THE UNIVERSITY *of* LIVERPOOL

Advanced Local Prediction and Its Applications in Power and Energy Systems

Thesis submitted in accordance with the
requirements of the University of Liverpool
for the degree of Doctor of Philosophy

in

Electrical Engineering and Electronics

by

Lei Zhu, B.Sc.(Eng.)

September 2014

Advanced Local Prediction and Its Applications in Power and Energy Systems

by

Lei Zhu

Acknowledgements

I would like to give my heartfelt thanks to my supervisors, Prof. Q.H. Wu and Dr. L. Jiang , whose encouragement, guidance and support enabled me to develop a deep understanding of my work. Without their consistent and illuminating instruction, my research work and my life could not proceed to this stage. The research skill, writing skill and presenting skill they taught me will benefit me throughout my life, as well as his inspiring insight into life philosophy and poetry.

I offer my regards and blessings to all of the members of the Intelligence Engineering and Industrial Automation Research Group, the University of Liverpool, especially to Dr. M.S. Li, Dr. W. Yao, Dr. D.Y. Shi, Dr. C.K. Zhang, and Mr. J. Chen. Special thanks also go to my friends, Y. Dong, S. Lin, X.Q. Yuan, H.C. Chang for their support and friendship. My thanks also go to the Department of Electrical Engineering and Electronics at the University of Liverpool and National Grid plc for providing the research facilities and valuable discussions that made it possible for me to carry out this research.

Last but not least, my thanks go to my beloved family for their loving considerations and great confidence in me through these years.

Abstract

Due to the global energy crisis and environmental concerns, the development of sustainable energy is considered by more and more countries. In order to make this target, energy demand management is significantly necessary in which forecasting the energy demand is the starting point. The accurate prediction of energy demand could help the energy sectors to make these operation decisions and policy properly.

A novel approach, which is the support vector regression based local predictor with false neighbor filtered (FNF-SVRLP), is proposed. This method is an amelioration of the support vector regression based local predictor (SVRLP). SVRLP is a powerful prediction method which employs phase reconstruction algorithms, such as the correlation dimension and mutual information methods used in time series analysis for data preprocessing. Compared with the global prediction method, in a local prediction method, each predicting point has its own model constructed based on its nearest neighbors (NNs) reconstructed from the time series, and the fitness of NNs would mainly affect the model performance. However, it has been found that NNs may contain a class of false neighbors (FNs) which would decrease the fitting accuracy dramatically and lead to a poorer forecasting performance. Therefore, a new false neighbor filter is proposed to remove those false neighbors and keep the optimal nearest neighbors. Then, the FNF-SVRLP is proposed.

Wind power is one of the most popular renewable energy. The increasing penetration of wind power into the electric power grid accompanied with a series of challenges. Due to the uncertain and variable nature of wind resources, the output power of wind farms is hard to control, which could lead to the instability of the power grid operation and the unreliability of electricity supplies. In order to slove this problem, the FNF-SVRLP based short-term wind power perdition model is p-

resented. Through the comparison with the SVRLP based short-term wind power perdition and ARMA based short-term wind power perdition, it is found that the FNF-SVRLP based short-term wind power perdition model is much more accurate than the others.

Due to the fact that natural gas is cleanest burning of all fossil fuel, it can be considered as an important adjunct to renewable energy sources such as wind or solar, as well as a bridge to the new energy economy. Different from the wind power, the customer consumption behavior could effect the natural gas demand. Therefore, the customer behavior based "Advanced Model" with FNF-SVRLP is presented to undertake the natural gas prediction. The proposed FNF-SVRLP natural gas model is compared with the SVRLP and autoregressive moving average (ARMA) to show its superiority. In addition, a web sever based online natural gas demand perdition system has been set up to help the National Grid to obtain the accurate daily natural gas demand perdition easily and timely.

It is found that the most kinds of energy demand data are non-stationary, the internal regularity between predicting point and its nearest-neighbors are much more complex than the stationary dataset. In order to help the local predictor to capture the internal regularity between predicting point and its nearest-neighbors more accurately, the morphological filter is proposed. the morphological filter is applied to decompose the non-stationary dataset into several subsequences, ranked form the low frequency subsequence to the high frequency subsequence. Through this way, the local predictor could capture the non-stationary dataset more accurate, and improve the final performance of prediction. The morphological filter is applied to decompose the non-stationary into several subsequences, ranked form the low frequency subsequence to the high frequency subsequence. Through this way, the local predictor could capture the non-stationary dataset more accurate, and improve the final performance of prediction.

Moveover, an novel calculation method of structure element (SE) is introduced. Different form the conventional SE, this novel approach can optimize the scale and shape of SE to match the original signal. After that, a novel algorithm, which is mathematical morphology based local prediction with support vector regression

(SVRLP-MM) is proposed. The real-world wind speed data has been used to evaluate the performance of SVRLP-MM. The final results presented demonstrate that SVRLP-MM based wind speed prediction model can achieve a higher prediction accuracy than the SVRLP based model and ARMA model based model by using the same real-world wind speed data.

Declaration

The author hereby declares that this thesis is a record of work carried out in the Department of Electrical Engineering and Electronics at the University of Liverpool during the period from October 2010 to September 2014. The thesis is original in content except where otherwise indicated.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Background of energy demand prediction

Energy is a vital input for social and economic development of any nation. With the advent of industrialization and globalization, the demand for energy has increased exponentially. Fossil fuels in the form of coal, oil and natural gas comprise 80% of the world's energy use. It is said that if the current global energy consumption pattern continues, the world energy consumption will be increased by over 50% before 2030 [1]. Energy is essential for the functioning of all activities be it a developed or developing nation. It is estimated that industrial energy usage in developing countries accounts for 45-50% of the total commercial energy consumption. Such mainly and largely energy consumption by industrial units will finally result in global energy crisis. What' worse, the global warming and air pollution are another coming issues. For example, the smog, which is a type of air pollutant, has been resulted from the emitting significant clouds of smoke by largely burning coal in industrial or in a power-producing plant. Air pollution from this source has been report in England since the Middle Ages. Currently, the air pollution of this type is still a big problem in areas that generate significant smoke from burning coal, as witnessed by the 2013 autumnal smog in Beijing and north part of China, which closed roads, schools, and the airport [2].

Therefore, due to the global energy crisis and environmental concerns, the consen-

sus of the sustainable development is shared by more and more countries nowadays. In order to make this target, nations are encouraged to apply the clean energy and develop the renewable energy. Natural gas, one of clean but nonrenewable energy, plays a key role in many countries' energy mix. According to HI-Energy [3] and government report of UK [4], UKs 46% of total electricity (around 350,000 GWh) generation and approximately 70% of domestic heating in 2008 is produced by natural gas. However, natural gas is an non-renewable energy source, which has limited reserve on the planet and contributes to global warming. Thus, planning energy demand is becoming an important issue in the energy sector and forecasting its demand is the starting point, especially for natural gas demand. Natural gas demand prediction has been a very commonly addressed problem in nowadays.

On the energy generation side, wind power will be one of the most promising renewable energy. Wind power is a clean and inexhaustible energy, is safer and less expensive to exploit in comparison with other renewable energy sources. According to the European Wind Energy Association [5][6](EWEA), by the end of 2020, wind energy will meet 14% of the EUs total electricity consumption and achieve 30% lower emissions. However, the increasing penetration of wind power into the electric power grid accompanied with a series of challenges. Due to the uncertain and variable nature of wind resources, the output power of wind farms is hard to be controlled, which could lead to the instability of the power grid operation and the unreliability of electricity supplies. Therefore, to maintain a balanced generation and demand, wind power generation prediction i.e., wind speed prediction, is crucially important to the power grid operation. There are numerous literatures on natural gas demand prediction (NGDP), wind speed prediction (WSP) and wind power prediction (WPP). However, the NGDP, WSP and WPP are still very important research issues in energy demand prediction and management area. Due the complex affecting factors, big data, and some other internal relationships, accurate estimating the future gas demand or wind power generation with the historical data has remained a difficulty up to now.

1.2 Objectives

Considering the complexity of the historical data and the uncertainty of the influencing factors such as weather and environments, global methods, on which all data of the whole time series are involved in modeling experiments, have been developed with phase space reconstruction and proposed to capture complicated historical data characteristics. They can exhibit a good performance in analysing and predicting the short term evolution in phase spaces reconstructed from analysis of the original time series.

The support vector regression based local prediction (SVRLP) method is developed by the University of Liverpool, 1997. The SVRLP method combines the powerful regression method SVR with local predictor. According to the our prior work [7], local prediction methods based on phase reconstruction normally perform better than global methods based on phase reconstruction. With a local prediction method, each predicting point has its own model constructed based on its nearest reconstructed from the time series, and the fitness of nearest neighbors (NNs) would mainly affect the model performance. However, it has been found that NNs may contain a class of false neighbors (FNs) which would decrease the fitting accuracy dramatically and lead to a poorer forecasting performance. This means that not all NNs are suitable for use in local prediction and some of them should be filtered [8]. One of the objectives of this research work is to overcome the problems in the local predictor. Therefore, a false neighbors filter which combine the Euclidean distance and the exponential separation rate is introduced to overcome the drawback.

With the purpose of the sustainable development and enhance the efficiency and stability, energy demand prediction, especially for natural gas demand and wind power generation is very important. Applying the proposed method to predict the real-world natural gas demand and real-world wind power(speed) are the objective of this work as well.

Local modeling is sensitive to the internal regularity between targeted point and its nearest-neighbors. The datasets, such as wind power, wind speed and natural gas demand, which are non-stationary and complex, the internal regularity between predicting point and its nearest-neighbors are much more complex than the station-

ary dataset. In order to improve the identifiable accuracy of the regularity of local predictor, and further to improve the final accuracy of prediction performance, the idea of decomposing the original non-stationary data into several subsequences is applied. In order to decomposing original data into subsequences with different frequency (ranked from low frequency to high frequency), the optimal filter should be decided. The most conventional filters are based on the wavelet, however, the mathematical morphology, which is powerful tool for image/signal processing, is not applied in this area. Compared with the Wavelet, the mathematical morphology is much easier to process and to calculate. Therefore, another objective of this work is to develop a mathematical morphology based filter to precess the decomposition of the original non-stationary data, and improve the accuracy of the final prediction results.

1.3 Thesis outline

The thesis is structured as follows:

Chapter 2 This chapter gives a literature survey for energy demand forecasting problem. The factors affecting the energy demand are introduced firstly, followed by description of the various forecasting methods.

Chapter 3 In this chapter, the method of Support Vector Regression-Based Local Prediction with False Neighbors Filtered is introduced. Firstly, the basic of concept of SVR is given, together with the limitations of original local prediction. In order to overcome the drawback of the original local prediction, the false neighbors filter is presented. Then combined the optimal nearest neighbors (NNs) and the powerful SVR to organize the novel approach which is named as support vector regression-based local prediction false neighbors filtered (FNF-SVRLP).

Chapter 4 The basic idea of phase-space reconstruction of the time series is introduced. In order to reconstruct phase-space of the time series, the embedding

dimension and the time delay constant must be estimated from the time series. A method is proposed for the estimation of the embedding dimension and time delay constant is introduced. On the other side, the wind power prediction has been received much attention with the fast development of the renewable energy. In this chapter, the FNF-SVRLP is applied to undertake short-term wind power prediction. The proposed predication method not only combines the powerful SVR with the reconstruction properties of time series, but also overcomes the drawback of the original local predictor by removing false neighbours. The proposed method (FNF-SVRLP) is evaluated with the real world wind power data, and the final performance is compared with the support vector regression based local predictor (SVRLP) and the autoregressive moving average (ARMA). The results demonstrate that the proposed method can achieve a better performance than the other method.

Chapter 5 In this chapter, the FNF-SVRLP is applied to perform short-term prediction of natural gas demand. During the natural gas demand prediction, the accuracy-based method is introduced to estimate the embedding and delay. At first, one unified model, named "Standard Model (SM) is presented to process the whole daily dataset. Then, in order to make further improvement of prediction accuracy, the customer behavior difference within one week is considered and three models are constructed based on datasets of Monday, Saturday and other days of the week, and results in a customer behavior based "Advance Model (AM). The developed FNF-SVRLP based AM has been applied successfully to predict natural gas demand for National Grid, United Kingdom (UK). It outperforms the SVRLP and the autoregressive moving average (ARMA) methods using the real data from National Grid. An web-based online server has been used to implement the algorithm developed, which can process the real data for National Grid on a daily basis. The configuration and interface of this online server have been described briefly. The overall performance of the online system outperform the National Grid's prediction model.

Chapter 6 This chapter investigates a novel prediction method, called the Mathematical Morphology-based Local Predictor with Support Vector Regression (SVRLP-MM) and its application of wind speed predictions. The basic concept about the mathematical morphology is introduced firstly. The decomposition schemes are discussed in detail. In addition, the optimal structural element (SE) calculation method is presented, through this method the SE can be modified according to different input dataset. Then, the implementation of the proposed SVRLP-MM is described. The test case with real wind speed is presented to show the effectiveness of the proposed SVRLP-MM method by comparing its performance with other methods.

Chapter 7 Finally, the thesis in concluded in this chapter and suggestions for future work are given.

1.4 Major contributions

The major contributions of this research include devolvement of advanced local prediction methods and their applications to energy demand prediction, in particular for natural gas demand predication and wind power prediction. Moreover, a web server based online prediction system has been implemented based on the algorithm proposed and contribute successfully to the daily operation of National Grid. The contributions are summarised as follows.

- An ameliorated local predictor is proposed. The main feature of the local prediction is the fitness of NNs, which would mainly effect the final performance of prediction. However, we has found that NNs may contain a class of false neighbors (FNs) which would decrease the fitting accuracy dramatically and cause the reduction of the forecasting performance. Therefore, the proposed false neighbors filter is applied to remove those false neighbors and leave the optimal nearest neighbors. Then, through the combination of the optimal NNs and the powerful SVR, the novel approach which named as support vector regression-based local prediction false neighbors filtered (FNF-SVRLP) is presented.

- The FNF-SVRLP method is applied to short-term wind power prediction with success. During the computation procedure of the WPP, the embedding dimension and the time delay are computed firstly, then the phase space reconstruction is applied to the wind power data. In the local prediction, there are some difference between the SVRLP and the proposed FNF-SVRLP. Compared with the SVRLP, the proposed method not only applies the Euclidean distance but also the exponential separation rate to modify nearest neighbors, and only the optimal nearest neighbors can be selected as training samples. The final results presented demonstrate that the proposed model can achieve a higher prediction accuracy than the SVRLP model and ARMA model using the same real world wind power data.

- The FNF-SVRLP based Advanced Model for short-term gas demand prediction is proposed. The forecasted gas demand can aid the market in making efficient decisions in balancing supply and demand, and reducing costs. The accuracy-based method is proposed to estimate the value of embedding dimension and the time delay, then the phase space reconstruction is applied to process the gas dataset. The final results presented demonstrate that the FNF-SVRLP can achieve a higher prediction accuracy than that of the SVRLP method and ARMA method using the same real world gas data. Moreover, the study on the customer behavior of different day has been tested, and results indicate that the customer behavior can affect the final forecasted performance.

- The web server based online natural gas demand prediction system has been implemented. This online system is based on the FNF-SVRLP method, and built with the friendly online interface. This online system is built for the daily operation of the National Grid. The engineers of the National Grid can access the online system via internet, and the prediction data obtained from the online system has been used in the daily operation and management of the National Grid.

- A new approach which is SVRLP-MM is presented The proposed approach combines a powerful regression algorithm which is SVR with a morphologi-

cal filter based local predictor. We applied the morphological filter to decompose the original wind speed time series into several subsequences, the subsequences contains "noise" parts and "baseline" part, then applied the SVRLP to predict each subsequence separately. The proposed approach is applied to the wind speed prediction, and the final results presented demonstrate that the proposed model can achieve a higher prediction accuracy than the SVR model, SVRLP model and ARMA model using the same real world wind speed data.

- An optimal SE calculation method is presented. By using optimal SE calculation method, the value and shape of the SE would be modified according to different dataset. The SE is considered as one of the key component of the morphology analysis. Generally, only when the scale and shape of the signal are matched to those of SE, the signal can be reserved well. SE should approach morphological features of the signal as far as possible.

1.5 Auto-bibliography

1. L. Zhu, M.S. Li, Q. H. Wu, L. Jiang, J. S. Smith. Support Vector Regression-Based Short-Term Wind Power Prediction With False Neighbors Filtered, International Conference on Renewable Energy Research and Applications, Madrid, Spain, 2013, pp. 740-744.

2. L. Zhu, Q. H. Wu, L. Jiang. Mathematical Morphology-based Short-term Wind Speed Forecasting Using Support Vector Regression, The 5th IEEE PES Innovative Smart Grid Technologies Europe (ISGT Europe); Istanbul, Turkey, 2014, pp. 1-6.

3. L. Zhu, M. S. Li, Q. H. Wu, L. Jiang. Short-term Natural Gas Demand Prediction Based On Support Vector Regression With False Neighbors Filtered, *Energy*, vol. 80, pp. 428-436, 2015.

4. L. Zhu, M. S. Li, Q. H. Wu, L. Jiang. Short-term Wind Speed Prediction Based On Morphological Support Vector Regression, *Applied Energy*(Submitted)

Chapter 2

Literature Review and Background of Energy Demand Prediction

This chapter describes the main factors affecting the energy demand and reviews the current popular prediction methods in energy demand prediction areas.

2.1 Factors affecting the energy demand

Generally, to all kinds of energy demand, they could be affected by many different kinds of factors. Those factors can be generally classified as weather, calendar,economical, and random factors. The effects of these factors are listed as follows to provide a basic understanding of the characteristic of energy demand , which would be helpful to find the hidden relationship between those factors and the final energy demand, reduce the noise in the dataset and eliminate the unnecessary input dataset.

- Weather: These factors include temperature, humidity, cloud cover, light intensity and so on. Different kinds of weather will come with different natural affect. Taken the London average wind speed as an instance, the average wind speed during the winter period (December to February) is about 5.6 m/s, and during the summer period (June to August), the average wind speed reduced to 3.5 m/s. What' more, the change of the weather causes the change

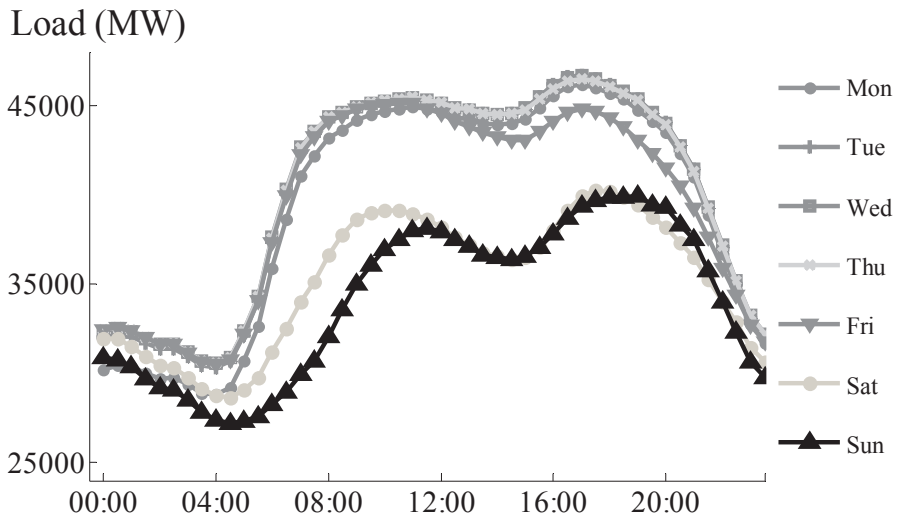


Figure 2.1: Average intra-day profile for each day of the weekl

of consumers comfort feeling and in turn the usage of some appliances such as gas heaters and air conditioners. To the most weather-sensitive energy demand, for example natural gas demand, its demand was highest during the coldest months of winter and lowest during the warmest months of summer. The main driver for this primary cycle of natural gas demand is the need for residential and commercial heating. As expected, heating requirements are highest during the coldest months and lowest during the warmest months.

- Calendar: The calendar factors contains the day of the week, season, sunrise/sunset, etc. In other words, the calendar factors can be partly considered as the customer behavior. For example, Fig. 2.1 presents the average load cycle for each day of UK form 1^{st} Jan. 2001. to 31^{st} Dec. 2009. From Fig. 2.1, it can bee seen that the load is clearly much lower at the weekends than on the weekdays. The load profile for Saturdays differs from Sundays, and for the weekdays, there are some differences, most notably on Mondays and Friday afternoons.

- Economic: The economics factor play an important role to the many kinds of energy demand. For example, to the natural gas, some countries like UK,

which is highly depend on import natural gas from Norway , the price of natural gas in Norway will effect the import demand of UK. To the load part, the relationship between electricity price and load profile is even stronger with the development of modern electricity markets [9, 10]. Although time of use pricing and demand-side management had arrived before deregulation, the volatility of spot markets and incentives for consumer to adjust loads are potentially of a much greater magnitude. At low price, elasticity is still negligible, but at times of extreme conditions, price-induced rationing is a much more likely scenario in a deregulated market compared to that under central planning.

- Other random factors: The modern energy system is composed of numerous users. The startup and shutdown of the large consumers, such as steel mill, which is an important source of random disturbance and always lead to an obvious impulse to the load curve. The startup and shutdown time of these users is quite random, i.e., there is no obvious rule of when and how they get power from the grid. In addition, special events are another source of random disturbance. A typical special event is, for example, a large social event in the winter, the indoor heating which supplied by the natural gas, would lead the natural gas demand increase dramatically.

2.2 Classification of conventional energy demand prediction method

Energy demand prediction is helpful to the effective utilization of the energy resources, reliability in supply and energy consumption. It can benefit the combined heat and power systems, renewable energy systems, integrated energy systems and independent power delivery systems. Cost effective options, commercially viable alternatives and environmental friendly solutions are desired to be explored. An accurate demand prediction plays a key role on the planning, implementing, and monitoring activities of energy utilization that and also to encouraging consumers

to modify their level and pattern of energy usage. To every nations, energy demand is found to be closely linked to energy price and GDP. Energy models are developed using macro economic variables to forecast the energy demand. This helps in planning and drafting policies for energy management on the demand side. Energy demand management can help in achieving self sufficiency and cost effectiveness and provide for a sustainable economic development. Energy demand prediction should thus help in

- planning for the future requirement, identifying conservation measures
- identification and prioritization of energy resources, optimized energy utilization, strategies for energy efficiency improvements
- framing policy decisions
- identification of strategies for reduced emission

2.2.1 Time series method

Time series models are the most simplest of models which uses time series trend analysis for extrapolating the future energy requirement. Times series models have been used for decades in different field such as digital signal precessing, economics, as well as energy demand prediction. The autoregressive moving average (ARMA), autoregressive integrated moving average (ARIMA) are both the traditional examples of the time series methods. Ergin and Shi employed four competing approaches based on the ARMA method for forecasting of wind speed as well as wind direction and strived to determine the best performing model while marking comparisons among them according to the mean absolute error (MAE) [11]. Erdogdu focused on the characteristics of demand and estimated short and long-run price and income elasticities of sectoral natural gas demand in Turkey and forecasted future growth in this demand using an ARIMA modeling, comparing these results with official projections [12]. Bargur and Mandel have examined the energy consumption and economic growth using trend analysis for Israel [13]. Gonzales et al. have forecast

energy production and consumption in Asturias-Northern Spain [14]. A semi statistical cyclic pattern analysis was used for forecasting the primary energy demand for Turkey. The results are found to be similar to Winter's exponential smoothing technique [15]. Hunt et al. investigated the energy demand in sectoral basis for the UK using time series approach [16]. Three time series models, namely, Grey-Markov model, Grey-Model with rolling mechanism, and singular spectrum analysis (SSA) are used to forecast the consumption of conventional energy in India. Grey-Markov model has been employed to forecast crude-petroleum consumption while Grey-Model with rolling mechanism to forecast coal, electricity (in utilities) consumption and SSA to predict natural gas consumption [17]. Sourcewise analysis is also carried out to determine the future demand. The consumption of oil and price is forecast under three scenarios: Parabolic, linear and chaotic behavior [18]. Aras and Aras [19] used the first-order autoregressive time-series model to predict the natural gas requirement for Eskisehir.

A time-series-based decision support system that integrates data management, model base management, simulation, graphic display, and statistical analysis to provide near-optimal forecasting models for electricity peak load forecasting in UAE is developed. The model base includes a variety of time-series techniques, such as exponential smoothing, BoxCJenkins (BJ), and dynamic regression [20].

Gonzalez-Romera et al. [21] used trend extraction method to examine the electric energy consumption for Spain. In the field of interval time-series (ITS) forecasting, different techniques have been developed. Arroyo et al. [22]have developed three exponential smoothing methods for ITS forecasting.

Himanshu and Lester [23] have used time series analysis for predicting electricity demand in Sri Lanka. Electrical power requirement for Jordon is predicted using models that account for trend, monthly, seasonal and cyclic dynamics [24]. Amarawickrama and Hunt [25] have presented a time series analysis of electricity demand in Sri Lanka. Various time series estimation methods were used to analyse using past electricity consumption. They have used income and price elasticities to predict the future electricity consumption in Sri Lanka.

2.2.2 Regression method

Energy forecasts are very important in the framing of energy of environment policies. Regression models have been used to forecast the coal, oil, gas, electricity requirement [26]. ONeill and Desai [27] analyse the accuracy in the projections of US energy consumption presented by Energy Information Administration (EIA). GDP and energy intensity (EI) are used in the projection of energy requirement. It is found that the GDP projections are consistently too high while EI projections are consistently too low. This tends to underestimate the future energy consumption. Linear and nonlinear effect of energy consumption on economic growth for Taiwan is examined by Lee and Chang [28]. It is found that a threshold regression provides a better empirical model than the standard linear model.

Jannuzzi and Schipper [29] have examined the of electrical energy consumption for the residential sector in Brazil. It was found that the increase in electricity demand was faster than the income. Dynamic relationship between electricity consumption and weather, price, and consumer income are examined by Harris and Lon-Mu [30] using 30 years data series from south east USA. Electricity demand based on the intensity of consumption is developed [31] and [32] to predict the future requirement.

The influence of economic variables on the annual electricity consumption in N. Cyprus is examined [33]. Using multiple regression analyses, the relationship between energy consumption, the number of customers, the price of electricity and the number of tourists is determined. A linear regression model was used [34] to predict the electricity consumption for Turkey based on the population and percapita consumption rates. Tunc et al. [35] used the regression analysis to predict Turkey's electric energy consumption.

Bessec and Fouquau [36] have examined the non linear relationship between electricity demand and temperature in the European Union. A panel threshold regression with exponential and logistic functions is considered for the data collected from 15 European countries. An empirical model based on multivariate regression is developed [37] to predict the electricity requirement of Jordon's industrial sector. Industrial production outputs and capacity utilization were found to be two most

important variables that affect electrical power demand. The residential and commercial sector electricity consumption pattern in Hong Kong was examined [38]. Principal component analysis of five major climatic variablesłdry-bulb temperature, wet-bulb temperature, global solar radiation, clearness index and wind speedłwas conducted. It was found that sector-wide electricity consumption correlated with the corresponding two principal components determined using multiple regression technique.

A non parametric regression model [39] is used to assess the wind energy forecasts. The conditional price distribution is found to be non Gaussian. The forecasting models for electricity spot prices for which parameters are estimated by a least squares technique will not have Gaussian residuals.

2.2.3 Econometric method

Econometric models correlate the energy demand with other macro-economic variables. Samouilidis and Mitropoulos [40] have researched energy and economic growth in industrialized countries. Econometric models are developed to forecast energy consumption as a function of Gross National Product (GNP), energy price, technology, population for India [41]. Ramaprasad Sengupta [42] and Rao and Parikh [43] have been established that such models are effective in forecasting energy patterns in developing countries. Berndt and Watkins attempted to estimate an econometrics model Arsenault et al. [44] have predicted the total energy demand as a function of previous year's energy demand, price of energy, real income and heating day for the province of Quebec. Ordinary least square technique (OLS) is used and prediction is made sectorwise C residential, commercial, industrial and street lighting. Yearly data has been used for demand side projection. Energy forecast is influenced by weather conditions data.

Energy supply and demand for the Asia-Pacific region is analysed [45]. The demand is forecast for three scenarios C high, low, base case considering variations in economic performance, prices and fuel substitution at the national and regional level. Four factors are considered for each country C econometric factors (GDP, foreign trade) with oil prices, domestic oil prices, substitution. A bottom up country by

country approach is followed. Oil, natural gas, coal and electricity requirements are projected. The effect of price elasticities, income elasticities and technical efficiency on residential energy demand is studied for OECD countries using econometric energy models [46]. The energy requirement and CO2 emission for Greece is forecast using econometric models. Demand equations are derived for each sector of economic activity traded, non-traded, public and agricultural sector and for each type of energy C oil, electricity and solid fuels. The energy system is integrated so that all interactions between energy, prices and production factors are considered [47].

Sharma et al. [48] analysed the requirement of three major forms of commercial energy in the state of Kerala (viz electricity, petroleum products and coal). Sectorwise/productwise econometric demand models are generated using regression method. ZhiDong [49] has conducted an econometric study for China linking energy, economy and the environment. A three equation model [50] is used for energy modelling and forecasting energy demand in UK and Germany. An economic model considers the price of electricity, oil, gas, coal, total energy demand and technological progress. The statistical model has the economic model embedded in its equation along with the error correction term. The results from the two models are then processes for structural change and stability.

Energy consumption in industrial, transportation, residential and commercial is determined for China using the consumption of fuel in a sector taking the case of a well off society [51]. Sectoral energy related parameters are identified to determine the final energy consumption in the sector. Econometric modelling is used for energy forecasting. Rural, social and economic data is collected for six provinces in China [52]. A sectoral energy demand analysis and a forecasting model are developed. Variables such as GDP, per capita income, agricultural production output, industrial production output, capital investment are used.

A modified form of econometric model EDM (Energy Demand Model) is used by Gori and Takanen [53] to forecast the Italian energy consumption. The possible substitution of various energy resources is investigated. In addition, the long term electricity consumption pattern in Italy is examined using cointegration and stationary time series models. The primary energy demand in Japan is determined

by exploring the relationship between energy demand, GNP and real energy price [54]. The resulting econometric model is used to determine long run price elasticity and income elasticity. The model is utilized to forecast the energy consumption and CO2 emission.

Raghuvanshi et al. [55] determine the characteristics of the drivers of energy development for India. The primary energy consumption is decomposed as a product of three variables, population, per capita GDP and energy intensity of GDP. Similarly the CO2 emissions are decomposed as the product of the primary energy consumption and the carbon intensity of primary supply. Ramanathan [56] has used data envelopment analysis to analyse the patterns of efficiency in terms of world energy consumption, Gross Domestic Product (GDP) growth and CO2 emissions. The impacts of the changes in energy prices due to deregulation of prices is examined [57] on aggregate energy intensity and coal/oil/electricity intensity is studied. Price elasticities by energy type are determined.

The levels and types of demands for energy services in 2040 for Australia are determined by projecting the levels of economic activity [58]. Demand for 2040 is estimated by examining how energy intensity has been changing in each sector in recent years and this is used to project the future energy requirement. The changes in energy price elasticity and elasticities of substitution are examined [59] between energy and non-energy (capital and labour) sectors in China. It is found that accelerated market oriented reforms lead to energy efficiency improvements because the energy price elasticity declines, and elasticities of substitution and cross price elasticities between energy, capital and labour rise. An econometric model is developed to predict China's energy demand [60]. The energy requirement is forecast and an energy balance is presented for 2020 for China.

2.2.4 Intelligent algorithms

In recent times, thee intelligent algorithms have been widely used for energy demand projections. Neural network is used to model the energy consumption of appliances, lighting, and space-cooling in Canadian residential sector [61]. The energy consumption for Turkey is predicted using artificial neural-network (ANN)

technique [62]. Two models are used: population, gross generation, installed capacity and years are used in the input layer of the network for Model 1 and other energy sources are used in input layer of network for Model 2.

Gorucu and Gumrah [63] have used ANN to predict the gas requirement for Ankara. GNP, population and vehicle kilometre are used as input parameters in training neural network model for predicting the transport energy demand for Turkey [64]. The best network architecture is selected using the training and validation data set. The final network is tested using the test data. The transport energy consumption in Thailand is determined using the national gross domestic product, population and the numbers of registered vehicles as independent variables [203]. Log-linear regression models and feed-forward neural network models are used in the study. Brown et al.[121] and Brown and Matin [122] used ANN to predict natural gas demand and demonstrated a better accuracy than the model with liner regression. Khotanzad and Elragal [123] and Khotanzad et al. [124] combined ANN with different methods and proposed a two-stage prediction models, which showed that different combinations of ANN models can improve prediction accuracy. In Serbia, Ivezić [125] applied the daily gas demand and daily minimal and maximal temperature as inputs to do the short-term NGPD with ANN. Azadeh et al. [126] applied an ANN-based algorithm, which is ANFIS, to predict natural gas demand in the Iranian network. They applied the same day gas demand in the previous year as an additional input with other conventional inputs.

Researchers have argued that green energy can be considered as a catalyst for energy security, sustainable development, and social, technological, industrial and economic development. In paper [72], it analysis the world green energy consumption through artificial neural networks (ANN). The world primary energy consumption including fossil fuels such as coal, oil and natural gas is also considered.

Arcaklioglu [73] used three different models to train the ANN. In Model 1, energy indicators such as installed capacity, generation, energy import and energy export, in Model 2, GNP was used and in the Model 3, GDP was used as the input in ANN. The output of the network is net energy consumption (NEC). It is found that the ANN approach presents greater accuracy when economic indicators namely

GNP, GDP are used for prediction. The energy demand for South Korea is estimated using a feed forward multilayer perception, error back propagation algorithm [74]. The model considered gross domestic product, population, import and export. The results are compared with the multiple linear and exponential regression energy demand models.

Pao [75] examines the following linear models: the exponential smoothing model (Winters), the exponential form of the generalized autoregressive conditional heteroscedasticity (EGARCH) and seasonal EGARCH (SEGARCH) models, the combined Winters with volatility EGARCH model (WARCH) and ANN non linear model. Based on the above models, two hybrid non linear models SEGARCH C ANN and WARCH C ANN are developed to predict Taiwan's consumption of electricity and petroleum. The models are validated using root mean square error (RMSE), mean absolute error (MAE) and mean absolute percentage error (MAPE). The results indicate the hybrid models give better accuracy and among the hybrid models WARCH-ANN is the better model.

The extent to which an economy relies upon imports in order to meet its energy needs is defined as Energy dependency (ED). Turkey's energy dependency is determined using ANN based on basic energy indicators and sectoral energy consumption [76]. Two models have been used. In Model 1, main energy indicators such as total production of primary energy per capita, total gross electricity generation per capita and final energy consumption per capita were used in the input layer of the ANN while sectoral energy consumption per capita was used in Model 2. A global optimization method called Modal Trimming Method is used to identify the values of model parameters [77]. In addition, the trend and periodic change are removed from time series data on energy demand. The converted data is used as the main input to a neural network. Furthermore, predicted values of air temperature and relative humidity are considered as additional inputs to the neural network, and their effect on the prediction of energy demand is investigated.

The Greek long-term energy consumption is predicted using ANN multilayer perception model. The input variables chosen are yearly ambient temperature, installed power capacity, yearly per resident electricity, consumption, gross domestic

product [78]. Energy consumption in Turkey is modelled based on socio-economic and demographic variables (gross domestic product-GDP, population, import and export amounts, and employment) using artificial neural network (ANN) and regression analyses. The models are validated using relative errors and RMSE [79].

Cadenas and Rivera applied ANN to the hourly wind speed time series and aimed to enhance prediction accuracy developing a model for each month of the year [80]. Then, a comprehensive study was presented by investigating the performances of three different ANN types, namely, Adaptive Linear Element (ADALINE), Feed Forward Back-Propagation (FFBP) and Radial Basis Function (RBF) for 1-h-ahead wind speed predictions and as a result, the authors indicated that different structures and model parameters can yield different forecast accuracies for the same wind data in terms of various evaluation criteria [81].

Chapter 3

Support Vector Regression-Based Local Prediction with False Neighbors Filtered

3.1 Introduction

Support Vector Regression is a powerful methodology which derived from Support Vector Machines (SVMs). It is based upon statistical machine learning and can be utilized to achieve nonlinear mapping from sample space to feature space through kernel functions. Moreover, SVR replaces the empirical risk minimization which is generally employed in the classical methods such as ANNs, with a more advantageous structural risk minimization principle. Therefore, SVR can achieve an outperforming of fitting accuracy for chaotic time series prediction [7].

Local prediction methods can normally perform better than global methods for chaotic time series prediction [7]. In local prediction methods, each predicting point has its own model constructed by its nearest neighbors (NNs) which were found in the neighborhood of the phase space reconstructed from the time series, and the fitness of NNs would mainly effect the final performance of prediction. However, it has been found that NNs may contain a class of false neighbors (FNs) which would decrease the fitting accuracy dramatically and cause the reduction of the forecasting

performance, so that not all of NNs are suitable for using in the local prediction, and these FNs should be filtered [8].
In this chapter, a false-neighbors filtered local predictor based on a proven powerful algorithm, which is the SVR combined with space reconstruction of time series, is introduced. The resulting predictor is referred as SVR based local prediction with false neighbors filtered (FNF-SVRLP).

3.2 Support Vector Regression

3.2.1 The foundation theory of SVR

Suppose there is a set of training data $\{x_i, y_i\}_{i=1}^{N}$ where each $x_i \in \Re^d$ denotes the inputs space of the sample and has a corresponding target value $y_i \in \Re$ for $i = 1, ..., N$, where N corresponds to the size of the training data. SVR looks for an approximation function $f(x)$ that has at most ε from the targets for all the training data and is as flat as possible for good generalization. This means that, we do not care about errors as long as they are less than the ε deviation.
Let the function $f(x)$ is linear as follows:

$$f(x) = \langle \omega, x \rangle + b \tag{3.2.1}$$

where $\langle ., . \rangle$ denotes the dot product, ω contains the coefficients that have to be estimated from the data and b is real constant.
The flatness of the function (3.2.1) implies seeking a small ω, through minimising the square norm $\|\omega\|^2$. Minimsing $\|\omega\|^2 = \langle \omega, x \rangle$ is equivalent to maximising the distance between the data point and the approximation function[1]. The coefficients ω and b can thus be estimated by minimising the regularized risk function [82].

$$R_{SVR} = R_{emp} + \frac{1}{2}\|\omega\|^2 = \frac{C}{N}\sum_{i=1}^{N} L_\varepsilon\left(y_i, f(x_i)\right) + \frac{1}{2}\|\omega\|^2 \tag{3.2.2}$$

where R_{SVR} and R_{emp} represent the regression and empirical risks, respectively while $L_\varepsilon\left(y_i, f(x_i)\right)$ is ε insensitive loss function. C is the regularization constant which determines the trade-off between the flatness of f and its accuracy in capturing the training data.

In the regularized risk function given by equation (3.2.2), the regression risk (test set error), R_SVR, is the possible error committed by the function f in predicting the output corresponding to a new (test) example input vector. In equation (3.2.2), the first term $\frac{C}{N}\sum_{i=1}^{N} L_\varepsilon\left(y_i, f(x_i)\right)$ denotes the empirical error (termed "training set error"), which is estimated by the the ε insensitive loss function. The second item, $\frac{1}{2}\|\omega\|^2$, is the regularization term.

To estimate ω and b, equation (3.2.2) is converted to the primal function by introducing slack variables ξ_i,ξ_i^*. Hence we have the following optimisation problem for SVR [83]:

$$\min_{\omega,b,\xi_i,\xi_i^*} \frac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) \tag{3.2.3}$$

$$\begin{cases} y_i - \langle\omega, x_i\rangle - b \le \varepsilon + \xi_i^* \\ \langle\omega, x_i\rangle + b - y_i \le \varepsilon + \xi_i^* \\ \xi_i, \xi_i^* \ge 0 \end{cases}$$

where $(\xi)_i$ is the lower training error ($\xi_i{}^*$ is the upper) subject to the ε-insensitive tube. SVR avoid under-fitting and over-fitting of the training data by minimising the regularisation term $\frac{1}{2}\|\omega\|^2$ as well as the training error $C\sum_{i=1}^{N}(\xi_i + \xi_i^*)$. The ξ-insensitive loss function which introduction by Vapnik [83] enforces the distance between the approximation function and examples no more that ξ. This loss function is shown in following figure in which the slope is determined by C and defined as follows:

$$|\xi|_\varepsilon = \begin{cases} 0 \ \text{ if } \ |\xi| < \varepsilon \\ |\xi| - \varepsilon \ \text{ otherwise} \end{cases} \tag{3.2.4}$$

3.2.2 Primal and dual optimisation

The Lagrange multipliers technique can be applied to solve equation (3.2.3) as following [84]:

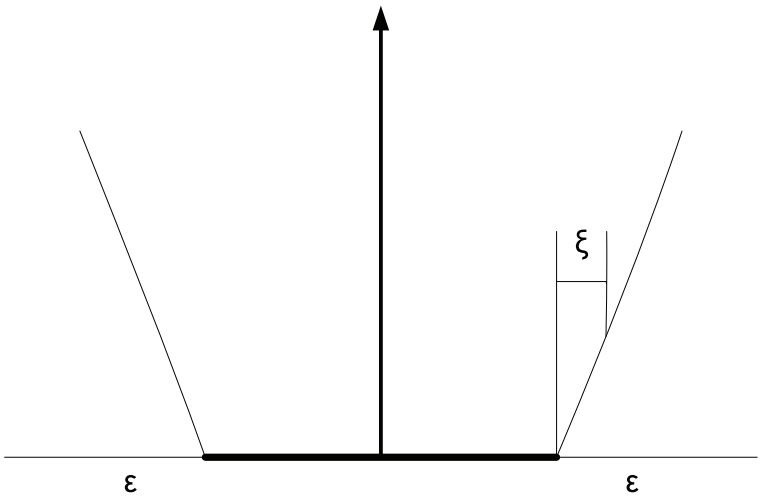


Figure 3.1: The ξ-insensitive loss function

$$\begin{aligned} L = \tfrac{1}{2}\|\omega\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*) - \sum_{i=1}^{N}\alpha_i(\varepsilon + \xi_i - y_i + \langle\omega, x_i\rangle + b) \\ - \sum_{i=1}^{N}\alpha_i^*(\varepsilon + \xi_i^* - y_i + \langle\omega, x_i\rangle - b) - \sum_{i=1}^{N}(\eta_i\xi_i + \eta_i^*\xi_i^*) \end{aligned} \tag{3.2.5}$$

Here L is the Lagrangian and the feature $\alpha_i, \alpha_i^*, \eta_i$ and η_i^* are Lagrange multiplier with:

$$\alpha_i, \alpha_i^*, \eta_i, \eta_i^* \geq 0 \tag{3.2.6}$$

Equation (3.2.5) is known as the primal objective function. The solution of this function is obtained by solving the dual objective function when the gradient of L with respect to ω, b, ξ_i and ξ_i^* is equal to 0, therefore, we have

$$\frac{\partial L}{\partial b} = \sum_{i=1}^{N}(\alpha_i^* - \alpha_i) = 0 \tag{3.2.7}$$

$$\frac{\partial L}{\partial \omega} = \omega - \sum_{i=1}^{N}(a_i - a_i^*)x_i = 0 \tag{3.2.8}$$

$$\frac{\partial L}{\partial \xi_i} = C - \alpha_i - \eta_i = 0 \tag{3.2.9}$$

$$\frac{\partial L}{\partial \xi_i^*} = C - \alpha_i^* - \eta_i^* = 0 \tag{3.2.10}$$

By substituting equations (3.2.7)-(3.2.10) into (3.2.5), the dual optimisation function can be obtained as follows :

$$\max_{\alpha_i, \alpha_i^*} \begin{cases} -\frac{1}{2} \sum_{i,j=1}^{N} (\alpha_i - \alpha_i^*)(\alpha_j - \alpha_j^*) \langle x_i, x_j \rangle \\ -\varepsilon \sum_{i=1}^{N} (\alpha_j + \alpha_j^*) + \sum_{i=1}^{l} y_i (\alpha_j - \alpha_j^*) \end{cases} \tag{3.2.11}$$

$$\text{subject to} \begin{cases} \sum_{i=1}^{N} (\alpha_j - \alpha_j^*) = 0 \\ \alpha_j, \alpha_j^* \in [0, C] \end{cases}$$

In deriving equation (3.2.11) the dual variables η_i, η_i^* were eliminated through conditions (3.2.9) and (3.2.10) ($\eta_i = C - \alpha_j$ and $\eta_i^* = C - \alpha_j^*$), as these variables did not appear in the dual objective function anymore but only were presented in the dual feasibility conditions. Therefore, the support vector expansion comes form equation (3.2.8) which can be rewritten as follows [85]:

$$\omega = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) x_i \ \text{ thus } \ \hat{f}(x) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \tag{3.2.12}$$

In order to calculate b, the KarushoKuhn-Tucker (KKT) conditions can be used [86][87] . These condition will be stated in the SVR optimisation subsection. Allowing inequality constraints, the KKT approach to nonlinear programming generalises the method of Lagrange multipliers, which have allowed only equality constrains. The KKT conditions state that the product between dual variables and constraints has to vanish at the optimal solution.

$$\alpha_i(\varepsilon + \xi_i - y_i + \langle \omega, x_i \rangle + b) = 0 \tag{3.2.13}$$

$$\alpha_i^*(\varepsilon + \xi_i^* - y_i + \langle \omega, x_i \rangle + b) = 0 \tag{3.2.14}$$

and

$$(C - \alpha_i)\xi_i = 0 \tag{3.2.15}$$

$$(C - \alpha_i^*)\xi_i^* = 0 \tag{3.2.16}$$

Consequently, when $\alpha_i, \alpha_i^* \in (0, C)$, we have $\xi_i, \xi_i^* = 0$. This allows us to conclude that

$$b = y_i - \langle \omega, x_i \rangle - \varepsilon \text{ for } \alpha_i \in (0, C) \tag{3.2.17}$$

$$b = y_i - \langle \omega, x_i \rangle + \varepsilon \text{ for } \alpha_i \in (0, C) \tag{3.2.18}$$

From equation (3.2.13) and equation3.2.14, the Lagrange multipliers are nonzero only if $\left| \hat{f}(x_i) - y_i \right| > \varepsilon$. Therefore we have a sparse expansion of ω in terms of x_i. The examples correspond to the nonzero Lagrange multipliers are called support vectors (SV). The support vector expansion can be rewritten as follows:

$$\hat{f}(x_i) = \sum_{i=1}^{N} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b = \sum_{i \in SV} (\alpha_i - \alpha_i^*) \langle x_i, x \rangle + b \tag{3.2.19}$$

The equivalence between SVR and sparse approximation has been pointed out by Girosi [90] where the same solution can be obtained from both SVR and spars approximation by solving the same quadratic programming problem.

3.2.3 Nonlinear Support Vector Regression

The Next step is to make the support vector regression algorithm nonlinear. This could be achieved by using a nonlinear mapping (ϕ) to map the low dimensional input space into a high dimensional feature space (F) (Fig. 3.2) via a function, $\phi : \Re^d \to F$. Then the following estimate function is used to make linear regression in the feature space [91] as:

$$f(x) = \langle \omega, \phi(x) \rangle + b \tag{3.2.20}$$

where $\phi(x)$ denotes the high dimensional feature space which is nonlinearly mapped from the input space. ω contains the coefficients that have to be estimated from the data and b is a real constant.

Figure 3.3 shows an example of a nonlinear regression function with an ε-insensitive loss function. The variables ξ_i, ξ_i^* measure the cost of the errors on the training points. These variables are equal to zero for all points inside the ε [91].

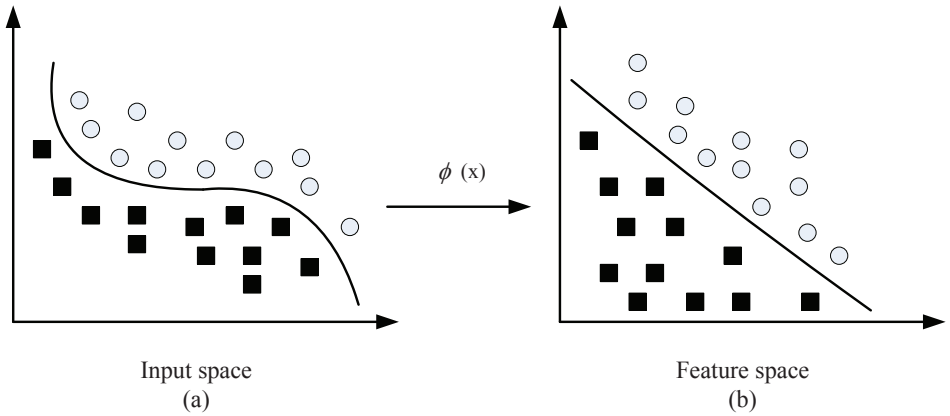


Figure 3.2: Transformation process illustration of s SVR Model. A nonlinear mapping function $\phi(x)$ defined to mapping a nonlinear problem in two dimensional input space (a) to linear problem in two dimensional feature space (b).

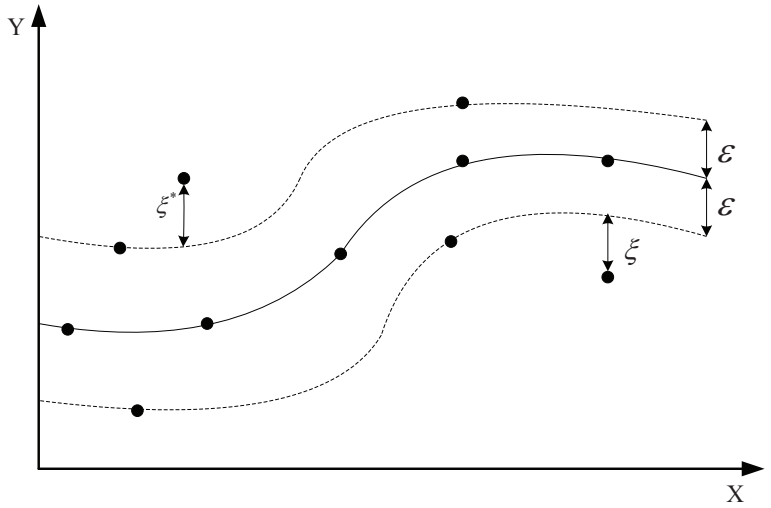


Figure 3.3: The ε-insensitive loss function for a nonlinear regression function. The solid line is the approximation function and the dashed line is the contour of the margin. The points lying on or outside

The nonlinear SVR solution based on ε-insensitive loss function is given by [89]:

$$\max_{\alpha_i,\alpha_i^*} \begin{cases} -\frac{1}{2}\sum_{i,j=1}^{N}(\alpha_i-\alpha_i^*)(\alpha_j-\alpha_j^*)Q(x_i,x_j) \\ -\varepsilon\sum_{i=1}^{N}(\alpha_j+\alpha_j^*)+\sum_{i=1}^{l}y_i(\alpha_j-\alpha_j^*) \end{cases} \tag{3.2.21}$$

$$\text{subject to} \begin{cases} \sum_{i=1}^{N}(\alpha_i-\alpha_i^*)=0 \\ \alpha_i,\alpha_i^*\in[0,C] \end{cases}$$

where $Q(x_i,x_j)=\langle\phi(x_i),\phi(x_i)\rangle$ is the kernel function that is the inner product of the point $\phi(x_i)$ and $\phi(x_j)$ mapped into feature space.

The two parameters C and ε are free parameters selected by the user. The complexity of SVR method depends on these parameters, therefore they must be tuned simultaneously. The regression output takes the following form [89]:

$$\hat{f}(x)=\sum_{i=1}^{N}(\alpha_j-\alpha_j^*)Q(x_i,x_j)+b=\sum_{i\in SV}(\alpha_j-\alpha_j^*)Q(x_i,x_j)+b \tag{3.2.22}$$

The ε-insensitive loss function is attractive because unlike to quadratic and Huber cost functions, where all the data points will be support vectors, the algorithm solution can be sparse[88].

We can briefly review the basic properties of the SVR algorithm for the regression as described so far. First, the input low dimensional input space (for which a prediction is to be made) is mapped into high feature space by a map ϕ, This equivalents to calculate kernel functions. Finally, the dot products are added up using the weights. This, plus the constant term b yields the final prediction output.

3.2.4 Kernel function

As stated, the nonlinear SVR utilisers the fact that the kernel $Q(x_i,x_j)$ becomes a dot product on the feature space F in constrast to a dot product of the input space of the linear case. The dot product on the feature space is denoted as $\langle\phi(x_i),\phi(x_i)\rangle_F$. The idea of the kernel function is to enable operations to be performed in the input space after rather than the potentially high dimensional feature space. Therefore, the inner product does not need to be evaluated in the feature space [88]. The following theorem of functional analysis which based upon Reproducing Kernel Hilbert Space

[90] shows that an inner product in feature space has an equivalent kernel in input space which provided certain conditions hold.

$$Q(x_i, x_j) = \langle \phi(x_i), \phi(x_i) \rangle \tag{3.2.23}$$

If Q is a symmetric positive definite function, which satisfies Mercers Conditions,

$$Q(x_i, x_j) = \sum_{z=1}^{\infty} \alpha_z \phi_z(x_i) \phi_z(x_j), \text{ where } \alpha_z \geq 0 \tag{3.2.24}$$

and

$$\int \int Q(x_i, x_j) a(x_i) a(x_j) dx_i dx_j > 0, \quad \text{where} \quad \int (\psi)^2(x) d(x) < \infty \tag{3.2.25}$$

then the kernel represents a legitimate inner product in feature space.
There are different types of kernel functions that satisfy Mercers conditions for SVR. They can be defined as follows:

- Linera kernel: The liner kernel is the simplest function of all kernel functions.

$$Q(x_i, x_j) = \langle x_i, x_j \rangle \tag{3.2.26}$$

- The Gaussian radial basis function kernel:

$$Q(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \tag{3.2.27}$$

- The polynomial kernel: A polynomial kernel of degree p is defined as:

$$Q(x_i, x_j) = (\langle x_i, x_j \rangle + 1)^p \tag{3.2.28}$$

- The hyperbolic tangent kernel:

$$Q(x_i, x_j) = \tanh\left(\text{scale} \langle x_i, x_j \rangle + \text{offset}\right) \tag{3.2.29}$$

- The Laplace radial basis function kernel:

$$Q(x_i, x_j) = \exp\left(-\frac{\|x_i - x_j\|}{2\sigma^2}\right) \tag{3.2.30}$$

- The linear splines kernel:

$$Q(x_i, x_j) = 1 + \langle x_i + x_j \rangle + \langle x_i, x_j \rangle \min(x_i, x_j) - \frac{x_i + x_j}{2}(\min(x_i, x_j))^2 + \frac{1}{3}(\min(x_i, x_j))^3 \tag{3.2.31}$$

- The Additive kernel: More complicated kernel can be obtained by forming summing kernels, since the sum of two positive definite functions is positive definite.

$$Q(x_i, x_j) = \sum_z Q_z(x_i, x_j) \tag{3.2.32}$$

The Gaussian and Laplace radial basis function kernels are general-purpose kernels used when there is no prior knowledge about the data. Whereas, the linear kernel is useful when dealing with large sparse data vectors as is usually the case in text categorisation. In addition, the polynomial kernel is popular in image processing [94].

3.2.5 SVR optimisation

In the ε-insensitive loss function SVR algorithm, a large quadratic programming (QP) problem (equation (3.2.21)) must be solved which gives a unique global minimum. The QP problem can be expressed in matrix notation as follows:

$$\min_{\alpha_i, \alpha_i^*} \frac{1}{2}\beta^T Q \beta + c^T \beta \tag{3.2.33}$$

$$\text{subjectto} \begin{cases} A\beta = 0 \\ \alpha_i, \alpha_i^* \in [0, C], i = 1, ..., N \end{cases}$$

where

$$\beta = \begin{bmatrix} \alpha_i \\ \alpha_i^* \end{bmatrix}, \quad c = \begin{bmatrix} \varepsilon + y \\ \varepsilon - y \end{bmatrix} \tag{3.2.34}$$

$$Q \begin{bmatrix} Q & -Q \\ -Q & Q \end{bmatrix}, \quad A = \left\{ \underbrace{1, ..., 1}_{N}, \underbrace{-1, ..., -1}_{N} \right\} \tag{3.2.35}$$

$y = [y_1, ..., y_N]^T$ and $Q = Q(x_i, x_j)$ for $i, j = 1, ..., N$.

According to Fletcher [95], when Q is a positive definite matrix and the constraints

of equation (3.2.21) are linear, the solution of equation (3.2.21) must satisfy the KKT condition as follows;

$$\begin{aligned} \beta_i = 0 &\Leftrightarrow y_i \hat{f}_i \geq 1, \\ 0 < \beta_i < C &\Leftrightarrow y_i \hat{f}_i = 1, \\ \beta_i = C &\Leftrightarrow y_i \hat{f}_i \leq 1, \end{aligned} \tag{3.2.36}$$

where $\hat{f}_i = \hat{f}(x_i)$ is the output of the SVM for the $i_t h$ training example.
Two algorithms will be introduced to solve the large quadratic programming problem. They are Decomposition Algorithm (DA) [97] and Sequential Minimal Optimization (SMO) [96] [85].

Decomposition algorithm

The Decomposition Algorithm (DA) solves a sequence of small quadratic programming sub-problems instead of solving the large quadratic programming problem at once [97]. It is based on the observations that a sequence of quadratic programming sub-problems which at least always contains one example violating the KKT conditions will eventually converge to the optimal solution [98]. Osuna [98] suggested keeping a constant size matrix for every quadratic programming sub-problem, which implies adding and deleting the same number of examples in each iteration.
In the DA, the index of the training set is partitioned into two sets. The first one is called a working set (G) while the second on is called a correcting set (E). Therefore, β, y, c and Q from equation (3.2.33) can be aggraded properly as follows:

$$\beta = \begin{bmatrix} \beta_G \\ \beta_E \end{bmatrix}, c = \begin{bmatrix} c_G \\ c_E \end{bmatrix}, y = \begin{bmatrix} y_G \\ y_E \end{bmatrix}, Q = \begin{bmatrix} Q_{GG} & Q_{GE} \\ Q_{EG} & Q_{EE} \end{bmatrix} \tag{3.2.37}$$

The dual objective function can be rewritten involving the two sets G and E as follows:

$$\min_{\alpha_i, \alpha_i^*} \frac{1}{2} \beta_G^T Q_{GG} \beta_G - \beta_G^T \left(c_G Q_{GE} \beta_E \right) \tag{3.2.38}$$

$$\text{subject to} \begin{cases} \langle y_G, \beta_G \rangle + \langle y_E, \beta_E \rangle = 0 \\ 0 \leq \beta_G \leq C \end{cases}$$

where C is a column vector with all elements equal to C.
At each step, n elements exchange between set G and E where at least one variable violating the KKT conditions is moved form E to G. Then the sub-problem (3.2.38) which involving the new working set is solved. The cycle repeats until no example violates the KKT conditions. Note that n is arbitrary.

Sequential minimal optimisation

Sequential Minimal Optimisation (SMO) is a special case of the DA. The SMO was derived in [96] and applied to text categorisation problems. Then, Smola et al. [85], has generalised SMO for solving the regression problems.
The SMO is derived by taking the idea of DA to its extreme and optimising a minimal set of just two points at each iteration ($n = 2$). The advantage of SMO comes from the fact that the optimisation problem for two data points admits an analytical solution, eliminating the need to use an iterative OP solver, which is hard to program, as part of algorithm.
Unfortunately, keeping the size of working set equal to 2 leads to more sub-problems which need to be solved by using SMO, but each sub-problem can be solved very quickly due to the existence of analytical solutions. Therefore, the overall training time of the SMO is less than the DA. In addition, the SMO does not require extra matrix storage due to the fact the QP sub-problem can be solved analytically.
There are two main components of SMO [96]. They are the analytical solution for the sub-problem and a heuristic strategy for choosing which two examples are to be optimised which corresponds to exchanging two examples form G to E and reverse. The implementation of the SMO is straightforward. The details and pseudo code of the SMO for regression can be found in [85].

3.3 Local Prediction with False Neighbors Filtered

Local prediction methods based on phase reconstruction normally perform better than global methods based on phase reconstruction. With the local prediction method, each predicting point has its own model constructed based on its nearest

neighbors (NNs) which are found in the neighborhood of the phase space reconstructed from the time series, and the fitness of NNs would mainly affect the model performance. However, it has been found that NNs may contain a class of false neighbors (FNs) which would decrease the fitting accuracy dramatically and lead to a poor modeling performance. This means that not all NNs are suitable for use in local prediction and some of them should be filtered [8]

3.3.1 Local Prediction

In global predictor, a prediction model is trained based on the entire data history and used to predict the energy demand at a specific time with a fixed data window. To overcome the the drawbacks of the global predictors, the local predictors can be used. In last few decades, the local predictor approach has interested many researchers to solve the nonlinear time series prediction problem such as McNames, et al. introduced the local liner model for very short dataset. Therefore, this model was used in to generate the winning entry of the K. U. Leuven time series competition. Lau, et al [7]. combined the strength of SVR and local predictor, and proposed a better algorithm for nonlinear time series prediction. More details is discussed in Section 3.4.1.

The local prediction method relies on a set of nearest neighbors which evolves similarly in the reconstructed phase space. The set of nearest neighbors D_N can be obtained as follow:

$$\begin{aligned} &D_N = \{z_i, y_i\}, \\ &\begin{cases} z_i = x_j, \\ y_i = x_i + n, \end{cases} \text{when } \|x - x_j\| < r, j = 1, ..., M, \end{aligned} \tag{3.3.1}$$

where n is the prediction horizon, x belongs to the testing sample and r is the radius enclosing the neighbors in reconstructed state space, and $\|x - x_j\| = \sqrt{\sum_{t=1}^{m} [x(t) - x_j(t)]^2}$ is the Euclid distance in the phase space. The local prediction method takes the nearest neighbors D_N as the training data set of the prediction model.

3.3.2 False Neighbors Filter

The local predictor relies on a set of nearest neighbors which evolves similarly in the reconstructed phase space. The choice of choosing neighbors is effected and limited by the finite size of the data set, by the stochastic noise, and especially by the complex structure of the attractor. These limitation are the main source of errors in the analysis. Finding suitable and reasonable neighbors of one known point is one of the most important tasks in achieving reliable results. The original method to chose the subset (neighborhoods) is on the basis of the Euclidian distance between the testing data and the training data in the input space. For each query vector q, the K nearest neighbors $\{g_q^1, g_q^2, ..., g_q^K\}$ among the training inputs are chosen by using the Euclidian distance as the distance metric between the q and each g in the reconstructed time series. where

$$\left\|q - g_q^K\right\| > \left\|q - g_q^{K-1}\right\|$$

However, false neighbors could exist in the reconstruct state space after using a particular embedding method. Those FNs can be caused by improper embedding, such as an insufficient large embedding dimension. This effect is illustrated by Fig. 3.4. Suppose the predicting (reference) point is $X_n(t)$. By utilizing the original local predictor (Euclidian distance), two NNs are found, $X_{n+1}(t)$ and $X_{n+2}(t)$, and $\|X_n(t) - X_{n+2}(t)\| < \|X_n(t) - X_{n+1}(t)\|$. Note that, $X_{n+2}(t)$ is a true neighbor of $X_n(t)$. However, since the deviation between $X_{n+2}(t+1)$ and the predicting (target) point $X_n(t+1)$ is large, it would decrease the fitting accuracy dramatically if we choose $X_{n+2}(t)$ as a training sample. Actually, $X_{n+1}(t)$ is much better than $X_{n+2}(t)$, even $X_{n+2}(t)$ is closer to the reference point. Such neighbors like $X_{n+2}(t)$ are called false neighbors. If those points are selected as the training samples, the predicted performance will obviously reduce. There are many ways to calculate the false neighbors. In this paper, we present a simple and effective way which is applied with the exponential separation rate. In the reconstructed phase space, set $X(t)$ as the predicting (reference) point at time t, and $X_j(t)$ is the j^{th} neighbor of $X(t)$ in the phase space, where $1 \leq j \leq N$. The exponential separation rate between the

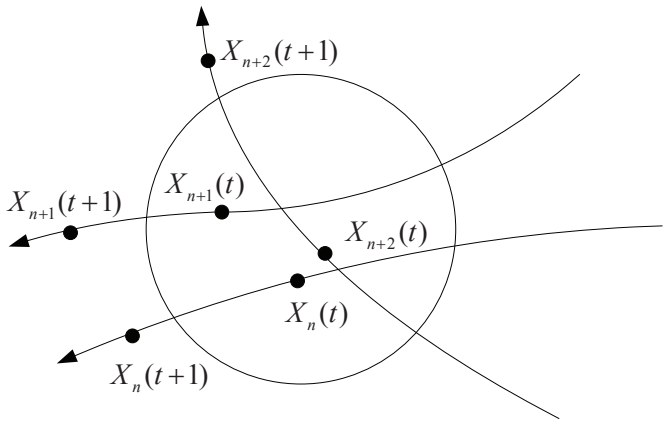


Figure 3.4: The schematic diagram of affecting forecast on evolutionary track of neighbor points. In this figure, $X_{n+2}(t)$ is the false neighbour point of $X_n(t)$

track of $X(t)$ and the track of $X_j(t)$ is,

$$\xi_j = \ln |d_j(t)/d_j(t-1)| \tag{3.3.2}$$

where $d_j(t-1)$ is the the Euclidean distance between $X(t-1)$ and $X_j(t-1)$ at time $t-1$. The proposed method employs the Euclidean distance d_j and exponential separation rate ξ_j to determine the validity of $X_j(t)$, which is $1 \leq j \leq N$ and $\xi_j \leq \varphi$. This method ensures that $X_j(t)$ is one of the nearest neighbors with a very small exponential separation rate, and the number of N and φ can be modified(enlarged or reduced) in order to achieve K optimal nearest neighbors.

3.4 Implementations of the Proposed FNF-SVRLP

3.4.1 SVRLP

This method is a general SVR based local predictor [7], which is presented by us previously. The SVRLP method can be summarized as follows: Firstly, τ and m are used to reconstruct the time series. Then, the local predictor (Euclidean distance) is applied to find the K nearest neighbors (could contain FNs) for each query vector. Thirdly, the K nearest neighbors are used to train the SVR to obtain support vectors and corresponding coefficients. Following is the brief process of SVRLP:

- The first step is to reconstruct the time series by applying the delay coordinate embedding methodology. The key part of this step is to obtain the embedding dimension m and the time delay constant τ. They can be finished by utilizing the correlation dimension method and the mutual information method respectively.

- The second step of this procedure is to find the K nearest neighbor $\left\{X_Z^1, X_Z^2, X_Z^3, ...X_Z^K\right\}$ for test. These points should be satisfied this rule: the distance from each of them to Z (the query point) has to be matched the function below:

$$d\left(X, Z\right) = \sqrt{\sum\nolimits_{j=1}^{d} \left[x(t_1 - (j-1)m - Z(t_2 - (j-1)m)\right]^2}$$

- When finishing these steps, it should be started to train the SVR algorithm. In the step 3, the value of $\{x(Z_l + T)\}_{l=1}^K$ (T is the prediction step) will be set as the objective value. Using the MATLAB optimization toolbox and complied code, the support vectors and their weight coefficients can be obtained.

- Achieve the prediction result of the point $x(t+T)$ of the corresponding vector Z. And the step 2 and 3 will be repeated if the future values of different Z are all required.

3.4.2 FNF-SVRLP

From Section 3.3.2, it is found that the false neighbors can exist in the reconstruct state space, and if those FNs are selected as the training samples, the accuracy of local modeling will be weaken. Therefore, it is necessary to optimize those NNs with false neighbors filtered before we apply them to train the SVR. The steps for gas demand prediction based on the proposed method can be summarized as following:

- Step 1 Reconstruct the time series: Load the multivariate time series dataset $X = (x_t(t), x_2(t), ..., x_i(t))$, $(t = 1, 2, ...T)$. Determining the embedding dimension m and time delay constant τ. Then, reconstruct the multivariate time series using these values.

- Step 2 Form a training and validation data: The input dataset after reconstruction $\tilde{X}$ is divided into two parts, that is a training $\tilde{X}_{tr}$ and validation $\tilde{X}_{va}$. The size of the training dataset is N_{tr} while the size of the validation dataset is N_{va}

- Step 3 Determine the nearest neighbors: For a query point $x_q(t)$ at time t, choosing the N nearest neighbors. $\{z^1_{x(t)_q}, z^2_{x(t)_q}, z^3_{x(t)_q}, ...z^N_{x(t)_q}\}$ of this query point using Euclidian distance between $x_q(t)$ and each point in $\tilde{X}_{tr}(1 < N \ll N_{tr})$

- Step 4 Calculate the exponential separation rate: For the N nearest neighbors $\{z^1_{x(t)_q}, z^2_{x(t)_q}, z^3_{x(t)_q}, ..., z^N_{x(t)_q}\}$, each nearest neighbor can achieve its correspondent exponential separation rate via Equation 3.3.2, which is $\{\xi_{z^1_{x(t)_q}}, \xi_{z^2_{x(t)_q}}, \xi_{z^3_{x(t)_q}}, ..., \xi_{z^N_{x(t)_q}}\}$.

- Step 5 Achieve K optimal nearest neighbors: Firstly, the Equation 4.3.1 is applied to calculate the number of K. To the i^{th} optimal nearest neighbors, $1 \leq i \leq K < N$ and $\xi_{z^i_{x(t)_q}} < \varphi$. The value of N and φ are modified until achieving K optimal nearest neighbors.

- Step 6 Train SVR: The K optimal nearest neighbors of query point are used to train the SVR algorithm.

- Step 7 Calculate the prediction value of the current query point using Equation (8).

- Step 8 Then, the steps 3 to 7 can be repeated until the future values of different query points are all acquired.

Fig. 3.5 presents the computation procedure for the proposed method.

3.5 Conclusion

In this chapter, an ameliorated local predictor has been proposed. The core of the modeling of the local prediction is the fitness of NNs, which mainly effect the final performance of prediction. However, due to the existence of false neighbors (FNs),

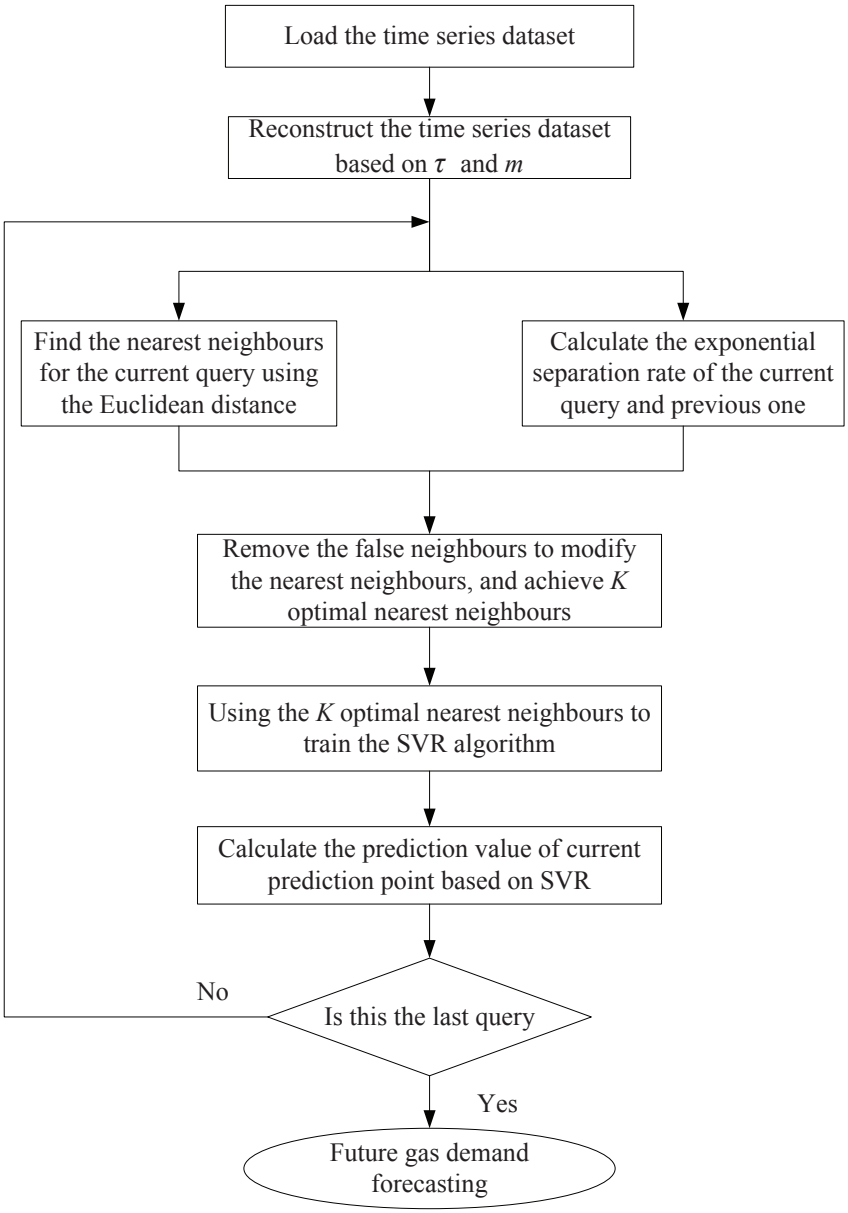


Figure 3.5: Flowchart of the proposed model

the fitting accuracy would be decreased dramatically and thus cause the reduction of the forecasting performance. The proposed approach is applying the false neighbors filter to remove those false neighbors and obtain the optimal nearest neighbors. The SVR is combined with NNs filter and result in a novel approach which is named as support vector regression-based local prediction false neighbors filtered (FNF-SVRLP).

Chapter 4

FNF-SVRLP Based Wind Power Prediction

4.1 Introduction

In recent years, the interest in the utilization of renewable energy has been significantly developed due to the global energy crisis and environmental concerns. In comparison with other renewable energy, wind power, a clean and inexhaustible energy, is safer and less expensive to exploit. Therefore, an increasing number of countries in the European (EU) are devoted to developing wind farms. According to the reports, published by European Wind Energy Association [5][6](EWEA), by the end of 2020, wind energy will meet 14% of the EU's total electricity consumption and achieve 30% lower emissions. However, the increasing penetration of wind power into the electric power grid accompanied with a series of challenges. Due to the uncertain and variable nature of wind resources, the output power of wind farms is hard to control, which could lead to the instability of the power grid operation and the unreliability of electricity supplies. Therefore, wind power prediction (WPP), particularly short-term (a few hours to a few minutes) WPP, is crucially important to grid operation. Utilization of short-term WPP allows scheduled operation of wind turbines and conventional generators, thus ensure the stability and efficiency of the power grid at optimal operating cost [99].

To deal with WPP, various methods have been utilized to build the prediction models. They can be classified into several categories, such as physical methods and statistical methods. The physical method is based on numerical weather prediction (NWP). This method is using abundant physical considerations as input variables to forecast the wind information. This kind of method performs well in long-term prediction. The statistical method aims to find the relationship between input data and the forecasted output data directly by a statistical analysis of previous time series data. The statistical methods usually have advantages in short-term prediction. Autoregressive moving average (ARMA) [100] is one of the tractional statistical methods. Support Vector Regression (SVR) is a powerful methodology derived from Support Vector Machines (SVMs), it is based upon statistical machine learning and can be utilized to achieve nonlinear mapping from sample space to feature space through kernel functions. Moreover, the SVR can achieve an outperformance of fitting accuracy for chaotic time series prediction [127]. Recently, it has already been utilized to predict short-term wind speed with success [102][103].
According to previous research [104], local prediction methods can normally perform better than global methods for chaotic time series prediction. In local prediction methods, each predicting point has its own model constructed by its nearest neighbours (NNs) which were found in the neighbourhood of the phase space reconstructed from the time series. However, it has been found that NNs may contain a class of false neighbours (FNs) which are unsuitable for use in the local prediction, and these FNs can cause the reduction of the forecasting performance [105].
In this chapter, the proposed FNF-SVRLP is used to predicted the wind power, and the results are compared with the original SVRLP and ARMA to show its superiority.

4.2 Phase Space Reconstruction

As for a chaotic system, Packard [106] believed that the phase space can be reconstructed from a univariate time series, since the univariate time series contains the information of all the variables in this dynamic system. Takens [107] and Sauer

[108] developed the embedding theorem and supported Packard's idea. The theorem regards a one-dimensional chaotic time series as compressed information of a higher dimension. Therefore chaotic features can be extracted by extending the one-dimensional time series to higher dimensional one. A univariate time series $(x_i)_{i=1}^{K}$, when K is the length of time series, is obtained by monitoring a scalar signal of a D-dimension chaotic attractor for a finite time T and with finite precision, the phase space R^D of the attractor can be reconstructed by using delay coordinate defined as

$$X(t) = [x(t), x(t+\tau), ..., x(t+(m-1)\tau)]^T \tag{4.2.1}$$

subject to: $t = 1, ..., M, M = K - (m-1)\tau$, and $m = integer(2d+1)$, where m is the embedding dimension, d is the correlation dimension which is an estimation of the dimension of the chaotic attractor D, τ is the delay constant and T denotes the vector transpose. After reconstructing one-dimensional to a higher dimensional time series, the obtained time series is called a reconstructed time series. In practices, both the dimension of the attractor and the delay constant must be estimated from the time series.

4.2.1 Estimate the delay constant

In order to reconstruct the phase space, it is necessary to estimate the delay constant τ. In principle, any choice of the delay constant τ is acceptable in the limit of an infinite amount of data. However, for a finite amount of data, the choice of τ can not be too small or too large, or else it will result in redundancy or irrelevance in the reconstructed phase space. The mutual information method developed by Fraser and Swinney [109] based on research by Shaw and others [110] is used to catch a reasonable choice of τ. To give a brief introduction of method used, the following case is discussed. In this case, the evolution vectors are of the form

$$(x, y) = [x(t), x(t+\tau)], t = 1, ..., N. \tag{4.2.2}$$

The mutual information function is defined by averaging the mutual information between two sets of measurements of x and y as a function of τ. This average

mutual information is defined by

$$I_{xy}(\tau) = H(x) + H(y) - H(x, y), \tag{4.2.3}$$

where

$$\begin{aligned} H(x) &= -\sum P_x(x) \log[P_x(x)], \\ H(y) &= -\sum P_y(y) \log[P_y(y)], \\ H(x, y) &= -\sum P_{xy}(x, y) \log[P_{xy}(x, y)], \end{aligned} \tag{4.2.4}$$

where $H(\cdot)$ is the entropy function which represents the average mutual information, $P_x(x)$ and $P_y(y)$ are probability density distribution of x and y, and $P_{xy}(x, y)$ is the joint probability density distribution of x and y. The reasonable choice of τ is the value that yields the first minimum of $I_{xy}(\tau)$.

4.2.2 Estimate the embedding dimension

To estimate the dimension of a dynamic system from the a given time series, the correlation dimension method can be used. The correlation dimension can be determined from the correlation integral which is defined as follows:

$$C(r) = \frac{1}{M^2} \sum_{i,j=1}^{M} \theta[r - \|X(i) - X(j)\|], \tag{4.2.5}$$

where $\theta(\cdot)$ is the Heaviside step function:

$$\theta(x) = \begin{cases} 0 & \text{if} \quad x \leq 0, \\ 1 & \text{if} \quad x > 0. \end{cases} \tag{4.2.6}$$

Obviously, the correlation integral depends on the size of r. When r is large, r_{ij} (the distance between two points) is smaller that r, then $C(r) = 1$; when r is small, r_{ij} is larger than r, then $C(r) = 0$. Neither of the above cases can reflect the inner properties of the dynamic system, they are insignificant. So it is important to choose an appropriate r according to different parameters. When $r \to 0$, the relation between the correlation integral $C(r)$ and r is

$$\lim_{r \to 0} C(r) \propto r^d, \tag{4.2.7}$$

and choose an appropriate r, the correlation dimension d is given by

$$d = \frac{\log(C(r))}{\log(r)} \tag{4.2.8}$$

4.3 Optimal value of nearest neighbors

In order to establish the local predictor, it is important to choose value of the nearest neighbors, K. There are some methods used in literatures to determine this parameter, such as cross validation [138] and bootstrap [139]. This parameter should be low for high density datasets and should be high for low density ones. We calculate K using a systematic method proposed by us in previous work [140] as follows:

$$K = \text{round}\left(\frac{\partial}{H \times k_{\max} \times D_{\max}} \sum_{i=1}^{H} \sum_{k=1}^{k_{\max}} D_k(x_i)\right) \tag{4.3.1}$$

where H is the total number of training points, $k_{\max}$ is the maximum number of nearest neighbors, $D_k(x_i)$ is the distance between each training point x and its nearest neighbors while $D_{\max}$ is the maximum distance, $\frac{1}{H \times k_{\max} \times D_{\max}} \sum_{i=1}^{H} \sum_{k=1}^{k_{\max}} D_k(x_i)$ is the average distance around the points, which is inversely proportional to the local densities and ∂ is a constant. The two constants $k_{\max}$ and ∂ are parameters with very low sensitivity. $k_{\max}$ can be chosen as a percentage of the number of training points (H) for efficiency while ∂ can be chosen as a percentage.

4.4 Case study

The data set used to evaluate the performance of the proposed FNF-SVRLP based WPP model is the real world wind power data from China, it was collected from wind power plants in the Jing-Jin-Tang area. The quarter hourly wind power data are the main inputs. The data set was the wind power from Apr. 16th, 2011 to Aug. 15th, 2011, which covers four months. The first three months (from Apr. 16th, 2011 to Jul. 15th, 2011) was used as training data and the fourth month (from Jul. 16th, 2011 to Aug. 15th, 2011) was used as testing data.

For all models, the prediction performance was quantified by the mean absolute

percentage error (MAPE) and the normalized mean square error (NMSE). They can be defined as

$$\text{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{y_i} \times 100 \tag{4.4.1}$$

$$\text{NMSE} = \frac{1}{\Delta^2 N}\sum_{i=1}^{N}(\hat{y}_i - y_i)^2, \Delta^2 = \frac{1}{N-1}\sum_{i=1}^{N}(y_i - \bar{y})^2 \tag{4.4.2}$$

where N is the size of testing dataset, $\hat{y}_i$ is the forecasted wind power, y_i is the actual wind power , $\bar{y}$ is the mean of the actual wind power, and i means the test instance index.

4.4.1 Models analyzed in the case study

1. ARMA model : ARMA was proposed by Box and Jenkins in 1976. ARMA is short for Autoregressive moving average. It has been widely used as the prediction approach . The basic principles of ARMA is analyzed in [100].

2. SVRLP model : This model is a general SVR based local predictor. NNs which are found by the local predictor (Euclidean distance) are all used as training samples. Those NNs may contain some FNs which would cause a decrease in final accuracy.

3. Proposed FNF-SVRLP model : The way of selecting training samples for this proposed model is different from the original SVRLP. In this model, not only the Euclidean distance but also the exponential separation are all judgemental features. Using this method, the probability of FNs is reduced, and the final accuracy is improved. Fig. 4.1 presents the computation procedure for the proposed model for wind power prediction.

4.4.2 Numerical Results and Discussion

In order to implement a good model, there are several important parameters to choose, such as the embedding dimension m and the time delay τ. Using the mutual information method and the correlation dimension method, which were described in

Table 4.1: MAPE and NMSE of different models

	MAPE	MAPE Impro.	NMSE	NMSE Impro.
FNF-SVRLP	4.39	-	0.012	-
SVRLP	5.75	23.65%	0.023	47.8%
ARMA	9.23	52.43%	0.065	81.7%

Section II, can determine the τ and m respectively. The final optimal τ is 2 and m is 9.

To evaluate the performance of the proposed FNF-SVRLP, a comparison with ARMA and SVRLP was conducted. The results of the comparison is shown in Table 4.1 and depicted in Fig.4.4 to Fig. 4.7.

These results show that the FNF-SVRLP performs better than ARMA and SVRLP. The MAPE of FNF-SVRLP is 4.39%, which is an improvement over the ARMA model and the SVRLP model by 52.43% and 23.65% respectively. Moveover, the NMSE of FNF-SVRLP at 0.012 is smaller than the others. Fig.5 shows results of the wind power prediction from the 16^{th} to 19^{th} July by different models in comparison with the actual wind power.

From previous results, it indicates that the accuracy of the SVR based local predictor and the importance of filtering false neighbours. Through filtering false neighbours, to leave the optimal neighbours, and using those optimal neighbours as training samples improves the accuracy.

4.5 Conclusion

In this chapter, the FNF-SVRLP has been applied to short-term wind power prediction. The embedding dimension and the time delay are estimated firstly. Then the phase space reconstruction is applied to the wind power data. In the local prediction, there are some difference between the SVRLP and the proposed FNF-SVRLP. Compared with the SVRLP, the proposed method not only applies the Euclidean distance but also the exponential separation rate to modify nearest neighbours, and only the optimal nearest neighbours can be selected as training samples. The final results presented demonstrate that the proposed model can achieve a higher prediction accuracy than the SVRLP model and ARMA model using the same wind power data.

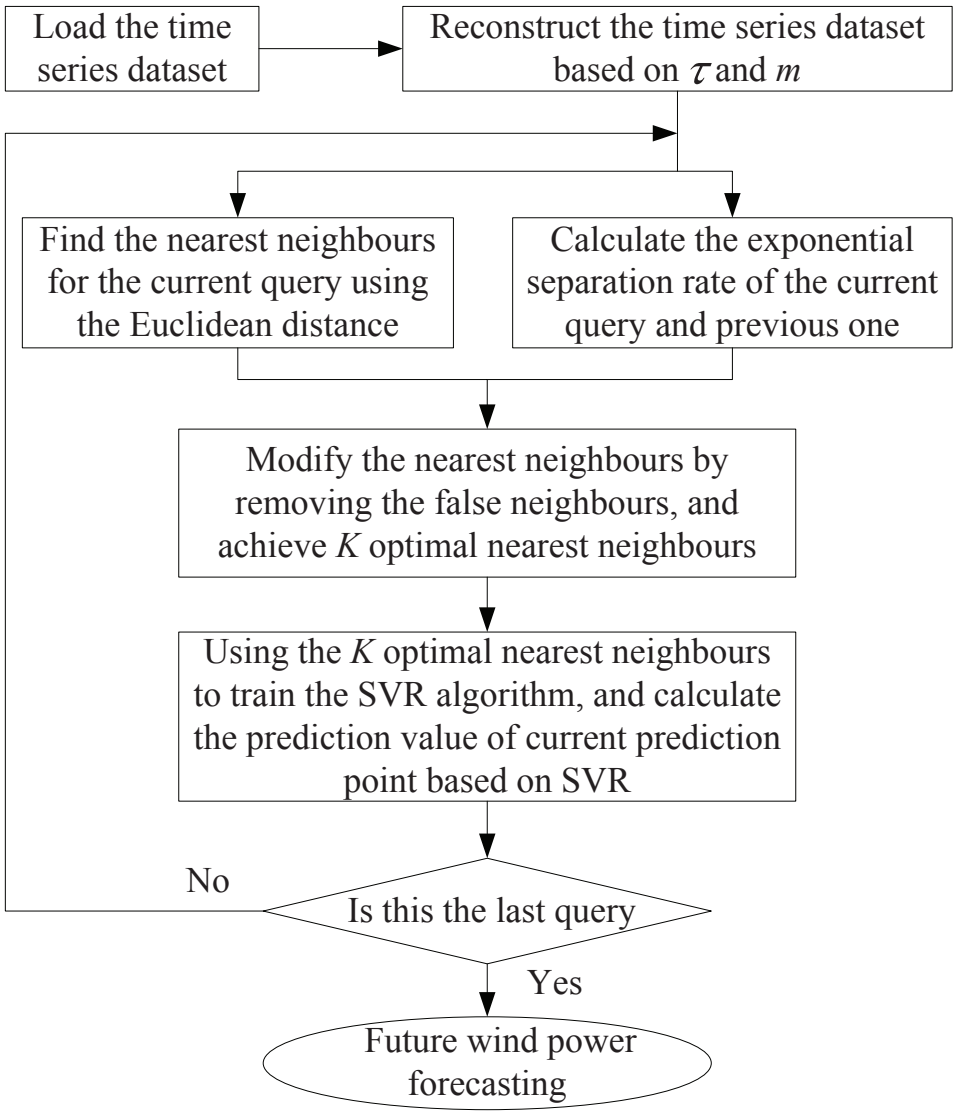


Figure 4.1: Flowchart of the proposed model

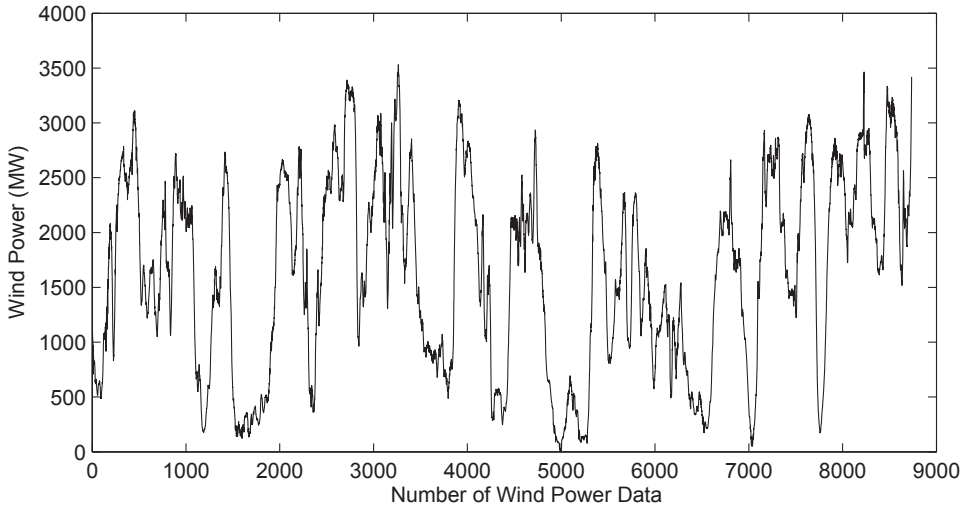


Figure 4.2: Actual wind power data used in this chapter

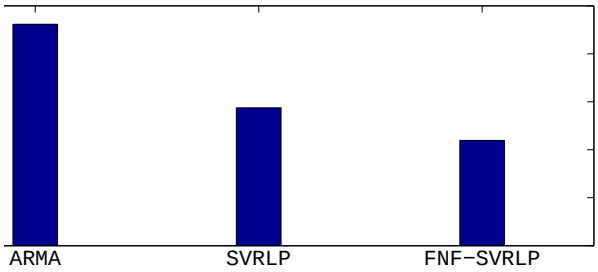


Figure 4.3: Comparison of FNF-SVRLP model and other models using the dataset of wind power

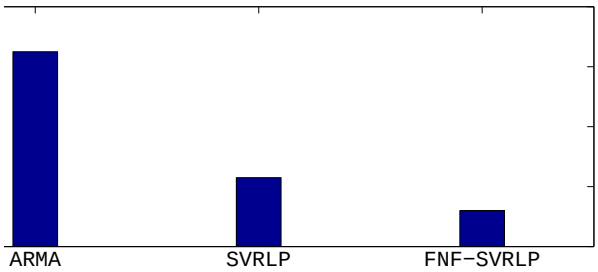


Figure 4.4: Comparison of FNF-SVRLP model and other models using the dataset of wind power

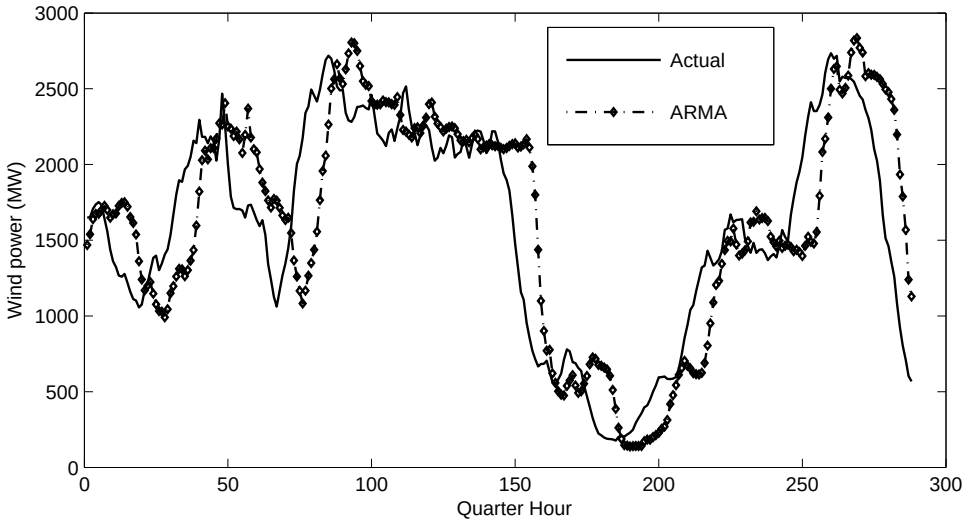


Figure 4.5: Result of wind power prediction from 16^{th} to 19^{th} July by ARMA model in comparison of actual wind power

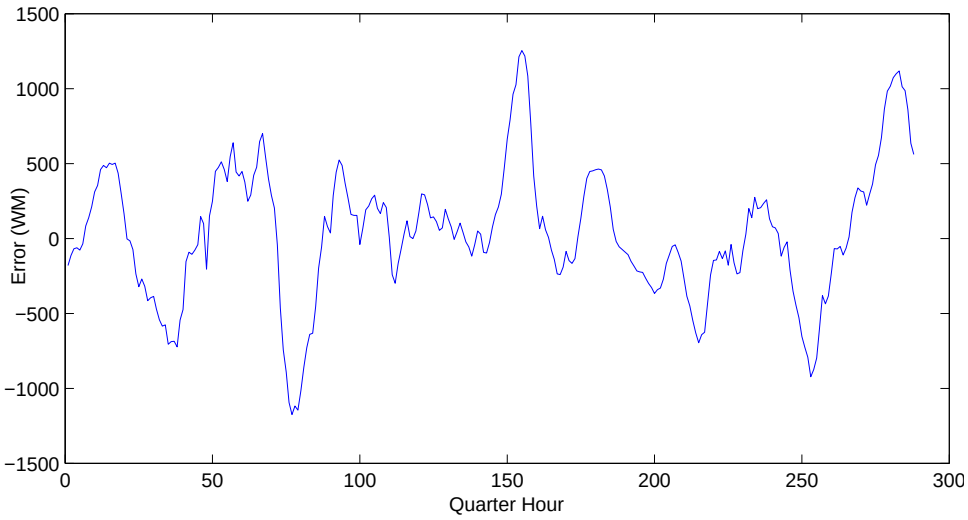


Figure 4.6: Error of the results of wind power prediction from 16^{th} to 19^{th} July by ARMA model in comparison of actual wind power

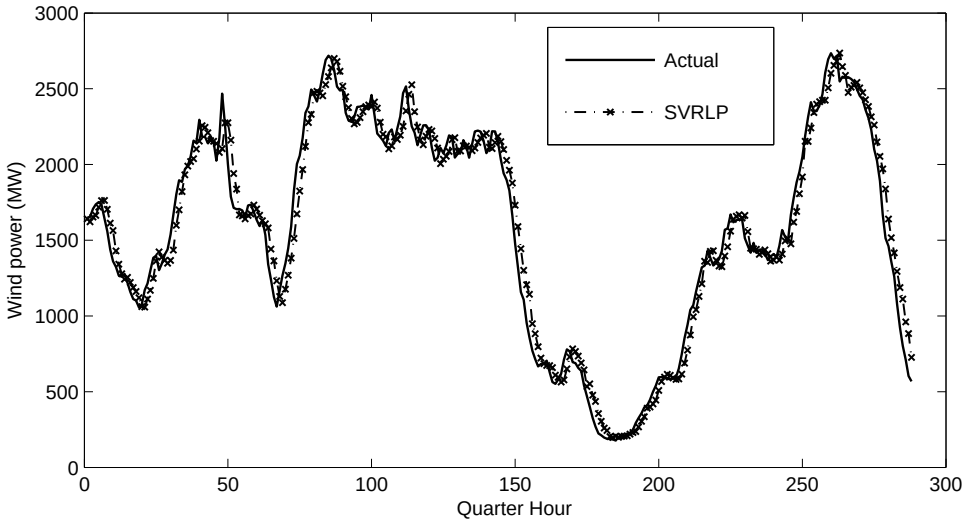


Figure 4.7: Result of wind power prediction from 16^{th} to 19^{th} July by SVRLP model in comparison of actual wind power

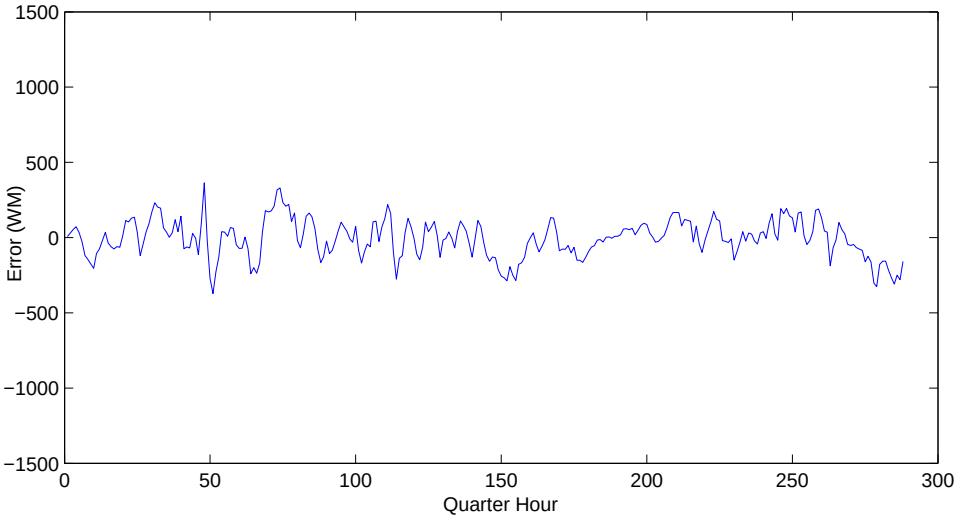


Figure 4.8: Error of the results of wind power prediction from 16^{th} to 19^{th} July by SVRLP model in comparison of actual wind power

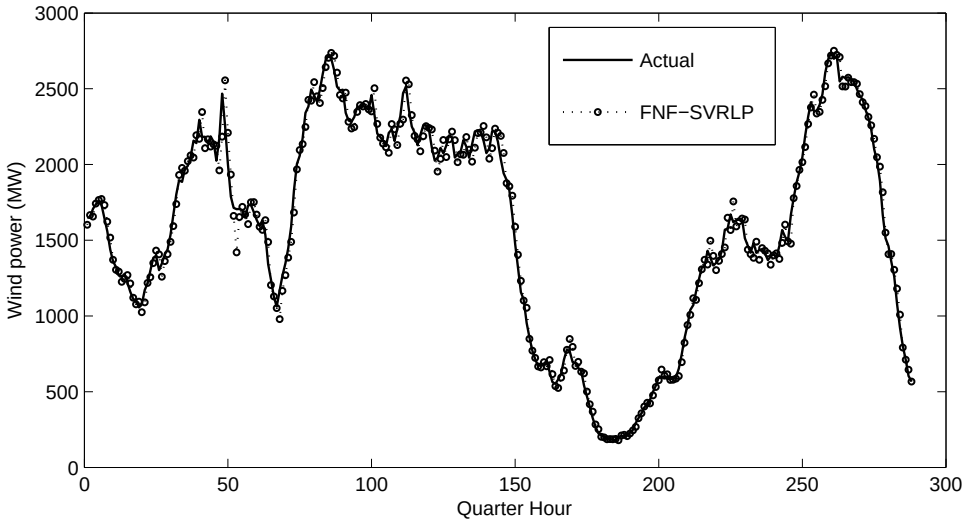


Figure 4.9: Result of wind power prediction from 16^{th} to 19^{th} July by FNF-SVRLP model in comparison of actual wind power

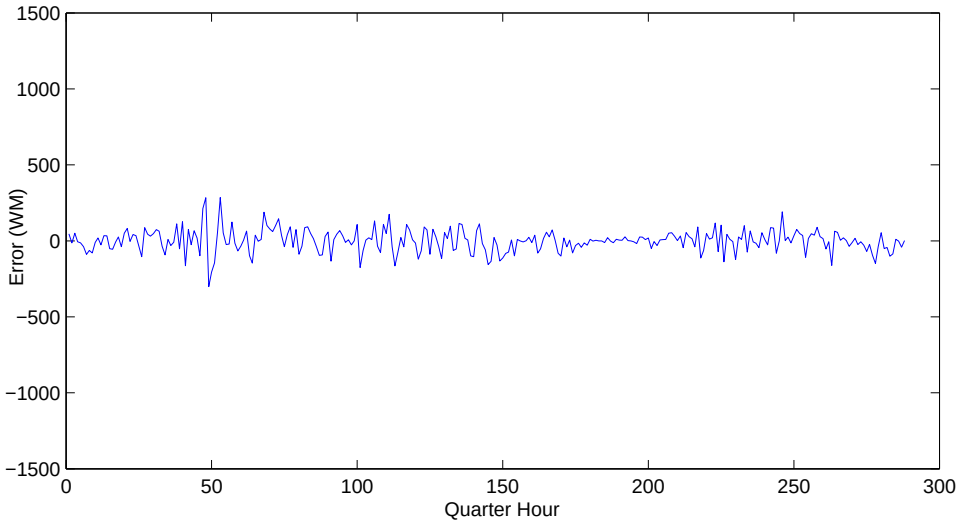


Figure 4.10: Error of the results of wind power prediction from 16^{th} to 19^{th} July by FNF-SVRLP model in comparison of actual wind power

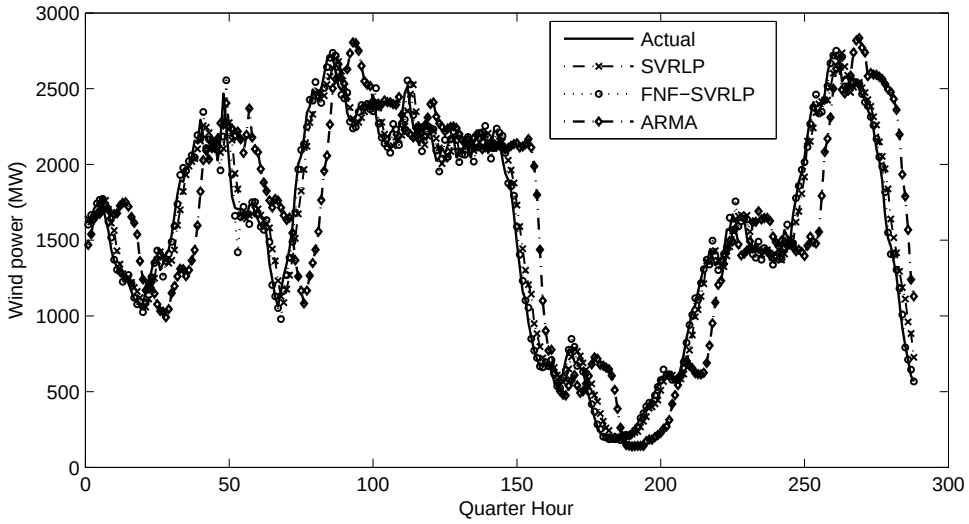


Figure 4.11: Results of wind power prediction from 16^{th} to 19^{th} July by different models in comparison of actual wind power

Chapter 5

FNF-SVRLP Based Natural Gas Demand Prediction

5.1 Introduction

Energy is a vital element to the development of society and economy of all contries[111]. Energy modeling and forecasting has attracted increasing attention within energy sectors due to the environmental concerns and increasingly stringent government polices on energy source. Natural gas as a primary energy source at the present time, was first commercially used around 1785, and has since been used widely in Europe [112][113], with UK becoming the largest market in Europe in the past decade. According to HI-Energy [114] and government report of UK [115], UK's 46% of total electricity (around 350,000 GWh) generation and approximately 70% of domestic heating in 2008 is produced by natural gas. It plays a key role in UK's energy mix. However, natural gas is an non-renewable energy source, which has limited reserve on the planet and contributes to global warming. Thus, planning energy demand is becoming an important issue in the energy sector and forecasting its demand is the starting point, especially for natural gas demand. At the same time, there are some countries, for example, UK, which is highly dependent on the import of gas, the accurate gas demand prediction can directly lower the purchase costs for distributors as well as for final consumers. Therefore, accurate gas demand predic-

tion could enhance the efficiency of natural gas usage with an optimal purchase cost and lower the emissions of global warming gas.

Pervious researches in the area of natural gas demand prediction (NGDP) can be roughly classified as the long-term NGPD and the short-term NGPD. The long-term prediction focus on forecasting the gas demand in a long time period, i.e. annual basis or monthly basis. And the main purpose of long-term prediction is always force on trend of natural gas demand, which would be helpful to the policy planning or investment et al. Similar to long-term prediction, an accurate short-term NGPD is equally important because that distributors are required by their suppliers to provide the amount of natural gas they would need for the coming hours or days within a regulated tolerance interval. Support vector machine (SVM) is a powerful machine learning method based on statistical learning theory [127]. The empirical risk minimization (ERM) principle, which is generally employed in traditional ANN, is replaced by structural risk minimization (SRM) principle in SVM. The most important concept of SRM is the utilization of minimizing an upper bound to the generalization error instead of minimizing the training error. On the basis of this principle, SVM will be equivalent to solving a linear constrained quadratic programming problem, so that the solution of SVM is always unique and globally optimal. With the introduction of Vapniks ε-insensitive loss function, SVM has been extended to solve the regression problems called support vector regression (SVR) [128]. Recently, SVR has been applied to various applications with excellent performances [129][130].

Because of the complexity of the historical gas demand and the uncertainty of the influencing factors, such as weather, economical, and other random factors, the time series reconstruction technique could be applied to the gas demand forecasting. According to the authors' work [7], local prediction methods based on phase reconstruction normally perform better than global methods based on phase reconstruction. With the local prediction method, each predicting point has its own model constructed based on its nearest neighbors (NNs) which are found in the neighborhood of the phase space reconstructed from the time series, and the fitness of NNs would mainly affect the model performance. However, it has been found that NNs

may contain a class of false neighbors (FNs) which would decrease the fitting accuracy dramatically and lead to a poor modeling performance. This means that not all NNs are suitable for use in local prediction and some of them should be filtered [8]. In this chapter, FNF-SVRLP is examined for its applicability in short-term gas demand forecast. FNF-SVRLP is an ameliorated SVRLP, which improve the local predictor (LP) by optimizing NNs. Those NNs play important roles in improving accuracy of local modeling. False neighbors filtered algorithm that uses combination of exponential separation rate and Euclid distance is proposed to optimize NNs. We proposed an unified model at first, named "Standard Model" (SM), which evenly treats the whole dataset. Then the customer behavior based "Advanced Model" (AM) is presented. The AM is the combination of three models, which are Mon. Model, Sat. Model, and other days of a week Model. The proposed FNF-SVRLP is verified by using operational gas dataset collected from National Grid and is compared with the SVRLP and ARMA to show its superiority.

5.2 Time Series Reconstruction

Packard [131] indicated that the phase space can be reconstructed from a univariate time series, since the latter contains the information of all the variables in this dynamic system. Takens [132] and Sauer [133] developed the embedding theorem and supported Packard's idea. The theorem regards a one-dimensional time series as compressed information of a higher dimension. An univariate one-dimensional time series $x(t)$ for $t = 1, 2, 3...T$ can be extracted by extending $x(t)$ to a vector $X(t)$ in a m-dimensional space as follows:

$$X(t) = [x(t), x(t-\tau), x(t-2\tau), ..., x(t-(m-1)\tau)] \tag{5.2.1}$$

where m is the embedding dimension, τ is the delay constant. After reconstructing one-dimensional to a higher dimensional time series, the obtained time series is called a reconstructed time series. There are numbers of methods to calculate the attractor and the delay constant. For example, the correlation dimension method [134] is the most popular method for determining m, and the mutual information

[135] is a good method for the selection of τ. To the multivariate time series, it is difficult to apply the above methods, we present an accuracy-based method to estimate the embedding dimension and the delay time, progress shown in Section 5.3.2.

5.3 Case Study

Case studies are based on the recorded gas data from Nation Grid, UK. At first, the FNF-SVRLP based Standard Model (Section 5.3.3) is proposed for processing the whole gas dataset. Then, based on the results obtained from the Standard Model, it is found that the accuracy of the proposed Standard Model could be improved by considering the customer behavior difference within one week. We upgrade the Standard Model to the Advanced Model (Section 5.3.4), which includes three models (Mon. Model, Sat. Model, and other days of a week Model). Details would be described in following parts.

5.3.1 Data Description

The dataset used in the research is provided by National Grid. There is a data item explorer on National Grid's website [136], it provides the market participants and shippers with information relative to gas. After numbers of testings, following datasets which are obtained from the data item explorer are selected as our experimental data:

- NTS, National Transmission System (NTS) demand refers to the amount of gas used by gas consumers directly connected to the NTS and all local distribution zones (In UK, there are totally 13 local distribution zones. Shown in Fig. 5.1).
- CWV, Composite Weather Variable (CWV) is a single measure of weather for each local distribution zone, which takes into account not only temperature, but also wind speed, effective temperature and pseudo seasonal normal effective temperature.

Figure 5.1: UK Local Distribution Zones

Table 5.1: Dataset Description

Data set	Start Time Stamp	End Time Stamp	Description
A	01/01/2009	31/12/2012	Total dataset; 1461 days
B_1	01/01/2009	31/12/2011	Train dataset; 1095 days
B_2	01/01/2012	31/12/2012	Test dataset; 366 days

The NTS gas demand and 13 CWVs are recorded at one day interval. Time scale of the dataset is starting from Jan. 1st, 2009 to Dec. 31th, 2012 (Dateset A in Table 5.2), which covers four years. Then Dataset A are divided into two subsets, Dataset B_1 and Dataset B_2. We utilize the Dataset B_1 which contains the first three years(from Jan. 1st, 2009 to Dec. 31th, 2011) to develop a prediction model for NTS gas demand. Dataset B_2 (from Jan. 1st, 2012 to Dec. 31th, 2012) is used to test the prediction performance of the model learned from Dataset B_1.

For all performed models, we quantified the prediction performance with mean absolute percentage error (MAPE) and mean absolute error (MAE). They can be defined as

$$\mathrm{MAPE} = \frac{1}{N}\sum_{i=1}^{N}\frac{|\hat{y}_i - y_i|}{y_i} \times 100 \tag{5.3.1}$$

$$\mathrm{MAE} = \frac{1}{N}\sum_{i=1}^{N}|\hat{y}_i - y_i| \tag{5.3.2}$$

where N is the size of testing dataset, $\hat{y}_i$ is the forecasted gas demand, y_i is the actual gas demand , $\bar{y}$ is the mean of the actual gas demand, and i represents the test instance index.

5.3.2 Parameters Setting

Embedding Dimension and Delay Constant

In this section, the accuracy-based method is introduced to estimate the embedding dimension m and the delay time τ. We denote day i as D_i, then one day before D_i is D_{i-1} and one day after D_i is D_{i+1}. In order to forecast the NTS gas demand G_{i+1} of D_{i+1}, we summarize the prior-known variables in Table 5.2.

Table 5.2: Obtained Variables

	Target	Prior-known				
Day	D_{i+1}	D_i	D_{i-1}	D_{i-2}	...	D_1
NTS gas demand	G_{i+1}	G_i	G_{i-1}	G_{i-2}	...	G_1
Actual CWVs	-	$C_i(13)$	$C_{i-1}(13)$	$C_{i-2}(13)$	...	$C_1(13)$
Forecast CWVs	-	$C^f_{i+1}(13)$	-	-	...	-

Table 5.3: MAPE of different value of embedding dimension with $\tau = 1$

M	m	$\tau = 1$	MAPE
1	27	D_i	4.1%
2	41	D_i D_{i-1}	3.8%
3	55	D_i D_{i-1} D_{i-2}	3.5%
4	69	D_i D_{i-1} D_{i-2} D_{i-3}	3.4%

In Table 5.2, G_i is NTS gas demand of day D_i, $C_i(13)$ are the actual CWVs of 13 local distribution zones of day D_i and $C^f_{i+1}(13)$ are the forecasted CWVs of 13 local distribution zones of day D_{i+1}. All the above data can be obtained from National Grid's website. Then we apply SVRLP based SM model with Dataset B_1 to help us to estimate embedding dimension and delay constant. Results are presented in Table 5.3 and Table 5.4. The threshold MAPE value of 3.5% achieved in computation has been considered as a good quality result. A lower threshold value leads to more predictors. A large number of predictors can result in inferior performance of extracted models due to "the course of dimensionality" principle [137]. Therefore, the embedding dimension m is 55 and the delay time τ is 1.

5.3.3 Standard Model

The Standard Model is an unified model, which is applied to process the whole dataset. We denote the week j as W_j, then one week before W_j is W_{j-1} and one week after W_j is W_{j+1}. The input dataset for Standard Model is shown in Table 5.5.

Table 5.4: MAPE of different value of embedding dimension with $\tau = 2$

M	m	$\tau = 2$	MAPE
1	14	D_{i-1}	5.4%
2	28	D_{i-1} D_{i-3}	5.1%
3	42	D_{i-1} D_{i-3} D_{i-5}	4.7%
4	56	D_{i-1} D_{i-3} D_{i-5} D_{i-7}	4.3%

Table 5.5: Input dataset for Standard Model

Target	Input		
.	.	.	.
Mon.(W_{j+1})	Sun.(W_j)	Sat.(W_j)	Fri.(W_j)
Sun.(W_j)	Sat.(W_j)	Fri.(W_j)	Thu.(W_j)
Sat.(W_j)	Fri.(W_j)	Thu.(W_j)	Wen.(W_j)
Fri.(W_j)	Thu.(W_j)	Wen.(W_j)	Tue.(W_j)
Thu.(W_j)	Wen.(W_j)	Tue.(W_j)	Mon.(W_j)
Wen.(W_j)	Tue.(W_j)	Mon.(W_j)	Sun.(W_{j-1})
Tue.(W_j)	Mon.(W_j)	Sun.(W_{j-1})	Sat.(W_{j-1})
Mon.(W_j)	Sun.(W_{j-1})	Sat.(W_{j-1})	Fri.(W_{j-1})
Sun.(W_{j-1})	Sat.(W_{j-1})	Fri.(W_{j-1})	Thu.(W_{j-1})
.	.	.	.

Comparison among SM with ARMA, SM with SVRLP and SM with FNF-SVRLP are presented in following part.

Results

In order to evaluate the performance of the proposed FNF-SVRLP, comparison with ARMA and SVRLP are conducted. The comparison is shown in Table 5.6 and Fig. 5.2, and we can see that SVRLP is much better than AMRA, and FNF-SVRLP is the best among of all three methods. The MAPE of FNF-SVRLP is 3.8%, it

Table 5.6: MAPE and MAE of different methods with the Standard Model

	MAPE	MAPE Impro.	MAE	MAE Impro.
FNF-SVRLP	3.8%	-	9.2	-
SVRLP	4.4%	13.7%	10.4	11.5%
ARMA	6.1%	27.9%	13.6	23.5%

improves the MAPE over ARMA method and SVRLP method by 27.9% and 13.7%. Moveover, the MAE of FNF-SVRLP is 9.2 mcm (million cubic meters), which is smaller than others. Using the MAE of SVRLP subtract the MAE of FNF-SVRLP, the result we achieved is about 1.2 mcm, which means that the FNF-SVRLP could save 1.2 mcm more gas demand per day than SVRLP. Fig. 5.3 and Fig. 5.4 show the MAPE and MAE of every day of the week (Monday-Sunday) during the testing period. These results confirm the superiority of the FNF-SVRLP. In addition, the MAPE and MAE of the whole testing data of 12 months are calculated, shown in Table 5.6. From this table, it can be seen that the performance of SVRLP with FNF is much better than the SVRLP. The aggregated distribution of MAPE's is illustrated in Fig. 5.5. From the above results, it clearly indicates that the accuracy of the SVR based local predictor and the importance of filtering false neighbors. Through filtering false neighbors, to leave the optimal neighbors, and using those optimal neighbors as training samples improves the accuracy.

5.3.4 Advanced Model

According to results obtained from the Standard Model (Section 5.3.3), it can be found that the accuracy of Saturday and Monday are always lower than the other days with a week (Shown in Fig. 5.3 and Fig. 5.4). In UK, Monday is the starting of the weekday, and Saturday is the starting of the weekend, so that the customer behavior of the two days will be different from the other days. However, the customer behavior is an unmeasurable factor, in order to eliminate its impact, we presented the Advanced Model, which includes Mon. model (only contains the dataset relative to Mon.), Sat. model (only contains the dataset relative to Sat.) and other days

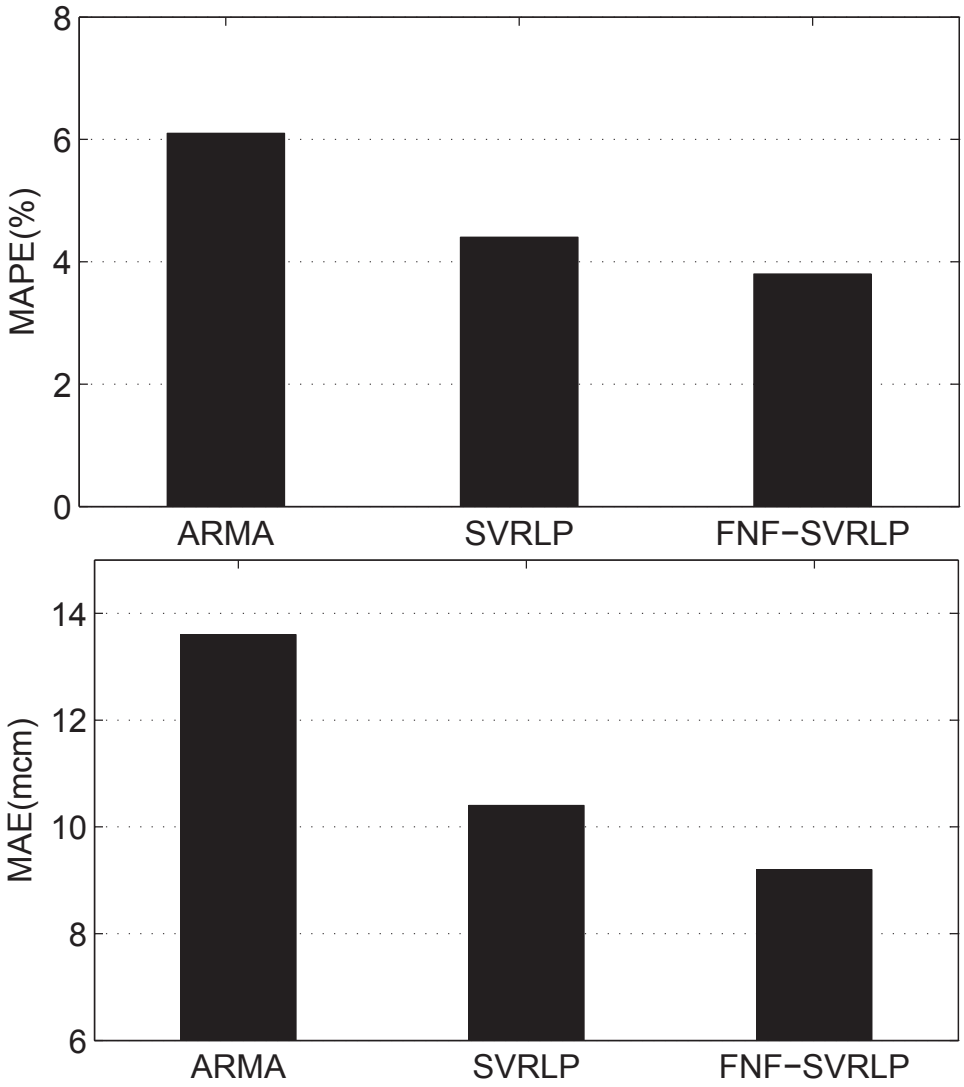


Figure 5.2: Comparison of FNF-SVRLP method and other methods

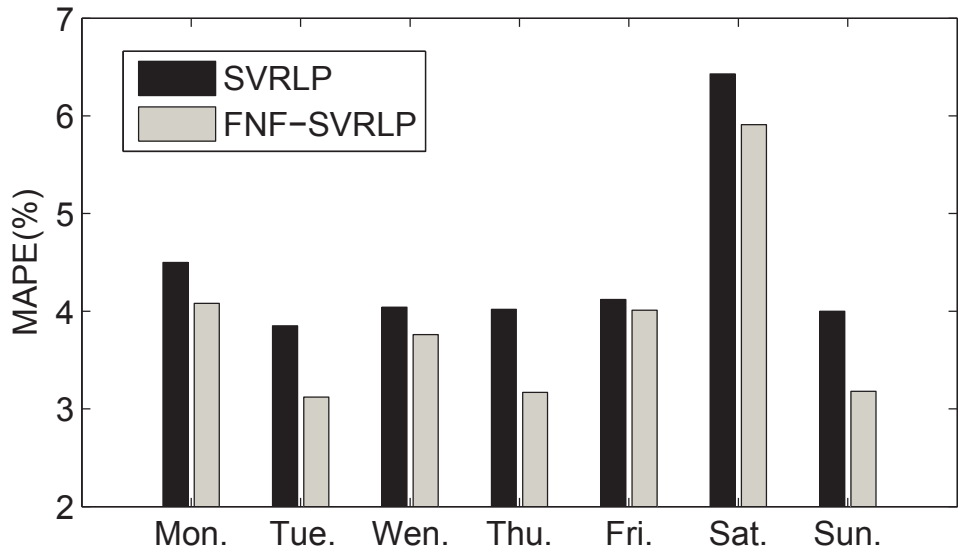


Figure 5.3: Average prediction MAPE of every day of the week during the whole testing period

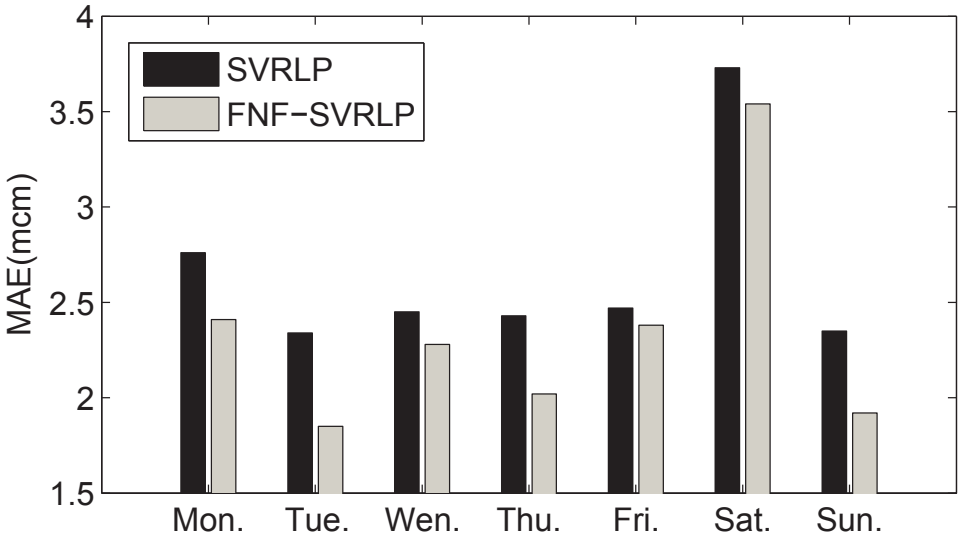


Figure 5.4: Average prediction MAE of every day of the week during the whole testing period

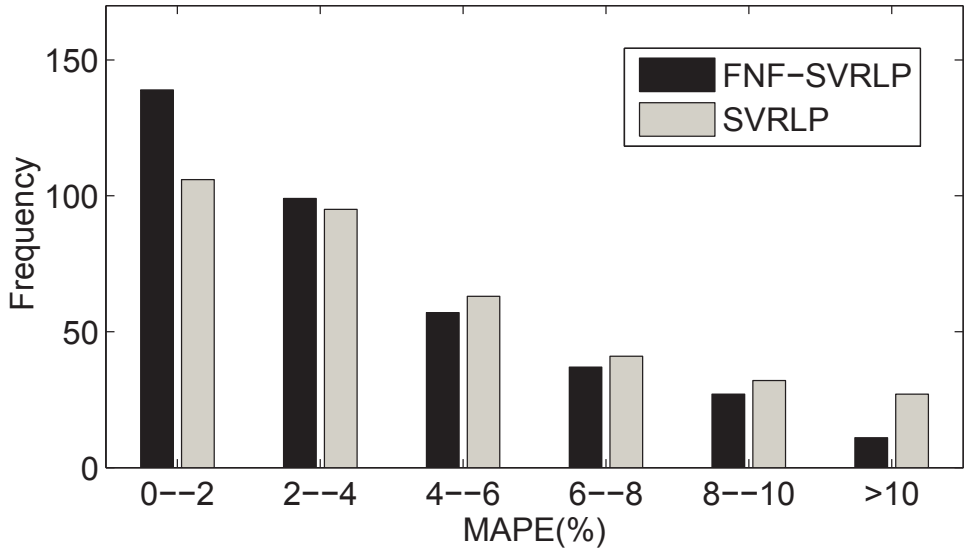


Figure 5.5: The aggregated distribution of MAPE's for FNF-SVRLP and SVRLP during the whole testing period.

Table 5.7: Average prediction MAPE and MAE of every month during the whole testing period

Month	MAPE(%)		MAE(mcm)	
	SVRLP	FNF-SVRLP	SVRLP	FNF-SVRLP
1	3.2	3.0	9.8	9.2
2	5.1	4.3	17.0	14.8
3	2.9	2.5	7.6	6.5
4	4.2	3.9	10.4	9.6
5	3.6	3.1	7.8	6.8
6	4.9	3.7	8.9	6.7
7	4.0	3.6	7.2	6.4
8	3.9	3.7	6.7	6.3
9	8.8	7.1	12.2	9.8
10	7.5	6.6	16.9	14.8
11	2.6	2.4	7.1	6.4
12	4.8	4.5	14.3	13.4

Table 5.8: Input dataset for Mon. Model

Target	Input		
.	.	.	.
Mon.(W_{j+1})	Sun.(W_j)	Sat.(W_j)	Fri.(W_j)
Mon.(W_j)	Sun.(W_{j-1})	Sat.(W_{j-1})	Fri.(W_{j-1})
Mon.(W_{j-1})	Sun.(W_{j-2})	Sat.(W_{j-2})	Fri.(W_{j-2})
.	.	.	.

Table 5.9: Input dataset for Sat. Model

Target	Input		
.	.	.	.
Sat.(W_{j+1})	Fri.(W_{j+1})	Thu.(W_{j+1})	Wen.(W_{j+1})
Sat.(W_j)	Fri.(W_j)	Thu.(W_j)	Wen.(W_j)
Sat.(W_{j-1})	Fri.(W_{j-1})	Thu.(W_{j-1})	Wen.(W_{j-1})
.	.	.	.

model (contains the remaining exclude Mon. and Sat.), shown in Table 5.8-5.10, respectively. Comparison between SVRLP based AM and FNF-SVRLP based AM are presented in the results part.

Table 5.10: Input dataset for Other days Model

Target	Input		
.	.	.	.
Tue.(W_{j+1})	Mon.(W_{j+1})	Sun.(W_j)	Sat.(W_j)
Sun.(W_j)	Sat.(W_j)	Fri.(W_j)	Thu.(W_j)
Fri.(W_j)	Thu.(W_j)	Wen.(W_j)	Tue.(W_j)
.	.	.	.

Table 5.11: Prediction results of Standard Model and Advanced Model with different methods under the same testing gas dataset

	MAPE(%)		MAE(mcm)	
	SVRLP	FNF-SVRLP	SVRLP	FNF-SVRLP
Standard Model	4.4	3.8	10.4	9.2
Advanced Model	4.2	3.6	9.7	8.6

Results

The results obtained from Standard Model show that the FNF-SVRLP provides significantly better forecast results than ARMA and SVRLP. However, it is found that the MAPE of Mon. and MAPE of Sat. are the top two highest MAPE among the seven days as well as MAE. Considering the customer behavior of the two days would be different from the other days, we presented a three-models combined Advanced Model to overcome this problem. In Table 5.11, it contains the prediction results of Standard Model and Advanced Model using two different methods. It shows that no matter which method is applied, the predicted results of Advanced Model are always much better than the Standard Model. Moveover, the combination of the FNF-SVRLP method and the Advanced Model outperforms the other combinations. From Fig. 5.6 and Fig. 5.7, it can be clearly seen that there are some improvements on the accuracy of seven days, especially on the Monday and Saturday. From above results, it shows the Advanced Model performs much better than the Standard Model, which indicates that the customer behavior can affect the forecasting performance.

5.4 Web server based implematation

National Grid produces NTS demand forecast and LDZs demand forecast, which is published on Nation Grids Information Exchange website to inform the market participants and shippers the demand. Through the conventional National Grid's prediction model, Nation Grid generates an individual day-ahead gas demand forecast for 13 LDZs and for each NTS direct customer, such as the separate power

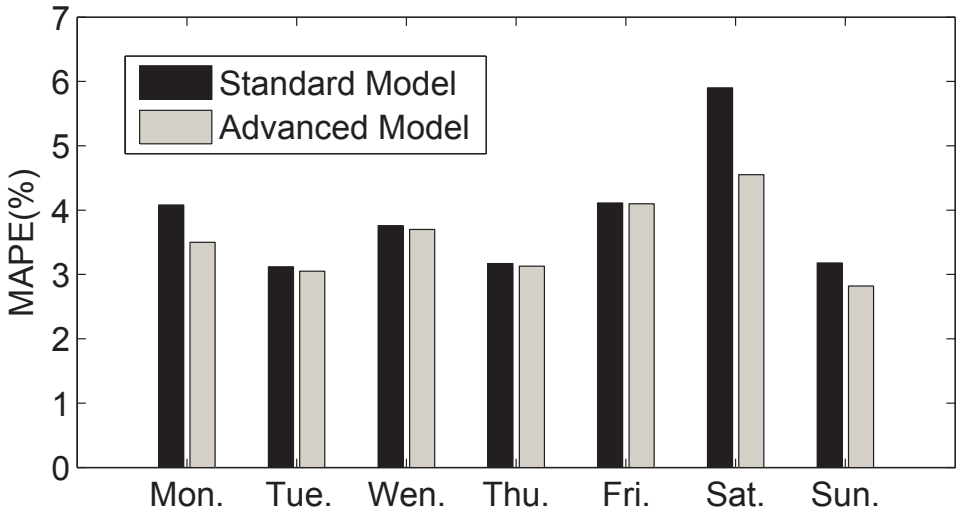


Figure 5.6: Comparison of average prediction MAPE of every day of the week between Standard Model and Advanced Model by using FNF-SVRLP method.

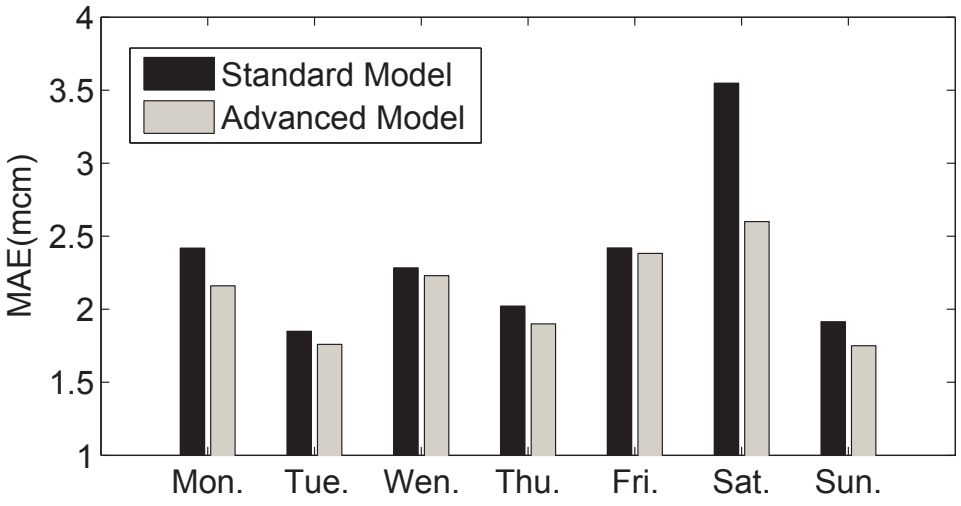


Figure 5.7: Comparison of average prediction MAE of every day of the week between Standard Model and Advanced Model by using FNF-SVRLP method.

station, industry load, etc, based on 13 Composite Weather Variables. The day-ahead natural gas demand is delivered to the website at 13:00 each day. Our online Liverpool Natural Gas Demand Prediction System employs the FNF-SVRLP based Advanced Model. Compared with the conventional model (National Grid's model), our online prediction model not only reduced the forecast complexity but also increased the accuracy.

5.4.1 Background of online prediction

The operation server

The server is purchased from DELL, with four 3.10GHz Intel Xeon CPU and 8 GB of RAM.

The basic software of the online prediction system

Windows Server 2012 Windows Server 2012 is full-featured 64-bit version of Windows Server that enables core IT resources, such as file and print sharing, remote access, and security. It provides a network foundation from which the user can centrally manage settings on computers that are based on the Windows operating system, and upon which the user can run the most popular business applications. It also provides a familiar Windows user experience that helps to manage users and safeguard business information. Because Windows Server 2012 comes pre-installed with the server hardware, we do not need to separately obtain and then install the hardware and operating system. Windows Server 2012 Foundation is supported by an extensive network of certified professionals who can to provide service for the Windows Server network. By design, Windows Server 2012 Foundation uses the Windows Server Catalog.

APS.net ASP.NET is an open source server-side Web application framework designed for Web development to produce dynamic Web pages. It was developed by Microsoft to allow programmers to build dynamic web sites, web

applications and web services. It was first released in January 2002 with version 1.0 of the .NET Framework, and is the successor to Microsoft's Active Server Pages (ASP) technology. ASP.NET is built on the Common Language Runtime (CLR), allowing programmers to write ASP.NET code using any supported .NET language. The ASP.NET SOAP extension framework allows ASP.NET components to process SOAP messages. ASP.NET is in the process of being re-implemented as a modern and modular web framework, together with other frameworks like Entity Framework.

Visual Studio Professional 2013 Visual Studio is an integrated development environment (IDE) from Microsoft. It is used to develop computer programs for Microsoft Windows, as well as web sites, web applications and web services. Visual Studio uses Microsoft software development platforms such as Windows API, Windows Forms, Windows Presentation Foundation, Windows Store and Microsoft Silverlight. It can produce both native code and managed code. Visual Studio includes a code editor supporting IntelliSense as well as code refactoring. The integrated debugger works both as a source-level debugger and a machine-level debugger. Other built-in tools include a forms designer for building GUI applications, web designer, class designer, and database schema designer. It accepts plug-ins that enhance the functionality at almost every levelłincluding adding support for source-control systems (like Subversion) and adding new toolsets like editors and visual designers for domain-specific languages or toolsets for other aspects of the software development lifecycle (like the Team Foundation Server client: Team Explorer). Visual Studio supports different programming languages and allows the code editor and debugger to support (to varying degrees) nearly any programming language, provided a language-specific service exists. Built-in languages include C, C++ and C++/CLI (via Visual C++), VB.NET (via Visual Basic .NET), C ♯(via Visual C), and F ♯ (as of Visual Studio 2010). Support for other languages such as M, Python, and Ruby among others is available via language services installed separately. It also supports XML/XSLT, HTML/XHTML, JavaScript and CSS.

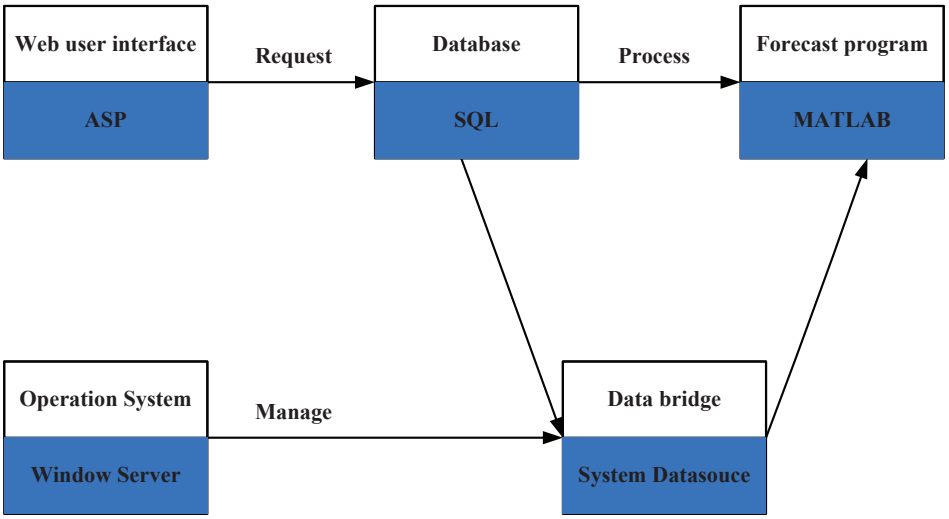


Figure 5.8: The design of the online prediction system

MATLAB 2012 MATLAB (matrix laboratory) is a multi-paradigm numerical computing environment and fourth-generation programming language. Developed by MathWorks, MATLAB allows matrix manipulations, plotting of functions and data, implementation of algorithms, creation of user interfaces, and interfacing with programs written in other languages, including C, C++, Java, and Fortran. Although MATLAB is intended primarily for numerical computing, an optional toolbox uses the MuPAD symbolic engine, allowing access to symbolic computing capabilities. An additional package, Simulink, adds graphical multi-domain simulation and Model-Based Design for dynamic and embedded systems.

5.4.2 Design of the online prediction system

The Fig. 5.8 shows the briefly design of the online prediction system.

5.4.3 Outline of the online prediction system

The engineers of National Grid in the National Grid House, Warwick, can easily connect to our server (Liverpool) via our online prediction system. The engineers

just need enter the current data, the system will do the calculation on the background, and then shows the final predicted results on new page directly.

5.4.4 Overall performance of the online prediction system

The following results are obtained from the online recording system, the recording system contain all information about the data, the input data, the out data and so on. The presented time period is from 01/01/2013 to 31/12/2013.

To the NTS

Due the Non Disclosure Agreement with the National Grid, only the sample results can be presented. The results are shown in Fig. 5.10-5.14. Fig. 5.10 is the one day predicted results. Fig. 5.11 is the two days predicted results. Fig. 5.12 is the one day predicted results. Fig. 5.13 is the three days predicted results. Fig. 5.14 is the four days predicted results. Fig. 5.15 is the five days preditied results.

To the LDZs

To the LDZs, in UK, there are totally 13 LDZs. The performance for different LDZs are shown in Fig. 5.15-5.20.

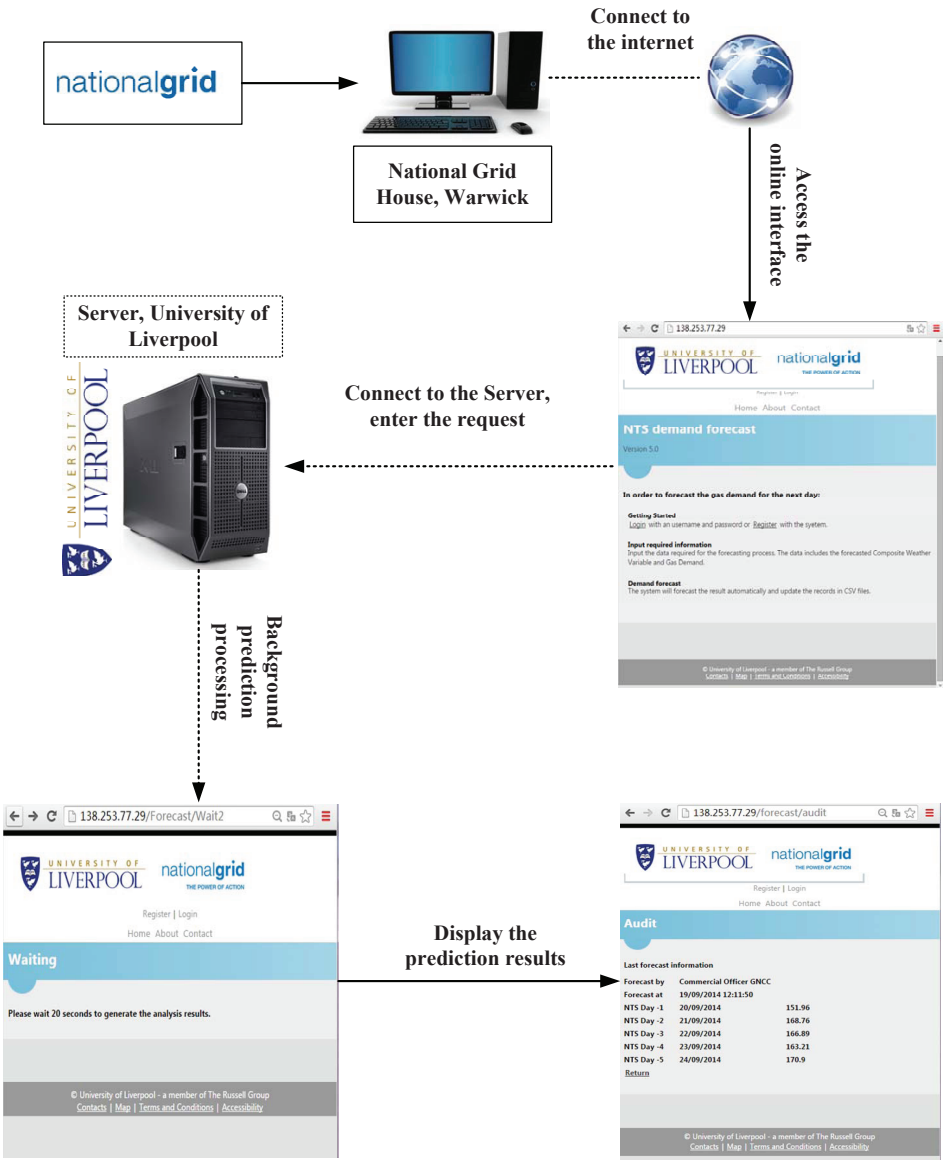


Figure 5.9: The outline of the online prediction system

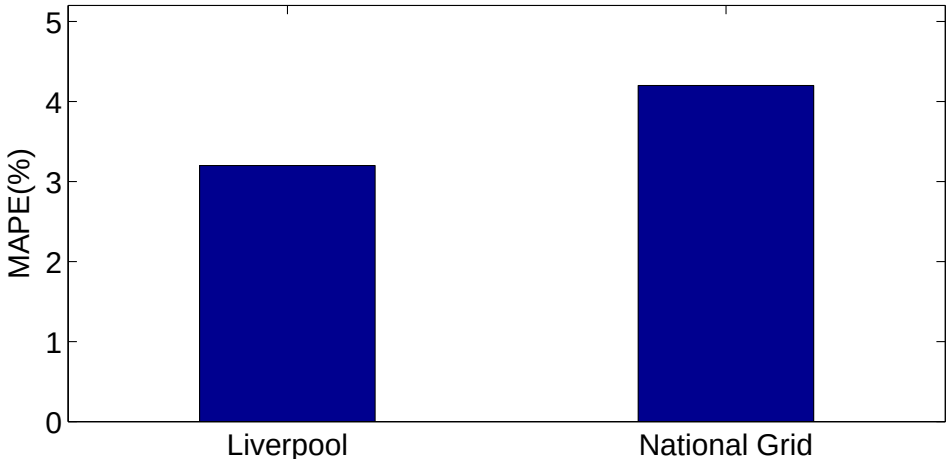


Figure 5.10: Comparsion between Liverpool online model and National Grid model for one day ahead prediton of NTS natual gas demand

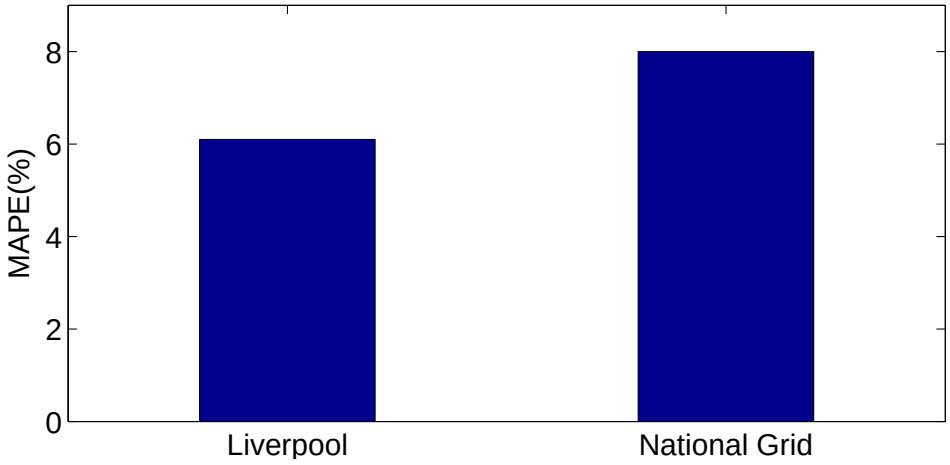


Figure 5.11: Comparsion between Liverpool online model and National Grid model for two days ahead prediton of NTS natual gas demand

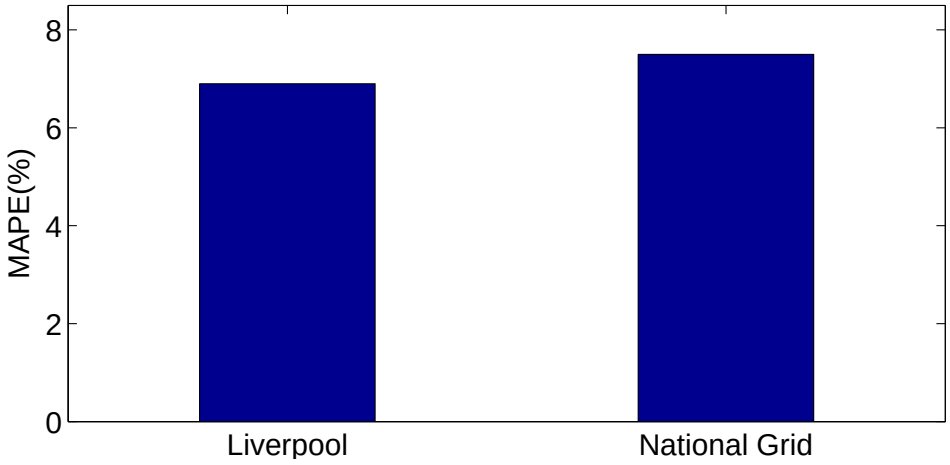


Figure 5.12: Comparsion between Liverpool online model and National Grid model for three days ahead prediton of NTS natual gas demand

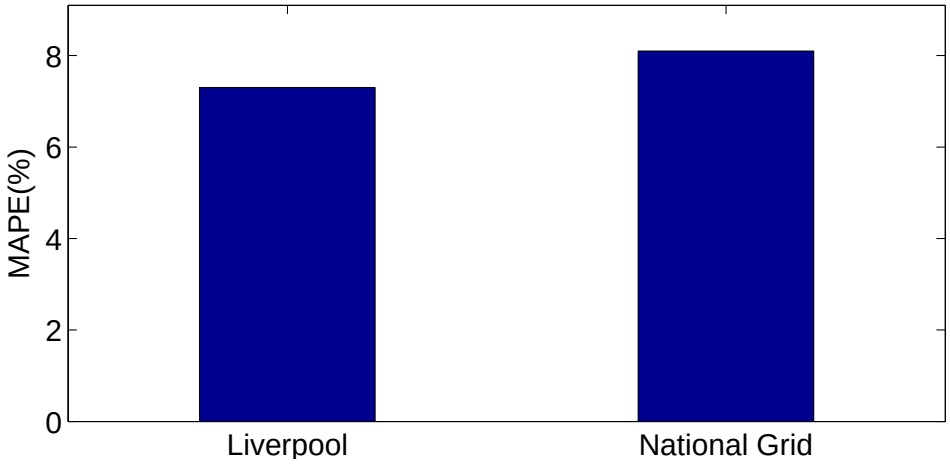


Figure 5.13: Comparsion between Liverpool online model and National Grid model for four days ahead prediton of NTS natual gas demand

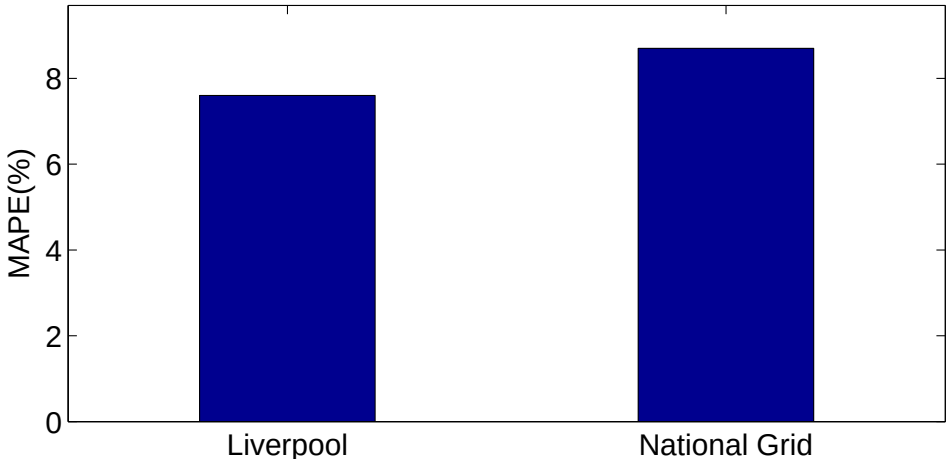


Figure 5.14: Comparsion between Liverpool online model and National Grid model for two days ahead prediton of NTS natual gas demand

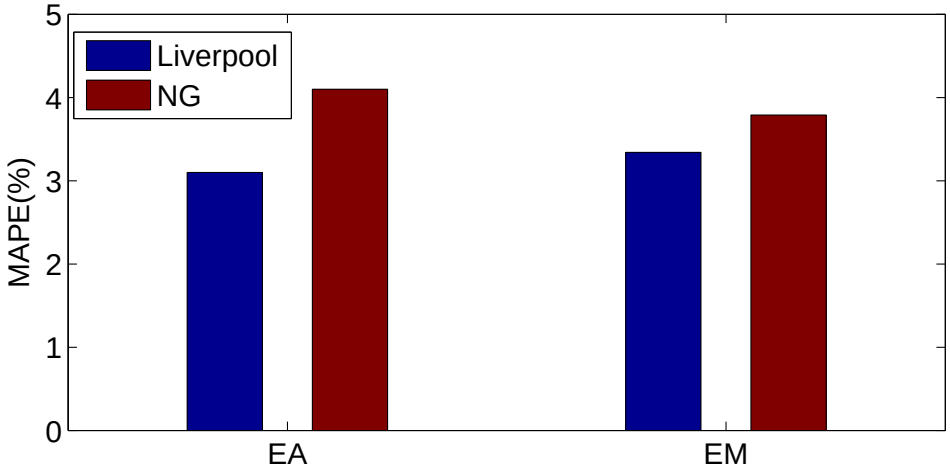


Figure 5.15: Comparsion between Liverpool online model and National Grid model for two LDZs gas demand prediction

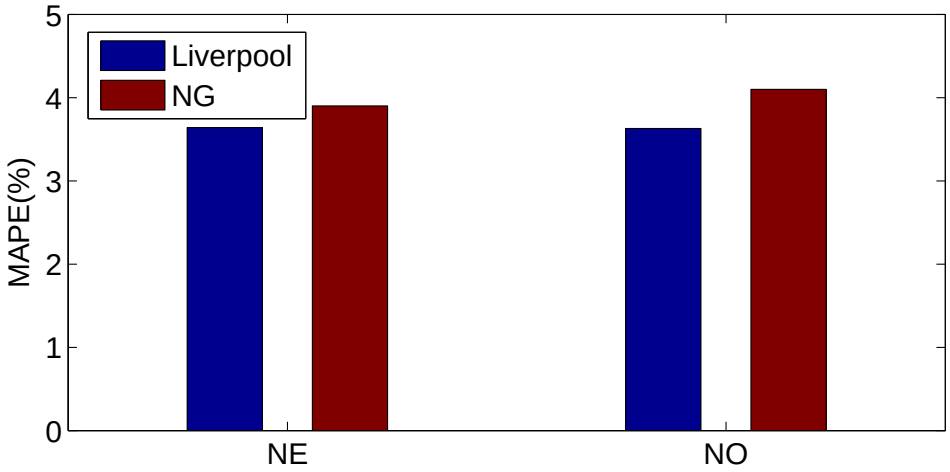


Figure 5.16: Comparsion between Liverpool online model and National Grid model for two LDZs gas demand prediction

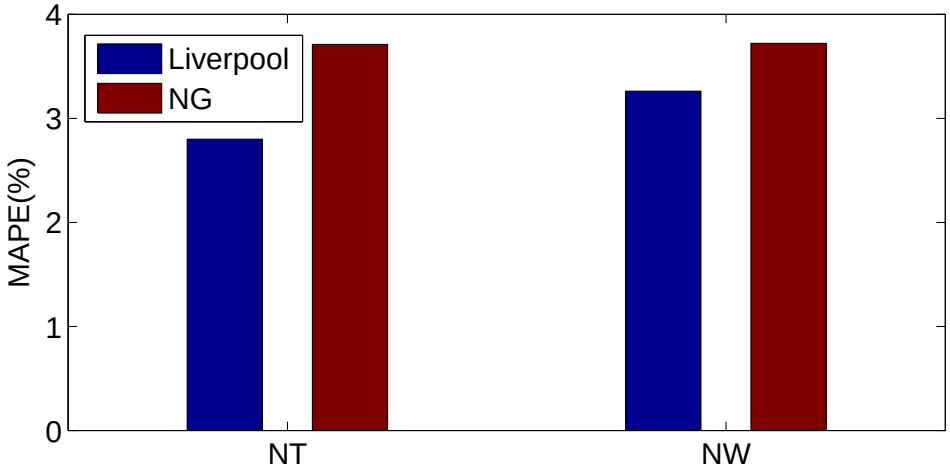


Figure 5.17: Comparsion between Liverpool online model and National Grid model for two LDZs gas demand prediction

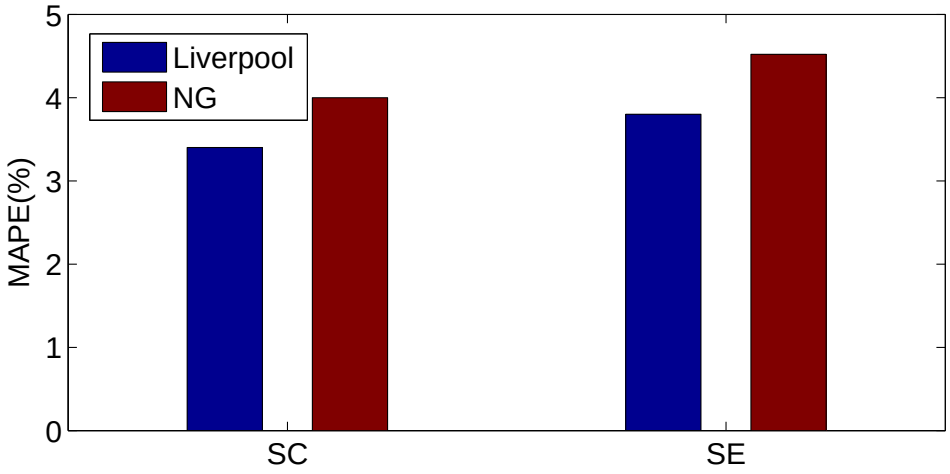


Figure 5.18: Comparsion between Liverpool online model and National Grid model for two LDZs gas demand prediction

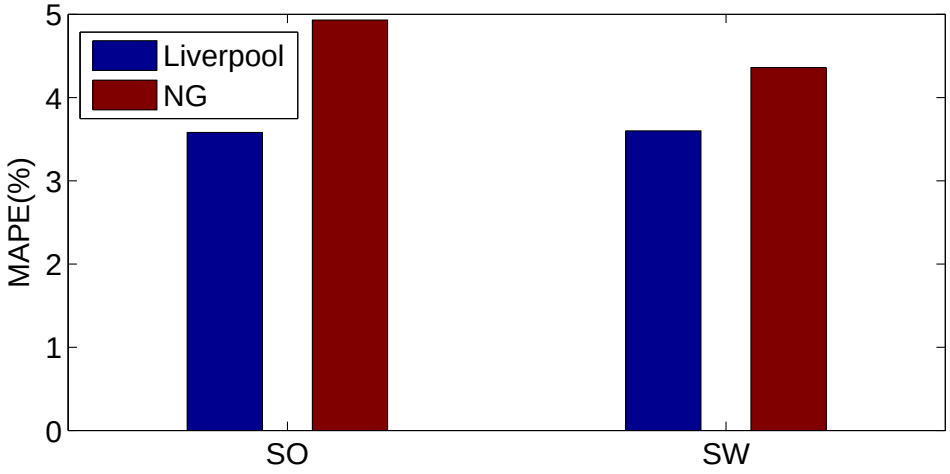


Figure 5.19: Comparsion between Liverpool online model and National Grid model for two LDZs gas demand prediction

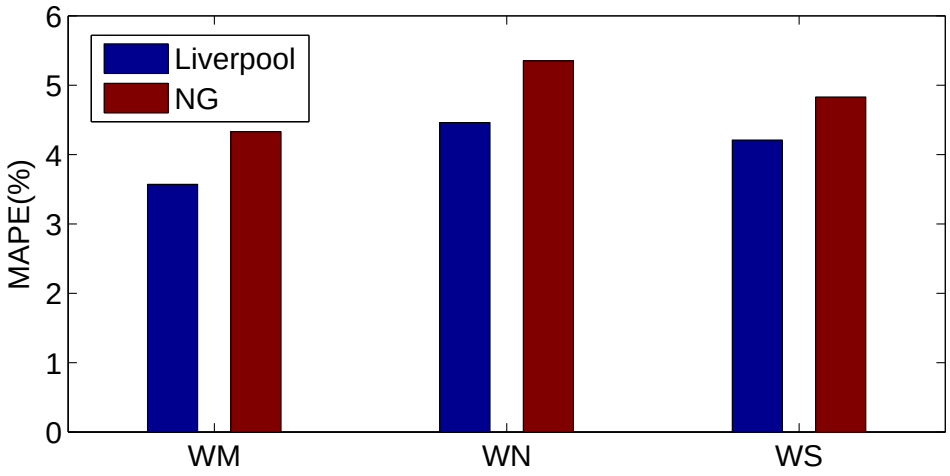


Figure 5.20: Comparsion between Liverpool online model and National Grid model for two LDZs gas demand prediction

5.5 Conclusion

In this chapter, we have presented a FNF-SVRLP based Advanced Model for short-term gas demand prediction. The forecasted gas demand can aid the market in making efficient decisions in balancing supply and demand, and reducing costs. The FNF-SVRLP combines the SVR based local predictor with a false-neighbors filter. During the computation procedure of the gas demand forecasting, the embedding dimension and the time delay are computed firstly through the accuracy-based method, then the phase space reconstruction is applied to process the gas dataset. In the local predictor, there are some difference between the SVRLP and the proposed FNF-SVRLP. Compared with the SVRLP, the FNF-SVRLP not only applies the Euclidean distance but also the exponential separation rate to verbify the nearest neighbors, and only the optimal nearest neighbors can be selected as training samples. The final results presented demonstrate that the FNF-SVRLP can achieve a higher prediction accuracy than that of the SVRLP method and ARMA method using the same real world gas data. Moreover, the study on the customer behavior of different day has been tested, and results indicate that the customer behavior can affect the final forecasted performance. In addition, the FNF-SVRLP based Advanced Model is applied to built the online natural gas demand prediction system for

National Grid, and overall obtained results outperform the National Grid's model.

Chapter 6

Mathematical Morphology-based Local Predictor with Support Vector Regression

6.1 Introduction

In previous chapters, we have introduced that the Support Vector Regression (SVR) which derived from the Support Vector Machine (SVM), can be utilized to achieve nonlinear mapping from sample space to the feature space through kernel functions. Since the SVR can make a better performance for chaotic time series prediction [127], it has already been utilized to predict the wind power (Chapter 4) and natural gas demand(Chapter 5). We also presented that the local methods can perform much better than global methods when doing the chaotic time series prediction [104]. Each predicting point of local methods has its own model constructed by its nearest neighbors which were found in the neighborhood of the phase space reconstructed from the time series, and the reliability of the constructed model would affect the final predicted performance. However, the reliability of the constructed model is sensitive to the internal regularity between predicting point and its nearest-neighbors. What' more, the datasets, such as wind power, wind speed and gas demand, which are non-stationary, the internal regularity between predicting point and

its nearest-neighbors are much more complex than the stationary dataset. In order to improve the identifiable accuracy of the regularity of local predictor, the morphological filter is applied to decompose the non-stationary into several subsequences, ranked form the low frequency subsequence to the high frequency subsequence. Through this way, the local predictor could capture the non-stationary dataset more accurate, and improve the final performance of prediction.

In this chapter, a mathematical morphology based local predictor with a proven powerful algorithm, which is the SVR combined with space reconstruction of time series, is introduced. Dataset is decomposed into several subsequences through the mathematical morphology. Then, a SVR algorithm is developed as a local predictor is applied to do the subsequences prediction. The forecasting results of each subsequences are summarized to achieve the final predicted result. The resulting predictor is referred as SVR based local predictor with mathematical morphology (SVRLP-MM). At the final, a real-world wind speed data which collected from Coffey wind farm is used to evaluated the proposed prediction method. And results demonstrated that the predicted performance of SVRLP-MM method is better than SVRLP method and the autoregressive moving average (ARMA) method with the same real-world wind data.

6.2 Basic concepts of mathematical morphology

When first introduced, morphology was based on the set theory, which meant it was limited to quantitatively describing shape and size for binary images. In 1978, Nakagawa and Rosenfeld linked the binary dilation and erosion operations to the maximum and minimum filter applied to grey-value images[141]. The notion of umbra was also introduced by Sternberg around 1978, which made it possible to apply all mathematics developed for binary images to grey-scale images[142]. As a result of the pioneering work, MM has achieved the status of a powerful tool image processing with applications in many area, and later on the MM is widely used in signal processing area. Given a set G defined in $(\mathbb{E})^d$, where $(\mathbb{E})^d$ is d-dimensional Euclidean space[143], the following definitions can be developed [144]:

- Translation: The translation of G by a vector x, denote as G_x, is defined as:

$$G_x = \{g : g = a + x; a \in G\} \tag{6.2.1}$$

- Reflection: The reflection of G based on its origin, denoted as $\check{G}$, is defined as:

$$\check{G} = \{-g : g \in G\} \tag{6.2.2}$$

- Complement: The complement of G, denoted as G^c, is defined as:

$$G^c = \{g : g \notin G\} \tag{6.2.3}$$

Letting two sets, $A, B \subseteq (\mathbb{E})^d$ represents the binary input set and the binary SE respectively, the dilation of A by the SE B can be developed as $A \bigoplus B$ with the definition of [144]:

$$A \oplus B = \bigcap_{b \in B} (A + b) \tag{6.2.4}$$

The oppsite of dilation is known as erosion. The same as dilation, given two sets $A, B \subseteq \mathbb{E}^d$, A is the input set and B is the binary SE. The erosion can be developed as $A \oplus B$, its definition is:

$$A \ominus B = \bigcap_{b \in B} (A - b) \tag{6.2.5}$$

Generally speaking, A is the image being processed, while B servers as an SE that slides as probe across image A and interacts with each pixel of A. Obviously, the size of B should be much smaller than that of A. To have a clear view of this process, we give an example in Fig. 6.1 to show how dilation and erosion function between a binary image and an SE. Here, the origin of B is set at (0,0). Figure 6.1 illustrates an important property of dilation and erosion - duality, which means that applying dilation to A is equivalent to apply erosion to its complement A^c. This property can be expressed as:

$$\mathrm{A} \oplus \mathrm{B} = (\mathrm{A}^c \ominus \mathrm{B}^c) \tag{6.2.6}$$

$$\mathrm{A} \ominus \mathrm{B} = (\mathrm{A}^c \oplus \mathrm{B}^c) \tag{6.2.7}$$

Proof:

$$\begin{aligned} \mathrm{A}^c \oplus \mathrm{B} &= \bigcup_{b\in B} (I-A)_b = \bigcup_{b\in B} (I-A)_{-b} \\ &= I - \bigcap_{b\in B} (A)_{-b} \\ &= I - A \ominus B = (A \ominus B)^c \\ \Rightarrow A \ominus B &= (A^c \oplus B)^c \end{aligned}$$

The property of duality illustrates that the processing of dilation and erosion is not reversible and there is no inverse transform for the operators. As we can see from the following sections, applying dilation and erosion alternately actually produces a pair of new operations. Another property of dilation and erosion is the distributivity:

$$A \ominus B \ominus C = A - (B \oplus C) \tag{6.2.8}$$

Proof:

$$\begin{aligned} A \ominus B \ominus C &= (A^c \oplus B)^c \ominus C \\ &= [(A^c \oplus B) \oplus C]^c \\ A \ominus (B \oplus C) &= [A^c \oplus (B \oplus C)]^c \end{aligned}$$

$$\begin{aligned} &\because (A^c \oplus B) \oplus C = A^c \oplus (B \oplus C) \\ &\therefore A - B - C = A - (B \oplus C) \end{aligned}$$

6.2.1 Basic Morphological Operators

Set Representations of Functions

In order to extend morphological operators to functions, the functions are represented by their umbra [146], which is defined as:

$$U(f) = \{(x, a) | a \leq f(x)\} \tag{6.2.9}$$

Hence, a d-dimensional function $f(x)$ is represented by a $(d+1)$-dimensional set. Obviously, the umbra is the set of points below the surface represented by $f(x)$. After getting the umbra, binary morphological operators can be applied to the signal.

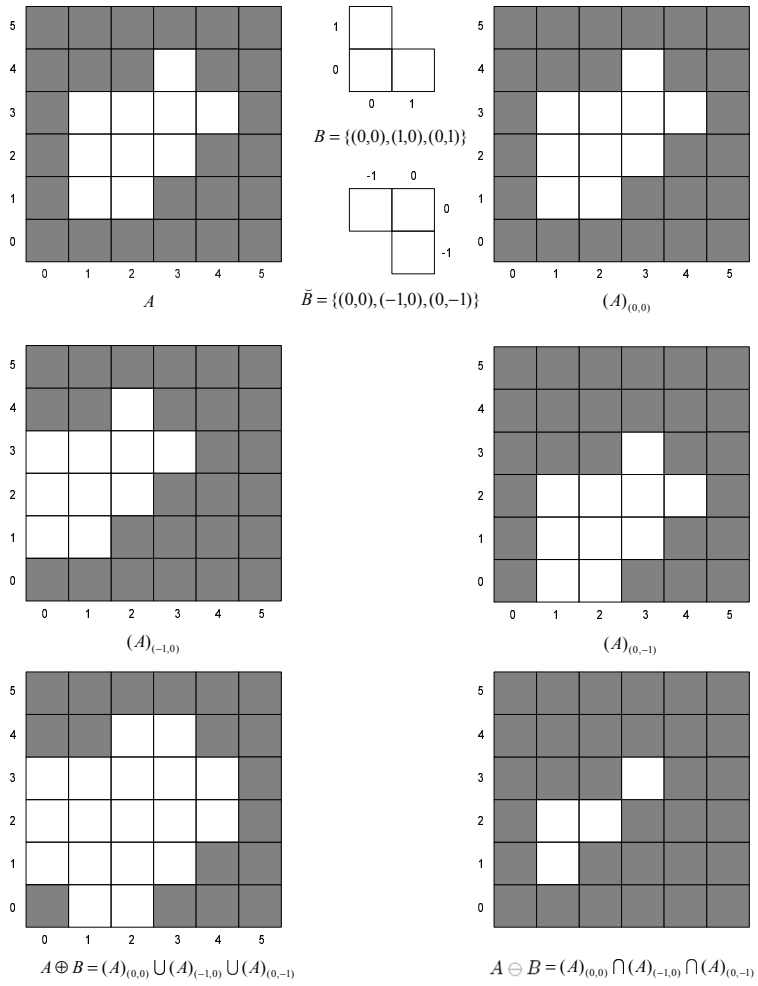


Figure 6.1: Binary dilation and erosion of a binary image

In general, the umbra set extends to $a = -\infty$, and the function f can be reconstructed from its umbra since:

$$f(x) = \max\{a|(x,a) \in U(f)\}, \forall x \tag{6.2.10}$$

Figure 6.2 shows, as an example, the umbra of a sinusoidal function, where the umbra of f(x) is the shaded region. We can easily show that $f \leq g \Leftrightarrow U(f) \subseteq U(g)$. Some definitions for grey-scale operations based on sets are defined as follows [147]:

- Grey-scale union: The union of two functions f and g, denoted by $f \bigvee g$, is defined as:

$$(f \bigvee g)(x) = f(x) \bigvee g(x). \tag{6.2.11}$$

  There is a one-to-one correspondence between the union of functions and the set union:

$$U(f \bigvee g) = U(x) \bigcup U(g). \tag{6.2.12}$$

- Grey-scale intersection: The intersection of two functions f and g, denoted by $f \bigwedge g$, is defined as:

$$(f \bigwedge g)(x) = f(x) \bigwedge g(x). \tag{6.2.13}$$

  A similar one-to-one correspondence exists for the function and the set intersection:

$$U(f \bigwedge g) = U(x) \bigcap U(g). \tag{6.2.14}$$

- Grey-scale transpose: The transpose $\check{f}$ of a function f is defined as:

$$\check{f}(x) = f(-x). \tag{6.2.15}$$

- Grey-scale complement: The complement f^c of a function f is defined as:

$$f^c(x) = -f(x.) \tag{6.2.16}$$

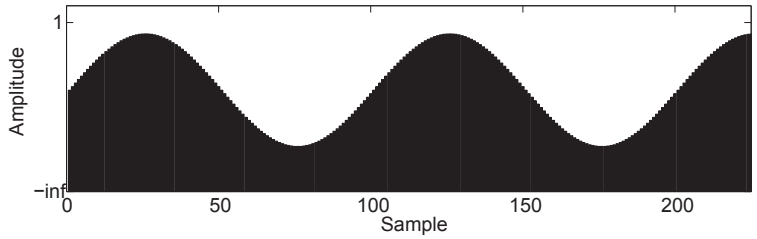


Figure 6.2: Umbrae $U(f)$ of a sinusoidal function f

For an eight-bit grey-scale digital image, each pixel of which can have $2^8 = 256$ possible values to carry the full and only information about its intensity, the image can be represented as a set whose components are in $\mathbb{Z}^3$. In this case, two components of each element of the set refer to the coordinates of a pixel, and the third corresponds to its discrete intensity value. For a signal, the set is defined in $\mathbb{Z}^2$ with each element corresponding to a sample of the digitised signal. Similarly, the first component of each element represents the coordinate and the second represents its value. Sets in a higher dimensional space can contain other attributes, such as the colour information of an image.

Grey-Scale Dilation and Erosion

In order to use MM in signal processing where most signals are not binary, morphological operators should be extended to a grey-scale level. Instead of performing dilation and erosion by union and intersection as in the binary case, they are performed by algebraic addition and subtraction in the grey-scale case. Let f denote a signal and g denote an SE, and the length of g be considerably shorter than that of f [149]. Dilation and erosion are defined as follows:

$$f \oplus g(x) = \max_s \{f(x+s) + g(s) | (x+s) \in D_f, s \in D_g\} \tag{6.2.17}$$

$$f \ominus g(x) = \max_s \{f(x+s) - g(s) | (x+s) \in D_f, s \in D_g\} \tag{6.2.18}$$

where D_f , D_g are the definition domains of f and g, respectively. For example, suppose the SE, g, has a length of five samples with its origin in the middle. In this

case, the domain of g is given by $D_g = -2, -1, 0, 1, 2$. The dilation and erosion of f by g are therefore calculated from

$$f \oplus g(x) = \max\{f(x-2) + g(-2), f(x-1) + g(-1), f(x) + g(0), \\ f(x+1) + g(1), f(x+2) + g(2)\}$$

and

$$f \ominus g(x) = \min\{f(x-2) - g(-2), f(x-1) - g(-1), f(x) - g(0), \\ f(x+1) - g(1), f(x+2) - g(2)\}$$

respectively. Intuitively, dilation can be imagined as swelling or expanding, while erosion can be thought of as a shrinking procedure.

As explained previously, the SE is a small set used to probe the signal under study. A simple case is that the SE has the form of $g(s) \equiv 0, s \in D_g$, which is referred to as a 'flat SE. Hence, definitions of dilation and erosion degrade to:

$$f \oplus g(x) = \max_s\{f(x+s) + g(s) | (x+s) \in D_f, s \in D_g\} \tag{6.2.19}$$

$$f \ominus g(x) = \min_s\{f(x+s) + g(s) | (x+s) \in D_f, s \in D_g\} \tag{6.2.20}$$

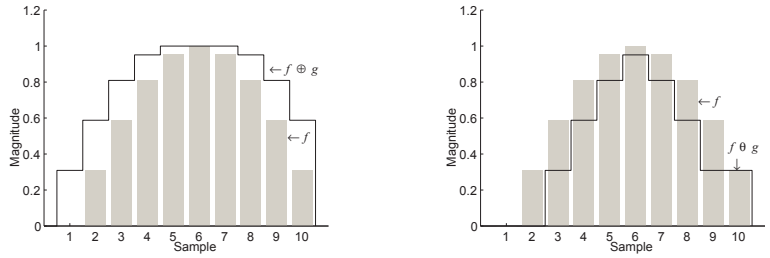
The function of g is to indicate which samples are involved when processing the current sample. For a binary signal, the SE, g, must be flat. The dilation and erosion of a one-dimensional signal are illustrated in Figs. 6.3a and b, respectively. Both operations use a flat SE of length 3: $g(-1) = g(0) = g(1) = 0$.

Figure 6.3 demonstrates another property of dilation and erosion, i.e. that they are increasing transforms. The property can be expressed as follows. For two signals, f_1 and f_2, and an arbitrary SE, g, we have:

$$f_1 \leq f_2 \Rightarrow \begin{cases} f_1 \oplus g \leq f_2 \oplus g \\ f_1 \ominus g \leq f_2 \ominus g \end{cases}. \tag{6.2.21}$$

The ordering relation between dilation and erosion can be expressed as the erosion of a signal by an SE being less than or equal to its dilation by the same SE: $f \ominus g \leq f \oplus g$. If the SE contains its origin, which means processing a sample of the signal within a window that contains the sample, the following ordering exists:

$$f \ominus g \leq f \leq f \oplus g. \tag{6.2.22}$$



(a) (b)

Figure 6.3: Grey-scale dilation and erosion of an one-dimensional signal. (a) Dilation. (b) Erosion.

6.2.2 Morphological Filters

Definitions of Morphological Filters

Morphological filters are non-linear signal transforms that locally modify the geometrical features of signals or image objects. The idempotents and increasing properties are necessary and sufficient conditions for a transform, ψ, to be a morphological filter: ψ is a morphological filter $\Leftrightarrow$ ψ is increasing and idempotent. The property of idempotents implies that applying a morphological filter twice to a signal is equivalent to applying it only once: ψ is idempotent $\Leftrightarrow$ $\psi\psi$=ψ. The increasing property ensures that the ordering relation on signals is preserved after being filtered if the same SE is employed.

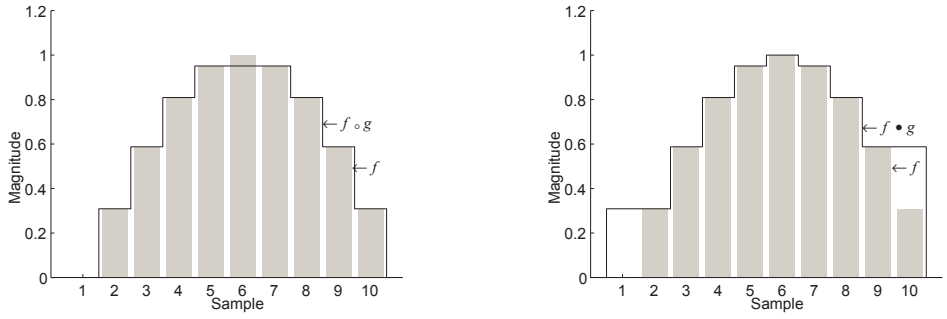
Opening and Closing

Opening is an operator that performs dilation on a signal eroded by the same SE. The definition is given as follows:

$$f \circ g = (f \ominus g) \oplus g \tag{6.2.23}$$

where f is the signal, g is the SE, and $\circ$ denotes the opening operator. Opening can recover most structures lost by erosion, except for those completely erased by erosion. Closing, on the other hand, can be defined by its duality as:

$$f \bullet g = (f \oplus g) \ominus g. \tag{6.2.24}$$



(a) (b)

Figure 6.4: Grey-scale opening and closing of an one-dimensional signal. (a) Opening. (b) Closing.

Usually, opening and closing are also denoted by operators γ and ϕ, respectively. The results of performing opening and closing on the signal used in the previous section by the same SE are illustrated in Fig. 6.4. Morphological opening and closing are both increasing transforms:

$$f_1 \leq f_2 \Rightarrow \begin{cases} \gamma(f_1) \leq \gamma(f_2) \\ \phi(f_1) \leq \phi(f_2) \end{cases} . \tag{6.2.25}$$

Moreover, successive applications of openings or closings do not further modify the signal, which means that they are both idempotent transforms:

$$\begin{aligned} \gamma\gamma &= \gamma \\ \phi\phi &= \phi \end{aligned} \tag{6.2.26}$$

Apparently, opening and closing fulfill the conditions of morphological filters. Opening and closing are also a pair of dual transforms:

$$A \circ B = (A^c \bullet B)^c \tag{6.2.27}$$

$$A \bullet B = (A^c \circ B)^c \tag{6.2.28}$$

Proof:

$$\begin{aligned} (A \circ B)^c &= (A - B \oplus B)^c = (A - B)^c - B \\ &= A^c \oplus B - B = A^c \bullet B \\ &\Rightarrow A \circ B = (A^c \bullet B)^c \end{aligned}$$

Alternating Sequential Filters

Opening and closing are the basic morphological filters and new filters can be designed from their sequential combinations, such as an opening followed by a closing or vice versa. In fact, all the following combinations are filters: $\gamma\phi$, $\phi\gamma$ and $\phi\gamma\phi$. Moverover, for these new filters, the ordering relations of

$$\gamma \leq \gamma\phi\gamma \leq \begin{matrix}\gamma\phi \\ \phi\gamma\end{matrix} \leq \phi\gamma\phi \leq \phi \tag{6.2.29}$$

are always satisfied [1]. The pair of dual filters $\gamma\phi$ and $\phi\gamma$ are called opening-closing and closing-opening filters, and they have almost the same filtering effects. Therefore, in practice, usually only one of them is employed.

In some applications, such as when the objects under processing are over a wide range of sizes, there is a need to alternatively use openings and closings with an SE of an increasing size. This sequential application of the basic operators is called an alternating sequential filter. Since the four types of sequential combinations of opening and closing are all morphological filters, four alternating sequential filters can be developed hereafter. Let γ_i and ϕ_i be a pair of dual operators with an SE of size i. Suppose the size of SE increases from i to j. Therefore, the four types of alternating sequential filters are given as:

$$f_{aoc} = (\gamma_j\phi_j)\cdots(\gamma_i\phi_i) \tag{6.2.30}$$

$$f_{aco} = (\phi_j\gamma_j)\cdots(\phi_i\gamma_i) \tag{6.2.31}$$

$$f_{aoco} = (\gamma_j\phi_j\gamma_j)\cdots(\gamma_i\phi_i\gamma_i) \tag{6.2.32}$$

$$f_{acoc} = (\phi_j\gamma_j\phi_j)\cdots(\phi_i\gamma_i\phi_i) \tag{6.2.33}$$

6.3 Multi-resolution Decomposition Schemes

Multi-resolution decomposition schemes provide convenient and effective approaches for signal processing. The core idea is to remove high frequencies in a signal, thus to obtain a reduced-scale version of the original signal [3]. Therefore,

by repeating this procedure, a collection of coarse signals at different levels are produced. Meanwhile, a collection of detail signals can be constructed by subtracting the coarse signal at level $i+1$ from the coarse signal at level i. Let $J \subseteq N$ be an index set indicating the levels in a multi-resolution decomposition scheme; J can be finite or infinite. Therefore, for a domain $V_j(j \in J)$, signals of level j belong to V_j. The analysis operator $\psi^{\uparrow}$ decomposed a signal in the direction of increasing, i.e. $\psi_j^{\uparrow} : V_j \to V_{j+1}$. On the other hand, the synthesis operator $\psi^{\downarrow}$ proceeds in the direction of decreasing, i.e. $\psi_j^{\downarrow} : V_{j+1} \to V_j$. In others words, the analysis operator $\psi^{\uparrow}$ maps a signal to a higher level and reduces information, while $\psi^{\downarrow}$ maps it to a lower level. Composing analysis operators successively, a signal at any level i can be transferred to a higher level j. The composed analysis operation is denoted as $\psi_{i,j}^{\uparrow}$ and is defined to be:

$$\psi_{i,j}^{\uparrow} = \psi_{i,j}^{\uparrow}\psi_{i,j}^{\uparrow}...\psi_{j}^{\uparrow}, j > i \tag{6.3.1}$$

It maps an element in V_i to an element in V_j. It should be noted that here the operations are carried from right to left. Likewise, the composed synthesis operator:

$$\psi_{j,i}^{\downarrow} = \psi_{i}^{\downarrow}\psi_{j,i}^{\downarrow}...\psi_{j-1}^{\downarrow}, j > i \tag{6.3.2}$$

takes the signal back from level j to level i. Finally, the composition operator that takes a signal from level i to level j and then back to level i is defined as:

$$\hat{\psi}_{i,j}^{\uparrow} = \psi_{j,i}^{\downarrow}\psi_{i,j}^{\uparrow}, j > i \tag{6.3.3}$$

where $\hat{\psi}_{i,j}$ is viewed as an approximation operator because analysis operator $\psi_j^{\uparrow}$ reduces signal information and, in general, the synthesis operator $\psi_j^{\downarrow}$ cannot compensate the information lot in the analyzing procedure.

Analysis and synthesis operators play an important role in the construction of a decomposition scheme. If there is no information lost during the analysis and synthesis procedure, then this decomposition scheme would be perfect.

6.3.1 Morphological Pyramid Transform

Pyramid transform is a multi-resolution decomposition scheme that does not influence the decomposition when analysis and synthesis steps are repeated. A

premise of pyramid transform is the so-called pyramid condition [50]. The analysis and synthesis operators $\psi_j^{\uparrow}$ and $\psi_j^{\downarrow}$ are said to satisfy the pyramid condition if $\psi_j^{\uparrow}\psi_j^{\downarrow} = \mathrm{id}$, where id represents the identity operator.

Although analysis operator $\psi_j^{\uparrow}$ is the left inverse of synthesis operator $\psi_j^{\downarrow}$, it is in general not the right inverse of the latter, which means that $\psi_j^{\downarrow}\psi_j^{\uparrow}(x)$ is only an approximation of $x \in V_j$. Denoting the approximation of x by $\hat{x}$, i.e. $\hat{x} = \hat{\psi}_{j,j+1}(x) = \psi_j^{\uparrow}\psi_j^{\downarrow}(x) \in \hat{V}_j$, it is assumed that there exists a subtraction operator $(x, \hat{x}) \mapsto x \dot{-} \hat{x}$ mapping $V_j \times \hat{V}_j$ into a set Y_j, as well as an addition operator $(\hat{x}, y) \mapsto x \dot{+} y$ mapping $\hat{V}_j \times Y_j$, where $y = x \dot{-} \hat{x}$ is the detail signal that contains information about x that is not present in $\hat{x}$. If a signal x can be reconstructed from its approximation $\hat{x}$ and the detail signal y, i.e. $\hat{x} \dot{+} (x \dot{-} \hat{x}) = x$, the signal can be decomposed recursively as:

$$\begin{aligned} x \to \{y_0, x_1\} \to \{y_0, y_1, x_2\} \to \cdots \\ \to \{y_0, y_1, \cdots, y_i, x_{j+1}\} \to \cdots \end{aligned} \tag{6.3.4}$$

where

$$\begin{cases} x_0 = x \in V_0 \\ x_{j+1} = \psi_j^{\uparrow}(x_j) \in V_{j+1}, j \geq 0. \\ y_i = x_j - \psi_j^{\downarrow}(x_j + 1) \in Y_j \end{cases} \tag{6.3.5}$$

This scheme is called the pyramid transform and the original signal x_0 can be exactly reconstructed from x_{j+1} and y_j,y_{j-1},...,y_0 by means of the backward recursion $x_j = \psi_j^{\downarrow}(x_{j+1}) + y_j$. Morphological pyramid is a special pyramid transform satisfying the following conditions [50]:

- All domains V_j are complete lattices;
- The pair $(\psi_j^{\uparrow}, \psi_j^{\downarrow})$ is an adjunction between V_j and V_{j+1}.

To satisfy the pyramid condition, $\psi_j^{\uparrow}$ is an erosion and $\psi_j^{\downarrow}$ is a dilation.

6.3.2 Morphological Wavelet

The morphological wavelet engages two analysis operators and one synthesis operator. A formal definition of the morphological wavelet is presented as follows. Assume that there exit sets V_j and W_j. V_j is referred to as the signal space at level

j and W_j denotes the detail space at level j. The two analysis operators are used to decompose a signal in the direction of increasing j. The signal analysis operator, $\psi_j^{\uparrow}$, maps a signal form V_j to $V_j + 1$, i.e. $\psi_j^{\uparrow} : V_j \to V_{j+1}$, while the detail analysis operator, $w_j^{\uparrow}$, maps it from V_j to W_{j+1}, i.e. $w_j^{\uparrow} : V_j \to W_{j+1}$. On the other hand, a synthesis operator proceeds in the direction of decreasing j, denoted as $\Psi_j^{\downarrow}$: $V_{j+1} \times W_{j+1} \to V_j$. In order to yield a complete signal representation, the mappings $(\psi_j^{\uparrow}, \omega_j^{\uparrow})$: $V_j \to V_{j+1} \times W_{j+1}$ and $\Psi_j^{\downarrow}$: $V_{j+1} \times W_{j+1} \to V_j$ should be inverses of each other, which means that the following condition:

$$\Psi_j^{\downarrow}(\psi_j^{\uparrow}(x), \omega_j^{\uparrow}(x)) = x, x \in V_j \tag{6.3.6}$$

should be fulfilled. This is called the perfect reconstruction condition, and

$$\begin{cases} \psi_j^{\uparrow}(\Psi_j^{\downarrow}(x,y)) = x, x \in V_{j+1}, y \in W_{j+1} \\ \omega_j^{\uparrow}(\Psi_j^{\downarrow}(x,y)) = y, x \in V_{j+1}, y \in W_{j+1} \end{cases} \tag{6.3.7}$$

where x is called the approximation signal and y is the detail signal. Therefore, decomposing an input signal $x_0 \in V_0$ with the following recursive analysis scheme:

$$\begin{aligned} x_0 \to \{x_1, y_1\} \to \{x_2, y_2, y_1\} \to \cdots \\ \to \{x_j, y_j, y_{j-1}, \cdots, y_1\} \to \cdots \end{aligned} \tag{6.3.8}$$

where

$$\begin{cases} x_{j+1} = \psi_j^{\uparrow}(x_j) \in V_{j+1} \\ y_{j+1} = \omega_j^{\uparrow}(x_j) \in W_{j+1} \end{cases} \tag{6.3.9}$$

x_0 cab be exactly reconstructed form x_j and y_j, y_{j-1}, ..., y_1 by means of the following recursive synthesis scheme:

$$x_{j-1} = \Psi_{j-1}^{\downarrow}(x_j, y_j) \tag{6.3.10}$$

This signal representation scheme is referred as the morphological wavelet decomposition scheme and is illustrated in Fig. 6.5.

6.4 Advanced Multi-resolution Morphological Filter with Optimal Structure Element

In previous Sections, the basic contents about the mathmatic morphology and morphological decomposition are presented. In fact, nowadays the multi-resolution

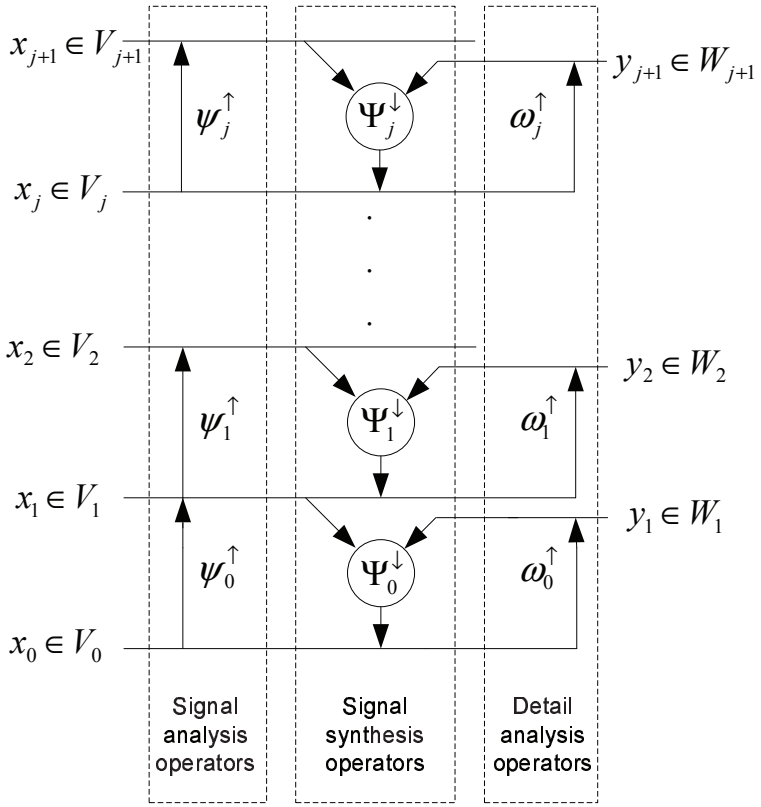


Figure 6.5: Samples stages of the morphological wavelet decomposition scheme

decomposition scheme is considered as a powerful tool for signal processing. In the following parts, an advanced multi-resolution morphological filter based on morphological wavelet with optimal structure element is presented.

6.4.1 Algorithm Description

In practice, morphological filters are chosen based on different application scenarios of signal processing. Sometimes it is difficult to obtain the prior knowledge of the positive or negative impulsive features from a signal, especially when it has both the positive and negative impulsive features. Open-closing and close-opening are two commonly used morphological filters derived from opening and closing, defined as:

$$\begin{aligned} OC(f) &= f \circ g \bullet g \\ CO(f) &= f \bullet g \circ g \end{aligned} \tag{6.4.1}$$

where

$$\begin{aligned} f \circ g &= f \ominus g \oplus g \\ f \bullet g &= f \oplus g \ominus g \end{aligned} \tag{6.4.2}$$

Combination of Open-Closing (OC) and Close-Opening (CO) can solve this case, such as the average filter formulated below:

$$\vartheta(x) = [OC(f) + CO(f)]/2 \tag{6.4.3}$$

The $f(x)$ is the input signal and the $\vartheta(x)$ is the output signal after filtered. In this filter, the structure element (SE) functions as the SE functions as a moving window through the signal to extract the high-frequency components. If the length of the SE is longer than the width of the noise in a segment of the signal, the noise can be eliminated. In prior-works, SE is always determined by trail and error, in this chapter, an optimal SE is introduced, more details is presented in following section. Hybridising the above ϑ and the morphological wavelet scheme, a multi-resolution morphological filter is constructed with the analysis operators $\psi_j^{\uparrow}$ and $\omega_j^{\uparrow}$:

$$\psi_j^{\uparrow}(x_j) = x_{j+1} = \vartheta(x_j) \tag{6.4.4}$$

$$\omega_j^{\uparrow}(x_j) = y_{j+1} = id - \vartheta(x_j) \tag{6.4.5}$$

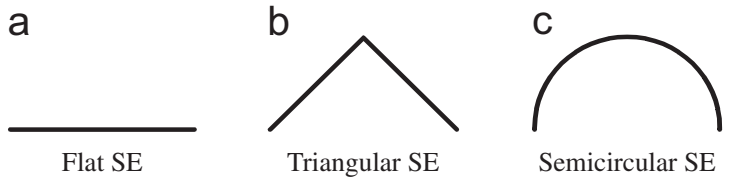


Figure 6.6: SEs of various shapes:(a) flat SE; (b) triangular SE; and (c) semicircular SE

$$\vartheta(x_j) = (OC(x_j) + CO(x_j))/2 \tag{6.4.6}$$

where $x_j \in V_j$, $x_{j+1} \in V_{j+1}$, $y_{j+1} \in W_{j+1}$, and id denotes the identity transform:

$$f, \mathrm{id}(f) = f$$

6.4.2 Optimal Structure Element Selection

After selecting the morphological filter, the SE is the next key component of the morphology analysis to be defined. Generally, only when the scale and shape of the signal are matched to those of SE, the signal can be reserved well. Therefore, the shape, length (domain) and height (amplitude) of SE should be selected according to the signal to be analyzed. The shapes of SE can vary from regular to irregular curves, such as flat, triangle, semicircle, and so on. Fig. 6.6 illustrates some of the common SEs.

As there are limited rules or guidelines for choosing the optimum SE for various morphological filter, the traditional SE selection method is adopted by prior knowledge. This is simple, and easy to be implemented. However, the prior knowledge is not always available for some applications and sometimes it is impossible to gain such knowledge before hand. Based on the [158], an optimal morphological SE calculation method is proposed.

Step 1: Select the shape of SE and the morphological filter. As previous introduced that, the OC-CO filter is selected as the morphological filter. Then, When

defining the shape of SE, an important principle shall be followed, which is that the SE should approach morphological features of the signal as far as possible. In other words, the shape of the SE should be similar to the shape of original signal. For most current energy demand dataset, such as natural gas demand, wind power demand, and electrical load demand, if we tread them as 2-D signal in time domain, the shape of their wave can be only loosely classify as triangular or dome-like. If only one fixed shape of the SE is used, the final result after filter may not good. Therefore, we presented a optimal SE shape selection method to overcome the problem. Suppose that there is an input signal $f(n)$ and a sequential multiple SE aggregation $g_{triangular}, g_{semicircular}$. The definitions of the OC-CO filter are:

$$\Phi_{oc}[f(n)] = \mathrm{MAX}(\mathrm{OC}(g_{\mathrm{triangular}}), \mathrm{OC}(g_{\mathrm{semicircular}})) \tag{6.4.7}$$

$$\Phi_{oc}[f(n)] = \mathrm{MAX}(\mathrm{OC}(g_{\mathrm{triangular}}), \mathrm{OC}(g_{\mathrm{semicircular}})) \tag{6.4.8}$$

Because the output of the OC filter is smaller and the output of the CO filter is larger than the original signal, we take the maximum output of the OC filter and the minimum outcome of the CO filter in multiple SE filtering, then take the average value of both outputs, this result could be much approach the practical signal. The framework of the OC-CO filter is shown in Fig. 6.7.

Step 2: Calculate the value of SE λ When defining the value SE, the length and width of original signal should be considered. According to the Zhang et al.[158], in order to find the suitable value SE, the local maximum and minimum peak values $h_{\max}$ and $h_{\min}$ of the original signal should be calculated as well as the minimum and maximum of peak intervals, $i_{\max}$ and $i_{\min}$. Then the minimum and maximum of length scales of SE can be achieved as following:

$$\begin{aligned} l_{\min} &= \lceil (i_{\min} - 1)/2 \rceil \\ l_{\max} &= \lfloor (i_{\max} - 1)/2 \rfloor \end{aligned} \tag{6.4.9}$$

where $\lceil\ \rceil$ denotes the operator of round toward infinity, and $\lfloor\ \rfloor$ the operator of round toward minus infinity. Then the width and height of SE under different

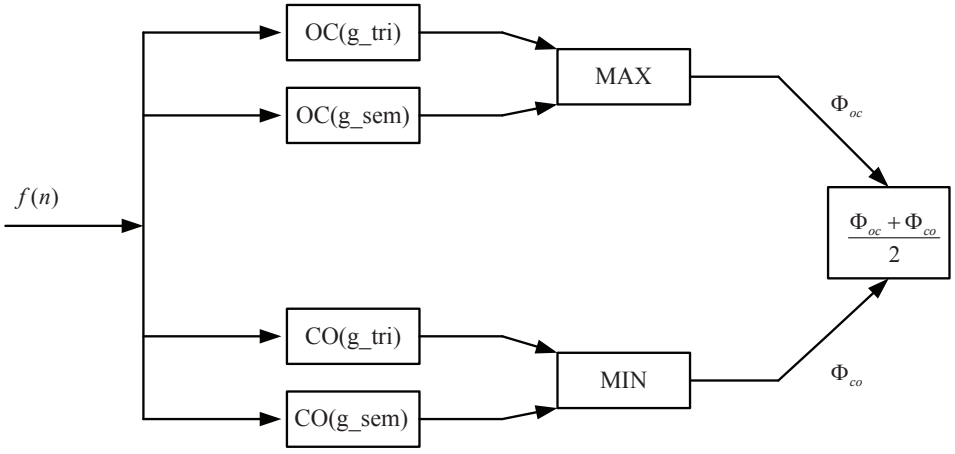


Figure 6.7: Framework of the OC-CO filter

size j can be calculated as following:

$$\begin{aligned} l_j &= l_{\min} + j, (j = 0, 1, 2, ..., l_{\max} - l_{\min}) \\ h_j &= \delta \times [h_{\min} + j \times (h_{\max} - h_{\min})/(l_{\max} - l_{\min})] \end{aligned} \tag{6.4.10}$$

where, δ is the magnitude coefficient $(0 \leq \delta \leq 1)$. In this Chapter, triangular shape and semicirular shape are both selected as the structuring elements.

$$\begin{aligned} \mathrm{SE}_{\mathrm{trig}} &= h_j \times [0\ 1\ \cdots\ l_{j-1}\ l_j\ l_{j-1} \cdots 1\ 0] \\ \mathrm{SE}_{\mathrm{demo}} &= h_j \times \sin(\tfrac{\pi}{2 \times l_j}[0\ 1\ \cdots\ l_{j-1}\ l_j\ l_{j-1} \cdots 1\ 0]) \end{aligned} \tag{6.4.11}$$

6.4.3 Advanced Multi-resolution Morphological Filter

The OC-CO filter with optimal SE is the key feature to achieve the advanced morphological filter, then the following figure shows the multi-resolution method, In Fig 6.8, the original signal x_0 is filtered 4 times. In order to explain it more clearly, we pre-set the x_0 as a sinusoidal signal polluted by white noise. The signal is filtered by a basic SE, which is flat SE with the length of j+1 (j=1,2,3,4). The results are shown in Fig. 6.8, which includes the decomposed signals x_j ("baseline parts")and the detail signals y_j at each level j ("noise parts"). From the Fig. 6.9, it can be clearly seen the changing from the x_0 to x_4, the signal x_4 is more smooth than the original x_0. To the common multi-resolution method, the ward "multi" is

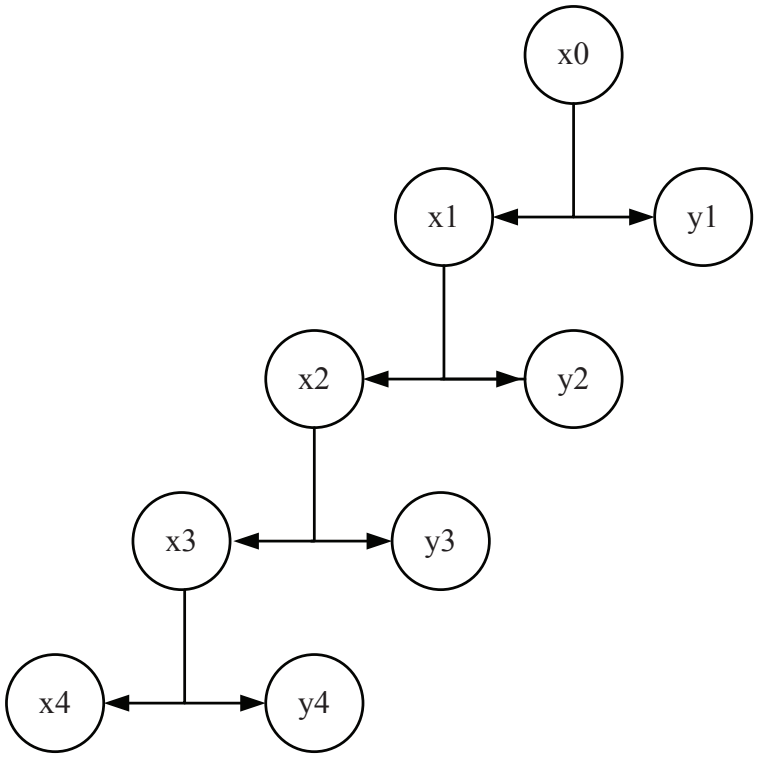


Figure 6.8: Computational process of Morphological Filter

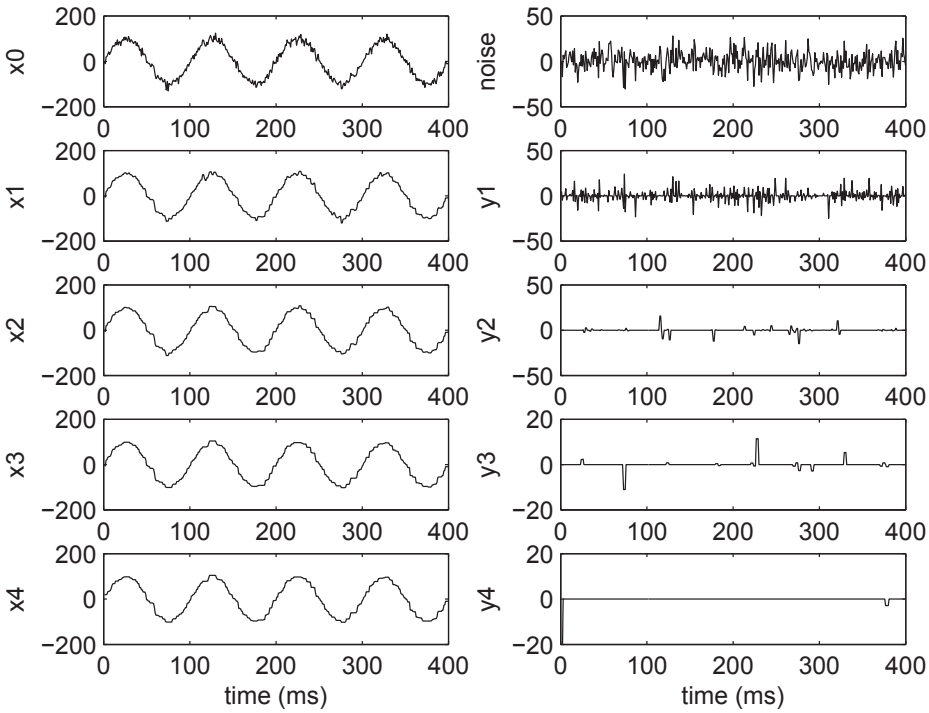


Figure 6.9: Performance of multi-resolution morphological filter on a sinusoidal signal

referred to the multiple filtering times, we denoted the times as N, in order to achieve an reasonable value N, we applied the signal-to-noise ration (SNR) to evaluate the amount of noise remaining in each level of the signal. The value of SNR is defined as:

$$SNR = 10 \times \log_{10} \frac{\mathrm{S}_\sigma}{\mathrm{N}_\sigma} \tag{6.4.12}$$

where S is noise-free signal and N is the noise. For a given signal X, X_σ is defined as:

$$X_\sigma = \sum_{l=0}^{L-1} \left(X(l) - \mu x\right)^2 \tag{6.4.13}$$

where μx and L are the mean and length of X respectively. The values if SNR of the original signal and the decomposed signals are shown in Table 6.1. As it can be seen form Tab. 6.1. SNR increases as the level ascends, but the increment is descending. At level 4, the improvement is not evident, which implies that it is enough after 3

Table 6.1: Performance of multi-resolution morphological filter

Signal	SNR (dB)	Increment
x_0	33.61	-
x_1	41.08	22.3%
x_2	42.95	4.5%
x_3	43.08	0.3%
x_4	43.17	0.2%

times filtered. Therefore, to the common application, if the filering times is N, for each $i = 1, 2, 3...N$ level, there is a corresponding SNR_i,then applied the

$$\eta_i = (SNR_i - SNR_{i-1})/SNR_{i-1}$$

, used the η to compare the pre-set increment value κ, if $\eta_i \leq \kappa$, filtering process will stop, if not, the filtering process keep running, until the $\eta_i \leq \kappa$. The proposed SVRLP-MM model is presented in Fig. 6.10.

6.4.4 Case study

Dataset

The dataset used in this chapter is the wind speed dataset which downloaded from the Illinois Wind, a website dedicated to information for Illinois residents who interested in wind as a source of renewable energy. We choose the data collected from the Coffey wind farm as our testing targets, shown in Fig. 6.11. The wind speed measuring point is 30 meters high from ground, every 10 minutes for one record, and the time scale of wind data is from 17/04/2005 to 18/04/2006, totally amount of the wind speed dataset is 58485.

Parameters

We used the 100 points which from the 58386^{th} to 58485^{th} of the wind speed time series to test the performances of proposed method, and we applied those 58485

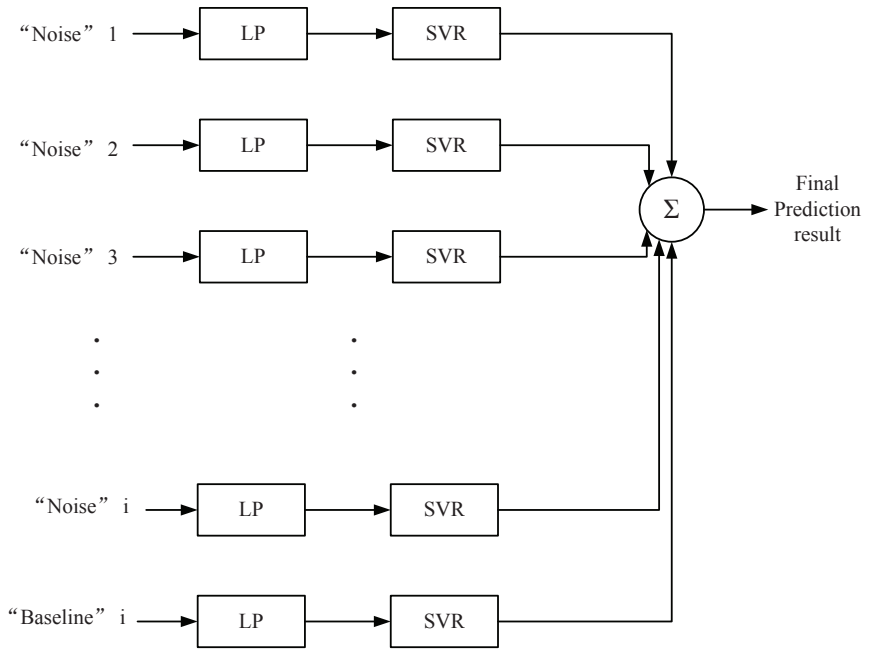


Figure 6.10: SVRMP-MM forcasting model

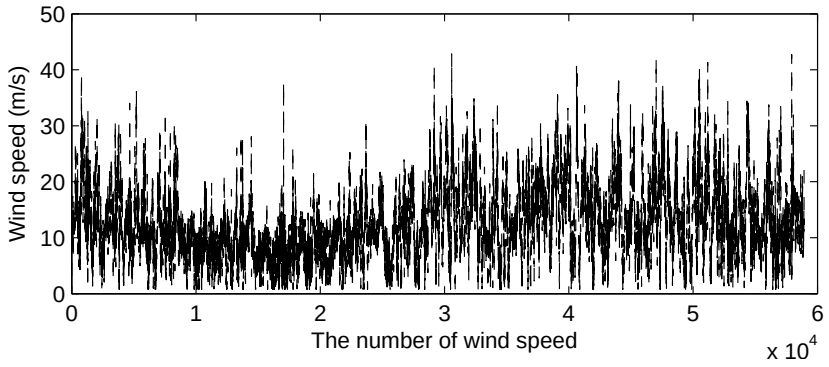


Figure 6.11: Real-world wind speed data used in this chapter

points before as historic data set. Firstly, in order to determine phase space reconstruction parameters, we randomly select 25 parts of the wind speed time series, and used them to do the phase space reconstruction parameters analysis. The results are shown in Fig. 6.12 and Fig. 6.13 we can see that the most often embedding dimen-

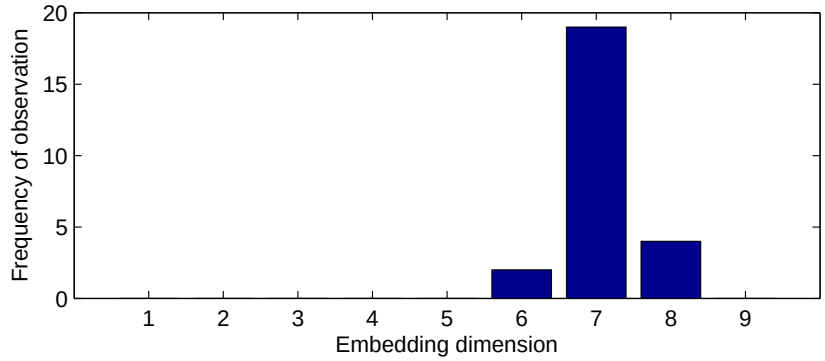


Figure 6.12: Observation times about the embedding dimension delay of the 25 random wind speed time series parts

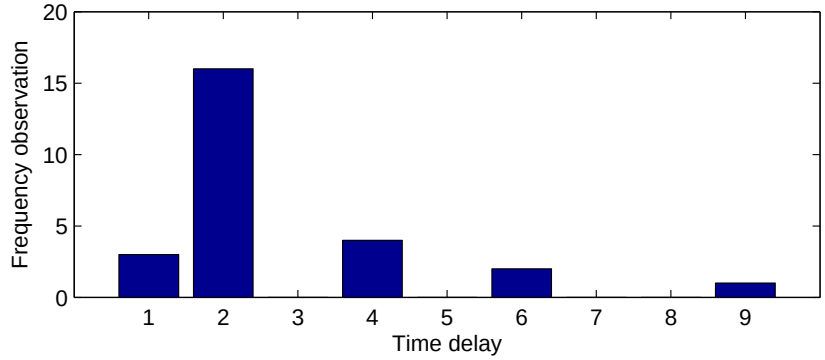


Figure 6.13: Observation times about the sampling time delay of the 25 random wind speed time series parts

sion and sampling time delay constant are 7 and 2 respectively. In order to verify the accuracy of the selection, the SVRLP with different input dimension are tested (the sampling time delay is 2). And result shows in the Fig. 6.14.

From the Fig. 6.14, we can see that the MAPE is the lowest when the sampling time delay is 2 and the embedding dimension is 7. Therefore, the sampling delay time 2 and embedding dimension 7 are fixed in the following tests. Also, choosing

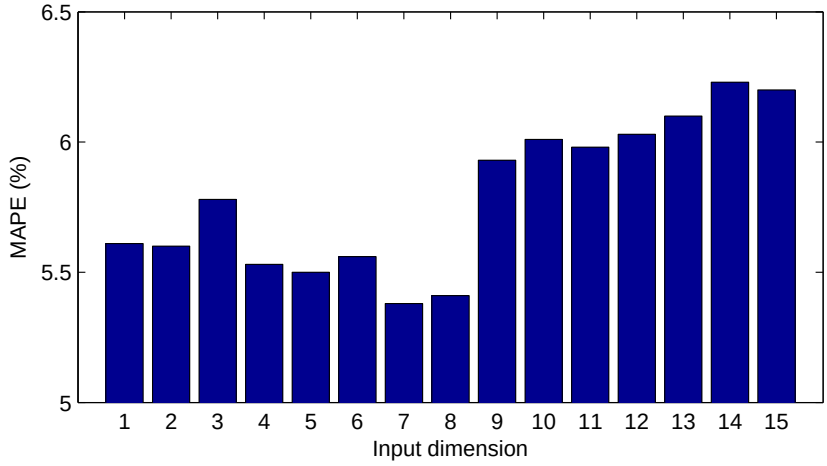


Figure 6.14: The MAPE of different dimension inputs by the SVRLP model

K is very important step in local prediction model. To calculate the best value of K for each dataset, Equation (4.3.1) is used. In addition, the parameter $k_{\max}$ and α are fixed for all test cases in this Chapter at 30% of N and 120, respectively.
Then, the wind speed dataset is treated as the signal and the proposed morphological filter is applied to it. According to the SNR results, which summarized in the Table. 6.2, we can find that the SNR increases as the level ascends, but the increment from level 7 to level 8 is quite small, it means that it is enough to perform the decomposition seven times. Therefore, to this case, the wind speed is filtered 7 times. The Figure shows results of morphological decomposition (In order to keep definition of figure, only 500 points of wind speed time series are demonstrated). From this figure, we can clearly find that shape of the x_7 is quite similar to the x_7, but with a more smooth feature.

6.4.5 Results

All of the prediction results of different models are summarized in Tab.6.3 and Fig.6.16-6.19. Form the Tab.6.4, it clearly shows that the SVRLP-MM is much better than others, it improves the accuracy and reduces the MPE. Moreover, it is also necessary to investigate the absolute percentage error (APE) of each samples to see the stability of models. In Fig. 6.20, the mainly absolute percentage error (APE)

Table 6.2: Performance of multi-resolution morphological filter for the wind speed dataset

Signal	SNR (dB)	Increment
x_1	18.79	-
x_2	22.92	21.9%
x_3	25.79	13.0%
x_4	27.98	8.4%
x_5	29.58	5.7%
x_6	30.81	4.2%
x_7	31.53	2.3%
x_8	31.81	0.8%

of SVRLP-MM is also lowest among all models. In other words, the SVRLP-MM is more stable than others.

6.5 Conclusion

In this chapter, we have presented a SVRLP-MM model. The proposed approach combines a powerful regression algorithm which is SVR with a morphological filter based local predictor. During the computation procedure, the embedding dimension and the time delay are computed, and the phase space reconstruction is applied to the dataset. We use the morphological filter to decompose the original wind speed time series into several subsequences, the subsequences contains "noise" parts and "baseline" part, then applied the SVRLP to predict each subsequence separately. The proposed approach is applied to the wind speed predicito, and the final results presented demonstrate that the proposed model can achieve a higher prediction accuracy than the SVR model, SVRLP model and ARMA model using the same real world wind speed data.

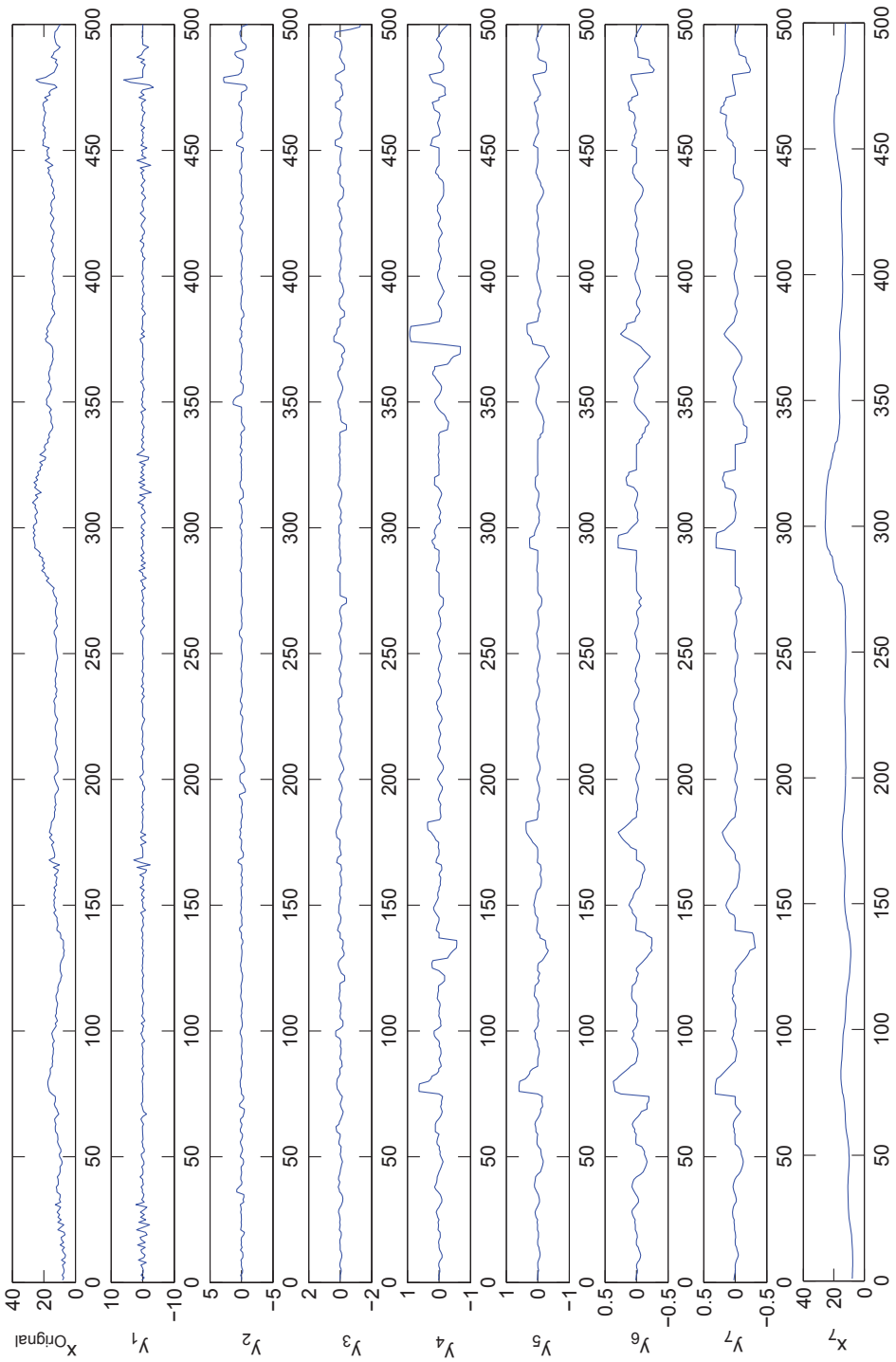


Figure 6.15: Result of morphological decomposition

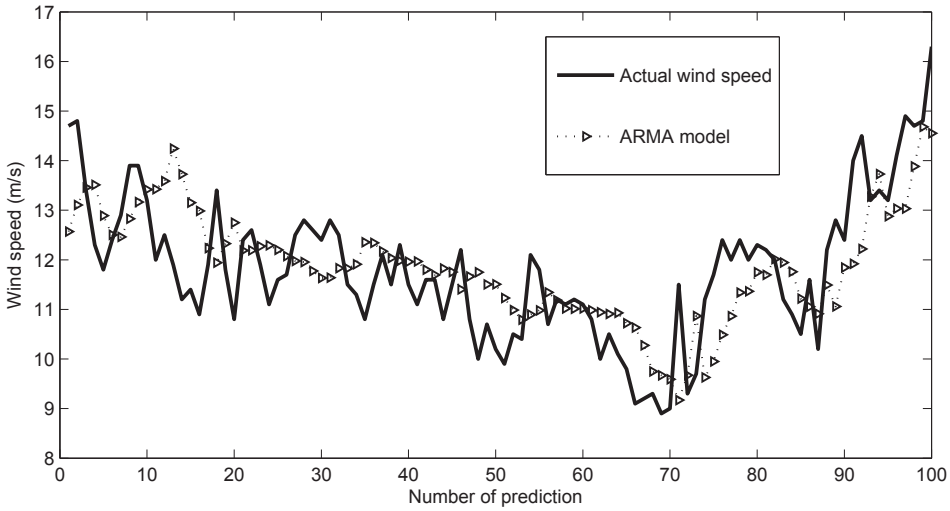


Figure 6.16: Results of wind speed prediction by ARMA model in comparison of actual wind speed

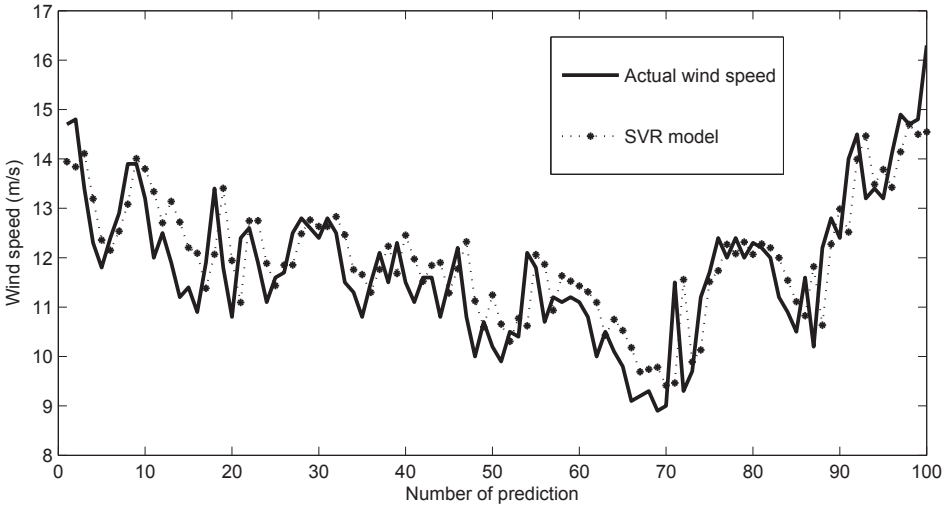


Figure 6.17: Results of wind speed prediction by SVR model in comparison of actual wind speed

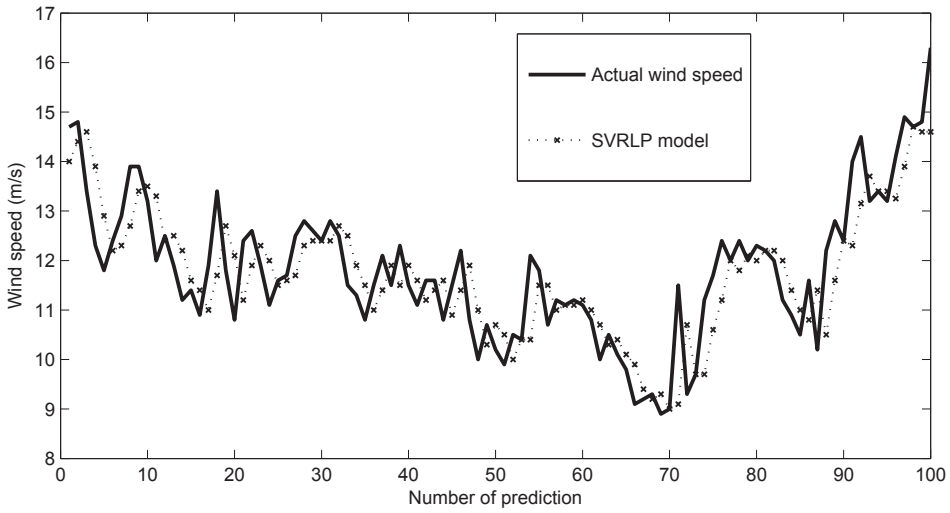


Figure 6.18: Results of wind speed prediction by SVRLP model in comparison of actual wind speed

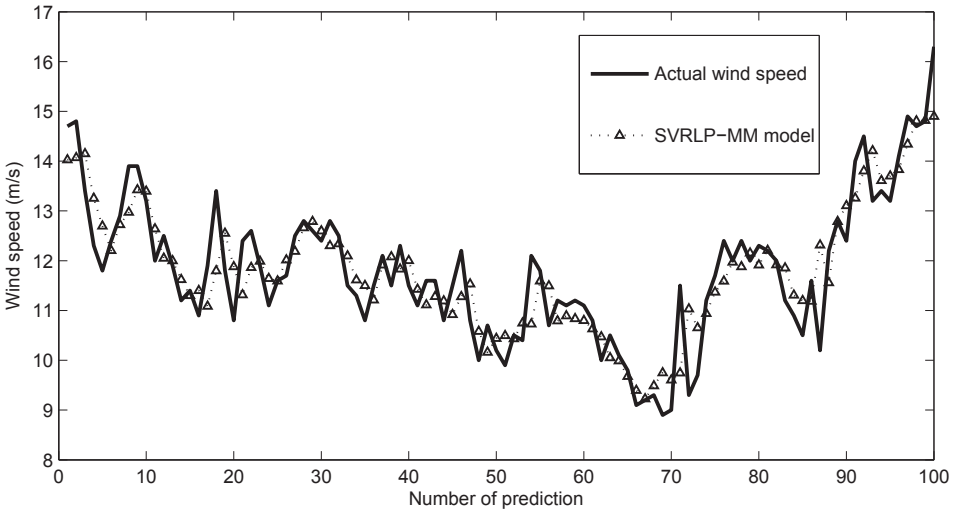


Figure 6.19: Results of wind speed prediction by SVRLP-MM model in comparison of actual wind speed

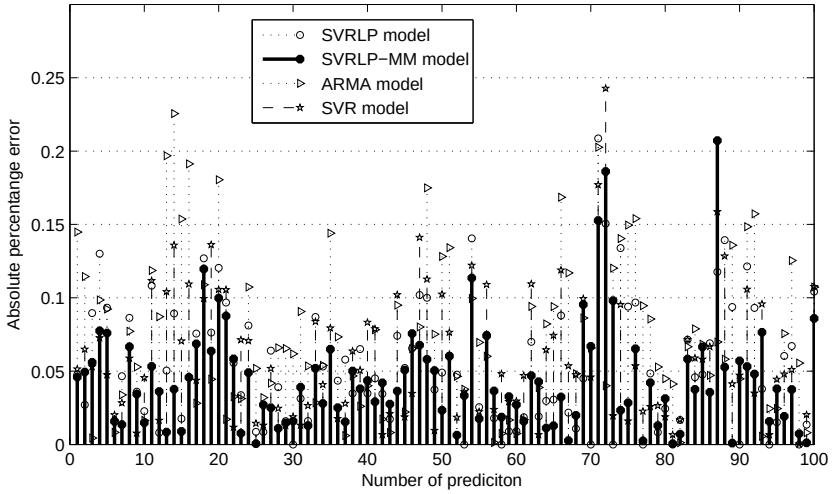


Figure 6.20: Absolute percentage error comparison among different prediction models

Table 6.3: Predicted result and the original wind speed time series

No. of prediction	Original speed	Predicted value				Predicted error			
		SVRLP-MM	SVRLP	SVR	ARMA	SVRLP-MM	SVRLP	SVR	ARMA
1	14.7	14.0	14.0	13.9	12.5	0.04	0.04	0.05	0.14
2	14.8	14.0	14.4	13.8	13.1	0.04	0.02	0.06	0.11
3	13.4	14.1	14.6	14.1	13.4	0.05	0.08	0.05	0.00
4	12.3	13.2	13.9	13.1	13.5	0.07	0.13	0.07	0.09
5	11.8	12.6	12.9	12.3	12.8	0.07	0.09	0.04	0.09
6	12.4	12.2	12.2	12.1	12.5	0.01	0.01	0.02	0.00
7	12.9	12.7	12.3	12.5	12.4	0.01	0.04	0.02	0.03
8	13.9	12.9	12.7	13.0	12.8	0.06	0.08	0.05	0.07
9	13.9	13.4	13.4	14.0	13.1	0.03	0.03	0.00	0.05
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
.	.	.	.	.	.	.	.	.	.
97	14.9	14.3	13.9	14.1	13.0	0.03	0.06	0.05	0.12
98	14.7	14.8	14.7	14.6	13.8	0.00	0.00	0.00	0.05
99	14.8	14.8	14.6	14.4	14.6	0.00	0.01	0.02	0.00
100	16.3	14.8	14.6	14.5	14.5	0.08	0.10	0.10	0.10

Table 6.4: MAPE and MPE of different models

	MAPE	MAPE Impro.	MPE	MPE Impro.
SVRLP-MM	4.43	-	18.61	-
SVRLP	5.38	17.7%	20.87	10.8%
SVR	5.94	22.5%	24.27	23.3%
ARMA	7.51	41.0%	22.56	17.5%

Chapter 7

Conclusion and Future Work

This chapter concludes the thesis. It summarizes the major achievements of the presented research work in the filed of new algorithm development for energy demand prediction. In this chapter, possible direction for future investigation are also indicated.

7.1 Conclusion

Due to the global energy crisis and environmental concerns, the energy demand management is becoming an important issue in the energy sector and forecasting the energy demand is the starting point. The original local predictor is widely used to predict the energy demand, and outperforms the global predictor. In local prediction, each predicting point has its own model constructed by its nearest neighbors (NNs) which were found in the neighborhood of the phase space reconstructed from the time series, and the fitness of NNs would mainly effect the final performance of prediction. However, it has been found that NNs may contain a class of false neighbors which would decrease the fitting accuracy dramatically and cause the reduction of the forecasting performance, so that not all of NNs are suitable for using in the local prediction, and these FNs should be filtered. Therefore, the false neighbor filter is presented to remove those false neighbor and leave the optimal nearest neighbors. In addition, the Support Vector Regression is a powerful methodology which de-

rived from Support Vector Machines, it is based upon statistical machine learning and can be utilized to achieve nonlinear mapping from sample space to feature space through kernel functions. Moreover, SVR replaces the empirical risk minimization which is generally employed in the classical methods such as ANNs, with a more advantageous structural risk minimization principle. Therefore, SVR can achieve an outperforming of fitting accuracy for chaotic time series prediction. Then, the local predictor with false neighbor filtered is combined the SVR to proposed a novel method which is the FNF-SVRLP. The FNF-SVRLP is an ameliorated SVRLP. Compared with the SVRLP, the FNF-SVRLP not only applies the Euclidean distance but also the exponential separation rate to verbify the nearest neighbors, and only the optimal nearest neighbors can be selected as training samples.

The FNF-SVRLP based short-term wind power prediction is introduced. Wind power is widely used in the world. However, the increasing penetration of wind power into the electric power grid accompanied with a series of challenges. Due to the uncertain and variable nature of wind resources, the output power of wind farms is hard to control, which could lead to the instability of the power grid operation and the unreliability of electricity supplies. The accurate short-term wind power prediction could help to solve this problem. During the computation procedure of the WPP, the embedding dimension and the time delay are computed firstly, then the phase space reconstruction is applied to the wind power data. Then the FNF-SVRLP is applied to predict the wind power data. The final results presented demonstrate that FNF-SVRLP based WPP model can achieve a higher prediction accuracy than the SVRLP based WPP model and ARMA model based WPP model.

Natural gas, one of the most important energy, is considered as the clean energy in most counties. However, natural gas is an non-renewable energy source, which has limited reserve on the planet and may contribute to global warming. Thus, planning energy demand is becoming an important issue in the energy sector and forecasting its demand is the starting point, especially for natural gas demand. Before the FNF-SVRLP is applied to do the short-term natural gas demand prediction, the natural gas data should be reconstructed in time series. The dataset of natural gas which contains both the demand information and weather information, are the

multivariate time series. Then, an accuracy-based method is proposed to determine the embedding dimension and the delay time for the natural gas data. Through investigating the natural gas data which obtained from National Grid, it is found that the customer behavior can affect the final forecasted performance. Therefore, the customer behavior based "Advanced Model" (AM) is presented. The FNF-SVRLP based AM for short-term natural gas demand prediction is verified by using operational gas dataset collected from National Grid and is compared with the SVRLP and ARMA to show its superiority. In addition, the online natural gas demand prediction system is set up. The engineers of Nation Grid can easily access the online system via internet in anywhere. The FNF-SVRLP based AM is installed in the online system, and the overall performance shows that our online natural gas prediction system is better than the National Grid's.

After numerous researches, we find that the most kinds of energy demand data are the non-stationary, the internal regularity between predicting point and its nearest-neighbors are much more complex than the stationary dataset. In order to help the local predictor to capture the internal regularity between predicting point and its nearest-neighbors more accurately, the morphological filter is proposed. the morphological filter is applied to decompose the non-stationary into several subsequences, ranked form the low frequency subsequence to the high frequency subsequence. Through this way, the local predictor could capture the non-stationary dataset more accurate, and improve the final performance of prediction. In addition, an novel calculation method of SE is introduced. Different form the conventional SE, this novel approach can optimize the scale and shape of SE to match the original signal. After that, a novel algorithm, which is SVRLP-MM is proposed. The real-world wind speed data has been used to evaluate the performance of SVRLP-MM. The final results presented demonstrate that SVRLP-MM based wind speed model can achieve a higher prediction accuracy than the SVRLP based model and ARMA model based model by using the same real world wind speed data.

7.2 Future work

This study has presented some novel techniques for solving the energy demand prediction problem. Although this study has achieved reasonable results, there are still many aspects which can be developed to improve the forecasting accuracy and increase the application areas.

In this study, the natural gas demand and wind power has been investigated in detail, there are some other important energy need to be research. For example, the electricity load. The forecasting of the future electricity load is one of the most important aspects of effective management of electric power systems. Therefore, the novel approaches which were introduced in this study can be applied to the load prediction. What' more, the prediction methods are not only applied in energy area, they can be applied to many other areas, such as the marking, the biosciences and so on. Therefore, the proposed prediction methods can be applied in many other areas.

Based on the experience of the web server based online natural gas demand prediction system, this web server based model can be extend the applications to other grid operators or other countries. According to the characteristic of the research object, special design and optimization can be applied, and care must be taken in defining the model parameters in view of the criteria that are set up by gas market operators in these countries.

As the numerous factors of the energy demand, prediction is never 100% accurate. Inaccuracies in energy demand prediction, for example, the natural gas demand prediction, may cause an increase in the cost of operating the gas system. The economic impact analysis of energy demand is not included in this study. Therefore, the economical impact of the energy demand, can be considered in the future.

References

[1] CSIRO and The Natural Edge Project. *Energy transformed: sustainable energy solutions for climate change mitigation*. 2007. p.6.

[2] Smog, *[Online]. Available:http://en.wikipedia.org/wiki/Smog*.

[3] How electricity is generated in the UK.*[Online]. Available: http://www.hienergy.org.uk/Renewables/Why-Renewable-%20Energy/Howelectricity- is-generated-in-the-UK.htm*

[4] Electricity generation and supply figures for Scotland, Wales, Northern Ireland and England, 2008 to 2011.*[Online].https://www.gov.uk/government/uploads/system/ uploads/attachment data/file/65841/7345-elec-gen-2008-2011-etarticle.pdf/*

[5] European Wind Energy Association (EWEA). Wind energy and EU climate policy, Belgium, October 2011.*[Online]. Available:http://www.ewea.org*.

[6] European Wind Energy Association (EWEA). Powering the energy debate, Belgium, 2010.*[Online]. Available:http://www.ewea.org*.

[7] K. W. Lau, Q. H. Wu, Local prediction of non-linear time series using support vector regression, *Pattern Recognition*,2008, 41(5): 1556-1564.

[8] X.F. Gong, C.H. Lai, Improvemen of local prediction of chaotic time series. *Physical Review E*, 1999;60(5):5463-5468.

[9] H. Chen, C.A. Canizares and A. Singh, ANN-based short-term load forecasting in eletricity markeds, In *Proceedings of the IEEE Power Engineering*

Society Transmission and Distribution Conf., pages 496-499,2001 Columbus, OH, USA.

[10] P. Mandal, T. Senju, and T. Funabashi, Neural networks apporach to forecast several hour ahead electricity prices and loads in deregulated market. In *Energy Conversion and Management*, 2006, 47:2128-2142.

[11] E. Erdem, J. Shi, ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy* 2011;88:1405-1414

[12] E. Erdogdu. Natural gas demand in Turkey, *Applied Energy* 2010;87(1):211-219.

[13] J. Bargur, A. Mandel, Energy consumption and economic growth in Israel:trend analysis (1960C1979). In: *Proceedings of the third international conference on energy use management*. 1981.

[14] C.S. Gonzales, B.J. Xiberta, C.H. Llaneza, Forecasting of energy production and consumption in Asturias (Northern Spain). *Energy* 1999;24:183-198.

[15] V. Ediger, H. Tathdil, Forecasting the primary energy demand in Turkey and analysis of cyclic patterns. *Energy Conversion and Management* 2002;43(4):473-87

[16] J.G. Hunt, G. Judge, Y. Ninomiya, Underlying trends and seasonality in UK energy demand: a sectoral analysis. *Energy Economics* 2003;25:93-118.

[17] U. Kumar, V.K. Jain, Time series models (Grey-Markov. Grey Model with rolling mechanism and singular spectrum analysis) to forecast energy consumption in India. *Energy* 2010;35(4):1709-1716.

[18] F. Gori, D. Ludovisi, P.F. Cerritelli, Forecast of oil price and consumption in the short term under three scenarios: parabolic, linear and chaotic behaviour. *Energy* 2007;32:1291-1296.

[19] H. Aras, N. Aras. Forecasting residential natural gas demand. *Energy Sources* 2004;26(5):463-472

[20] M.A. Badri, A. Al-Mutawa, D. Davis, EDSSF: a decision support system (DSS) for electricity peak-load forecasting. *Energy* 1997;22(6):579-589.

[21] E. Gonzalez-Romera, M.A. Jaramillo-Maron, D. Carmona-Fernandez, Monthly electric energy demand forecasting based on trend extraction. *IEEE Transactions on Power Systems* 2006;21(4):1946-1953.

[22] J. Arroyo, A. Sarabia, Exponential smoothing methods for interval time series. *In: ESTSP 07: first European symposium on time series prediction (TSP). Proceedings of symposium. 2007*. 231-240.

[23] A.A. Himanshu, C.H. Lester, Electricity demand for Sri Lanka: a time series analysis. *Energy* 2008;33:724-739.

[24] Al-Shobaki S, Mohsen M, Modeling and forecasting of electrical power demands for capacity planning. *Energy Conversion and Management* 2008;49(11):3367-3375.

[25] H.A. Amarawickrama, L.C. Hunt, Electricity demand for Sri Lanka: a time series analysis. *Energy* 2008;33(5):724-739.

[26] H. Farahbalhsh,V.I. Ugursal, A.S. Fung, A residential enduse energy consumption model for Canada. *Energy Research* 1998;22:1133-43.

[27] B.C. ONeill, M. Desai, Accuracy of past projections of US energy consumption. *Energy Policy* 2005;33(8):979-93.

[28] C.C. Lee, C.P. Chang, The impact of energy consumption on economic growth: evidence from linear and nonlinear models in Taiwan. *Energy* 2007;32(12):2282-94.

[29] G. Jannuzzi, L. Schipper, The structure of electricity demand in the Brazilian household sector. *Energy Policy* 1991;19(9):879-91.

[30] J.L. Harris, L. Lon-Mu, Dynamic structural analysis and forecasting of residential electricity consumption. *Forecasting* 1993;9:437-55.

[31] A.T. Furtado, S.B. Suslick, Forecasting of petroleum consumption in Brazil using the intensity of energy technique. *Energy Policy* 1993;21(9):958-68.

[32] C.M. Harry. Trends in Dutch energy intensities for the period 1969-1988. *Energy* 1998;23(10):815-22.

[33] F. Egelioglu, A.A. Mohamad, H. Guven, Economic variables and electricity consumption in Northern Cyprus. *Energy* 2001;26(4):355-62.

[34] Z. Yumurtaci, E. Asmaz, Electric energy demand of Turkey for the year 2050. *Energy Sources* 2004;26(12):1157-64.

[35] M. Tunc, U. Camdali, C. Parmaksizoglu, Comparison of Turkeys electrical energy consumption and production with some European countries and optimization of future electrical power supply investments in Turkey. *Energy Policy* 2006;34(1):50-9.

[36] M. Bessec, J. Fouquau, The non-linear link between electricity consumption and temperature in Europe: a threshold panel approach. *Energy Economics* 2008;30(5):2705-21.

[37] A. Al-Ghandoor, I. Al-Hinti, J.O. Jaber, S.A. Sawalha, Electricity consumption and associated GHG emissions of the Jordanian industrial sector: empirical analysis and future projection. *Energy Policy* 2008;36(1):258-67.

[38] J.C. Lam, H.L.Tang, D.H.W. Li. Seasonal variations in residential and commercial sector electricity consumption in Hong Kong. *Energy* 2008;33(3):513-23.

[39] T. Jonsson, P. Pinson, H. Madsen. On the market impact of wind energy forecasts. *Energy Economics* 2010;32(2):313-20.

[40] J.E. Samouilidis, C.S. Samouilidis. Energy and economic growth in industrialized countries. *Energy Economics* 1984:191-201.

[41] L. Suganthi, T.R. Jagadeesan. A modified model for prediction of India future energy requirement. *Energy and Environment* 1992;3(4):371-86.

[42] R. Sengupta. Energy modelling for India: towards a policy for commercial energy. New Delhi: Study Report of Planning Commission, Government of India; 1993.

[43] R.D. Rao, J. Parikh. Forecast and analysis of demand for petroleum products in India. *Energy Policy* 1996;24(6):583-92.

[44] E. Arsenault, J.T. Bernard,C.W. Carr, Genest-Laplante E. A total energy demand model of Quebec, Forecasting properties. *Energy Economics* 1995;17(2):163-71.

[45] D. Intarapravich, C.J. Johnson, B. Li, S. Long, S. Pezeshki, W. Prawiraatmadja, et al. Asia-Pacific energy supply and demand to 2010. *Energy* 1996;21(11):1017-39.

[46] R. Haas, L. Schipper. Residential energy demand in OECD-countries and the role of irreversible efficiency improvements. *Energy Economics* 1998;20(4):421-42.

[47] N.M. Christodoulakis, S.C. Kalyvitis, D.P. Lalas, S. Pesmajoglou. Forecasting energy consumption and energy related CO2 emissions in Greece: an evaluation of the consequences of the Community Support Framework II and natural gas penetration. *Energy Economics* 2000;22:395C 422.

[48] D.P. Sharma, Nair PS Chandramohanan,R. Balasubramanian. Demand for commercial energy in the state of Kerala, India: an econometric analysis with medium-range projections. *Energy Policy*2002;30(9):781-91.

[49] L. ZhiDong. An econometric study on Chinas economy, energy and environment to the year 2030. *Energy Policy* 2003;31:1137-50.

[50] I.D. McAvinchey, A. Yannopoulos. Stationarity, structural change and specification in a demand system: the case of energy. *Energy Economics* 2003;25(1):65-92.

[51] W. Lu, Y. Ma. Image of energy consumption of well off society in China. *Energy Conversion and Management* 2004;45:1357-67.

[52] M. Yang, X. Yu. Chinas rural electricity marketła quantitative analysis. *Energy* 2004;29(7):961-77.

[53] F. Gori, C. Takanen. Forecast of energy consumption of industry and household and services in Italy. *Heat Technology* 2004;22(2):115-21.

[54] L.C. Hunt, Y. Ninomiya. Primary energy demand in Japan: an empirical analysis of long-term trends and future CO2 emissions. *Energy Policy* 2005;33(11):1409-24.

[55] S.P. Raghuvanshi, A. Chandra,A.K. Raghav. Carbon dioxide emissions from coal based power generation in India. *Energy Conversion and Management* 2006;47:427-41.

[56] R. Ramanathan. A multi-factor efficiency perspective to the relationships among world GDP, energy consumption and carbon dioxide emissions. *Technological Forecasting and Social Change* 2006;73:483-94.

[57] L. Hang, M. Tu. The impacts of energy prices on energy intensity: evidence from China. *Energy Policy* 2007;35(5):2978-88.

[58] H. Saddler, M. Diesendorf, R. Denniss. Clean energy scenarios for Australia.*Energy Policy* 2007;35:1245-56.

[59] Y. Fan, H. Liao, Y.M. Wei. Can market oriented economic reforms contribute to energy efficiency improvement? Evidence from China. *Energy Policy* 2007;35(4):2287-95.

[60] F.G. Adams, Y. Shachmurove. Modeling and forecasting energy consumption in China: implications for Chinese energy demand and imports in 2020. *Energy Economics* 2008;30(3):1263- 78.

[61] M. Aydinalp, A.S. Fung. Modeling of the appliance, lighting, and space-cooling energy consumptions in the residential sector using neural networks. *Applied Energy* 2002;71(2):87-110.

[62] A. Sozen, E. Arcaklioglu, M. Ozkaymak. Turkeys net energy consumption. *Applied Energy* 2005;81(2):209-21.

[63] F.B. Gorucu, F. Gumrah. Evaluation of forecasting of gas consumption by statistical analysis. Energy Sources 2004;26:267-76.

[64] Y.S. Murat, H. Ceylan. Use of artificial neural networks for transport energy demand modeling. *Energy Policy* 2006;34(17):3165-72.

[65] T. Limanond,S. Jomnonkwao,A. Srikaew. Projection of future transport energy demand of Thailand. *Energy Policy* 2011;39:2754-63.

[66] R.H. Brown, P. Kharouf, X. Feng, L.P. Piessens, D. Nestor, Development of feed-forward network models to predict gas consumption. *In: IEEE international conference on neural networks - conference proceedings*, vol. 2; 1994. p. 802-5.

[67] R.H. Brown, I.Matin, Development of artificial neural network models to predict daily gas consumption. *In: IECON proceedings (industrial electronics conference)*, vol. 2; 1995. p. 1389-94.

[68] A. Khotanzad, H.Elragal. Natural gas load forecasting with combination of adaptive neural networks. *In: Proceedings of the international joint conference on neural networks*, vol. 6; 1999. p. 4069-72.

[69] A. Khotanzad, H. Elragal, T.L. Lu. Combination of artificial neural-network forecasters for prediction of natural gas consumption. *IEEE Transactions Neural Networks 2000*;11(2):464-73.

[70] Ivezić D. Short-term natural gas consumption forecast. *FME Transactions* 2006;34;165-169.

[71] Azadeh A, Asadzadeh SM, Ghanbari A. An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: uncertain and complex environments. *Energy Policy* 2010;38(3):1529-36.

[72] K. Ermis, A. Midilli,I. Dincer,M.A. Rosen. Artificial neural network analysis of world green energy use. *Energy Policy* 2007;35(3):1731-43.

[73] A. S*ö*zen,E. Arcaklioglu. Prediction of net energy consumption based on economic indicators (GNP and GDP) in Turkey. *Energy Policy* 2007;35(10):4981-92.

[74] Z.W. Geem, W.E. Roper. Energy demand estimation of South Korea using artificial neural network. *Energy Policy* 2009;37(10):4049-54.

[75] H.T. Pao. Forecasting energy consumption in Taiwan using hybrid nonlinear models. *Energy* 2009;34(10):1438-46.

[76] A. S*ö*zen. Future projection of the energy dependency of Turkey using artificial neural network. *Energy Policy* 2009;37(11):4827-3.

[77] R. Yokoyama,T.Wakui, R. Satake. Prediction of energy demands using neural network with model identification by global optimization. *Energy Conversion and Management* 2009;50(2):319-27.

[78] L. Ekonomo, Greek long-term energy consumption prediction using artificial neural networks. *Energy* 2010;35:512-7.

[79] M. Kankal, A.AkpInar, M.I. Komurcu,T.S.Ozsahin. Modeling and forecasting of Turkeys energy consumption using socio-economic and demographic variables. *Applied Energy* 2011;88:1927-39.

[80] E. Cadenas, W. Rivera. Short term wind speed forecasting in La Venta, Oaxaca, Mexico, using artificial neural networks. *Renew Energy* 2009;34:274-8.

[81] G. Li, J. Shi. On comparing three artificial neural networks for wind speed forecasting. *Appl Energy* 2010;87:2313-20.

[82] K. Y. Chen and C. H. Wang. Support vector regression with genetic algorithms in forecasting tourism demand, *Tourism Management*, 28: 215-226, 2007.

[83] V. N. Vapnik, The natural of statistical learning theory. Springer, New York, 1995.

[84] B. Scholkopf, C. Burges, and V. Vapnik. Extracting support data for a given task. In *Proceedings of the First Int. Conf. on Knowledge Discovery and Data Mining, Menlo Park*, pages 252-257. AAAI Press, 1995.

[85] A. J. Smola and B. Scholkopf. A tutorial on support vector regression. *Neuro COLT Technical Report NC-TR-98-030, Royal Holloway College, Univesity of London*, 1998.

[86] W. Karush. Minima of functions of several variables with inequalities as side constraints. Master's thesis, Deptartment of Mathematics, University of Chicago, 1939.

[87] H. W. Kuhn and A. W. Tucker. Nonlinear programming. In *In Proc. of second Berkeley Symposium on Mathematical Statistics and Probabilitics*, pages 481-492, University of California Press,1951.

[88] S. R. Gunn. Support vector machines for classification and regression. *Technical Report Image Speech and Intelligent Systems Research Group*, University of Southampton, 1997.

[89] S. Haykin. *Neural networks, A comprehensive foundation*. Printic-Hall, Inc., 1999.

[90] F. Girosi. An equivalence between sparse approximation and support vector manchines. *Neural Comutation*, 10(6): 1455-1480, August 1998.

[91] N. Christiani and J. S. Taylor. *An introdcution to support vector machines and other kernel-based learning methods*. Cambridge University Press, Cambridge, 2000.

[92] N. Aronszajn. Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68(3):337-404, May 1950.

[93] B. Scholkopf and A. J. Smola. *learning with kernels-support vector machines, regularization, optimization and beyond*. The MIT Press, Cambridge, 2002.

[94] A. Karatzogluo, D. Meyer, and K. Hornik. Support vector machines in R. *Jouranl of Statistical Software*, 15(9):1-28, April 2006.

[95] R. Fletcher. *Practical methods of optimization*. John Wiley and Sons, 1987.

[96] J. C. Platt. Sequential minimal optimization : A fast algorithm for training support vector machines. *Microsoft Research Technical Report MSR-TR-98-14*, 1998.

[97] T. Joachims. Text categorization with support vector machines: Learning with many relevant features. *Leture Notes in Computer Science*, 1398:137-142,1998.

[98] E. Osunna, R. Freund, and F. Girosi. An inproved training algorithm for support vector machines. In *Proceedings of IEEE Workshop on Neural Networks for Signal Processing*, pages 276-285,1997.

[99] E. E. Elattar. Short term wind power prediction using evilutionart optimized local support vector regression, in *Innovative Smart Grid Technologies (ISGT Europe)*. 1-7, 2001.

[100] L. Kamal and Y. Z. Jafri, Time series models to simulate and forecast hourly averaged wind speed in Wuetta, Pakistan, *Solar Enengy*, vol. 61, no. 1, pp. 23-32, 1997.

[101] K. W. Lau and Q. H. Wu, Local prediction of non-linear time series using support vector regression, *Pattern Recognition*, vol. 41, no. 5, pp. 1556-1564, 2008.

[102] P. Zhao, J. R. Xia, Y. P. Dai and J. X. He, Wind speed prediction using support vector regression, in *Industrial Electronics and Applications (ICIEA)*, 2010. pp. 882-886.

[103] Y. Wang, D. L. Wu, C. X. Guo and Q. H. Wu, Short-term wind speed prediction using support vector regression, in *Power and Energy Society General Meeting*, 2010. pp. 1-6.

[104] J. D. Farmer, and J. J. Sidorwich. Predicting chaotic time series, *Phys. Rev. Lett*, vol. 59, pp. 845-848, 1987.

[105] F. X. Gong and C. H. Lai. Improvement of the local prediction of chaotic time series, *Phys. Rev. E*, vol. 60, no.5 pp. 5463-5468, 1999.

[106] N. H. Packard, J. P. Cratchfield, J. D. Farmer, and R. S. Shaw, Geomertry from a time series, *Phys. Rev. E*, vol. 45, no.9 pp. 712-726, 1980.

[107] F. Takens, Detecting Strange Attractors in Turbulence, *Lecture Notes in Mathematics*, Berlin: Springer, vol. 898, pp.366-381, 1981.

[108] T. Sauer, J. Yorke, and M. Csdagli, Embedology , *Sat. Phys.*, vol. 65, no. 3/4, pp.579-616, 1991.

[109] A. M. Fraser, and H. L. Swinney, Independent coordinate for strange attractors from mutual information, *Phys. Rev.*, vol. 33, no. 2, pp.1134-1140, 1986.

[110] R. Shaw, Strange attractors, chaotic behavior, and information flow, *Zeitschrift Naturforschung Teil A*, vol. 36, pp.80, 1981.

[111] J. Parikh, P. Purohit,P. Maitra . Demand projections of petroleum products and natural gas in India. *Energy* Oct.2007;32(10):1825-1837.

[112] B. Soldo. Forecasting natural gas consumption. *Applied Energy* 2012;92:26-37.

[113] A.T. Almeida, A.C. Lopes, A. Carvalho, J. Mariano, et al. Examining the potential of natural gas demand-side measures to benefit customers, the distribution utility, and the environment: two case studies from Europe. *Energy* 2004;29:979-1000.

[114] How electricity is generated in the UK.*[online] http://www.hi-energy.org.uk/Renewables/Why-Renewable-%20Energy/How-electricity-is-generated-in-the-UK.htm*

[115] Electricity generation and supply figures for Scotland, Wales, Northern Ireland and England, 2008 to 2011.*https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/65841/7345-elec-gen-2008-2011-et- article.pdf/*

[116] H. Sarak, A. Satman. The degree-day method to estimate the residential heating natural gas consumption in Turkey: a case study. *Energy* 2003;28:929-939.

[117] J. Siemek, S. Nagy, S. Rychlicki. Estimation of natural-gas consumption in Poland based on the logistic-curve interpretation. *Applied Energy* 2003;75(1-2):1-7.

[118] R. Gutierrez, A. Nafidi, R.G. Sanchez. Forecasting total natural-gas consumption in Spain by using the stochastic Gompertz innovation diffusion model. *Applied Energy* 2005;80(2):115-24.

[119] P. Potočnik, M. Thaler, E. Govekar, I. Grabec, A. Poredos. Forecasting risks of natural gas consumption in Slovenia. *Energy Policy* 2007;35:4271-4282.

[120] E. Erdogdu, Natural gas demand in Turkey. *Applied Energy* 2010;87(1):211-219.

[121] R.H. Brown, P. Kharouf, X. Feng, L.P. Piessens, D. Nestor. Development of feed-forward network models to predict gas consumption. *In: IEEE international conference on neural networks - conference proceedings*, vol. 2; 1994. p. 802-5.

[122] Brown RH, Matin I. Development of artificial neural network models to predict daily gas consumption. *In: IECON proceedings (industrial electronics conference)*, vol. 2; 1995. p. 1389-94.

[123] A. Khotanzad, H. Elragal. Natural gas load forecasting with combination of adaptive neural networks. *In: Proceedings of the international joint conference on neural networks*, vol. 6; 1999. p. 4069-72.

[124] A. Khotanzad, H. Elragal, T.L. Lu. Combination of artificial neural-network forecasters for prediction of natural gas consumption. *IEEE Transactions Neural Networks 2000*;11(2):464-73.

[125] Ivezić D. Short-term natural gas consumption forecast. *FME Transactions* 2006;34;165-169.

[126] A. Azadeh, S.M. Asadzadeh, A.Ghanbari. An adaptive network-based fuzzy inference system for short-term natural gas demand estimation: uncertain and complex environments. *Energy Policy* 2010;38(3):1529-36.

[127] K. Salahshoor, M. Kordestani, M.S. Khoshro. Fault detection and diagnosis of an industrial steam turbine using fusion of SVM (support vector machine) and ANFIS (adaptive neuro-fuzzy inference system) classifiers. *Energy* 2010;35(12):5472-5482.

[128] W.Y. Zhang, W.C. Hong,Y. Dong, G.Tsai, J.T. Sung, G.F. Fan, Application of SVR with chaotic GASA algorithm in cyclic electric load forecasting. *Energy* 2012;45:850-858.

[129] H. Taghavifar, A. Mardani, A comparative trend in forecasting ability of artificial neural networks and regressive support vector machine methodologies for energy dissipation modeling of off-road vehicles. *Energy* 2014;66:569-576.

[130] W.C. Hong. Load forecasting by seasonal recurrent SVR (support vector regression) with chaotic artificial bee colony algorithm. *Energy* 2011;36:5568-5578.

[131] N.H. Packard, J.P. Cratchfield, J.D. Farmer, R.S. Shaw. Geomertry from a time series. *Physical Review E* 1980;45(9):712-726.

[132] F. Takens, Detecting Strange Attractors in Turbulence. *Lecture Notes in Mathematics. Springer* 1981;898:366-381.

[133] T. Sauer, J. Yorke, M. Csdagli. Embedology. *Journal of Statistical Physics* 1991; 65(3/4):579-616.

[134] P. Grassberger, I. Procaccia, Estimation of the kolmogorov entropy from a ahaotic signal. *Physics Review A* 1983;28:2591-2593.

[135] W. Liebert, H.G. Schuster. Proper choice of the time delay for the analysis of chaotic time series. *Physics Letter* 1989;142:107-111.

[136] Data Item Explorer. *http://marketinformation.natgrid.co.uk/gas/*

[137] A. Kusiak, H. Zheng, Z. Song, Short-Term Prediction of Wind Farm Power: A Data Mining Approach. *IEEE Transactions On Energy Conversion* 2009;24(1):125-136.

[138] A.T. Lora, J.C. Riquelme, J.L.M. Ramos, Influence of kNN based load forecasting errors on optimal energy production. *Lecture Notes in Computer Science (Springer Berlin)* 2003;2902:189-203.

[139] A. Sorjamaa, N. Reyhani, A. Lendasse, Input and structure selection for k-NN approximator. *Lecture Notes in Computer Science (Springer Berlin)* 2005;3512:985-992.

[140] E.E. Elattar, J.Y. Goulermas, Q.H. Wu. Generalized Locally Weighted GMDH for Short Term Load Forecasting. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*. 2012;42(3):345-356.

[141] Y. Nakagawa and A. Rosenfeld. A note on the use of local min and max oprations digital picture processing. *IEEE Trans. Systems Man Cybernet SMC, 8(8):632-635,1978*

[142] H.J.A.M. Heijmans. *Morphological Image Operators*. Academic Press, Boston, 1994.

[143] J.L. Kelly. *General Topology*. Springer-Verlag,1975.

[144] H. Minkovski. Volumen und oberflache. *Math, Annalen*, 57:447-495,1903.

[145] E. Styvaktakis, M. H. J. Bolen, and I. Y. H. Gu. Expert system for classification and analysis of power system events. *IEEE Power Engineering Review*, 22(2):64-64, February 2002.

[146] S.R. Sternberg. Grayscale morphology. *Computer Vision, Graphics, and Image Processing*, 35(3):333-355, 1986.

[147] R. van den Boomgaard. *Mathematical morphology: Extensions towards computer vision*. PhD thesis, University of Amsterdam, 1992.

[148] T. Y. Ji, Z. Lu, and Q. H. Wu. Optimal soft morphological filter for periodic noise removal using a particle swarm optimiser with passive congregation. *Signal Processing.*

[149] M. H. Sedaaghi and Q. H. Wu. Real-time implementation of grey-scale morphological operators. *Electronics Letters*, 33(21):1761-1763, July 1997.

[150] A. Gasteratos, I. Andreadis, and P. Tsalides. Fuzzy soft mathematical morphology. *Vision, Image Signal Processing, IEE Proceedings*, 145(1):41-49, 1988.

[151] M.S. Hamid, N.R. Harvey, and S. Marshall. Genetic algorithm optimisation of multidimensional grey-scale soft morphological filters with applications in archive film restoration. *Circuits and Systems for Video Technology, IEEE Transactions*, 13(5):406-416, May 2003.

[152] P.K. Ghosh and K. Deguchi. *Mathematics Of Shape Description: A Morphological Approach to Image Processing and Computer Graphics.* John Wiley and Sons, Asia, 2008.

[153] J. Goutsias and H. J. A. M. Heijmans. Nonlinear multiresolution signal decomposition schemes - Part I: Morphological pyramids. *IEEE Transactions on Image Processing*, 9(11):1862-1876, 2000.

[154] P. Maragos. Pattern spectrum and multiscale shape representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(7):701- 716, July 1989.

[155] G.J.F. Banon, J. Barrera, and U. de Mendonça Braga-Neto (Eds.). *Mathematical Morphology and its Applications to Signal and Image Processing: Proceedings of the 8th International Symposium on Mathematical Morphology*. Ministério da Ciência e Tecnologia, Rio de Janeiro, Brazil, 2007.

[156] L.J. Zhang, J. W. Xu, J. H. Yang, D. B. Yang, and D. D. Wang. Multiscale morphology analysis and its application to fault diagnosis. *Mechanical Systems and Signal Processing*, 22:597-6102008.

[157] P. Sun, Q. H. Wu, An Improved Morphological Approach to background normalization of ECG Signals. *IEEE Transcation on Bimedical Engneering*, 50(1): 117-121,2003.

[158] L.J. Zhang, J.W. Xu, J.H. Yang, D.B. Yang, and D.D. Wang, Multiscale morphology analysis and its application to fault diagnosis. *Mechanical System and Signal Processing*, 22:597-610,2008.