Kwame Nkrumah University of Science and Technology

College of Science

Department of Mathematics

Dynamic Linear State Space Model for Forecasting Peak and Short-Term Electricity Demand using Kalman Filtered Monte Carlo Method

A Thesis submitted to the Department of Mathematics through the National Institute for Mathematical Sciences, Ghana in partial fulfillment of the requirements for the award

of

Master of Philosophy degree
(Scientific Computing and Industrial Modeling)

By
FRANK KOFI OWUSU
August, 2018

Facilitated By:
National Institute for Mathematical Sciences, Ghana

Declaration

I hereby declare that, this thesis is the result of my own original research and that no part of it has been submitted to any institution or organization anywhere for the award of a degree. All inclusions from the work of others have been duly acknowledged.

Frank Kofi Owusu	....	....
Student (PG7283616)	Signature	Date
Certified by:		
Dr. P. Amoako-Yirenkyi		
______________________________	....	....
Member, Supervisory Committee	Signature	Date
Dr. Nana Kena Frempong		
______________________________	....	....
Member, Supervisory Committee	Signature	Date
Dr. Akoto Yaw Omari-Sasu	....	....
Member, Supervisory Committee	Signature	Date
Prof. (Mrs.) Atinuke Olusola Adebanji		
______________________________	....	....
Head of Department	Signature	Date

Abstract

Electricity has become a major part of human life, especially in our part of the world. It is one of the most used energy across the world. Due to the fast changing world, the demand for electricity keeps on increasing from time to time yet there is not any efficient way of storing this energy for future use. So operators are very cautious about the amount to release and also to meet the demand of their consumers. For this reason, load forecasting has become a main integrated section in energy management and production. This research seeks to look at Short-Term Load Forecasting. The objective is to forecast the peak demand and total energy generated or electricity demand. So the Seemingly Unrelated Time Series Equations Model which models the level or state and trend in the system was used for the study. A Markov Chain Monte Carlo (MCMC) method, Gibbs Sampler, together with the Kalman Filter and Kalman Smoother, the Forward Filtering Backward Sampling with Gibbs Sampler Algorithm were applied to the model to simulate for the variances also to predict the peak demand the next day's peak and electricity demand. The running ergodic mean showed the convergence of the MCMC process and hence the posterior means of the variances were estimated. The one-step-ahead forecast showed a Mean Absolute Percentage error (MAPE) of 3.696% error in the peak demand forecast and a 4.235% error in the electricity demand forecast. The forecast for the next day was about 2187MW and 44090MW for the peak and electricity demands respectively.

For further studies, the model can be extended to capture seasonal components.

Contents

List of Tables

List of Figures

Dedication

This thesis is dedicated to my mother, Ms.Monica Donkor.

Acknowledgements

This thesis became a success with the kind of support and help of many individuals. I would like to extend my sincere thanks to all of them. Firstly,I want to thank God Almighty for the wisdom He bestowed upon me, the strength and good health throughout this research period. I owe my deepest gratitude Dr. Nana Kena Frempong. Without his encouragement and support, this study would hardly have been completed. I also express my warmest gratitude to the NIMS-KNUST supervisory team, especially Dr.Peter Amoako Yirenkyi and Dr. Omari Sasu. Completion of this thesis looked unachievable without their immense support and advice. Also very grateful to Dr.Wahab Iddrisu for his contribution to this research. I owe a great debt of gratitude to my senior colleagues Reindolf Borkor, Rhydal Esi Eghan, Francis Bukari and Kofi Prempeh for their wonderful guidance. Many Thanks and appreciation to my colleagues and people who offered a helping hand towards the success of this research.

List of Abbreviations/Acronyms

AR	Autoregressive
STLF	Short Term Load Forecasting
DLM	Dynamic Linear Models
FFBS	Forward Filtering and Backward Sampling
MAPE	Mean Absolute Percentage Error
MAE	Mean Absolute Error
MSE	Mean Square Error
MA	Moving Average
RMSE	Root Mean Square Error
NRMSE	Normalize Root Mean Square Error
ES	Exponential Smoothing
STSM	Structural Time Series Models(Methods)
SUTSE	Seemingly Unrelated Time Series Equations
PACF	Partial Autocorrelation Function
ARMA	Autoregressive Moving Average
ARIMA	Autoregressive Integrated Moving Average
AIC	Akaika’s Information Criterion
BIC	Bayesian Information Criterion
MMPF	Multi-Model Partitioning Filter
PCA	Principal Component Analysis
AI	Artificial Intelligence
NN	Neural Network

GDP	Gross Domestic Product
RBA	Radial Basis Algorithm
FFBS	Forward Filtering Backward Sampling
GRIDCo	Ghana Grid Company Limited
MW	Megawatt
GSR	Global Solar Radiation
FLR	Fuzzy Linear Regression
GRNN	General Regression Neural Network
CCHP	Combined Cooling Heating and Power
DBN	Dynamic Bayesian Network

Chapter 1
Introduction

1.1 Background of Study

A reliable and continuous supply of electrical energy is necessary for the functioning of today's complex societies. There are three major stages in power supply from the power plant to end users of the electricity;namely, generation, transmission and distribution(Dwijayanti and Hagan, 2013).

Due to the increasing consumption and obstruction of various kinds, and the extension of existing electrical transmission networks all across the world, power systems are operating closer and closer to their limits. Therefore the probability of occurrences of overloading, equipment failures and blackout tend to increase. Moreover, another problem in this sector is that the electrical energy cannot be stored. So the electricity is only generated when needed (Tai, 2009).

The most challenging part of this power system is for the operator to supply the electricity to meet the demand of the consumers at all times at a cheap cost as possible. Due to the unstable load demand or fluctuations in the demand each day, the prediction of the electric or load demand for the next few hours,days,weeks,months or even years is very key to the system operator for good and quality planning and scheduling of electricity generation and supply.

Forecasting is the backbone of any planning process in all fields of interest. It provides a useful tool to coordinate operations and business controls which has great impact on decision making. Electricity load forecasting or electricity demand forecasting is one of the major areas considered in designing a power transmission and distribution systems (Al-Musaylh et al., 2018).

Electric load forecasting, or shorter load forecasting, comes under a variety of synonyms: electricity load forecasting; electricity demand forecasting; consumption forecasting; electricity load prediction; power demand prediction; load demand prediction; and load estimation etc. By definition, electric load forecasting is a realistic estimate of future demand of power (Matijaˇs, 2013).

Load forecasting has become a necessity due to its ability for the operator to have prior knowledge the electricity or load requirements. In effective management of electricity system, the most important component is the load or demand forecasting. Load forecasting is categorized in three main classes; namely,

1. Short-term load forecasting
2. Medium-term load forecasting
3. Long-term load forecasting

Short term load forecasting (STLF) basically predicts the load for every hour, day, and week. STLF is essential for the economic generation dispatch of power systems, energy transactions scheduling, load flows estimation and making decisions that can help avoid overloading or oversupply. This means that power generation can be planned according to the predicted value decreasing the risk of equipment failures such as transformer and transmission lines. The possibility of blackouts occurring is also reduced as well as losses in revenue of power utilities.

Medium-term load forecasting is predictions of demand for a week to a year in the future. It is necessary for the maintenance and network development work performed by power utilities. It involves the forecasting of systems load and sometimes bulk power interchange levels. Peak load demand is the main focus in this forecast but off-peak demand is also predicted.

Long-term load forecasting is for much longer duration into the future. The forecast period may last up to twenty years. Because of long duration of this forecast, it is much more difficult to forecast off-peak load demand and so the main emphasis is on the annual peak system load. Long-term load forecasting is essential in determining the load growth of a particular area. The peak load forecasted helps the power utilities in establishing whether upgrades of particular networks need to be performed. The capacity in terms of electrical power in that area may be exceeded thus making long term forecast, a key factor in planning operations of a utility.

1.2 Problem statement

Electricity is a clean energy compared to other energy generation sources. With rapid development of technology, need for electric energy has gradually increased across the world including Ghana(Gyamfi et al., 2018). However, there is no effective way of storing electrical energy, hence every generated energy supplied must be consumed(Tai, 2009). This creates the need for daily forecast. So every electricity operator is particular about the amount of energy to supply to meet the demand of the consumers and to reduce cost of transmission as well. Due to difficulties in transmission, the expected amount of energy at the demand end keeps fluctuating and this makes the problem stochastic. Yet the models employed by the Ghana Grid Company Limited (GRIDCo) are purely deterministic hence generating inaccurate demand values. If the right information about the demand is not given, the distributor is likely to oversupply or under supply energy to consumers leading to higher production cost of supply. Also, under supplying electricity might lead to blackout and low productivity of consumers.

1.3 Objectives

The main objective of this study is to develop a model to forecast the daily peak demand and daily total energy to be generated.

1.3.1 Specific Objectives

1. To predict the daily peak load and electricity demand using the Seemingly Unrelated Time Series Equations Model.

2. To forecast the next day's peak demand and total energy to be generated .

1.4 Methodology

1. Daily peak load data together with the daily energy generated data from February 2, 2014 to August 2, 2018, were taken from Ghana Grid Company Limited (GRIDCo).

2. The Seemingly Unrelated Time Series Equations Model was adopted to model the system.

3. The forward filtering and backward sampling Gibbs sampler algorithm was used to estimate the variance and states in the system. The one-step-ahead forecast values and the next day's peak and energy to be generated were estimated through the same algorithm.

1.5 Justification

Energy, especially electricity is an important utility in Ghana. It is one of the major component in the economic growth of every country since the consumers are also the contributors to the economy(Altinay and Karagol, 2005). Consumers always want electricity for production and other activities. Therefore, power failure will affect the productivity which will lead to poor performance of the economy. On the other hand, electricity operators always want to satisfy the demand of their consumers and also minimize the cost of transmission and supply. If the supply is above the demand, there will be a large amount of energy wasted thereby increasing cost of production. High-precision short-term load forecasting provides a great foundation and useful tool for electricity operator to make right decisions and to develop best power generation planning, scheduling and other activities. It is important to study the recent growing trend of the short-term load forecasting method and master relevant advanced technology, needed for improving the accuracy of short-term load forecasting. It would be a success to show the possibility of doing a research on this subject in Ghana. A Bayesian time series model which will be employed for this research is expected to offer better results than most classical methods due to its probabilistic nature and its ability to capture uncertainties in the data . This method has proven useful in fields like Finance and efficient at reducing errors involved in forecasting.

1.6 Thesis Organization

The study is organized into five(5) chapters as follows:

1. In chapter one, the background of this research, problem statement and the objectives for the study will be presented.

2. Literature review is discussed in chapter two.
3. Chapter three discusses some univariate and multivariate Dynamic Linear models, Kalman Filter and Smoother methods and also the Gibbs Sampling algo-

   rithm.

4. Implementation of the models and algorithms discussed in chapter three will be made available through chapter four.

5. Conclusion and Recommendation are in chapter 5.

1.7 Remarks

This chapter was centered on the background of the whole research along with the problem to solved. The targets for the research were set straight through the objectives for the research and also spelt out the organization of the whole work. In the next chapter, reviews will be done on already existing work on this research especially methods which had been applied in this field.

Chapter 2

LITERATURE REVIEW

2.1 Introduction

This chapter seeks to capture the reviews on the various stages of electricity production, types of load forecasting and some models which has been applied so far in load forecasting.

2.2 Overview of Electric Power Production

Electric power system is known as a real-time energy delivery system because power is always generated, transported and supplied at the instance the electric switch is turned on (Dwijayanti and Hagan, 2013). Basically, electricity goes through three main stages before it gets to the final consumer; namely,

1. The Generation Stage:

   This is the starting point of the electric power supply. At this process, power plants produce electrical energy by transforming other sources of energy, such as solar, heat, hydraulic,wind and fossil fuel.

2. The Transmission Stage:

   This is the second phase of the process. The generated energy from the power plants is transformed to high voltage electrical energy which is transmitted through the transmission lines. At this point, energy is transported from distant stations to load centers then to substations.

3. The Distribution Stage:

   This is the final stage where by the energy is stepped down again to be supplied to the various homes and other consumers.

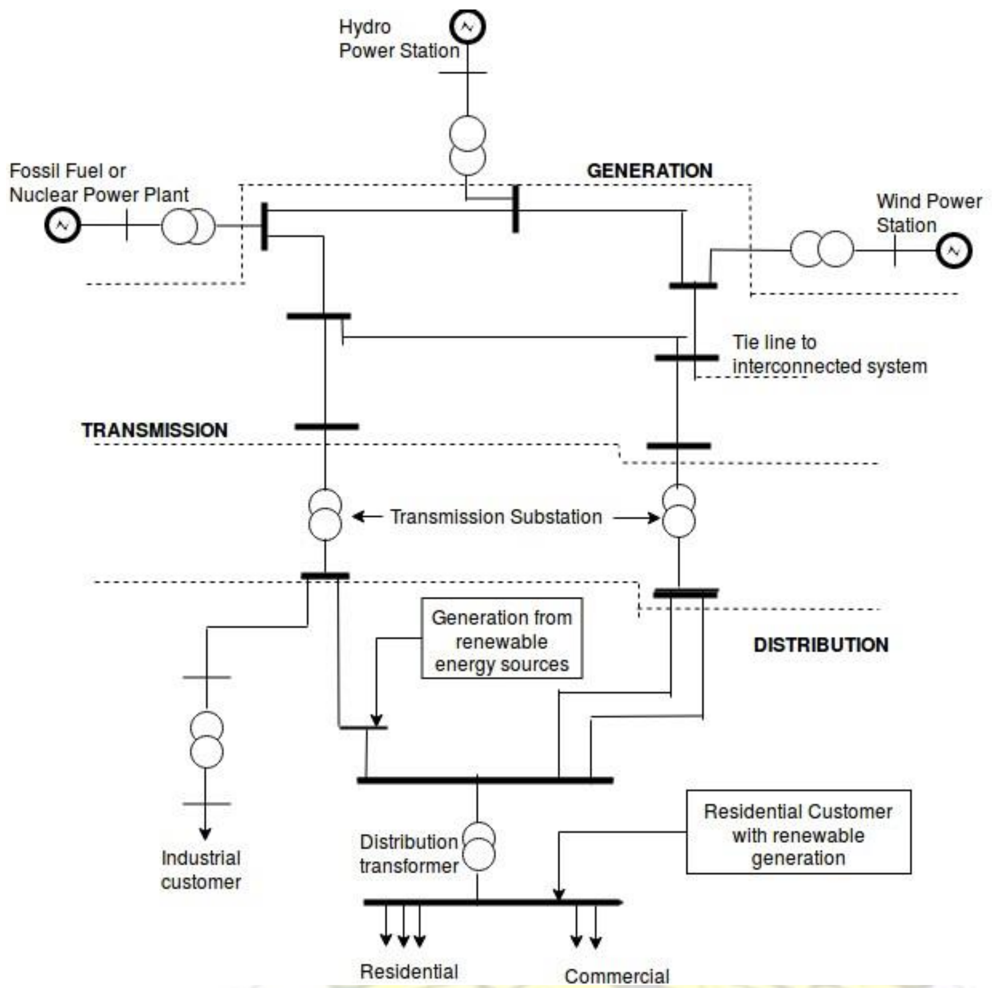


Figure 2.1: A Configuration of an Power System (Kyriakides and Polycarpou, 2007)

2.3 Classification of Load Forecasting

Load forecasting is classified into four major groups based on the time duration (Mishra, 2009).

1. Long-term load forecasting: This deals with load demand forecast with lead time of more than one year.

2. Medium-term load forecasting: It involves forecasting the load demand with lead time from one week to a year.
3. Short-term load forecasting: This deals with forecasting the load demand with lead time of 24 hours through to 168 hours.

4. Very Short-term load forecasting: This involves load demand forecast with lead time shorter than one day.

2.4 Short-Term Load Forecasting Models

According to Yang (2006), the approaches adopted in researching in this field can be categorized into two main methods;namely

1. Statistical Methods
2. Artificial Intelligence (AI) methods.

AI methods basically try to imitate humans way of thinking and reasoning to be able to acquire knowledge from past experience and to forecast the future load values while statistical methods uses equations showing the relationship between load and its relative factors after training the historical data. There are actually a lot of research done in modeling and forecasting energy and peak demand, showing the importance and the need in developing a more accurate model.

Every model comes with its pros and cons and none of these methods have supremacy over the other (Mirlatifi, 2016). There is always going to be a model better in a region more than another model. The energy demand models can also be classified based on the technique used in the forecasting as shown in the Figure 2.2.

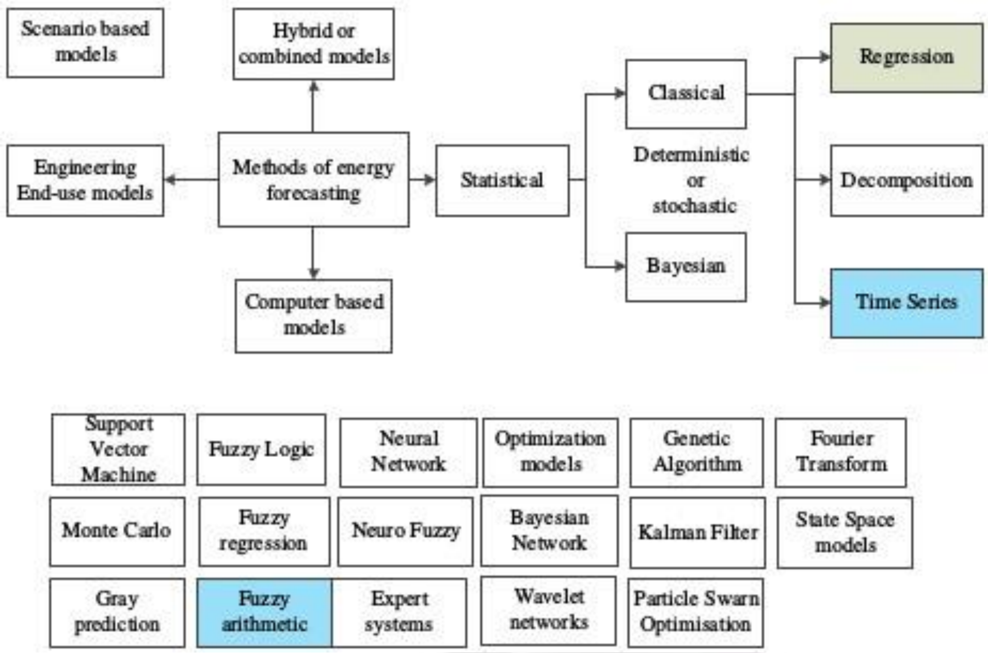


Figure 2.2: A Configuration of a Power System (Mirlatifi, 2016)

2.5 Time Series Models

Basically, the time series methods assume that the future or predicted values of the peak demand and the daily electricity generation or demand depends on historical data of the demands and also makes the assumption that the data follows certain stationary patterns. It follows that these patterns depend on trends and seasonal variations (Feinberg and Genethliou, 2005). These kinds of models started from the deterministic point of view which later developed through the idea of stochastic processes. The Autoregressive Integrated Moving Average (ARIMA) models, Exponential Smoothing (ES) models, Autoregressive (AR) models, the Moving Average models and the Structural Time Series Models (STSM) are some of the well-known or used time series models.

2.5.1 Autoregressive (Integrated) Moving Average

The Autoregressive Moving Average (ARMA) and Autoregressive Integrated Moving Average (ARIMA) models are the extensions of the Autoregressive (AR) models which assumes that the future value depends on the historical data and the Moving Average (MA) models which estimates the future error based on the past shocks in the system. It does this by combining these two models, AR and MA models and assumes that the error terms in the model follows the white noise process.

The ARMA and the ARIMA models with the help of the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF) selects the order of the MA and AR components respectively. With the Akaike's Information Criterion (AIC), multi-model partitioning filter (MMPF) and the Bayesian Information Criterion among others, the order of the model can be selected.

These methods have been applied in load forecasting and also in peak demand forecasting. A univariate ARMA model was applied in Greece for a long term electricity demand load forecasting with the use of the MMPF for model selection(Pappas et al., 2008). In Saudi Arabia, Eastern region, an electricity consumption forecasting model was developed using the ARIMA model based on Box-Jerkins which used a transfer function to overcome the effects of sudden changes in weather parameters (AbdelAal and Al-Garni, 1997). The ARIMAX model developed by Liu et al. (2015) for forecasting cooling, heating and electrical load in Victoria, Canada was used in the designing of a CCHP system (Mirlatifi, 2016). The seasonal ARIMA model and the

Box and Jenkins transfer function model was applied by Hagan (1977), Hubele and Cheng (1990) for STLF. Alfares and Nazeeruddin (2002) made a detailed review of the ES methods which considered these models as special case of the ARIMA and even the state space models(Mirlatifi, 2016). This method has been used as a traditional method for peak demand forecasting (Alfares and Nazeeruddin, 2002).

Li and Hong (2014) used the ES method in a demand response algorithm predicting required energy of appliances. The Exponential smoothing applied by Taylor et al. (2006) for forecasting the daily demand peak of Rio de Janeiro outperformed the ARIMA , Principal Components Analysis (PCA) and the Neural Network (NN) methods used. These models require the data to be stationary. Non-stationary data needs to be transformed before used.This might lead to loss of some information from the original data.

2.6 Regression Analysis

Regression model relates a dependent variable, electricity demand to some predictors or independent variable. It looks simple and easy to use. Bianco et al. (2009) applied a linear regression which used inputs like Gross Domestic Product (GDP),electricity consumption data, population and GDP per capita in long term electricity consumption forecasting in Italy.Goia et al. (2010) also used a functional regression for peak demand forecasting of a district which is a kind of generalization for multiple regression model. Regression models have challenges if the predictors are correlated. This will lead to Multicollinearity.

2.7 Neural Network

Artificial Neural Network (ANN) has proved to be excellent in scheduling problems which has made the method very popular (Mirlatifi, 2016). It has been used severally in load forecasting. The Neural Network stems its basis from the activities of the brain. It tries to mimic the operations of the human brain and based on this process, various combinations of networks have been utilized in solving practical problems.

Raza and Khosravi (2015) presented an overview on the use of Artificial Intelligence (AI) methods in electrical load forecasting in order to get the smart grid and smart buildings concept. In order to solve the problem of over-fitting and curse of dimensionality effects within this method, Cecati et al. (2015) proposed a radial basis function (RBF) training algorithm to help resolve such problems. The daily load consumption of a large building like hospital facilities, was forecasted using backpropagation Artificial Neural Network by determining the cofactors or regressors as: whether conditions, load, time of the day and special days like holidays(Bagnasco et al., 2015). Powell et al. (2014) applied the ANN to forecast heating, cooling and electricity load up to 24 hours pertaining to a large scale district. It was also applied in forecasting peak electricity demand up to an hour for a large government building (Grant et al., 2014). Cecati et al. (2015) did a review on some well-known ANN algorithms used for a 24 hour forecasting of electricity load.The feed-forward neural network with a back-propagation algorithm was implemented by Bhattacharyya and

Thanh (2004) to predict the short term load demand of northern areas of Vietnam. Lo and Wu (2003) proposed a method which can be used in evaluating the risk involved in the supply industry in UK which identified no strong correlation existing between the local demand and weather based on their choice of sample used for the study (Mirlatifi, 2016). ANN has proven very using in prediction but it requires large size of data for the training process to be effective in prediction.

2.8 Support Vector Machine

Support Vector Machine (SVM) is a popular machine learning method which was invented in the year 1963 and later developed into a more useful algorithm in different fields. It's basically divided into two parts;

1. Support Vector Classification (SVC): This actually is used for classification problems.

2. Support Vector Regression (SVR): This deals with modeling and prediction.

In this method, the training of the dataset is done in an iterative manner and it does so by mapping the data to a higher dimensional space rather than the original space. To achieve a much better results, the training data size can be increased (Mirlatifi, 2016).

The Support Vector Machine (SVM) was used for in forecasting the short term electrical load (Mohandes, 2002) in Eastern Saudi Arabia and it was established that unlike the NN or AR, the trained data wasn't limited in SVM. The SVM was also applied in time series forecasting for a medium term electric load forecasting (Chen et al., 2004). It was established that performance was improved based on proper segmentation and large error can be incurred due to incorrect data predictions. Medium Term Electricity demand forecasting was done in Italy with a linear regression model and SVM applied forecast of seasonal climatic temperature in the predictions (De Felice et al., 2015). The Support Vector Regression (SVR) was applied for a long term modeling and prediction of the consumption of electricity from 2007 to 2026 in Turkey

(Kavaklioglu, 2011). In Taiwan, the SVM was applied together with ARIMA and GRNN where the SVM outperformed the other two methods(Pai and Hong, 2005). The parameter selection was done using the Simulated Annealing (SA) algorithm and later applied the SVM method (Pai and Hong, 2005). Forecasting was done by Global Solar Radiation (GSR) in Iran for the solar power system design and implementation which came out with the SVR outperforming the Fuzzy Linear Regression (FLR)(Mirlatifi, 2016). Just like the other AI techniques, this method requires a larger dataset for the training process to do effective classification or prediction.

2.9 Fuzzy Models

Zadeh introduced the idea of Fuzzy set theory in the year 1965 (Zadeh, 1965). This has become a major field in solving problems with uncertainty. This set is actually an extension of the classical set and this set, every member carries some probability. This set theory has been grouped into several subdomains which includes Fuzzy regression, fuzzy logic, fuzzy control and fuzzy pattern recognition. The fuzzy method is one of the most used methods in modeling and forecasting of energy.

2.9.1 Fuzzy Logic

The fuzzy logic was designed to mimic the day to day human and animal activities. It is able to solve problems with uncertainty and imprecision where the usual discrete logic will fail. Unlike the well known logic, specifically the Boolean logic which uses 0 and 1 for false and true statements, the fuzzy logic allows a varying value from 0 to 1 and the antecedents and consequences are referred to as fuzzy propositions.

This method was used in Bahia state, Brazil(Pereira et al., 2015). In Turkey, Kucukali and Baris (2010) used the fuzzy logic for yearly electricity demand forecast which concluded that GDP affects the annual demand.

2.9.2 Fuzzy Regression

Fuzzy Linear regression was applied in Iran for solar radiation prediction(Ramedani et al., 2014). An intelligent algorithm for energy consumption for Iran incorporated the fuzzy regression (Azadeh et al., 2010).

2.10 Bayesian Methods

This is a probabilistic method. The Bayes' theorem can be applied to include probability in the demand of appliances as part of electricity demand algorithm (Li and Hong, 2014). This method can be applied in model selection (Box et al., 1995);(Pappas et al., 2008). It can also be applied in AI to solve problems with some level of uncertainty. This technique is referred to as Bayesian Networks which tackles problems with randomness, uncertainty or both(Mirlatifi, 2016). The idea of this method was applied in the field of renewable energy, mostly in hydroelectricity and wind power (Borunda et al., 2016). The Dynamic Bayesian Network (DBN) was used in wind power prediction of a wind farm located in Mexico for a 5 hour time horizon (Ibargu¨engoytia et al., 2014).

2.11 Error Accuracy Measure

These are methods used in literature to measure the performance of models used in forecasting.

Mean Absolute Error (MAE)

$$MAE = \frac{1}{T}\sum_{t=1}^{T} |p_t - p_t'| \qquad (2.11.1)$$

Mean Square Error (MSE)

$$MSE = \frac{1}{T}\sum_{t=1}^{T}(p_t - p_t')^2 \qquad (2.11.2)$$

Root Mean Square Error (RMSE)(Pappas et al., 2008)

$$RMSE = \sqrt{\frac{1}{T}\sum_{t=1}^{T}(p_t - p_t')^2} \qquad (2.11.3)$$

Normalized Root Mean Square Error (NRMSE) (Pai and Hong, 2005)

$$NRMSE = \frac{\sqrt{\sum_{t=1}^{T}(p_t - p'_t)^2}}{\sum_{t=1}^{T} p_t} \quad (2.11.4)$$

Mean Absolute Percentage Error (MAPE)(Azadeh et al., 2010)

$$MAPE = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{p_t - p'_t}{p_t}\right| \quad (2.11.5)$$

where p_t – Actual values

p'_t – Forecast values

T – Number of actual or forecast values.

2.12 Remark

Review of existing work was done in this chapter. The concentration was based on the energy generation process, the classification of load forecasting and a few methods which have already been applied in this field. The measure of accuracy of forecasting was also discussed. With this in place,the chapter seeks to discuss the method to be adopted for this research.

Chapter 3
Methodology

3.1 Introduction

This chapter contains Dynamic Linear Models (DLMs) and techniques employed to carry out analysis on forecasting daily energy generated and the peak demand based on data from the Ghana Grid Company Limited (GRIDCo) from February 2, 2014 to August 2,2018. The details of Kalman filter and Smoother which are adopted in the DLMs is being discussed in this chapter.

The algorithm chosen for this study in estimating the parameters in the adopted model, Seemingly Unrelated Time Series Equations (SUTSE) model been discussed in this chapter is the Markov Chain Monte Carlo (MCMC) methods, specifically Gibbs sampler. The Forward Filtering and Backward Sampling (FFBS) algorithm is employed with the Gibbs sampler,using a multivariate state space model to generate one-step-ahead forecast of the total daily energy generated together with the daily peak load and to make inferences about the variances in the system.

3.2 Bayes' Theorem

Bayes' theorem is the basis of the DLMs. This comes with the idea of conditional probabilities. Assuming E and F are events (a subset of the sample space which contains outcomes that had occurred) and $P(F) > 0$, then the conditional probability of E given that F has occurred is given by

$$P(E \cap F)$$

$$P(E|F) = \frac{\qquad}{P(F)} \tag{3.2.1}$$

Hence,the joint probability of both events happening can be given by

$$P(E \cap F) = P(F)P(E|F) \tag{3.2.2}$$

Another important concept in the fundamentals of the Bayes' theorem is the concept of mutual exclusiveness. If $E = (E \cap F) \cup (E \cap F^c)$ then $(E \cap F)$ and $(E \cap F^c)$ are said to mutually exclusive. So,

$$P(E) = P(E \cap F) + P(E \cap F^c) \tag{3.2.3}$$

$$P(E) = P(E|F)P(F) + P(E|F^c)P(F^c) \tag{3.2.4}$$

$$P(E) = P(E|F)P(F) + P(E|F^c)[1 - P(F)] \tag{3.2.5}$$

$P(E)$ is referred to as the marginal probability.

Bayes' theorem also referred to as **the theorem of inverse probability** is a consequence of the joint probability equation which states that

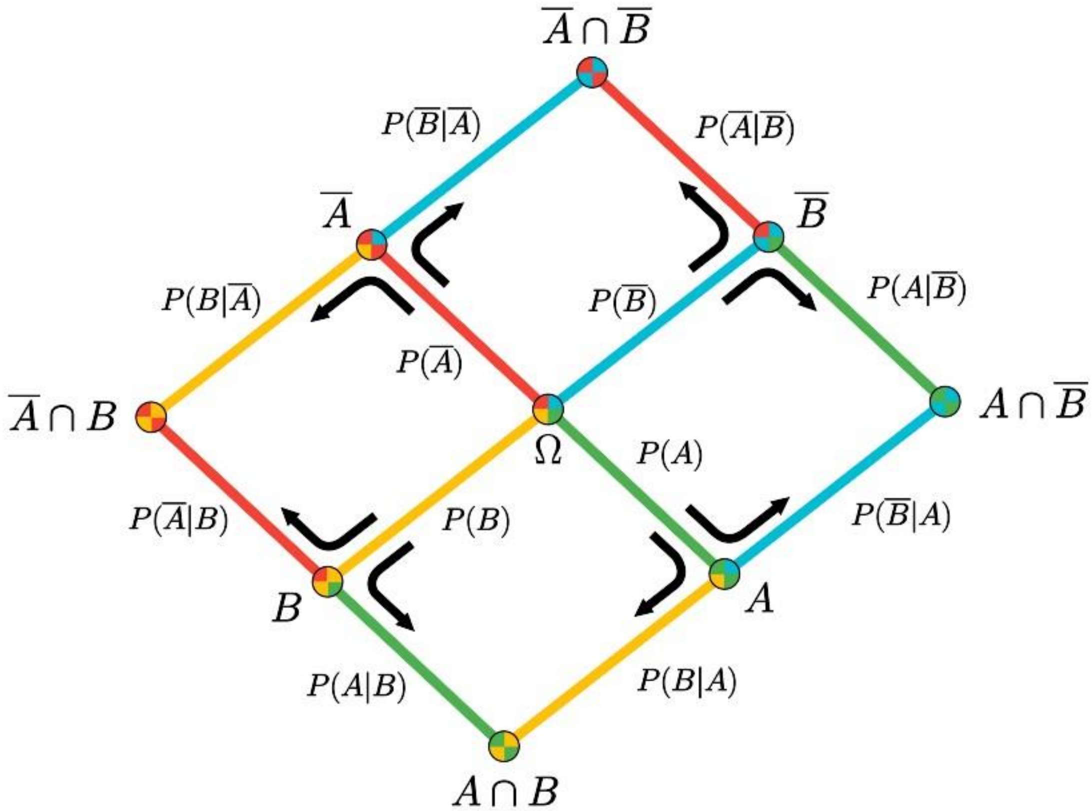
$$P(E|F) = \frac{P(E|F)P(E)}{P(F)} \tag{3.2.6}$$

where

1. $P(E|F)$ is a conditional probability. Thus the likelihood of event E occurring given that F is true.
2. $P(F|E)$ is a conditional probability which is the likelihood of event F occurring given that E is true.
3. P(E) and P(F) are known as marginal probabilities. They are the respective probability of observing E and F independently.

4. So Bayes' theorem basically describes the probability associated with an event, based on some prior knowledge of conditions that might have some relation with the event.

Figure 3.1 shows the Bayes' Theorem framework as discussed above.



$$P(A|B) \cdot P(B) = P(A \cap B) = P(B|A) \cdot P(A)$$

Figure 3.1: Bayes' Framework

The diagram in Figure 3.2 illustrates the application of the Bayes' theorem.

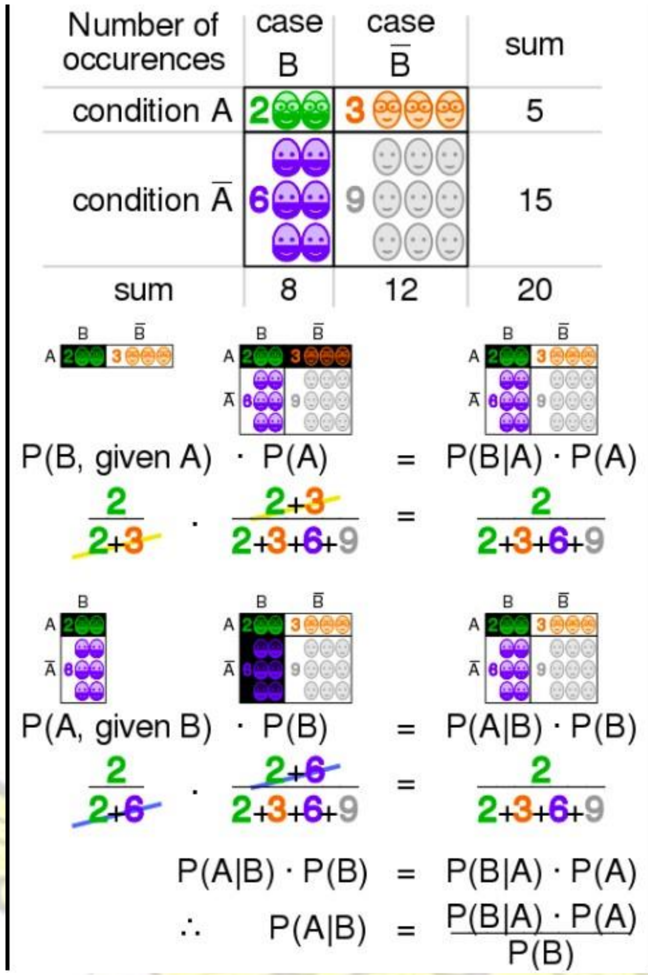


Figure 3.2: Example Illustrating the Bayes' Theorem

Assuming the quantity of interest is a vector θ of the parameters of a model and y is an observed data. So for the Bayes' theorem in the general form, the Bayes formula for computing the conditional density of θ given y:

$$\pi(\theta|y) = \frac{\pi(y|\theta)\pi(\theta)}{\pi(y)} \quad (3.2.7)$$

where $\pi(y) = \int_{\theta \in \Theta} \pi(y|\theta)\pi(\theta)$ is the marginal distribution, $\pi(\theta)$ is a prior distribution and $\pi(y|\theta)$ is termed as the likelihood.

3.3 State Space Models

State space models picture a time series as the output of some dynamic system which has been perturbed by some random disturbance.These models can be used for both univariate and multivariate time series.The fundamental concept of state space models is to use some latent variables to estimate and predict the observations.It does this using a recurrence approach.

The usual time series methods works best when there are regular patterns in trend and seasonality(Petris et al., 2009). In using these methods, the dataset is required to be stationary. With state space models, the data are not required to be stationary because the non-stationarity and any irregularities in the trend will be catered for in the model(Petris et al., 2009). It can be formulated in such a way to account for structural breaks or changes, outliers and level shifts.

It is able to naturally analyze data that shows instability in mean level and even in variance, structural breaks and sudden jumps due to it's ability to model each component in the system (Petris et al., 2009).

With Bayesian approach where joint probability is of much interest, computations are made simple as possible. For simplicity, Markovian dependence which is a simpler form is assumed. $(Y_t)_{t\geq 1}$ is a Markov Chain if, for any $t \geq 1$,

$$\pi(y_t|y_{1:t-1}) = \pi(y_t|y_{t-1}) \tag{3.3.1}$$

Therefore, Y_t and $Y_{1:t-2}$ are conditionally independent given y_{t-1}. The finite dimensional distribution can be written as

$$\pi(y_{1:t}) = \pi(y_1)\prod_{j=2}^{t}\pi(y_j|y_{j-1}) \tag{3.3.2}$$

In state space model, it is assumed that there exist an unobservable Markov chain θ_t referred to as **the state process** and that the measurement Y_t is an inexplicit measure of θ_t. Formally, this kind of model consists of an $\mathbb{R}^p$-valued time series $(\theta_t : t = 1,2,\ldots)$ and an $\mathbb{R}^m$-valued time series $(Y_t : t = 1,2,\ldots)$; which satisfies the assumptions in figure 3.3:

A1. (θ_t) is a Markov process.

A2. Conditionally on (θ_t), the Y_t's are individualistic and Y_t depends solely θ_t.

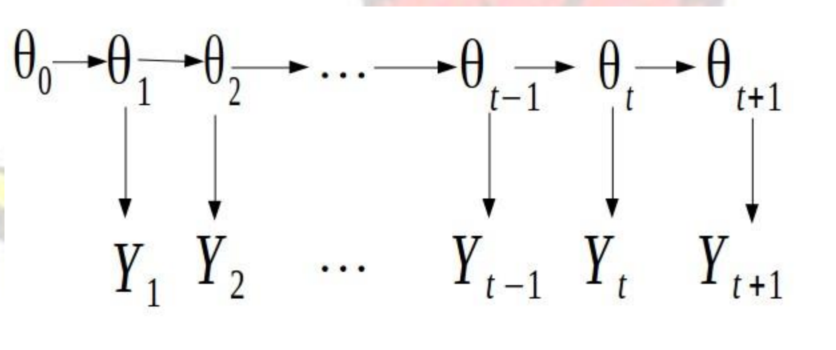


Figure 3.3: Dependence structure for a state space model

For any $t \geq 0$ and $\pi(\theta_0)$ as the prior distribution and $\pi(\theta_t | \theta_{t-1})$ and $\pi(y_t | \theta_t)$ as conditional densities,

$$\pi(\theta_{0:t}, y_{1:t}) = \pi(\theta_0) \times \prod_{j=1}^{t} \pi(\theta_j | \theta_{j-1}) \pi(y_j | \theta_j) \qquad (3.3.3)$$

By integrating equation (3.3.3) out θ_j's, joint density of the observations can be obtained.

So,

$$\pi(y_t | \theta_{0:t-1}, y_{1:t-1}) = \pi(y_t | \theta_t) \tag{3.3.4}$$

Similarly,

$$\pi(\theta_t | \theta_{0:t-1}, y_{1:t-1}) = \pi(\theta_t | \theta_{t-1}) \tag{3.3.5}$$

3.4 Dynamic Linear Models (DLMs)

State space models described by a prior distribution for θ_0 with the observation and state equations

$$Y_t = h_t(\theta_t, v_t) \tag{3.4.1}$$

$$\theta_t = g_t(\theta_{t-1}, w_t) \tag{3.4.2}$$

for some arbitrary functions g_t and h_t, and v_t and w_t are individualistic sequences of independent random vectors.g_t and h_t are described as linear functions when dealing with Linear space models. With Gaussian Linear Models, the idea of Gaussian distribution is factored in the model.

The Gaussian linear state space models,also known as Dynamic Linear Models (DLMs) can be defined by a prior distribution which is normally distributed for the p-dimensional state vector at time $t = 0$,

$$\theta_0 \sim \mathrm{N}_p(m_0, C_0) \tag{3.4.3}$$

with the observation and evolution equations for each time $t \geq 1$

$$Y_t = F_t\theta_t + v_t \qquad v_t \sim \mathrm{N}_p(0, V_t) \tag{3.4.4}$$

$$\theta_t = G_t\theta_{t-1} + w_t \qquad w_t \sim \mathrm{N}_p(0, W_t) \tag{3.4.5}$$

where

G_t and F_t are known matrices (of order $p \times p$ and $m \times p$ respectively). $(v_t)_{t\geq 1}$ and $(w_t)_{t\geq 1}$ are two independent Gaussian random vectors which are not dependent on each other and has a mean of zero and known variance matrix $(V_t)_{t\geq 1}$ and $(W_t)_{t\geq 1}$

respectively. Equation (3.4.4) is referred to as the **Observation Equation** while Equation (3.4.4) is known as the **State Equation or System Equation**.

A simple model for univariate time series $(Y_t : t = 1,2,...)$ is the **Random Walk plus Noise Model** which is also known as **Local Level Model** is defined as

$$Y_t = \mu_t + v_t \qquad v_t \sim N(0,V) \tag{3.4.6}$$

$$\mu_t = \mu_{t-1} + w_t \qquad w_t \sim N(0,W) \tag{3.4.7}$$

where the error sequences (v_t) and w_t are not dependent both within and between them.The Local level model is a DLM with $p = m = 1$ (thus $\mathbb{R}^1$) and $F_t = G_t = 1$.

Intuitively, this is appropriate for time series showing no clear trend or seasonal variation.

Another well-known DLM is the **Linear Growth Model** or **Local Linear Trend**, which has an observation equation similar to that of the local level model. It also incorporates a time varying slope in the dynamics of the model.

$$Y_t = \mu_t + v_t \qquad v_t \sim N(0,V) \tag{3.4.8}$$

$$\mu_t = \mu_{t-1} + \beta_{t-1} + w_{t,1} \qquad w_{t,1} \sim \mathcal{N}(0,\sigma^2_\mu) \tag{3.4.9}$$

$$\beta_t = \beta_{t-1} + w_{t,2} \qquad w_{t,2} \sim N(0,\sigma^2_\beta) \tag{3.4.10}$$

with uncorrelated errors v_t,$w_{t,1}$ and $w_{t,2}$. The linear trend model DLM has

$$\theta_t = \begin{pmatrix} \mu_t \\ \beta_t \end{pmatrix} \tag{3.4.11}$$

$$G = \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} \tag{3.4.12}$$

$$W = \begin{pmatrix} \sigma_\mu^2 & 0 \\ 0 & \sigma_\beta^2 \end{pmatrix} \quad (3.4.13)$$

$$F = \begin{pmatrix} 1 & 0 \end{pmatrix} \quad (3.4.14)$$

μ_t is explained as the local level and β_t interpreted as the local growth rate. Intuitively, this model assumes that the current level μ_t linearly changes through time and the growth rate is likely to evolve over time.

3.5 Filtering

The main focus of this process is computing the filtering densities $\pi(\theta_t|y_{1:t})$ in the state space models.

3.5.1 Proposition 1 (Filtering Recursions)

For a generalized state space model defined by the assumptions (A1) and (A2) stated in Figure 3.3, the following statements hold.

1. The one-step-ahead predictive density for the state can be computed from the filtered density $\pi(\theta_{t-1}|y_{1:t-1})$ according to

$$\pi(\theta_t|y_{1:t-1}) = \int \pi(\theta_t|\theta_{t-1})\pi(\theta_{t-1}|y_{1:t-1})d\theta_{t-1} \quad (3.5.1)$$

2. The one-step-ahead predictive density for the observations can be computed from the predictive density for the states as

$$\pi(y_t|y_{1:t-1}) = \int \pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})d\theta_t \quad (3.5.2)$$

3. The filtering density can be computed from the densities in equations (3.5.1) and (3.5.2) as

$$\pi(\theta_t|y_{1:t}) = \frac{\pi(y_t|\theta_t)\pi(\theta_t|y_{1:t-1})}{\pi(y_t|y_{1:t-1})} \qquad (3.5.3)$$

3.5.2 Kalman Filter

Kalman filter is an optimal estimation algorithm. It is an iterative method which make use of a set of equations and consecutive data inputs in estimating the true value, position, velocity etc of an object and the measured value contains unpredicted or random error, or uncertainty or variations. Mostly used when there are estimated measurements rather than the actual measurements or the observations are thought of to have some errors in them. It is also used to combine measurements to estimate the state of an object.

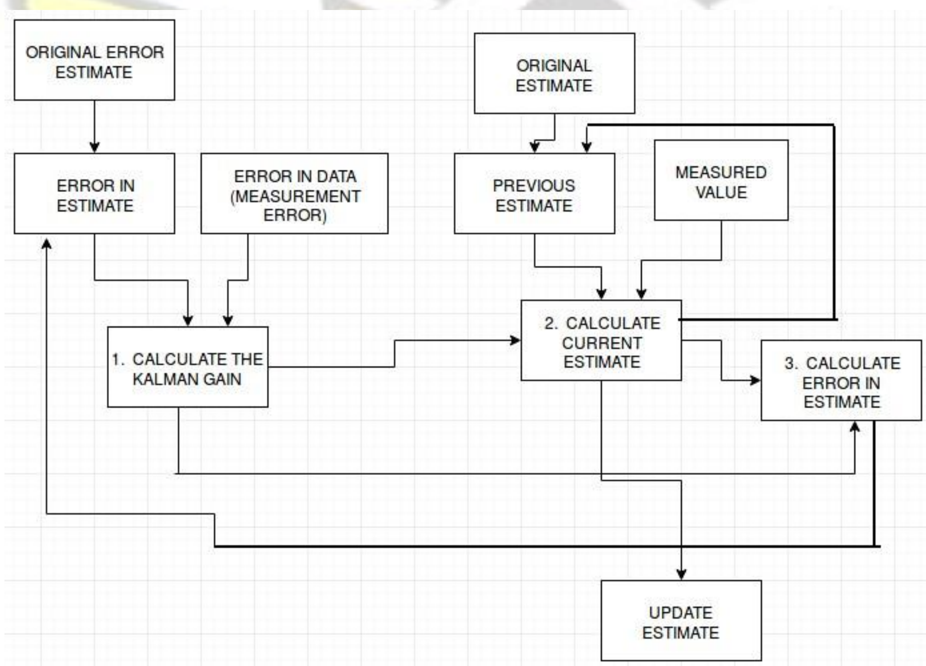


Figure 3.4: Single Measured Value Kalman Filter

The concentration on this method is focused on three main equations. The Kalman gain (KG) given in equation (3.5.4) below is basically the weight placed either on the measured value or the estimated value,

$$KG = \frac{E_{EST}}{E_{EST} + E_{MEA}} \tag{3.5.4}$$

where E_{EST} is the error in the estimate, E_{MEA} is the error in measurements and $0 \le KG \le 1$. If $KG = 1$ or very close to 1, then the measurements are accurate and estimates are unstable. The measurements are said to be inaccurate and the estimates being stable if KG is closer to zero.

The current estimate EST_t is given by:

$$EST_t = EST_{t-1} + KG[MEA - EST_{t-1}] \tag{3.5.5}$$

where EST_{t-1} is the previous estimate and MEA is the measurement. If KG is closer to 0 then the error in the measurement is large. Hence, more weight is placed on the previous estimate. If KG is closer to 1 then the error in the estimate is large and the error in the measurement is less. Hence, more weight is placed on the measurement. The deviation in the current estimate is given by equation (3.5.6).

$$E_{EST_t} = (1 - KG)E_{EST_{t-1}} \tag{3.5.6}$$

The diagram in Figure 3.5 shows a multi-dimensional flowchart for Kalman filter.

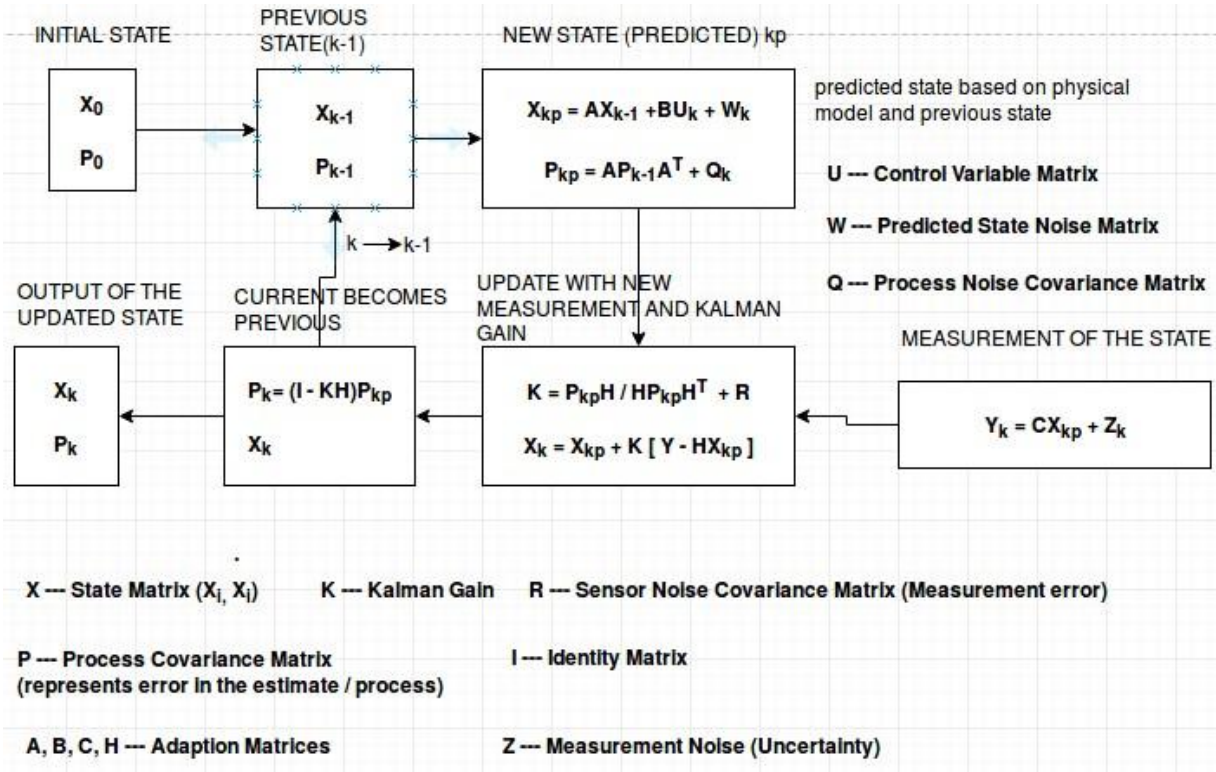


Figure 3.5: Kalman Filter for Multi-Dimensional Model

3.5.3 Kalman Filter for DLMs

For DLMs, the Gaussian distribution plays a key role in these kind of models. So the marginal and conditional distributions follow the Gaussian distribution. Hence, the relevant parameters here are the means and variances. Kalman filter gives the solution to the DLMs.

3.5.4 Proposition 2 (KALMAN FILTER)

Consider the DLM

$$\theta_0 \sim N_p(m_0, C_0) \tag{3.5.7}$$

$$Y_t = F_t\theta_t + v_t \qquad v_t \sim N_p(0, V_t) \tag{3.5.8}$$

$$\theta_t = G_t\theta_{t-1} + w_t \qquad w_t \sim N_p(0, W_t) \tag{3.5.9}$$

for $t \geq 1$. Let $\theta_{t-1}|y_{1:t-1} \sim N_p(m_{t-1}, C_{t-1})$ Then the following statements hold.

1. The one-step-ahead predictive distribution of θ_t given $y_{1:t-1}$ is Gaussian, with parameters

$$a_t = E(\theta_t|y_{1:t-1}) = G_t m_{t-1} \quad (3.5.10)$$

$$R_t = Var(\theta_t|y_{1:t-1}) = G_t C_{t-1} G'_t + W_t \quad (3.5.11)$$

2. The one-step-ahead predictive distribution $Y_t|y_{1:t-1}$ is Gaussian, with parameters

$$f_t = E(Y_t|y_{1:t-1}) = F_t a_t \quad (3.5.12)$$

$$Q_t = Var(Y_t|y_{1:t-1}) = F_t R_t F'_t + V_t \quad (3.5.13)$$

3. The filtering distribution $\theta_t|y_{1:t}$ is Gaussian, with parameters

$$m_t = E(\theta_t|y_{1:t}) = a_t + R_t F'_t Q_t^{-1} e_t \quad (3.5.14)$$

$$C_t = Var(\theta_t|y_{1:t}) = R_t - R_t F'_t Q_t^{-1} F_t R_t \quad (3.5.15)$$

where $e_t = Y_t - f_t$

The gain matrix given by

$$K_t = R_t F'_t Q_t^{-1} \quad (3.5.16)$$

gives the weight of the correction term.Thus, the observation variance V_t influences the weight placed on the current data point through Q_t and on $R_t = Var(\theta_t|y_{1:t-1}) = G_t C_{t-1} G'_t + W_t$.

3.6 Smoothing

3.6.1 Proposition 3(Smoothing Recursion)

For a generalized state space model given by the Markovian and Conditional independence assumptions, the following statements hold.

1. Conditional on $y_{1:T}$,the state sequence $(\theta_0,\ldots,\theta_T)$ has backward transition probabilities given by

$$\pi(\theta_t|\theta_{t+1},y_{1:T}) = \frac{\pi(\theta_{t+1}|\theta_t)\pi(\theta_t|y_{1:t})}{\pi(\theta_{t+1}|y_{1:t})} \tag{3.6.1}$$

2. The Smoothing distributions of θ_t given $y_{1:T}$ can be computed according to the following backward recursion in t, starting from $\pi(\theta_T|y_{1:T})$:

$$\pi(\theta_t|y_{1:T}) = \pi(\theta_t|y_{1:t})\int \frac{\pi(\theta_{t+1}|\theta_t)}{\pi(\theta_{t+1}|y_{1:T})}\pi(\theta_{t+1}|y_{1:T})d\theta_{t+1} \tag{3.6.2}$$

3.6.2 Proposition 4(Kalman Smoother)

For a DLM defined by (3.5.7)-(3.5.9), if $\theta_{t+1}|y_{1:T} \sim N(s_{t+1},S_{t+1})$,then $\theta_t|y_{1:T} \sim N(s_t,S_t)$,where

$$s_t = m_t + C_tG'_{t+1}R^{-1}_{t+1}(s_{t+1} - a_{t+1}) \tag{3.6.3}$$

$$S_t = C_t - C_tG'_{t+1}R^{-1}_{t+1}(R_{t+1} - S_{t+1})R^{-1}_{t+1}G_{t+1}C_t \tag{3.6.4}$$

3.7 Forecasting

Through the process of Kalman filtering, the one-step-ahead forecast distributions for the states and measurements or observations are given as byproducts as shown in proposition 1. The magnitude of the gain matrix plays an important role in the process of determining how sensitive f_t is, the one-step-ahead predicted value is to the previous observation y_{t-1}. This is because for each t, f_t is a linear function of the filtering mean m_{t-1}.

3.7.1 Proposition 5 (Forecasting Recursion)

For a generalized state space model which is defined by the Markovian and the conditional independence assumptions, the following statements hold. 1. The k-steps-ahead predictive distribution of the state is given as

$$\pi(\theta_{t+k}|y_{1:t}) = \int \pi(\theta_{t+k}|\theta_{t+k-1})\pi(\theta_{t+k-1}|y_{1:t})d\theta_{t+k-1} \tag{3.7.1}$$

2. The k-steps-ahead predictive distribution of the observation is given as

$$\pi(y_{t+k}|y_{1:t}) = \int \pi(y_{t+k}|\theta_{t+k})\pi(\theta_{t+k}|y_{1:t})d\theta_{t+k} \tag{3.7.2}$$

3.8 Bayesian Inference

Bayesian approach considers the unknown parameters, say ψ as a random vector. The assumptions, Markov process and conditional independence underlying state space models for the processes (Y_t) and (θ_t) are assumed to hold conditionally on the parameters ψ. So, the prior knowledge is established based on a probability law $\pi(\psi)$.Thus, for $n \geq 1$,

$$(\theta_0,\theta_1,...\theta_n,Y_1,...,Y_n,\psi) \propto \pi(\theta_0|\psi)\prod_{t=1}^{n}\pi(y_t|\theta_t,\psi)\pi(\theta_t|\theta_{t-1},\psi) \tag{3.8.1}$$

From the concept stated above, the filtering density can be defined by

$$\pi(\theta_t|y_{1:t}) = \int_{\psi\in\Theta} \pi(\theta_t|\psi,y_{1:t})\pi(\psi|y_{1:t})d\psi \qquad (3.8.2)$$

All the unknown states' history up to time t can be reconstructed; given the observations or data $y_{1:t}$, inference on $\theta_{0:t}$ and ψ can be expressed by the joint posterior density

$$\pi(\theta_{0:t},\psi|y_{1:t}) = \pi(\theta_{0:t}|\psi,y_{1:t})\pi(\psi|y_{1:t}) \qquad (3.8.3)$$

One challenge about the normal or recursive method for Bayesian inference is that computations are mostly intractable. However, the introduction of the MCMC methods and sequential Monte Carlo algorithms has provided efficient tools for approximating posterior distributions of interest. Gibbs sampler which is an MCMC method has been widely used in approximating the joint posterior distribution $\pi(\theta_{0:t},\psi|y_{1:t})$ (Petris et al., 2009). This methods requires iteratively simulating from full conditional distributions $\pi(\theta_{0:t}|\psi,y_{1:t})$ and $\pi(\psi|\theta_{0:t},y_{1:t})$. Gibbs sampler can also be used to approximate the filtering density, $\pi(\theta_t|y_{1:t})$ and the marginal smoothing density $\pi(\theta_s|y_{1:t}), s < t$. These simulated samples from the posterior distribution can be used as input to generate a sample from the predictive distribution of states and observables, $\pi(\theta_{t+1:t+k},y_{t+1:t+k}|y_{1:t})$. So,

$$\pi(\theta_{t+1:t+k},y_{t+1:t+k},\psi,\theta_t|y_{1:t}) = \pi(\theta_{t+1:t+k},y_{t+1:t+k}|\psi,\theta_t).\pi(\psi,\theta_t|y_{1:t}) \qquad (3.8.4)$$

The simulations from the posterior automatically solves the filtering, smoothing and forecasting problems for a Dynamic Linear Model with unknown parameters.

3.9 Gibbs Sampler

Gibbs sampling,which is a Markov Chain Monte Carlo (MCMC) method was named after the physicist,Josiah Willard Gibbs, based on a comparison between the sampling algorithm and statistical physics. The description of this algorithm was done by Stuart and Donald Geman in 1984, about eighty years after the death of Willard Gibbs. Suppose the unknown parameter ψ is multidimensional,hence the posterior distribution will be in a multivariate form. Thus $\psi = (\psi^{(1)},...,\psi^{(k)})$. Let the target density be represented as $\pi(\psi) = \pi(\psi^{(1)},...,\psi^{(k)})$. This algorithm starts from an arbitrary

point $\psi_0 = (\psi_0^{(1)},...,\psi_0^{(k)})$ in the parameter space and each component is updated at a time by drawing ψ^{i}, $i = 1,...,k$ from the necessary conditional distribution. The Gibbs sampler still works even when one or more of the components $\psi^{(i)}$ itself is multidimensional (Petris et al., 2009). In such a situation, the Gibbs sampler in turn updates ”blocks” of components of the parameter ψ drawn from relevant conditional distributions, given the other components. Below is the Gibbs Sampling algorithm.

Algorithm 1 : Gibbs Sampler

---

0. Initialize the starting point $\psi_0 = (\psi_0^{(1)},...,\psi_0^{(k)})$;
1. for j= 1,...,N:

   1.1 generate $\psi_j^{(1)}$ from $\pi(\psi^{(1)}|\psi^{(2)} = \psi_{j-1}^{(2)},...,\psi^{(k)} = \psi_{j-1}^{(k)})$

   1.2 generate $\psi_j^{(2)}$ from $\pi(\psi^{(2)}|\psi^{(1)} = \psi_j^{(1)},\psi^{(3)} = \psi_{j-1}^{(3)},...,\psi^{(k)} = \psi_{j-1}^{(k)})$

   ...

   1.k generate $\psi_j^{(k)}$ from $\pi(\psi^{(k)}|\psi^{(1)} = \psi_j^{(1)},...,\psi^{(k-1)} = \psi_j^{(k-1)})$

MCMC methods have the ergodic property. Hence, guarantees that the stationary distribution of the samples generated under Gibbs sampling is the target joint posterior of interest(Gilks et al., 1996). Therefore, ergodic means can be estimated using the law of large numbers given in equation (3.9.1)(Yildirim, 2012).

$$E_{\pi}[g(\psi)] \approx \frac{1}{N}\sum_{i=1}^{N} g(\psi_j) \tag{3.9.1}$$

where $\pi \rightarrow$ posterior distribution of interest $g(\psi) \rightarrow$ desired expectation

$g(\psi_j) \rightarrow j^{th}$ simulated sample from π

$N \rightarrow$ Number of iterations of simulated Markov Chains.

3.9.1 Forward Filtering Backward Sampling (FFBS)

Forward Filtering Backward Sampling (FFBS) is basically the simulation version of the smoothing recursions. The joint distribution of $\theta_{0:T} | y_{1:T}$ as

$$\pi(\theta_{0:T} | y_{1:T}) = \prod_{t=0}^{T} \pi(\theta_t | \theta_{t+1:T}, y_{1:T}) \tag{3.9.2}$$

where the last factor in the product is the filtering distribution of θ_T, $\pi(\theta_T | y_{1:T})$ is $N(m_T, C_T)$. Equation (3.9.2) suggests that in obtaining a draw from the left hand side of the equation, θ_T can be drawn from $N(m_T, C_T)$ and then, for $t = T-1, T-2, \ldots, 0$,

[1] . Run Kalman filter.

[2] . Draw $\theta_T \sim N(m_T, C_T)$.

[3] . For $t = T-1, \ldots, 0$, draw $\theta_t \sim N(h_t, H_t)$

recursively draw θ_t from $\pi(\theta_t | \theta_{t+1:T}, y_{1:T})$. This process is termed as the forward filtering process. Then the next step is the smoothing process which does the backward sampling process. Thus, drawing θ_t from a $N(h_t, H_t)$ where

$$h_t = m_t + C_t G'_{t+1} R^{-1}_{t+1}(\theta_{t+1} - a_{t+1}) \tag{3.9.3}$$

$$H_t = C_t - C_t G'_{t+1} R^{-1}_{t+1} G_{t+1} C_t \tag{3.9.4}$$

The FFBS algorithm can be done by simply undergoing the Kalman filter for the forward filtering and a modified Kalman Smoother for the backward sampling process. Below is the summary of the Forward Filtering Backward Sampling algorithm.

Algorithm 2 : Forward Filtering Backward Sampling

3.10 Forward Filtering Backward Sampling with Gibbs Sampler

Algorithm 3 : Forward Filtering Backward Sampling in a Gibbs Sampler

---

0. Initialize: set $\psi = \psi^{(0)}$ 1.
For $i = 1, ..., N$:

a. Draw $\theta_{0:T}^{(i)}$ from $\pi(\theta_{0:T} | y_{1:T}, \psi = \psi^{(i-1)})$ using FFBS.

b. Draw $\psi^{(i)}$ from $\pi(\psi | y_{1:T}, \theta_{0:T} = \theta_{0:T}^{(i)})$.

3.11 Application of Gibbs Sampler in Local Level Model

The **Random Walk plus Noise model or Local Level Model** is defined as: $\mu_0 \sim N(m_0, C_0)$

$$Y_t = \mu_t + v_t \qquad v_t \sim N(0, V) \tag{3.11.1}$$

$$\mu_t = \mu_{t-1} + w_t \qquad w_t \sim N(0, W) \tag{3.11.2}$$

where the error sequences (v_t) and (w_t) are not dependent both within and between them.

This Dynamic Linear Model as it stands now is a univariate model with V and W as the observation and system variances respectively.

The **inverse-gamma family** is chosen as the family of priors for the variances. Let $\psi_1 = V^{-1}$ and $\psi_2 = W^{-1}$ and assume that ψ_1 and ψ_2 are a priori independent, with

$$\psi_i \sim G(a_i, b_i), \qquad i = 1,2 \tag{3.11.3}$$

$a_i, b_i (i = 1,2)$ which are the parameters of the prior can be defined to fit the viewpoint of the analyst about the unknown precisions, which are expressed in terms of their means and variances. The Gibbs sampler will draw the states $\theta_{0:T}$ and the parameters ψ_1 and then ψ_2. These draws are generated random quantities from full conditional distributions. Thus, these quantities are drawn from distributions of the quantity of interest given all other random variables in the model, including the observations. FFBS algorithm can be used to generate $\theta_{0:T}$ by setting ψ_1 and ψ_2 to their most current simulated value. The basic approach relies on the fact that the full conditional distribution is proportional to the joint distribution of all the random variables involved. Considering ψ_1,

$$\pi(\psi_1|\psi_2,\theta_{0:T},y_{1:T}) \propto \pi(\psi_1,\psi_2,\theta_{0:T},y_{1:T}) \propto \pi(y_{1:T}|\theta_{0:T},\psi_1|\psi_2)\pi(\theta_{0:T}|\psi_1,\psi_2)\pi(\psi_1,\psi_2) \tag{3.11.4}$$

$$\propto \prod_{t=1}^{T} \pi(y_t|\theta_t, \psi_1) \prod_{t=1}^{T} \pi(\theta_t|\theta_{t-1}, \psi_2)\pi(\psi_1)\pi(\psi_2) \quad (3.11.5)$$

ψ_1 and ψ_2 are conditionally independent given $\theta_{0:T}$ and $y_{1:T}$.

$$\psi_1|\theta_{0:T}, y_{1:T} \sim G\left(a_1 + \frac{T}{2}, b_1 + \frac{1}{2}\sum_{t=1}^{T}(y_t - \theta_t)^2\right) \quad (3.11.6)$$

$$\psi_2|\theta_{0:T}, y_{1:T} \sim G\left(a_2 + \frac{T}{2}, b_2 + \frac{1}{2}\sum_{t=1}^{T}(\theta_t - \theta_{t-1})^2\right) \quad (3.11.7)$$

3.11.1 Constant Unknown Variances: the d-Inverse Gamma Prior

The time-invariant observation variance, V and system covariance, W assumption is a common decision being made. W is usually a diagonal matrix. Working with precisions φ, d-inverse-gamma prior assumes that

$$V_t = \phi_y^{-1}, \qquad W_t = diag(\phi_{\theta,1}^{-1}, ..., \phi_{\theta,p}^{-1}) \quad (3.11.8)$$

$\phi_y^{-1}, \phi_{\theta,1}^{-1}, ..., \phi_{\theta,p}^{-1}$ all have independent Gamma distributions. The product of $d = (p+1)$ Inverse-Gamma densities is the prior on the vector of the variances $\phi_y^{-1}, \phi_{\theta,1}^{-1}, ..., \phi_{\theta,p}^{-1}$.

The prior hyperparameters are fixed by expressing a prior guess on the unknown precisions, $E(\varphi_y) = a_y$ and $E(\varphi_{\theta,i}) = a_{\theta,i}$, with prior uncertainty given the the prior variances $Var(\theta_y) = b_y, Var(\varphi_{\theta,i}) = b_{\theta,i}$, where $i = 1,...,p$. Hence, the Gamma priors can be parameterized as

$$\varphi_y \sim G(\alpha_y, \beta_y), \qquad \varphi_{\theta,i} \sim G(\alpha_{\theta,i}, \beta_{\theta,i}) \qquad i = 1,...,p \quad (3.11.9)$$

with

$$\alpha_y = \frac{a_y^2}{b_y}, \qquad \beta_y = \frac{a_y}{b_y}, \qquad \alpha_{\theta,i} = \frac{a_{\theta,i}^2}{b_{\theta,i}}, \qquad \beta_{\theta,i} = \frac{a_{\theta,i}}{b_{\theta,i}}, \qquad i = 1, ..., p. \quad (3.11.10)$$

Given the observations $y_{1:T}$, the joint posterior of the unobservable states $\theta_{0:T}$ and the unknown parameters $\psi = (\phi_y, \phi_{\theta,1}, ..., \phi_{\theta,p})$ is proportional to the joint density

$$\pi(y_{1:T}, \theta_{0:T}, \psi) = \pi(y_{1:T} \mid \theta_{0:T}, \psi) \cdot \pi(\theta_{0:T} \mid \psi) \cdot \pi(\psi) \tag{3.11.11}$$

$$= \prod_{t=1}^{T} \pi(y_t|\theta_t, \phi_y) \cdot \prod_{t=1}^{T} \pi(\theta_t|\theta_{t-1}, \phi_{\theta,1}, ..., \phi_{\theta,p}) \cdot \pi(\theta_0) \cdot \pi(\phi_y) \cdot \prod_{i=1}^{p} \pi(\phi_{\theta,i}) \tag{3.11.12}$$

The sampling of the states will be done using FFBS. ϕ_y is a gamma distribution with full conditional distribution

$$\phi_y | \cdots \sim \mathcal{G}\left(\alpha_y + \frac{T}{2}, \beta_y + \frac{1}{2} SS_y\right) \tag{3.11.13}$$

with $SS_y = \sum_{t=1}^{T} (y_t - F\theta_t)^2$

$$\phi_{\theta,i} | \cdots \sim \mathcal{G}\left(\alpha_{\theta,i} + \frac{T}{2}, \beta_{\theta,i} + \frac{1}{2} SS_{\theta,i}\right) \tag{3.11.14}$$

with $SS_{\theta,i} = \sum_{t=1}^{T} \left(\theta_{t,i} - (G_t\theta_{t-1})_i\right)^2$.

3.12 Seemingly Unrelated Time Series Equations(SUTSE) Model

This model is used for multivariate time series. For the purpose of this research, each series is modeled based on the linear growth model. So each of the series will have its state vector comprising of the level and slope component and for simplicity, the system's error variance matrix is assumed to be diagonal. This implies that the evolution of level and slope is conditioned by independent random inputs. To be specific, two series $(m = 2)$ will be considered for the model. The state vector

$\theta_t = (\mu_{1,t}, \mu_{2,t}, \beta_{1,t}, \beta_{2,t})$ has the system equation

$$\begin{bmatrix} \mu_{1,t} \\ \mu_{2,t} \\ \beta_{1,t} \\ \beta_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_{1,t-1} \\ \mu_{2,t-1} \\ \beta_{1,t-1} \\ \beta_{2,t-1} \end{bmatrix} + \begin{bmatrix} w_{1,t} \\ w_{2,t} \\ w_{3,t} \\ w_{4,t} \end{bmatrix} \quad (3.12.1)$$

where $(w_{1,t}, w_{2,t}, w_{3,t}, w_{4,t})' \sim N(0, W)$ and

$$\begin{bmatrix} Y_{1,t} \\ Y_{2,t} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \theta_t + \begin{bmatrix} v_{1,t} \\ v_{2,t} \end{bmatrix} \quad (3.12.2)$$

represents the observation equation for the bivariate time series with $v_{1,t}, v_{2,t} \sim N(0, V)$, where $\mu_{i,t}$ – **the Local Level** $\beta_{i,t}$ – **the Local Growth Rate** for $i = 1,2$. So the SUTSE model can be written as

$$y_t = (F_t \otimes I_m)\theta_t + v_t \qquad v_t \sim N(0,V) \tag{3.12.3}$$

$$\theta_t = (G_t \otimes I_m)\theta_{t-1} + w_t \qquad w_t \sim N(0,W) \tag{3.12.4}$$

where $\otimes$ is the Kronecker product.

3.12.1 Gibbs Sampling for SUTSE model

Consider a SUTSE system where each series is described through a linear growth model. The precisions $\Phi_0 = V^{-1}, \Phi_1 = W_\mu^{-1}$ and $\Phi_2 = W_\beta^{-1}$ which are the prior distributions and are independent Wishart,namely $\Phi_j \sim W(v_j, S_j), j = 0,1,2$. The Wishart hyperparameters can be expressed as

$$v_0 = \frac{\delta_0 + m - 1}{2} = \frac{\delta_0 + 1}{2} \tag{3.12.5}$$

$$v_j = \frac{\delta_j + p_j - 1}{2} = \frac{\delta_j + 1}{2} \qquad j = 1,2 \tag{3.12.6}$$

$$S_0 = \frac{V_0}{2}, \qquad S_1 = \frac{W_{\mu,0}}{2} \qquad S_2 = \frac{W_{\beta,0}}{2}. \tag{3.12.7}$$

So, if $\delta_j > 2, j = 0,1,2$,then

$$E(V) = \frac{1}{\delta_0 - 2}V_0, \qquad E(W_\mu) = \frac{1}{\delta_1 - 2}W_{\mu,0} \qquad E(W_\beta) = \frac{1}{\delta_2 - 2}W_{\beta,0} \tag{3.12.8}$$

The parameters δ_i give an idea or relevant information about the uncertainty in the prior. From the full conditional distributions, we have

$$E(V \mid y_{1:T}, W_\mu, W_\beta, \theta_{0:t}) = \frac{\delta_0 - 2}{(\delta_0 + T) - 2}E(V) + \frac{T}{(\delta_0 + T) - 2}\frac{\sum_{t=1}^{T}(y_t - F_t\theta_t)(y_t - F_t\theta_t)'}{T} \tag{3.12.9}$$

So if δ_i is closer to 2 then a smaller weight of the prior is in the updating. Now with the Gibbs sampling from the joint posterior $\pi(\theta_{0:T}, \Phi_0, \Phi_1, \Phi_2 | y_{1:T})$ in turn generates the state vectors $\theta_{0:T}$ and the precision Φ_0, Φ_1, Φ_2. The full conditional of the precisions are:

$$\Phi_0 | \ldots \sim G\left(\frac{\delta_0 + 1 + T}{2}, \frac{1}{2}(V_0 + SS_y)\right) \tag{3.12.10}$$

$$\Phi_1 | \ldots \sim G\left(\frac{\delta_1 + 1 + T}{2}, \frac{1}{2}(W_{\mu,0} + SS_1)\right) \tag{3.12.11}$$

$$\Phi_2 | \ldots \sim G\left(\frac{\delta_2 + 1 + T}{2}, \frac{1}{2}(W_{\beta,0} + SS_2)\right) \tag{3.12.12}$$

where

$$SS_{t,i} = (\theta_{t,i} - G_{t,i}\theta_{t,i-1})(\theta_{t,i} - G_{t,i}\theta_{t,i-1})^{1} \tag{3.12.13}$$

and

$$SS_y = \sum_{t=1}^{T} (y_t - F_t\theta_t)(y_t - F_t\theta_t) \tag{3.12.14}$$

3.13 Remark

In this chapter, the adopted model, SUTSE model was discussed together with the forward filtering backward sampling Gibbs Sampler Algorithm which will be used in estimating the unknown parameters and forecasting the states and observations in the next chapter.

Chapter 4

Simulation, Analysis and Discussion of Results

4.1 Introduction

This chapter discusses the descriptives of the data used for the research and also discusses the outputs for the data analysis. One major assumption made is that the supply or energy generation is equal to the electricity demand.

4.2 Descriptives

There were 1644 observations each from the daily peak demand data and the daily electricity generation from February 2,2014 to August 2,2018 collected from the Ghana Grid Company Limited. The minimum peak demand value recorded over this period was 1228MW and a maximum value of 2433MW. The electricity generation data recorded a maximum value of 50832MW and a minimum value of 15552MW. The mean values of the peak demand and the generation are 1815.881MW and 35917.48MW respectively. The standard deviations in the peak and generation data recorded were 228.9MW and 4569.109MW respectively. Also, the Peak demand had a coefficient of variation of 12.6% while the electricity demand had a coefficient of 12.7%. This implies that the variations in both datasets relative to their means are approximately the same.

4.3 Time Series Plot of Peak Demand and Electricity Demand

Figure 4.1 shows the peak demand which demonstrates a high level of noise in the system. It also shows a decreasing trend through to the middle of 2015 and started increasing from that point.

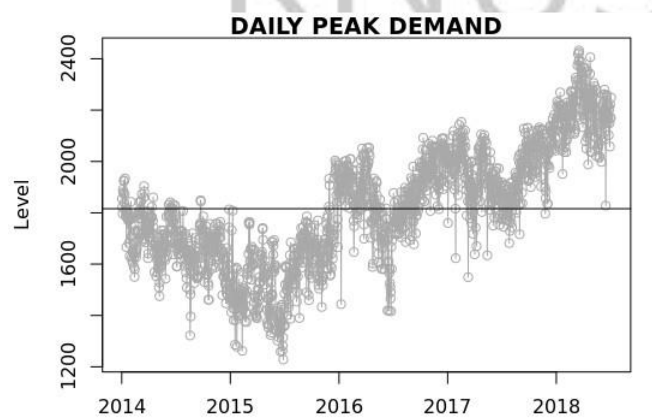


Figure 4.1: Daily Peak Demand Plot

Figure 4.2 shows the daily total energy generation which also demonstrates a high level of noise in the system. There are also few data points which are likely to be outliers which were as a result of power shortage challenges in the country. Figure 4.2 also shows slight decreasing trend through to mid 2015 and started increasing again from the latter part of 2017.

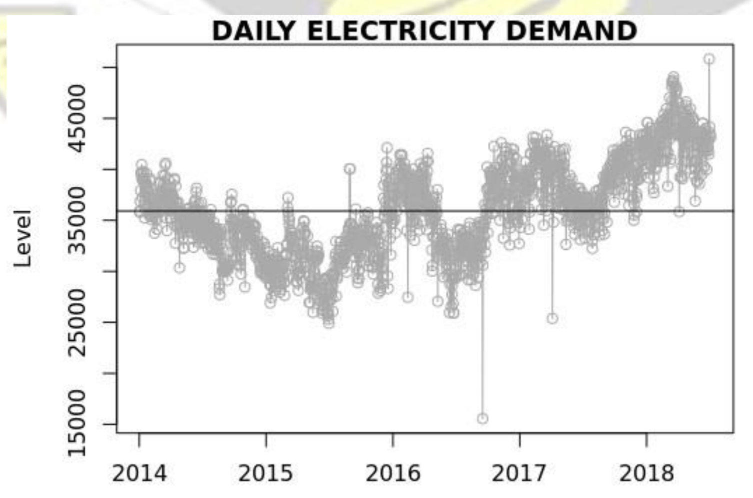


Figure 4.2: Daily Electricity Demand Plot

4.4 Markov Chain Monte Carlo Simulation for Variance Estimation

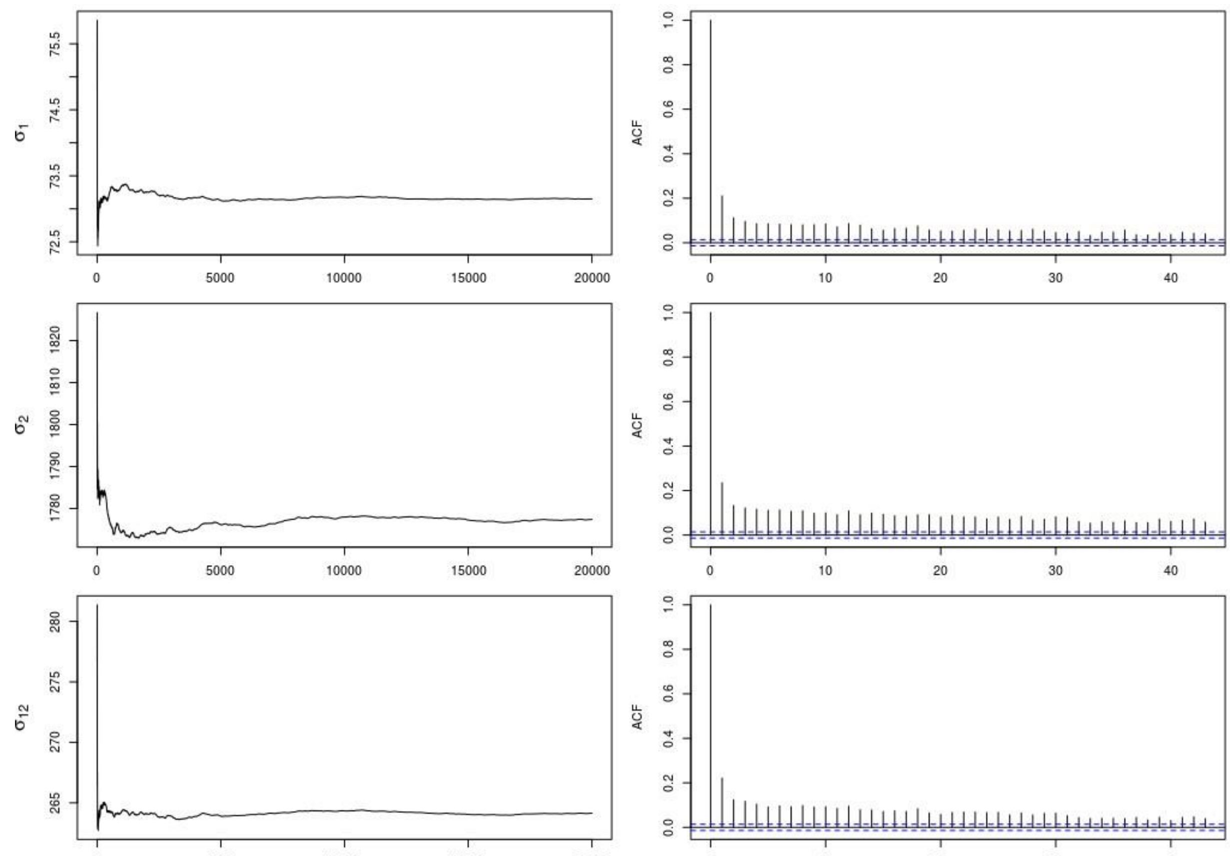


Figure 4.3: MCMC ergodic means and autocorrelation for the covariance matrix V

A few diagnostic plots obtained from the MCMC output for the covariance V is shown in Figure 4.3. It shows the running ergodic means of the parameters and the estimated autocorrelation plots. The ACF plots in Figure 4.3 also demonstrates how weak the correlations between the days grow when the days are far apart. Since the simulation converges, the posterior means or the ergodic means for the observation covariance matrix were estimated as shown in equation (4.4.1). Also, the deviations in (parenthesis) involved in the Monte Carlo were estimated. Thus,

$$E(V|y_{1:T}) = \begin{pmatrix} 5352.923 & 69829.474 \\ (2.5078) & (52.566) \\ 69829.474 & 3160403.255 \\ (52.566) & (1679.427) \end{pmatrix} \qquad (4.4.1)$$

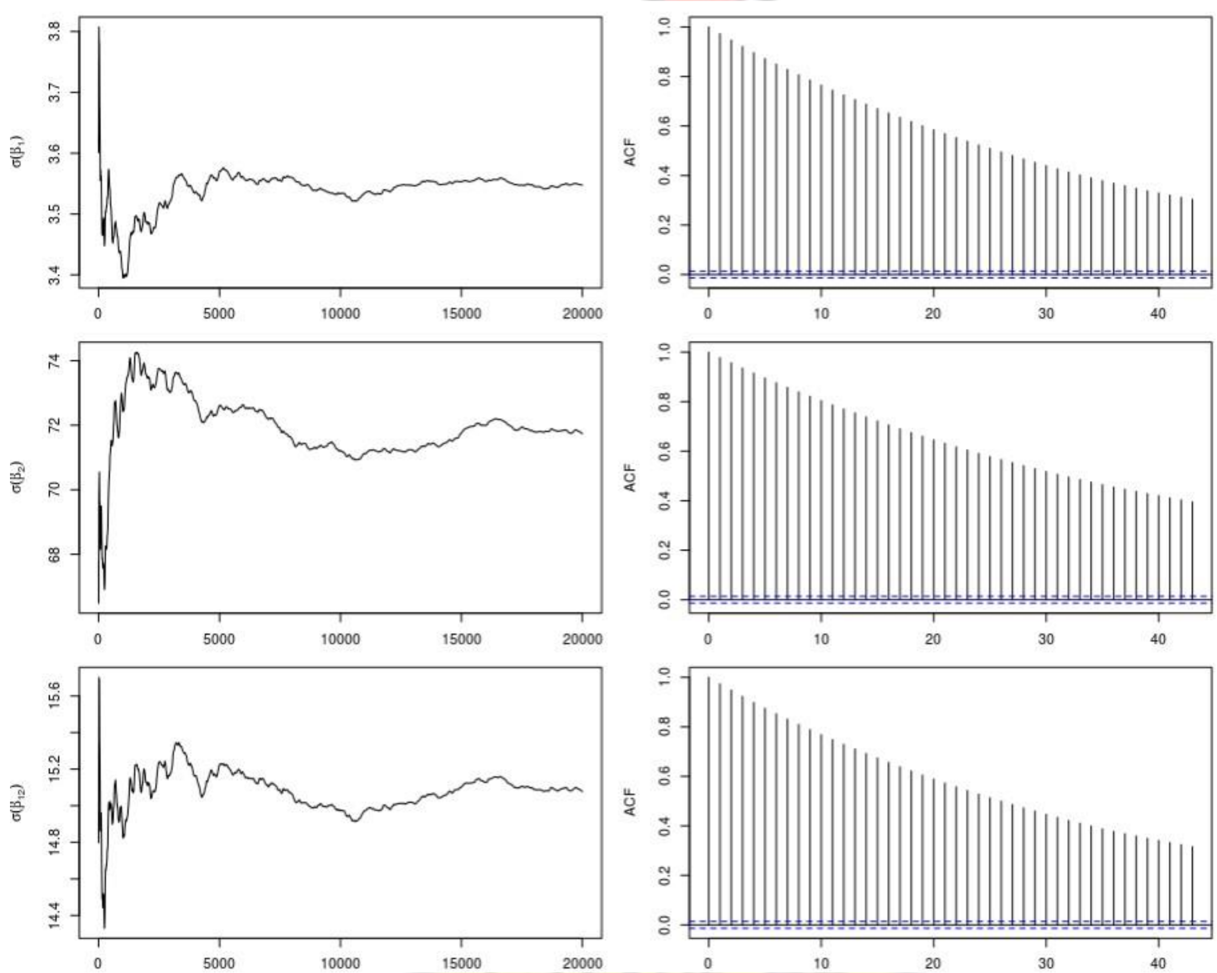


Figure 4.4: MCMC ergodic means and its corresponding autocorrelation for the covariance matrix W_β

Figure 4.4 also shows similar diagnostic plots obtained from the MCMC output for covariance matrix in trend component of the system. It also shows the running ergodic means of the parameters and the estimated autocorrelation plots. The ACF plots in Figure 4.4 demonstrates how weak the correlations between the days grows when the days are far apart from each other. The posterior means or ergodic means for the trend component in the system together with the Monte Carlo deviations can be estimated from the MCMC output due to convergence of the Monte Carlo

process. Thus,

$$E(W_\beta|y_{1:T}) = \begin{bmatrix} \underset{(0.1657)}{12.730} & \underset{(3.211)}{230.371} \\ \underset{(3.211)}{230.371} & \underset{(82.454)}{5220.107} \end{bmatrix} \qquad (4.4.2)$$

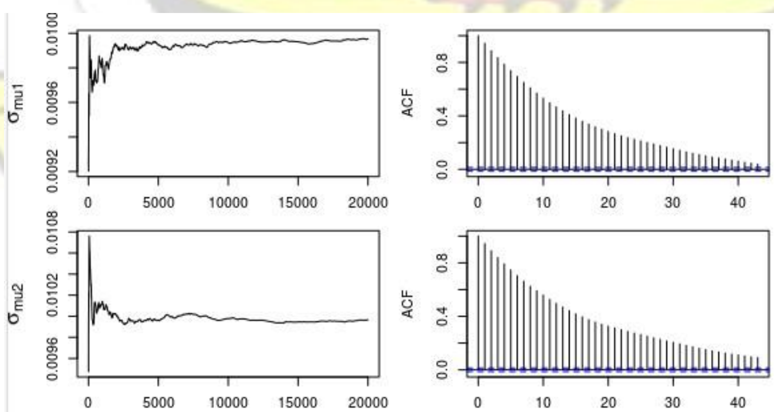


Figure 4.5: MCMC ergodic means and its corresponding autocorrelation for the covariance matrix W_μ

As shown in Figures 4.3 and 4.4, the ACF plots in Figure 4.5 also show how weak the correlations between the days grows when the days are far apart from each other. Since Seemingly Unrelated Time Series Equations Model involves integrated random walks, the Ergodic means or posterior means were expected to approach zero as shown in matrix equation (4.4.3).

$$E(W_\mu | y_{1:T}) = \begin{bmatrix} 9.982080e-05 & 5.173045e-07 \\ (5.288852e-07) & (5.288852e-07) \\ 5.173045e-07 & 9.984811e-05 \\ (5.288852e-07) & (6.293606e-07) \end{bmatrix} \quad (4.4.3)$$

4.5 Filtering Density and One-Step-Ahead Forecasting Plots for the Peak Demand

Figure 4.6 shows the estimated means of filtering distribution of the states of the system. Thus, $E(\theta_t|y_{1:t})$. It shows that the unobserved state distribution given the observations of the system which has a relation with the Kalman gain which tells what information to dwell upon most in the one-step-ahead forecast of the observations.

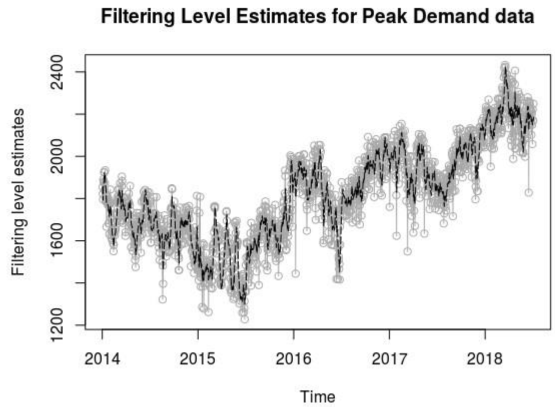


Figure 4.6: Filtering Density for the Daily Peak Demand

Figure 4.7 shows the one-step-ahead predictive distribution for the observations. The MAPE value involved in this forecasting was 0.03696 which is 3.696% error in the forecast.

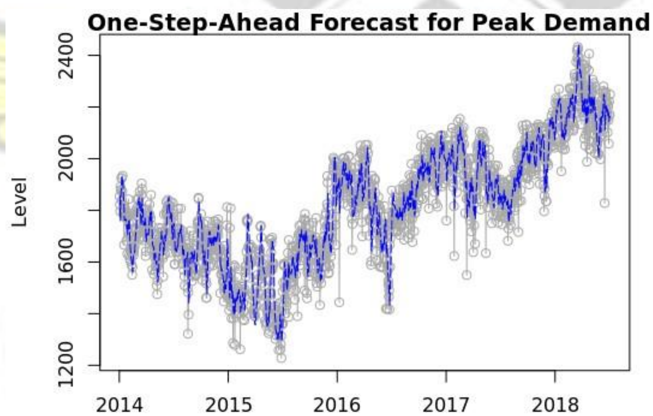


Figure 4.7: One-Step-Ahead Forecast for Daily Peak Demand

4.6 Filtering Density and One-Step-Ahead Forecasting Plots for the Electricity Demand

The unobserved states density for the daily total generation resulting from the filtering process is shown in Figure 4.8.

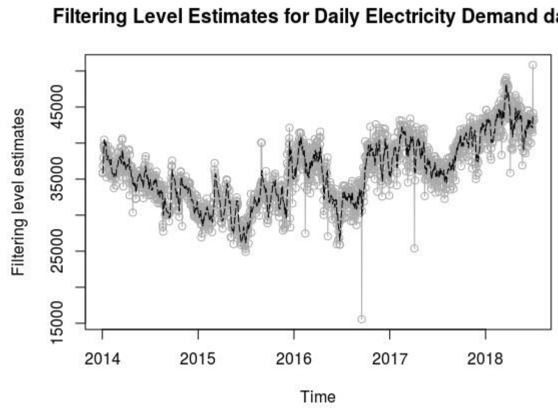


Figure 4.8: Filtering Density for the Daily Total Generation

Figure 4.9 also displays the one-step-ahead predictive distribution for the observations from the total energy generation. The MAPE value involved in this forecasting was about 0.04235 which is 4.235% error in the forecast.

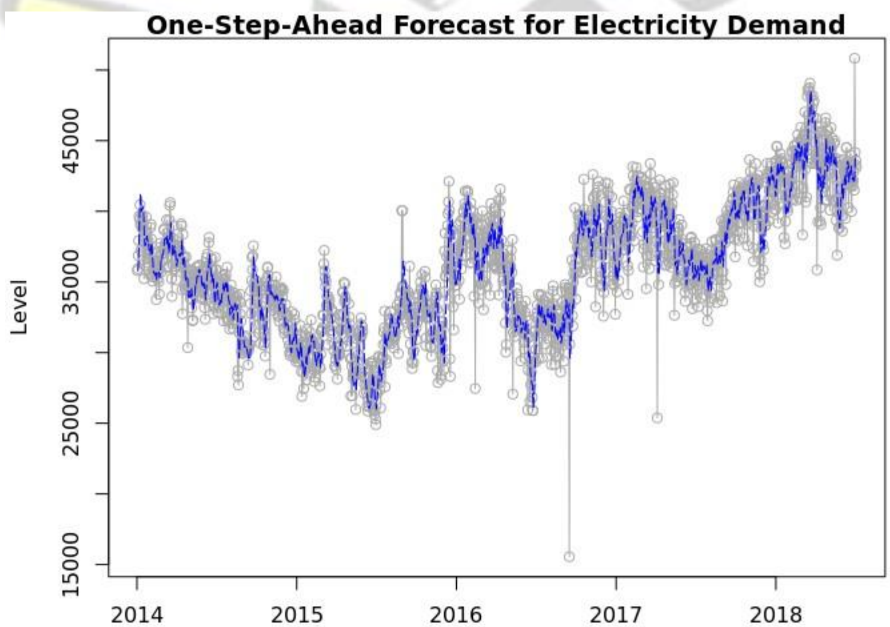


Figure 4.9: One-Step-Ahead Forecast for the Daily Electricity Demand

4.7 Forecasting

Table 4.1: Forecast of Peak demand and Electricity Demand or generation

	3rd August,2018
Peak Demand	2187MW
Electricity Demand	44090MW

4.8 Remark

This chapter discuss some basic descriptives from the data used and also made some inferences based on the MCMC method used. It went on to discuss the one-stepahead forecast and measures the accuracy of the adopted method in the previous chapter which turned out to be very efficient. The forecast value for the next day,3rd August,2018 was given in this chapter. In the next chapter, conclusions and recommendations will be made based on what was observed in this chapter.

Chapter 5

CONCLUSION and RECOMMENDATIONS

5.1 Introduction

The main purpose for this study was to forecast the peak demand and total electricity generation or demand. In view of this, the data collected was visualized and a summary of the data was taken to understand the data. There were two series used for the study;namely,

1. Daily peak demand data

2. Daily total electricity or energy generation which also referred to as electricity demand data.

1644 observations each were used for this study. The minimum peak demand value recorded was 1228MW and the maximum value recorded was 2433MW. For the electricity generation, the lowest generation recorded was 15552MW and the maximum was 50832MW. The mean peak demand and electricity demand were 1816MW and 35917MW respectively. Also, the Peak demand had a coefficient of variation of 12.6% while the electricity demand had a coefficient of 12.7%. This implies that the variations in both datasets relative to their means are almost the same.

5.2 Summary of Results

The adopted model, Seemingly Unrelated Time Series equation was able to capture the various aspects of the electricity transmission including the losses involved through the process of electricity production which plays a major role in this sector. The Forward Filtering Backward Sampling was used in the Gibbs Sampler, a Markov Chain Monte Carlo Process method to estimate the posterior means of the unknown variances with 50,000 simulations. The ACF plots showed weakness in correlation between days far apart from each other. Due to the convergence of the simulation, it was legit to estimate the posterior mean of the variances from the MCMC process and the deviations involved in the estimation. The covariance for the state matrix showed less deviations in the estimated values but the covariance for the observation equations projected a larger variance between the second series with itself of 3160404.255 and a higher deviation of 1679.427. The nonzero covariance value in the observation covariance matrix showed that the Peak demand and the Electricity demand are correlated.

The forward filtering which is estimated by Kalman filter was used to generate the state distribution, thus $\pi(\theta_t|y_{1:t})$ which portrayed the actual state of the system. The one-step-ahead forecast was used to test the performance of the model in estimating the true observations. To confirm this, the Mean Absolute Percentage Error (MAPE) was used to measure the accuracy of the model.An error of about 3.696% was estimated from the peak demand one-step-ahead forecast. With the energy or electricity demand, the MAPE recorded was about 4.235%. The predicted peak demand for the next day,3rd August,2018 is 2187MW and that of the electricity demand is 44090MW.

5.3 Conclusion

A Dynamic Linear State Space Model, specifically the Seemingly Unrelated Time Series (SUTSE) model was adopted for this study due to its ability to measure the losses involved in the electricity production. The model captures the random errors and the systematic errors or losses using probability distributions and the local trends in the system.

The model showed a high level of accuracy by being able to predict about 96% accuracy of the daily peak demand and about 95% accuracy of the daily electricity demand.

Furthermore, the model together with the Forward Filtering Backward Sampling (FFBS) Gibbs Sampler method made it possible for forecasting the next day's peak demand and total electricity to be generated.

5.4 Recommendation for Further Studies

1. The model does not capture the seasonality component in the system.It is recommended that future work incorporate an equation to capture the seasonal component.

2. The forward filtering of the algorithm used the linear Kalman filter which might fail if the system is highly nonlinear. Based on this, the extended Kalman filter or the unscented Kalman filter which is able to capture the nonlinearity is recommended.

3. The Gibbs sampler is very powerful in estimating posterior distribution of parameters but if there is a new observation made then the sampler needs to go through the whole simulation process again.To resolve this, it is recommended to apply the sequential Monte Carlo Methods.

REFERENCES

Abdel-Aal, R. and Al-Garni, A. (1997), "Forecasting monthly electric energy consumption in eastern Saudi Arabia using univariate time-series analysis," *Energy*, 22, 1059–1069.

Al-Musaylh, M. S., Deo, R. C., Adamowski, J. F., and Li, Y. (2018), "Short-term electricity demand forecasting with MARS, SVR and ARIMA models using aggregated demand data in Queensland, Australia," *Advanced Engineering Informatics*, 35, 1–16.

Alfares, H. K. and Nazeeruddin, M. (2002), "Electric load forecasting: literature survey and classification of methods," *International journal of systems science*, 33, 23–34.

Altinay, G. and Karagol, E. (2005), "Electricity consumption and economic growth: evidence from Turkey," *Energy economics*, 27, 849–856.

Azadeh, A., Saberi, M., and Seraj, O. (2010), "An integrated fuzzy regression algorithm for energy consumption estimation with non-stationary data: a case study of Iran," *Energy*, 35, 2351–2366.

Bagnasco, A., Fresi, F., Saviozzi, M., Silvestro, F., and Vinci, A. (2015), "Electrical consumption forecasting in hospital facilities: An application case," *Energy and Buildings*, 103, 261–270.

Bhattacharyya, S. C. and Thanh, L. T. (2004), "Short-term electric load forecasting using an artificial neural network: case of Northern Vietnam," *International journal of energy research*, 28, 463–472.

Bianco, V., Manca, O., and Nardini, S. (2009), "Electricity consumption forecasting in Italy using linear regression models," *Energy*, 34, 1413–1421.

Borunda, M., Jaramillo, O., Reyes, A., and Ibargu¨engoytia, P. H. (2016), "Bayesian networks in renewable energy systems: A bibliographical survey," *Renewable and Sustainable Energy Reviews*, 62, 32–45.

Box, G., Jenkins, G., and Reinsel, G. (1995), "Time series analysis: forecasting and control. 2008," *Google Scholar*.

Cecati, C., Kolbusz, J., R´oz˙ycki, P., Siano, P., and Wilamowski, B. M. (2015), "A novel RBF training algorithm for short-term electric load forecasting and comparative studies," *IEEE Transactions on industrial Electronics*, 62, 6519–6529.

Chen, B.-J., Chang, M.-W., et al. (2004), “Load forecasting using support vector machines: A study on EUNITE competition 2001,” *IEEE transactions on power systems*, 19, 1821–1830.

De Felice, M., Alessandri, A., and Catalano, F. (2015), “Seasonal climate forecasts for medium-term electricity demand forecasting,” *Applied Energy*, 137, 435–444.

Dwijayanti, S. and Hagan, M. (2013), “Short term load forecasting using a neural network based time series approach,” in *Artificial Intelligence, Modelling and Simulation (AIMS), 2013 1st International Conference on*, pp. 17–22, IEEE.

Feinberg, E. A. and Genethliou, D. (2005), “Load forecasting,” in *Applied mathematics for restructured electric power systems*, pp. 269–285, Springer.

Gilks, W. R., Richardson, S., and Spiegelhalter, D. J. (1996), “Introducing markov chain monte carlo,” *Markov chain Monte Carlo in practice*, 1, 19.

Goia, A., May, C., and Fusai, G. (2010), “Functional clustering and linear regression for peak load forecasting,” *International Journal of Forecasting*, 26, 700–711.

Grant, J., Eltoukhy, M., and Asfour, S. (2014), “Short-term electrical peak demand forecasting in a large government building using artificial neural networks,” *Energies*, 7, 1935–1953.

Gyamfi, S., Diawuo, F. A., Kumi, E. N., Sika, F., and Modjinou, M. (2018), “The energy efficiency situation in Ghana,” *Renewable and Sustainable Energy Reviews*, 82, 1415–1423.

Hagan, M. (1977), “Identification techniques of Box and Jenkins applied to the problem of short-term load forecasting,” *IEEE PES SM, Mexico City, Mexico (1977-7)*.

Hubele, N. and Cheng, C.-S. (1990), “Identification of seasonal short-term load forecasting models using statistical decision functions,” *IEEE Transactions on Power Systems*, 5, 40–45.

Ibargu¨engoytia, P. H., Reyes, A., Romero-Leon, I., Pech, D., Garc´ıa, U. A., Sucar, L. E., and Morales, E. F. (2014), “Wind power forecasting using dynamic Bayesian models,” in *Mexican International Conference on Artificial Intelligence*, pp. 184–197, Springer.

Kavaklioglu, K. (2011), “Modeling and prediction of Turkeys electricity consumption using Support Vector Regression,” *Applied Energy*, 88, 368–375.

Kucukali, S. and Baris, K. (2010), “Turkeys short-term gross annual electricity demand forecast by fuzzy logic approach,” *Energy policy*, 38, 2438–2445.

Kyriakides, E. and Polycarpou, M. (2007), “Short term electric load forecasting: A tutorial,” in *Trends in Neural Computation*, pp. 391–418, Springer.

Li, X. H. and Hong, S. H. (2014), “User-expected price-based demand response algorithm for a home-to-grid system,” *Energy*, 64, 437–449.

Liu, M., Shi, Y., and Fang, F. (2015), “Load forecasting and operation strategy design for CCHP systems using forecasted loads,” *IEEE Transactions on Control Systems Technology*, 23, 1672–1684.

Lo, K. and Wu, Y. (2003), “Risk assessment due to local demand forecast uncertainty in the competitive supply industry,” *IEE Proceedings-Generation, Transmission and Distribution*, 150, 573–581.

Matijaˇs, M. (2013), “Electric Load Forecasting using Multivariate Meta-learning,” Ph.D. thesis, Fakultet elektrotehnike i raˇcunarstva, Sveuˇciliˇste u Zagrebu.

Mirlatifi, A. M. (2016), “Electricity Peak Demand Forecasting for Developing Countries,” .

Mishra, S. (2009), “Short Term Load Forecasting Using Computational Intelligence Methods,” Ph.D. thesis.

Mohandes, M. (2002), “Support vector machines for short-term electrical load forecasting,” *International Journal of Energy Research*, 26, 335–345.

Pai, P.-F. and Hong, W.-C. (2005), “Support vector machines with simulated annealing algorithms in electricity load forecasting,” *Energy Conversion and Management*, 46, 2669–2688.

Pappas, S. S., Ekonomou, L., Karamousantas, D. C., Chatzarakis, G., Katsikas, S., and Liatsis, P. (2008), “Electricity demand loads modeling using AutoRegressive Moving Average (ARMA) models,” *Energy*, 33, 1353–1360.

Pereira, C. M., de Almeida, N. N., and Velloso, M. L. (2015), “Fuzzy modeling to forecast an electric load time series,” *Procedia Computer Science*, 55, 395–404.

Petris, G., Petrone, S., and Campagnoli, P. (2009), “Dynamic linear models,” in *Dynamic Linear Models with R*, pp. 31–84, Springer.

Powell, K. M., Sriprasad, A., Cole, W. J., and Edgar, T. F. (2014), “Heating, cooling, and electrical load forecasting for a large-scale district energy system,” *Energy*, 74, 877–885.

Ramedani, Z., Omid, M., Keyhani, A., Khoshnevisan, B., and Saboohi, H. (2014), “A comparative study between fuzzy linear regression and support vector regression for global solar radiation prediction in Iran,” *Solar Energy*, 109, 135–143.

Raza, M. Q. and Khosravi, A. (2015), "A review on artificial intelligence based load demand forecasting techniques for smart grid and buildings," *Renewable and Sustainable Energy Reviews*, 50, 1352–1372.

Tai, H. F. (2009), "Case study of short-term electricity load forecasting with temperature dependency," Ph.D. thesis, Universiti Malaysia Pahang.

Taylor, J. W., De Menezes, L. M., and McSharry, P. E. (2006), "A comparison of univariate methods for forecasting electricity demand up to a day ahead," *International Journal of Forecasting*, 22, 1–16.

Yang, J. (2006), "Power system short-term load forecasting," *Elektrotechnik und Informationstechnik der Technischen Universitat: Thesis for Ph. d degree.–Germany, Darmstadt*.

Yildirim, I. (2012), "Bayesian inference: Gibbs sampling," *Technical Note, University of Rochester*.

Zadeh, L. A. (1965), "Information and control," *Fuzzy sets*, 8, 338–353.