2017

FORECASTING THE WORKLOAD WITH A HYBRID MODEL TO REDUCE THE INEFFICIENCY COST

Xinwei Pan
University of Kentucky, xinwei.pan@uky.edu
Digital Object Identifier: https://doi.org/10.13023/ETD.2017.201

Xinwei Pan, Student

Dr. Wei Li, Major Professor

Dr. Haluk E. Karaca, Director of Graduate Studies

FORECASTING THE WORKLOAD WITH A HYBRID MODEL TO REDUCE THE INEFFICIENCY COST

THESIS

A thesis submitted in partial fulfillment of the
requirement for the degree of Master of Science in Mechanical Engineering
in the College of Engineering
at the University of Kentucky

By
Xinwei Pan
Lexington, Kentucky
Director: Dr. Wei Li, Assistant Professor of Mechanical Engineering

Lexington, Kentucky
2017

ABSTRACT OF THESIS

FORECASTING THE WORKLOAD WITH A HYBRID MODEL TO REDUCE THE INEFFICIENCY COST

Time series forecasting and modeling are challenging problems during the past decades, because of its plenty of properties and underlying correlated relationships. As a result, researchers proposed a lot of models to deal with the time series. However, the proposed models such as Autoregressive integrated moving average (ARIMA) and artificial neural networks (ANNs) only describe part of the properties of time series. In this thesis, we introduce a new hybrid model integrated filter structure to improve the prediction accuracy. Case studies with real data from University of Kentucky HealthCare are carried out to examine the superiority of our model. Also, we applied our model to operating room (OR) to reduce the inefficiency cost. The experiment results indicate that our model always outperforms compared with other models in different conditions.

KEYWORDS: ARIMA; Artificial Neural Networks; Hybrid model; Time series forecasting; Operating room scheduling.

Xinwei Pan
Student's Signature

04/10/2017
Date

FORECASTING THE WORKLOAD OF SURGEONS WITH A HYBRID MODEL TO REDUCE THE INFFICIENT COST

By

Xinwei Pan

Dr. Wei Li
Director of Thesis

Dr. Haluk E. Karaca
Director of Graduate Studies

04/10/2017
Date

ACKNOWLEDGMENTS

I would like to thank my advisor Dr. Li first. His efforts and experiences help me a lot to perfect my case study and complete thesis writing. Whenever I had a trouble in researching, Dr. Li could always provide efficient advises to steer me solve the problems.

I also would like to thank my committee members: Dr. Jawahir and Dr. Badurdeen. I am grateful to them for their critical comments on this thesis. I thank all of my lab mates, Amin, Feidi, Honghan, and Vivek, their helps on my research and life make me work more efficiently.

Finally, I would like to express my sincerest gratitude to my parents. They encourage me when I am lost, support me when I have to face some difficulties. No matter what happens, they will always stand behind me.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1 Introduction

1.1 Background and motivation

With the growth of demand in the healthcare industry, healthcare cost exceeds $2 trillion in these years (Cheng et al., 2015), and preventive cost control is becoming one of the most important challenges for healthcare. The majority of the annual budget of healthcare spend on operating room (OR) (Wijngaarden, Scholten, & Wijk, 2012); among the cost of OR, the salaries of staff (include surgeons, nurses, and anesthesia, etc.) account for the largest proportion (Alex Macario, Terry S., Brian, & Tom, 1995). Therefore the managers in a hospital prefer using the OR block time and related resources or equipment efficiently (Hosseini & Taaffe, 2015). Most industry information resources indicate that the acceptable utilization for the OR should between 75% and 80%. For example, the American Hospital Association uses 75% as a guideline

Under the capitated hospital contracts, anesthesia groups need to provide all services of anesthesia to a given number of patients who are insured, and surgical groups are paid with a fixed salary each month. Under both capitated revenue and salaried settings, the anesthesia team prefers minimizing the cost of staff by allowing the surgical group to allocate efficient block time to finish its surgical cases without too much overtime or idle time. If allocating too much block time to the surgical groups, it results in extra idle time, which is a waste of recourses; oppositely, if underestimating the block time, it will generate over-utilize cost (Dexter, Macario, Qian, & Traub, 1999).

As a result, the estimation of the demand (number of patients) or workload (total case time during a period) for OR cases is central to OR block planning and scheduling problems. However, underestimating demand has a series of negative impact on the hospital, such as extending the staff's working time, increasing the waiting list of patients, and reducing patients' satisfactory. On the contrary, overestimating demand also causes some troubles, such as lowering OR utilization, occupying the resources unnecessarily. Plenty of inefficient costs will be generated by the deviation between forecast value and actual value.

Although most hospitals predict the workload or block time by surgeons' estimation, the surgeons' estimation is not accurate (Fügener, Schiffels, & Kolisch, 2015). Therefore, an accurate prediction of workload is valuable to OR block scheduling. From the perspective of operating room manager, it is much more beneficial if they hire the fewest staff and keep the quality of service at the same time (Dexter et al., 1999). Under the settings above, the day or week on which to perform cases should be selected to match the time period, in which the human resources should be scheduled to operating room. From the perspective of the OR staff, an accurate prediction of the workload will smooth their workload which helps to keep them a good working condition and emotion. To some extent, it is beneficial to improve OR utilization and benefits.

From the perspective of a whole hospital system operating over a long time period, we can carry out a long-term forecast procedure (e.g., one to three years). It helps the hospital to determine the budget, which is used to deal with the demand in the following years (e.g., constructing more operating rooms, purchasing medical instruments). An accurately predetermined volume is so important as the hospital wish to ensure that both the quantity and quality of service delivered by the budget meet a predetermined set of volumes and standards (Dredge, 2004).

1.2 Challenges in time series forecasting

Demand forecasting with time series is useful for OR block planning and scheduling, but challenging. Forecasting the time series is regarded as one grand challenge in modern science. (Cheng et al., 2015). Obtaining an efficient and accurate forecasting value of time series are still challenging problems, because of the different mixed relationships of non-linearity, linearity and periodic behaviors presented. In statistic, time series can be thought of non-linear dynamic systems if they are marked as non-linear causal factors such as irregular, non-periodic, asymmetric, multimodal and lagged variable relations, predictions of changes in state space and time irreversibility, etc. (Fan & Yao, 2013).

Predicting workload of operating room has been a crucial and challenging application of time series based models. This modeling method is useful in practice while little information is available on the underlying relationships of time series when no available statistical models are able to relate the objective variable to other explanatory variables.

As the time series contains different kinds of features, such as linear and nonlinear, it presents properties which are difficult for us to capture and describe precisely with one single model (G P Zhang, 2003). In fact, the most recent models do not perform well in predicting the workload of the OR. We want to build a new model to improve the forecast accuracy, which can be applied to OR workload forecasting. A hybrid model could be more efficient when predicting the workload for healthcare. The advantages of this approach come from the following aspects (G P Zhang, 2003):

Usually, it is hard to tell where the time series comes from, or in practice, to determine if a forecast model is able to obtain a better performance among several models. As a result, it is a difficulty for researchers to choose an optimal model for their unique situation. In practice, several different models are applied and compared, and then the model, which generates smaller errors among several methods is selected. However, the choice of models is not necessarily based on the minimum error for current data because of many underlying factors for future application, such as the variability of the sample, uncertainty of models and structural changes. When combining different approaches, model selection can be based on additional efforts.

Besides, time series are seldom only contain linearity or non-linearity in the real word. They usually present both properties. Under this condition, both ARIMA and neural networks are not sufficient for modeling and forecasting because when ARIMA models deal with non-linear relationships, it is inefficient; whereas neural network models alone cannot model and forecast both linear and non-linear patterns simultaneously. Therefore, while combining the ARIMA and ANN models, the underlying relationships among time series can be described and captured in a more efficient way.

The last point is that in the literature about time series modeling and forecasting, none of the methods is the best in each case. This is mainly because application problems in the real world are usually complicated in reality, none single forecasting method can describe the same pattern in the same way. Many empirical studies (Makridakis et al., 1993; Makridakis & Hibon, 2000), including several large sample size predictive competition, show that hybrid models, which are combining several different methods, the forecasting accuracy could be improved. Thus, hybrid models are quite useful because they increase the chances of capturing underlying patterns in time series and improve the performance of whole model. A few empirical studies have shown that hybrid models outperform a single model with prediction accuracy increased. Additionally, the hybrid model has more robustness to possible structural changes in time series.

1.3 Problems and shortcomings in literature

Because of the complexity of time series, the specific structure of the measured patterns is still unknown, and in most cases, it is not certain. As what we presented in previous paragraph, it is generally agreed that the past and future observations are related by linearity or simple nonlinearity. However, the actual time series indicates non-linear properties and can be regarded as non-stationary as well. Thus, some classical and crucial linear methods, such as autoregressive (AR) and moving averages (MA), are not able to capture and describe the nonlinear relationships because of the disadvantages in predicting the complicated patterns.

The current predictive model cannot be used universally, because a single model captures and presents only part of properties. The various properties of time series pose a significant challenge to obtain accurate results. The time series appearances from such a complicated system exhibits non-periodic pattern even in a stable condition. Besides, since real-world time series often present underlying relationships, the relationships obtained from them tend to present numerous forms of non-stationary. However, the methods proposed in the literature emphasis predicting linear and fixed time series (Cheng et al., 2015).

From the related research, some authors apply time series prediction to OR demand, but they did not obtain acceptable accuracy using their models (Dexter et al., 1999). For some other papers, they focus on estimating the duration or variation of individual cases based on the past mean value (A Macario & Dexter, 1999). Nevertheless, the application did not show persuasive results to convince the reviewers that their approach was beneficial to the hospital. Although the bias from this method is close to 0, the absolute error is 1.3 hours (the average duration is equal to 3.06 hours) (Dexter et al., 1999). To a certain extent, these studies cannot help us to achieve useful results in predicting the demand of OR.

Over the past a few decades, several developments were made in capturing patterns from the time series and the nonlinear structure of the underlying system. Recent reviews of time series prediction models (De Gooijer & Hyndman, 2006) provide extend coverage of important prediction models and some other models; and neural network models. However, there is still no comprehensive review of the latest developments in the prediction models for nonlinear and non-stationary complex systems. The primary objective in this thesis is to construct a hybrid model for predicting time series in the real world, which improves the accuracy of forecasting and obtains a better performance when applied on OR.

1.4 Contribution

The contribution of this thesis is as following aspects: (1) we develop a new Hybrid-Filter model to solve problems in time series forecasting, which is used to obtain accurate results; (2) we carry out a case study to examine the priority of our model when compared with classical models; (3) another case study to examine the validation of our model in an application problem.

1.5 The structure of this thesis

This thesis is organized as follows: In Chapter 2, we will carry out literature review on forecast models; a newly built model is presented in Chapter 3; and in Chapter 4, we provide two case study, one is for comparing the accuracy with three models, another is for the performance in an application. In Chapter 5, we will summarize this thesis and talk about the future work.

Chapter 2 Literature Review

The time series is defined as a series of data, which are co-related with each other over consecutive times. A univariate time series only contains a single variable. Similarly, when over one variables are under consideration, it is termed as a multivariate time series. We mainly carry out literature review on univariate time series forecasting instead of multivariate time series in this section.

2.1 Difficulties in time series forecasting

The prediction and modeling for time series are fundamentally important in various practical areas. As a result, a lot of research works on this topic have been active for several decades. Many crucial forecasting methods are developed in the literature to obtain better results and validity when applied to prediction. To investigate these forecasting methods, we will discuss some classical models, as well as their inherent advantages and disadvantages.

In recent years, manufacturing and healthcare systems faced a trade-off problem of high quality products and services with limited and finite resources. We can solve this problem sufficiently through statistical methods. Particularly, the accurate forecasting demand are necessary for lots of activities, fore example, budgeting, marketing, financial planning and inventory control. The overall decision-making process can benefit from accurate demand forecasting and modeling tools.

Whenever we try to model a process behavior, we have to deal with the natural variation it presents. This natural change is confused with the errors while measuring the model. In the case of data collected from some different time series, we must also consider the correlation among the data presented in the recorded value. These kind of approach has been shown to be useful for the characterization and prediction of such time-dependent processes.

One of the key challenges to forecasting theories and practice is that time series data are not stationary in reality, because they always contain different features and various factors that may change established relationships and patterns . In Makridakis and Hibon's M-3 competition (Makridakis & Hibon, 2000), authors draw four conclusions after comparing various forecast models with different forecast horizon and measures, and one critical view is that the accuracy of the hybrid methods outperforms specific individual methods, and the errors generated by hybrid models are relatively smaller compared with other models.

In the back-propagation (BP) neural network model, there are a lot of neurons in different layers, which are able to capture complex phenomena. The principle of the model is to adjust the parameters iteratively and minimize the prediction error. When dealing with small scale data, Liu et al. (Liu, Kuo, & Sastri, 1995) indicated that the neural network model generated better results compared with the other statistical model. According to previous studies, although the BP neural network model ignores the prediction error when building the model, it can still generate good prediction. The errors between output nodes and objective value provide feedback to update the weights and threshold, but the errors do not affect the value of the input nodes. Therefore, this lost part of features by the ANN could be crucial (Tseng, Yu, & Tzeng, 2002).

Many papers suggest that combine two different methods and take advantages of both models is a possible solution to obtain better accuracy than a single model. Thus, in this paper, we combined ANNs and ARIMA together, and integrate a filter structure to further improve the accuracy. The newly built hybrid model benefits from different positive properties of ANNs and ARIMA. In next section, we will introduce our method in details.

2.2 Models in time series forecasting

In this section, we are going to present some classical methods used to de describe the time series. Some of the following models are used as control groups in chapter 4, and compare the results from these models with our new Hybrid-Filter model.

2.2.1 Exponential smoothing method

In 1957, Holt (Holt, 2004) proposed a forecasting model for extending simple exponential smoothing method. The model has the following properties: (1) reducing the weighting parameter put on older data and giving more significance to recent data; (2) less computing complexity; and (3) less required data. This model takes the advantages of these properties to smooth current random fluctuations and to continuously revise seasonal trend. In 1960, Peter (Peter, 1960) extended Holt's method by adding seasonality to the former algorithm, which is known as the Holt-Winter method. Taylor (Taylor, 2003) improved Holt-Winter method by introducing multiple-seasonality trends. However, as these models contain trend equations, which are fixed when forecasting multi-step ahead future data, they cannot be used to forecast the turning point.

2.2.2 Autoregressive Integrated Moving Averages (ARIMA)

During the past few years, the research field of modeling time series draw lots of researchers' attention. The main purpose of modeling the time series is to obtain and describe the relationships among past data to build an available method that captures the underlying relationships in the time series (R.K, Agrawal, Adhikari, R.K, & Agrawal, 2013). And then use the previous method to predict the value or condition in the future. Because of the indispensable importance of time series, we should give appropriate attention to adapting the appropriate model to the underlying time series in various practical areas, for example, business, economics and engineering. The accurate time series prediction depends on the fitness of the model. Researchers have done a lot of effort over the years to develop efficient models to improve prediction accuracy. Therefore, a variety of important prediction models has been developed in the literature during the past decades.

A classical time series model is Autoregressive Integrated Moving Average (ARIMA) model. The most important assumption of implementing the model is that the time series is considered as linearity and follows a specific probability distribution. What’s more, the Autoregressive (AR)， moving average (MA) and autoregressive moving average (ARMA) are subclasses of ARIMA model. Box and Jenkins (Box & Jenkins, 1971) proposed a fairly successful change in the ARIMA model for seasonal time series modeling, which is seasonal ARIMA (SARIMA). The ARIMA model is popular because of its flexibility, it can simply describe and interpret time series, and the associated Box-Jenkins method is used to minimize the error between model and actual value. However, there are still some limitations in these models, for example, assumption of normal distribution of original data, which are deficient in many practical conditions. To solve this shortcoming, some methods for nonlinear time series are proposed in the literature; however, these are not as straightforward and simple as the ARIMA model from the point of view for implementation.

The feature of a time series based demand which is stochastic has been predicted and modeled during the past a few decades frequently by ARIMA. ARIMA model is such classic that it is used as a complicated benchmark for comparing a couple of alternative methods within a univariate framework (Abraham & Nath, 2001; Darbellay & Slama, 2000; Taylor & Buizza, 2003).

Taylor et al. (Taylor, de Menezes, & McSharry, 2006) compared several univariate forecast methods, for example, seasonal ARMA methods, exponential smoothing for double seasonality, neural networks, regression method with principal component analysis (PCA) and benchmark method $\widehat{y_t} = y_{t-1}$. To measure the accuracy of these models, they use mean absolute percentage error (MAPE). As a result, the exponential smoothing method performs the best, leading to the conclusion that simple and robust methods can outperform sophisticated alternatives.

From the result, the neural network generate a bad result. Taylor thought it was not due to over-fitting or an overly complex structure of neural networks (Taylor et al., 2006), these are two general drawbacks when modeling with the neural network. The bad results may come from the following aspect: they did not divide the data into two categories, because weekday and weekend follow different distribution. Another possible reason is that the sample size used to train an ANNs is too small, the nonlinear relationship may be concealed.

2.2.3 Artificial neural networks (ANNs)

From the past a few decades, artificial neural networks (ANNs) drawn lots of attention in the field of prediction. Although it is based on the initial biological inspiration, later ANN was used in some other practical fields, such as for the purpose of prediction and classification. When applied to the prediction of time series, the superiority of ANN is inherent in nonlinear modeling, and there is no pre-estimation of the statistical distribution of the observed results. Appropriate models are adaptively formed according to the given data. For this reason, ANN is data driven and adaptive. Recently, a variety of researches were carried out in the field of the application of time series prediction with neural network. Some import points about the application of time series forecasting with ANNs was presented by Zhang et al. in 1998 (Guoqiang Zhang & Eddy Patuwo, 1998). Several neural network methods are proposed in the literature. One of these methods is the multi-layer perceptions (MLPs), which only contains a single hidden layer Feed Forward Network (FNN) (G P Zhang, 2003). Another form of FNN is the Time Lagged Neural Network (TLNN) (Kihoro, J. M., R. O. Otieno, 2004). In 2008, C. Hamzacebi (Hamzaçebi, 2008) constructed a new NN model, which is applied to prediction on seasonal patterns. This method is quite easy, and the experiments were proved to be useful and effective in predicting time series which contain seasonality. Besides, there are many different neural network structures in the literature due to the ongoing research work in this area.

2.2.4 Support Vector Machine (SVM)

Recently, an important methods on the prediction of time series is Vapnik's (Vapnik, 1998) support vector machine (SVM). Although initially SVM was designed to solve the classifying the patterns, it was later widely used in lots of different areas, for example, regression analysis, and time series analysis. Therefore, the SVM method becomes a crucial model to deal with nonlinear data, especially for the modeling and forecasting. The main points of the SVM model is applying the structural risk minimization (SRM) theory to find the optimal decision rules. Only the subset of training data determines the results to a specific patterns. Another crucial feature of SVM comes from the following point, the training here is equal to solving the problem of linear constrained quadratic optimization. Unlike the neural network model, SVM generated results are always an optimal solution. Another feature of SVM is that we can independently control the quality, regardless of the size of the input data. Generally, in the SVM methods, mapping the input nodes to a higher dimensional space, with some special functions, called support vector kernel, even in high-dimensional often produce good promotion. In the past few years, researchers have developed many SVM prediction models.

Suykens and Vandewalle (J. A. K. Suykens & Vandewalle, 1999) applied SVM on pattern recognition problems. The authors build an SVM model base on the least square version for classification problems. The ridge regression interpretation of support vector has been given in for the function estimation problem, which considers type constraints instead of inequalities from the classical SVM approach.

Cao and Tay (Cao & Tay, 2003) carried out a research on financial time series forecasting. Their experiment results show that both SVM and regularized Radial basis function (RBF) neural network minimize the regularized risk function. What's more, for dealing with the vibration in the financial data, they proposed adaptive parameters method, which is updating larger weights on the latest value and smaller weights on older value. However, the adaptive parameters and weights cannot be guaranteed to be an optimal solution. Thus, an efficient method to search the best combinations of parameters should be investigated in the future work.

For extending the least square based SVM, Suykens and Vandewalle (J. Suykens & Vandewalle, 2000) introduced recurrent LS-SVM's for static nonlinear function estimation problems. They use finite numbers of case to assign large penalties to extreme value in this paper. The authors indicated that LS-SVM's is able to generalize well, despite there are only small size of training data sets. Their works provide some fresh aspects to modeling the time series which presents nonlinearity.

Raicharoen et al. (Raicharoen, Lursinsap, & Sanguanbhokai, 2003) applied Critical SVM (CSVM) to approximate functions for signal processing and time series. The CSVM consists of the SVM, the Perceptron, and the Nearest Neighbor Algorithm. Their algorithm presents significant performances on the chaotic Mackey-Glass time series. Moreover, the CSVM is easy to implement. Farooq et al. (Farooq, Guergachi, & Krishnan, 2007) proposed a model based on Least-Squares SVM to forecast long term chaotic time series. The results of case studies indicate that this methods obtains better results compared with related methods for Mackey Glass-30 long term chaotic time series prediction task.

2.2.5 Autoregressive Conditional Heteroskedastic (ARCH)

Garcia et al.(Garcia, Contreras, van Akkeren, & Garcia, 2005) proposed a new method based on the generalized autoregressive conditional heteroskedastic (GARCH) to predict one-step-ahead electricity prices in hours, which is used to modeling and prediction. The average forecast errors, which are measured with forecast mean square error (FMSE), are around 9%. When consider the complexity in time series for recoded electricity price, the errors are acceptable. Besides, the GARCH model outperforms

ARIMA model when volatility and spikes present. Considering the demand as an explanatory variable, the GARCH model improves the performance as well.

2.3 Hybrid models in time series forecasting

The hybrid prediction model is a well-constructed method for obtaining better accuracy of forecasting. This hybrid approach utilizes the advantages from different forecasting methods, such as the ability to capture different time series patterns (de Menezes, W. Bunn, & Taylor, 2000).

There are a lot of methods available for us to combine similar models together. Draper (Draper, 1995) developed a new method to gain stable results by the average of different models. Likewise, Zou and Yang (Zou & Yang, 2004) used the aggregated forecast through exponential re-weighting method (AFTER) to mix the forecasts from the different individual autoregressive moving average (ARMA) models. Nevertheless their research only focus on combinations of linear methods with different parameters. The advantage of combining similar models together is to obtain a stable model with high robustness and enhance the strong point of the similar models (Yu, Wang, & Lai, 2005). To the contrary, the disadvantages of these models are also augmented.

To deal with the previous drawbacks, combining different kinds of models has been studied during past a few decades. Tseng et al. (Tseng et al., 2002) developed a new method which combines BP-ANN model and seasonal time series ARIMA (SARIMA). A crucial limitation of SARIMA model comes from the following aspect: too many of historical data is required. However, as various factors of uncertainty changing rapidly, we always need to forecast with limited historical data. To make up the shortcomings of SARIMA, the author hybrid neural network, which performs well with small sample size with SARIMA. What's more, their model takes advantage of SARIMA to help the NN model to reduce the forecasting errors. The result also supports the author's model as their hybrid SARIMA-BP model outperforms other models with the lowest error. However, as the precondition of SARIMA, this model cannot be applied to non-seasonal time series.

Zhang (G P Zhang, 2003) also suggested that many studies, including some predictive experiments with large sample size, show that when a few different predictive models are combined together, the accuracy of the prediction is usually better than a single model and there is no need to find a "true" or "best" model (Makridakis et al., 1993). In this paper, the author combining two models, ARIMA model ANNs, the hybrid method takes the advantages of ARIMA model which captures linear features and ANNs model which focuses on nonlinear features. The results of an empirical case study with three real datasets, evaluated by MSE (mean squared error) and MAD (mean absolute deviation), show that their hybrid model outperforms each single model are used separately.

Similarly, Pai and Lin (Pai & Lin, 2005) suggested that by combining various prediction methods, they can capture lots of types of patterns of data sets at the same time. From literature, the conclusion from a variety case studies indicates that when combining different forecasting models, the results obtain from the hybrid models are better than the results come from single methods, on average (Makridakis, 1989). A hybrid model which combining AIRMA and SVM model was presented, the experimental results show the overall errors between forecasting values and actual values can be significantly reduced with a hybrid model.

2.4 Accuracy measurements in time series forecasting

Many evaluation methods of forecasting developed in the past several years. However, the most of these measures cannot be universally applicable, as some measures can be infinite and others can be undefined in different cases, under this condition, these measures would generate misleading results and make reviewers confused (Hyndman & Koehler, 2006).

From the papers in the past, most textbooks and papers (e.g., Bowerman, O'Connell, & Koehler, 2004) use MAPE (Mean Absolute Percentage Error) as an evaluation measurement. In contrast, Makridakis (Makridakis, et al., 1998) warn against the use of the MAPE in some specific conditions. Other measuremetns such as median relative absolute error (MRAE), Geometric Mean Relative Absolute Error (GMRAE) and Median Absolute Percentage Error (MdAPE) were recommended by Fildes (Fildes, 1992). The MdRAE, Symmetric Median Absolute Error (sMAPE) and Symmetric Median Absolute Percentage Error (sMdAPE) were presented in the M3-competition (Makridakis & Hibon, 2000). In addition, some other researchers (Tseng et al., 2002; G. Peter Zhang, 2007) evaluated their results with MSE, MAE (Mean Absolute Error) and MAPE. To determine choose which measures to evaluate forecast errors, Hyndman and Koehler (Hyndman & Koehler, 2006) compared these measures with different data sets.

The authors suggest that the new proposed scaling error method MASE (Mean Absolute Scaled Error) can be a standard method of measuring error. This method is applicable in a wide range unless it is always finite and well defined in all cases where historical data is equal and irrelevant. When the value of MASE is greater than 1, the prediction result is worse than the simple method. However, in some cases, the existing measurement method is still preferred. For example, when all the data is on the same order of magnitude, MAE can be the preferred method of measuring accuracy as it is well interpretable; when all the observations are much greater than zero, MAPE is a better choice because of a simpler calculation difficulty.

2.5 Application of time series forecasting on operating room (OR)

Dexter et al. (Dexter et al., 1999) proposed a case study to forecast the workload of surgical group's elective cases for allocating block time. The author compared the configuration of time series forecasting, for instance, the number of previous time periods of observations applied to optimize the model in prediction, and then generate an optimal method comparing the error. The results generated by time series analysis were used to calculate the upper bound of cases which will be finished in the future. Dexter suggested that their results could help managers to deal with the variations in future OR demand. What's more, this method can be applied to compute how many numbers of OR block time should be allocated to surgeons, which is used to minimize the labor costs. However, the error generated by their method is over 20% evaluated by MAPE, and some more sufficient forecast models should be applied to improve the accuracy. Also, a case study about how this method help to reduce labor cost should be carried out to convince us.

To increase the OR efficiency, Stepaniak et al. (Stepaniak, Heij, Mannaerts, De Quelerij, & De Vries, 2009) provided a statistical model to forecast the duration of surgical cases. From the obtained results, Stepaniak et al. suggested that the scheduling of OR cases can be improved by applying a 3-parameter logarithmic model with surgical effect and by using a surgeon's prior guess for a rarely observed CPT. The underestimation and overestimation of OR time and OR inefficiency were significantly reduced in each case from the results.

In next section, we are going to introduce our hybrid model.

Chapter 3 Methodology

Analysis of historical data with time series provides more insights of the data than simple statistics, such as averages or means, variance or standard deviation, skewness, and kurtosis. In this section, we present two popular methods, the autoregressive integrated moving average (ARIMA) method and the artificial neural networks (ANNs) method, for demand forecasting with time series.

3.1 Autoregressive Integrated Moving Average (ARIMA)

ARMA(p, q) is a simple combination of AR(p) and MA(q) models. At first, we introduce the AR(p) model and MA(q) model. In an AR(p) model, we assume the unknown value y_t consists of past observations y_{t-i} with accumulated errors ε_t and a constant c, and the AR model is as following:

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \varepsilon_t \tag{3-1}$$

where φ_i is a parameter and, p is the number of autoregressive terms, $i = 1,2,\dots,p$. If $p = 1$, we call it first-order autoregressive model, in this condition y regress on itself lagged by one period (see equation 3.2).

$$y_t = c + \varphi_1 y_{t-1} + \varepsilon_t \tag{3-2}$$

In some papers, the AR model is specified in terms of lag operators. The lag operator L is defined as:

$$L y_t = y_{t-1} \tag{3-3}$$

Combining equation 3.1 and 3.2, then we obtain:

$$\varepsilon_t = \varphi(L)y_t = (1 - \sum_{i}^{p} \varphi_i L^i)y_t \tag{3-4}$$

Similarly, the MA(q) model is:

$$y_t = \mu + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{3-5}$$

$$y_t = \theta(L)\varepsilon_t = (1 + \sum_{j=1}^{q} \theta_j L_j)\varepsilon_t \tag{3-6}$$

where μ is the expectation value of y_t, usually it is the average of past q values, θ_j is a coefficient for $j = 1,2, \dots, q$, q is the number of lagged forecast error, ε_t describes the stochastic error. When $q = 1$, we call it first-order moving average model, it is defined as:

$$y_t = \mu + \theta_1 \varepsilon_{t-1} + \varepsilon_t \tag{3-7}$$

We can obtain a more general and useful model ARMA(p, q) by simply combining AR(p) and MA(q) models because ARMA(p, q) is able to capture more features of time series. The new model is as following:

$$y_t = c + \sum_{i=1}^{p} \varphi_i y_{t-i} + \sum_{j=1}^{q} \theta_j \varepsilon_{t-j} + \varepsilon_t \tag{3-8}$$

We can also describe ARMA model in a lag operator notation. Thus the ARMA(p, q) model are given as follows:

$$\varphi(L)y_t = \theta(L)\varepsilon_t \tag{3-9}$$

If a time series has constant mean and standard deviation when the time changes, we call it stationarity or stationary time series. When the state of time series is non-stationary and we applied ARMA model, some problems occur, such as the prediction results are biased and invalid. Thus the prerequisite of application of ARMA model described above is stationary time series. Then ARMA model is proposed to deal with non-stationary time series. For changing non-stationary data to stationary, in ARIMA model we compute the differences of original data y_t and make it a new data set y_t^*. The mathematical formulation of ARIMA(p, d, q) can be represented as follows:

$$\varphi(L)(1-L)^d y_t^* = \theta(L)\varepsilon_t \tag{3-10}$$

The integer d is the level of differencing. If $d = 0$, we obtain:

$$y_t^* = y_t \tag{3-11}$$

In this condition, the ARIMA is the same as ARMA model. If $d = 1$:

$$y_t^* = y_t - y_{t-1} \tag{3-12}$$

When d=2, the y_t^* is defined as:

$$y_t^* = (y_t - y_{t-1}) - (y_{t-1} - y_{t-2}) = y_t - 2y_{t-1} + y_{t-2} \tag{3-13}$$

As only if the time series is stationary, then we are able to apply ARMA model to forecast the future condition of a time series. Therefore, we carry out an Augmented Dickey-Fuller test (ADF test), which is applied to examine whether unit roots present in a time series or not. If the unit root exists, then the time series is non-stationary. After constructing a new time series by calculating the differences of previous time series, we test the stationarity. These procedures repeated until the null hypothesis is rejected, which means we have already obtained a stationary time series, the times of differencing is recorded as d. As the time series is the differences of original time series, we need to return to the original condition after we obtain forecast value.

To determine the value of the lag operators of AR(*p*) and MA(*q*), we need to set up a range of *p* and *q* (usually they are around *T*/10, ln(*T*) or *T*^0.5, *T* is the period) first, from 1 to N (both *p* and *q* are integers). We calculate the value of Akaike information criterion (AIC) of every combination of *p* and *q* (totally N*N combinations), the combination of *p* and *q* with a minimum value of AIC is reserved as the lag operators of ARIMA (*p*, *d*, *q*) model.

For example, assuming we are going to construct an ARIMA model with the 15 past observations. At first, we should carry out an ADF test to examine the stationarity of the time series.

Table 3.1 statistic of ADF test (H=0, null hypothesis, H=1, alternative hypothesis)

H	*p*-Value
1	0.0183

The results from Table 3.1 show that because the confidence level is at 5% level (*p*-value is less than 0.0183), the null hypothesis is rejected. If the *p* value is larger than 0.05, then we should compute the differences of the original time series and make them a new data set. This procedure will repeat until the data is stationary.

The range of *p* and *q* is from 1 to 3, which means we have 9 combinations of *p* and *q* values. We compare the value of AIC which is calculated with different combinations of *p* and *q*. The combination of *p* and *q* is recorded with the minimum value of AIC. Thus, an ARIMA(*p*, *d*, *q*) is completed and ready for forecasting the demand of OR in the following days.

Table 3.2: All combinations of p and q with AIC values.

P	*Q*	AIC
1	1	15.5497
1	2	15.6258
1	3	15.1653
2	1	15.6497
2	2	15.7079
2	3	15.0611
3	1	15.5661
3	2	16.4248
3	3	16.9767

From Table 3.2, we find the minimum value of AIC is 15.0611, the corresponding combination of p and q is 2 and 3. Therefore, the final ARIMA model should be ARIMA(2,0,3), which is equal to ARMA(2,3) model as the original time series is stationary.

3.2 Artificial Neural Networks

Classic forecasting models include exponential smoothing (e.g., holt-winter method), and ARIMA, etc. One of the typical features of these models is that they are only able to capture linear properties of data with a time series. To solve the limitation of ARIMA model, researchers construct several models to capture the unique features in time series. For instance, the autoregressive conditional heteroscedastic (ARCH) model (Engle, 1982), and the threshold autoregressive (TAR) model (Tong, 1983). Although these models improve the ability to describe nonlinear time series, we have limited development in nonlinear modeling (De Gooijer & Kumar, 1992). This is because these methods are constructed for specific time series, and those method may be not efficient when applied to other types of time series.

Compared with other methods of modeling nonlinear time series, an ANN model has a crucial advantage that it can be used to model a lot of functions with goodaccuracy by using universal estimators (G P Zhang, 2003). Such advantage also comes from processing the data information in parallel. Different from ARIMA model, no prior assumption (e.g., stationary time series) is required before constructing the model. Instead, the ANN model is mainly determined by the properties of time series, such as periodicity, seasonality, etc. Therefore, ANNs have been regarded as an optional forecasting model because of its strength of handling with nonlinear patterns.

A classical feedforward network, which contains one hidden layer is most widely used for modeling and forecasting data with a time series (Guoqiang Zhang & Eddy Patuwo, 1998). This structure is a network contains three layers (i.e., one input layer, one hidden layer and one output layer) of connected neurons. We describe the structure between output (y_t) and the inputs (y_{t-1}, y_{t-2},…, y_{t-p}) by the following mathematical representation:

$$y_t = \alpha_0 + \sum_{j=1}^{q} \alpha_j g(\beta_{0j} + \sum_{i=1}^{p} \beta_{ij} y_{t-i}) + \varepsilon_t \tag{3-14}$$

where α_j are the weights between hidden layer and output layer, and β_{ij} are weights connect input layer and hidden layer (j = 0, 1, 2,…, q; i =0, 1, 2,…, p; j =1, 2,…, q); p is the number of input nodes and q is the number of hidden nodes. The equation (3-14) express one output node in the output layer of which is mostly used for day-ahead prediction. The sigmoid function is mostly used as the hidden layer activative function g(·), that is,

$$g(x) = \frac{1}{1 + e^{-x}} \quad (3\text{-}15)$$

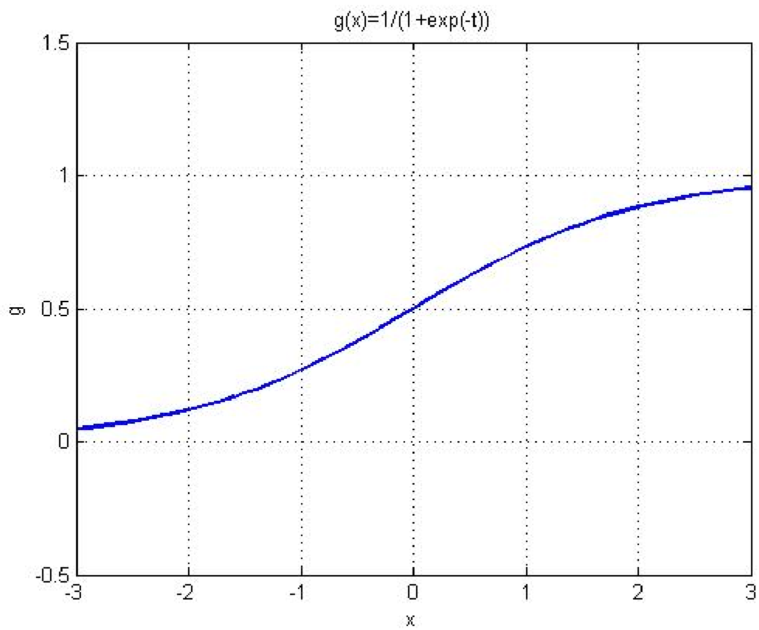

Figure 3.1: Shape of sigmoid function

With a transfer function g(·), the ANNs is able to describe the nonlinear property in time series, and then mapping the features from input layer to hidden layer. Hence, the ANN model, in fact, is used to map the data from the past observations (y_{t-1}, y_{t-2},…, y_{t-p}) to the future value y_t in a nonlinear function, i.e.,

$$y_t = f(y_{t-1}, y_{t-2},\ldots, y_{t-p}, w) + \varepsilon_t \quad (3\text{-}16)$$

where w are the weights of all coefficiceints in neural works, and f(·) is a function to map past observations to objective value through weights and activate function. Therefore, the neural network is able to model nonlinear time series. The equation given by (3-14) is a most generally used neural networks. When the nodes are large enough, this model can be used to model any type of nonlinear time series. In practice, a neural network with small numbers of nodes is able to forecast the out-of-sample data well. An overfilled model is consistent with the model established, but the generalization of the sample data is poor. Both *p* and *q* may be the most crucial parameters in the ANN model because these two parameters are quite important to determine the structure of the neural networks, besides, there are generally agreed principle to determine the value of *q*. However, no theory is able to determine the value of *p*. Therefore, experiments case studies are often carried out to determine the optimal combinations of *p* and *q*. As long as the values of *p* and *q* are fixed, the network can be trained. As in the ARIMA model establishment, the parameters are determined on the overall accuracy criterion such as minimize the mean square error.

In our model, we construct a three-layers ANN model, the number of input layer nodes is five, the hidden layer has twenty nodes and one node in the output layer. We use another sigmoid function instead of equation (3.13) as the transport function between layers, the formula is as following:

$$tansig(n) = {}^{2}\!/_{1 + e^{-2*n}} - 1 \qquad (3\text{-}17)$$

When calculating the parameters of ANN, we choose algorithm ‘traingdx’, it means that a training function according to gradient descent momentum and an adaptive learning rate, the objective of this function is to revise the parameters and then minimize the errors. As BP-ANN always generates different parameters, we repeat constructing models for 10 times and pick the model with minimum MSE after training data. For accelerating the computing procedure, we set up a training object as 1e-4.

We extract 250 historical data from UK HealthCare, the sample size for training the ANN is 240 (except the last 10 data). Every time when we training the networks, we use 15 groups of past data, 12 groups for training set and 3 groups for validation set. In each group, there are 5 data for input and 1 datum for output, each observation is used 6 times (5 times for input nodes, 1 time for output node) repeatedly. Thus totally 80 data were used to train a single networks.

3.3 Hybrid-Filter

Although different forecasting models are suitable in different background settings, none of them can be universally applied to the variety of circumstance. For example, ARIMA is successful when used in long-term and linear data sets, but perform badly in short-term and nonlinear data sets. Both theatrical and empirical studies drawn a conclusion that hybrid model outperforms single forecasting models (Pai & Lin, 2005). To take advantages of both ARIMA and ANNs, we combine these two models. Thus, the newly hybrid model can capture different aspects of various patterns.

In the real world, a time series always presents hybrid properties, and it is reasonable to assume that time series consists of linear components and nonlinear components. So that it is a combination of linear components and nonlinear components as follows:

$$y_t = \beta_1 L_t + \beta_2 N_t + \beta_0 \tag{3-18}$$

Here, y_t is actual value at time period t, L_t denotes the linear part, and N_t denotes nonlinear part, β_1 and β_2 are coefficients, β_0 is the error. L_t and N_t are computed from past observations. In this case study, we use ARIMA to compute L_t with historical data and obtain N_t from ANNs.

Assuming that we have obtained a series of y_i, L_t and N_t (e.g., i=t–15, t–14,…, t–1), then we regress these L_t and N_t on y_i. However, there must be some residuals between y_i and L_t and N_t. Therefore, we also need to model the residual. After that we get a regression model, that is:

$$\widehat{y_i} = \beta_1 \hat{L}_i + \beta_2 \widehat{N}_i + \beta_0 \qquad (3\text{-}19)$$

Here, β_0 denotes the error; $\hat{L}_i$ is the estimate linear value, $\widehat{N}_i$ is the estimate nonlinear value. We can just plug in $\hat{L}_t$ and $\widehat{N}_t$ to the regression function above while computing $\widehat{y_t}$.

However, not all $\widehat{y_t}$ obtained from regression model are final outputs. For further reducing the error of the hybrid model, we lead in the concept of control chart Russo (Russo, Camargo, & Fabris, 2012) presented an application of control charts ARIMA for autocorrelated data, this method can be integrated into the filter structure to deal with the time series data which is obviously autocorrelated. A filter structure F is integrated into the hybrid model to filter 'potential inaccurate' $\widehat{Y}_t$.

To determine whether it is accurate or inaccurate, we use the concept of quality control charts. While $\widehat{y_t}$ is between $\mu + a * \sigma$ and $\mu - b * \sigma$, we deem it is accurate; oppositely, while the potential forecast value $\widehat{y_t}$ is larger than $\mu + a * \sigma$ (μ is mean demand of previous days, σ is standard deviation of previous a few days, a is a coefficient) the $\hat{y}_t$ is judged 'inaccurate', and then we use another value as the final output. Similarly, when $\hat{y}_t$ is less than $\mu - b * \sigma$, we use the value to replace $\hat{y}_t$ as well.

Figure 3.2 shows how our model works. We obtain the upper bound and lower bound by past observations. When the output $\hat{y}_t$ exceeds upper bound or lower bound, the filter structure picks 'inaccurate data' (e.g., point 2, 4 and 9 in figure 2) and we use output generated by the mean of past observations to replace the original output $\hat{y}_t$.

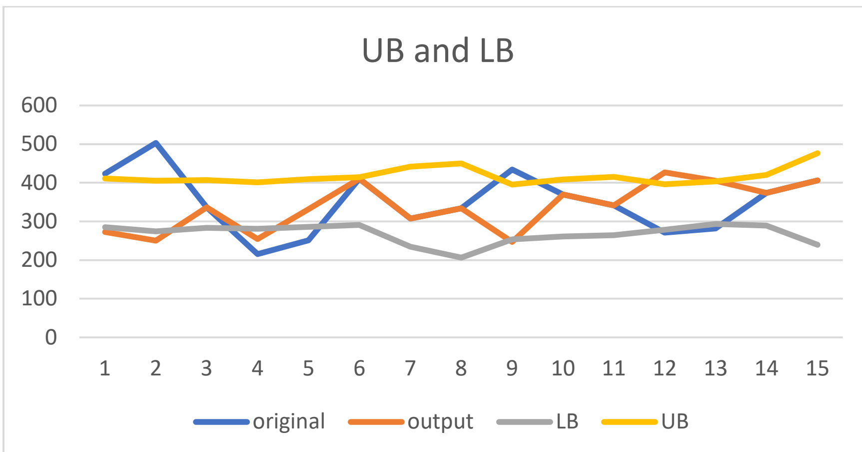

Figure 3.2: Upper bound and lower bound of workload

An optional filter structure we take into consideration is Logistic Regression. The formula of logistic regression is as following:

$$P(Y = 1) = \frac{e^{\beta_0 + \beta_1 X}}{1 + e^{\beta_0 + \beta_1 X}} \tag{3-20}$$

Also, it can be expressed in Log format:

$$log\left({P(Y = 1)} \Big/ {1 - P(Y = 1)} \right) = \beta_0 + \beta_1 X \tag{3-21}$$

Logistic regression is a model applied to categorical dependent variables, such as lose/win, fail/pass or low/high. In our model, we use logistic regression to judge the forecast value, which comes from the hybrid model, is inaccurate or accurate.

In summary, there are three steps for obtaining the final outputs:

Step 1, using ARIMA to compute linear part of time series, $\hat{L}_i$, and using ANNs to compute nonlinear part, $\hat{N}_i$;

Step 2, regressing $\hat{L}_i$ and $\hat{N}_i$ on y_i, we can obtain $\hat{y}_i$ by plugging $\hat{L}_i$ and $\hat{N}_i$ in regression model;

Step 3, judging $\hat{y}_i$ obtained from step 2 is 'accurate' or not, using past mean value to replace inaccurate data.

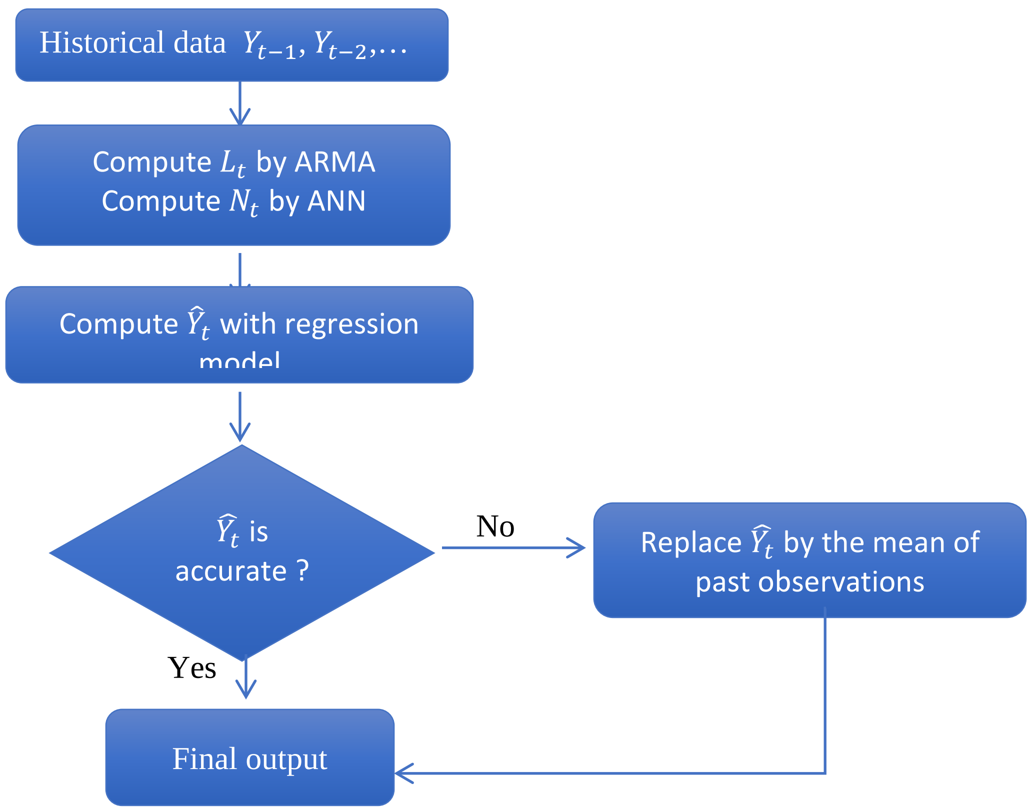

Figure 3.3: Flow chart of Hybrid-Filter model

At time t, we use historical data to enumerate the value of α from 0 to 3 and record the best result with α; this α will be applied in next time. This procedure will repeat before determining the accurate of $\hat{Y}_t$ every time.

For allocating the resources of the hospital, the managers need to obtain the forecast value several days or weeks in advance, to make the results meaningful, we will forecast from every single day to every 10 days (2 weeks). As the properties of ANNs, every time we will obtain a different result, therefore, whenever we use ANNs to forecast, we will repeat 10 times and use the mean value of these 10 series data to compare with other methods.

In this thesis, we use Matlab R2014a to construct the Hybrid-Filter model.

3.4 Evaluation method

There are lots of evaluation methods for measuring the errors. However, not all of them are suitable for this paper. The first one is Mean Forecast Error (MFE), it presents as:

$$MFE = \frac{1}{n}\sum_{t=1}^{n} e_t \qquad (3\text{-}22)$$

e_t : the error between actual and forecast value;

In MFE, it shows the forecast bias, when the MFE equals to zero, the performance is best, but the effects of positive and negative errors cancel out.

Mean Absolute Error (MAE) is another measure, the formula is:

$$MAE = \frac{1}{n}\sum_{t=1}^{n} |e_t| \qquad (3\text{-}23)$$

It presents the magnitude of errors, which are between forecasting values and actual values. When the MAE computed from a set of data is smaller, the better the result is. However, MAE is dependent on the scale of observations; we cannot apply it when compare the different scales of data.

$$MSE = \frac{1}{n}\sum_{t=1}^{n} e_t^2 \qquad (3\text{-}24)$$

The Mean Squared Error (MSE) will generate extreme large errors because of the square. The feature of this evaluation method is that the weights of large errors are much more expensive compared with value with small errors.

As all data in this case study are positive and much greater than zero, the MAPE may be preferred for evaluating the accuracy of the forecast (Hyndman & Koehler, 2006). Therefore, in this case study, we use MAPE (mean absolute percentage error) as the measurement of accuracy. The formula is as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} |e_{t/y_t}| \qquad (3\text{-}25)$$

y_t : denotes the actual value.

Compared with other measurements, MAPE is independent of the scale of measurement and extreme forecast errors will not be panelized.

To evaluate the inefficient cost, we construct functions and they are described as follows:

$$C = O_{jk}^{i} + h * u_{jk}^{i} \qquad (3\text{-}26)$$

$$O_{jk}^{i} = \max\{t_{jk}^{i} - x_{jk}^{i}, 0\} \quad (3\text{-}27)$$

$$u_{jk}^{i} = \max\{x_{jk}^{i} - t_{jk}^{i}, 0\} \quad (3\text{-}28)$$

C : inefficient cost

O_{jk}^{i} : over-time of surgeon *k* in *i*-th day in week *j*

u_{jk}^{i} : idle time of surgeon *k* in *i*-th day in week *j*

t_{jk}^{i} : actual workload of surgeon *k* in *i*-th day in week *j*

x_{jk}^{i} : assigned workload of surgeon *k* in *i*-th day in week *j*

h : cost ratio over-time cost / idle time cost.

We will calculate the inefficient cost of several surgical groups with forecast workload, and compare the inefficient generated by different methods.

In next section, we will present empirical experiments with the results.

Chapter 4 Case Study

In this section, we carry out two case studies to provide solid results to support our Hybrid-Filter model. The first case study is to prove the superiority of our Hybrid-Filter model; we compare the results with ARIMA and ANNs. Also, a new time series data are generated randomly follow the features of original time series. The objective is to prove that our model can be applied in different condition.

The second case study, we defined 'inefficient cost' to evaluate the performances in the application. Using three forecast methods, we generate three inefficient cost. The results are presented to support the point that our Hybrid-Filter model is efficient in this application problem.

4.1 Case study on forecasting accuracy

4.1.1 Data generation

We have collected 365-day historical OR data from UK Healthcare. Extracting the data on weekends and holidays, data of 245 days are left. Given,

i : index for cases i=1,…,I;

j : index for ORs j=1,…,J;

d : index for days d=1,2,3,4,5;

w : index for weeks w=1,…,W;

p_{wd}^{ij} : processing time of case i in OR j of d-th day w-th week, in minute;

y_{wd} : forecast total time of cases of w-th week, in minute.

The duration of a case is defined as:

$$p_{wd}^{ij} = p_{wdl}^{ij} - p_{wde}^{ij}, \tag{4-1}$$

where p_{wdl}^{ij} denotes the time when the surgery is completed, p_{wde}^{ij} denotes the time when the surgery is start.

4.1.2 Normality tests

As mentioned before, we integrate the concept of Statistical Process Control Charts to improve our forecast accuracy further. Because of the concept of Statistical Process Control Charts, if a series of data generated by stationary process that follow a normal distribution, we can assume that about 68% of future data to fall within ± 1 standard deviation around the mean value, 95% of future observations which are between ± 2 standard deviation around the mean value, and approximate 99% of future data fall within ± 3 standard deviation around the mean value. Under these principles, we can filter extreme values, which is more likely to generate large errors between forecast value and actual value. Thus the assumption of normal distribution is necessary when we construct a quality control charts.

From the original data of UK healthcare, the total case time per day is 1.5107e4 ± 1.5784e3 min. To test the normality of original time series, test and Kolmogorov-Smirnov test are carried out to test the normality using SPSS 19.0.

Jarque-Bera test (JB-test) is applied to determine whether the sample data present the same skewness and kurtosis as the normal distribution (Jarque & Bera, 1987). The statistic of JB test is described as following:

$$JB = \frac{n-k+1}{6}(S^2 + \frac{1}{4}(C-3)^2) \quad (4\text{-}2)$$

Where n denotes the number of observations; S is the skewness of sample data; C is the value of kurtosis of sample data, and k is the number of regressors.

In Kolmogorov–Smirnov (KS) test, the probability distributions of one-dimensional sample data are used to compare a sample with a reference probability distribution. The following tables present the statistic of the sample data.

Table 4.1: Statistics of total case time per day

		Statistic	Std. Error
Mean		15107.03	100.84
95% Confidence Interval for Mean	Lower Bound	14908.39	
	Upper Bound	15305.65	
5% Trimmed Mean		15105.05	
Median		15133.00	
Variance		2491350.38	
Std. Deviation		1578.40	
Minimum		10507.00	
Maximum		19806.00	
Range		9299.00	
Interquartile Range		2081.00	
Skewness		-.028	.156
Kurtosis		.128	.310

In statistics, we use skewness to measure the asymmetry of the probability distribution about its mean. The value of skewness can be positive and negative. The skewness can be computed as:

$$\gamma_1 = E\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mu_3}{\sigma^3} = \frac{E[(X-\mu)^3]}{\left(E[(X-\mu)^2]\right)^{3/2}} \tag{4-3}$$

Where X is a random variable from sample data, μ is the mean value, σ is the standard deviation, E is the expectation operator, μ_3 denotes the third central moment. The *n-th* central moment is defined as:

$$\mu_n = \mathrm{E}[(X - \mathrm{E}(X))^n] = \int_{-\infty}^{+\infty} (x-\mu)^n f(x)dx \tag{4-4}$$

Similarly, kurtosis is a measurement of the 'tailedness' of a distribution. In another word, Kurtosis describes the shape of a probability distribution. The formula to calculate kurtosis is as following:

$$Kurt[X] = \frac{\mu_4}{\sigma^4} = \frac{E[(X-\mu)^4]}{(E[(X-\mu)^2])^2} \tag{4-5}$$

Generally, the value of skewness and Kurtosis between -2 and +2 is considered acceptable in order to prove the normality of sample data using SPSS. Obviously, -0.028 and 0.128 are within the range.

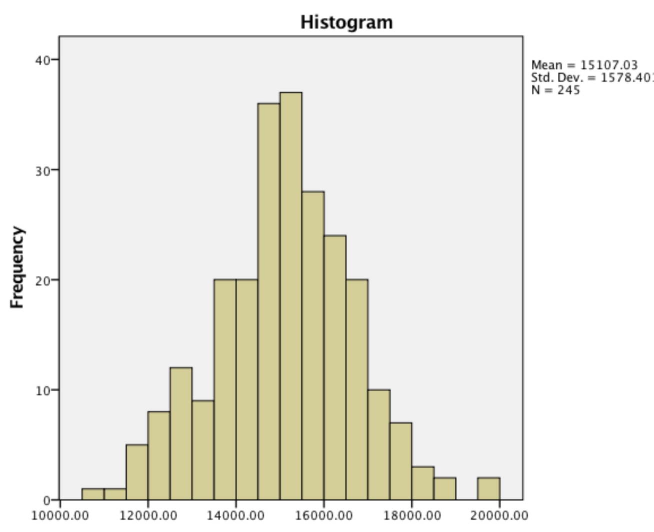

Figure 4.1: Histogram of total case time per day

From histogram in Figure 4.1, the total case time per day is approximately symmetrical around the mean value. Figure 4.1 also supports the result of skewness in Table 4.1. Thus we consider it as an approximate normal distribution time series.

Table 4.2: Percentiles of original data

	5	10	25	50	75
Weighted Average	12252.30	12859.40	14088.00	15133.00	16169.00
Tukey's Hinges			14089.00	15133.00	16155.00

	90	95
Weighted Average	17040.60	17708.00
Tukey's Hinges		

Table 4.3: Extreme values of original data

		Case Number	Value
Highest	1	213	19806.00
	2	138	19616.00
	3	163	18897.00
	4	140	18882.00
	5	193	18243.00
Lowest	1	15	10507.00
	2	236	11389.00
	3	23	11745.00
	4	17	11878.00
	5	204	11894.00

Table 4.4: Tests of Normality

Kolmogorov-Smirnov			Shapiro-Wilk		
Statistic	df	Sig.	Statistic	df	Sig.
.044	245	.200	.996	245	.698

From Table 4.4 the sigma value of Kolmogorov-Smirnov test is 0.200 which is larger than 0.05. The result shows there are no significant differences between the normal distribution and the sample data. Thus the decision of Kolmogorov-Smirnov test is retaining the null hypothesis at 5% level. Similarly, the sigma value of Shapiro-Wilk test is also larger than 0.05, and the null hypothesis, which the sample data is a normal distribution, is maintained.

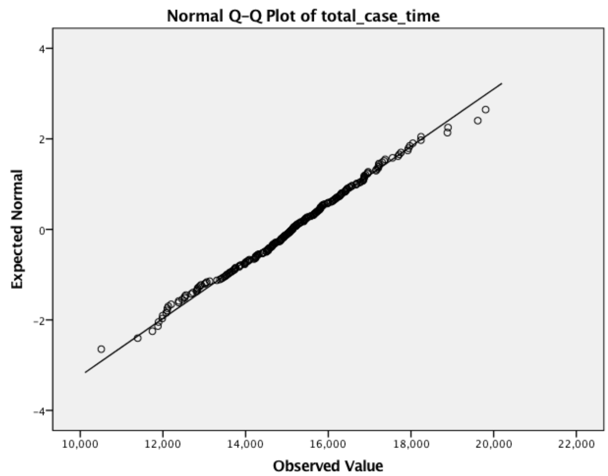

Figure 4.2: Normal Q-Q Plot

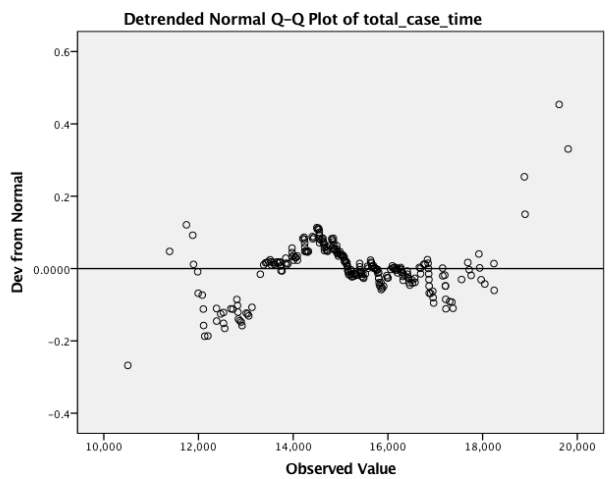

Figure 4.3: Detrended Normal Q-Q Plot

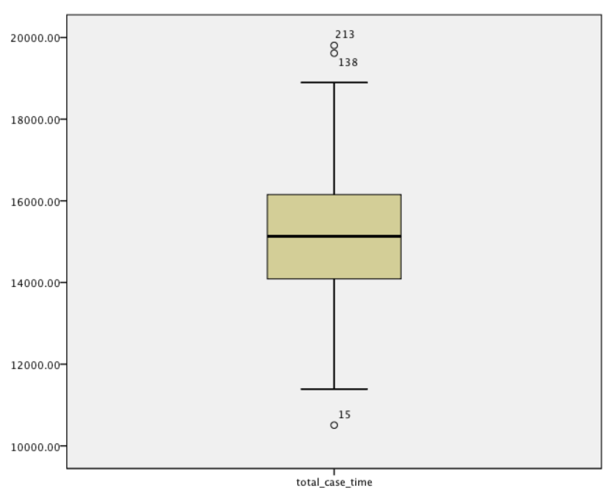

Figure 4.4: Box plot of total case time per day

From Figure 4.2, we can find the points presented the normal Q-Q plot distribute around the diagonal. The Q-Q plots in Figure 4.2 verify the conclusion above.

4.1.3 Results of case study one

For examining the applicability and generality of the Hybrid-Filter, we use the above features of the original data to generate a new random time series. There are 300 observations in the new set of time series, and it follows normal distribution with the mean value of 1.5084e4 and standard deviation of 1.6002e3.

In the first case study, we are going to investigate how the hybrid model performs among three models (i.e., ARIMA, ANNs & Hybrid filter). We compare the results generated by ARIMA, ANNs and Hybrid filter. As we mentioned in the previous chapter, the errors between forecast value and actual value are evaluated by MAPE. The MAPE can be described as follows:

$$MAPE = \frac{1}{n}\sum_{t=1}^{n} |e_{t/y_t}| \qquad (4\text{-}6)$$

e_t: the error between actual and forecast value;

y_t: the actual value.

As the ANNs model always output different values, we repeat the ANN model 10 times to obtain 10 series of data, and then compute the average value of these data. Thus the final outputs of ANNs are the mean value of 10 series of data. The original time series is denoted by S1, and the new randomly generated time series is denoted by S2.

Table 4.5 presents the MAPE when the three methods are applied on one-step ahead forecasting (i.e., forecast one day in advance). Although we apply our Hybrid-

Filter model on a different time series with similar features, the performance of our model is still better than the other two methods.

Table 4.5: Comparing the MAPE of different methods & time series

	ARIMA	ANNs	Hybrid-Filter
MAPE of S1	11.52%	10.64%	8.51%
MAPE of S2	12.50%	10.99%	9.78%

Table 4.6: Comparing the MAPE of S2

	ARIMA	ANNs	Hybrid-Filter
MAPE	12.50%	10.99%	9.78%
MAPE≤5%	23.93%	28.93%	30.57%
MAPE≤10%	48.57%	57.86%	60.00%
MAPE≤15%	67.97%	79.62%	80.00%
MAPE≤20%	81.85%	85.35%	89.43%
MAPE>20%	18.15%	14.65%	10.67%

From Table 4.6, the MAPE generated by ARIMA is 12.50%, 10.99% by ANNs, and 9.78% by Hybrid-Filter model. Comparing with ARIMA, the error generated by ANNs reduces 12.08%, and the error of the Hybrid-Filter model reduces 21.76%.

In another view, the Hybrid-Filter model obtains more accurate data than ARIMA and ANNs, the percentages of MAPE is less than 5% are 23.93%, 28.93%, 30.57%, the percentages of MAPE which is between 5% and 10% are 24.64%, 28.93%, 29.43%. Also, when we focus on the extreme value, our model avoids extreme large bias from the actual value as well, because of the percentages of MAPE larger than 20% are only 10.67%; the percentage reduces 41.21% (compared with ARIMA) and 27.17% (compared with ANNs). Obviously, as the Hybrid-Filter model generates more accurate data and less inaccurate data, our model outperforms among these 3 models.

For allocating the resources of the hospital, the managers need to obtain the forecast value several days or weeks in advance. Therefore, we are going to investigate how these three models perform while we forecast the workload from 1 day to 10 days (2 weeks) ahead. The results are measured by MAPE.

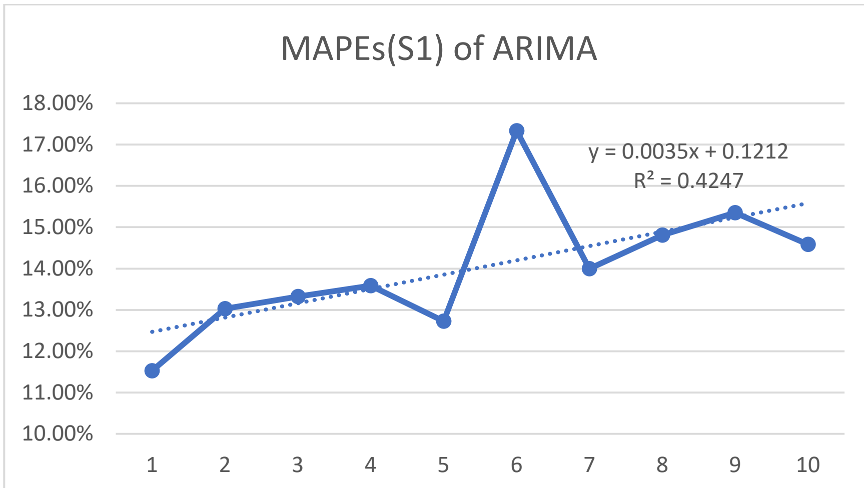

Figure 4.5: MAPEs (S1) of ARIMA with different days in advance

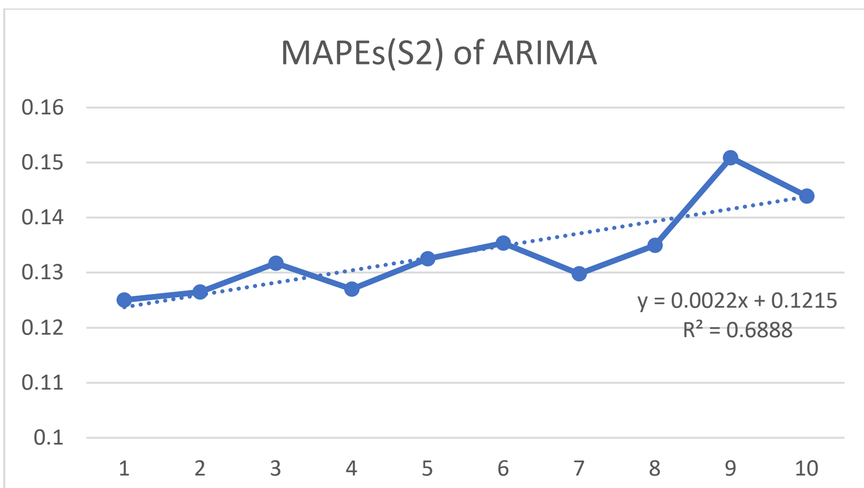

Figure 4.6: MAPEs (S2) of ARIMA with different days in advance

From Figure 4.5 and 4.6, when we apply ARIMA model, the error will increase while the forecast period increase. Both of the above two linear regression models show a strong increasing trend with a positive coefficient. That means when we want to make a long-term forecast with ARIMA model, the error will go very high so that ARIMA model is inapplicability in multi-step forecasting.

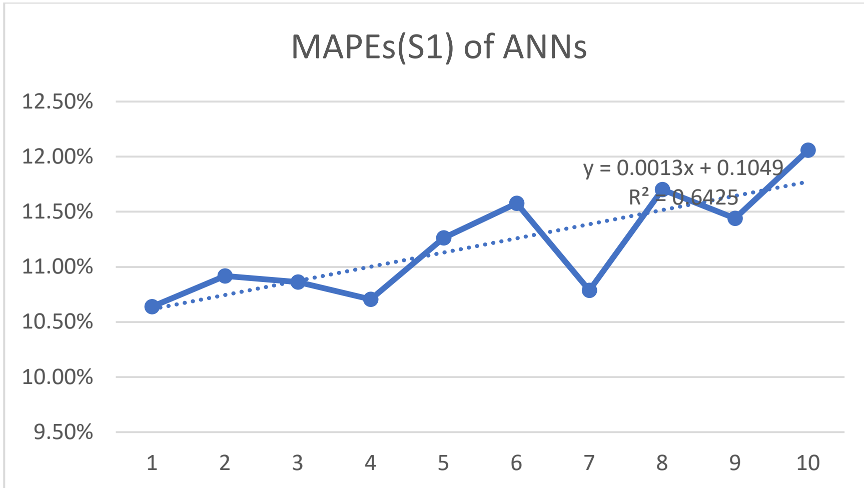

Figure 4.7: MAPEs (S1) of ANNs with different days in advance

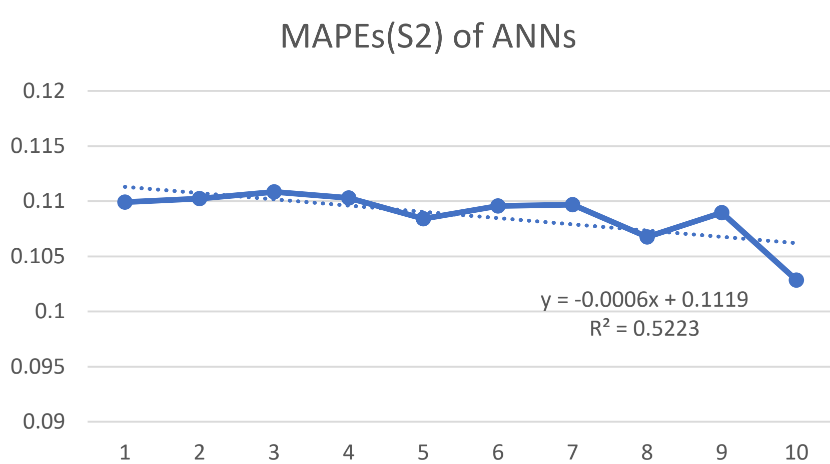

Figure 4.8: MAPEs (S2) of ANNs with different days in advance

Figure 4.7 and 4.8 present opposite trend of the MAPEs generated in different cases. However, the coefficients are relatively small compared with ARIMA. To some extent, we consider the MAPEs of ANNs are stochastic and without any linear trend.

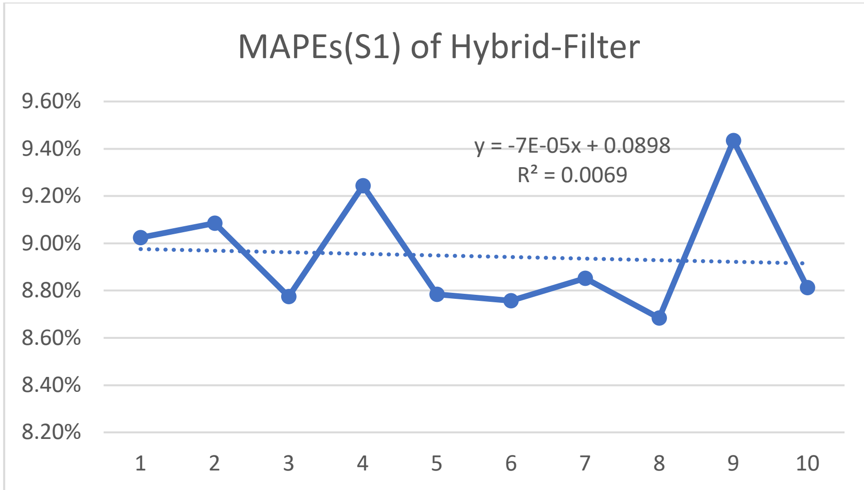

Figure 4.9: MAPEs (S1) of Hybrid-Filter with different days in advance

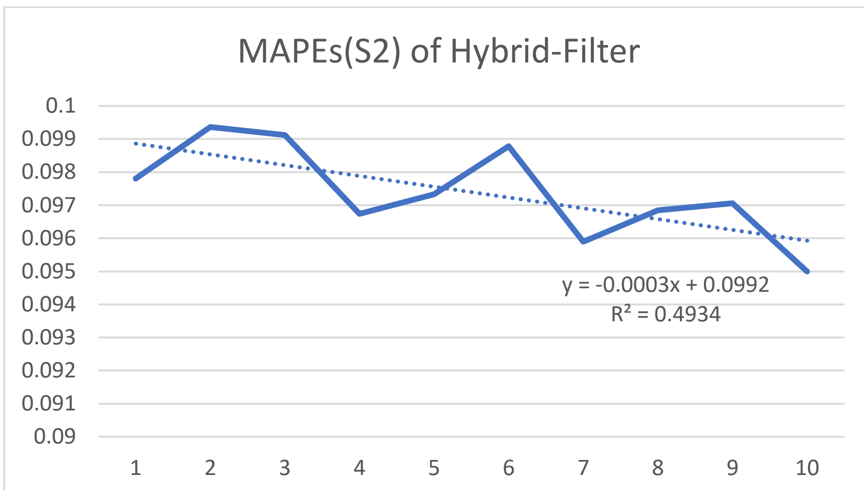

Figure 4.10: MAPEs (S2) of Hybrid-Filter with different days in advance

From figures 4.8 and 4.9, although the trend of these two models is declining, the absolute value of coefficient is quite small compared with ARIMA. Generally, the MAPEs keep consistent when the number of days in advance increase. Thus, compared with ARIMA, Hybrid-Filter and ANNs are better when applied to multi-days in advance forecasting.

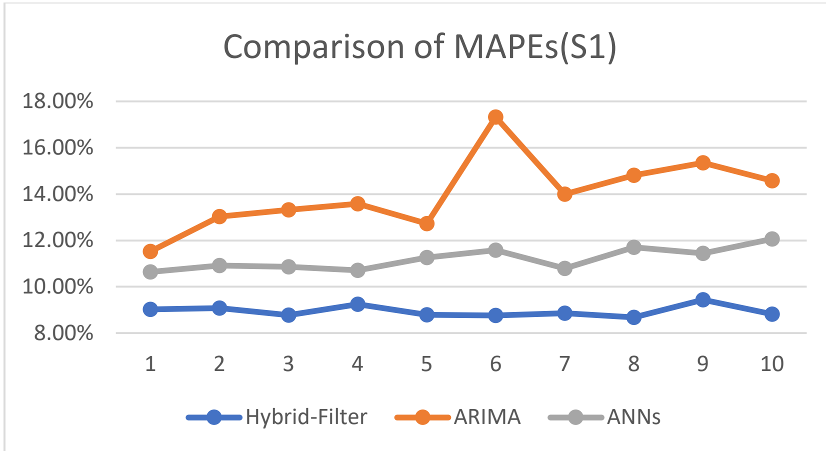

Figure 4.11: MAPEs (S1) of different methods

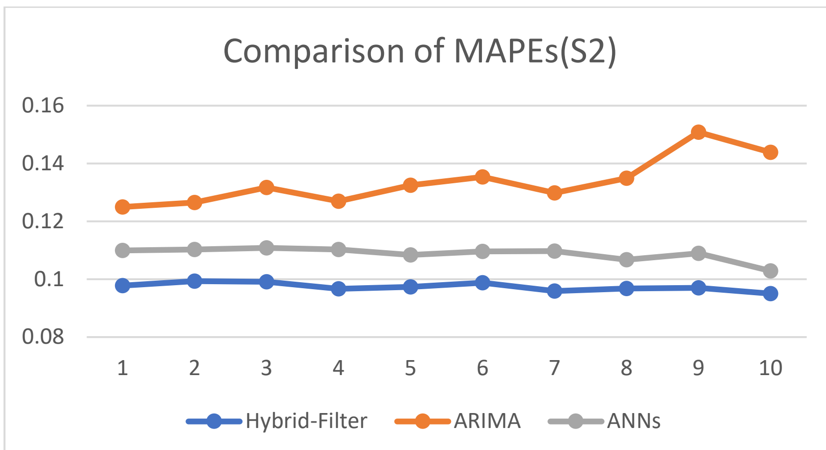

Figure 4.12: MAPEs (S2) of different methods

Table 4.7: Comparison of MAPEs (S1) with different days ahead

Days in advance	ARIMA	ANNs	Hybrid-Filter
1	11.53%	10.64%	9.02%
2	13.03%	10.92%	9.09%
3	13.32%	10.86%	8.77%
4	13.59%	10.70%	9.24%
5	12.73%	11.26%	8.78%
6	17.33%	11.58%	8.76%
7	13.99%	10.79%	8.85%
8	14.81%	11.70%	8.68%
9	15.35%	11.44%	9.44%
10	14.58%	12.06%	8.81%

Table 4.8: Comparison of MAPEs (S2) with different days ahead

Days in advance	ARIMA	ANNs	Hybrid-Filter
1	12.50%	10.99%	9.78%
2	12.65%	11.02%	9.94%
3	13.17%	11.08%	9.91%
4	12.70%	11.03%	9.67%
5	13.25%	10.84%	9.73%
6	13.54%	10.96%	9.88%
7	12.98%	10.97%	9.59%
8	13.50%	10.68%	9.68%
9	15.09%	10.90%	9.71%
10	14.39%	10.29%	9.50%

From Figure 4.11 and 4.12, it is obviously that our model always performs best among these three methods. The differences between ARIMA and Hybrid-Filter model become larger and larger. Predictably our model will maintain the superiority. Table 4.7 and Table 4.8 present the MAPEs of three models in different forecast period. As the forecast period increase, the accuracy of ANNs and Filter keeps steady while the MAPE of ARIMA increase very high.

Table 4.9: P values of one tail t-test (S1)

Days in advance	ARIMA_Hybrid-Filter	ANNs_Hybrid-Filter
1	0.0146	0.036586
2	0.0017	0.037991
3	0.0001	0.014469
4	0.0010	0.074817
5	0.0018	0.022846
6	0.0000	0.003817
7	0.0001	0.029218
8	0.0000	0.004413
9	0.0000	0.035826
10	0.0000	0.004236

Table 4.10: P values of one tail t test (S2)

Days in advance	ARIMA_Hybrid-Filter	ANNs_Hybrid-Filter
1	0.0448	0.2602
2	0.0004	0.0729
3	0.0000	0.0609
4	0.0002	0.0333
5	0.0002	0.0686
6	0.0002	0.0712
7	0.0002	0.0324
8	0.0001	0.0814
9	0.0000	0.0531
10	0.0000	0.1239

Table 4.9 and Table 4.10 present the P values of one tail t-test. The null hypotheses are the MAPEs generated by ARIMA are smaller than Hybrid-Filter model, and the MAPEs generated by ANNs are smaller than Hybrid-Filter model. As is shown in the table, in time series S1, the MAPEs of ARIMA are always significantly larger than MAPEs of Hybrid model at 95% confidence level. In time series S2, except one-step ahead forecasting, the MAPEs of ARIMA are significantly larger than MAPEs of Hybrid model at 95% confidence level as well. When compared with ANNs, from Table 4.9, we found that in half cases, our model improves the performance significantly at 95% confidence level; 9 of 10 cases improve the accuracy significantly at 90% confidence level. But in time series S2, only 2 cases improve the forecasting accuracy at 90% confidence level. As the time series S2 is generated randomly, some underlying relationships from original time series S1 are broken. Despite the fact that we obtained worse performances in time series S2 compared with S1, the results did not present negative effect on our model.

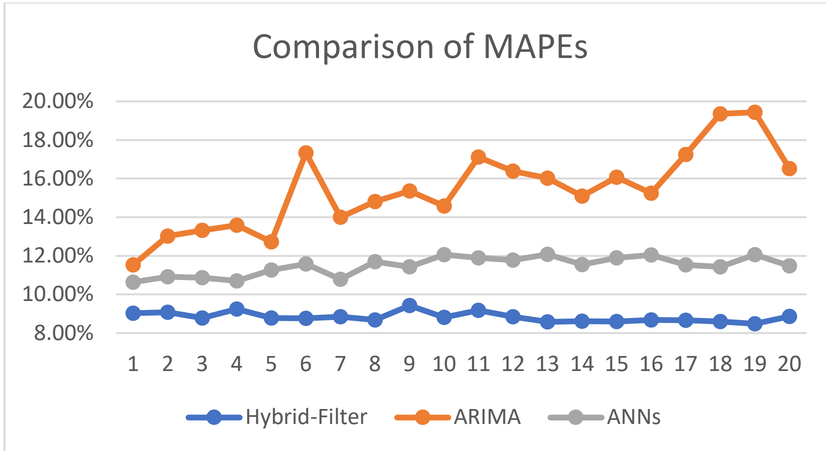

Figure 4.13: Extension to 20 steps of MAPEs of different methods

To investigate if the MAPEs will converge, we extend the steps from 10 to 20. From Figure 4.13, we found that the MAPEs of ARIMA are around 19% when the steps extend to 20, and they still have an increasing trend. The MAPEs of ANNs are around 12%, compared with the result of one step, the errors also increase about 13%. The errors of Hybrid-Filter model are more stable compared with the other two models, because the difference of MAPEs between 1 step and 20 steps is only 1.77%.

4.2 Case study on application to Operating room

In this case study, we are going to reduce the over-utilized cost and idle time cost of OR with forecasting the workload of a surgical group. There were 189 surgeons in operations during 6.3.2013 to 5.30.2014 in UK healthcare. To simplify the problem, we choose surgeons ID# 2744, 1822 and 677 who operated the most surgeries among 189 surgeons.

In practice, there are two methods for surgery scheduling, which are block scheduling and open scheduling (Cardoen, Demeulemeester, & Belien, 2008). In the case of block scheduling system, each surgeon allocates the block time according to a periodic schedule, which is weekly or monthly, until the next period (multiple weeks or months).

In this condition, while the surgery is determined to be completed within the block time, the surgeons will schedule the case in their blocks. In the case of an open scheduling system, the surgeons request block times to OR for each case, and add the scheduled case time to an available OR on a first-come, first-serve basis. In this thesis, we follow the open scheduling method in this case study.

4.2.1 Evaluation method

To evaluate the results, we define an ‘inefficient cost'; the cost can be described as following:

$$\mathrm{C} = h * O_{wk}^{d} + u_{wk}^{d} \tag{4-6}$$

$$O_{jk}^{i} = \max\{t_{wk}^{i} - x_{wk}^{i}, 0\} \tag{4-7}$$

$$u_{jk}^{i} = \max\{x_{wk}^{i} - t_{wk}^{i}, 0\} \tag{4-8}$$

Where, O_{wk}^{d} denotes over-time of surgeon *k* in *d*-th day in week *w*, u_{wk}^{d} denotes over-time of surgeon *k* in *d*-th day in week *w*, t_{wk}^{d} is the actual workload of surgeon *k* in *d*-th day in week *w* and x_{wk}^{d} is the assigned workload of surgeon *k* in *d*-th day in week *w*, *h* is the cost ratio of over-time over idle time. In most cases, *h* is around 1.75 (Hosseini & Taaffe, 2015).

4.2.2 Results of case study

We forecast the workload of each surgeon and calculate the inefficient cost with an actual workload, and then compare the ‘inefficient' cost obtained by different forecast methods.

As surgeons only worked three or two days per week, so that after we exclude holidays and weekends, totally 273 observations are available (40 weeks). Among 273 observations, the first 15 data are used as historical data for training models. By the end of the week, day *T*, we forecast the surgeon's workload in next week (workload on *T*+1, *T*+2, *T*+3), and then the surgeons apply the operating room with an estimating value of their workload.

Table 4.11: MAPE of different methods

Forecasting method	MAPE
ARIMA	30.49%
ANNs	24.65%
Hybrid-Filter	21.52%

At first, we compare the error of these three methods. From Table 4.11, we found that the ARIMA generated the largest error, 30.49%. Comparing with ARIMA, the MAPE of Hybrid-Filter decreases 29.42% and comparing with ANNs; it decreases 12.70%. Obviously, our model outperforms among these methods and improves the performance significantly.

As mentioned before, an optional filter is logistic regression model. To examine which filter is better in application problem, we integrate logistic regression model into the hybrid model instead of the control charts. In this case study, we define the MAPE is larger than 10% as 'inaccurate' output; oppositely, the MAPE which is less than 10% is 'accurate' output.

The first method is using logistic model to replace the original 'filter structure', named Hybrid-Logistic model. Another is combining filter and a logistic regression model, called Hybrid-Filter & Logistic model. In Hybrid-Filter & Logistic model, we use original filter structure to pick the value, which exceeds lower bound or upper bound first and then looking for ‘inaccurate' with a logistic regression model. The third method is Hybrid-Filter model, which is proposed in the previous section.

Table 4.12: MAPE of different filter

Forecast method	MAPE
Hybrid-Logistic	22.32%
Hybrid-Filter & Logistic	25.17%
Hybrid-Filter	21.52%

From Table 4.10, the MAPE generated by Hybrid-Logistic model is 22.32%, improved 9.45% compared with ANNs and 26.8% compared with ARIMA. The MAPE of Hybrid-Filter & Logistic is 25.17%, even worse than ANNs because the function of filter and logistic is overlapping. As the performances of Hybrid-Logistic and Hybrid-Filter are close, only 3.58% difference between these two methods, it is difficult to tell which one is better. However, in an application problem, we can compare the performance of each method and choose a better one.

Table 4.13: Inefficient cost of different methods (h is from 1 to 2)

h	ARIMA	ANNs	Hybrid-Filter
1.00	19518.29	15993.93	13162.89
1.25	22061.79	17878.81	14876.96
1.50	24605.28	19763.69	16591.04
1.75	27148.77	21648.56	18305.12
2.00	29692.27	23533.44	20019.2

Table 4.14: Comparing results with ARIMA and ANNs (h is from 1 to 2)

h	(ARIMA-Hybrid)/ARIMA%	(ANNs-Hybrid)/ANNs%
1.00	32.56%	17.70%
1.25	32.57%	16.79%
1.50	32.57%	16.05%
1.75	32.57%	15.44%
2.00	32.58%	14.93%

As the value of h affects the inefficient cost, we compare the results with different h. Table 4.13 and 4.14 present the results of inefficient cost generated by three forecast models. From the above tables, the Hybrid-Filter model always performs best with the lowest inefficient cost. When the h increases, the differences between ANNs and hybrid model decrease. In Table 14, we compute the percentage differences between each two methods; the results show that the over-time cost generated by our model is larger than ANNs', or, the ANNs generate more idle time cost compared with the Hybrid-Filter mode. This information can be used to further improve our model.

In this case study, the Hybrid-Filter model takes advantages of each single forecasting method, ARIMA and ANNs. The ability of modeling linear and nonlinear time series contribute to obtaining the better performances. Besides, a filter structure is integrated into the model to further improve the performance. The results prove the effectiveness of the Hybrid-Filter model. Thus, when we apply our Hybrid-Filter model to multi-steps-ahead forecasting in complicated time series, our model outperforms other models in practice.

Chapter 5 Conclusion and future works

5.1 Conclusion and results

Modeling and predicting data with time series are essential in various areas of practice, for example, the product demand in mass-manufacture for resource allocation, the short-term electricity demand for the electricity supply system, the rainfall and river flow for flood prevention, etc. As a result, research on this topic has been active for decades, and many classical models proposed to obtain better performance and validity of predicting data with time series, such as Autoregressive Integrated Moving Average (ARIMA), artificial neural network (ANN), etc.

Allocating necessary resources to operating rooms (OR) based on a high accuracy of demand forecasting is important to OR management, which provides sufficient OR block times for surgeons to complete their cases and reduces unnecessary waiting times for patients. However, data with time series have different features, such as linearity and non-linearity, which are difficult to capture or describe by using a single or simple modeling and predicting method (G P Zhang, 2003). It does not exist a general model can be applied to every particular problem.

Besides, because most modeling methods focus on one-step-ahead forecasting, the performance in mid-term forecasting (i.e., weekly and biweekly forecasting) is quite bad. However, in the OR management system, each surgeon assigned blocks of time in a cyclic schedule (weekly or monthly) until the next review period (Cardoen et al., 2008). Therefore, for allocating the resources of the hospital, the managers want to obtain the demand several days or weeks in advance. A model, which can be applied to mid-term or even long-term forecasting (weeks in advance) with high accuracy, is critical under this condition.

As mentioned before, to describe and model the underlying relationship of time series, many classical models are proposed, one basic models is ARIMA model. Compared with moving average (MA) and autoregressive (AR) model, ARIMA is more flexible to model the feature of the time series. The main drawback of ARIMA is that the model always treats the time series as linear. In other words, a linear relationship is assumed among the time series. Thus nonlinear patterns cannot be captured and modeled by the ARIMA model. Obviously, there does not exist pure linear time series in the real world.

To solve the limitation of ARIMA model, researchers construct several models to capture the nonlinearity in time series. For instance, the autoregressive conditional heteroscedastic (ARCH) model (Engle, 1982), and the threshold autoregressive (TAR) model (Tong, 1983). Although these models improve the ability to describe nonlinear time series, we have limited development in nonlinear modeling (De Gooijer & Kumar, 1992). These methods are constructed for specific time series, and they are not efficient when applied to other types of time series. Then ANNs is proposed to overcome these shortcomings, compared with ARCH and TAR, the most significant feature of ANNs is that ANNs are universally applied model. This is because the ANNs is flexible to model various time series. The flexibility of ANNs comes from the parallel processing of the information from the data (G P Zhang, 2003). In addition, there is no precondition when applying ANNs to time series modeling.

Many empirical studies indicate that the accuracy of forecasting can be improved by combining several different models (Makridakis et al., 1993; Makridakis & Hibon, 2000). Therefore, combining different models is a feasible choice of which increase the chances of modeling different properties in time series and improve the predictive accuracy.

The advantage of combining similar models together is to obtain a stable model with high robustness and enhance the strong point of the similar models (Yu et al., 2005). Since we mentioned in the previous paragraph, to deal with the different features in time series, we build a hybrid model, which combines the ARIMA and ANNs, to take advantages of each single model. The hybrid model is as following:

$$y_t = \beta_1 L_t + \beta_2 N_t + \beta_0 \tag{5.1}$$

Here, y_t is actual value at period t, L_t denotes the linear part, and N_t denotes nonlinear part, β_1 and β_2 are coefficients, β_0 is the error. L_t is obtained from ARIMA, and N_t is computed from ANNs.

To further improve the forecast accuracy when applied to mid-term forecasting, we lead in the concept of Statistical Process Control Chart. Using the past observations, we can calculate the upper bound and lower bound with the past mean value and standard deviation. The extremely large or small value, which exceeds the upper bound and lower bound obtained from the previous procedure, is considered as inaccurate value. We call this structure 'filter', the function of which is to find the potential inaccurate data, and then replacing them with the mean value of several past observations.

We also provide an optional filter structure, which is logistic regression model. Logistic regression is a model applied to categorical dependent variables, such as lose/win, fail/pass or low/high. Here we use logistic regression to judge the forecast value is inaccurate or accurate. While applying a logistic regression model, a threshold value is set up to distinguish inaccurate data from forecast value, and then the average of past observations is used to replace the inaccurate value.

The filter structure is integrated into our hybrid model to improve the accuracy in mid-term prediction. It is beneficial to obtain better results, because the filter structure helps to exclude extreme large and small value, which is more likely to generate relative high errors.

We carry out two case studies to examine the priority and applicability of our Hybrid-Filter model proposed in previous chapter. In the first case study, to evaluate the error, mean absolute percentage error (MAPE) is used because of its properties, such as scale independent and non-penalty with extreme values. A new time series is generated randomly following the features of original time series, the results from two series data will be compared to prove that our Hybrid-Filter model can be used in different condition with high accuracy.

Three methods are proposed in this case study, from the results of Table 5.1 and 5.2, compared with ARIMA, the MAPE is reduced by 35.49% on average with different days in advance in Table 5.1 and 26.94% in Table 5.2. When compared with ANNs, the MAPE is reduced by 19.95% and 10.43% in different cases. Notice that the results in Table 5.1 are better than Table 5.2. This is because some underlying relationships of time series 1are captured by our models, but as time series 2 is generated randomly, the underlying relationships are broken. Therefore, we achieve worse results with less efficient information in Table 5.2.

Table 5.1: Comparison of MAPEs (S1) with different days ahead

Days in advance	ARIMA	ANNs	Hybrid-Filter
1	11.53%	10.64%	9.02%
2	13.03%	10.92%	9.09%
3	13.32%	10.86%	8.77%
4	13.59%	10.70%	9.24%
5	12.73%	11.26%	8.78%
6	17.33%	11.58%	8.76%
7	13.99%	10.79%	8.85%
8	14.81%	11.70%	8.68%
9	15.35%	11.44%	9.44%
10	14.58%	12.06%	8.81%

Table 5.2: Comparison of MAPEs (S2) with different days ahead

Days ahead	ARIMA	ANNs	Hybrid-Filter
1	12.50%	10.99%	9.78%
2	12.65%	11.02%	9.94%
3	13.17%	11.08%	9.91%
4	12.70%	11.03%	9.67%
5	13.25%	10.84%	9.73%
6	13.54%	10.96%	9.88%
7	12.98%	10.97%	9.59%
8	13.50%	10.68%	9.68%
9	15.09%	10.90%	9.71%
10	14.39%	10.29%	9.50%

The results present in Table 5.1 and 5.2 prove the following points: (1) in different cases, our Hybrid-Filter model always performs best with the smallest MAPEs among three methods; (2) different from ARIMA, when the number of days in advance increasing, the MAPEs of our model keep steady instead of increasing.

In the second case study, we extend to an application problem. The objective is to reduce the inefficient cost, which is generated by the error between forecasting workload and actual workload. The inefficient cost is defined as following:

$$\mathrm{C} = h * o_{jk}^{i} + u_{jk}^{i} \tag{5.2}$$

o_{jk}^{i} is the overtime generated by surgeon k in i-th day in week j, u_{jk}^{i} is the idle time generated by surgeon k in i-th day in week j, h denotes the cost ratio of overtime and idle time.

Table 5.3: Inefficient cost of different methods

h	ARIMA	ANNs	Hybrid-Filter
1.00	19518.29	15993.93	13162.89

From the results in Table 5.3, compared with ARIMA, our Hybrid-Filter model reduces the inefficient cost by 32.56% and the inefficient cost reduce 17.70% of the cost generated by ANNs. This case study provides sufficient results to convince us that when using our model to solve an application problem, it is still valid and better than other proposed methods.

5.2 Limitations and future work

Despite we achieve the best performance among the three methods using our Hybrid-Filter model, there are still some limitations in our case study. At first, too many factors have effects on the final outputs, such as the number of past observations used in ARIMA, ANNs and filter structure. It is difficult for us to obtain an optimal combination of these numbers, and then applied the combination to our model to generate optimal results.

What's more, since our model is based on the combination of ARIMA and ANNs, thus when we take the advantages of these two models, the disadvantages of these models are also augmented. The disadvantages of ARIMA and ANNs have negative effects on the final forecast outputs as well. For example, the computation complexity of ANNs makes our model take too much time to compute.

In the following research, we are going to investigate how to improve the forecast efficiency. In other words, we expect higher accuracy with less computation complexity and less historical data. Also we are searching a method or principle to determine the numbers of historical data used in time series model. Thus we can achieve optimal results as long as following the principle. Besides, we expect to integrate more explanatory variables to describe the workload.

Reference:

Abraham, A., & Nath, B. (2001). A neuro-fuzzy approach for modelling electricity demand in Victoria. *Applied Soft Computing*, *1*(2), 127–138. https://doi.org/10.1016/S1568-4946(01)00013-8

Cao, L. J., & Tay, F. H. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting. *IEEE Transactions on Neural Networks / a Publication of the IEEE Neural Networks Council*, *14*(6), 1506–18. https://doi.org/10.1109/TNN.2003.820556

Cardoen, B., Demeulemeester, E., & Belien, J. (2008). Operating room planning and scheduling: A literature review. *Journal of the Operational Research Society*, *201*(3), 2–44. https://doi.org/10.1016/j.ejor.2009.04.011

Cheng, C., Sa-Ngasoongsong, A., Beyca, O., Le, T., Yang, H., Kong, Z. (James), & Bukkapatnam, S. T. S. (2015). Time series forecasting for nonlinear and non-stationary processes: a review and comparative study. *IIE Transactions*, *47*(10), 1053–1071. https://doi.org/10.1080/0740817X.2014.999180

Darbellay, G. A., & Slama, M. (2000). Forecasting the short-term demand for electricity: Do neural networks stand a better chance? *International Journal of Forecasting*, *16*(1), 71–83. https://doi.org/10.1016/S0169-2070(99)00045-X

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 Years of Time Series Forecasting. *International Journal of Forecasting*, *22*(3), 443–473. https://doi.org/10.1016/j.ijforecast.2006.01.001

De Gooijer, J. G., & Kumar, K. (1992). Some recent developments in non-linear time series modelling, testing, and forecasting. *International Journal of Forecasting*, *8*(2), 135–156. https://doi.org/10.1016/0169-2070(92)90115-P

de Menezes, L. M., W. Bunn, D., & Taylor, J. W. (2000). Review of guidelines for the use of combined forecasts. *European Journal of Operational Research*, *120*(1), 190–204. https://doi.org/10.1016/S0377-2217(98)00380-4

Dexter, F., Macario, A., Qian, F., & Traub, R. D. (1999). Forecasting surgical groups' total hours of elective cases for allocation of block time. *Anesthesiology*, *91*(5), 1501–8. https://doi.org/10.1097/00000542-199911000-00044

Draper, D. (1995). Assessment and propagation of model uncertainty. *Journal of the Royal Statistical Society: Series B (Methodological)*. https://doi.org/10.2307/2346087

Dredge, R. (2004). Hospital Global Budgeting. Washington, DC, *World Bank*: 1-54.

Engle. (1982). Autoregressive Conditional Heteroscedacity with Estimates of variance of United Kingdom Inflation. Journal of Econometrica. https://doi.org/10.2307/1912773

Farooq, T., Guergachi, A., & Krishnan, S. (2007). Chaotic time series prediction using knowledge based Green's Kernel and least-squares support vector machines. *2007 IEEE International Conference on Systems, Man and Cybernetics*, 373–378. https://doi.org/10.1109/ICSMC.2007.4414023

Fan, J., & Yao, Q. (2013). *Nonlinear time series: nonparametric and parametric methods*. New York: Springer-Verlag New York.

Fildes, R. (1992). The evaluation of extrapolative forecasting methods. *International Journal of Forecasting*, *8*(1), 81–98. https://doi.org/10.1016/0169-2070(92)90009-X

Fügener, A., Schiffels, S., & Kolisch, R. (2015). Overutilization and underutilization of operating rooms-insights from behavioral health care operations management. *Health Care Management Science*. https://doi.org/10.1007/s10729-015-9343-1

Garcia, R. C., Contreras, J., van Akkeren, M., & Garcia, J. B. C. (2005). A GARCH forecasting model to predict day-ahead electricity prices. *IEEE Transactions on Power Systems, 20*(2), 867–874. https://doi.org/10.1109/TPWRS.2005.846044

Guoqiang Zhang, B., & Eddy Patuwo, M. Y. H. (1998). Forecasting with artificial neural networks: The state of the art Guoqiang. *International Journal of Forecasting*, *14*(1), 35–62. https://doi.org/10.1016/S0169-2070(97)00044-7

Hamzaçebi, C. (2008). Improving artificial neural networks' performance in seasonal time series forecasting. *Information Sciences*, *178*(23), 4550–4559. https://doi.org/10.1016/j.ins.2008.07.024

Holt, C. C. (2004). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, *20*(1), 5–10. https://doi.org/10.1016/j.ijforecast.2003.09.015

Hosseini, N., & Taaffe, K. M. (2015). Allocating operating room block time using historical caseload variability. *Health Care Management Science*, *18*(4), 419–430. https://doi.org/10.1007/s10729-014-9269-z

Hyndman, R. J., & Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting*, *22*(4), 679–688. https://doi.org/10.1016/j.ijforecast.2006.03.001

Jarque, C. M., & Bera, A. K. (1987). A Test For normality of Observations and Regression Residuals. *International Statistical Review*, *55*(2), 163–172.

Kihoro, J. M., R. O. Otieno, and C. W. (2004). Seasonal Time Series Forecasting: a

Comparative Study of Arima and Ann Models. *Frican Journal of Science and Technology (AJST) - Science and Engineering Series*, *5*(2), 41–49.

Liu, M., Kuo, W., & Sastri, T. (1995). An Exploratory Study of a neural Network Approach for Reliability Data Analysis. *Quality and Reliability Engineering …*. Retrieved from http://onlinelibrary.wiley.com/doi/10.1002/qre.4680110206/abstract

Macario, A., & Dexter, F. (1999). Estimating the duration of a case when the surgeon has not recently scheduled the procedure at the surgical suite. *Anesthesia and Analgesia*, *89*(5), 1241–1245. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/10553843%5Cnzotero://attachment/10197/

Macario, A., Terry S., V., Brian, D., & Tom, M. (1995). Where Are the Costs in Perioperative Care? Analysis of Hospital Costs and Charges for Inpatient Sugical Care. *Anesthesiology*, *83*(6), 1138–1144.

Makridakis, S. (1989). Why combining works? *International Journal of Forecasting*, *5*(4), 601–603. https://doi.org/10.1016/0169-2070(89)90017-4

Makridakis, S., Chatfield, C., Hibon, M., Lawrence, M., Mills, T., Ord, K., & Simmons, L. F. (1993). The M2-competition: A real-time judgmentally based forecasting study. *International Journal of Forecasting*, *9*(1), 5–22. https://doi.org/10.1016/0169-2070(93)90044-N

Makridakis, S., & Hibon, M. (2000). The M3-Competition: results, conclusions and implications. *International Journal of Forecasting*, *16*(4), 451–476. https://doi.org/10.1016/S0169-2070(00)00057-1

Pai, P.-F., & Lin, C.-S. (2005). A hybrid ARIMA and support vector machines model in stock price forecasting. *Omega*, *33*, 497–505.

https://doi.org/10.1016/j.omega.2004.07.024

Peter, W. (1960). Forecasting Sales by Exponentially Weighted Moving Averages. *Management Scinece*, *6*(3), 324–342.

R.K, Agrawal, R. A. K., Adhikari, R., R.K, & Agrawal, R. A. K. (2013). An Introductory Study on Time Series Modeling and Forecasting Ratnadip Adhikari R. K. Agrawal. *arXiv Preprint arXiv:1302.6613*, *1302.6613*, 1–68.

Raicharoen, T., Lursinsap, C., & Sanguanbhokai, P. (2003). Application of critical support vector machine to time series prediction. *Circuits and Systems, 2003. ISCAS'03. Proceedings of the 2003 International Symposium On. Vol. 5. IEEE, 2003.*, *5*, 741–744. https://doi.org/10.1109/ISCAS.2003.1206419

Russo, S. L., Camargo, M. E., & Fabris, J. P. (2012). Applications of Control Charts Arima for Autocorrelated Data. *Practical Concepts of Quality Control*, 31–53. https://doi.org/10.5772/50990

Stepaniak, P. S., Heij, C., Mannaerts, G. H. H., De Quelerij, M., & De Vries, G. (2009). Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: A multicenter study. *Anesthesia and Analgesia*, *109*(4), 1232–1245. https://doi.org/10.1213/ANE.0b013e3181b5de07

Suykens, J. A. K., & Vandewalle, J. (1999). Least Squares Support Vector Machine Classifiers. *Neural Processing Letters 9*, 293–300. https://doi.org/10.1023/A:1018628609742

Suykens, J., & Vandewalle, J. (2000). Recurrent Least Squares Support Vector Machines. *IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS—I: FUNDAMENTAL*

THEORY AND APPLICATIONS, *47*(7), 1109–1114. https://doi.org/10.1109/81.855471

Taylor, J. W. (2003). Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential Smoothing Short-Term Electricity Demand Forecasting Using Double Seasonal Exponential. *Journal of Operational Research Society*, *54*(0), 799–805. https://doi.org/10.1057/palgrave.jors.2601589

Taylor, J. W., & Buizza, R. (2003). Using weather ensemble predictions in electricity demand forecasting. *International Journal of Forecasting*, *19*(1), 57–70. https://doi.org/10.1016/S0169-2070(01)00123-6

Taylor, J. W., de Menezes, L. M., & McSharry, P. E. (2006). A comparison of univariate methods for forecasting electricity demand up to a day ahead. *International Journal of Forecasting*, *22*(1), 1–16. https://doi.org/10.1016/j.ijforecast.2005.06.006

Tong, H. (2003). *Non-linear time series: a dynamical system approach*. Oxford: Clarendon.

Tseng, F. M., Yu, H. C., & Tzeng, G. H. (2002). Combining neural network model with seasonal time series ARIMA model. *Technological Forecasting and Social Change*, *69*(1), 71–87. https://doi.org/10.1016/S0040-1625(00)00113-X

Wijngaarden, J. D. H. Van, Scholten, G. R. M., & Wijk, K. P. Van. (2012). Strategic analysis for health care organizations : the suitability of the SWOT-anaylsis. *International Journal of Health Planning and Management*, *27*(July 2015), 34–49. https://doi.org/10.1002/hpm

Yu, L., Wang, S., & Lai, K. K. (2005). Time series forecasting with multiple candidate models : Selecting or combining ? *, *18*(1), 1–18.

Zhang, G. P. (2003). Time series forecasting using a hybrid ARIMA and neural network model. *Neurocomputing*, *50*, 159–175. https://doi.org/10.1016/S0925-2312(01)00702-0

Zhang, G. P. (2007). A neural network ensemble method with jittered training data for time series forecasting. *Information Sciences*, *177*(23), 5329–5346. https://doi.org/10.1016/j.ins.2007.06.015

Zou, H., & Yang, Y. (2004). Combining time series models for forecasting, *20*, 69–84. https://doi.org/10.1016/S0169-2070(03)00004-9

Vita

NAME:
Xinwei Pan

EDUCATION:
Bachelor of Engineering 2013
Department of Mechanical and Electronic Engineering
Nanjing University of Aeronautics and Astronautics, Nanjing, China

RESEARCH INTERESTS:

Time series forecast model

Statistical analysis

SKILLS:
Engineering software: Matlab; Solidworks; Creo; ANSYS.
Statistical software: SPSS.