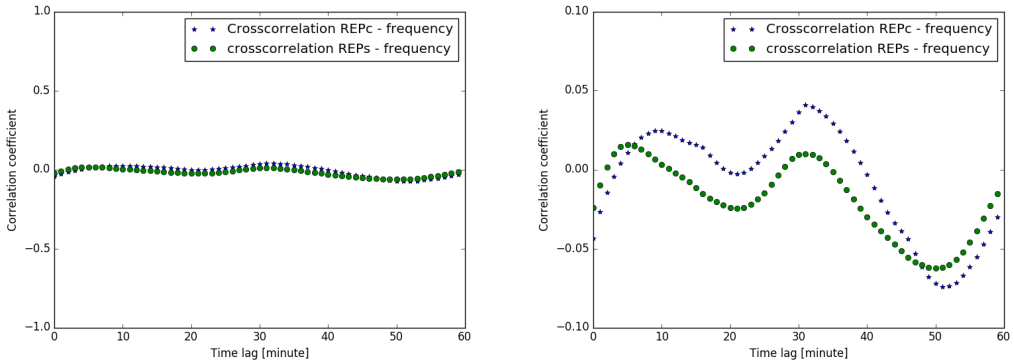


Figure 3-8: Cross-correlation of frequency deviation from $f_{nom} = 50Hz$ and REP for an hour in lags. There is no evidence that the quantities are correlated

A plausible explanation for this lack of correlation between grid frequency and REP is the fact that REP is governed by SCR activation. The PCR is used to stabilise the grid frequency after a frequency deviation occurs anywhere in Europe, while the TSO employs SCR only to correct for any local imbalance as measured by cross-border electricity flows. As the IGCC also deals with imbalance netting between adjacent TSO control areas, it might be possible to find a correlation between frequency deviation and availability of IGCC, but investigating whether this is possible is not within the scope of this thesis. As no significant evidence has been found of a correlation between grid frequency deviation and REP, the grid frequency will not be used as an exogenous input to REP forecasting models.

Table 3-8: Signals published by the TSO and KNMI which are related to the REP and their inclusion in or exclusion from the prediction model

Signal	Time resolution	Included	Reason for exclusion
Activated SCR	Minute	Yes	-
Activated TCR	Minute	No	No influence on the regulatory state
IGCC	Minute	Yes	-
Bid price ladder characteristics	PTU	No	Unobservable
Temperature	Hour	No	Correlation too low
Wind speed	Hour	No	Correlation too low
Grid frequency	10 Seconds	No	Correlation too low

In this section available data sources have been investigated in order to select inputs to the REP forecasting models described in Chapter 2. As endogenous inputs the three, four and five minutes delayed REP time series values will be used. This corresponds to the three most recent observations published by the TSO. The findings on exogenous inputs which will be used in REP forecasts are summarized in Table 3-8. Exogenous inputs which will be used are the most recently observed SCR values, trend in SCR, availability of the option to net imbalances through the IGCC and amount of imbalance netting taking place. The next section describes the REP model identification and forecasting experiment which will be used to assess model accuracy.

3-3-5 Prediction of the Real-time electricity price

In this section the experiment regarding REP model identification and subsequent forecast is described. The primary aim of forecasting REP in this thesis work is to reduce overall cost of operation of electrical assets in a demand response portfolio, for which forecast accuracy is not necessarily the most descriptive metric. To extend the applicability of identified REP models to general use cases and acquire insight into the real-time electricity market system dynamics, however, it is required to forecast REP as accurately as possible.

As REP data is published by the TSO with a frequency of once per minute, acquiring enough data to perform an identification experiment by real-time observation is infeasible. Therefore, historical data which can be downloaded from the website of the TSO[9] is used. As 2015 is at the moment of writing the most recent complete year of which historical data is available, REP data of the year 2015 is used to identify prediction models and evaluate their performance.

In system identification it is normal to split the data into training and validation sets, for instance using 2/3 of the data samples to identify a model and the remaining data to evaluate performance. In this experiment, data is split in a more skewed fashion: the first two weeks of 2015 are used to identify a model, while the remainder of 2015 is used for validation purposes. This choice for training on two weeks of data was made because of the rule of thumb in identification experiment design that the length of the identification experiment should be approximately ten times the longest time constant of interest in the system[49]. As was seen in Section 3-3-4, the longest time constant of interest in the system is a day. Two weeks constitute fourteen times this time constant, which satisfies the rule of thumb.

This project is carried out at Peeeks B.V. which mainly uses Python as software platform to operate its systems. To be compatible with the existing software framework of the company, the models described in Chapter 2 are implemented in Python 3.4.

SARIMAX-GARCH, NARX and GFNARX all use least squares optimisation methods to estimate their parameters and as such, the Root-Mean-Square Error (RMSE) metric, described by Equation 3-14, is a fair method to compare model performance. As reference prediction model the naïve model $\hat{y}(t) = y(t-3)$ is used.

$$RMSE = \sqrt{\frac{\sum_{t=1}^{T} ||y(t) - \hat{y}(t)||}{T}} \tag{3-14}$$

In this section, the Real-time Electricity market Price (REP) time series has been analysed. From the relation between REP and Settlement Price (SP), it has been inferred that it is necessary to forecast REP_c and REP_s separately. The deterministic trend in REP stemming from the mid price P_m has been subtracted from the series and they have been clipped to avoid model overfitting on outliers. Based on REP autocorrelation and available data on exogenous factors influencing REP, model inputs have been selected. As endogenous inputs the three, four and five minutes delayed REP time series values will be used. This corresponds to the three most recent observations published by the TSO. Exogenous inputs which will be used in both REP_c and REP_s forecasts are the most recently observed SCR values, trend in SCR, availability of the option to net imbalances through the IGCC and amount of imbalance netting taking place. The models described in Chapter 2 are trained on historical data of the first two weeks of 2015 and subsequently used to predict REP for the entire year 2015. Results of this REP prediction experiment are reported in Section 4-2-2. In this thesis work, the main aim of REP prediction is to reduce cost of operation of electrical assets within a demand response portfolio. The next section proposes a benchmark method based on operation of a typical asset within such a portfolio.

3-4 Thermal control of a Cold Storage Warehouse

In this section, a benchmark method is proposed which simulates application of simple control strategies to a modelled CSW. The CSW is a prime example of an asset with a thermal buffer, which can be used in a demand response portfolio. In order to produce sensible results, a heat loss model based on first principles is included. This section is structured as follows: first, the heat loss model is described, after which flaws in dimensioning of CSW models found in literature are corrected based on load data from an actual CSW. Then, the cost of CSW operation is described and ultimately, the resulting benchmark is presented in which it is the objective to minimise cost of CSW operation based on historical data of the year 2015.

3-4-1 First principles modelling of a cold storage warehouse

A CSW typically stores frozen products at an approximate maximum temperature of -18°C. This temperature ensures product quality does not degrade while stored, so it is imperative

to keep the temperature below this threshold when dealing with long term storage of frozen goods.

Because the temperature in a CSW is lower than the environmental temperature, there will be heat losses to the environment. Heat losses through all outside surface areas are governed by conduction, but the walls connected to outside air also suffer from convective losses. Moreover, there is heat production whenever there is activity inside, because fork-lift trucks and people produce heat. Based on these considerations, the thermal energy balance of a CSW is given by Equation 3-15. It is assumed that the ground has the same temperature as the outside air.

$$\frac{dE_T}{dt} = \sum_i U_i A_i (T_{out} - T) - V_C \left(\delta \eta P_C + P_{prod}\right) \tag{3-15}$$

with

E_T Thermal energy: $E_T = C_C V_C T$

C_C Heat capacity of the CSW $[J/m^3K]$

V_C Volume of the CSW $[m^3]$

T Temperature $[K]$

t Time $[s]$

U_i Heat transfer coefficient through surface i $[W/m^2K]$

A_i Area of surface i $[m^2]$

T_{out} Environment temperature $[K]$

δ Dummy variable for the activation of cooling power ($\delta = 1$ if cooling, $\delta = 0$ otherwise)

η Coefficient of Performance (COP) of the cooling power. $COP = |Q_C|/P_C t$, with as theoretical limit $COP_{max} = \dfrac{T}{T_{out} - T}$, based on the first law of thermodynamics

Q_C Actual heat subtracted from the CSW $[J/m^3]$

P_C Rated power of the cooling installation $[J/m^3 s]$

P_{prod} Heat production by activity inside the CSW $[J/m^3 s]$

Discretizing to obtain a time series model, Equation 3-15 becomes:

$$T(k+1) = T(k) + \frac{t_s}{C_C}\left(\sum_i \left(\frac{U_i A_i}{V_C}(T_{out}(k) - T(k))\right) - \delta \eta P_C + P_{prod}\right) \tag{3-16}$$

with t_s the time resolution of the series.

3-4-2 Realistic dimensioning of model parameters

Throughout literature, various models of CSWs have been used. As representatives, the models used in an MSc internship previously performed at Peeeks [13] will be considered. The parameters of these CSW models are given in Table 3-9.

Table 3-9: CSW parameters for three models [10, 11, 12] completed with parameters determined in [13]

	Lukasse et al, 2009 [10]	Verzijlbergh and Lukszo, 2013 [11]	Stoeckle, 2001 [12]
Size (l x w x h) [m]	20 x 25 x 5	200 x 190 x 30	95.94 x 95.94 x 9.6
Volume V_C $[m^3]$	2500	$1.14 \cdot 10^6$	$8.84 \cdot 10^4$
Heat capacity C_C $[J/m^3K]$	$5 \cdot 10^5$	7895	$5.57 \cdot 10^5$
Heat transfer coefficient U $[J/m^2Ks]$	$U_f = 0.21$	0.2	0.3
(floor, walls, roof)	$U_w = 0.16$	0.2	0.26
	$U_r = 0.13$	0.2	0.26
Cooling power P_C $[J/m^3s]$	21.7	1.75	13.6
Rated cooling power P^* $[kW]$	54	1995	1202
COP η	1.5	3	2.6
Heat production P_{prod} $[J/m^3s]$	20	0	3.6

In the experiment performed in [13], the nominal behaviour of these CSWs is to have δ take values corresponding to:

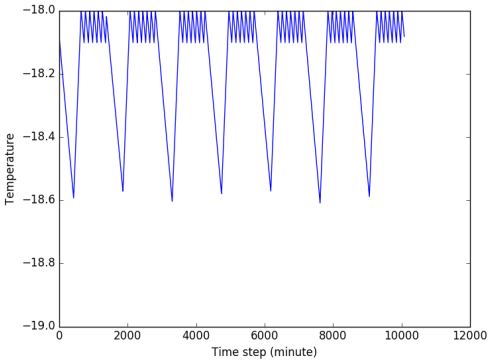
daytime (07.00 - 23.00) δ is regulated to keep the temperature $-18.1°C <= T < -18.0°C$. If $T < -18.1°C$, $\delta = 0$

nighttime (23.00 - 07.00) δ is regulated to keep the temperature between $-21.9°C >= T > -22°C$. If $T > -21.9°C$, $\delta = 1$

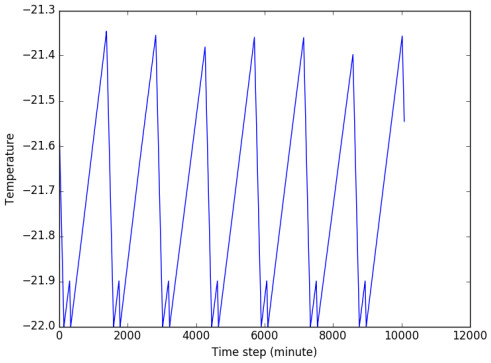
This leads to temperature profiles as seen in Figure 3-9. It can be seen that the CSW obtained from [10] operates within a really small range of the maximum temperature allowed. This CSW thus has very little flex available. On top of that, it is small, with a total rated power of $P^* = 54kW$. On the other hand, the CSW modelled by [11] is larger than any realistic scenario and does not take into account heat production within the storage itself. Moreover, the cooling is too efficient, causing the cooling power to be active slightly above ten percent of the time, which is unrealistic. The CSW which is described by [12] is the opposite of the first one, it is a large CSW which operates within a range around its minimum temperature, which again leaves very little room for flexible operation.

To arrive at a CSW model which has realistic operational parameters and has enough flexibility in the thermal buffer to serve as benchmark, a real world CSW, owned by A2 coldstorage, is regarded. This A2 CSW has the following specifications:

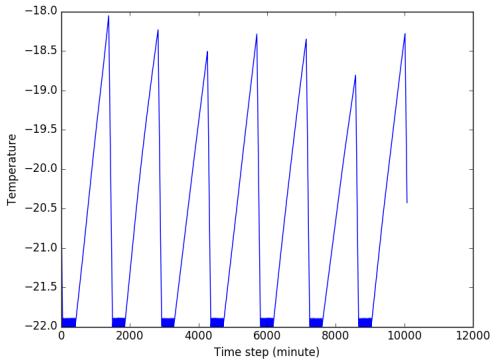
Volume V_C $[m^3$ $]$ $1.3 \cdot 10^5$



(a) Nominal operation of the CSW described by [10] from 08-01-2015 through 14-01-2015



(b) Nominal operation of the CSW described by [12] from 08-01-2015 through 14-01-2015



(c) Nominal operation of the CSW described by [11] from 08-01-2015 through 14-01-2015

Figure 3-9: Temperature profiles of the CSW models under nominal operation

Rated cooling power P^* $[kW\ ]$ 440

fraction of time actively cooling (2014) $\mathcal{E}(\delta)$ 0.65

The other operational parameters, such as heat loss and production, are not known.

Starting from the most realistic CSW described in Table 3-9, the cooling power P_C is scaled to have $P^* = 1MW$. This is for ease of interpretability, since the REP is settled in terms of Eur/MWh. The other parameters are also slightly adjusted in a manner which ensures that the fraction of time actively cooling $\mathcal{E}(\delta) \approx 0.6$, which is of the correct order of magnitude. The final CSW model parameters are given in Table 3-10.

Table 3-10: Operational parameters of the benchmark CSW model

Size (l x w x h) [m]	95.94 x 95.94 x 9.6
Volume V_C $[m^3]$	$8.84 \cdot 10^4$
Heat capacity C_C $[J/m^3K]$	$2.57 \cdot 10^5$
Heat transfer coefficient U $[J/m^2Ks]$	$U_f = 0.4$
(floor, walls, roof)	$U_w = 0.4$
	$U_r = 0.4$
Cooling power P_C $[J/m^3s]$	11.32
Rated cooling power P^* $[kW]$	1000
COP η	1.9
Heat production P_{prod} $[J/m^3s]$	10
Maximum Temperature allowed T_{max} $[°C]$	-18.0
Minimum Temperature allowed T_{min} $[°C]$	-22.0

3-4-3 Cost of operation and Naïve control

To obtain a benchmark for the cost of operating the CSW as modelled in the previous paragraph, a baseline cost has to be established. In the reference case, perfect consumption forecast is assumed, which means that all electricity is bought at the APX day-ahead auction[50]. This way, no imbalance is generated. The CSW consumption forecast is based on the control regulations given in [13], which keep the temperature $-22.0°C \leq T \leq -18.0$, while trying to adhere to the control scheme $\delta = 1$ from 23.00-06.59, $\delta = 0$ otherwise, each day.

A reference control method which does utilise the imbalance market and the possibility of buying electricity during times of favourable SPs, is created by regulating the CSW without modelling the REP. The REP forecasts in this reference method are the most recently obtained delayed REP values as published by the TSO[9], and as such, this method is called naïve forecasting.

With the operational limits given in Table 3-10, the naïve forecasting CSW control method exhibits the following properties and behaviour:

- Average power consumption required: $\bar{L} = 0.6MW$

- Maximum load: $L_{max} = 1MW$
- Minimum load: $L_{min} = 0MW$
- Ramp speed: $\frac{\Delta L}{\Delta t} = \alpha MW/min$; in our modelled CSW, $\alpha = \infty$
- If the REP for overconsumption, $REP_c \geq \Delta P_{threshold}$, the asset will ramp up towards maximum load. $\delta = 1$
- If the REP for oversupply, $REP_s- \geq \Delta P_{Threshold}$, the asset will ramp down towards minimum load. $\delta = 0$
- If $\Delta P_{threshold} \geq max(REP_c, REP_s)$, the asset will ramp up or down corresponding to its current temperature, which is a function of its internal imbalance $\Delta E = \int_0^T L(t)dt - \bar{L}T$. If $T > -18.5°C$, $\delta = 1$; if $T < -19.5°C$, $\delta = 0$
- If $T <= -22.0$, $\delta = 0$
- If $T >= -18.0$, $\delta = 1$
- If none of the above apply, follow the baseline behaviour

Generalising cost of operation to include REP forecasts, the overall block scheme of determining the costs of operating the simulated CSW is shown in Figure 3-10. Deviations from baseline operation, the electricity of which is bought at the day-ahead Amsterdam Power Exchange (APX) auction, are settled at the relevant SP (see Section 3-3-2) and the total cost is equal to the reference cost modified by the imbalance transactions.

3-4-4 Benchmarking cost of thermal control of a cold store

If no overriding control signal is given, the modelled CSW is run according to the principles described in [13]:

If the current time is between 07.00 and 22.59:

- If $T < -18.1$, $\delta = 1$
- If $T >= -18.0$, $\delta = 0$

If the current time is between 23.00 and 06.59:

- If $T > -21.9$, $\delta = 1$
- If $T <= -22.0$, $\delta = 0$

The energy required to operate the CSW in this way is bought on the APX day-ahead market and is assumed to be perfectly forecast. No imbalance is generated in the process. Operating the CSW according to this regulation scheme for the entire year 2015, leads to total electricity costs given in Table 3-11.

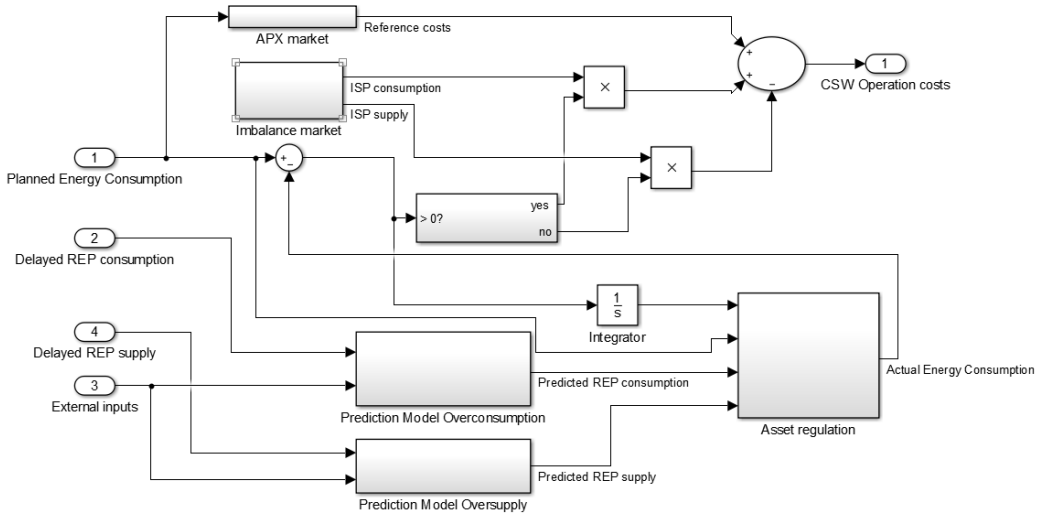


Figure 3-10: Block scheme of the overall cost of operation benchmark method. The Real-time electricity market prices (REP) are determined from their delayed values and external inputs. These predictions, along with the planned consumption and current internal consumption surplus or deficit are used to determine actual consumption levels. The deviations from the planned usage are settled at the relevant Imbalance Settlement Prices SP

Table 3-11: Cumulative electricity costs of operating the CSW in 2015 based on historical data. Nominal operation refers to buying all energy at the APX day-ahead auction with perfect forecasts. The three minute delay corresponds to the delay at which the TSO publishes REP data

	$Cost_{tot}$ $[Eur]$	$Cost_{avg}$ $[Eur/MWh]$
Nominal operation	195682	37.12
Naïve strategy (3 minutes delay)	162547	30.86
Peeeks strategy (3 minutes delay)	153076	29.03

When testing CSW regulation with REP forecasts, deviations from the forecasted usage in the reference case are settled at the appropriate SP. The performance measure is the total costs of acquiring electricity to run the CSW. To provide contrast, the electricity costs of CSW operation when utilizing the control strategy currently implemented at Peeeks is also shown in Table 3-11, as well as the naïve control implementation which is described in the previous paragraph. The results of applying REP predictions to regulate CSW cooling power are reported in Section 4-3-2.

This chapter introduced the experiments which are used in this thesis. To show the contribution of GFNARX to fuzzy modelling literature, a model identified with the GFNARX algorithm is applied to a literature benchmark. The benchmark method has been described in Section 4-1 and the results are presented in Section 3-1. GFNARX builds its network structure online, which can result in overparametrisation if there is no method to prune insignificant rules from the fuzzy rule base. To show that GFNARX is capable of identifying a significant subset of rules, a literature experiment in which the task is to weed out redundant and inconsequential rules from an existing rule base has been described in Section 3-2. The

results of this rule subset selection are presented in Section 4-2.

The aim of this thesis is to achieve accurate, real-time and interpretable prediction of the electricity market price, which is published by the TSO in the form of a time series, REP. GFNARX is proposed to achieve this aim. Based on the real-time market specifics with regard to Settlement Prices (SPs), the two components of REP, REP_c and REP_s, are forecast separately. As endogenous inputs the three, four and five minutes delayed REP time series values will be used. This corresponds to the three most recent observations published by the TSO. Exogenous inputs which will be used in both REP_c and REP_s forecasts are the most recently observed SCR values, trend in SCR, availability of the option to net imbalances through the IGCC and amount of imbalance netting taking place. Other exogenous inputs which were thought to be influential are not included due to the following:

Wind speed The sampling rate of available historical data is too low to yield a significant correlation to REP. A preliminary test including the KNMI hourly data on wind speed as an exogenous input variable did not result in an improvement of model accuracy

Temperature As with wind speed, the hourly sampling rate of available historical data is too low to yield a significant correlation to REP. As the correlation is lower than that of wind speed, the KNMI hourly data on temperature has not been used as exogenous input

Frequency The real-time electricity market is mainly used to maintain the electricity grid frequency at $f_{nom} = 50Hz$. However, disturbances to the grid frequency occurring anywhere in Europe are resolved with PCR which does not influence REP. The TSO activates SCR and TCR to resolve local disturbances, which are not necessarily proportional to the overall grid disturbance. As a result, no evidence has been found of a significant correlation between measured grid frequency and REP

TCR TCR is one of the three control products the TSO employs to balance the electricity grid. Together with SCR, activation of TCR governs REP. As described in Section 3-3-2 however, the regulation state which influences SP is completely determined by SCR, rather than TCR. Therefore, information contained within the TCR time series is redundant

As no REP prediction benchmark method is known from literature, SARIMAX-GARCH (Section 2-2)and NARX (Section 2-3) models are used to create a comparison. Models are identified based on historical data of the first two weeks of 2015 and tested on prediction accuracy throughout the rest of 2015. The economic value of REP prediction is benchmarked with a newly created benchmark method which simulates application of simple control strategies on a Cold Storage Warehouse (CSW) based on REP predictions. The results of REP prediction accuracy and economic value are presented in Section 4-3.

Chapter 4

Results

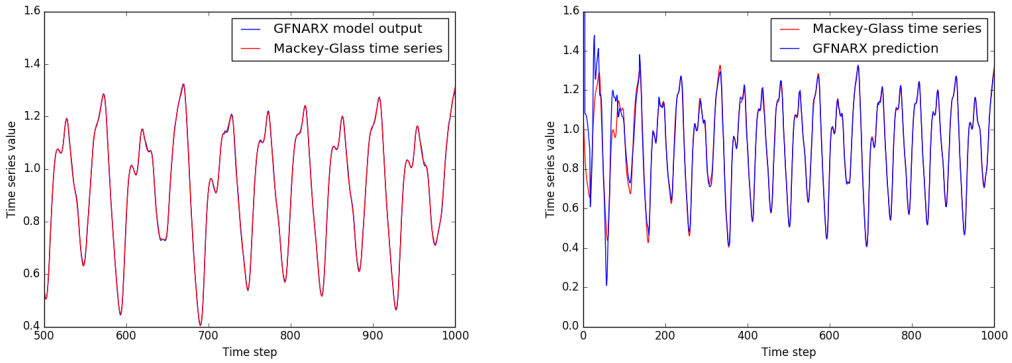
In this chapter, the results of the experiments described in Chapter 3 are presented. The Generalised Fuzzy Neural Network formulation of a Non-linear Auto-Regressive with eXogenous inputs model structure (GFNARX) algorithm described in Section 2-4-4 is used to forecast the Real-time Electricity market Price (REP). As there is no known benchmark method for REP forecasts, a benchmark method simulating a Cold Storage Warehouse (CSW) has been developed in Section 3-4. Two models from time series literature, described in Sections 2-2 and 2-3, have been implemented to serve as reference results on this benchmark.

Benchmarking GFNARX performance with a self-developed benchmark method does not establish that GFNARX contributes to fuzzy time series forecasting literature. As such, it is first used to forecast a time series benchmark popular in fuzzy time series models literature. The results of this literature benchmark prediction are presented in Section 4-1. As determining the appropriate amount of fuzzy rules in GFNARX is part of the training process, a rule pruning method which identifies and deletes inconsequential rules is required. Validation of the activation-based rule pruning method used in GFNARX is described in Section 4-2. The actual REP prediction and its application to load regulation of a CSW model is then described in Section 4-3.

4-1 Mackey-Glass chaotic time series

In this thesis GFNARX is proposed with which to forecast REP. Performance of GFNARX within the context of time series prediction has not yet been established and is topic of this section. REP forecasting has not been studied extensively yet, and there are no benchmark methods known from literature for prediction of REP. Therefore, a popular time series prediction benchmark method known from fuzzy modelling literature, described in Section 3-1, is used to support the contribution of GFNARX to fuzzy modelling literature.

In this experiment, 1000 input-output data pairs are available. The first 500 samples are used



(a) Test data prediction with offline training **(b)** Full prediction with online training

Figure 4-1: Mackey-Glass time series prediction with GFNARX

as training data after which the final 500 samples are predicted without applying any changes to the model. The Mackey-Glass time series and GFNARX model predictions on the test data are shown in Figure 4-1a. After 500 training steps the GFNARX model predictions are hardly distinguishable from the real time series.

Modelling accuracy of GFNARX is compared to various (fuzzy) neural networks found in literature in Table 4-1. As GFNARX is intended to be used as online data-driven modelling algorithm, the performance of GFNARX with online training up to the latest measured output $y(k-6)$ has also been included and shown in Figure 4-1b. To keep the comparison fair, the training window is limited to 500 training samples. In this way, the amount of training samples is kept equal to the original benchmark method where training is performed offline on the first 500 data samples. It can be seen that the accuracy of the proposed GFNARX algorithm is comparable to state-of-the-art fuzzy neural networks which have higher computational complexity or can only be trained offline.

The very best model found in literature (IRSFNN-FuL[51]) uses a trigonometric functional link expansion to create local higher order nonlinear models and reports an accuracy which is significantly better than that of any other model found in literature. However, this IRSFNN-FuL model includes prior knowledge about the range spaces of inputs to statically map these to the domain $[0, 1]$ in advance. In a dynamically changing environment like the real-time electricity market for which GFNARX is proposed, it is not trivial to accomplish a sensible mapping which correctly represents the total input range spaces at all times, as statistical properties of inputs can change over time. Although it might be possible to create a dynamic mapping to the domain $[0, 1]$, it is known from image reconstruction literature that this can result in alteration of observable contrast[70]. In the least squares weight estimation framework of GFNARX, this would impact the singular values of the regression matrix, which in turn influences the model parameter estimation. Investigating whether such a dynamic mapping can stably be achieved or whether it can enhance accuracy of identified models, is outside of the scope of this thesis.

Implementing such a functional link consequent inference model in the GFNARX algorithm

Table 4-1: Performance of GFNARX and other models on prediction of the Mackey-Glass time series sorted by RMSE with the best model at the top

Paper	**Model name**	**RMSE**
Lin, Chang and Lin, 2013 [51]	IRSFNN-FuL	0.0002
Yilmaz and Oysal, 2010 [52]	FWNN	0.0023
This thesis	**GFNARX (online training)**	**0.0026**
Juang, Lin and Tu, 2010 [53]	RSEFNN-LF	0.0031
Juang and Tsao, 2008 [54]	SEIT2FNN	0.0034
Chen, Yang and Dong, 2006 [55]	LLWNNH	0.0036
Lin, Chang and Lin, 2013 [51]	IRSFNN-TSK	0.0039
This thesis	**GFNARX**	**0.0049**
Paul and Kumar, 2002 [56]	SuPFuNIS	0.0057
Juang, Chung and Hsu, 2007 [57]	CSPSO	0.0064
Russo, 2000 [58]	GEFREX	0.0067
Lin, Chen and Lin, 2008 [59]	SEELA	0.0068
Wu and Er, 2000 [60]	D-FNN	0.008
Lin, Chen and Lin, 2009 [61]	FLNFN-CCPSO	0.0084
Kim and Kasabov, 1999 [62]	HyFIS	0.01
Soto et al., 2013 [63]	Ensemble type-2 ANFIS	0.0119
Juang, 2002 [64]	TRFN-S	0.0124
Cho and Wang, 1996 [65]	RBF-AFS	0.0128
Pulido et al., 2013 [66]	Ensemble ANN	0.0173
Sapankeveych and Sankar, 2013 [67]	SVR-PSO	0.0379
Nauck and Kruse, 1999 [68]	NEFPROX	0.0533
Gaxiola et al., 2014 [69]	Type-2 FNN	0.055

to predict the Mackey-Glass time series, did not result in an improvement in model accuracy. Applying the same trigonometric functional link expansion to the GFNARX inference mechanism as was used in [51], RMSE of the prediction was increased to $RMSE_{FL} = 0.0061$, which constitutes a 24.5% relative increase in error. This might be due to the Linear Least Squares (LLS) weight estimator being incapable of producing unbiased weight parameter estimates in the case where the consequent inference model is nonlinear. This warrants further investigation though, which is not within the scope of this thesis.

It is possible to apply a functional link consequent model which does not make assumptions about the input and output space, to the proposed algorithm. This requires all possible input combinations $u_i u_j$ to be fed through the network, resulting in $(N_u^2 - N_u)/2$ weights being required per rule. Convergence is then slow and the computational load renders it intractable for online dynamic estimation. Trying a functional link expansion with all possible input combinations, no model convergence has been achieved within the training data set. The computation time also extended to over a day as opposed to approximately one minute with the linear TSK-style inference. This linear Takagi-Sugeno-Kang (TSK)-style model interpolates in a nonlinear manner described by Equation 2-30 between local affine models.

Although a short experiment with functional link inference models did not yield an improvement in this thesis, in theory it should be possible to improve the prediction accuracy of continuous signals by mapping the inputs to a higher order feature space. Taylor's theorem suggests that for continuous functions which are at least twice differentiable, higher accuracy is achieved by approximating with higher order polynomials than first order affine models. On the other hand, regression also becomes slower, as more parameters are involved. For example, when using local quadratic polynomials as inference basis functions, the amount of regression parameters approximately doubles. This leads to the connection weights w_j in Equation 2-30 being described by Equation 4-1 and the regression matrix width (Equation 2-37) $\nu = n_r(2n_u + 1)$

$$w_j = k_{0,j} + \sum_{i=1}^{N_u} k_{i,j} u_i + \sum_{i=N_u+1}^{N_u} k_{i,j} u_i^2 \tag{4-1}$$

with

$k_{i,j}$ Individual weights

u Input vector of length N_u

The application of different functional link basis functions which incorporate higher order dynamics into the GFNARX inference mechanism still requires more study.

In this section, the accuracy of GFNARX has been benchmarked on a reference time series prediction problem posed in literature. The RMSE of dynamically trained GFNARX predictions ranks among the top fuzzy models found in literature. Only the fuzzy neural network proposed by [51] with a nonlinear inference mechanism is reported to have a significantly better performance. GFNARX performance is thus comparable to state-of-the-art fuzzy neural networks found in literature while using computationally cheaper means. In this thesis the

output was constructed from interpolating between local affine models. A possible improvement to GFNARX which was not within the scope of this thesis is to find an efficient nonlinear inference structure in which the output is constructed from local nonlinear models as was done in the IRSFNN-FuL network[51]. The next section will describe results on another literature experiment in which inconsequential rules are pruned from an existing corrupted rule base.

4-2 Activation-based rule subset selection

In this section, results of the experiment described in Section 3-2 are given. The experiment is to find an importance ordering of rules in an existing corrupted rulebase with as goal to assign the least importance to redundant and inconsequential rules. Then, if a subset of rules is chosen to represent the system in a more sparse way, the reduced rulebase will not contain these redundant or inconsequential rules. GFNARX ranks rule importance by the simple metric of fraction of activation, instead of assessing explanatory influence on the model fit. The rule ranking by fraction of activation is discussed in Section 4-2-1. As the importance ordering obtained by GFNARX does not resemble the ordering obtained by using orthogonal decomposition methods or rank-revealing methods as described by [7], the prediction accuracy of GFNARX when using a subset of rules based on the importance ordering is scrutinised in Section 4-2-2.

4-2-1 Ranking rules based on fraction of activation

In this section, the capability of the GFNARX method of ranking rule importance to assess which rules are least influential with regards to model predictions, is investigated. Starting with the 25 numbered rules in the initial rulebase given in the modeling problem of [8] (see Table 3-1), the first 1000 noisy data pairs obtained by the data generating system described in Equations 3-3 and 3-4 are used to initialise the GFNARX regression matrix, after which the rules are ordered by their fraction of activation f_{act}. This ordering, along with the importance orderings obtained by using the methods described in [7], are shown in Table 4-2.

It can be seen from Table 4-2 that the bottom five rules in the importance ordering based on f_{act} correspond to the bottom five rules in the importance ordering of the P-QR method which is acclaimed by [7] to yield the best generalising performance. As the experiment was designed to have three redundant and 2 non-firing rules, this is the desired result. To select an appropriate cut-off threshold K_{act}, a closer look is taken at the fractions of activation of the redundant and inconsequential rules. Table 4-3 shows f_{act} for the bottom 10 rules in the importance ordering shown in Table 4-2. Based on these values, $K_{act} = 0.025$ is chosen as threshold for fraction of rule activation.

It can be seen though that the importance ordering obtained by f_{act} differs from the ones obtained by output contribution (OLS, E-OLS) and rank-revealing methods (SVD-QR, P-QR). This can partly be attributed to the random disturbance, which results in slightly different rule activation fractions each time the experiment is run, but there is no evidence yet that the ordering makes sense. The rule ordering obtained by SVD-QR is claimed by [7] to be qualitatively close to a random approach in selecting rules, so lack of resemblance to this ordering is not a bad thing. The orderings obtained by OLS and E-OLS however, are

Table 4-2: Importance ordering of the rules, referred to by number, provided in Table 3-1 according to various ordering methods (most important at top). It is known that there are 3 duplicate and 2 non-firing rules, so the bottom five rules are deleted from the rule base

SVD-QR	E-OLS	OLS	P-QR	f_{act}
25	5	5	24	8
4	24	24	25	6
19	25	25	6	12
7	16	16	15	21
3	8	8	23	23
24	21	21	8	22
8	23	23	4	25
13	11	11	12	18
23	3	3	11	9
14	22	22	14	2
21	7	6	18	19
17	9	7	19	7
22	19	15	22	14
18	4	19	17	11
12	14	4	7	4
9	18	9	21	17
11	1	17	3	3
2	17	13	13	15
5	13	18	9	24
16	12	12	2	13
15	2	1	20	1
10	6	2	5	5
20	10	14	16	10
6	15	10	1	16
1	20	20	10	20

Table 4-3: Fraction of activation f_{act} for the bottom 10 rules in the importance ordering shown in Table 4-2

Rule number	17	3	15	24	13	1	5	10	16	20
f_{act}	0.030	0.029	0.028	0.026	0.025	0.020	0.004	0.000	0.000	0.000

based on their estimated contributions to the overall model fit. In P-QR rules are ordered in terms of generalisation ability, as columns from the regression matrix which contribute the most are picked first. The f_{act} importance ordering is also based on generalisation ability, so the ordering should intuitively resemble the one obtained with P-QR. This is not the case. Redundant and non-firing rules are assigned the lowest importance though, which establishes the capability of f_{act} importance ordering to filter out corrupted rules from a rule base. The next section investigates prediction accuracy using a subset of rules chosen based on the f_{act} based importance ordering to show that the ordering makes sense.

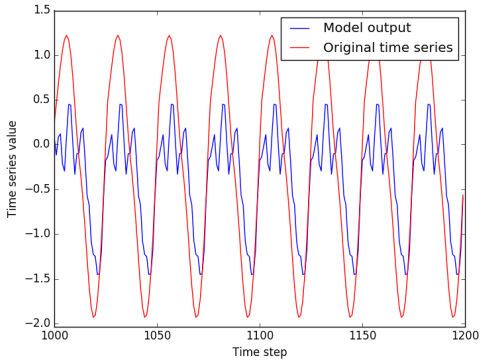
4-2-2 Prediction accuracy

In the previous section, it was shown that f_{act} based ordering of rule importance results in assigning the least importance to redundant and non-firing rules as expected. The importance ordering obtained, however, differs from that obtained by other importance ranking methods described in [7]. Therefore a closer look is taken at the importance ordering obtained with GFNARX in this section.

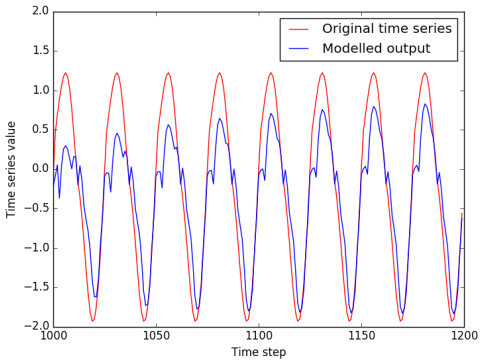
Estimating GFNARX weight parameters based on the noisy training data, results in a model fit on the evaluation data as shown in Figure 4-2a. It can be seen that the weights are not estimated correctly, resulting in a bad fit. Due to the extremely low signal-to-noise ratio the GFNARX algorithm, which assumes a high signal-to-noise ratio, is unable to produce a good estimation of the plant output. When estimating the parameters online, the model fit increases in accuracy during the duration of the evaluation experiment, when the disturbance is no longer random. Figure 4-2b shows this increasingly accurate model fit with online parameter estimation for the rulebase with $f_{act} \geq 0.025$, which constitutes the cut-off between redundant and significant rules in this experiment (see Table 4-3).

To then establish that the obtained rule importance ordering makes sense, model accuracy is tested starting with only the most important rule and subsequently adding rules according to the ordering shown in Table 4-2. GFNARX weight parameters are estimated in an online fashion, while the structure remains unchanged. Table 4-4 shows final model accuracy on the evaluation data when using a subset of rules. The result is not quite apparent when including the noisy data, but when the noisy data is discarded, it is seen that the selected rules generalise very well and fit accuracy is very high even with a small subset of rules.

In this section, the ability of the proposed f_{act} method of rulebase pruning to reject redundant or insignificant rules is shown. On top of that, the importance ordering obtained by this method shows good generalisation capabilities. Evaluation data is fitted well even with a small subset of the rulebase selected from an importance ordering obtained from training data, as long as the weights are estimated correctly. In the REP time series considered in the real-time electricity market, the published REP time series is noise-free, so the inability of the algorithm to properly identify a model under noisy circumstances is not thought to be of consequence in the intended use case. The ability of GFNARX to properly identify a model and extend predictions to out-of-sample data in a noise-free environment has been



(a) Model fit based on parameters estimated from noisy training data



(b) GFNARX model accuracy steadily increases by continuing to estimate weight parameters online

Figure 4-2: GFNARX prediction for the evaluation data of the experiment described in Section 3-2

Table 4-4: Model accuracy with online weight estimation from either all data samples or just the samples with sinusoidal disturbance when using a subset of rules according to the f_{act} importance ordering of Table 4-2. The metric used is Root-Mean-Square Error

Amount of rules	Noisy+sinusoidal (RMSE)	Sinusoidal only (RMSE)
1	0.45	0.230
2	0.38	0.145
3	0.36	0.066
4	0.26	0.030
5	0.24	0.012
6	0.23	0.009

benchmarked in the previous section. The next section will report on application of GFNARX to forecast REP, which is the aim of this project.

4-3 Real-time market price time series prediction

In this section results of forecasting the Real-time Electricity market Price (REP) are presented. As prediction accuracy and economic value of using predictions are not correlated in a one-to-one fashion, results pertaining to accuracy and value are presented separately. Section 4-3-1 describes the prediction of REP itself, while section 4-3-2 presents results of operating the CSW described in Section 3-4 based on REP predictions.

4-3-1 Model accuracy

Forecasting REP time series has not been done in literature. In this section, the accuracy of REP predictions obtained by GFNARX are investigated and benchmarked against NARX and SARIMAX-GARCH predictions.

As described in Section 3-3-5, the training data consists of the first two weeks of January 2015, the rest of 2015 is used as test data. Prediction accuracies of the various models are given in Table 4-5. These values are visualised in Figure 4-3. As the time series data is published by the Transmission System Operator (TSO) with a 3 minute delay, the shortest prediction horizon considered, is three steps ahead. It can be seen that REP_c prediction errors are consistently higher than REP_s prediction errors. This is to be expected due to REP_c being more volatile and the pre-processing of the series, where REP is clipped at $REP_{c,max} = 200 Eur/MWh$, while $REP_{s,max} = 100 Eur/MWh$, as described in Section 3-3-3.

Figure 4-4 shows a single peak event of the REP_c signal, along with its prediction by the models. It is clear that model predictions are more accurate than the naïve reference for each model, as all models show a quicker response to changes in the REP_c than the naïve model. The nonlinear networks show a quicker response than SARIMAX-GARCH, with GFNARX fitting the REP_c shape best, as NARX predictions have more overshoot. This is not an artefact of this single peak, as is seen in Figure 4-5 where REP_c prediction over the course of half a day is shown.

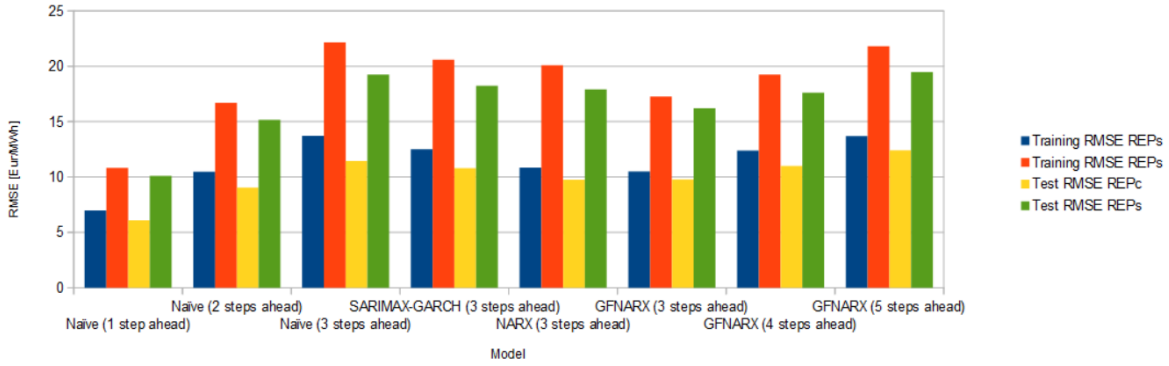


Figure 4-3: Prediction accuracies of various models for the training data and test data

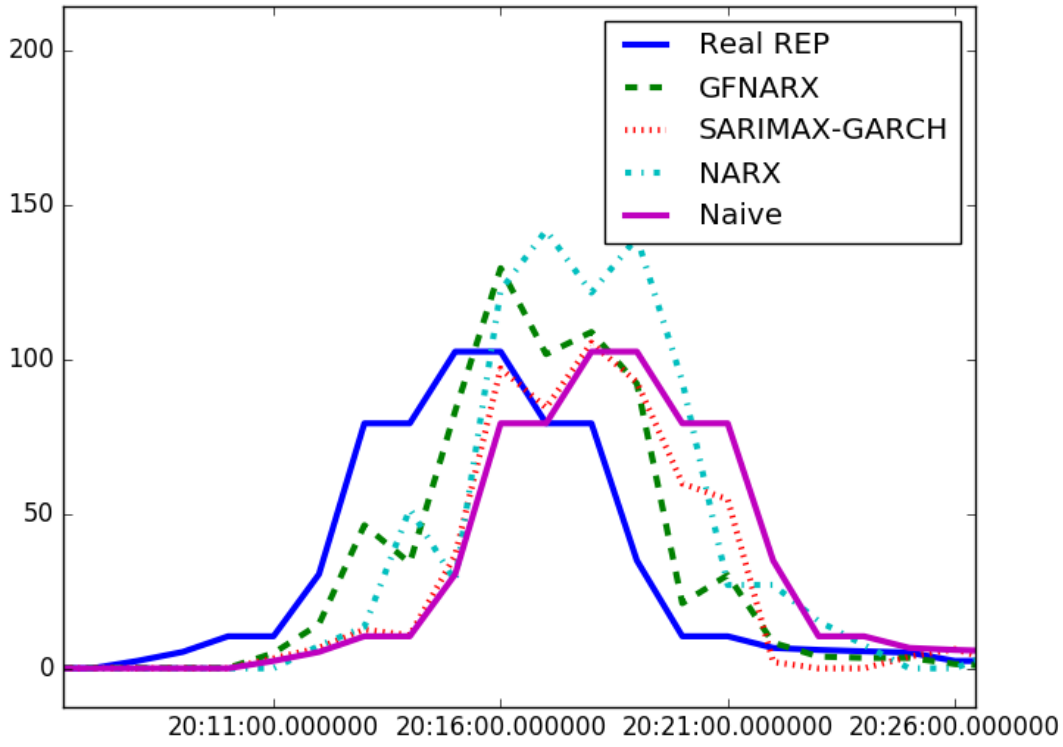


Figure 4-4: Predictions of REP_c for a single market peak event by SARIMAX-GARCH, NARX and GFNARX. The solid lines denote the real REP_c signal and its naïve forecast. All models outperform the naïve reference forecast

Table 4-5: Prediction errors of the various models on training data and test data of REP_c and REP_s. It can be seen that GFNARX is the most accurate model and predicting 5 steps ahead still yields more accurate predictions than the naïve reference. All values are given in [Eur/MWh]

Model (prediction horizon)	Training RMSE REP_s	Training RMSE REP_c	Test RMSE REP_s	Test RMSE REP_c
Naïve (3 steps)	13.71	22.13	11.44	19.25
SARIMAX-GARCH (3 steps)	12.50	20.58	10.81	18.22
NARX (3 steps)	10.84	20.09	9.77	17.90
GFNARX (3 steps)	10.51	17.25	9.78	16.21
GFNARX (4 steps)	12.38	19.25	11.00	17.61
GFNARX (5 steps)	13.69	21.80	12.42	19.47

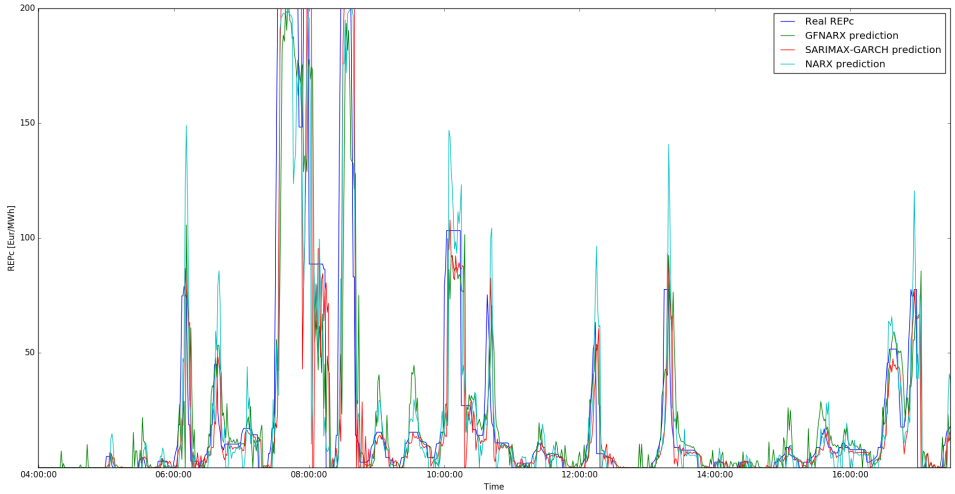


Figure 4-5: Predictions of REP_c by SARIMAX-GARCH, NARX and GFNARX over the course of half a day. It can be seen that the predicted values differ significantly from each other with NARX predictions producing the largest overshoot

There are two things of particular note in the REP predictions:

RMSE on training data is higher than on test data This is due to the first two weeks of January, which are used as training data, being more volatile than average. The standard deviation SD_c of REP_c equals $SD_{c,training} = 39.80 Eur/MWh$, while $SD_{c,test} = 32.35 Eur/MWh$. Similarly, $SD_{s,training} = 20.77 Eur/MWh$ and $SD_{s,test} = 16.54 Eur/MWh$

Model performance is consistent, with GFNARX outperforming the benchmark models Only REP_s test data prediction sees NARX slightly outperforming GFNARX, but overall GFNARX yields the most accurate predictions of both training and evaluation data

In this section, REP prediction accuracy has been investigated for GFNARX, NARX and SARIMAX-GARCH models. All of the models beat the naïve reference model which assumes $\hat{y}(t) = y(t-3)$, with GFNARX yielding the best accuracy. If GFNARX prediction is extended to five steps ahead, the forecasts are still more accurate than this naïve reference. The next section describes results of using REP predictions to apply thermal control to a simulated Cold Storage Warehouse.

4-3-2 Application of predictions to naïve control of a cold storage warehouse

In the previous section, model accuracy with regards to REP predictions has been investigated. As these REP predictions are only useful if one uses them to trade on the real-time electricity market, this Section benchmarks the result of adjusting electricity usage in a simulation of a CSW according to the forecasted REP. This CSW model simulation has been described in Section 3-4. Deviations from the reference case, which entails the electricity program being perfectly forecast and the electricity transactions transpiring at the Amsterdam Power

Exchange (APX) day ahead price, are traded at their corresponding Settlement Price (SP). The relation between SP and REP can be found in Section 3-3-2.

The rest of this section is structured as follows. First, the actual result in terms of cost of operation is given, before discussing the applicability of this result to other types of assets and other periods of time.

Cost of CSW operation Reduction in cost of CSW operation is what is benchmarked in this simulation. As described in Section 3-4, the models are trained on two weeks of data and subsequently tested on data of the rest of the year 2015. This constitutes an out-of-sample period which is 25 times as large as the training period. This coincides with the training and test periods of the prediction accuracy benchmark, the results of which are presented in Section 4-3-1.

The cumulative electricity costs of operating the simulated CSW in 2015 based on model predictions, are given in Table 4-6. The Naïve strategy refers to regulating the asset directly according to the most recently published REP values with the strategy currently implemented at Peeeks, the company where this project is carried out. The hypothetical situation where the TSO would publish REP with a delay less than three minutes has also been included.

Table 4-6: Cumulative electricity costs of operating the CSW in 2015 based on model predictions. Whenever it is not explicitly mentioned, the strategy used is the Peeeks strategy. The models are ordered by descending costs

Strategy (prediction horizon)	$Cost_{tot}$ $[Eur]$	$Cost_{avg}$ $[Eur/MWh]$	Cost reduction [%]
Nominal operation	195682	37.12	0.0
Naïve strategy (3 minutes)	162547	30.86	16.9
SARIMAX-GARCH prediction (3 minutes)	159883	30.45	18.3
GFNARX prediction (6 minutes)	156412	29.68	20.1
GFNARX prediction (5 minutes)	151584	28.75	22.5
GFNARX prediction (4 minutes)	150190	28.49	23.3
NARX prediction (3 minutes)	150134	28.48	23.3
Naïve prediction (2 minutes)	149538	28.36	23.6
Naïve prediction (1 minute)	146624	27.81	25.1
GFNARX prediction (3 minutes)	145871	27.67	25.5
Perfect forecast	142563	27.03	27.1

When regarding the prediction horizon corresponding to the actual delay at which REP is published, nearly the same model ranking is obtained as would be expected from model accuracy (see Table 4-5). However, SARIMAX-GARCH predictions perform poorly when used to control electricity usage of the simulated CSW, as these predictions show slow response to market events (see Figure 4-5).

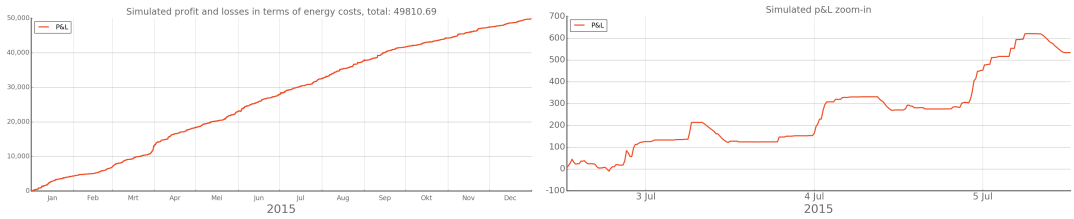
Whereas none of the models was able to outperform naïve forecasts at a hypothetical two minute prediction horizon in terms of RMS prediction error, controlling the CSW according to predictions obtained by GFNARX results in a lower cost of operation. As GFNARX predictions are quick at following the trend, yet tend to overshoot the actual REP, this result

was to be expected. The cost of operation with GFNARX predictions is also lower than using the control strategy in conjunction with a naïve forecast in the hypothetical situation where the REP publish delay is only one minute. As the RMS prediction error in this same hypothetical situation is only 62% of the RMS error of GFNARX at a three step delay, it is clear that GFNARX predictions capture REP trends better than is reflected in the prediction accuracy alone.

The cost reduction obtained by using the hypothetical situation of perfectly forecast REP data is the optimal cost reduction obtainable with the strategy used. Controlling CSW cooling power according to GFNARX forecasts results in 93.8% of this optimal cost reduction, which is considered a good result. It outperforms controlling according to the hypothetical situation of 1 minute delay in REP publishing.

Robustness of prediction method To establish that the 25% reduction in costs of operating the simulated CSW according to REP forecasts by GFNARX is obtained in a stable way, the profit and loss progression throughout the year is investigated. Figure 4-6a shows a rolling cumulative profit line throughout the year 2015. It can be seen that over the course of the year the reduction in costs increases steadily. On shorter time scales the profit line fluctuates a bit more and sometimes losses are incurred. Figure 4-6b shows the profit line for part of July 2015. Although sometimes dynamically controlling the CSW results in losses, it is apparent that profitable actions occur more frequently and with larger impact.

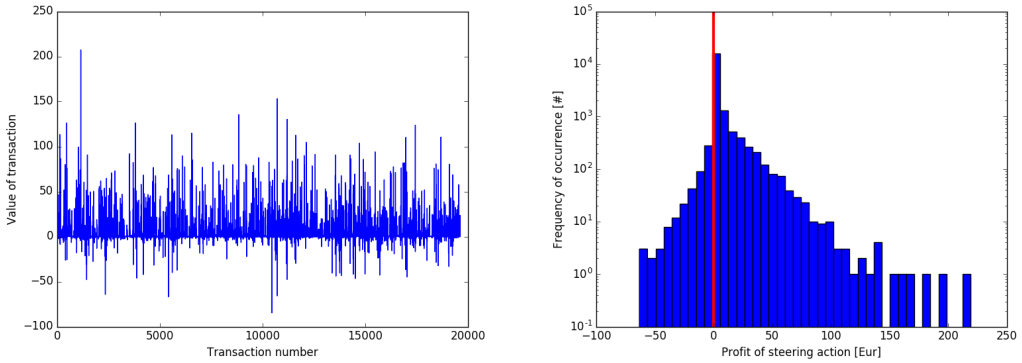
This is also confirmed when looking at Figure 4-7. Figure 4-7a shows the value of all individual transactions on the real-time electricity market. It is clear that the magnitude of profitable transactions on the real-time electricity market exceeds the magnitude of disadvantageous transactions. The distribution of transaction values can be seen in Figure 4-7b. The chance of regulating according to REP predictions resulting a profitable transaction occurring, vastly exceeds the chance of incurring a loss when doing so.



(a) Profit and loss line for the year 2015 when regulating the CSW according to GFNARX REP forecasts

(b) Profit and loss line for 3 days in July 2015 when regulating the CSW according to GFNARX REP forecasts

Figure 4-6: Cumulative profit over time



(a) Value per transaction **(b)** Distribution of value per transaction

Figure 4-7: Value of individual transactions on the real-time electricity market. (a) shows the value in sequence per transaction while (b) shows the probability distribution. The rightmost bin agglomerates all transactions netting a profit more than 5 Euro, while the leftmost bin does the same for all transactions incurring losses greater than 5 Euro

There are two reasons that some transactions on the real-time electricity market turn out negative:

Due to the standard behaviour of the modelled CSW If the threshold price premium $\Delta P_{threshold} \geq max(REP_c, REP_s)$, the asset will ramp up or down corresponding to its current temperature, which is a function of its internal imbalance. The internal balance is specified by the cumulative electricity usage deviation from its forecasted value. A full description of the standard CSW behaviour can be found in Section 3-4

Due to the discretisation of ISP It is possible due to the difference in time resolution between REP and SP that until halfway through a PTU, the cooling motor will be given advice to scale up its electricity consumption. However, if at the end of the PTU the single instance of $REP_c > 0$ causes the SP to be settled according to the regulation state $S = 2$ pricing scheme (see Section 3-3-2), this means that the extra electricity consumption throughout the Program Time Unit (PTU) is settled at a disadvantageous price

The year 2015 is not an exception. Although the market does change over time, the fact that there is added value to predicting REP is consistent over the years. Using the same GFNARX trained on the first 2 weeks of January 2015 to predict REP in 2014 and 2016 yields 17.8% and 25.0% reduction in costs when compared to nominal operation of the CSW. The costs are summarized in Table 4-7.

Applicability to other types of assets A CSW is an asset which has a relatively large thermal buffer and is thus well suited to be an asset in a demand response portfolio. Although the CSW model has been derived from first principles and not specifically catered to fit REP predictions, the model does make a few assumptions which are not applicable to all possible assets in such a portfolio:

- Infinite ramp speed in load. The cooling power is either on or off and there are no states in between

Table 4-7: Cost of operating the simulated CSW according to REP predictions with GFNARX over the years

Year	Mode of operation	Total costs [Eur]	Average costs [Eur/MWh]	Cost reduction [%]
2014	Nominal operation	202376	38.14	0.0
2014	GFNARX prediction	166165	31.32	17.9
2015	Nominal operation	195682	37.12	0.0
2015	GFNARX prediction	145871	27.67	25.5
2016	Nominal operation	156558	29.65	0.0
2016	GFNARX prediction	117477	22.26	25.0

- Full availability. The cooling motor can switch between off and on states without limit as long as the temperature stays within operational bounds. In practice there will often be a limit to the amount of times a device can switch states within a short period

Both of these issues pertain to persistence in REP. If changes from REP_c spikes to REP_s spikes and vice versa occur frequently, then changes in regulation advice can outpace the switch limit of a device. Moreover, in case of a cooling motor with finite ramp speed, rapid changes in REP cannot be tracked. There are situations where the electricity grid is clearly in a state of overconsumption or oversupply and hence have a regulation state (see Section 3-3-2 for specifics about regulation states) which persists for long times. As can be seen in Figure 4-8 however, the grid state most frequently changes between subsequent PTUs.

To then establish whether REP predictions can be used in conjunction with other assets, the cost reduction obtained by using GFNARX predictions to operate a CSW with a motor which takes five minutes to fully ramp between states is investigated. The resulting cost of operation is $Cost = 150229Eur$. This is indeed slightly more than the $Cost = 145871Eur$ under ideal motor conditions, but the cost reduction is still 23.2%. Thus even when motor conditions or asset limitations are not ideal, a cost reduction can be achieved when predicting REP.

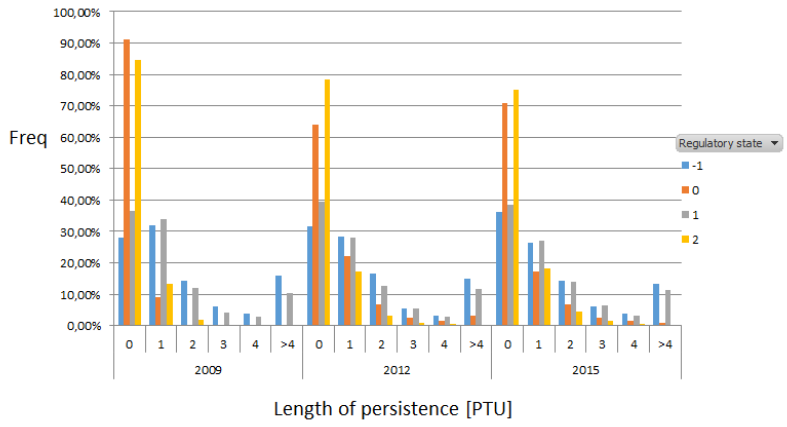


Figure 4-8: Normalized occurence of regulation state length of persistence. The sudden increase in persistence of regulation state $S = 0$ from 2012 onwards is attributed to the introduction of IGCC. It can be seen that the state most frequently changes from PTU to PTU

Chapter 5

Conclusion

5-1 Conclusions

The aim of this project is to compensate for the delay with which Real-time Electricity market Price (REP) is published by the Transmission System Operator (TSO) by using models to forecast REP.

From this vantage point fuzzy logic models, which have been developed to incorporate human reasoning into modelling, have been investigated. Currently existing fuzzy models are often trained offline with batch data. The initial amount of data clusters is overestimated and subsequently shrunk to the optimal amount. In a dynamic environment like the electricity market, the optimal amount of clusters need not be static in time. An existing approach which adaptively builds a model based solely on incoming data is Generalised Fuzzy Neural Network (G-FNN). However, G-FNN has some flaws which are fatal to its accuracy. Data clusters are formed around the first point attributed to the cluster instead of optimised to represent the cluster in its entirety. Moreover, there is no cap on the amount of rules generated, which can lead to instability due to singularity of the regression matrix due to overparametrisation.

In this thesis, Generalised Fuzzy Neural Network formulation of a Non-linear Auto-Regressive with eXogenous inputs model structure (GFNARX) is proposed as an improvement to G-FNN. The improvements constitute:

Bounding the amount of rules If addition of a rule would render the regression matrix singular, it is not added. Instability due to overparametrisation is avoided by doing so

Moving average cluster centring When attributing a data point to an existing cluster, the cluster centre is recalculated as the average position of all points within the cluster. In this way, cluster locations optimised in an online fashion

Normalised distance as cluster separation criterion G-FNN uses Euclidian distance as distance norm between adjacent clusters, but using a normalised distance instead yields a better description of the feature space

Identifying unnecessary rules by fraction of activation The metric to assess rule importance in G-FNN overestimates the importance of redundant rules. Assessing rule importance by their overall fraction of activation is computationally less expensive and solves this problem

The following conclusions are made about four research aims posed in Section 1-3:

1. First of all it is investigated whether the proposed data-driven fuzzy modelling technique is capable of producing time series predictions which are comparable in accuracy to computationally more expensive models in literature. There is no known benchmark method for REP forecasting. To then verify whether GFNARX produces accurate models, it has been used on a benchmark time series prediction experiment which is popular in fuzzy modelling literature, the Mackey-Glass time series. GFNARX model predictions result in a Root-Mean-Square Error (RMSE) of $RMSE = 0.0026$, which ranks GFNARX performance on this benchmark near the top of fuzzy modelling algorithms found in literature.

2. Secondly, it is investigated which observable endogenous and exogenous factors can be used as influential inputs to a REP model identified through time series modelling techniques. As endogenous inputs the three, four and five minutes delayed REP time series values are used. This corresponds to the three most recent observations published by the TSO. Exogenous inputs which need to be used in REP forecasts are the most recently observed Secondary Control Reserve (SCR) values, trend in SCR, availability of the option to net imbalances through the International Grid Control Cooperation (IGCC) and amount of imbalance netting taking place

3. The third research challenge is to find which model structure of the ones selected in [14] (SARIMAX-GARCH, NARX, GFNARX) predicts REP most accurately. Historical data of the year 2015 has been used with which to forecast REP. The first two weeks of data are used as training data, the rest of the year as evaluation data. All three models yield predictions which are more accurate than the naïve reference which assumes the most recently observed REP value to be equal to the current value. Out of these three models, GFNARX yields the most accurate REP forecasts. Up to five steps ahead prediction, REP forecasts by GFNARX are more accurate than the naïve reference

4. The final research aim is to incorporate REP predictions in the demand response control strategy of Peeeks and investigate how much of the negative effects of the delay in publishing of REP can be mitigated when using REP forecasts to control the cooling motor of a Cold Storage Warehouse (CSW). In this thesis, a CSW model has been derived from literature models and data from an actual CSW in order to create a simulation benchmark for economical value of REP forecasts. The baseline cost is defined by avoiding use of the real-time market at a strategic resource as is done in literature. All electricity is bought at the day-ahead auction. Using this benchmark method with historical data of 2015, a 25.5% reduction in electricity costs is achieved

by using GFNARX to forecast REP. For historical data of 2014 and 2016, reductions in cumulative electricity costs of 17.8% and 25.0% are observed. It is concluded that dynamically controlling the CSW based on REP forecasts using GFNARX yields a robust reduction in electricity costs. Controlling the CSW cooling motor based on the naïve reference forecast yields a 16.9% reduction in costs compared to the baseline. The relative reduction in costs by using GFNARX to forecast REP is thus 10.3% of the naïve reference model costs

The aim of this project is to compensate for the delay with which REP is published by means of forecasting. The REP is published by the TSO as a time series with an approximate delay of three minutes. The hypothesis was that negative effects of this delay can be compensated for by forecasting REP. Controlling a simulated CSW according to GFNARX forecasts of REP, the average electricity costs are $27.67 Eur/MWh$. The reference which uses the delayed observations as naïve forecast acquires electricity at an average cost of $30.86 Eur/MWh$. If REP would be perfectly forecast, the average costs of acquiring electricity would be $27.03 Eur/MWh$. This means that 83% of the effect of delay is compensated for in this simulation.

5-2 Recommendations for further research

In this thesis GFNARX has been proposed as an improvement to G-FNN with the aim of achieving accurate, real-time and interpretable prediction of the electricity market price. Although the added value of using GFNARX to predict REP has been shown, there are still some topics which can benefit from further research.

- The cut-off for fraction of activation $f_{act} \geq 0.025$ has been selected based on a literature experiment. The assumption that this cut-off is rigorous enough to reject redundant rules from the rulebase requires more analysis. This could be tested by devising more experiments in which rulebases are corrupted
- In the analysis of exogenous variables which are potentially related to REP, it has been seen that REP correlation to publicly available data on temperature and wind speed is low due to the low sampling frequency of these weather conditions. As electricity supply by renewable energy sources and electricity consumption are dependent on weather conditions, it is plausible to assume that a correlation between weather conditions and REP exists. This correlation needs to be studied further by using weather data sampled at a higher frequency
- The actual control input which determines SCR activation is the measured cross-border electricity flow. If data on these cross-border flows can be observed, it might be possible to obtain more accurate REP forecasts by predicting the cross-border flow and reverse engineering REP
- Applications of REP forecasts have not received enough attention in literature. This thesis investigates application of REP forecasts to thermal control of a simulated CSW which results in a reduction in electricity costs, but a challenge in the future will be to

find and document other applications of REP forecasts. A useful application could be to apply REP forecasts to control the charging process of an electric car if time is not the critical factor

Bibliography

[1] M. Ibrahim, "Energy forecasting using model parameter estimation," May 27 2004. US Patent App. 10/302,599.

[2] J. M. P. Menezes and G. A. Barreto, "Long-term time series prediction with the narx network: an empirical evaluation," *Neurocomputing*, vol. 71, no. 16, pp. 3335–3343, 2008.

[3] R. Babuška, "Fuzzy and neural control, disc course lecture notes," 2001.

[4] "Fuzzy inference systems." http://www.cs.princeton.edu/courses/archive/fall07/cos436/HIDDEN/Knapp/fuzzy004.htm. Accessed: 2016-12-19.

[5] Y. Gao and M. J. Er, "Narmax time series model prediction: feedforward and recurrent fuzzy neural network approaches," *Fuzzy sets and systems*, vol. 150, no. 2, pp. 331–350, 2005.

[6] L. K. Szeto, A. W.-C. Liew, H. Yan, and S.-s. Tang, "Gene expression data clustering and visualization based on a binary hierarchical clustering framework," *Journal of Visual Languages & Computing*, vol. 14, no. 4, pp. 341–362, 2003.

[7] M. Setnes, *Complexity Reduction in Fuzzy Systems*. PhD thesis, Delft University of Technology, 2001.

[8] J. Yen and L. Wang, "Simplifying fuzzy rule-based models using orthogonal transformation methods," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 29, no. 1, pp. 13–24, 1999.

[9] TenneT, "Balans-delta igcc." http://www.tennet.org/bedrijfsvoering/Systeemgegevens_uitvoering/Systeembalans_informatie/balansdeltaIGCC.aspx. Accessed: 30-01-2017.

[10] L. Lukasse, J. Broeze, and S. van der Sluis, "Optimal control and design of a cold store using dynamic optimization," *Optimal Control Applications and Methods*, vol. 30, no. 1, pp. 61–75, 2009.

[11] R. A. Verzijlbergh and Z. Lukszo, "Conceptual model of a cold storage warehouse with pv generation in a smart grid setting," in *Networking, Sensing and Control (ICNSC), 2013 10th IEEE International Conference on*, pp. 889–894, IEEE, 2013.

[12] R. Stoeckle, *Refrigerated warehouse operation under real-time pricing*. PhD thesis, University of Wisconsin-Madison, 2001.

[13] M. Buitenhuis, "Modelling the flex in cold stores," 2015.

[14] I. de Hoogt, "Literature study," 2016.

[15] M. H. Albadi and E. El-Saadany, "A summary of demand response in electricity markets," *Electric power systems research*, vol. 78, no. 11, pp. 1989–1996, 2008.

[16] R. Van Der Veen and L. De Vries, "The impact of microgeneration upon the dutch balancing market," *Energy Policy*, vol. 37,7, pp. 2788–2797, July 2009.

[17] TenneT, *Preparation of E-Programmes & T-Forecast*, 2010.

[18] R. Van Der Veen, A. Abbasy, and R. Hakvoort, "Agent-based analysis of the impact of the imbalance pricing mechanism on market behavior in electricity balancing markets," *Energy Economics*, vol. 34, pp. 874–881, 2012.

[19] TenneT, "The imbalance pricing system as at 01-01-2001, revised per 26-10-2005," 2015.

[20] G. Klæboe, A. L. Eriksrud, and S.-E. Fleten, "Benchmarking time series based forecasting models for electricity balancing market prices," *Energy Systems*, vol. 6, no. 1, pp. 43–61, 2015.

[21] G. E. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time series analysis: forecasting and control*. John Wiley & Sons, 2015.

[22] L. Tang and H. Xu, "Two necessary and sufficient conditions of mean ergodicity about covariance stationary process," in *2010 International Conference on Computer Application and System Modeling (ICCASM 2010)*, vol. 10, pp. V10–244, IEEE, 2010.

[23] R. Schlittgen and B. H. Streitberg, *Zeitreihenanalyse*. Oldenbourg Verlag, 2001.

[24] T. Bollerslev, "Generalized autoregressive conditional heteroskedasticity," *Journal of econometrics*, vol. 31, no. 3, pp. 307–327, 1986.

[25] R. Weron, "Electricity price forecasting: A review of the state-of-the-art with a look into the future," *International Journal of Forecasting*, vol. 30, pp. 1030–1081, 2014.

[26] E. Sentana, "Quadratic arch models," *The Review of Economic Studies*, vol. 62, no. 4, pp. 639–661, 1995.

[27] W. Enders, *Applied Econometric Time Series, 4th Edition*. John Wiley & Sons Ltd, 2014.

[28] G. P. Zhang, "Time series forecasting using a hybrid arima and neural network model," *Neurocomputing*, vol. 50, pp. 159–175, 2003.

[29] T. Lin, B. G. Horne, P. Tiňo, and C. L. Giles, "Learning long-term dependencies in narx recurrent neural networks," *Neural Networks, IEEE Transactions on*, vol. 7, no. 6, pp. 1329–1338, 1996.

[30] H. T. Siegelmann, B. G. Horne, and C. L. Giles, "Computational capabilities of recurrent narx neural networks," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 27, no. 2, pp. 208–215, 1997.

[31] G. Cybenko, "Approximation by superpositions of a sigmoidal function," *Mathematics of control, signals and systems*, vol. 2, no. 4, pp. 303–314, 1989.

[32] H. Wu, "Global stability analysis of a general class of discontinuous neural networks with linear growth activation functions," *Information Sciences*, vol. 179, no. 19, pp. 3432–3441, 2009.

[33] J. Heaton, *Introduction to neural networks with Java*. Heaton Research, Inc., 2008.

[34] H. Gavin, "The levenberg-marquardt method for nonlinear least squares curve-fitting problems," 2016.

[35] E. H. Mamdani, "Application of fuzzy logic to approximate reasoning using linguistic synthesis," *IEEE transactions on computers*, vol. 100, no. 12, pp. 1182–1191, 1977.

[36] H.-J. Zimmermann, "Fuzzy mathematical programming," in *Fuzzy Sets, Decision Making, and Expert Systems*, pp. 71–124, Springer, 1987.

[37] R. Babuška and H. Verbruggen, "Constructing fuzzy models by product space clustering," in *Fuzzy model identification*, pp. 53–90, Springer, 1997.

[38] H. Roubos and M. Setnes, "Compact and transparent fuzzy models and classifiers through iterative complexity reduction," *IEEE Transactions on Fuzzy Systems*, vol. 9, no. 4, pp. 516–524, 2001.

[39] L. A. Zadeh, "Outline of a new approach to the analysis of complex systems and decision processes," *IEEE Transactions on systems, Man, and Cybernetics*, no. 1, pp. 28–44, 1973.

[40] P. C. Mahalanobis, "On the generalized distance in statistics," *Proceedings of the National Institute of Sciences (Calcutta)*, vol. 2, pp. 49–55, 1936.

[41] G. H. Golub and C. Reinsch, "Singular value decomposition and least squares solutions," *Numerische mathematik*, vol. 14, no. 5, pp. 403–420, 1970.

[42] M. C. Mackey, L. Glass, *et al.*, "Oscillation and chaos in physiological control systems," *Science*, vol. 197, no. 4300, pp. 287–289, 1977.

[43] 50Hertz Transmission GmbH and Amprion GmbH and Elia System Operator NV and TenneT TSO B.V. and TenneT TSO GmbH and TransnetBW GmbH, "Potential cross-border balancing cooperation between the belgian, dutch and german electricity transmission system operators," October 2014.

[44] "Information on the international expansion of the grid control cooperation by addition of the dutch control block," January 2012.

[45] Elia BE, "Frequency restoration process - cross-border balancing project elia-tennet," 2014.

[46] Rijksdienst voor Ondernemend Nederland, "Tabellen stand van zaken sde+ 2015," December 2015.

[47] H. Kamp, "Basisbedragen voor sde+ 2015," 2014.

[48] "Mains frequency." http://www.mainsfrequency.com/.

[49] M. Verhaegen and V. Verdult, *Filtering and system identification: a least squares approach*. Cambridge university press, 2007.

[50] "Apx day ahead auction." https://www.apxgroup.com/trading-clearing/day-ahead-auction/. Accessed: 13-01-2017.

[51] Y.-Y. Lin, J.-Y. Chang, and C.-T. Lin, "Identification and prediction of dynamic systems using an interactively recurrent self-evolving fuzzy neural network," *IEEE transactions on neural networks and learning systems*, vol. 24, no. 2, pp. 310–321, 2013.

[52] S. Yilmaz and Y. Oysal, "Fuzzy wavelet neural network models for prediction and identification of dynamical systems," *IEEE transactions on neural networks*, vol. 21, no. 10, pp. 1599–1609, 2010.

[53] C.-F. Juang, Y.-Y. Lin, and C.-C. Tu, "A recurrent self-evolving fuzzy neural network with local feedbacks and its application to dynamic system processing," *Fuzzy Sets and Systems*, vol. 161, no. 19, pp. 2552–2568, 2010.

[54] C.-F. Juang, C.-F. Lu, and Y.-W. Tsao, "A self-evolving interval type-2 fuzzy neural network for nonlinear systems identification," *IFAC Proceedings Volumes*, vol. 41, no. 2, pp. 7588–7593, 2008.

[55] Y. Chen, B. Yang, and J. Dong, "Time-series prediction using a local linear wavelet neural network," *Neurocomputing*, vol. 69, no. 4, pp. 449–465, 2006.

[56] S. Paul and S. Kumar, "Subsethood-product fuzzy neural inference system (supfunis)," *IEEE Transactions on Neural Networks*, vol. 13, no. 3, pp. 578–599, 2002.

[57] C.-F. Juang, I.-F. Chung, and C.-H. Hsu, "Automatic construction of feedforward/recurrent fuzzy systems by clustering-aided simplex particle swarm optimization," *Fuzzy sets and systems*, vol. 158, no. 18, pp. 1979–1996, 2007.

[58] M. Russo, "Genetic fuzzy learning," *IEEE transactions on evolutionary computation*, vol. 4, no. 3, pp. 259–273, 2000.

[59] C.-J. Lin, C.-H. Chen, and C.-T. Lin, "Efficient self-evolving evolutionary learning for neurofuzzy inference systems," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 6, pp. 1476–1490, 2008.

[60] S. Wu and M. J. Er, "Dynamic fuzzy neural networks-a novel approach to function approximation," *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, vol. 30, no. 2, pp. 358–364, 2000.

[61] C.-J. Lin, C.-H. Chen, and C.-T. Lin, "A hybrid of cooperative particle swarm optimization and cultural algorithm for neural fuzzy networks and its prediction applications," *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, vol. 39, no. 1, pp. 55–68, 2009.

[62] J. Kim and N. Kasabov, "Hyfis: adaptive neuro-fuzzy inference systems and their application to nonlinear dynamical systems," *Neural Networks*, vol. 12, no. 9, pp. 1301–1319, 1999.

[63] J. Soto, P. Melin, and O. Castillo, "A new approach for time series prediction using ensembles of anfis models with interval type-2 and type-1 fuzzy integrators," in *2013 IEEE Conference on Computational Intelligence for Financial Engineering & Economics (CIFEr)*, pp. 68–73, IEEE, 2013.

[64] C.-F. Juang, "A tsk-type recurrent fuzzy network for dynamic systems processing by neural network and genetic algorithms," *IEEE Transactions on Fuzzy Systems*, vol. 10, no. 2, pp. 155–170, 2002.

[65] K. B. Cho and B. H. Wang, "Radial basis function based adaptive fuzzy systems and their applications to system identification and prediction," *Fuzzy sets and systems*, vol. 83, no. 3, pp. 325–339, 1996.

[66] M. Pulido, P. Melin, and O. Castillo, "Optimization of ensemble neural networks with type-2 fuzzy response integration for predicting the mackey-glass time series," in *Nature and Biologically Inspired Computing (NaBIC), 2013 World Congress on*, pp. 16–21, IEEE, 2013.

[67] N. I. Sapankevych and R. Sankar, "Constrained motion particle swarm optimization and support vector regression for non-linear time series regression and prediction applications," in *Machine Learning and Applications (ICMLA), 2013 12th International Conference on*, vol. 2, pp. 473–477, IEEE, 2013.

[68] D. Nauck and R. Kruse, "Neuro-fuzzy systems for function approximation," *Fuzzy Sets and Systems*, vol. 101, no. 2, pp. 261–271, 1999.

[69] F. Gaxiola, P. Melin, F. Valdez, and O. Castillo, "Interval type-2 fuzzy weight adjustment for backpropagation neural networks with application in time series prediction," *Information Sciences*, vol. 260, pp. 1–14, 2014.

[70] J. M. DiCarlo and B. A. Wandell, "Rendering high dynamic range images," in *Electronic Imaging*, pp. 392–401, International Society for Optics and Photonics, 2000.

Glossary

List of Acronyms

DCSC Delft Center for Systems and Control

NARX Non-linear Auto-Regressive with eXogenous inputs

SARIMAX Seasonal Auto-Regressive Integrated Moving Average with eXogenous inputs

SARIMAX-GARCH Seasonal Auto-Regressive Integrated Moving Average with eXogenous inputs and Generalized Auto-Regressive Conditional Heteroskedasticity

SCR Secondary Control reserve Power

TSO Transmission System Operator

WSS Wide Sense Stationary

IGCC International Grid Control Cooperation

GFNARX Generalised Fuzzy Neural Network formulation of a Non-linear Auto-Regressive with eXogenous inputs model structure

ARMA Auto-Regressive Moving Average

ARIMA Auto-Regressive Integrated Moving Average

GARCH Generalized Auto-Regressive Conditional Heteroskedasticity

MA Moving Average

AR Auto-Regressive

LLS Linear Least Squares

PACF Partial AutoCorrelation Function

G-FNN Generalised Fuzzy Neural Network

ANN	Artificial Neural Network
SP	Settlement Price
REP	Real-time Electricity market Price
REPP	Real-time Electricity market Price Premium
ERR	Error Reduction Ratio
TSK	Takagi-Sugeno-Kang
CSW	Cold Storage Warehouse
COP	Coefficient of Performance
APX	Amsterdam Power Exchange
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
PTU	Program Time Unit
PCR	Primary Control Reserve
SCR	Secondary Control Reserve
TCR	Tertiary Control Reserve
KNMI	Koninklijk Nederlands Meteorologisch Instituut
BRP	Balance Responsible Party
RMSE	Root-Mean-Square Error
SDE	Stimulering Duurzame Energieproductie