UNIVERSITY OF PARDUBICE

FACULTY OF ELECTRICAL ENGINEERING
AND INFORMATICS

DOCTORAL THESIS

2018 Obiora Sam Ezeora

University of Pardubice

Faculty of Electrical Engineering and Informatics

Dynamic Stochastic Modeling Methods for Optimization of Environmental Measurements

Obiora Sam Ezeora

Doctoral Thesis

2018

I hereby declare that I have written the Doctoral thesis on my own.

All the literary sources and the information used are listed in the bibliography.

I was familiar with the fact that rights and obligations arising from the Act No. 121/2000 Coll., the Copyright Act, apply to my thesis, especially the fact that the University of Pardubice has the right to enter into a license agreement for use of the thesis as a school work pursuant to § 60, Section 1 of the Copyright Act, and the fact that should this thesis be used by me or should a license be granted for the use to another entity, the University of Pardubice is authorized to claim a reasonable fee to cover the costs incurred during the making of the thesis, up to the real amount thereof.

I agree with the reference-only disclosure of my thesis in the University Library.

In Pardubice on January 11, 2018

Obiora Sam Ezeora

ACKNOWLEDGEMENTS

I use this opportunity to recognize and thank those who have contributed towards the success of this work. I first thank my supervisors – prof. Ing. Petr Musilek, Ph.D., and Mgr. Jana Heckenbergerová, Ph.D. I appreciate all their support and guidance.

I am grateful to my wife and children for their patience and understanding throughout the course of this programme. I also acknowledge the contribution of thesis' reviewers and members of the doctoral examination committee.

I would also like to express my thanks to the science and research officer of the faculty, Ms Ilona Kučerová for her professional excellence and support. My fellow students and other staff at the faculty are also remembered and acknowledged.

ABSTRACT

Minimization of energy consumption of environmental measurement systems is important to ensure their extended operational lifetime and low maintenance cost. This needs to be realized without sacrificing on data quality. One possible way to achieving this is the use of energy-aware sampling techniques such as adaptive and event-triggered sampling. In this work, new methods based on these sampling techniques have been developed. The first method produces stochastic models that accurately predict missed and future data with minimal energy. The method also determines the optimal sampling interval. The second method utilizes new type of event-triggered mechanism that adjusts sampling interval so that it adapts to the changes in measurement data. Algorithms have been developed and all methods demonstrated using field data. Obtained results have been thoroughly analyzed from the perspective of approximation error and energy savings. Models have been validated and favorable results obtained. High R-squared values and low values of mean square normalized error have been obtained. Battery lifetime is extended by more than 87% when sampling interval increases from 15 to 30 seconds. Furthermore, about 45% daily savings of energy consumption of analog-to-digital converter has been achieved in a case study analysis involving the new algorithm, an ADC and field data.

KEYWORDS

Time series, sampling interval, environmental variables, stochastic, Box-Jenkins, sensor, energy consumption, data quality

TITLE

Dynamic Stochastic Modeling Methods for Optimization of Environmental Measurements

CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND SYMBOLS

Abbreviations and Acronyms	Meaning
ADC	Analog-to-Digital Converter
SAR	Successive Approximation Register
DAC	Digital-to-Analog Converter
DAS	Data Acquisition System
FPGA	Field Programmable Gate Array
ISM	Industrial-Scientific-Medical
RF	Radio Frequency
WSN	Wireless Sensor Network
TCP	Transmission Control Protocol
IP	Internet Protocol
PAR	Photosynthetically Active Radiation
ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal Autoregressive Integrated Moving Average
FARIMA	Fractional Autoregressive Integrated Moving Average
FGM	Fractional Gaussian Model
RNN	Recurrent Neural Network
HMM	Hidden Markov Model
DMCM	Discrete-time Markov Chain Model
CTMC	Continuous-time Markov Chain Model
SDE	Stochastic Differential Equation Model
SVM	Support Vector Machine
AAM	Analytical Approach Model
ACF	Autocorrelation Function
PACF	Partial Autocorrelation Function
ANOVA	Analysis of Variance
FFT	Fast Fourier Transform
AIC	Akaike Information Criterion
MLP	Multilayer Perceptron
RNN	Recurrent Neural Network
MSE	Mean Square normalized Error
ECG	Electrocardiogram
PhD	Doctor of Philosophy
EU	European Union
US	United States of America
SNR	Signal to Noise Ratio
SINAD	Signal to Noise and Distortion Ratio
ENOB	Effective Number of Bits
FOM	Figure of Merit
Cov	Covariance
Var	Variance
iid	Independent Identically Distributed

WN	White Noise
ARCH	Autoregressive Conditional Heteroskedasticity
GARCH	Generalized Autoregressive Conditional Heteroskedasticity
EGARCH	Exponential Generalized Autoregressive Conditional Heteroscedasticity
NAR	Non-linear Autoregressive
PPM	Poisson Process Model
NMA	Non-linear Moving Average
MLE	Maximum Likelihood Estimation
AIC	Akaike Information Criterion
AICC	Corrected Akaike Information Criterion
BIC	Bayesian Information Criterion
MPE	Maximum Permissible Error
PAR	Photosynthetically Active Radiation
nm	Nanometer
CI	Confidence Interval
LL	Likelihood
MLE	Maximum Likelihood Estimation

TERMINOLOGY

Observations	Set of values obtained when an environmental variable is measured.
Process	Series of physical, chemical and biological activities involved in measurement of environmental variable.
Variable	Physical, chemical or biological property of the environment that is observed (measured) over time at a given location. Observed values may be discrete or continuous. An example is air temperature.
Environmental measurement:	Comprises of monitoring and evaluation of environmental variable. During monitoring, variable of interest is observed. During evaluation, observed value is processed to give a final settled value within given tolerance band.
Sampling	Act of reading and collecting data from the environment using sensors.
Sampling interval	Interchangeable with sampling time interval. Represents time interval separating successive observations.
Sensing time	Time taken for sensor to measure and attain a final settled value for the variable measured. Comprises of measurement time, start-up time and response time. Measurement time is negligibly small compared to other components. Therefore, a sensor that is already in operation has a fairly constant sensing time.
Duty-cycle	Ratio of length of time sensor spends in active mode to sampling interval.
Stochastic process	Mathematical relation that describes probability structure of time series; implying that a set of values (observations) is sample realization of the stochastic process that produced it.
White noise	Sequence of uncorrelated values with mean of zero and finite variance.
Measurement error	Difference between measured value of variable and its true value.
Residual	Difference between measured value and its model-predicted value.

CHAPTER ONE
INTRODUCTION

1.1 Background Information

There is huge focus on energy management of self-powered electronic systems such as those used in environmental and health measurement systems. This is due to need for extended battery lifetime and low life-cycle cost requirement of measurement systems. Therefore energy is required to be managed. Energy management is classified based on whether additional energy is supplied to systems or whether available energy is conserved and efficiently utilized. The former is termed *energy supplying* while the latter is termed *energy conservation* [1].

Energy conservation approach is however, most favored and widely adopted as better energy management approach. This is because energy supplying approach is faced with environmental or security of supply challenges. First is doubt if sufficient energy supply from renewable energy sources would be available to meet current energy supply needs by 2020. This doubt is caused by belief that European Union (EU) 2020 target of having 20% total energy production from renewable energy sources may not be achieved [2]. This is a threat to security of supply. Second is the environmental challenge facing energy supply through conventional energy sources. For above reasons, it becomes imperative that available energy be efficiently utilized. Based on methods used in realizing energy savings, energy conservation is classified into three main parts: data-oriented, node-oriented and network-oriented methods [1, 3]. Taxonomy of energy management technique is shown below.

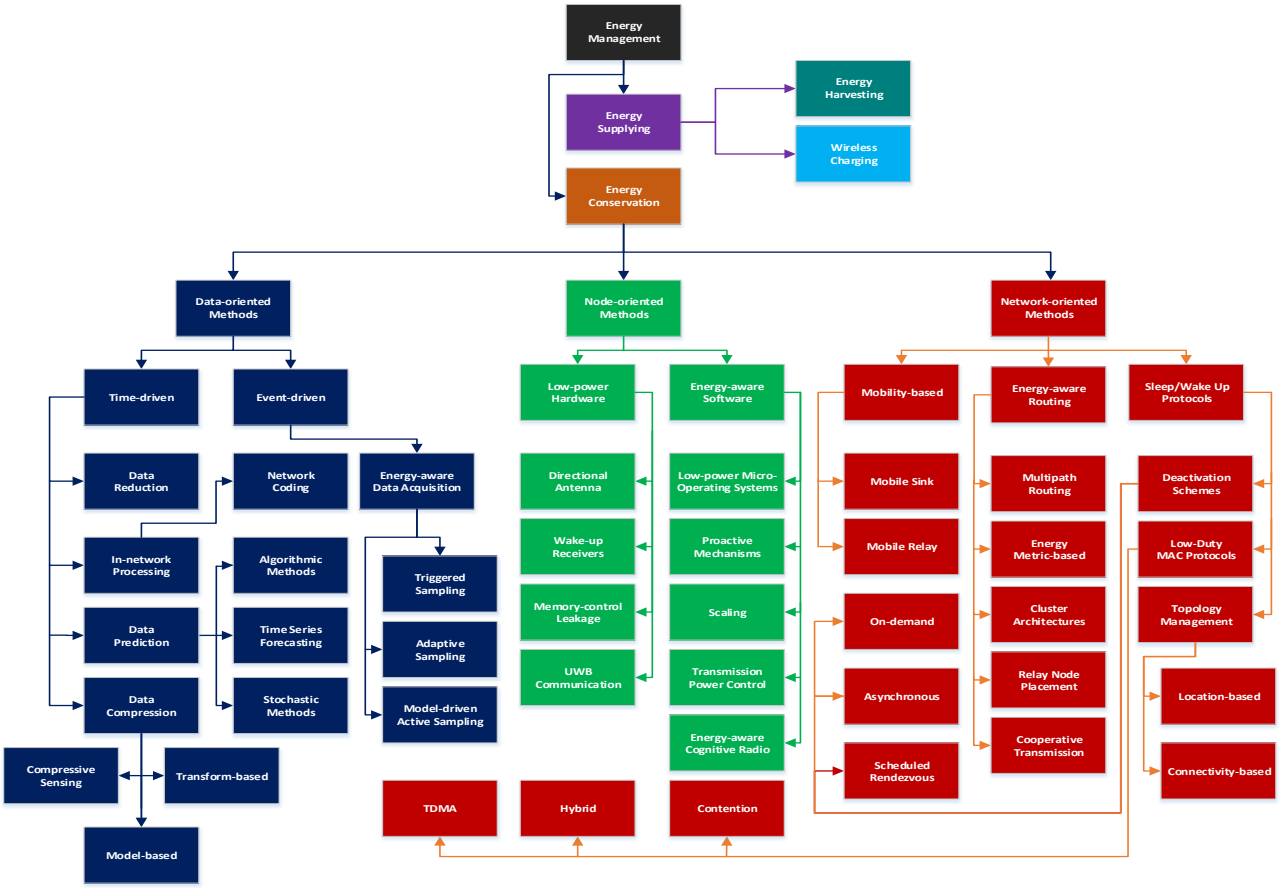


Figure 1: Taxonomy of energy management technique [1, 3]

This work focuses on measurement optimization of environmental variables to ensure that variables are measured with minimal energy without significant loss of data. One way of achieving this is the use event-driven sampling methods such as adaptive and event-triggered sampling. These methods are energy-aware data acquisition methods. They work to minimize energy consumption while sampling. This would be seen in figure 1. Event-triggered sampling exercises control in sampling by ensuring that sensor processes and transmits data whenever significant change in data occurs. Adaptive sampling on the other hand reduces number of samplings by adjusting sampling frequency. This allows sampling at low frequencies.

When sampling with low frequency, accurate stochastic models are determined so that missed and future values are estimated with sufficient accuracy. It is also required that models be made simple. They should operate with simple software programs that minimize energy consumption of measuring device without the occurrence of aliasing. Consequently, an optimal sampling interval is required.

Aliasing is a major limitation of discrete-time sampled variables. It is said to have occurred whenever there is loss of significant amount of data. Loss of significant amount of data causes difficulty in reconstructing the original signal. It also causes loss of vital data. The latter effect is more important in this work since measured data are processed and used as digital input to other microelectronic devices. It is not intended to be reconstructed to its original analogue signal.

Therefore when variable sampled with low sampling frequency is interpolated for missed values, model used should give values that are close to those which would have been obtained if a higher sampling frequency were used. Thus, knowledge of appropriate stochastic model and optimal sampling interval are needed while sampling at low frequency. They help ensure that trade-off between aliasing and energy consumption is balanced.

1.2 Description of Environmental Variables

Environmental measurement involves assessment of physical, biological or chemical factors that affect health and quality of life within the human environment. It comprises of environmental monitoring and evaluation. During monitoring, the environmental variable or factor of interest is observed. During evaluation, the observed (or monitored) variable is processed in order to obtain a final settled value that is within a tolerance band.

While performing environmental measurements, disturbances may be imposed on the environment. Consequently, an environmental measurement may be destructive or non-destructive. Destructive measurement results in samples being physically removed from the environment. Non-destructive measurement which is also known as non-invasive measurement does not involve removal of samples from environment. Due to advances in sensors and transducers technologies, non-destructive measurements are becoming popular. There are two main techniques employed in non-destructive measurements. They are remote sensing and liquid-solid (or gas-solid) sensing [4].

Remote sensing uses earth's electromagnetic waves to perform measurement while liquid-solid or gas-solid sensing convert changes in parameters at interface to electrical response. Liquid-solid or gas-solid sensing is commonly used in environmental measurement systems where variable of interest is measured and displayed, or stored locally within the system. It is

also used in measurement systems enabled for wireless sensor network and in ubiquitous systems.

In this work, environmental variables that are of interest are - air temperature, photosynthetically active radiation (PAR), relative humidity and soil moisture. Air temperature is the action of individual air molecules at a certain time in a given location. It relates to the average kinetic energy of moving air molecules at a given time in a particular location. Small-scale actions of individual air molecules usually describe air temperature. This emanates from the kinetic theory of gases. On the other hand, air temperature description for large-scale effects of larger number of air molecules is based on thermodynamic principles. It follows three fundamental principles of thermodynamics; namely:

1. The temperature of an object can affect some properties of the object such as length, AC electrical resistance of the object, etc.
2. Two objects that have the same temperature are in thermodynamic equilibrium.
3. Two objects that have different temperatures, if in contact with each other will eventually attain thermodynamic equilibrium.

Photosynthetically active radiation (PAR) on the other hand is an electromagnetic radiation whose wavelength between is 400 and 700 nm [5]. It represents the amount of light available for photosynthesis. It is essential since it is required by plants for photosynthesis. This is among the reasons it is been measured. PAR values can vary from zero (mostly during night time) to its maximum value during daytime. This variation is usually significant during summer. This is why PAR values are usually logarithmic-transformed during data analysis. It helps stabilize variation in variance that occurs when daytime and nighttime PAR values are considered as dataset.

On the other hand, relative humidity is the ratio of water vapour density to saturated water vapour density of air at a given temperature. This ratio is usually expressed in percent. It is worth noting that vapour density is the same as vapour mass per unit volume. It then follows that relative humidity checks the amount of moisture or water vapour in a given volume of air relative to the amount in a saturated volume of air. This means that relative density depends on moisture content and air temperature.

Soil moisture on the other hand constitutes water that has been held up within the spaces of soil particles. It could be surface or root soil moisture. In surface soil moisture, water is being held within 10 cm from soil surface. In root soil moisture, water is held up within the root zone which is about 200 cm above the soil surface [6]. In this work, surface soil moisture was measured. Both surface and root soil moistures are useful to farmers since its selection depends on type of crop to be planted. This is one of the reasons soil moisture is measured.

Additionally, soil moisture plays key role in weather-pattern formation and precipitation. This it achieves through control of heat transfer and water exchange that takes place between the ground and atmosphere. By controlling the processes of transpiration and evaporation, soil moisture helps establish the weather pattern and precipitation.

Having explained the environmental variables that are of interest in this work, sampling techniques, measurement system and problem formulation are discussed in succeeding sections.

1.3 Sampling Techniques

Sampling is done while measuring environmental variables. Sampling is the act reading and collecting data from the environment using sensors. It may be random, systematic, exploratory, surrogate, composite or time-integrated [4]. In random sampling, variable is sampled randomly. Systematic sampling is a special type of random sampling where the sampling process is optimized in order to save energy, reduce sampling cost and improve reliability of measurements. Consequently, adaptive sampling is a type of systematic sampling. Another sampling technique is stratified sampling. Stratified sampling is a special type of random sampling where entire population is divided into strata. Sampling is done randomly so that stratum is formed. It continues until entire population is classified.

On the other hand, exploratory sampling involves conducting one or two observations so as to identify the presence or absence of variable of interest. Surrogate sampling is used where cost of observing and analyzing variable of interest is high, while there is alternative low-cost variable that could be measured. Data relating to the variable of interest are therefore deduced from the measured low-cost variable. For instance, the cost of analyzing sodium (Na) and chlorine (Cl) ions in soil sample is more expensive than measuring electrical conductivity. A cost-effective approach is therefore to measure the electrical conductivity of soil water extract [4].

Composite sampling is another cost-effective method of sampling environmental variables. It is a special type of exploratory sampling. In composite sampling, point measurement taken for specific time interval is considered as the average value and consequently, used for subsequent time intervals. Time-integrated sampling is another sampling method. It is commonly used in weather stations where environmental variables such as air temperature are measured frequently but reported hourly or daily with an average value [4].

It is worth noting that sampling techniques discussed above are classified as probability sampling techniques. This is because members of the population have equal chance of occurring or being selected. This contrasts non-probability sampling techniques where members of a given population do not have equal chance of occurring or being selected. Examples of non-probability sampling techniques include – quota sampling, accidental sampling, purposive sampling, etc.

It is also worth noting that while taking measurements, errors are prone to occur. Three types of errors exist - random, instrument calibration and systematic errors. Random errors are due to spread of measured data about a true value. It may be caused by unknown or unpredictable changes that occur during measurement. Examples of such changes include - electrical circuit noise in measuring instrument, changes in wind speed, etc.

Random errors are defined statistically by variance or standard deviation about a true value. They decrease as the number of measurements increases. Therefore, random errors impact most on precision of measurement. On the other hand, instrument calibration error is concerned with range of detection of the instrument. There is always uncertainty with calibration range of the measuring instrument. Certain assumptions are usually made during instrument calibration. As a result, some adjustments have been made in order to obtain optimized calibration curves [4].

Systematic errors, also known as constant errors, usually arise from the use of measuring instrument. If they occur, they are normally caused by functional error with instrument or instrument being wrongly used. For instruments with linear responses, two types of systematic errors are common, namely – offset and scale factor errors. Offset error implies that the instrument does not read zero when quantity of the variable being measured is zero.

Scale factor error means that the change in value of variable or quantity read by the instrument is either greater than or less than the actual (true) change. Therefore, systematic error impacts most on accuracy of the measurement. That is, how close the measured value is to the true value of the variable being measured. This is partly mitigated by ensuring that during data post-processing, accuracy of instrument indicated by the manufacturer in technical data sheet is taken into consideration.

1.4 Environmental Measurement System

The variables were measured using Hobo sensors. These are continuous analog signals which are represented in digital domain through amplitude quantization or time quantization. Hobo sensors measure the analog signals which are converted to voltage signals at output. Using its resolution (in mV/^{o}C for air temperature), measured values are obtained. This is discussed in this section and demonstrated using an ADC driven analog temperature sensor.

In order to effectively monitor and evaluate environmental variables, automated equipment is required [4]. This basically consists of sensor or transducer that senses the variable and a data acquisition system (DAS). At present, microprocessor-controlled automated data acquisition system is the most commonly available data acquisition system. Its data transmission process may be via wireless communication protocol; consequently, resulting in an EMS enabled WSN. Alternatively, data transmission may involve TCP/IP-enabled devices that connect independently to the internet as in ubiquitous application systems.

A complete measurement process for automated equipment consists of a sensor - placed in the environment, signal conditioning circuit, an analog-to-digital (ADC) converter, counters, processors and storage memories. The sensor, signal conditioning circuit, together with ADC and digital counters constitute the sensing unit. The block diagram below shows the measurement process.

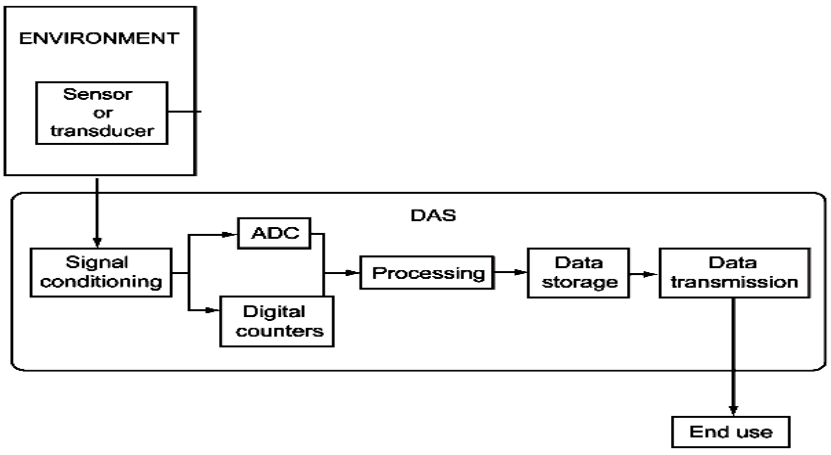


Figure 2: Measurement process of automated environmental measurement system [4]

For an environmental variable to be measured, sensing unit is required. The simplified sensing unit consists of sensor and analog-to-digital converter (ADC) only. Signal conditioning circuitry and its functions of filtering and amplification, in most cases, are embedded within the ADC. Similarly, the digital counter is part of the ADC's digital control logic unit. The ADC considered in this work is of successive approximation type. Sensing unit of a typical EMS enabled for WSN application is shown below.

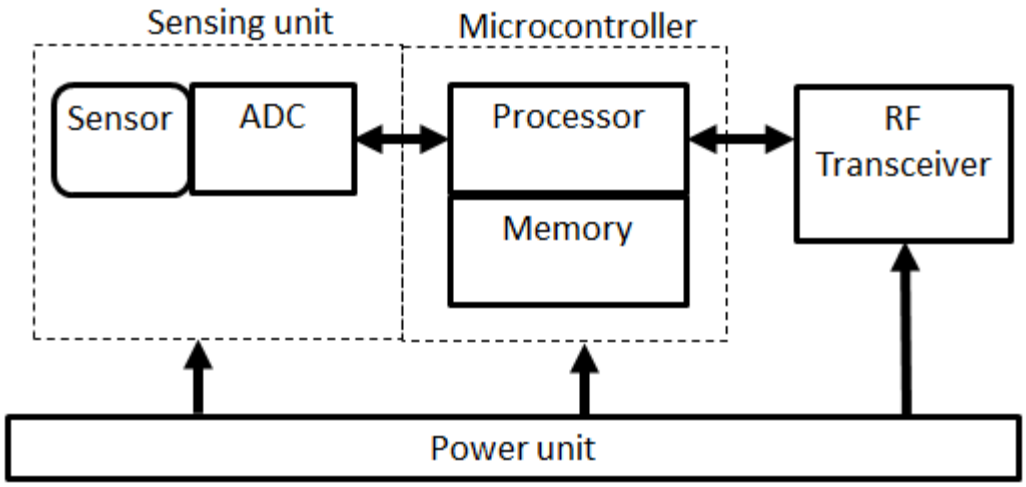


Figure 3: Sensing unit of EMS enabled for WSN applications [7].

The sensing unit measures, processes and relays processed input data either directly to end user (sink node) or indirectly via upstream modules such as microcontrollers, transceivers, digital-to-analog devices, etc. Autonomicity of the system demands that the complete system operate for relatively long time without human intervention. This implies the system should be able to operate with low need of maintenance, and with available energy for time length specified by customer. Simplified circuit drawing of a typical sensing unit is shown below. It consists of a single-ended voltage source SAR ADC driven by analog temperature sensor.

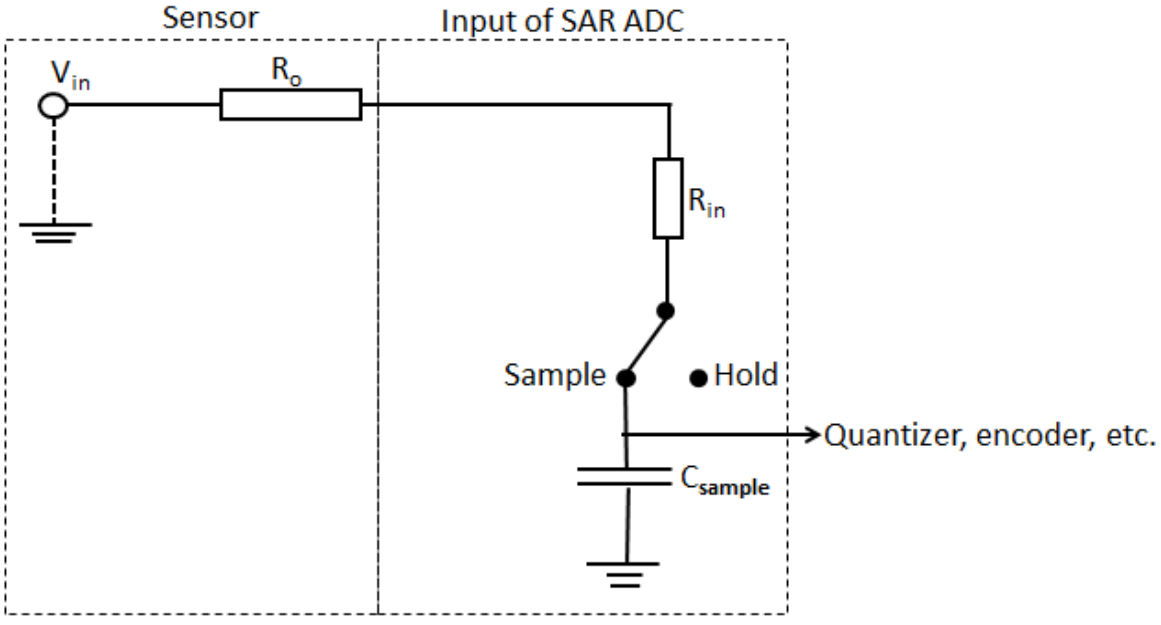


Figure 4: Schematic of a single-ended voltage source ADC driven by analog temperature sensor [8].

The sensor in figure 4 may be embedded in the ADC, or discrete and directly-coupled to the ADC. Regardless of sensor-ADC configuration selected, in this work, it is assumed that both sensor and ADC operate within one clock domain. This differentiates the chosen architecture from another architecture where sensor and ADC operate in different clock domains – one for sensor in range of few kHz and another for ADC in range of few MHz. The sensor considered is of analog type. Digital sensor was not considered because they usually have ADC embedded within them. This results in generation of digital output signals. Therefore, their consideration will increase the complexity of the model. Additionally, analog sensors are popular for systems used in environmental measurements due to their inherent small size, low cost and low power consumption capability [8].

For the ADC, a voltage source, single-ended mode SAR-ADC was considered. Successive approximation register (SAR) architecture was selected over other architectures such as sigma-delta, flash, pipelined, etc., due to its increased resolution and low-power consumption. The table below provides an overview of various ADC architectures and their characteristics.

Table 1: ADC architectures and their characteristics

ADC Architecture	**Resolution**	**Speed**	**Conversion time**	**Power consumption**
SAR	High	From medium to low	Increases with resolution	From medium to low
Flash	Low	High	Does not change with resolution	High
Pipelined	Medium	From medium to high	Increases with resolution	Medium
Sigma-Delta	High	Medium	Slow and decreases with resolution	From medium to low

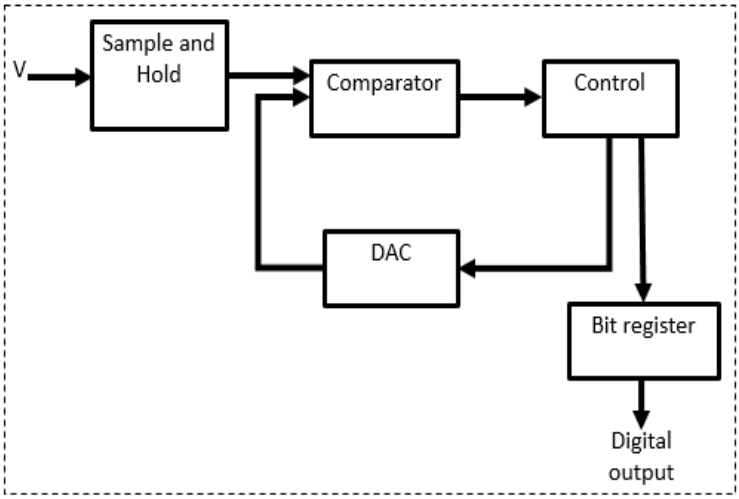


Figure 5: Block diagram showing components of SAR ADC [9].

SAR ADC uses the binary search method and consists of a high-speed comparator, a digital-to-analog converter (DAC), control logic and bit (control) register. The fully charged capacitor during sampling holds voltage, *V*. The voltage acts as input to SAR ADC. SAR architecture is based on binary search. The comparator compares the input voltage with half of the reference voltage coming from the DAC. If the input voltage is greater than half of the reference voltage, the most significant bit (MSB) is set (registered as 1) and the comparator computes for next MSB.

On the other hand, if the input voltage is less than half of the reference voltage, MSB is turned off (registered as 0) and the comparator computes for next MSB by comparing input voltage with updated reference voltage. In this case, the updated reference voltage is sum of last voltage that caused MSB to be turned on and the reference voltage for next bit position. The reference voltage for next bit position, in this case, is one-quarter of the first reference voltage. This continues until all bit positions are compared and data placed at data register for microcontroller to read.

1.5 Problem Formulation

Frequent sampling of environmental variables at sensor nodes and consequently, their analog-to-digital conversion process contribute in increasing total energy consumption and operating cost of measurement systems. They also create set-up and hold-time issues [10]. This is due to high data rate that occurs at analog-to-digital (ADC) output. High data rate at ADC output also demands that field-programmable gate arrays (FPGA) operate for longer time since they are required to capture all bits at ADC output. As a result, energy is not conserved.

Nyquist and Nyquist-Shannon sampling theorems did establish some bounds on how continuous-time signals should be recorded at discrete time instants. The theorems allow frequent sampling of variables thereby contributing in increased energy consumption of systems [1]. Furthermore, they do not provide simple and straight-forward way of determining accurate minimum sampling rates [11]. Before discussing further, the theorems are presented briefly.

Theorem 1.5.1: *Let f_s represent Nyquist sampling frequency and f_{max} the highest frequency contained in the signal. Nyquist sampling frequency theorem states that f_s must be at least twice of f_{max}. That is, sampling frequency should be at least two times greater than the highest frequency contained in the signal [11].*

In order to understand the above theorem, consider a cosinusoidal signal whose frequency is 1 Hz. This is shown in figure 6 below. According to Nyquist, the signal should be sampled with frequency of 2 Hz or more in order to avoid aliasing when reconstructed. Sampling the signal at 2 Hz, for example, captures all the peaks and troughs in the signal. Consequently, the original analog signal can conveniently be recreated without loss of information. However, sampling the signal with frequency lower than 2 Hz, say 1.5 Hz, captures only a few peaks and troughs in the signal. As a result, some data are lost and original signal cannot be recreated from sampled data. The above is shown in figures below.

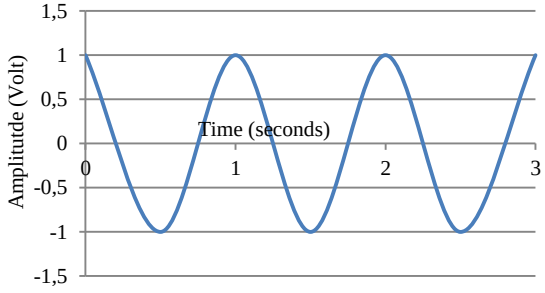


Figure 6: Original signal

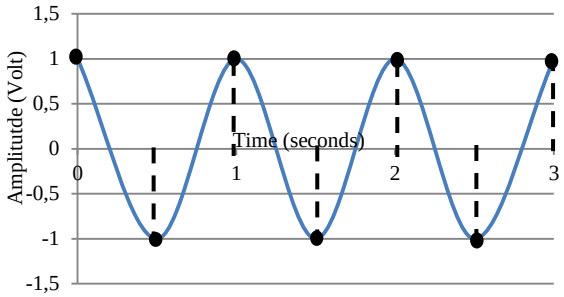


Figure 7: Sampling original signal at 2 Hz

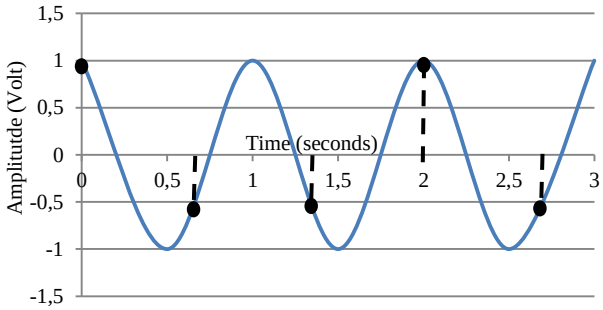


Figure 8: Sampling original signal at 1.5 Hz

Theorem 1.5.2: *Let B represent signal bandwidth of a band-limited signal and f_{ns} represent Nyquist-Shannon sampling frequency. Nyquist-Shannon sampling theorem states that f_{ns} must be at least twice of B. That is, sampling frequency should be at least two times greater than the signal bandwidth in order to avoid aliasing when reconstructed [11].*

Besides allowing frequent sampling of variable at sensor node, Nyquist and Nyquist-Shannon sampling theorems are faced with implementation challenges. Nyquist sampling theorem requires knowledge of highest frequency component of the signal. This is not always known a priori [1]. Consequently, power spectrum analysis of the signal has to be performed through fast Fourier transformation (FFT) or use of spectrum analyzer. For non-stationary signals with blowups, it becomes difficult to establish the actual highest frequency component. This usually leads to excessive oversampling.

On the other hand, Nyquist-Shannon sampling theorem demands that input signals be band-limited [10 - 12]. This constitutes a short-coming since no real world signal is truly band-limited. A truly band-limited signal has no energy outside some finite frequency band. This implies that the signal must extend infinitely in time [10, 11]. Although pure repetitive waveform with single frequency component is theoretically band-limited since it perfectly repeats itself over time, no real signal however exhibits such behavior.

Considering above challenges and need to sample less frequently for efficient utilization of energy, Nyquist sampling theorem is not commonly adopted [1]. It is for this reason that research is intense and novel methods such as compressive sensing emerging. These emerging methods aim to allow sampling below the Nyquist rate [1]. With reduced sampling rate, analog-to-digital converters can work with lower voltage levels. Also, output data rate is reduced. Consequently, energy consumption of the system is reduced [1, 10 and 13].

Additionally, use of lower sampling rates (rates below Nyquist rate) gives savings in memory space of sensor nodes. It also helps in reducing data transmission traffic in upstream stage (transceiver section) of measurement systems enabled for WSN application. Reducing

data transmission traffic makes it possible for lower frequencies of industrial-scientific-medical (ISM) band to be used. As a result, they give more range for a given power budget. Other benefits associated with using lower sampling rates include – reduced latency, reduced data hopping and bandwidth packing problems during data transmission. These help in improving system reliability and life cycle cost of the system.

While above discourse emphasized on need for available energy to be efficiently utilized through use of low sampling rates, it should be noted that trade-off exists between data quality and energy consumption when using low sampling rates [11, 12, and 14]. While low sampling rates help reduce energy consumption, they however result in loss of data. Consequently, there is need to balance this trade-off so that energy consumed is minimized without sacrificing to data loss. This constitutes the problem statement of this work. It is worth noting that the phrase "sampling at lower frequencies" implies sampling at rates below the Nyquist rate. Similarly, the phrase "without sacrificing to data loss" implies that all data lost due to aliasing while sampling at a lower frequency are accurately determined.

1.6 Research Methodology

The framework adopted for research methodology is that presented by Langhe [15]. The framework is a four-level structure with input questions that define the levels. It addressed important elements required in scientific research work. The elements include goals and objectives, mental models, research scope and type of study. Since goals are non-measurable while objectives are measurable, the goal of this research work is to investigate and determine dynamic stochastic modeling methods that are optimized for measurement of environmental variables.

Based on above goal, the objective is to determine novel stochastic modeling methods that enable sampling of environmental variables at lower frequencies without sacrificing to data loss. This is achieved through stochastic analysis involving real field datasets with energy consumption taken into consideration. Research questions such as how often environmental variables should be sampled in order to secure energy-efficiency were addressed. Novel method of determining optimal sampling interval was also presented.

Research mental model adopted was analytical. It comprised of predictive and empirical approaches. The research work was inductive and descriptive. It is inductive because it worked from specifics to generalizations. It is descriptive not because case-studies were involved, but also due to its "features-to-detect" characteristics. That is, the work moved from observation of features to detection.

CHAPTER TWO
The State of the Art

2.1 Sampling and Time Series Data Processing

Since most real world signals are continuous functions of time, they are required to be transformed into digital domain to enable compatibility with digital electronic components. Although it is easier to transmit analog signals through the air, they are still required to be digitalized so that digital electronic components can read, understand and manipulate them. Consequently, quantization is performed so that analog signals could be represented in the digital domain. Quantization could be amplitude or time quantization. Time quantization which is also known as sampling is the main focus in this context. It has been extensively discussed in preceding chapter. Nyquist and Nyquist-Shannon sampling theorems have also been discussed. They have been found not to be energy-efficient. Research is therefore intensive for energy-efficient sampling techniques that do not sacrifice to data loss.

Compressed sensing (also known as compressed sampling) has been found to be more energy-efficient than Nyquist-based sampling approaches [16, 17]. An improved version of compressed sensing which involves compressing while sampling made compressed sensing very competitive. Currently, it competes with adaptive sampling technique. Improved compressed sensing has also displaced the traditional compressed sensing method. Traditional compressed sensing samples first, thereafter compresses [1]. Other variants of compressed sensing method exist. However, the above-described improved version gives better results in terms of energy-efficiency and data quality.

For adaptive sampling technique, Al-Hoqani and Yang [16] proposed an adaptive sampling method that uses single exponential smoothing to perform predictions. In their work, Reno transmission control protocol (TCP Reno) was used for congestion control. Other variants of adaptive sampling technique exist.

Padhy et al [17] used Bayesian linear model to determine sampling rates. In their work, properties of Kullback-Leibler divergence were used to map values on sampling rate so that sensor adjusts sensing rate based on data it believes the system shall observe as next (incoming) data. Although their work adopted a mathematical approach to predict incoming data, their model and methodology differ from those used in this work. Also, sensor network topology used in their work differs from that used in this work. Their work used a mesh-network topology. Mesh-network topology is known to be susceptible to bandwidth and latency challenges [18].

Chatterjea and Havinga [19] used time series forecasting method to predict variability in measured values. In their work, sampling frequency changes whenever significant variability occurs in measured values. Time series models used in their work was completely linear and did not consider lagged errors as predictors. In this work, forecasts are linear functions of past data. They are also functions of lagged errors and considered lagged errors as predictors. Also, this work used a different reference condition in determining when models should be updated.

Law et al [20] developed an adaptive sampling method that uses Box-Jenkins approach to predict future values. In their work, when confidence interval of predicted value lies below a

specified error tolerance, predicted value is considered accurate enough and sensor skips sampling and goes for next sampling. The sensor samples if this condition is not fulfilled. Some significant differences exist between their work and that conducted in this research work. First, this research work established models that perform both forward and backward predictions. Second, this work samples with lower frequencies. This offers the capability of integrating lower frequencies of ISM band and their associated benefits. Third, this work used a different condition for adaptive sampling. The condition involves skip of subsequent sampling and continued use of predicted values if difference between immediate sampled and predicted values lies within a predefined residual threshold interval.

Furthermore, this work evaluates to determine whether or not next (succeeding) sampling should be measured while their work evaluates to determine if present sampling should be skipped. Moreover, if need for update of model occurs, the new method developed in this research work demands that model coefficients be re-estimated while keeping orders of model constant. As considered in this work, this holds for situations where dataset and its developed model accurately represent defined period or season. Thus values of the dataset are homogeneous. Least square errors in model coefficients will occur if the moving average component of the autoregressive integrated moving average model was not kept constant. Similarly, common-factor problem will arise if the autoregressive and integrated components of the autoregressive integrated moving average model were not kept constant.

Therefore, work presented in this contribution is simpler, less obscure and reliable. This is essential considering that simpler and less complex models give better energy efficiency.

2.2 Sensing Unit

This section reviews related works on sensing units used in stand-alone systems. This review is vital since energy consumption is among the parameters of focus in this work. Sensing unit which basically consists of sensor and analog-to-digital converter (ADC) has been described in section 1.4. Studies on low-power ADC have been on the increase.

O'Driscoll et al [21] showed that power consumption of an ADC can be adaptively reduced by reducing resolution of ADC. In their work, ADC resolution was varied from 8 to 3-bits and power consumption found to reduce from 0.90 mW to 0.23 mW.

Similarly, Schroeder [22] developed circuit architecture for integrated ADC that uses information-theoretic redundancy in input signals to reduce conversion workload in wireless sensor networks. The integrated ADC circuit architecture also performs data compression during ADC conversion. Consequently, savings in energy was achieved during radio transmission due to reduced workload and bandwidth.

Furthermore Chen [23] investigated the use of adaptive fuzzy resolution controller in adjusting sampling rate of ADC in wireless ECG health care monitoring system. In his work, a fuzzy resolution controller produced control signal that adaptively selects appropriate sampling clock for ADC so that sampling rate is adjusted.

In addition to the above, Artan et al [24] in their work on neural implants found that significant loss of ADC power occurs if sampling rate does not adapt to the significant value of the neurological signal being sampled. Their work also noted that reduction in ADC power consumption up to 62% is achieved if sampling rate is reduced for cases where values of sampled input signal are below a defined threshold value.

The work presented in thesis differs from prior arts in several ways. First, it does not require additional hardware such as comparators in the feedback control loop. Instead, it uses a software-only solution. Detection of change-point is performed statistically using data obtained at ADC output. When detected, the algorithm changes the frequency of ADC clock. This makes the system simpler and smarter. Consequently maintenance and operating costs will be reduced. This shall also reduce life cycle cost of sensor node.

2.3 Energy Consumption of Sensing Unit

In order to improve understanding of sensing unit and its energy consumption, an energy model was presented. This section reviews similar published works. It also presents distinctions between results in published works and those presented in this thesis work.

Hanfoug et al [25] presented a behavioral model of a current-source SAR ADC using MATLAB Simulink. Their model investigated both static and dynamic performance of an 8-bit current-mode SAR ADC. Non ideal factors such as switching noise, clock-feed through, flicker noise, clock jitter, etc., were taken into consideration. Simulation results for ideal model were compared with those for non-ideal model so that effects of non-ideality in ADC could be established. Work presented in this thesis report differs from their work in that it modeled a complete sensing unit; not just the ADC. Additionally, the ADC considered in this work is a voltage-source SAR ADC while theirs is a current-source SAR ADC.

Halgamuge et al [26] developed an energy model which is used to study energy consumption of certain components of a sensor node. For the sensing unit, their work did not consider energy consumed in converting analog input signal to b-bit packet. This is one of the differences between their work and that presented in this thesis. In this thesis, total energy consumed by ADC which comprises of energy consumed in switching the control logic, that consumed by DAC capacitor arrays and that consumed by comparator were all considered. Additionally, in this work, energy consumed during sampling mode and that dissipated by sensor output and ADC input resistances were taken into account.

Razzaque and Dobson [27] investigated sensing energy cost for certain off-the-shelf sensors. In their work, it was shown that energy consumption and cost of sensing operations are greater than that for certain radio frequency transceivers. It was also shown that compressed sensing and distributed compressed sensing give significant savings in sensing and overall energy cost of the system. Their work differs from that presented in this thesis in that it used digital sensors and did not consider energy consumed when analog-to-digital conversions are done.

Ginsburg [28] in his doctoral thesis presented energy models that were used to study energy consumption of SAR ADC and flash ADC. The ADCs were of the same bit-configuration. In terms of energy consumption, SAR ADC was found to be very competitive when compared against traditional high speed architectures such as flash ADC. Similar situation applies to Fredenburg et al [29] work. Both works differ from that presented in this thesis work for the reason that they did not consider sensor sub-unit in their energy model.

CHAPTER THREE
DATA ANALYSIS

3.1 Overview of Time Series Analysis

3.1.1 Time Series

Definition 3.1.1a: *Time series is defined as a sequential set of data points measured over successive times [30, 31]. That is, repeated measurements of the same phenomenon carried out over time.*

Example 3.1.1: Hourly air temperature values obtained between 09:00:15 and 07:00:15 on 2016-06-04 during measurement in a residential area in Edmonton constitute a time series. The series is presented as [18.2, 20.9, 24.1, 27.5, 30.3, 30.8, 22.4, 42.7, 26.6, 36.5].

A given time series may follow a probability model. The model describes joint distribution of random variables that constitute the time series. As a result, mathematical expression describing the probability model constitutes a stochastic process. Therefore, sequence of observations of a given time series are actually samples realizations of the stochastic process that produced them [31].

Time series and its analysis cannot be discussed without stating the central limit theorem.

Definition 3.1.1b: *The central limit theorem states that the distribution of sampling means approaches a normal distribution as sample size gets larger, notwithstanding shape of the population distribution [32].*

The implications of central limit theorem are:

1. The average of all the samples means will approach mean of population from where the samples were taken as sample size increases, notwithstanding the distribution of the data – whether normal or non-normal.
2. The average of all the samples means and standard deviation will approach standard deviation of population from where the samples were taken as sample size increases.

Time series dataset may be independent and identically distributed (iid). On the other hand, it may not be an iid. Independent and identically distributed (iid) dataset follows a random walk while non-iid time dataset follows a pattern [31]. Both iid and non-iid time series are described below.

3.1.2 IID and Non-IID Time Series

Definition 3.1.2a: *An independent and identically distributed time series is defined as collection of data that are identically distributed, and mutually independent. This means every finite subset of data within the time series is independent [33].*

Example 3.1.2a: Consider hourly air temperature data obtained between 09:00:15 and 07:00:15 on 2016-06-04 during measurement in a residential area in Edmonton. The sequence of values is already shown in Example 3.1.1. It constitutes an iid time series because dataset is uncorrelated and normal distributed with finite variance. At this point, it is worth noting that:

1. Uncorrelated dataset does not absolutely mean that dataset is independent. This is because dataset could be uncorrelated and still dependent [33].

2. Independence of data does not absolutely mean that given dataset is an iid. This is because dataset can be independently distributed and still not identical [33]. An independent dataset whose subsets have different probability distribution functions is not identically distributed.

It is for above reasons that sequence of values shown in Example 3.1.2 was investigated for iid by studying status of its autocorrelation coefficients and normality. Autocorrelation function (ACF) analysis was performed in order to determine correlation between elements of the sequence. Normality on the other hand was investigated using Chi-square test. They are shown in the preceding analysis where autocorrelation and chi-square are described.

Theorem 3.1.2a: *Members of a given time series are correlated if their autocorrelation coefficient is near zero and within the 95% confidence interval. On the other hand, they are uncorrelated if one or more of the autocorrelation coefficients is significantly non-zero [33].*

The ACF plot of Example 3.1.2 requires that autocorrelation coefficients of various orders be computed and plotted against corresponding lag. The equation below and table describe it.

$$r_t = \left(\frac{1}{(n-1)\left(S_{x_t} * S_{x_{t+1}}\right)}\right) \sum_{t,i=1}^{n-1} \left[(x_{t,i} - \dot{x}_t)(x_{t+1,i} - \dot{x}_{t+1})\right] \qquad (1)$$

where:

r_t represents autocorrelation coefficient of order t,

$x_{t,i}$ represents series of first observations for order t,

$\dot{x}_t$ represents sample mean of values contained in the series x_t,

$x_{t+1,i}$ represents series of last observations for order t,

$\dot{x}_{t+1}$ represents sample mean of values contained in the series x_{t+1},

S_{x_t} represents sample standard deviation of values contained in series x_t,

$S_{x_{t+1}}$ represents sample standard deviation of values contained in series x_{t+1},

n represents number of values contained in each of the series.

Table 2: Autocorrelation analysis table using hourly air temperature data

Time index	**Air temp (°C)**	**Air temp (°C), First order, x_1**		**Air temp (°C), Second order, x_2**		**Air temp (°C), …Ninth order, x_9**	
		First (n-1) values	Last (n-1) values	First (n-2) values	Last (n-2) values	First (n-9) values	Last (n-9) values
1	18.2	18.2	20.9	18.2	24.1	18.2	36.5
2	20.9	20.9	24.1	20.9	27.5		
3	24.1	24.1	27.5	24.1	30.3		
4	27.5	27.5	30.3	27.5	30.8		
5	30.3	30.3	30.8	30.3	22.4		
6	30.8	30.8	22.4	30.8	42.7		
7	22.4	22.4	42.7	22.4	26.6		
8	42.7	42.7	26.6	42.7	36.5		
9	26.6	26.6	36.5				
10	36.5						
Sample mean		27.06	29.09	27.11	30.11	-	-
Sample standard deviation		6.80	6.58	7.21	6.27	-	-
Autocorrelation coefficient (r)		-0.052		0.380		-	

Distribution of autocorrelation coefficients computed for various orders or lags is shown in the table below. Similarly, ACF plot with limit lines of 95% confidence interval is shown in figure below. Limit lines corresponding to 95% confidence interval were estimated using equation (2) [33].

Table 3: Summary of autocorrelation coefficient

Lag (order)	1	2	3	4	5	6	7	8
Autocorrelation coefficient (r)	-0.052	0.387	-0.075	-0.015	-0.049	-0.097	-0.337	-0.094

$$CI_{95} = \frac{\pm 1{,}96}{\sqrt{n}} \qquad (2)$$

where:
CI_{95} represents 95% confidence interval limit,
n represents number of samples.

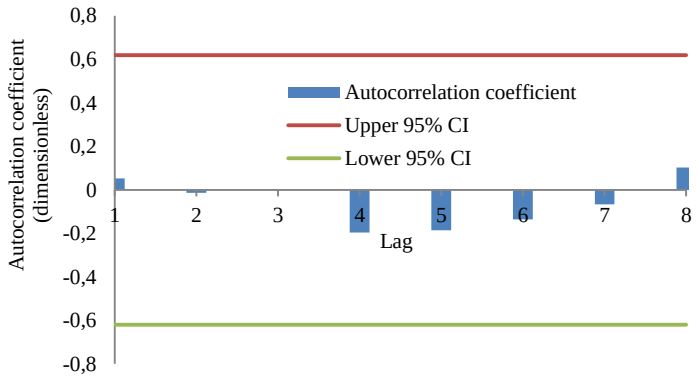


Figure 9: Autocorrelation function plot of an iid time series

In above figure, it would be seen that all autocorrelations are near zero and within lower and upper limits of 95% confidence interval. This shows that values are uncorrelated. Next is the use of Chi-square test to investigate normality of data.

Theorem 3.1.2b: *Assuming $z_1, \ldots, z_k$ represents set of independent observations from normal distributed population, then the distribution constituting sum of squares of the observations is a chi-square distribution [33].*

That is, a random variable, w, represented in equation (3) has a chi-square distribution χ^2(k) as long as, k is a positive integer and observations $\{z_1, \ldots, z_k\}$ taken from independent and normal distributed population.

$$Z_i = \sum_{i=1}^{k} {z_i}^2 \qquad (3)$$

The null hypothesis is that observations are taken from a normal-distributed dataset. Therefore if the null hypothesis is true, distribution of the test statistic follows a chi-square distribution. Pearson's chi-square test statistic is one of the commonly used test statistics in performing the above-described test for goodness of fit. It tests to determine whether observed data correspond sufficiently well to expected values. It is described in below equation.

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} \qquad (4)$$

where:

χ^2 represents chi-square test statistic,

O_i represents the observed frequency for bin i,

E_i represents the expected frequency for bin i,

n represents the number of bins,

Using Example 3.1.2, chi-square test for goodness of fit was performed. The steps are described below.

Step 1: Mean and standard deviation of the dataset were computed and noted.

Step 2: Dataset was grouped in bins. Frequency of each bin was noted. This constitutes the observed frequency.

Step 3: Using computed mean and standard deviation, normalized Gaussian distribution function was determined. For each bin, definite integral of normalized Gaussian distribution function was performed. Lower and upper limits of the definite integral were the lower and upper limits of the bins. Since normalized Gaussian distribution is the same as normal distribution, result obtained when total number of data is multiplied by definite integral of normalized Gaussian distribution over given bin is the expected frequency (cumulative frequency) for that bin. The results are shown in Table 4 while the equation below shows probability density function of normal distribution.

$$f(x) = \frac{e^{-(x-\mu)^2/(2\sigma^2)}}{\sigma * \sqrt{2\pi}} \qquad (5)$$

where;

$f(x)$ represents probability density function of normal distribution,

μ represents the mean of given dataset,

σ represents the standard deviation of given dataset.

Step 4: With expected frequency determined, chi-square value for each of the bins was computed using equation (4). The steps and results obtained are summarized in below table.

Table 4: Chi-square test of normality for iid time series

Time index	**Air temp. values in ^{o}C**	**Air temp. values in bins (^{o}C)**	**Air temp. observed frequency (O_i)**	**Air temp. expected frequency (E_i)**	**Chi-square value $= \frac{(O_i - E_i)^2}{E_i}$**
1	18.2	18 – 21	2	0.8218	1.6892
2	20.9	21 – 24	2	1.2456	0.4570
3	24.1	24 – 27	1	1.5791	0.2124
4	27.5	27 – 30	1	1.6747	0.2718
5	30.3	30 – 33	2	1.4856	0.1781
6	30.8	33 – 36	0	1.1024	1.1024
7	22.4	36 -39	1	0.6843	0.1457
8	42.7	39 – 42	0	0.3553	0.3553

9	26.6	42 – 45	1	0.1543	4.6356
10	36.5	45 – 48	0	0.0560	0.0560
				Sum	9.1035

Degree of freedom was determined as number of categories (or groups) minus one. In this case, there are ten groups or categories. Thus, degree of freedom was (10 – 1 = 9). Chi-squared test statistic of 9.1035 was obtained. This gives p-value that is between 0.25 and 0.5 when read from chi-square distribution table. Using Excel's "CHISQ.TEST" function, exact value of p-value was found to be 0.4277. *P-value* of 0.4277 is greater than standard significance level of 0.05 usually used in statistical hypothesis testing. Thus, the null hypothesis cannot be rejected. In order to reject the null hypothesis *P-value* less than 0.05 is required. Therefore, dataset tested for goodness-of-fit is normal distributed.

Definition 3.1.2b: *A non-iid time series is defined as collection of data that are not identically distributed and independent [33].*

The above definition implies that non-iid time series is dependent and non-identically distributed. This usually occurs in situations where data points are interdependent such as cases where data exhibit significant temporal or spatial correlations. Such an example is described below.

Example 3.1.2b: Consider values of photosynthetically active radiation (PAR) recorded every 30 minutes in a Brazilian forest from 23:30 on 2009-01-01 to 10:00 on 2009-01-02. Due to significant temporal correlation existing within the dataset, it constitutes non-iid time series. The analysis is discussed below. Histogram plot of dataset and its autocorrelation function plot are shown below.

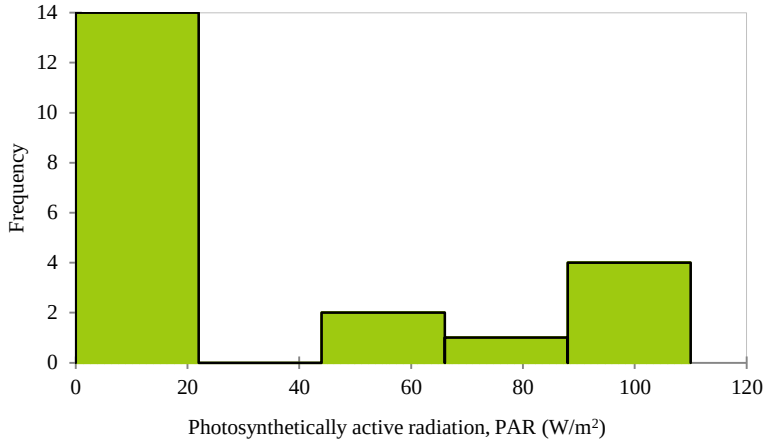


Figure 10: Histogram plot of non-iid time series

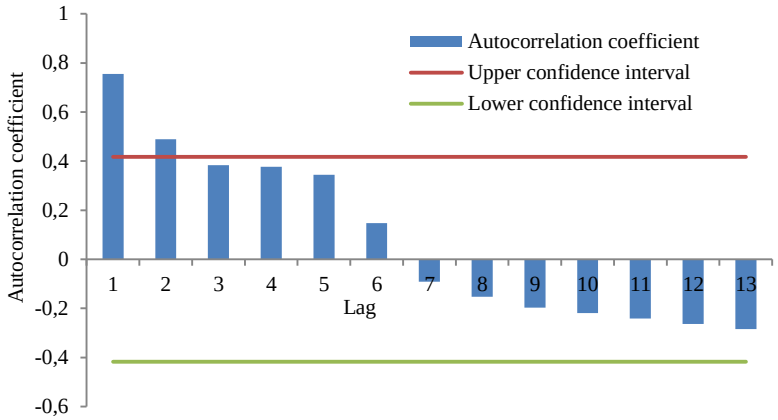


Figure 11: Autocorrelation function plot of a non-iid time series

Fitting a normal distribution curve to histograms in figure 10 appears to be difficult. This suggests the dataset is not normal distributed. In order to confirm this, chi-square test for normality was performed. This is summarized in the table below.

Table 5: Chi-square test of normality for a non-iid time series

PAR values in bins (W/m^2)	**PAR observed frequency (O_i)**	**PAR expected frequency (E_i)**	**Chi-square value = $\frac{(O_i-E_i)^2}{E_i}$**
1.2 – 21.2	15	4.1339	28.56112
21.2 – 41.2	0	4.1536	4.1536
41.2 – 61.2	2	3.7748	0.8345
61.2 – 81.2	1	2.4182	0.8318
81.2 – 100.2	2	1.1865	0.5577
100.2 – 120.2	2	0.4459	5.4165
		Sum	40.3552
		p-value	0.0067

P-value was calculated to be 0.0067. This is less than the significance value (0.05). Therefore the null hypothesis that dataset is normal distributed is rejected. Also, figure 11 shows that autocorrelation coefficient at lags 1 and 2 were significant and above upper confidence interval limit. This indicates that dataset is dependent and correlated.

Analyzing the dataset further revealed that logistic distribution fits best among all the distributions examined. *P-values* obtained during goodness-of-fit tests are shown in table below. Similarly, the figure below shows the dataset being fitted with logistic distribution.

Table 6: Goodness-of-fit test results for a non-iid time series

Distribution	**p-value**
Exponential	0.0000
Gamma	0.0050
Normal	0.0067
Log-normal	0.0040
Logistic	0.052
Weibull	0.0070

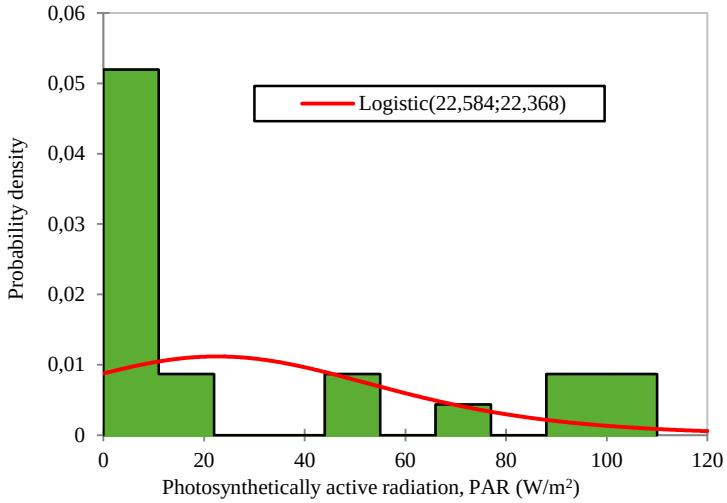


Figure 12: Dataset fitted with logistic distribution

3.1.3 Discrete-time and Continuous-time Series

Time series data may be discrete-time or continuous-time. In discrete-time series, observations are recorded at discrete time-points and values constitute a finite set. Continuous-time series on the other hand has observations that were continuously recorded over a closed time interval. Some discrete-time series are obtained by observing a continuous-time process at discrete observation times [33].

In systems or models where nonlinearity exists, continuous-time series are most suitable in describing the nonlinearity. This is due to reduced risk of missing data while sampling in continuous-time. Furthermore, some system dynamics are better investigated using continuous-time series. In such systems, time may be made granular during measurement.

3.1.4 Stationary and Non-stationary Time Series

Stationarity defines the quality of time series properties. This is essential since statistical properties of a given time series are required to be time-invariant so that model can be fitted to the series. Stationarity therefore plays an important role in time series analysis. Stationary time series may be weakly or strictly stationary.

Definition 3.1.4a: *Time series is said to be weakly stationary if its first and second orders of moments are time-invariant [33].*

First orders of moment are mean and variance while autocorrelation is the second order of moment. Thus, above definition implies that statistical properties of mean, variance and autocorrelation do not change with time in a weakly stationary time series. Such time series usually originate from a stationary process.

Definition 3.1.4b: *If t and h are positive integers, then time series $\{X_t\}$ is strictly stationary if joint distribution of $\{X_1, X_2, \dots, X_p\}$ is the same as that of $\{X_{1+h}, X_{2+h}, \dots, X_{p+h}\}$ [33].*

This means that joint distribution of a strict stationary time series only depends on difference, h, but not on time. It therefore implies that a strict stationary time series has its

higher orders of moments (e.g. skewness, kurtosis) being constant [33]. A strict stationary time series has the same structure when moved forward or backward in time.

It shall be noted that weak stationarity is a sufficient and necessary condition in linear time series analyses. In non-linear time series analyses, weak stationarity condition is not usually sufficient. Strict stationarity is usually required to be fulfilled. The example below has been provided to describe the concept of stationarity.

Example 3.1.4: Consider hourly values of air temperature obtained during measurement in a residential area in Edmonton from 09:00:15 on 2016-06-04 to 08:00:15 on 2016-06-05. The set of values obtained at time index, $t = \{1, 2, \ldots 24\}$ is shown below.

$$X_t = \{18.2, 20.9, 24.1, 27.5, 30.3, 30.8, 22.4, 42.7, 26.6, 36.5, 26.3, 22.4, 20.4, 16.7, 15.2, 14.4, 13.8, 12.7, 11, 9.9, 9.2, 9, 10.6, 13.1\}$$

If two subsets are drawn from the series, it would be seen that they have unequal mean and variance. However, if values of the subsets are logarithmic-transformed, it would be seen that their mean and variance become approximately the same (see Table 7). Therefore, the original series is not stationary but logarithmic-transformation of it has helped it become stationary. The table below summarizes the analysis

Table 7: Securing stationarity

	X_t = Air temperature values (^oC)		**Log-transformed air temperature values (^oC)**	
	Subset 1: first 12 datasets in the set	Subset 2: last 12 datasets in the set	Subset 1: first 12 datasets in set	Subset 2: last 12 datasets in set
	18.2	20.4	1.260071388	1.309630167
	20.9	16.7	1.320146286	1.222716471
	24.1	15.2	1.382017043	1.181843588
	27.5	14.4	1.439332694	1.158362492
	30.3	13.8	1.481442629	1.139879086
	30.8	12.7	1.488550717	1.103803721
	22.4	11.0	1.350248018	1.041392685
	42.7	9.9	1.630427875	0.995635195
	26.6	9.2	1.424881637	0.963787827
	36.5	9.0	1.562292864	0.954242509
	26.3	10.6	1.419955748	1.025305865
	22.4	13.1	1.350248018	1.117271296
Sample mean	27.39	13.00	1.43	1.11
Sample variance	47.86	11.44	0.011	0.011

In addition to above approach, stationarity can be checked through following tests:

1. Modulus of roots of the characteristic equation: If absolute value (modulus) of roots of model characteristic equation lies outside a unit circle, the model is stationary. This means that absolute value of roots of model characteristic equation must be greater than one.
2. Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test**:** Based on null hypothesis that time series is trend stationary with no noise component. This implies that the time series trend has only deterministic component. The alternative hypothesis is that time series has both trend and noise component. This implies it has stochastic component.
3. Augmented Dickey-Fuller (ADF) test: Based on null hypothesis that time series has unit roots. This means that model characteristic equation of the time series has roots

that lie within the unit circle. Consequently, the time series is not stationary. The alternative hypothesis is that time series does not have unit roots. This implies the roots lie outside the unit circle. Therefore the time series is stationary.

It is always recommended that ADF test be performed whenever KPSS test is used. It complements KPSS test [33].

3.1.5 Linear and Non-linear Time Series

Time series may be linear or non-linear. Linear time series are represented by models that investigate covariance structure of the series. According to [33] and [34], linear time series are represented by two main models:

1. Autoregressive models
2. Moving average models

Definition 3.1.5a: A time series is linear if it can be represented as shown below.

$$X_t = \sum_{j=-\infty}^{\infty} \theta_j \epsilon_{t-j} \qquad (6)$$

where:

$\{X_t\}$ represents time series,

$\{\theta_j\}$ represents sequence of constants,

$\{\epsilon_j\}$ represents sequence of uncorrelated random values with mean of zero and finite variance.

The sequence of values, $\{\epsilon_j\}$, described above does not need to be independent. It is similar to white noise and should not be confused with Gaussian (or strict) white noise.

Equation (6) also defines the moving average model. In simplified form, it can be written as shown below.

$$x_t = \epsilon_t + \theta_1\epsilon_{t-1} + \theta_2\epsilon_{t-2} + \theta_3\epsilon_{t-3} + \cdots + \theta_q\epsilon_{t-q} \qquad (7)$$

where:

q represents order of the model,

$\theta_1, \theta_2, \ldots, \theta_q$ represent model parameters,

ϵ_t represents sequence of uncorrelated random values with mean of zero and finite variance.

For moving average model of order 1, that is MA(1), equation (7) reduces to equation (8).

$$x_t = \epsilon_t + \theta_1\epsilon_{t-1} \qquad (8)$$

Linear time series defined above can also be represented in terms of autoregressive (AR) model since MA and AR models are invertible. That is, an MA model can be converted to an AR model and vice versa. This leads to Definition 3.1.5b presented below.

Definition 3.1.5b: Linear time series $\{X_t\}$ is said to come from an autoregressive model if it satisfies the equation shown below [33].

$$X_t = \sum_{j=1}^{p} \Phi_j X_{t-j} + \epsilon_t \qquad (9)$$

where:

$\{X_{t-j}\}$ represents past values of elements of the series,

p represents order of the model,

$\{\Phi_j\}$ represents sequence of constants or coefficients of model,

$\{\epsilon_t\}$ represents sequence of uncorrelated random values with mean of zero and finite variance.

In simplified form, equation (9) reduces to equation (10).

$$x_t = \Phi_1 x_{t-1} + \Phi_2 x_{t-2} + \Phi_3 x_{t-3} + \ldots \Phi_p x_{t-p} + \epsilon_t \qquad (10)$$

where:

$\Phi_1, \Phi_2, \Phi_3, \ldots, \Phi_p$ represent model parameters or coefficients,

p represents order of the model,

ϵ_t represents sequence of uncorrelated random values with mean of zero and finite variance.

Example 3.1.5:

Consider hourly values of air temperature obtained during measurement in a residential area in Edmonton from 09:00:15 on 2016-06-04 to 08:00:15 on 2016-06-05. The values were shown in preceding example (Example 3.1.4). Without using software program, an AR model was fitted to the dataset. Below are the steps and results obtained.

Step 1: Time series was detrended by taking the first difference $(X_{t+1} - X_t)$. The differenced series is shown below.

$$\{X_{t+1} - X_t\} = \{2.7, 3.2, 3.4, 2.8, 0.5, -8.4, 20.3, -16.1, 9.9, -10.2, -3.9, -2, -3.7, -1.5, -0.8, -0.6, -1.1, -1.7, -1.1, -0.7, -0.2, 1.6, 2.5\}$$

Step 2: Autocorrelation function (ACF) and partial autocorrelation function (PACF) of the differenced series were computed. They are shown below.

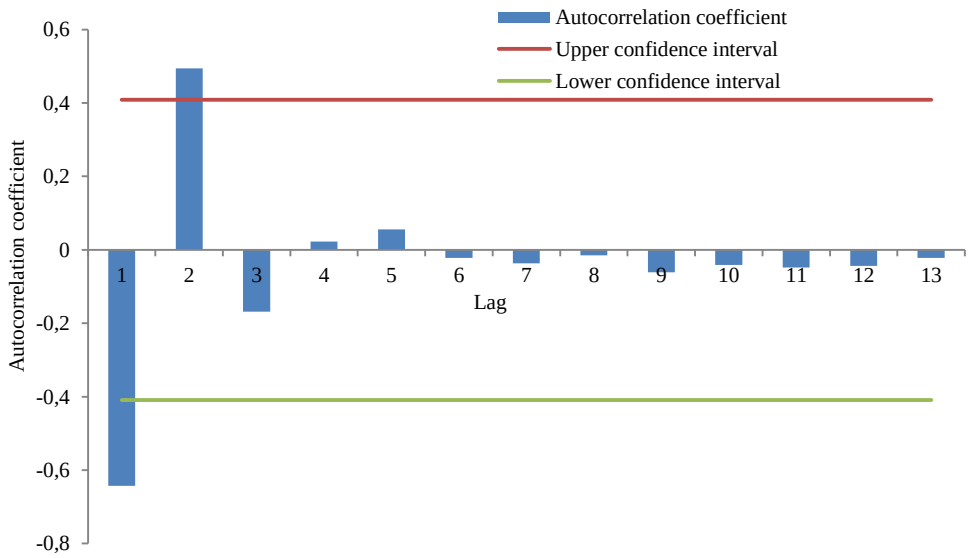


Figure 13: Autocorrelation function plot for Example 3.1.5

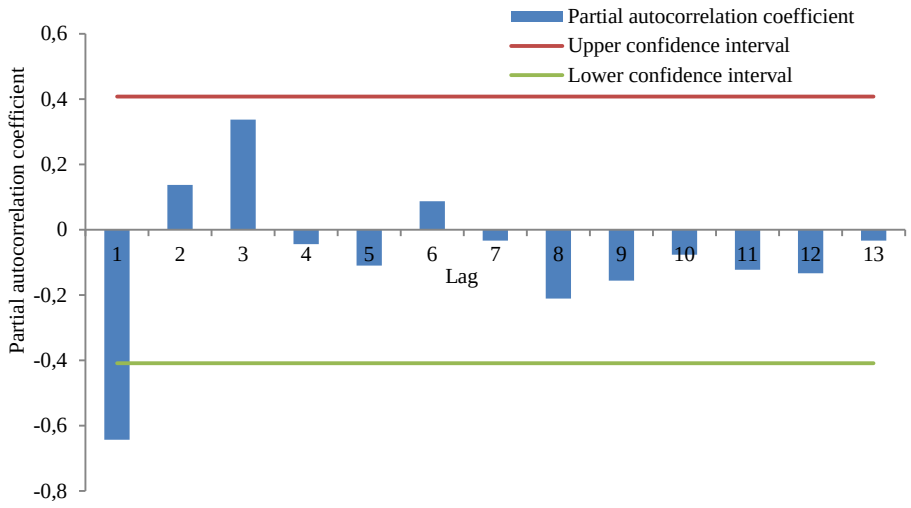


Figure 14: Partial autocorrelation function plot for Example 3.1.5

Step 3: The above ACF plot shows an AR(1) model. This is because autocorrelation coefficient is significant at lag 1. Also, it would be seen in PACF plot that partial autocorrelation coefficient is not significant at lag 2 and above. This further confirms an AR(1) model since partial autocorrelation coefficient of an AR(p) process is non-significant at lag (p + 1) and above.

Step 4: Considering the simplified equation of an AR(p) process, an AR(1) model is therefore represented below.

$$x_t = \Phi_1 x_{t-1} + \epsilon_t \qquad (11)$$

Model coefficient $\Phi_1$was determined by finding value of Φ_1 that minimizes sum of squares of residuals (SSR) subject to constraint that mean of residuals should be equal to zero. Setting initial guess of $\Phi_1$as zero and using values of x_t and x_{t-1}, residuals (ϵ_t) were computed using equation (11). The residuals were squared and summed to give SSR. Using Excel Solver, SSR was set as objective function. SSR was then minimized by varying $\Phi_1$until $\Phi_1$value which gives lowest possible SSR and zero mean residuals was determined. $\Phi_1$was found to be 0.989. Thus, model equation of the first differenced series becomes:

$$x_t = 0.989 x_{t-1} + \epsilon_t \qquad (12)$$

Plot of model predicted series against first differenced series is shown below. Also shown below is plot of model predicted series against the original (initial) time series.

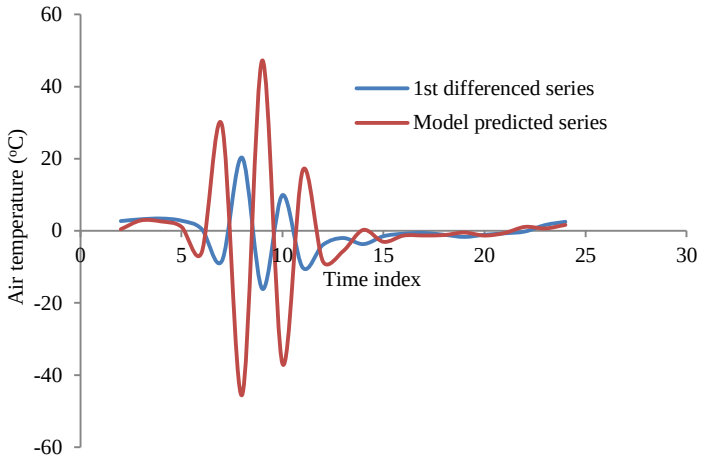


Figure 15: Plot of model predicted series against first differenced series

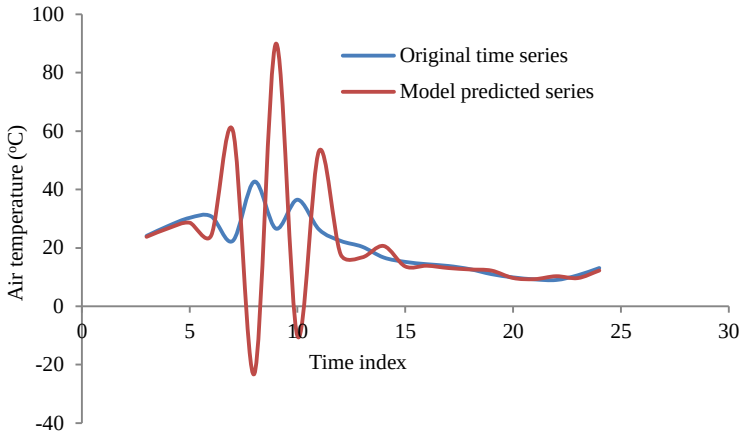


Figure 16: Plot of model predicted series against original (initial) time series

It would be seen in figure 16 that the model can predict the original series. Histogram plot of the residuals is shown below. Mean value of residuals was calculated as -0.01^{o}C and variance calculated as 151.5^{o}C.

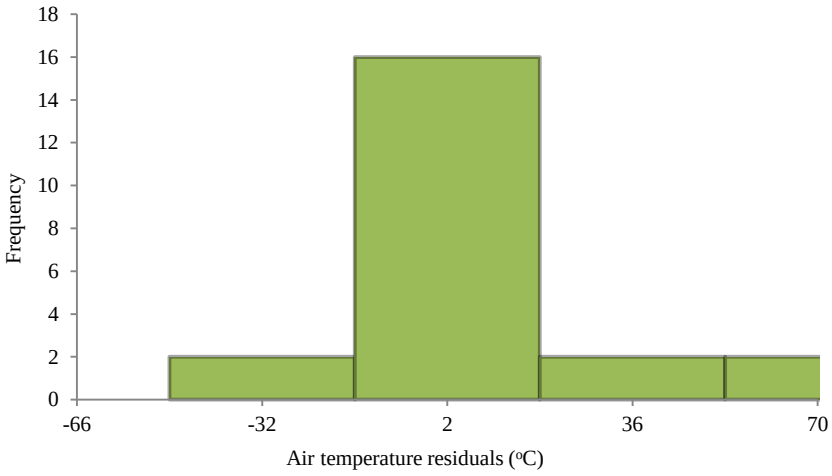


Figure 17: Histogram plot of residuals of AR(1) model - air temperature model

Non-linear time series are produced by non-linear dynamic stochastic processes [33]. They exhibit characteristics that are difficult to be modeled using linear time series models. Such characteristics include breaks and thresholds, time-dependent variances, asymmetric and aperiodic cycles, etc.

Some analyses involving non-linear time series require determination of test statistics of probability distribution of the given series. Such test statistics include - Lyapunov exponents, entropies and prediction errors, etc. This is quite a tedious task since most non-linear test statistics require their statistical estimates such as time-reversal invariance, etc., to be first determined [34]. Furthermore, this approach of analyzing non-linear time series through its test statistics is largely applied when time series model is low dimensional in state space. This means that state space representation of transition and observation equations of the series is of lower order. As a result, there is high focus on finding simplified approach for non-linear time series analysis.

In order to characterize stochastic parameters of non-linear time series, models based on threshold principle are usually considered. Such models include - threshold autoregressive, exponential autoregressive (EXPAR), smooth transition autoregressive (STAR), etc [35]. The models demand that regimes be created and dynamics associated with the regimes investigated. It is recommended that non-linearity tests be performed before selecting and building non-linear model. For this, Brock-Dechert-Scheinkman (BDS) test of non-linearity may be considered [36].

3.2 Approaches in Time Series Analysis

Time series can be analyzed using any of the following approaches:

- Classical-decomposition approach
- Parametric approach
- Non-parametric approach.

Classical-decomposition approach is a traditional approach that has long been used. It is also known as reduction-to-noise approach. It is based on the concept that components of a given time series can be gradually removed so that resulting component approximates to strict white noise. This is because strict white noise cannot be modeled. Functions representing all the eliminated (removed) components add up to constitute model for original time series.

Parametric approach involves the use of models [37]. It tends to fit model to given time series data. On the other hand, non-parametric approach does not involve the use of models. It is model-free and does not require model assumptions as well. This approach is mainly data-driven. It explores characteristic features of time series data and uses them for identification and characterization of time series data.

Above-mentioned approaches are treated in detail in preceding sections. Examples are provided to demonstrate their applications.

3.2.1 Classical-Decomposition Approach

Classical-decomposition approach considers times series as being composed of trend, cyclic, seasonal and random components. It aims to detrend time series data, and remove all seasonal and irregular components. The components are described in the equation below.

$$X_t = T_t + C_t + S_t + R_t \qquad (13)$$

where:
X_t represents values of the series,
T_t represents trend component of the series,
C_t represents cyclic component of the series,
S_t represents seasonal component of the series,
R_t represents random component of the series.

Classical-decomposition approach is composed of following steps:

1. Ensure that time series is stationary. If time series is not stationary, perform logarithmic or square-root transformations of the series. This helps in stabilizing the variance. If non-stationarity is due to variation in mean, perform differencing.
2. Fit model to stationarized time series. It is usually recommended to start with the simplest function, a linear function. Linear functions are easy to fit and help detrend the series. The resulting series (detrended series) obtained after fit of linear function is subjected to further analysis (see steps 3 and 4 below). This resulting (detrended) series is the difference between stationarized series and fitted linear function.
3. Check if detrended series has seasonal or cyclic components. If seasonal component is present, identify its period. If cyclic component is present, identify its number of cycles.
4. Fit function to detrended series using moving average or harmonic functions. Use of moving average or harmonic functions help remove seasonal and cyclic components. The function should be fitted such that resulting residual approximates to strict white noise. Resulting residual in this stage is the difference between fitted function and detrended series. Also analyze the resulting residuals so as to ascertain if they are independent with mean of zero and finite variance. This implies that residuals have been reduced to lowest possible form. A form similar to that of strict white noise since strict white noise cannot be modeled.
5. Time series model therefore constitutes sum of all fitted functions. Remember to transform back values read-out (obtained) using the model to ensure consistency, in form, with the initial (original) time series values.

Example 3.2.1 has been provided to demonstrate how classical-decomposition approach works following above-described steps.

Example 3.2.1: Consider hourly values of air temperature obtained during measurement in a residential area in Edmonton from 09:00:15 on 2016-06-04 to 08:00:15 on 2016-06-05. The values were shown in Example 3.1.4. They reproduced and shown below.

$$X_t = \{18.2, 20.9, 24.1, 27.5, 30.3, 30.8, 22.4, 42.7, 26.6, 36.5, 26.3, 22.4, 20.4, 16.7, 15.2, 14.4, 13.8, 12.7, 11.0, 9.9, 9.2, 9.0, 10.6, 13.1\}$$

Classical-decomposition approach is used to analyze data and fit model to the series. The steps are shown below:

Step 1*:* From Example 3.1.4, it would be seen that time series was logarithmic-transformed in order to achieve stationarity. The resulting series is shown below.

$$\log_{10}(X_t) = \{1.26, 1.32, 1.38, 1.44, 1.48, 1.49, 1.35, 1.63, 1.42, 1.56, 1.42, 1.35, 1.31, 1.22, 1.18, 1.15, 1.13, 1.10, 1.04, 0.10, 0.96, 0.95, 1.03, 1.12\}$$

Step 2: Linear function was fitted to above-stationarized series. The results are shown below.

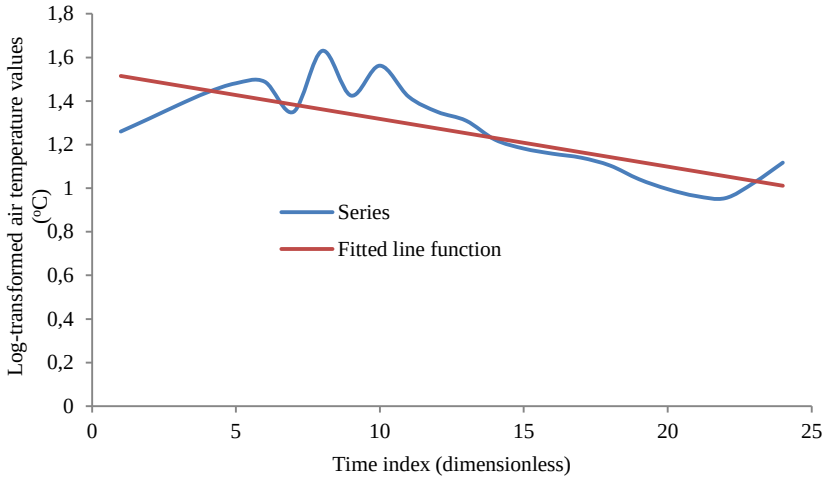


Figure 18: Straight line fitted to stationarized time series

Straight line being the simplest linear function was selected and fitted to the stationarized series. Its equation is shown below.

$$T_t = -0.0219t + 1.5369 \qquad (14)$$

where:
T_t represents the log-transformed air temperature values,
t represents time index.

Step 3: Detrended series is obtained by subtracting values of fitted straight line from the series. That is $(\log_{10}(X_t) - T_t)$. The resulting series is shown below.

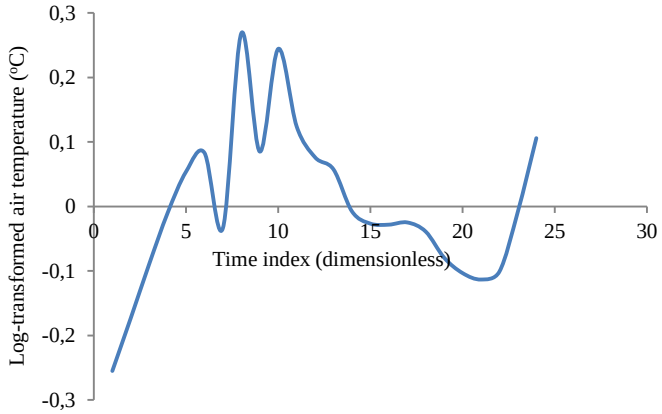


Figure 19: Plot of resulting series after straight line fit

Step 4: Irregular cyclic components are seen within the series shown in figure 19. In order to remove these components and smoothen the series, 4-point moving average was performed. The choice of 4-point moving average is due to the presence of four cycles. Also some cycles have only one or two time points in-between them. The results are shown in Table 8 and in figure 20. Also, below equation shows how fitted values were calculated using 4-point moving average.

$$Y = \frac{1}{n}\sum_{t=0}^{n-1} D_{t-i} \qquad (15)$$

where:

Y represents computed value of log-transformed air temperature (see Table 8),

n represents number of points of moving average which in this case is 4,

D represents values of detrended series (see Table 8),

t represents time index for which t at $n = 0$.

Table 8: Removal of irregular cycles from detrended series and smoothening of the series

Time index	Detrended values of log-transformed air temp., (D) in ^{o}C	Computed values of log-transformed air temp. using 4-point moving average, (Y) in ^{o}C	Values of poly-fitted function obtained after smoothening, (Y_c) in ^{o}C	Residuals of log transformed air temp., (D -Y_c) in ^{o}C
1	-0.254978379	Not applicable	Not applicable	Not applicable
2	-0.173027725	Not applicable	Not applicable	Not applicable
3	-0.089281213	Not applicable	Not applicable	Not applicable
4	-0.010089806	-0.131844281	-0.1505	0.1405
5	0.053895885	-0.054625715	-0.0597	0.1137
6	0.082879728	0.009351149	0.0105	0.0726
7	-0.033547214	0.023284648	0.0619	-0.0953
8	0.268508398	0.092934199	0.0963	0.1724
9	0.084837916	0.100669707	0.1155	-0.0304
10	0.244124899	0.140981000	0.1213	0.1231
11	0.123663539	0.180283688	0.1155	0.0085
12	0.075831565	0.132114480	0.0999	-0.0238
13	0.057089469	0.125177368	0.0763	-0.0189
14	-0.007948471	0.062159026	0.0465	-0.0541
15	-0.026945599	0.024506741	0.0123	-0.0389
16	-0.028550939	-0.001588885	-0.0245	-0.0036
17	-0.025158589	-0.022150899	-0.0621	0.0374
18	-0.039358198	-0.030003331	-0.0987	0.0598
19	-0.079893478	-0.043240301	-0.1325	0.0531
20	-0.103775213	-0.062046370	-0.1617	0.0584
21	-0.113746825	-0.084193429	-0.1845	0.0713
22	-0.101416387	-0.099707976	-0.1991	0.0982
23	-0.008477275	-0.081853925	-0.2037	0.1958
24	0.105363911	-0.029569144	-0.1965	0.3025

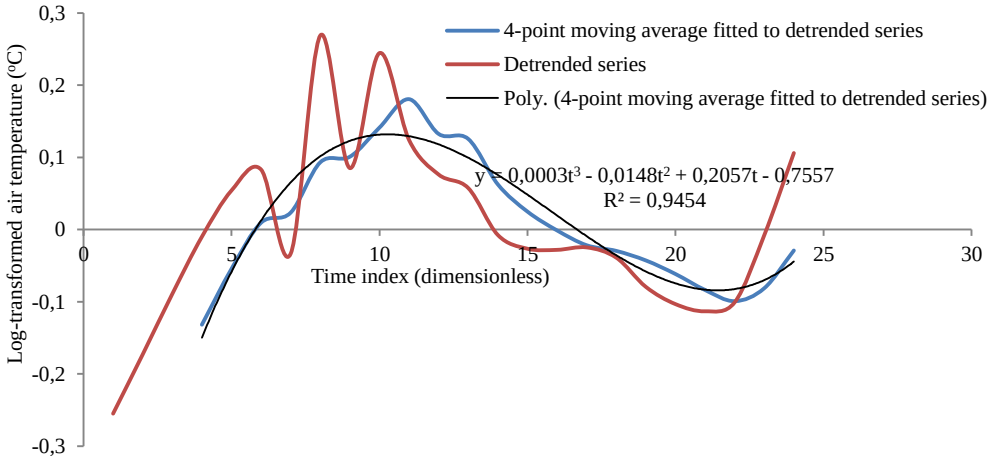


Figure 20: 4-point moving average fitted to detrended series

Step 5: In figure 20, it would be seen that the series obtained after performing the 4-point moving average analysis was smoothened. This resulted in poly-fitted function shown in

figure 20 and in equation (16). Values of poly-fitted function were subtracted from detrended series. The resulting residuals obtained have a mean value of 0 and variance of 0.009. They are also uncorrelated as could be seen in figures 21 and 22 where all autocorrelation and partial-autocorrelation coefficients are within confidence interval limits.

$$0.0003t^3 - 0.0148t^2 + 0.2057t - 0.7557 \qquad (16)$$

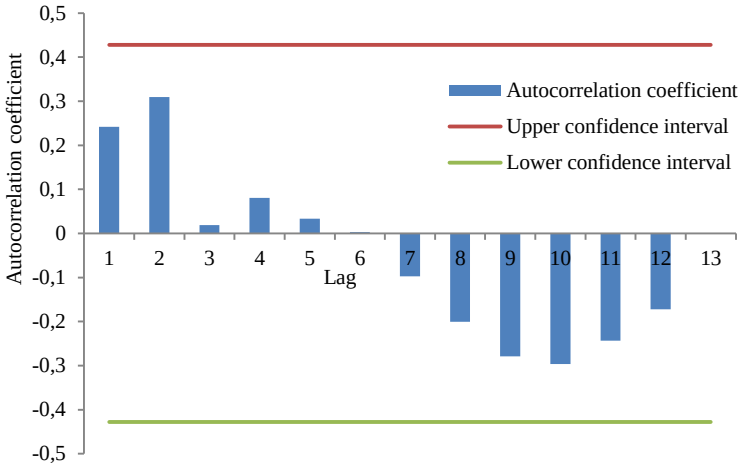


Figure 21: Autocorrelation function plot of residuals

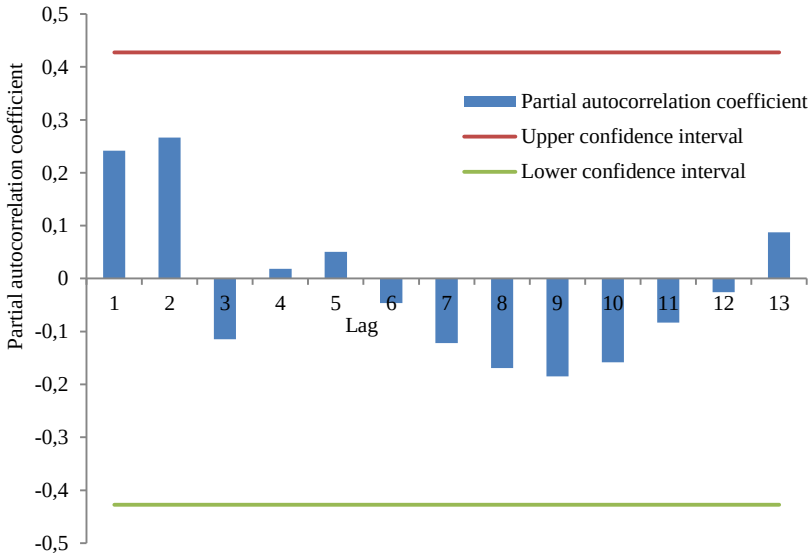


Figure 22: Partial autocorrelation function plot of residuals.

Step 6: Finally, the time series model becomes sum of equations (14) and (16). This results in equation (17) shown below. It should be noted that all values obtained using equation (17) are back-transformed using the inverse-logarithmic function. This is to ensure consistency, in form, with the original time series data.

$$0.0003t^3 - 0.0148t^2 + 0.1838t + 0.7812 \qquad (17)$$

3.2.2 Non-Parametric Approach

Non-parametric approach does not require model selection or assumptions on probability distribution before investigating dataset. It is usually used when information about underlying model or distribution is unclear or unavailable. Consequently, it is free from restricting assumptions about models. It investigates the internal structures, patterns and trends present in time series data. Therefore, it explores to obtain and interpret features that describe the given time series data.

Non-parametric approach has some short comings. First is its lack of continuity. Some statistical methods that are based on non-parametric approach lack continuity. For example, they cannot be used to predict future values. This requirement is essential in engineering applications such as filter operations of autonomous vehicles, where past data are used to estimate and program the future.

Bootstrap and Kernel estimation methods are good examples of methods that are based on non-parametric approach.

3.2.2.1 Bootstrap Method

Bootstrap method is a time series non-parametric approach. It is used to create replicate of a given time series. It computes replicate of time series statistics. Time series statistics whose replicate are computed could be range, mean, etc. Bootstrap method can also be used to estimate confidence intervals for the replicates. Bootstrapping requires that time series data fulfill independently and identically distributed (iid) condition. If time series data are not iid, successive differencing or data transformation are performed until the series fulfills iid requirement.

Bootstrap method comprises of following steps:

1. Ensure time series data are iid. If not, perform successive differencing, etc.
2. Create reshuffled series by reshuffling the differenced time series.
3. Create one-bootstrap replicate of the time series by adding only the first data of the (initial) original time series to the sum of preceding values of the reshuffled series.
4. Compute replicate of test statistics that are of interest.

Example 3.2.2.1 is provided to demonstrate how bootstrap method could be used.

Example 3.2.2.1: In order to ease understanding and reduced data size, consider air temperature data obtained in a Brazilian canopy forest between 0000 and 0600 hours on 2009-09-01. First, it is shown how one-bootstrap replicate can be created. Thereafter, it is shown how mean replicate of the time series is computed. The analysis is presented as follows.

Step 1: Compute mean of the original (initial) time series. In this example, mean is the statistic of interest. Sequence of original time series and its mean are shown below.

[21.33, 20.95, 20.19, 19.81, 19.04, 18.66, 18.28, 17.90, 17.52, 17.52, 17.14, 16.76, 16.38]
Mean = 18.5754

Step 2: Compute successive differences. In this case, first difference is sufficient. The result is shown below.

[-0.38, -0.76, -0.38, -0.77, -0.38, -0.38, -0.38, -0.38, 0, -0.38, -0.38, -0.38]

Step 3: Reshuffle the differenced series.

[-0.77, -0.38, -0.38, -0.38, 0, -0.38, -0.38, -0.38, -0.76, -0.38, -0.38, -0.38]

Step 4: Create one-bootstrap replicate of original (initial) time series.

[21.33, 21.33+(-0.77), 21.33+(-0,77)+(-0.38), 21.33+(-0.77)+(-0.38)+(-0.38), …]
This simplifies to sequence shown below:

[21.33, 20.56, 20.18, 19.8, 19.42, 19.42, 19.04, 18.66, 18.28, 17.52, 17.14, 16.76, 16.38]

Similarly, replicate of mean of original (initial) series becomes 18.8069.

Bootstrap method can also be applied by sampling in groups or blocks. In this case, random samples of blocks are selected. The samples are mixed to form bootstrap time series. The process is repeated numerous times so as to ensure that several bootstrap time series are obtained. These artificial series (i.e. the obtained bootstrap series) are then used to investigate properties of the original time series [38].

3.2.2.2 Kernel Estimation Method

Kernel estimation method is another time series non-parametric approach. It assumes that both conditional mean and variance depend on number of lags within the series. It estimates non-parametrically, the probability distribution function of a given time series dataset. It also estimates the series cumulative distribution function.

Considering sample of n observations taken from time series dataset of unknown distribution, kernel estimator of probability and cumulative distribution functions (PDF and CDF) are defined in equations shown below [39].

$$f_T(y) = \frac{1}{nh}\sum_{i=1}^{n} k\frac{(y - y_i)}{h} \qquad (18)$$

$$F_T(y) = \frac{1}{n}\sum_{i=1}^{n} H\frac{(y - y_i)}{h} \qquad (19)$$

where:
$f_T(y)$ represents kernel estimated PDF,
$F_T(y)$ represents kernel estimated CDF,
n represents number of observations,
h represents bandwith,
$H(.)$ represents kernel which takes the form of a CDF,
$K(.)$ represents the kernel - a bounded PDF that is symmetrical about origin.

The kernel - *K(.)* can be rectangular, quadratic or Gaussian weighting function. This implies that the weighting function is imposed on each observed point on x-axis.

Consequently, PDF estimate at a given point is the product of *(1/n)* and total height component of the weighting function [40]. However, choice of bandwidth is more important than choice of kernel. Therefore, degree of smoothing is varied by changing values assigned to bandwidth.

Example 3.2.2.2: Using air temperature series discussed in preceding example (Example 3.2.2.1), together with equation (18), kernel probability density distribution was computed.

Rectangular weighting function was considered. For standardized rectangular weighting function, Kernel function - $K(t)$ is always equal to $\frac{1}{2}$ [40]. Bin-width was determined as the quotient between range and number of bins. Range is the difference between maximum and minimum values in the series. Thus, 5 bins gave 0.99 bin-width while 3 bins gave 1.65 bin-width. The choice of number of bins was guided by Freedman-Diaconis' rule which is described in equation below [41].

$$B = 2 * IQR * N^{\left(-\frac{1}{3}\right)} \qquad (20)$$

where:

B represents bin-width,

IQR represents interquartile range,

N represents data size.

Selecting kernel bandwidth is an area of considerable research. Therefore, kernel bandwidth was computed using the conventional approach of iterating for bandwidth with start value of half bin-width. For each iteration, kernel density plot obtained was compared with histogram plot of the density function.

Histogram and kernel density plots for 5 bins are shown below. Bandwidth was found to be approximately 0.5.

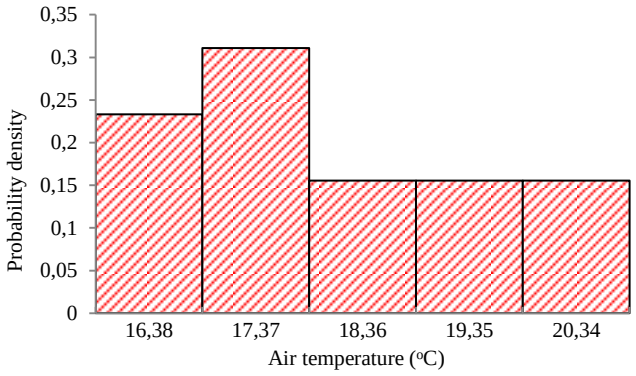


Figure 23: Histogram plot of density function for air temperature

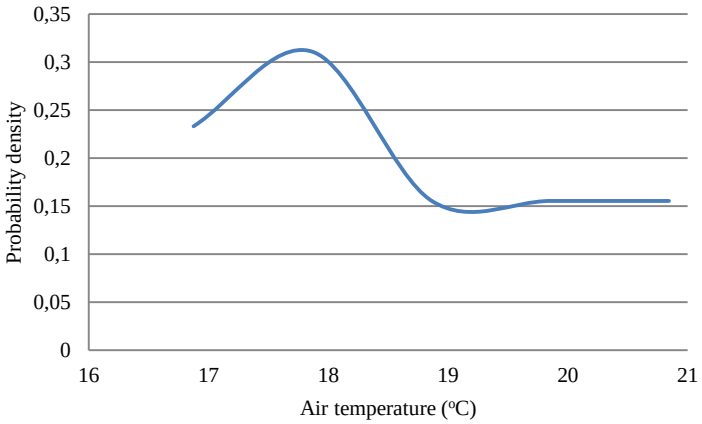


Figure 24: Kernel density plot for air temperature (bandwidth =0.5)

Using bandwidths greater than 0.5 does not give better density functions (see figure 25). They result in reduced area between density function and x-axis. Similarly, bandwidths less than 0.5 do not give better density functions. They result in increased area between density function and x-axis (see figure 25). The area between density function and x-axis is 0.83 for bandwidth of 0.6, and 1.7 for bandwidth of 0.3. Therefore, they do not fulfill important characteristics of probability density function which requires that area between density function and x-axis be equal to one.

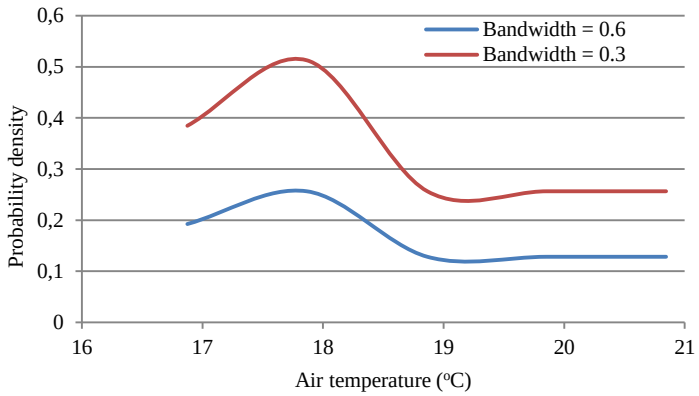


Figure 25: Kernel density plot for air temperature (bandwidths of 0.6 and 0.3)

3.2.3 Parametric Approach

State-space, artificial neural network and Box-Jenkins methods are considered as time series parametric approaches. Others are regression and variogram methods. They all require the use of models. Choice of model is usually made during start of analysis. This is usually done after a preliminary (also called preceding) analysis.

In dynamic systems where outputs change with time, models can be deterministic or stochastic. Deterministic models have their output determined only by past observations with no consideration to errors. Stochastic models take into account state-space and probability distribution of the parameters. As a result, stochastic models offer better predictive capability than deterministic models. Methods such as state-space, artificial neural network and Box-Jenkins methods are stochastic while variogram method is deterministic.

As general statement, dynamic stochastic models are seen as deterministic concepts that take into account stochastic information of uncertainty available for the concept. According to Gelfan [42], four groups of dynamic stochastic models exist. They are:

1. Dynamic-stochastic models with random inputs. These are responsible for stochasticity in temporal variations of meteorological factors such as air temperature, air humidity, etc.
2. Dynamic-stochastic models that are responsible for stochasticity in spatial variations of factors such as soil properties, topography, etc.
3. Dynamic-stochastic models that are responsible for stochastic properties in parameters used in describing relationships and processes.
4. Dynamic-stochastic models that are responsible for stochastic characteristics of measurements used in developing other models.

3.2.3.1 Regression

According to Helsel and Hirsch [43], regression can be performed due to following reasons:

1. In order to investigate relationship between two variables.
2. In order to remove variations in certain portions of measured variable.
3. In order to estimate values of a variable based on knowledge of another variable.

Regression involves variables of interest. The variables are called response and explanatory variables. If one explanatory variable is involved in a linear regression, the linear regression becomes a simple linear regression. However if more than one explanatory variable are involved, it becomes multiple linear regression. Simple linear regressions are represented by model equation described below.

$$Y = a + bX + \varepsilon \qquad (21)$$

where:
Y represents observations of the response (also called dependent) variable,
X represents observations of the explanatory (also called independent) variable,
a represents intercept on the ordinate,
b represents the slope,
ε represents the residuals.

Table 9 presents assumptions of linear regression and their applicability.

Table 9: Assumptions of linear regression (culled from [43] and [4])

	Purpose			
Assumption	Predict Y given X	Predict Y and Variance for the prediction	Obtain best linear unbiased estimator of Y	Test hypotheses, estimate confidence or prediction intervals
Model form is correct.	+	+	+	+
Data used to fit model represents data of interest.	+	+	+	+
Variance of residuals is constant (homoscedastic)		+	+	+
Residuals are independent.			+	+
Residuals are normally distributed.				+

Key: + indicates assumption is required.

Equation (21) is also used to test for trend. The null hypothesis is that slope, b is equal to zero. If slope is non-zero, the null hypothesis is rejected. Rejection of the null hypothesis means linear trend exists [43].

One of the methods of estimating coefficients of linear regression equation is least squares method. Least squares method aims to minimize sum of squared residuals from the regression line [4]. It is described in equation below.

$$\sum_{t=1}^{n} \varepsilon^2 = \sum_{t=1}^{n} (Y_t - a - bX_t)^2 \qquad (22)$$

where:

ε represents residuals,

Y represents observations of the response variable,

X represents observations of the explanatory variable,

a represents intercept on ordinate,

b represents slope.

It is important to mention an important measure used in regression analysis. It is called coefficient of determination of residuals or index determination of residuals. It is usually represented as R^2. It represents fraction of initial (original) variance of the series that is fitted with model. It is shown below.

$$R^2 = 1 - \frac{var(R_t)}{var(X_t)} \qquad (23)$$

where:

X_t represents the initial (orginial) series,

R_t represents residuals of the initial series,

$var(X_t)$ represents variance of the initial time series,

$var(R_t)$ represents variance of the residuals.

Value of R^2 varies from 0 to 1; where 0 indicates that fitted model is of no importance. R^2 value of 1 indicates that fitted model is of significant importance. An example has been provided to demonstrate how linear regression is used to test and identify trends in time series. The example also demonstrates how a linear model could be fitted to the series.

Example 3.2.3.1: Consider the time series discussed and used in Examples 3.1.4 and 3.2.1. The series is been used to demonstrate linear regression. Table 10 presents the series and results obtained.

Table 10: Linear regression analysis

Time index (X)	Air temp. (^{o}C) (Y)	XY	X^2	Y_{trend}	$Y - Y_{trend}$
1	18.2	18.2	1	30.61834	-12.41834
2	20.9	41.8	4	29.71204	-8.81204
3	24.1	72.3	9	28.80574	-4.70574
4	27.5	110.0	16	27.89944	-0.39944
5	30.3	151.5	25	26.99314	3.30686
6	30.8	184.8	36	26.08684	4.71316
7	22.4	156.8	49	25.18054	-2.78054
8	42.7	341.6	64	24.27424	18.42576
9	26.6	239.4	81	23.36794	3.23206
10	36.5	365.0	100	22.46164	14.03836
11	26.3	289.3	121	21.55534	4.74466
12	22.4	268.8	144	20.64904	1.75096
13	20.4	265.2	169	19.74274	0.65726
14	16.7	233.8	196	18.83644	-2.13644
15	15.2	228.0	225	17.93014	-2.73014
16	14.4	230.4	256	17.02384	-2.62384

17	13.8	234.6	289	16.11754	-2.31754
18	12.7	228.6	324	15.21124	-2.51124
19	11.0	209.0	361	14.30494	-3.30494
20	9.9	198.0	400	13.39864	-3.49864
21	9.2	193.2	441	12.49234	-3.29234
22	9.0	198.0	484	11.58604	-2.58604
23	10.6	243.8	529	10.67974	-0.07974
24	13.1	314.4	576	9.77344	3.32656
$\sum X$= 300	$\sum Y$ = 484.7	$\sum(XY)$ = 5016.5	$\sum X^2$ = 4900		$\sum$(Y - Y_{trend}) = 39.6

From equation of least square shown in equation (22), it implies that equations (24) and (25) shown below hold for intercept and slope.

$$\sum_{t=1}^{n} Y_t = na + b\sum_{t=1}^{n} X_t \qquad (24)$$

$$\sum_{t=1}^{n} (X_t * Y_t) = a\sum_{t=1}^{n} X_t + b\sum_{t=1}^{n} (X_t)^2 \qquad (25)$$

Using equations (24) and (25), the regression line coefficients were determined. This has been demonstrated in Table 10 above. For the example used in Table 10, intercept was calculated to be 31.5246 and slope calculated to be -0.9063. Also, the figure below shows plot of trend line and original time series.

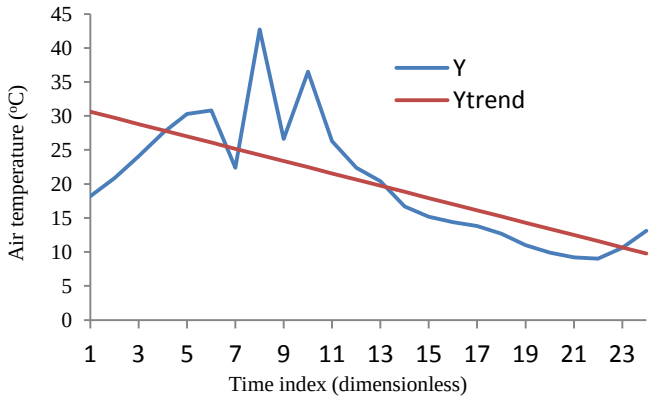


Figure 26: Fit of trend using linear regression method

Coefficient of determination was computed using equation (23). It was calculated to be 0.4985. Also, the residuals were normally distributed with mean of 0 and variance of 39.6.

Multiple linear regression analysis involves two or more explanatory (independent) variables. The equation below shows multiple regression of Y on X_1, X_2, X_3,…, X_k in standard form.

$$Y = b_0 + b_1X_1 + b_2X_2 + b_3X_3 + \cdots + b_kX_k + \varepsilon \qquad (26)$$

where:

b_0 represents intercept,

$b_1, \ldots, b_k$ represent model coefficients,

ε represents regression residual.

For nonlinear regression, dependent variable is a nonlinear function of model parameters. It is also a nonlinear function of independent variables. Commonly used nonlinear regression models include:

1. Asymptotic growth regression model. Its regression function is shown below.

$$\mu_Y(x) = \beta_1 + \beta_2 e^{(\beta_3 x)} \quad (27)$$

$\beta_1, \beta_2, \beta_3$ represent model parameters that are to be estimated.

2. Asymptotic decay regression model. Its regression function is shown below.

$$\mu_Y(x) = \beta_4 - \beta_5 e^{-(\beta_6 x)} \quad (28)$$

$\beta_4, \beta_5, \beta_6$ represent model parameters that are to be estimated.

3. Logistic regression model. Its regression function is shown below.

$$\mu_Y(x) = \frac{\beta_7}{1 + e^{-(\beta_8 + \beta_9 x)}} \quad (29)$$

$\beta_7, \beta_8, \beta_9$ represent model parameters that are to be estimated.

4. Gompertz regression model. Its regression function is shown below.

$$\mu_Y(x) = \beta_{10} e^{-e^{-(\beta_{11} + \beta_{12} x)}} \quad (30)$$

$\beta_{10}, \beta_{11}, \beta_{12}$ represent model parameters which are required to be estimated.

Other nonlinear regression models include: Weibull model, bi-exponential model, etc. According to Ratkowsky [44], main assumptions of nonlinear regression model include:

1. Mean of subpopulations of dependent (response) variable is a nonlinear function of model parameters.
2. Standard deviation of all subpopulations of the dependent variable is constant.
3. Every subpopulation of the dependent variable is Gaussian.
4. Sample size is selected either by random sampling or sampling with preselected values of the independent variable.
5. All sampled values are observed without error. This implies that usual (common) methods used in determining model parameters are applicable. However, if errors are not negligible, inferential analysis methods may be applied.

From above assumptions, it would be seen that assumptions of nonlinear regression are the same as those for linear regression with exception of first assumption. Solutions to nonlinear regression models are usually obtained using methods such as, transformation to linear models, methods of separable least squares and iterative methods such as use of Gauss-Newton algorithm.

3.2.3.2 The Variogram

The variogram is used to estimate the covariance structure of a time series. It measures the variation between values of a given time series. It is described in below equation.

$$2\gamma(h) = var(x_t - x_{t+h}) \quad (31)$$

where;

$2\gamma(h)$ represents the variogram,
x_t and x_{t+h} represent time series data at time points, i and $(i+h)$,
var represents variance.

Time series data are more similar when there is little variability in their difference, and less similar when there is greater variability in their difference. The improved equation for computing classical variogram is described in below equation [45].

$$\gamma(h) = \frac{1}{2N}\sum_{t=1}^{N(h)}(x_t - x_{t+h})^2 \qquad (32)$$

where;
x_t represents time series data at time index *t*,
$\gamma(h)$ represents the semivariance at separation distance *h*,
N represents number of data in the set *N(h)*,
cov represents covariance.

Variograms have low values at small separation distance. Their value increases with increasing distance until plateau is reached. The plateau is called sill and defined as point at which variogram reaches zero correlation distance [46]. The separation distance for which data are correlated is called range. Beyond the range is sill, and data are no longer correlated. Thus, range is defined as separation distance at which variogram reaches sill. The numerical value of variogram at zero separation distance is called nugget. Above-described terms are illustrated in figure below.

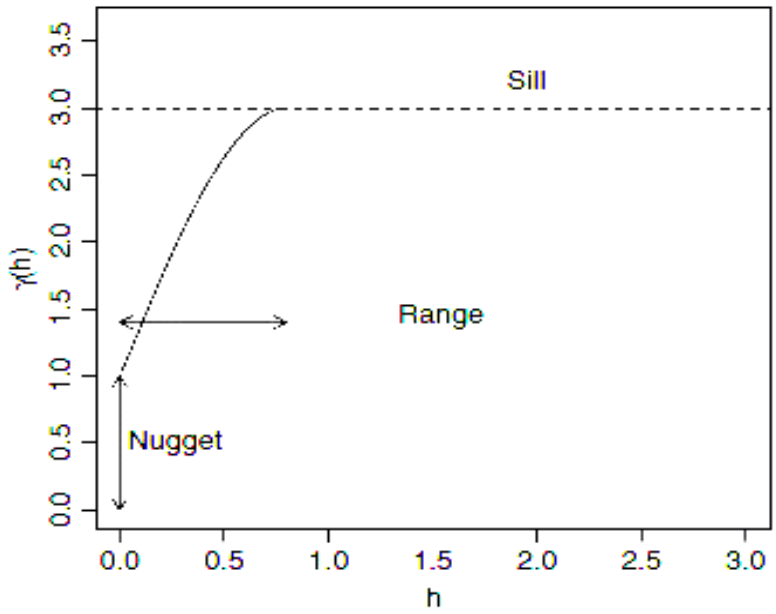


Figure 27: Semivariogram of a stationary process with nugget effect (Culled from [45]).

Variogram model of given time series is usually unknown. Therefore, following computational steps are required in order to determine it.

1. Perform preceding analysis by checking if given time series is stationary.
2. Determine experimental variogram (variogram cloud).
3. Process and determine values of classical variogram estimator.

4. Fit model to values of classical variogram estimator.
5. Validate variogram model.

Preceding analysis consists of check to determine if given time series is stationary. In order obtain appropriate variogram models, the analysis requires the given time series to be stationary. If the given time series is not stationary, logarithmic or square root transformation of data is performed until stationarity is achieved.

Experimental variograms are usually determined using time series data. As noted by Trauth [46], experimental variograms describe trends in time series data. This is useful in detecting outliers or anomalies. One shortcoming of experimental variogram is its difficulty in performing correlation analysis. To overcome this shortcoming, experimental variograms are replaced with classical variogram estimators.

Classical variogram estimators are computed using equation (32) described earlier. They are plotted and model fitted to them. Various variogram models exist. They include:

1. Power variogram model: This takes the form shown in equation below. If $a = 1$, it becomes a linear variogram model. Thus, linear variogram model is a special case of power variogram model.

$$\gamma(h) = c_0 + bh^a \qquad (33a)$$

where;
$\gamma(h)$ represents semivariance at separation distance *h*,
c_0 represents nugget,
b represents slope,
a represents index of the model.

2. Exponential variogram model: This takes the form shown in equation below.

$$\gamma(h) = c_0 + c_1\left(1 - e^{-\left(\frac{h}{a}\right)}\right) \qquad (33b)$$

where;
c_0 represents nugget,
$(c_0 + c_1)$ represent sill,
h represents separation distance,
a represents index of the model.

3. Spherical variogram model: This takes the form shown in either equation (33c) or (33d).

$$\gamma(h) = c_0 + c_1\left(\frac{3h}{2a} - \frac{1}{2}\left(\frac{h}{a}\right)^3\right) \text{ for } 0 < h \leq a \quad (33c)$$

$$\gamma(h) = c_0 + c_1 \text{ for } h \geq a \qquad (33d)$$

where;
c_0 represents nugget,
$(c_0 + c_1)$ represent sill,
h represents separation distance,
a represents index of the model.

Once variogram model is selected (i.e. chosen) and drawn (i.e. fitted), parameters of the model are determined using known values $\gamma(h)$ and h that are on fitted model. They are used to setup system of equations involving the unknown parameters. The number of equations is determined by the number of unknown parameters. Read-out values of sill and nugget from model, sometimes, help in simplifying the equations. Another method of determining values of variogram model parameters is use of ordinary least square method discussed in section 3.2.3.1. It evaluates for values of the parameters that minimize the sum of squared residuals. Iterative method is another method used in determining values of variogram model parameters.

As noted by Burgess and Webster [47], and Oliver and Webster [48], an optimal sampling strategy can be determined through plot of sampling frequency against kriging variance. This requires estimation of kriging variance or prediction error of blocks. Estimating kriging variance requires knowledge of classical variogram estimator, kriging weight and block size.

Example 3.2.3.2 has been provided to illustrate how empirical variogram is manually computed.

Example 3.2.3.2: Consider stationary time series shown in Table below. Variogram analysis is performed and detailed below.

Table 11: Variogram analysis

X_i	$(X_i-X_{i+1})^2$	$(X_i-X_{i+2})^2$	$(X_i-X_{i+3})^2$	$(X_i-X_{i+4})^2$	$(X_i-X_{i+5})^2$	$(X_i-X_{i+6})^2$	$(X_i-X_{i+7})^2$	$(X_i-X_{i+8})^2$	…	$(X_i-X_{i+13})^2$
0.8417	63.9088	210.7549	387.3378	556.7901	695.667	791.2856	839.313	841.6033		444.7839
8.836	42.5508	136.5766	243.4255	337.8685	405.4384	440.0177	441.6765	414.7577		113.8297
15.359	26.6617	82.4282	140.6145	185.2974	208.9037	210.0470	191.6148	159.1029		2.886941
20.523	15.3311	44.8177	71.3839	86.30410	87.03957	75.32504	55.50399	33.11312		33.87356
24.438	7.7234	20.5517	28.8853	29.3114	22.69093	12.49340	3.381553	0.044944		141.3364
27.217	3.0776	6.7361	6.9427	3.937843	0.57078	0.883976	8.946679	27.93757		272.4513
28.972	0.7074	0.7755	0.0529	0.997601	7.26033	22.51882	49.56019	89.61273		387.0230
29.813	0.0016	0.3733	3.3852	12.50047	31.20898	62.11016	106.2446	162.6747		457.4850
29.852	0.4232	3.5321	12.7813	31.65188	62.73432	107.0604	163.6839	229.5104		468.9000
29.202	1.5102	8.5533	24.7556	52.85290	94.02211	147.4622	210.2239	277.2891		418.9595
27.973	2.8754	14.0370	36.49489	71.70025	119.1263	176.0982	237.8720	297.9628		317.9838
26.277	4.2062	18.8825	45.85863	84.98627	133.9691	188.4415	242.2972	288.2661		
24.226	5.2647	22.2878	51.37879	90.69896	136.3407	182.6552	222.8303	249.7158		
21.932	5.8879	23.7500	52.25989	88.02192	125.8996	159.5927	182.4634	188.6118		
19.505	5.9873	23.0650	48.37898	77.33444	104.1726	122.7974	127.8505	116.0360		
17.058	5.5493	20.3275	40.28568	60.21139	74.55459	78.50314	69.30729	47.85319		
14.703	4.6350	15.9313	29.20214	39.42333	42.30852	35.63374	20.81093			
12.550	3.3801	10.5690	17.02305	18.93642	14.56567	5.803281				
10.711	1.9952	5.2322	6.315672	3.912484	0.325470					
9.299	0.7655	1.2113	0.31979	0.708964						
8.424	0.0509	0.0957	2.947746							
8.198	0.2863	3.7737								
8.733	1.9811									
10.141										
Sum (S)	204.7607	674.2625	1250.03	1833.447	2366.799	2818.73	3173.581	3424.092		3059.513
N(h)	23	22	21	20	19	18	17	16		11
$\gamma(h)=S/2h$	4.4513	15.3241	29.76262	45.83617	62.28417	78.29805	93.34061	107.0029		139.0688
Time index	1	2	3	4	5	6	7	8		13

Figure 28a shows plot of the variogram estimator. Figure 28b shows a power model fitted to variogram estimator while figure 28c shows an exponential model fitted to variogram estimator.

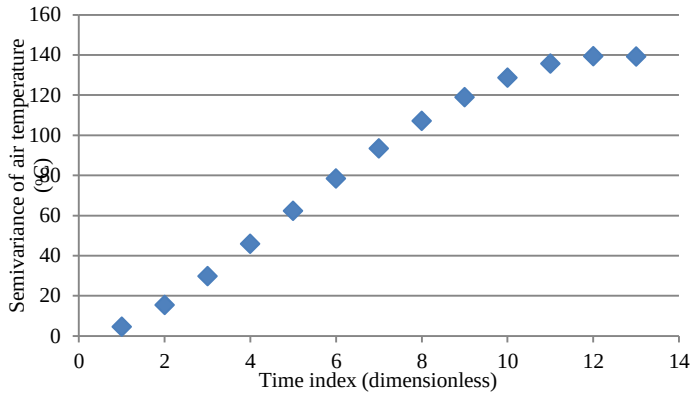


Figure 28a: Plot of classical variogram estimator

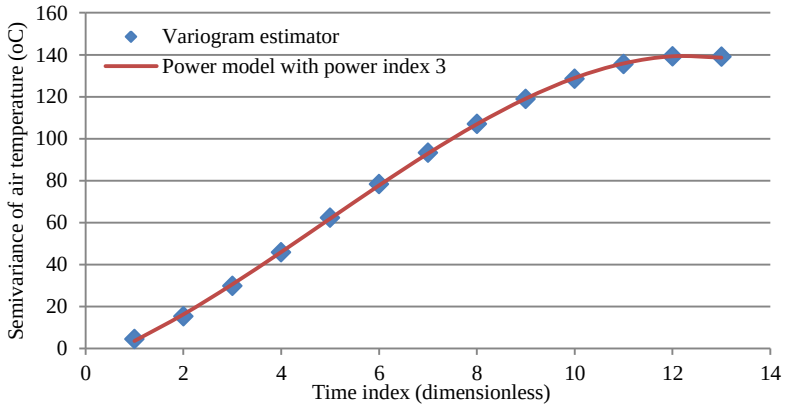


Figure 28b: Variogram estimator fitted with power model

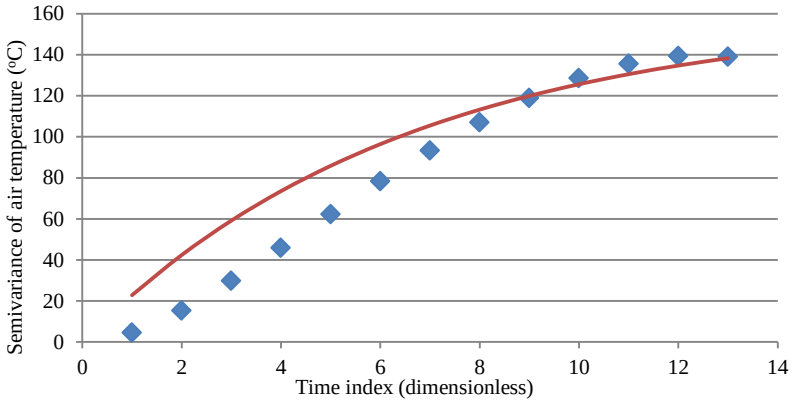


Figure 28c: Variogram estimator fitted with exponential model

Fitted power variogram model is described in equation (33e) while that for exponential variogram model is described in equation (33f).

$$\gamma(t) = -0.0954t^3 + 1.3978t^2 + 9.1466t - 6.8993 \qquad (33e)$$

$$\gamma(t) = 160\left(1 - e^{-\left(\frac{t}{6.5}\right)}\right) \qquad (33f)$$

It would be seen that the power variogram model fits better than the exponential variogram model. It gave an R-squared value of 99.99% while exponential variogram model gave an R-squared value of 92.68%. Both models have sill of 140. Power variogram model has a range of 13 while that for exponential variogram model is 12.

3.2.3.3 State-Space Methods

State-space methods model time series data into distinct components made up of trends, seasonality, regression elements and noise. These components when put together constitute a state-space model. This differs from Box Jenkins models where components are removed through differencing or transformations so that stationarity is achieved. Parameters of state-space model are often called hyper-parameters. Hyper-parameters are estimated using methods such as maximum likelihood estimation, expectation maximization (EM) algorithm, etc. [49].

The simplest of all state-space models is local level model. Local level model is basically an exponentially weighted moving average. Exponentially weighted moving average (EWMA) measures the degree of variation in time series data. It applies non-uniform weightings to a given time series. Its statistic is computed using the equation below [50].

$$E_t = \lambda X_t + (1-\lambda)E_{t-1}\,;\; t = 1, 2, \dots, n \qquad (34)$$

where:

E_t represents calculated exponentially weighted moving average at time *t*,

E_{t-1} represents exponentially weighted moving average at time *t-1*,

X_t represents time series observation at time *t*,

λ represents a constant which varies from 0 to 1.

λ values are usually between 0.2 and 0.3. High values of λ imply recent data are given more weights. λ values can be obtained from EWMA look-up table.

Example 3.2.3.3 illustrates how EWMA statistic (estimator) is computed. It also shows how EWMA model is fitted to time series data.

Example 3.2.3.3: Consider time series data described in Example 3.2.2.1. Using equation (34), exponentially weighted moving average statistic (E_t) is computed for each time-point. The series is shown in Table 12 where the EWMA analysis is summarized using $\lambda = 0.3$.

Table 12: Exponentially weighted moving average analysis

t	1	2	3	4	5	6	7	8	9	10	11	12	13
X_t	21.3	21.0	20.2	19.8	19.0	18.7	18.3	17.9	17.5	17.5	17.1	16.8	16.4
E_t	19.4	19.9	20.0	19.9	19.7	19.4	19.0	18.7	18.3	18.1	17.8	17.5	17.2

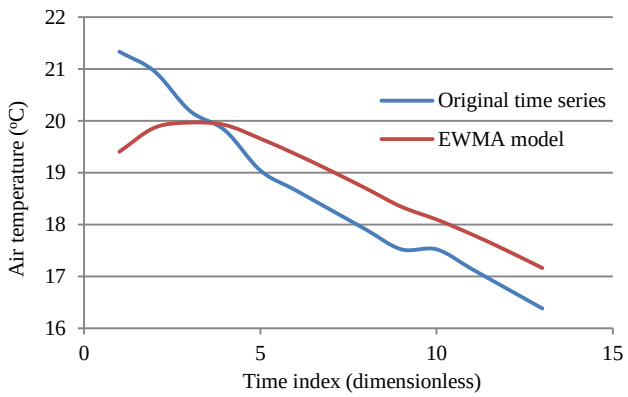


Figure 29: EWMA model for air temperature series

The EWMA model gave an R-squared value of 71.3%. This is high enough; thus, the model can be used fit missed and future values to the series.

Another state-space model is the Markov model. It is used to model randomly changing systems. Markov models are produced from discrete-time or continuous-time processes. Discrete-time Markov process is also called discrete-time Markov chain. It has finite discrete state-space and possesses Markov property.

Definition 3.2.3.3: Let $\{X_t: t \geq 0\}$ be states of a stochastic process with finite state space $S \subset \{0,1,2,\ldots\}$, the stochastic process is discrete-time Markov chain if it possesses Markov property defined as follows:

$$P(X_{t+1} = j | X_t = i; X_{t-1} = i_{t-1}, \ldots, X_0 = i_0) = P(X_{t+1} = j | X_t = i) \qquad (35a)$$

for all $t \geq 0; i, j, i_0, i_1, \ldots, i_{t-1} \subset S$;

where;

$P(X_{t+1} = j | X_t = i)$ represents probability of the next state, given the current state.

Thus, Markov property implies that probability of the next (future) state, given the current (present) state and the entire past states, depends only on the current state. Components of discrete-time Markov chain include set of states, set of transitions between states, and special start and end states. Transition between states is represented with an arc. The arc is associated to transition weight which defines the transition probability. Sum probability on all arcs leaving a state is equal to one.

Continuous-time Markov process, just like discrete-time Markov process, has states' transitions independent of the past. However, when in a certain state, it can remain in that state for an exponentially distributed amount of time before changing state. This is one of the differences between continuous-time Markov chain and discrete-time Markov chain. Another distinguishing factor is how time is being treated. Time is discrete in discrete-time Markov chain. This means time can take values such as {1, 2, …}. On the other hand, time is continuous in continuous-time Markov chain. This means time can take values like [0, ∞).

Hidden Markov model (HMM) is one of the variants of Markov chain. It is used in time series analysis. HMM is a sequence model that allocates each time series data to a class or group. This results in generation of sequence of classes. The classes then act as new states. The new states are not directly observable. They are hidden or said to be latent. HMM therefore constitutes a baseline model for treating sequential or temporal data. It is useful for state-estimation of model parameters and also for inferential analysis of data [51].

In order to understand the fundamentals of HMM, consider two different sets containing random values that are discrete, real and finite. The random values can be presented in Trellis diagram as shown in figure 30.

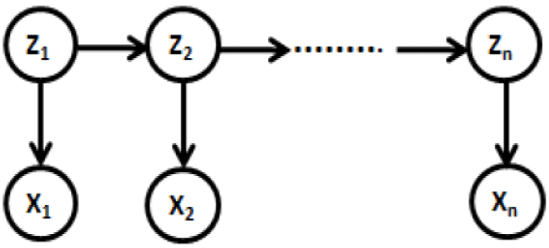


Figure 30: Trellis diagram for HMM representation

$(X_1, X_2, \dots X_n)$ are the observed values while $(Z_1, Z_2, \dots Z_n)$ are the hidden or latent values. In order to determine probability of the joint distribution, probability of hidden values and conditional probability of observed values are required. This is represented below.

$$P(X_1, \dots X_n, Z_1, \dots Z_n) = \big(P(X_1)\big)\Big(P\big(X_1 \,|\, Z_1\big)\Big)\prod_{t=2}^{n}\Big(P\big(Z_t \,|\, Z_{t-1}\big)\Big)\Big(P\big(X_t \,|\, Z_t\big)\Big) \quad (35b)$$

Three types of probabilities are contained in equation (35b). They are:

1. Initial state probability such as $P(X_1)$.
2. Emission probability such as $P\big(X_1 \,|\, Z_1\big)$. It is also called observation probability and represents the probability of each state within the set of observed values.
3. Transition probability such as $P\big(Z_t \,|\, Z_{t-1}\big)$. It represents the probability of moving from one state to the other.

According to Deng et al [37], Chen et al [52] and Box et al [53], state-space models when used in time series analysis are more expensive, less accurate and more computation-demanding than Box-Jenkins method and artificial neural networks. Therefore greater efforts were made elaborating Box-Jenkins methodology and artificial neural networks.

3.2.3.4 Box-Jenkins Method

Box-Jenkins method is another time series parametric approach. It is a systematic method of identifying, fitting, checking, and using integrated autoregressive, moving average time series models. It is appropriate for analyzing time series of medium to long length - at least 50 observations [54].

Box-Jenkins method adjusts time series to ensure stationarity through successive differencing of data. By so doing, it unravels autocorrelation and partial autocorrelation structures of the time series. This is essential since autocorrelation and partial autocorrelation analyses are vital in Box Jenkins methodology.

Box-Jenkins method produces stationary models. The resulting residuals obtained are stationary and independent. Box-Jenkins method does not require that resulting residuals be normal distributed. Residuals in Box-Jenkins analysis can assume any probability distribution such as Laplace probability distribution, etc. Autoregressive integrated moving average (ARIMA) model and its variants are based on Box-Jenkins methodology. Variants of ARIMA include autoregressive (AR), moving average (MA), autoregressive moving average (ARMA), etc. The mathematical equation below describes Box-Jenkins model:

$$X_t = \sum_{i=1}^{p} \Phi_i X_{t-i} + \epsilon_t - \sum_{j=1}^{q} \theta_j \epsilon_{t-j} \quad (36a)$$

where;

X_t represents predicted value,

Φ_i, θ_j represent model parameters

X_{t-i} represents observed time series values

ϵ_{t-j} represents uncorrelated random variables (residuals) with zero mean and finite variance.

In simplified form, the equation becomes:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \epsilon_t - \theta_1 \epsilon_{t-1} - \theta_2 \epsilon_{t-2} - \cdots - \theta_q \epsilon_{t-q} \quad (36b)$$

where;
X_t represents predicted value,
$X_{t-1}, X_{t-2}, \ldots X_{t-p}$ represent observed time series values,
$\Phi_1, \Phi_2, \ldots \Phi_p, \theta_1, \theta_2, \ldots \theta_q$ represent model parameters,
$\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-q}$ represent uncorrelated random variables with zero mean and finite variance.

In compact form, the equation becomes:

$$\Phi_p(B) X_t = \theta_q(B) \epsilon_t \quad (37)$$

where;
B represents backward difference operator, such that:

$$(B) X_t = X_{t-1} \quad (38a)$$

$$(B^n) X_t = X_{t-n} \quad (38b)$$

The above Box-Jenkins equations require that time series be stationary. If time series is not stationary, differencing of the series is required. Thus, the equations become:

$$\Phi_p(B)(1-B)^d X_t = \theta_q(B) \epsilon_t \quad (39)$$

$$\left(1 - \Phi_1 B - \Phi_2 B^2 - \cdots \Phi_p B^p\right)(1-B)^d X_t = \theta_q B = \left(1 - \theta_1 B - \theta_2 B^2 - \cdots \theta_q B^q\right) \epsilon_t \quad (40)$$

Box-Jenkins model becomes an autoregressive moving (AR) model if d and q are set to zero. The equation becomes:

$$(\Phi_p B) X_t = \epsilon_t \quad (41)$$

The term, $(\Phi_p B)$, represents the characteristic polynomial of the AR(p) process. In the same vein, if d and p are set to zero, Box-Jenkins model becomes an MA(q) process. Its equation is shown below.

$$X_t = (\theta_q B) \epsilon_t \quad (42)$$

Finite MA(q) process is always stationary since absolute values of roots of its characteristic equation are always greater than one.

Furthermore, if only differencing, "*d*" in Box-Jenkins equation is set to zero, it results in ARMA(p,q) process.

$$(\Phi_p B) X_t = (\theta_q B) \epsilon_t \quad (43)$$

Box-Jenkins method has one challenge. That is, in order to update model due to recent changes such as arrival of new data, Box Jenkins method requires that parameters of model be re-estimated. The reason is that variance in Box-Jenkins model is not modeled. In Box-Jenkins models, variance is unconditional and expected to remain constant at all time.

Having discussed Box-Jenkins model equations, Box-Jenkins systematic steps for identifying models, estimating parameters and checking fitted models are discussed. This is followed by an example which demonstrates how it works. The steps are:

Box-Jenkins Step 1: Establish stationarity – check if time series is stationary. If series is not stationary, perform successive differencing or logarithmic transformation in order to establish stationarity.

Box-Jenkins Step 2: Investigate autocorrelation function (ACF) and partial autocorrelation function (PACF) of the stationarized series. If ACF decays exponentially while PACF cuts off after few lags, then the process is an autoregressive (AR) process. On the other hand, if ACF cuts off after few lags while PACF decays gradually, the process becomes a moving average (MA) process. However, if both ACF and PACF decay gradually, an autoregressive moving average (ARMA) model is considered.

Box-Jenkins Step 3: Compute orders of the model and estimate parameters. This is achieved using any of the following methods - sample ACF and PACF analysis, maximum likelihood estimation (MLE) and Akaike information criterion (AIC).

Definition 3.2.3.4a: Let $\{X_t = X_1, X_2, \dots, X_n\}$ be random samples of a variable whose distribution depends on unknown parameters $\{\Phi_i = \Phi_1, \Phi_2, \dots, \Phi_m\}$ with probability mass function $f(X_t; \Phi_1, \Phi_2, \dots, \Phi_m)$, then:

1. Joint probability mass function of $\{X_t = X_1, X_2, \dots, X_n\}$ is given as:

$$L(\Phi_i) = L(\Phi_1, \Phi_2, \dots, \Phi_m) = \prod_{t=2}^{n} f(X_t; \Phi_1, \Phi_2, \dots, \Phi_m) \quad (44a)$$

   where $L(\Phi_i)$ represents the likelihood function.

2. Assuming $[k_1(X_1, X_2, \dots, X_n), k_2(X_1, X_2, \dots, X_n), \dots k_m(X_1, X_2, \dots, X_n)]$ maximizes the likelihood function, then:

$$\widehat{\Phi}_i = \{\widehat{\Phi}_1, \widehat{\Phi}_2, \dots, \widehat{\Phi}_m\} = k_i(X_1, X_2, \dots, X_n) \quad (44b)$$

   where $\widehat{\Phi}_i$ represents the maximum likelihood estimates of Φ_i, for $i = 1, 2, \dots, m$.

Therefore, if the process is assumed to be normally distributed with mean of zero and finite variance, the likelihood function is as shown below.

$$L(\Phi_i, \mu, \sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}} * \left(1 - {\Phi_i}^2\right)^{\frac{1}{2}} * e^{-\left(\frac{S(\Phi_i \mu)}{2\sigma^2}\right)} \quad (44c)$$

where;

$L(\Phi_i, \mu, \sigma^2)$ represents the likelihood function,

$\Phi_i = (\Phi_1, \Phi_2, \dots, \Phi_m)$ represents parameters to be estimated,

μ represents mean,

σ^2 represents variance,

n represents number of samples,

S represents summation.

Obtaining maximum likelihood estimates of the parameters is equivalent to minimizing the likelihood or log-likelihood function. Likelihood function is as shown above. Log-likelihood function in its simplified form is shown below.

$$\ln\left(\frac{1}{\sqrt{2\pi\sigma^2}} * e^{-\left(\frac{SSR}{2\sigma^2}\right)}\right) \quad (44d)$$

where;
LL represents log-likelihood function,
σ^2 represents variance,
SSR represents sum of squares of residuals.
SSR is also called residuals sum of squares. It is described below.

$$SSR = \sum_{t=1}^{n} ((\epsilon_t)^2) \qquad (44e)$$

where;
ϵ_t represents residuals,
SSR represents sum of squares of residuals.

In order to have log-likelihood function minimized, SSR must be minimized. How this is been performed is shown in Example 3.2.3.4 provided at the end of this section.

Definition 3.2.3.4b: Akaike information criterion (AIC) which compares quality of models is defined as shown below.

$$AIC = n * \ln\left(\frac{{\sigma_e}^2}{n}\right) + 2p \qquad (45a)$$

where;
AIC represents Akaike information criterion,
n represents number of observations,
p represents number of model parameters,
${\sigma_e}^2$ represents sum of squares of residuals.

If number of observations divided by number of model parameters $\left(\frac{\mathrm{n}}{\mathrm{p}}\right)$ is less than 40, improved version of AIC known as corrected Akaike information criterion (AICC) is recommended [33].

$$AICC = \left(N * \log\left(\frac{SS}{N}\right)\right) + \left(\frac{2N(p+q+1)}{N-p-q-2}\right) \qquad (45b)$$

where;
N represents number of observations after differencing,
SS represents sum of squares of differences,
p and *q* represent orders of the model.

This has been demonstrated in Example 3.2.3.4 provided at end of this section.

Box-Jenkins Step 4: Perform goodness-of-fit check to ensure model adequately represents the given time series. Check to ensure that residuals are independent and stationary. If residuals are not independent and stationary, Box-Cox transformation may be performed [55].

Box-Jenkins Step 5: Fitted model can be used to predict future values as well as missed values.

Example 3.2.3.4 has been provided to demonstrate above-discussed Box-Jenkins steps. It also demonstrates how maximum likelihood stimation and AIC analysis are performed.

Example 3.2.3.4: Consider air temperature series described in Example 3.2.3.3. The series is shown in Table 13. The series was differenced twice to secure stationarity (Table 13). All activities undertaken for an ARIMA model to be fitted to the series are discussed below.

Step 1: Secure stationarity.

Table 13: Air temperature time series for ARIMA analysis

t	1	2	3	4	5	6	7	8	9	10	11	12	13
X_t	21,3	21,0	20,2	19,8	19,0	18,7	18,3	17,9	17,5	17,5	17,1	16,8	16,4
$\Delta X_t^{(1)}$	N/A	-0,3	-0,8	-0,4	-0,8	-0,3	-0,4	-0,4	-0,4	0,0	-0,4	-0,3	-0,4
$\Delta X_t^{(2)}$	N/A	N/A	-0,5	0,4	-0,4	0,5	-0,1	0,0	0,0	0,4	-0,4	0,1	-0,1

Step 2: Investigate autocorrelation and partial autocorrelation plots - Autocorrelation function (ACF) and partial autocorrelation plots shown below.

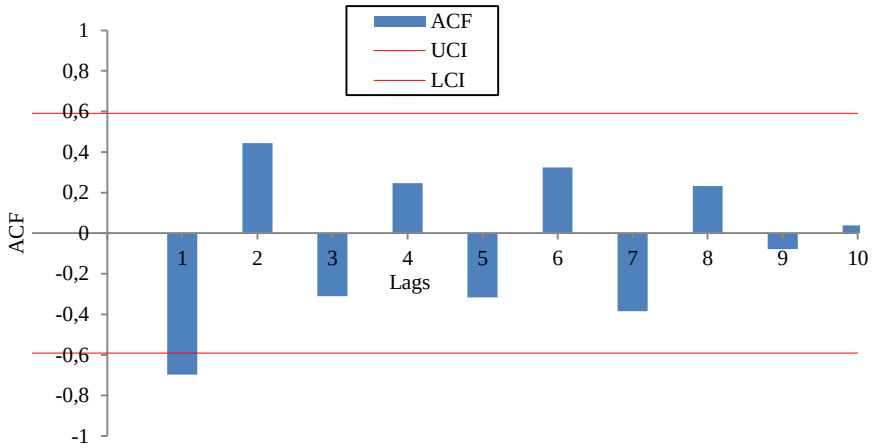


Figure 31: ACF plot of stationarized air temperature series

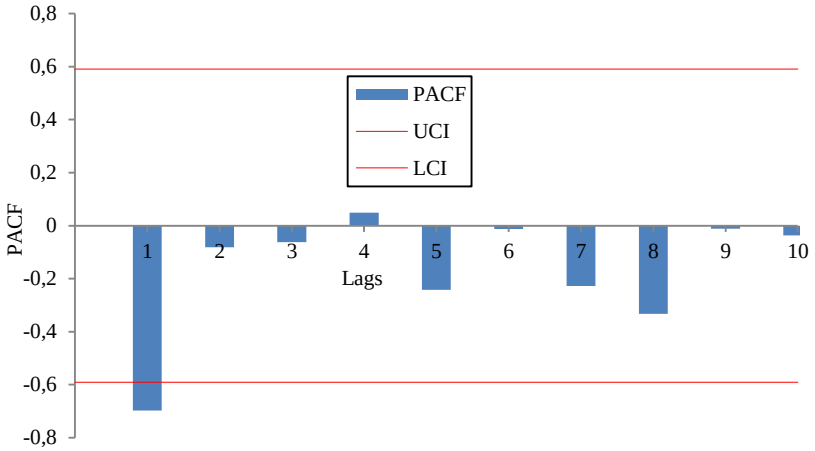


Figure 32: PACF plot of stationarized air temperature series

Step 3: Determine orders of model.

In above ACF plot, only one spike was significant at lag 1. This suggests an AR of order 1. Similarly, PACF plot shows one significant spike at lag 1. This suggests an MA of order 1. Therefore, the time series can be fitted with an ARIMA of orders (*p = 1, d = 2, q = 1*). *d = 2* because the time series was differenced twice.

Besides autocorrelation analysis method, corrected Akaike information criterion (AICC) method was used to verify if above-selected orders were the most suitable. AICC method requires that model with lowest AICC value be selected. AICC values were computed using AICC equation discussed above (equation 45b). Table 14 presents the models and their AICC values.

Table 14: Models and their AICC values

Model (p,d,q)	(1,2,1)	(1,2,2)	(1,2,3)	(1,2,4)	(1,2,5)	(2,2,1)
AICC value	37.23	46.23	61.23	91.23	181.23	58.56

It would be seen in Table 14 that model (1,2,1) has lowest AICC value. Therefore, it should be selected. This validates model selection result obtained earlier using autocorrelation analysis (ACF and PACF) method.

Step 4: Estimate model parameters (i.e. model coefficients).

Maximum likelihood estimation method was used to estimate the model parameters. In order to apply maximum likelihood estimation method, model has to be written in standard form since orders are already known (i.e. $p = 1$, $d = 2$, $q = 1$). Therefore, applying Box-Jenkins equation, it becomes:

$$\Phi_p(B)(1-B)^d X_t = \theta_q(B)\epsilon_t \Rightarrow (1-\Phi_1 B)(1-B)^1 X_t = (1-\theta_1 B)\epsilon_t \quad (46)$$

Simplifying further gives:

$$\epsilon_t = X_t - BX_t - \Phi_1 BX_t + \Phi_1 B^2 X_t + \theta_1 B\epsilon_t \quad (47a)$$

It further simplifies to:

$$\epsilon_t = X_t - X_{t-1} - \Phi_1 X_{t-1} + \Phi_1 X_{t-2} + \theta_1 \epsilon_{t-1} \quad (47b)$$

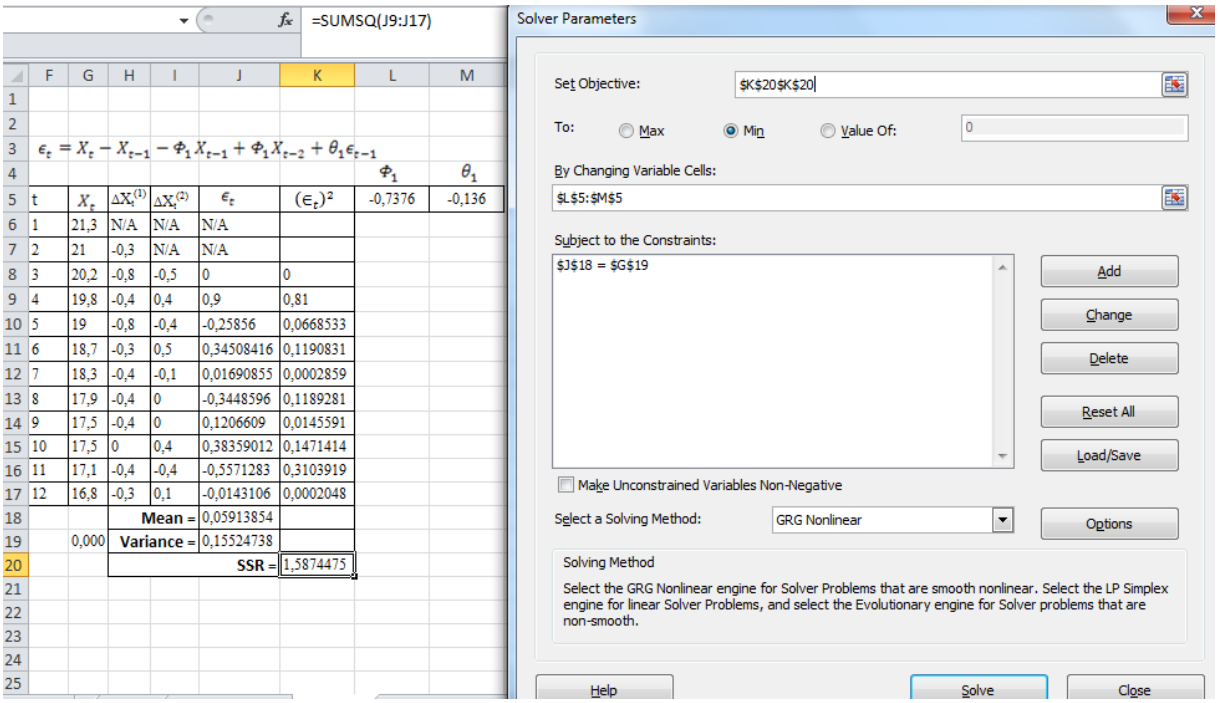
where;

Φ_1 and θ_1 represent model parameters that are to be estimated,

ϵ_t and ϵ_{t-1} represent residuals at time *t* and *t-1* respectively,

X_t, X_{t-1}, and X_{t-2} represent time series values at *t*, *t-1* and *t-2* respectively.

Using equation (47b) and setting initial values of Φ_1, θ_1 and ϵ_{t-1} to zero, values of ϵ_t (residuals) were computed. Square of each residual was evaluated and sum of all squared residuals evaluated. This resulted in sum of squares of residuals (SSR). Using Excel Solver, SSR was minimized by changing values of Φ_1 and θ_1, subject to constraint that mean of residuals should be zero. The results were -0.7376 for AR component and -0.136 for MA component. To access Excel Solver, open Excel and select Data on main menu. In Data menu, select Data Analysis and then Solver. Screen shot of above-performed analysis is shown below.



$\epsilon_t = X_t - X_{t-1} - \Phi_1 X_{t-1} + \Phi_1 X_{t-2} + \theta_1 \epsilon_{t-1}$

							Φ_1	θ_1
	t	X_t	$\Delta X_t^{(1)}$	$\Delta X_t^{(2)}$	ϵ_t	$(\epsilon_t)^2$	-0,7376	-0,136
	1	21,3	N/A	N/A	N/A			
	2	21	-0,3	N/A	N/A			
	3	20,2	-0,8	-0,5	0	0		
	4	19,8	-0,4	0,4	0,9	0,81		
	5	19	-0,8	-0,4	-0,25856	0,0668533		
	6	18,7	-0,3	0,5	0,34508416	0,1190831		
	7	18,3	-0,4	-0,1	0,01690855	0,0002859		
	8	17,9	-0,4	0	-0,3448596	0,1189281		
	9	17,5	-0,4	0	0,1206609	0,0145591		
	10	17,5	0	0,4	0,38359012	0,1471414		
	11	17,1	-0,4	-0,4	-0,5571283	0,3103919		
	12	16,8	-0,3	0,1	-0,0143106	0,0002048		
				Mean =	0,05913854			
		0,000		Variance =	0,15524738			
				SSR =	1,5874475			

Figure 33: Screen shot showing implementation of maximum likelihood estimation method using Excel

The model equation was determined by substituting values of obtained parameters into equation (47b). This resulted in equations shown below.

$$X_t = X_{t-1} - 0.7376X_{t-1} + 0.7376X_{t-2} + \epsilon_t + 0.1363\epsilon_{t-1} \quad (48a)$$

That is, $$X_t = 0.2624X_{t-1} + 0.7376X_{t-2} + \epsilon_t + 0.1363\epsilon_{t-1} \quad (48b)$$

Step 5: Compare fitted model to actual (original) time series. Also examine residuals.

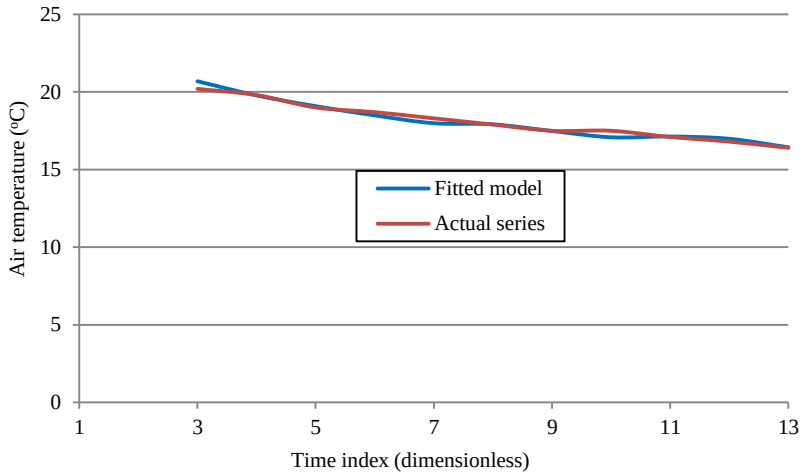


Figure 34: Plot of model fitted values and actual values

Figure 34 shows plot of model and actual (original) series. The model is stationary since absolute values of estimated parameters are less than one. The residuals are stationary with mean of 0.06 and variance 0.16. They are also uncorrelated as would be seen in its ACF and PACF plots shown in figures below. ACF and PACF plots show no significant autocorrelation occurred at any lag.

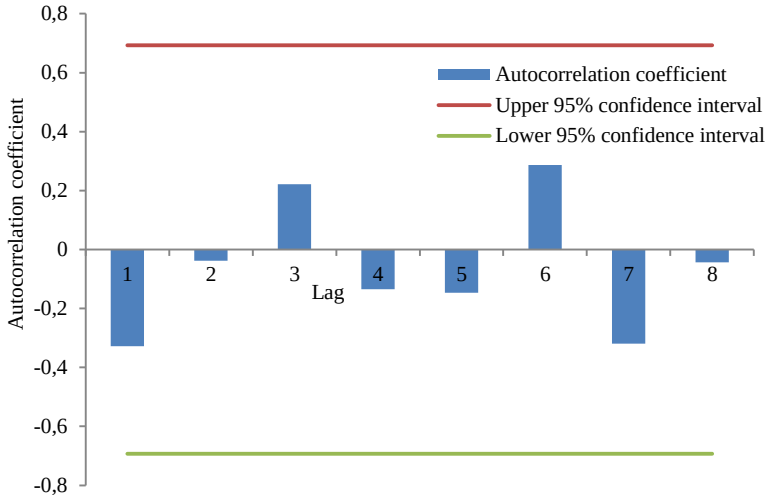


Figure 35a: ACF plot of air temperature residuals

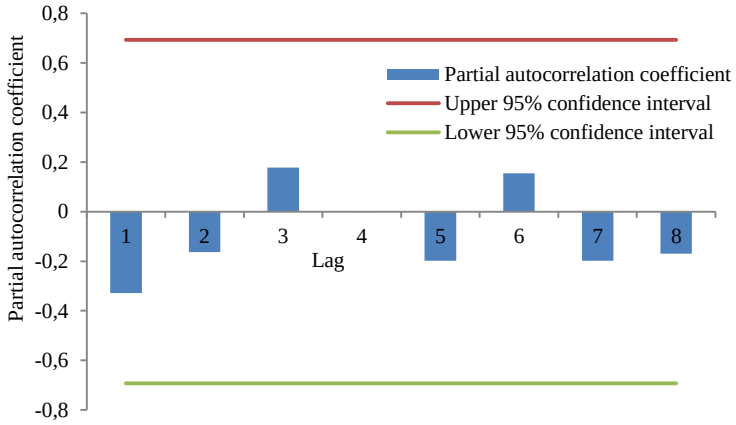


Figure 35b: PACF plot of air temperature residuals

3.2.3.5 Conditional Heteroscedasticity Methods

Conditional heteroscedasticity methods are time series parametric approach. Autoregressive conditional heteroscedasticity (ARCH) is its simplest form. Variants of ARCH include: generalized autoregressive conditional heteroscedasticity (GARCH), exponential generalized autoregressive conditional heteroscedasticity (EGARCH), among others. Conditional heteroscedasticity methods are used to model variance in time series data. They aim to ensure that recent changes and fluctuations in the time series are taken into account.

With conditional heteroscedasticity methods, variance of residuals becomes heteroscedastic (that is, not constant). They become conditionally dependent on past variances. As a result, they can be modeled autoregressively. ARCH model being the simplest conditional heteroscedastic method is usually used. Equation (49) shows the generalized form of ARCH model while equation (50) shows an ARCH(1) model [56].

$$\mathrm{Var}(a_t|a_{t-1},\dots a_n) = (\sigma^2)_t = \alpha_0 + \alpha_1(a^2)_{t-1} + \cdots + \alpha_n(a^n)_{t-n} \qquad (49)$$

$$(\sigma^2)_t = \alpha_0 + \alpha_1(a^2)_{t-1} \qquad (50)$$

where:

n represents number of data in the series,

α_0 and α_1 represent the coefficients. To secure positive variance, $\alpha_0 > 0$ and $\alpha_1 \geq 0$,

$(\sigma^2)_t$ represents variance of residuals at time *t*,
$(a^2)_{t-1}$ represents one-step lag of squared residuals.
The terms are demonstrated in the example that shall follow.

GARCH model uses values of past squared residuals and past variances to model present (current) variance in time series. Consequently, GARCH(*p*,*q*) model implies that past conditional variance describing the GARCH component is of order *p* while the past squared residual is of order *q*. An example is GARCH(1,1) which is described in equation below.

$$(\sigma^2)_t = \alpha_0 + \alpha_1(a^2)_{t-1} + \beta\,(\sigma^2)_{t-1} \qquad (51)$$

where:
α_0, α_1, and β represent the coefficients; To secure positive variance, $\alpha_0 > 0$ and $\alpha_1 \geq 0$,
n represents number of data in the series,
$(\sigma^2)_t$ represents variance of residual at time *t*,
$(a^2)_{t-1}$ represents one-step lag of squared residuals,
$(\sigma^2)_{t-1}$ represents variance of residuals at time *t-1*,
The terms are demonstrated in the example that shall follow.

One method of identifying if ARCH/GARCH model is by investigating if several clusters of variances (changes in variance) exist in squared time series plots. If squared time series is serially correlated, conditional variance will vary with time and cluster of variances will be seen. This helps ensure that series captures all new changes and fluctuations in variance [57].

Example 3.2.3.5: GARCH model analysis

Using air temperature series described in Example 3.2.3.4, GARCH model was evaluated as follows:

Step 1: Plot squared time series of the variable to determine if changes in variance exist. Also examine its ACF and PACF plots to determine if values are correlated. Actual (original) series together with its squares series is shown in table below. Also plot of squared time series is shown below.

Table 15: Air temperature time series for GARCH analysis

t	1	2	3	4	5	6	7	8	9	10	11	12	13
X_t	21.3	21.0	20.2	19.8	19.0	18.7	18.3	17.9	17.5	17.5	17.1	16.8	16.4
$(X_t)^2$	453.7	441	408	392	361	349.7	334.9	320.4	306.3	306.3	292.4	282.2	269

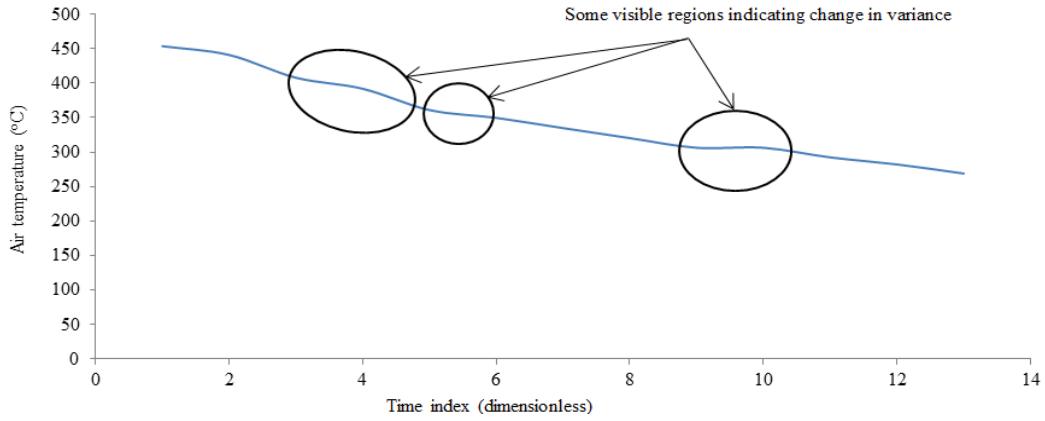


Figure 36: Squared time series plot of air temperature

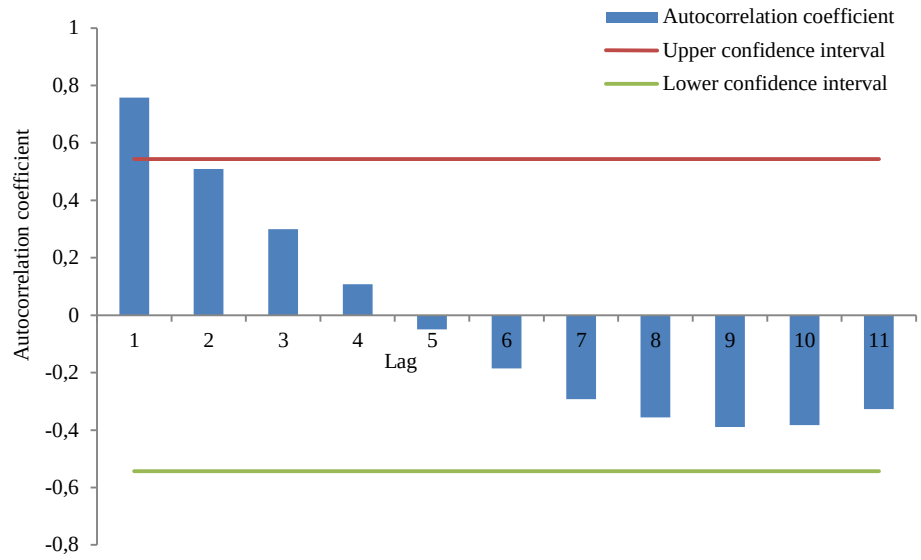


Figure 37a: ACF plot of squared time series

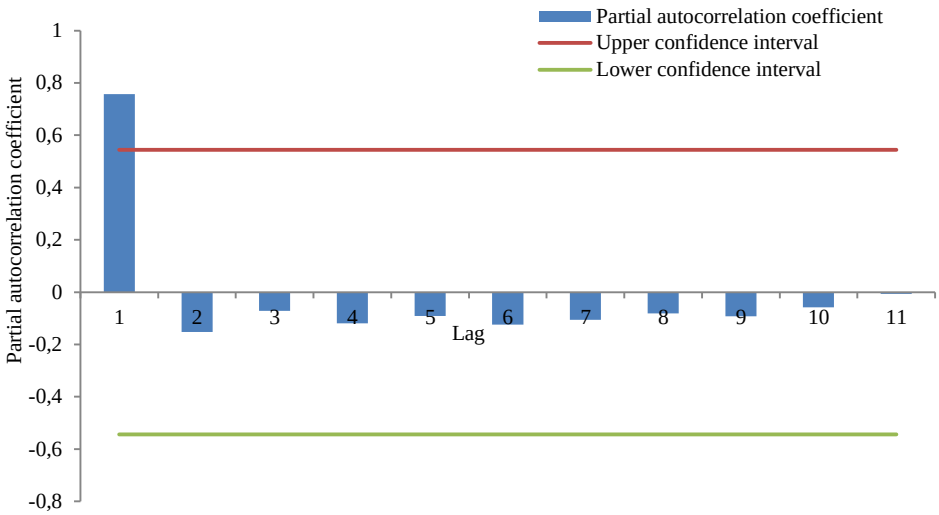


Figure 37b: PACF plot of squared time series

Squared time series of the variable is not stationary. It shows some changes in variance. Additionally, its ACF and PACF plots reveal existence of significant correlations at lag 1. Thus, GARCH model can be used to stabilize the variance.

Step 2: Identify orders of the model

It would be seen from above plots that autocorrelation coefficient is only significant at lag 1. With Box Jenkins method, it means order of AR component of the model is 1. Since ARCH(1) model is an AR(1) model impressed on squared residuals, it then means the series has ARCH(1) component [56].

In order to improve accuracy, past variance needs to be considered. Therefore, GARCH(1,1) which is GARCH corresponding model for ARCH(1) model with past variance consideration was selected.

Step 3: Determine coefficients of model.

In order to determine model coefficients, the differenced series which is called residuals in conditional heteroscedasticity analysis has to be determined. They were obtained by subtracting each of the time series data from its succeeding data in the series. For example, at time index 2, the residual was obtained by subtracting air temperature data at time index 1 from the value at time index 2. The complete residual series can be seen in the analysis table below. With residuals known, squared residuals series was computed.

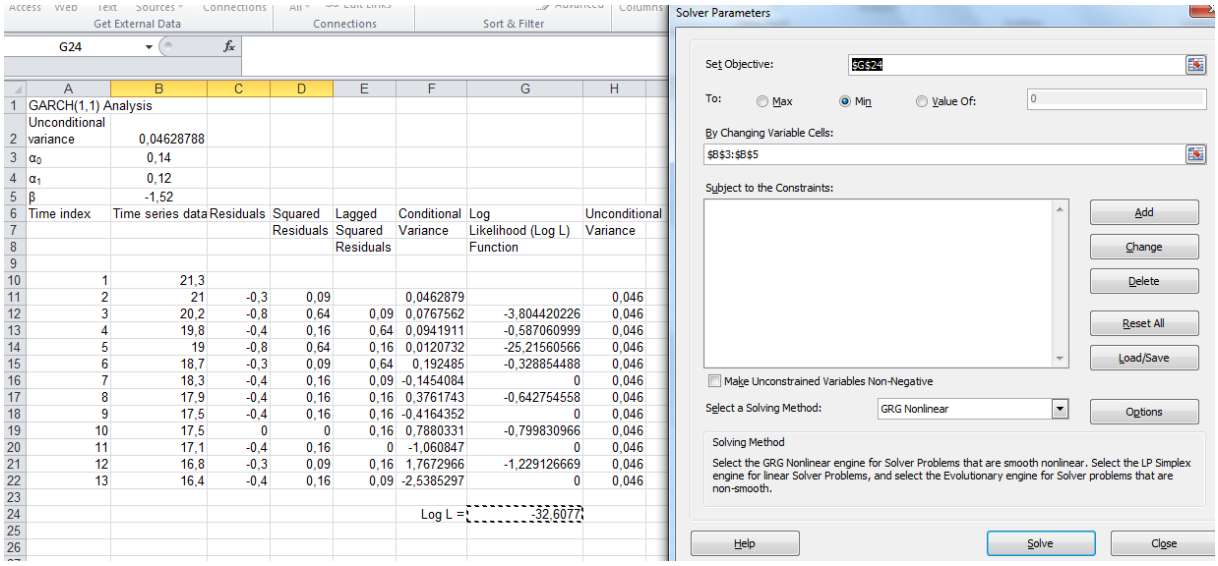
Next was to determine lagged squared residuals. This was obtained by shifting (lagging) squared residual series by one. For example, lagged squared residual at time index 3 is value representing squared residual at time index 2.

Next was to determine conditional variances. Their values were computed using GARCH(1,1) equation for conditional variance described earlier (i.e. equation 51). Initial values of coefficients were set to $\alpha_0 = 0.19, \alpha_1 = 0.09, \beta = -1.52$. They were not set to zero which is the common practice. This is because log-likelihood values do not exist if they were set to zero. Values of lagged squared residuals ($(a^2)_{t-1}$) were already known. Start value for variance of residuals at time *t-1* (that is $(\sigma^2)_{t-1}$) was set to value representing the conditional variance (i.e. variance of the time series data). Consequently, conditional variance for each time index was computed. This was done so that log-likelihood function could be determined and minimized. In line with maximum, likelihood estimation method discussed earlier, values of the coefficients (in this case α_0, α_1, and β) that give lowest possible value of log-likelihood function are the appropriate values (estimates) of the model coefficient.

Next was to compute values of log-likelihood function. This was done using its equation described earlier (i.e. equation 44d). Sum of log-likelihood values was determined and minimized using Excel Solver by changing values of α_0, α_1, and β. The analysis is summarized and shown in table below. Its screen shot is also shown below.

Table 16a: GARCH model analysis before minimization of log-likelihood function

Time index	Air temp. (^{o}C)	Residuals (^{o}C)	Squared residuals (^{o}C)	Lagged squared residuals (^{o}C)	Conditional variance of residuals	Unconditional variance of residuals	Values of log likelihood function
1	21.3	-	-	-	-	-	-
2	21.0	-0.3	0.09	-	0.046	0.046	-
3	20.2	-0.8	0.64	0.09	0.125	0.046	-2.4440
4	19.8	-0.4	0.16	0.64	0.057	0.046	-0.8900
5	19.0	-0.8	0.64	0.16	0.115	0.046	-2.6207
6	18.7	-0.3	0.09	0.64	0.072	0.046	-0.2287
7	18.3	-0.4	0.16	0.09	0.086	0.046	-0.6231
8	17.9	-0.4	0.16	0.16	0.071	0.046	-0.7227
9	17.5	-0.4	0.16	0.16	0.094	0.046	-0.5893
10	17.5	0.00	0.00	0.16	0.059	0.046	0.4935
11	17.1	-0.4	0.16	0.00	0.096	0.046	-0.5790
12	16.8	-0.3	0.09	0.16	0.055	0.046	-0.2871
13	16.4	-0.4	0.16	0.09	0.112	0.046	-0.5395
						Sum	*-9.0306*



	A	B	C	D	E	F	G	H
1	GARCH(1,1) Analysis							
2	Unconditional variance	0,04628788						
3	α_0	0,14						
4	α_1	0,12						
5	β	-1,52						
6	Time index	Time series data	Residuals	Squared Residuals	Lagged Squared Residuals	Conditional Variance	Log Likelihood (Log L) Function	Unconditional Variance
10	1	21,3						
11	2	21	-0,3	0,09		0,0462879		0,046
12	3	20,2	-0,8	0,64	0,09	0,0767562	-3,804420226	0,046
13	4	19,8	-0,4	0,16	0,64	0,0941911	-0,587060999	0,046
14	5	19	-0,8	0,64	0,16	0,0120732	-25,21560566	0,046
15	6	18,7	-0,3	0,09	0,64	0,192485	-0,328854488	0,046
16	7	18,3	-0,4	0,16	0,09	-0,1454084	0	0,046
17	8	17,9	-0,4	0,16	0,16	0,3761743	-0,642754558	0,046
18	9	17,5	-0,4	0,16	0,16	-0,4164352	0	0,046
19	10	17,5	0	0	0,16	0,7880331	-0,799830966	0,046
20	11	17,1	-0,4	0,16	0	-1,060847	0	0,046
21	12	16,8	-0,3	0,09	0,16	1,7672966	-1,229126669	0,046
22	13	16,4	-0,4	0,16	0,09	-2,5385297	0	0,046
24						Log L =	32,6077	

Figure 38a: Screen shot showing GARCH analysis after minimization of log-likelihood function

The model parameters were determined to be $\alpha_0 = 0.14, \alpha_1 = 0.12, \beta = -1.52$. Thus model equation is as shown in equation (52). The variance plot is also shown below.

$$(\sigma^2)_t = 0.14 + 0.12(a^2)_{t-1} - 1.52(\sigma^2)_{t-1} \qquad (52)$$

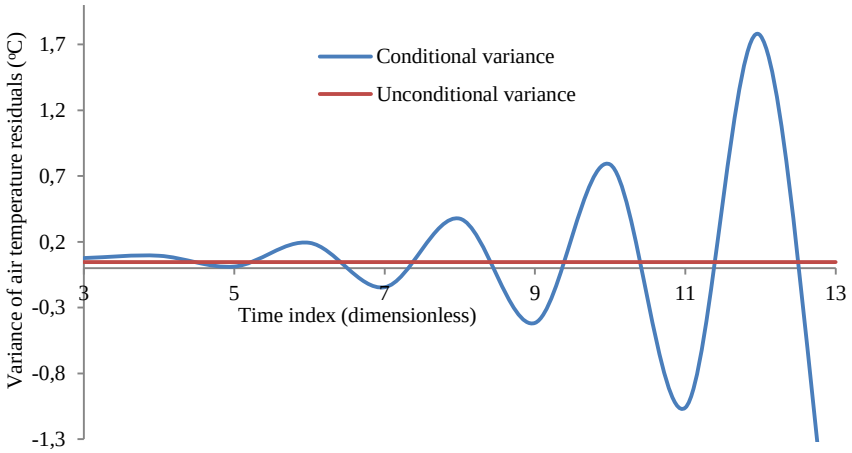


Figure 38b: Conditional and unconditional variances of air temperature residuals

Observe that unconditional variance cuts through the conditional variance model. This is typical in variance modeling since unconditional variance is one of the time-independent components of conditional variance. It approximates to weighted average of conditional variance if variance of conditional mean is neglected. This is in line with the law of total variance (Eve's law) which states that if variance of conditional mean is neglected due to low uncertainity in measured values of a variable, then unconditional variance becomes the weighted average of conditional variance [58].

Recall that what was modeled is variance of residuals, and air temperature values were differenced in order to obtain the residuals (see Table 16a). Then values obtained at various time instants using variance model equation 52 represent changes in present values of air temperature with respect to their immediate past values. This is the variance of residuals that was modeled. Consequently, model-computed air temperature values were determined. They

are shown in the table below. Plot of model-predicted values together with actual (original) time series and variance of residuals is also shown below.

Table 16b: Summary table for computed air temperature values

Time index	Actual air temp. (^oC)	Variance of residuals (^oC)	Model-computed air temp. (^oC)
1	21.3	-	-
2	21.0	-	-
3	20.2	0.0804	21.0804
4	19.8	0.0945	20.2945
5	19.0	0.0155	19.8155
6	18.7	0.1932	19.1932
7	18.3	-0.1429	18.5571
8	17.9	0.3764	18.6764
9	17.5	-0.4129	17.4871
10	17.5	0.7868	18.2868
11	17.1	-1.0559	16.4441
12	16.8	1.7642	18.8642
13	16.4	-2.5308	14.2692

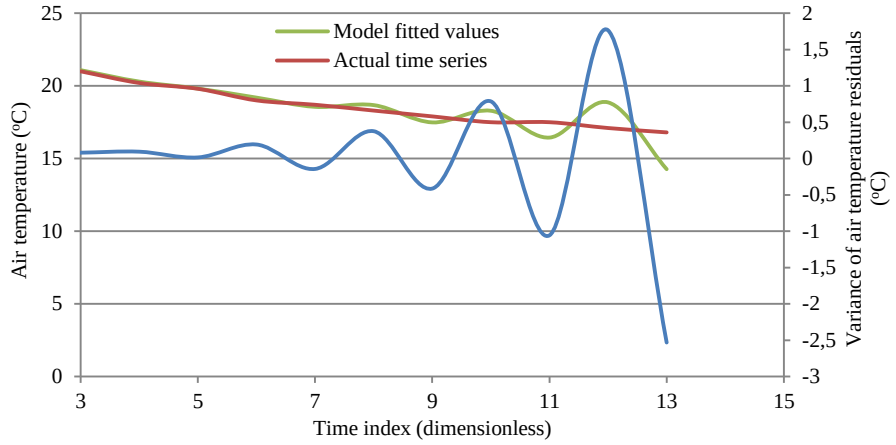


Figure 38c: Plot of model predicted air temperature values and original time series

It would be seen that model computed values did not fit well with actual time series data at higher time instants. This is due to significant variation of air temperature residuals at higher time instants. Although GARCH models for conditional variance while Box Jenkins models for conditional mean, Box Jenkins' ARMA model is still applied without performing differencing or data transformation in time series that exhibit time-dependency in variance. The resulting residuals would be non-stationary and correlated. The residuals are then modeled using GARCH. Thus, ARMA-GARCH model combination is common in engineering applications.

3.2.3.6 Artificial Neural Networks

Artificial neural networks are examples of time series parametric approach. They are classified into feedforward and recurrent networks. The presence of feedback connections in recurrent neural networks (RNN) is the main difference between feedforward and recurrent networks. The feedback connections provide RNNs the ability to process sequences, such as time series. They also endow the network with memory for saving past inputs. Therefore,

outputs of RNNs are influenced not only by current inputs, but also by history of past network inputs.

RNNs are classified into Elman, Jordan, time-delay and echo-state recurrent neural networks [59]. Elman network is becoming obsolete. In time-delay RNNs, additional parameters (abstract memories) are usually available. They are used to store inputs that arrive at hidden units at different times. Consequently, they influence subsequent inputs. RNN models may possess time-windows or intervals. The time-windows are time delays. It is for this reason that time-delay RNN models are considered as autoregressive models.

While artificial neural networks help understand pattern in time series data, they also help classify and predict missing and future data. They achieve this by training time series data into models. The models, having identified patterns and structures of the series, are then used to perform predictions. Each neural network is composed of input layers of neutrons, hidden layers of neutrons and output layers of neutrons. The neutrons are connected with lines which represents their weights. The weights are strength of the neural network. The weights, together with input-output transfer function, determine performance and behavior of neural networks.

Sigmoid transfer function shown in figure 39 was used as the input-output function (transfer function) for the units in the network. Sigmoid transfer function is popular and more commonly used over other functions (e.g. linear, threshold transfer functions, etc.). It has also the additional feature of representing real neutrons closer than other transfer functions. It also has the capability of forcing inputs into the range [0, 1]. Furthermore, its derivative is simple and easy to obtain [60]. This is required for back-propagation of the residuals into the units.

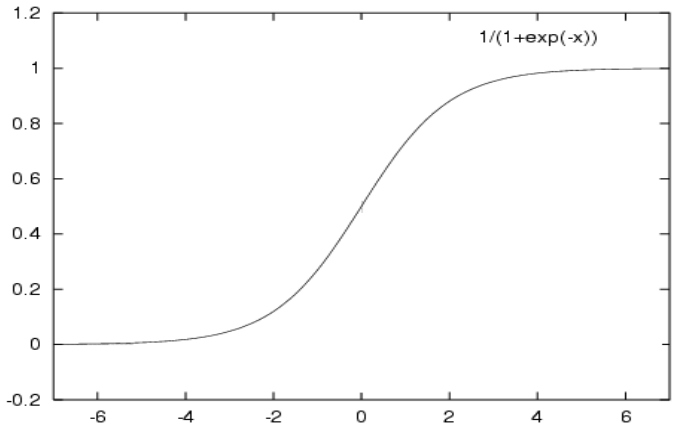


Figure 39: Sigmoid transfer function

Example 3.2.3.6: Consider time series described in Table 15 of Example 3.2.3.5. The time series data were trained into a neural network model using back-propagation method. The method produces an output from given input data by multiplying input with weight and transfer function. The produced output is compared with desired (expected) output so that residual is obtained. Obtained residual is squared and sum of squared residuals (SSR) processed. SSR is then minimized by varying values of weight. The above is presented in steps below.

Step 1: Present given time series and define applicable neural network structure. The given time series is shown in table below.

Table 17: Neural network output table for prediction

Time index	1	2	3	4	5	6	7	8	9	10	11	12	13
Actual (expected) air temp. value	21.3	21	20.2	19.8	19	18.7	18.3	17.9	17.5	17.5	17.1	16.8	16.4

There is one variable, which is air temperature. Thus, there is one input layer with one neutron, one output layer with one neutron and one hidden layer. Hidden layer has two neutrons, in line with rule of thumb that number of neutrons for hidden layer must be at least twice the number of neutron in input layer. Also, there are 13 time-points and input layer has 13 patterns which were mapped to 13 expected outputs. The neural network structure is therefore shown below.

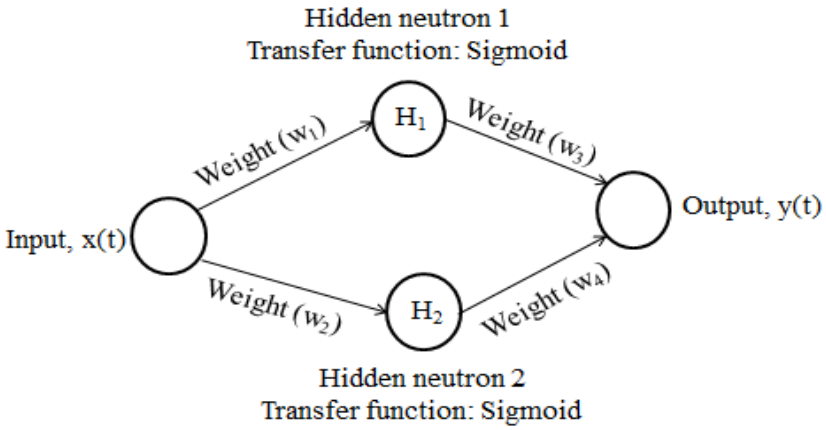


Figure 40: Neural network structure for Example 3.2.3.6

Step 2: Select transfer function and determine weights and biases. Also compute values of hidden neutrons. To achieve this, first assign random values to weights (w_1, w_2, w_3 and w_4) and biases. Sigmoid transfer function was selected and Excel's "RAND" function used to assign random values to weights and biases. Consequently, values of hidden neutrons were computed as follows:

$$H_1(t) = \frac{1}{1+e^{-x(t)}} * \left(\left(w_1 * x(t)\right) - b_1\right) \qquad (53a)$$

$$H_2(t) = \frac{1}{1+e^{-x(t)}} * \left(\left(w_2 * x(t)\right) - b_1\right) \qquad (53b)$$

where:
$H_1(t)$ represents value of hidden neutron 1 at time index (time-point) *t*,
$H_2(t)$ represents value of hidden neutron 2 at time index (time-point) *t*,
$x(t)$ represents input value at time index (time-point) *t*,
w_1 represents weight between input neutron and hidden neutron 1,
w_2 represents weight between input neutron and hidden neutron 2,
b_1 represents bias for hidden layer.

Note that in neural network, biases are not assigned to input layers because activations received from bias neutrons at inputs are overridden by input data. The above computation was done in Excel. Its screenshot is shown in figure 41a.

Step 3: Compute neural network output. Transfer function between neural network output and hidden neutron is usually linear. Thus, network output is computed using below equation.

$$y(t) = \big((H_1(t) * w_3) - b_2\big) + \big((H_2(t) * w_4) - b_2\big) \quad (53c)$$

where:

$H_1(t)$ represents value of hidden neutron 1 at time index (time-point) t,

$H_2(t)$ represents value of hidden neutron 2 at time index (time-point) t,

$y(t)$ represents output value at time index (time-point) t,

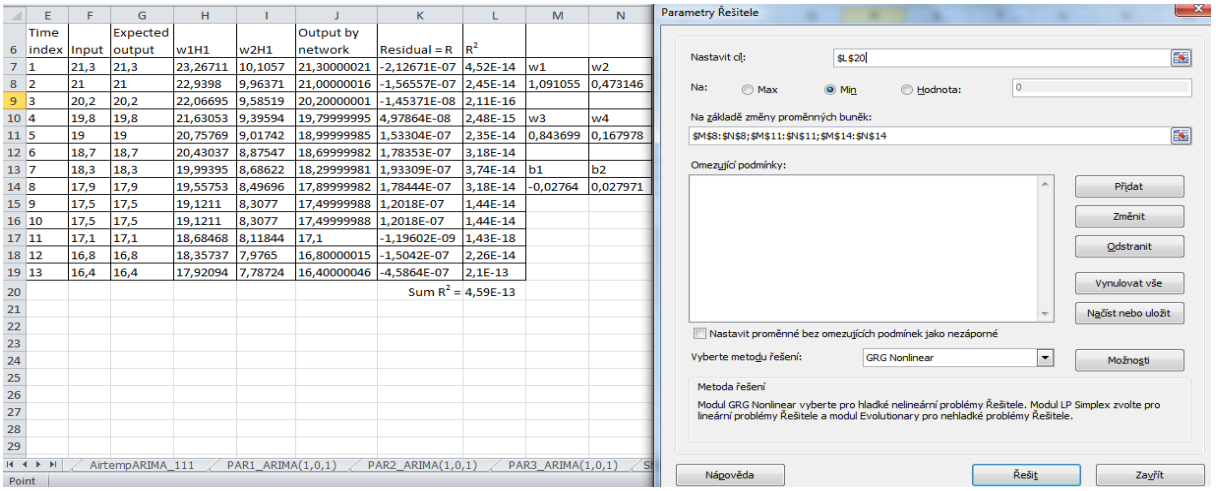
w_3 represents weight between hidden neutron 1 and output neutron,

w_4 represents weight between hidden neutron 2 and output neutron,

b_2 represents bias for output layer.

Step 4: Compute residuals, square of each residual and sum of squared residuals (SSR). Residual is difference between the actual (expected) output and output computed by neural network. Computed residuals were squared and sum of squared residuals determined. This is shown in screenshot of figure 41a.

Step 5: Minimize SSR by changing values assigned to weights and biases. This was done in Excel using Excel's Solver. To access Excel Solver, select "Data" in Excel main menu and then "Solver". Figure 41a shows screenshot summarizing all computations involved while figure 41b shows graph of outputs computed by neural network with actual (original) series.



	E	F	G	H	I	J	K	L	M	N
6	Time index	Input	Expected output	w1H1	w2H1	Output by network	Residual = R	R^2		
7	1	21,3	21,3	23,26711	10,1057	21,30000021	-2,12671E-07	4,52E-14	w1	w2
8	2	21	21	22,9398	9,96371	21,00000016	-1,56557E-07	2,45E-14	1,091055	0,473146
9	3	20,2	20,2	22,06695	9,58519	20,20000001	-1,45371E-08	2,11E-16		
10	4	19,8	19,8	21,63053	9,39594	19,79999995	4,97864E-08	2,48E-15	w3	w4
11	5	19	19	20,75769	9,01742	18,99999985	1,53304E-07	2,35E-14	0,843699	0,167978
12	6	18,7	18,7	20,43037	8,87547	18,69999982	1,78353E-07	3,18E-14		
13	7	18,3	18,3	19,99395	8,68622	18,29999981	1,93309E-07	3,74E-14	b1	b2
14	8	17,9	17,9	19,55753	8,49696	17,89999982	1,78444E-07	3,18E-14	-0,02764	0,027971
15	9	17,5	17,5	19,1211	8,3077	17,49999988	1,2018E-07	1,44E-14		
16	10	17,5	17,5	19,1211	8,3077	17,49999988	1,2018E-07	1,44E-14		
17	11	17,1	17,1	18,68468	8,11844	17,1	-1,19602E-09	1,43E-18		
18	12	16,8	16,8	18,35737	7,9765	16,80000015	-1,5042E-07	2,26E-14		
19	13	16,4	16,4	17,92094	7,78724	16,40000046	-4,5864E-07	2,1E-13		
20							Sum R^2 = 4,59E-13			

Figure 41a: Screenshot showing neural network computations in Excel

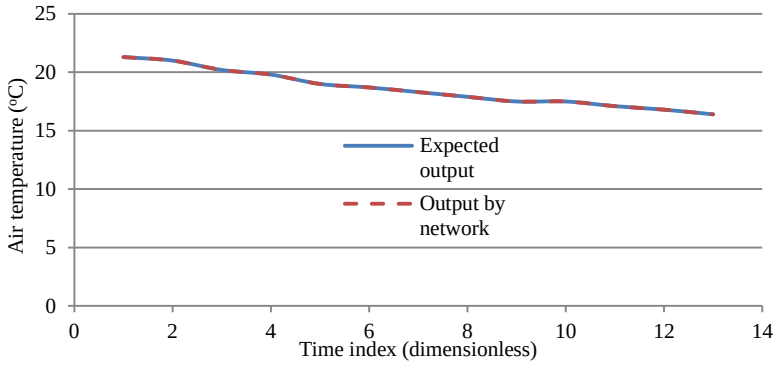


Figure 41b: Air temperature model using artificial neural network

R-squared value was computed as 99.9%, and outputs computed by neural network fit perfectly the actual time series.

3.2.4 Outlook on other Dynamic Stochastic Models

Dynamic stochastic models and their parametric qualities were discussed in section 3.2.3. Other dynamic stochastic models exist. According to Hipel and McLeod [55], they include:

1. Long memory models such as fractional autoregressive integrated moving average (FARIMA or ARFIMA), fractional Gaussian model (FGM), etc.
2. Point process models such as Poisson process model (PPM).
3. Stochastic differential equations (SDE).

In fractional autoregressive integrated moving average (FARIMA or ARFIMA), order of differencing takes non-integer values unlike ARIMA which takes only integer values. The remaining orders of FARIMA (i.e. *p*, *q*) are integer values. Even though order of differencing in FARIMA can take non-integer values between -0.5 and 0.5, it cannot predict values for non-integer time indexes. FARIMA is mostly used in analyzing time series with long-range dependence such as those with structural break or gap. They make the series to be fractionally integrated by allowing order of differencing assume a fractional value between -0.5 and 0.5. Model representation of FARIMA is shown below.

$$\emptyset(B)(1-B)^d(X_t-\mu)=\theta(B)\varepsilon_t \qquad (54a)$$

where:

μ represents mean

$\emptyset$ and θ represent model estimates.

Backward-shift operators are defined as:

$$\emptyset(B)=1-\emptyset_1B-\cdots-\emptyset_pB^p \qquad (54b)$$

$$\theta(B)=1+\theta_1B+\cdots+\theta_qB^q \qquad (54c)$$

Point process models are simple stochastic processes used to model observations that occur randomly. They are without memory and analyze observations over a definite period of time. Poisson distribution is an example of Poisson process model.

Stochastic differential equation (SDE) model is another dynamic stochastic model. It is based on diffusion process where both time and state variables are continuous. Diffusion process is a special type of Markov process with continuous small paths and no jumps. It should be recalled that a Markov process consists of three parts – drift (deterministic), random process and jump. Although Markov chain and Markov process exhibit Markov property (i.e. memorylessness), they differ in their index set. That is, Markov chain has discrete time while Markov process has continuous time.

SDE model is different from discrete-time Markov chain model which has both time and state variables discrete. It is also different from continuous-time Markov chain model which has continuous time but discrete state variables. SDE model can also be linear or reducible. Linear SDE model is common and easy to solve. Its representation is shown below.

$$dX_t=\emptyset(t,X_t)dt+\theta(t,X_t)dE_t \qquad (55)$$

with;

$$\emptyset(t,X_t)=\emptyset_1(t)X_t+\emptyset_2(t) \qquad (56a)$$

$$\theta(t,X_t)=\theta_1(t)X_t+\theta_2(t) \qquad (56b)$$

where,$\emptyset$ and θ represent model parameters; E_t represents noise component; and X_t represents time series state variables.

CHAPTER FOUR

Case Study I

4.1 Site Area and Sampling Design

Site area is within a forest canopy in central western part of Brazil. One sensor station, MC8 understory, was used to measure and record photosynthetically active radiation (PAR), air temperature, relative humidity and soil moisture. Another sensor station, MC8 PAR array, was used for four separate PAR measurements. MC8 understory operated with sampling rate of 30 seconds. Sampled average data were logged at 30 minutes interval, implying a resolution of 30 minutes from August 24, 2007 to May 31, 2014. MC8 PAR array sampled and logged at high frequency (30 seconds resolution) from midday March 16, 2007 to Mach 18, 2012.

4.2 Descriptive and Inferential Statistical Analysis

The two major components of statistical analysis – descriptive and inferential statistical analysis were used to numerically and graphically summarize field data, and also infer on probability distribution of the datasets. Descriptive statistical analysis of field data were performed and presented. Estimates of their test statistics were used in determining appropriate probability distribution functions that describe datasets.

The analysis investigated if observations followed any well-known discrete probability distribution. This is required since it forms basis for determining optimal sampling interval.

4.2.1 Probability Distribution Analysis

Probability distribution functions that fit field data were investigated. Non-parametric methods (such as kernel density estimation method) were not applied since it is usually recommended that normality and moments of distribution be investigated first.

As a result, normality analysis was performed. Besides examining whether observations follow normal distribution, normality analysis also provides information about first, second, third and fourth moments of distribution. These statistics give clue on nature of the characteristic function. Consequently, they contribute in determining the appropriate probability distribution function. In order to perform above analysis, variability among observations (data) needs to be reduced.

In order to reduce variability among data, they were classified based on natural cycles - daytime and nighttime datasets. This is common and generally accepted in temporal sampling programs where data are taken at regular intervals [4]. Moreover, due to changes in physical, chemical and biological properties that occur in the environment during diurnal cycles, daytime and nighttime measurements are seen as two different stochastic processes.

Therefore, field data described in section 4.1 were classified into daytime and nighttime datasets. Daytime dataset constitutes set of values taken from 06:13:08 to 17:43:08. Nighttime dataset constitutes set of values taken from 18:13:08 till 05:43:08 the following day. Data analysis revealed that nighttime datasets for photosynthetically active radiation (PAR) are constant and do not vary. Consequently, PAR values have to be measured once during nighttime. Daytime PAR values together with daytime and nighttime values of other variables

(air temperature, relative humidity and soil moisture) are normal distributed. This would be seen in histogram plots and chi-square tables shown below. Analysis involving daytime distribution has been shown as an example. The histograms represent typical daytime distribution of variables while chi-square tables test their goodness-of-fit to normal distribution. The null hypothesis is that dataset for each variable fits a normal distribution. Chi-square tests gave p-values that are greater than the statistical significance value (0.05). Thus, null hypotheses were accepted implying datasets were normal distributed.

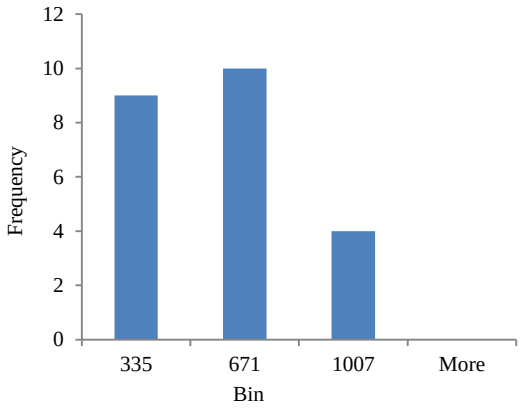


Figure 42a: Histogram plot of PAR

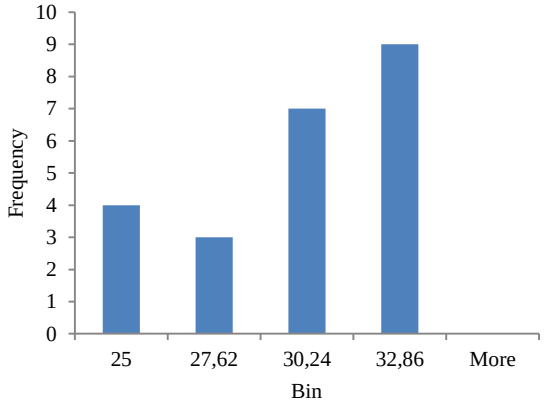


Figure 42b: Histogram plot of air temperature

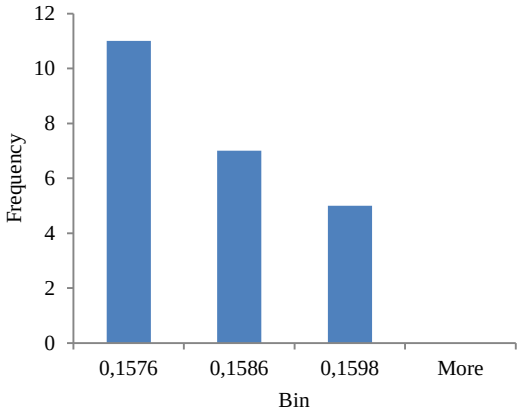


Figure 42c: Histogram plot of soil moisture

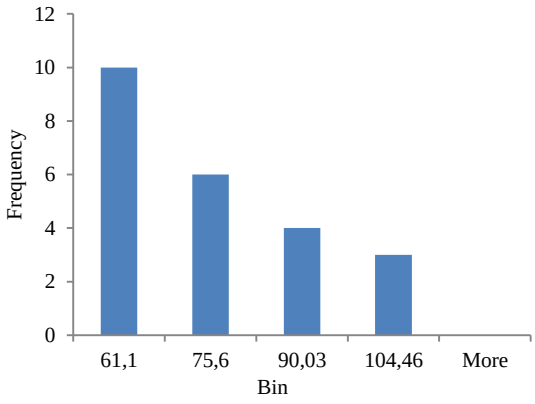


Figure 42d: Histogram plot of humidity

Table 18a: Chi-square test of normality for PAR

PAR values in bins (W/m^2)	PAR observed frequency (O_i)	PAR expected frequency (E_i)	Chi-square value = $\frac{(O_i-E_i)^2}{E_i}$
0 - 335	9	7.1553	0.4756
336 - 671	10	10.4666	0.0208
672 - 1007	4	3.7467	0.0171
		Sum	0.5135
Degree of freedom = 3 – 1 = 2		*p-value*	0.7736

Table 18b: Chi-square test of normality for air temperature

Air temperature values in bins (^{o}C)	Air temperature observed frequency (O_i)	Air temperature expected frequency (E_i)	Chi-square value = $\frac{(O_i-E_i)^2}{E_i}$
22.48 - 25.00	4	2.0424	1.8763
25.10 - 27.62	3	5.9754	1.4816
27.72 -30.24	7	7.8200	0.0860
30.34 - 32.86	9	4.5908	4.2348
		Sum	7.6787
Degree of freedom = 4 – 1 = 3		*p-value*	0.0531

Table 18c: Chi-square test of normality for soil moisture

Soil moisture values in bins (kg/kg)	Soil moisture observed frequency (O_i)	Soil moisture expected frequency (E_i)	Chi-square value = $\frac{(O_i-E_i)^2}{E_i}$
0.1565 - 0.1576	11	7.2864	1.8927
0.1577 - 0.1586	7	7.2450	0.0083
0.1587 - 0.1598	5	3.8410	0.3497
		Sum	2.2507
Degree of freedom = 3 – 1 = 2		*p-value*	0.3245

Table 18d: Chi-square test of normality for humidity

Humidity values in bins (g/g)	Humidity observed frequency (O_i)	Humidity expected frequency (E_i)	Chi-square value = $\frac{(O_i-E_i)^2}{E_i}$
46.75 - 61.17	10	5.5131	3,6517
61.18 - 75.6	6	8.5859	0,7788
75.61 - 90.03	4	5.5683	0,4417
90.04 - 104.46	3	1.4996	1,5012
		Sum	6.3734
Degree of freedom = 4 – 1 = 3		*p-value*	0.0948

The tables below show seasonal statistical estimates of nighttime and daytime measurements respectively.

Table 19: Nighttime seasonal descriptive statistics - summary table for Brazil data (PAR constant at 1.2 W/m^2)

Variable	**Statistics**	**Estimate**	**Standard Error**
Nighttime air temperature (^{o}C) - Spring (Sept 2010 - Nov 2010)	Mean	26.4032	0.08631
	Variance	32.525	-
	Skewness	0.239	0.037
	Kurtosis	-0.776	0.074
Nighttime humidity (g/kg) – Spring (Sept 2010 - Nov 2010)	Mean	51.6414	0.44467
	Variance	863.277	-
	Skewness	0.178	0.037
	Kurtosis	-1.256	0.074
Nighttime soil moisture (kg/kg) – Spring (Sept 2010 – Nov 2010)	Mean	0.104455	0.000533
	Variance	0.001	-
	Skewness	1.426	0.037
	Kurtosis	0.497	0.074
Nighttime air temperature (^{o}C) - Summer (Dec 2010 – Feb 2011)	Mean	22.0889	0.0397
	Variance	3.245	-
	Skewness	0.499	0.054
	Kurtosis	0.376	0.108
Nighttime humidity (g/g) – Summer	Mean	91.2118	0.23096

(Dec 2010 – Feb 2011)	Variance	109.83	-
	Skewness	-1.201	0.054
	Kurtosis	0.836	0.108
Nighttime soil moisture (kg/kg) - Summer (Dec 2010 – Feb 2011)	Mean	0.131942	0.000717
	Variance	0.001	-
	Skewness	0.258	0.054
	Kurtosis	-1.143	0.108
Nighttime air temperature (^{o}C) – Autumn (Mar 2011 – May 2011)	Mean	20.8653	0.0495
	Variance	5.54	-
	Skewness	-0.517	0.051
	Kurtosis	0.039	0.103
Nighttime humidity (g/g) - Autumn (Mar 2011 – May 2011)	Mean	93.2801	0.000712
	Variance	69.084	-
	Skewness	-1.74	0.051
	Kurtosis	3.731	0.103
Nighttime soil moisture (kg/kg) - Autumn (Mar 2011 – May 2011)	Mean	0.118066	0.000712
	Variance	0.001	-
	Skewness	0.727	0.051
	Kurtosis	-0.981	0.103
Nighttime air temperature (^{o}C) – Winter (June 2011 – Aug 2011)	Mean	18.341	0.07484
	Variance	12.833	-
	Skewness	0.516	0.051
	Kurtosis	-0.249	0.102
Nighttime humidity (g/g) - Winter (June 2010 – Aug 2011)	Mean	53.2205	0.40301
	Variance	372.097	-
	Skewness	-0.008	0.051
	Kurtosis	-0.864	0.102
Nighttime soil moisture (kg/kg) - Winter (June 2010 – Aug 2011)	Mean	0.072916	0.00004153
	Variance	0	-
	Skewness	1.114	0.051
	Kurtosis	1.073	0.102

Table 20: Daytime seasonal descriptive statistics estimates - summary table for Brazil field data

Variable	**Statistics**	**Estimate**	**Standard Error**
Daytime PAR (W/m^2) - Spring (Sept 2010 – Nov 2010)	Mean	650.809	12.37135
	Variance	337629	-
	Skewness	0.668	0.052
	Kurtosis	-0.832	0.104
Daytime air temperature (^{o}C) - Spring (Sept 2010 - Nov 2010)	Mean	29.5104	0.11737
	Variance	30.389	-
	Skewness	-0.475	0.052
	Kurtosis	-0.567	0.104
Daytime humidity (g/g) - Spring (Sept 2010 - Nov 2010)	Mean	41.7316	0.61824
	Variance	843.189	-
	Skewness	0.603	0.052
	Kurtosis	-0.93	0.104
Daytime soil moisture (kg/kg) - Spring (Sept 2010 – Nov 2010)	Mean	0.104419	0.000765
	Variance	0.001	-
	Skewness	1.386	0.052
	Kurtosis	0.367	0.104
Daytime PAR (W/m^2) - Summer (Dec 2010 – Feb 2011)	Mean	423.7972	9.64403
	Variance	210475.7	-
	Skewness	1.376	0.051
	Kurtosis	1.243	0.103
Daytime air temperature (^{o}C) - Summer (Dec 2010 – Feb 2011)	Mean	27.4812	0.08902
	Variance	17.933	-
	Skewness	0.001	0.051

	Kurtosis	-0.9	0.103
Daytime Humidity (g/g) - Summer (Dec 2010 – Feb 2011)	Mean	69.4829	0.51476
	Variance	599.636	-
	Skewness	-0.287	0.051
	Kurtosis	-1.069	0.103
Daytime soil moisture (kg/kg) - Summer (Dec 2010 – Feb 2011)	Mean	0.131987	0.000673
	Variance	0.001	-
	Skewness	0.173	0.051
	Kurtosis	-1.21	0.103
Daytime PAR (W/m^2) - Autumn (Mar 2011 – May 2011)	Mean	239.7565	6.31984
	Variance	86390.96	-
	Skewness	2.179	0.053
	Kurtosis	5.187	0.103
Daytime air temperature (^oC) - Autumn (Mar 2011 – May 2011)	Mean	25.3851	0.07951
	Variance	13.674	-
	Skewness	-0.629	0.053
	Kurtosis	-0.266	0.105
Daytime humidity (g/g) – Autumn (Mar 2011 – May 2011)	Mean	74.2216	0.45914
	Variance	455.982	-
	Skewness	-0.494	0.053
	Kurtosis	-0.646	0.105
Daytime soil moisture (kg/kg) - Autumn (Mar 2011 – May 2011)	Mean	0.118253	0.000726
	Variance	0.001	-
	Skewness	0.695	0.053
	Kurtosis	-1.08	0.105
Daytime PAR (W/m^2) – Winter (June 2011 – Aug 2011)	Mean	508.8828	8.808
	Variance	165479.9	-
	Skewness	0.59 8	0.053
	Kurtosis	-0.543	0.106
Daytime air temperature (^oC) – Winter (June 2011 – Aug 2011)	Mean	27.6505	0.13446
	Variance	38.561	-
	Skewness	-0.919	0.053
	Kurtosis	-0.121	0.106
Daytime humidity (g/g) - Winter (June 2010 – Aug 2011)	Mean	29.059	0.4993
	Variance	531.755	-
	Skewness	1.222	0.053
	Kurtosis	0.444	0.106
Daytime soil moisture (kg/kg) - Winter (June 2010 – Aug 2011)	Mean	0.071555	0.0000465
	Variance	0	-
	Skewness	1.111	0.053
	Kurtosis	0.924	0.106

One seasonal cycle comprises of four seasons – spring, summer, autumn and winter. For nighttime datasets, null hypothesis is that values are normal distributed. Similarly, null hypothesis for daytime datasets is that values are normal distributed. Some variables have statistical significances (p-values) below 0.05. Although some p-values were below 0.05, their null hypotheses were still accepted. This is because a variable is reasonably close to normal distribution, and considered being normal-distributed if both skewness and kurtosis lie between -1.5 and +1.5 [61, 62]. These are fulfilled as could be seen in Tables 19 and 20. The above normality test involving skewness and kurtosis is a rule of thumb within statistical methods of testing for normality. Diagnostic hypothesis tests such as use of chi-square test are also part of statistical methods. They are different from graphical methods such as use of histogram. Buttressing further, it is well-known and established that daytime and nighttime measurements of most environmental variables are normal distributed [63].

Considering that 20 is the minimum number of data required for good computation of first and second moments of a normal-distributed data [64], optimum sampling interval was computed using the equation below.

$$S_M = \frac{M_D}{S_Z} \qquad (57)$$

where;

S_M represents optimum sampling interval,

M_D represents total measurement duration,

S_Z represents sample size.

Using Brazil field data, 690 minutes was measurement duration for each daytime and nighttime variable. Optimum sampling interval was calculated as 34.5 minutes. The probability distribution function can then be determined.

A similar work was done by Kish [65]. Kish stated that optimum sampling interval within a given stratum is proportional to square-root of sample variance of the stratum.

$$S_M = k * \sqrt{Var_s} \qquad (58)$$

where;

S_M represents optimum sampling interval,

Var_s represents variance of the stratum,

k represents proportionality constant which depends on stratum sample size and total population.

4.2.2 Analysis of Variance and Homogeneity

In this section, one-way analysis of variance (ANOVA) and homogeneity of distributions were used to investigate if observations differ significantly from year to year. ANOVA evaluates if there is significant difference in means (averages). Comparing means of distributions is common since it helps in determining if measurements made for a particular year can stand-in for other years. This helps save energy if realized. For each variable, the null hypothesis was that no significant difference in mean values occurred over the years. Tables 21a, 21b, 21c and 22d summarize the results.

Table 21a: ANOVA results for PAR

		Mean values (M) (subset for alpha = 0.05)					Standard deviation (SD)	Coefficient of variation (SD/M)
Year	No of dataset	1	2	3	4	5		
2013	4269	158.5956					310.6729	1.958900
2009	4368	183.1706					325,4224	1.776608
2008	4301		241.4372				380,4278	1.575680
2011	4366			285.7036	285.7036		478.0617	1.673278
2012	4269				304.2241		494.9873	1.627048
2010	4366					337.663	529.5169	1.568182
Sig.		0.188	0.565	0.184	0.667	1		

Table 21b: ANOVA results for air temperature

		Mean values (M)		Standard deviation (SD)	Coefficient of variation (SD/M)
Year	No of dataset	1	2		
2011	4366	24.7816		5.7582	0.232358
2013	4269	24.853		4.78698	0.192612
2010	4366		26.4032	5.70306	0.215999
2012	4269		26.5770	6.0345	0.227057
2008	4301		26.7333	5.8539	0.218974
2009	4368		26.7643	5.04843	0.188626
Sig.		1	0.084		

Table 21c: ANOVA results for humidity

		Mean of values (M)				Standard deviation (SD)	Coefficient of variation (SD/M)
Year	No of dataset	1	2	3	4		
2012	4269	42.6054				33.11612	0.777275
2008	4301	42.9912				30.29296	0.703236
2010	4366		51.6414			29.38157	0.568954
2011	4366		52.4455			31.99193	0.610003
2009	4368			54.5219		27.57314	0.505726
2013	4269				80.4237	21.09720	0.262326
Sig.		1	0.993	1	1		

Table 21d: ANOVA results for soil moisture

		Mean of values (M)				Standard deviation (SD)	Coefficient of variation (SD/M)
Year	No of dataset	1	2	3	4		
2008	4301	0.086144				0.036117	0.419268
2012	4269		0.098416			0.049115	0.499054
2009	4368		0.098834			0.034936	0.353477
2010	4366			0.104455		0.035217	0.337146
2011	4366			0.105215		0.041529	0.394707
2013	4269				0.135233	0.039762	0.294024
Sig.		1	1	1	1		

From Tables 21a – 21d, it would be seen that none of the variables has same mean value for all years investigated. This indicates that none of the yearly observations could be used for all the years investigated. However, for certain years, some of the variables have approximately same mean value. This indicates homogeneity in mean. This means observations could be seen as being drawn from the same population. They possess the same distribution and consequently same distribution parameters such as mean.

Additional tests performed such as chi-squared test of homogeneity indicates that yearly distributions of the variables whose means (averages) belong to the same group or subset are homogeneous. Furthermore, estimates of coefficient of variation shown in above tables indicate that variability of yearly observations did not change significantly. This means that yearly observations have similar distribution. Air temperature has the lowest level of variability since its coefficient of variation is approximately constant at 0.2.

Key findings made in this section are summarized as follows:

1. None of the yearly observations could stand-in as observations for all the years investigated.

2. Some yearly observations of the variables have the same type of probability distribution.
3. Some yearly observations of the variables have the same distribution parameters such as mean. They are considered homogeneous and can stand-in for each other.

The variogram method was used to investigate the observations. This is discussed in the succeeding section.

4.2.3 Variogram Analysis

Yearly observations of the variables were analyzed using classical variogram. The aim was to determine how often variables should be measured. This is another form in which sampling interval can be represented. Results obtained are shown below.

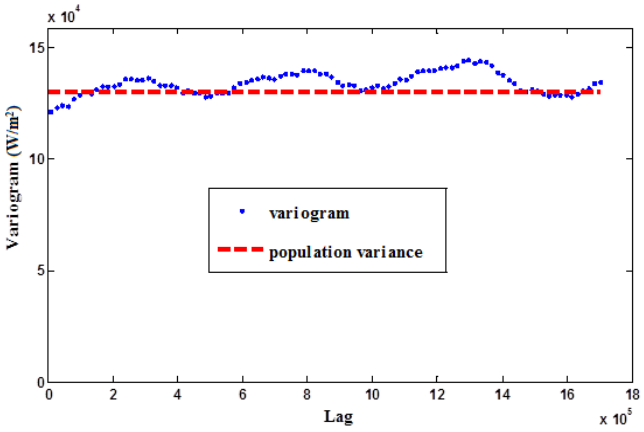


Figure 43a: Variogram of PAR

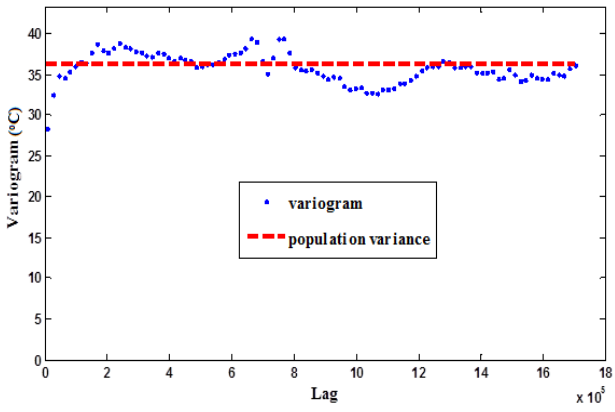


Figure 43b: Variogram of air temperature

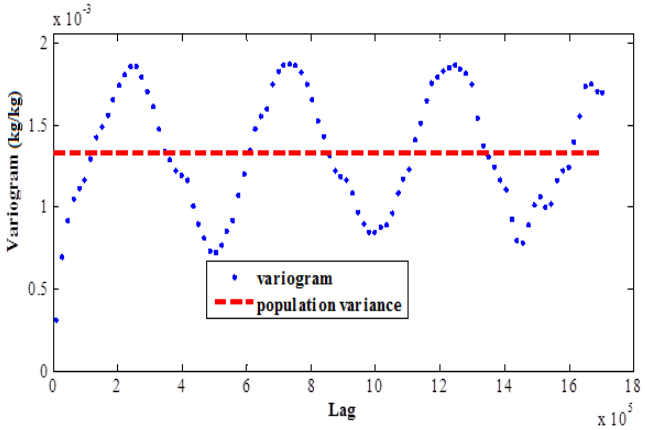


Figure 43c: Variogram of soil moisture

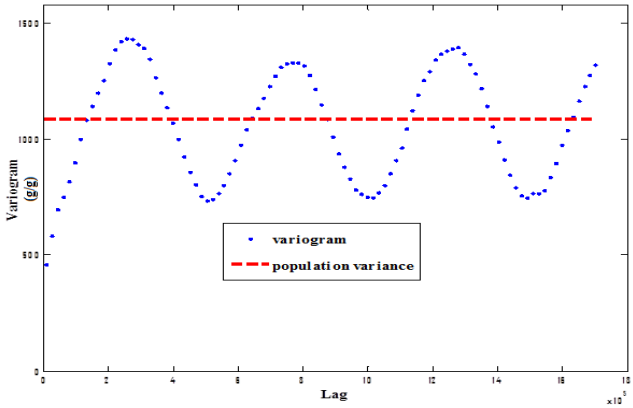


Figure 43d: Variogram of humidity

From above figures, it would be seen that:

1. PAR values are correlated at low lag with small percentage change in variances. This is due to high nugget and reduced variability among values. Percentage change in variances was further reduced within the amplitude region. Thus, nighttime values are more similar and less correlated than daytime values. Consequently, nighttime values should be measured *less frequently* - about twice during nighttime period. Daytime values should be measured *frequently* - about six or seven times during daytime period.

1. Air temperature values are correlated at low lag with significant percentage change in variances. This indicates values are not as similar as they were for nighttime PAR. However, within the region after amplitude, variability was reduced. Thus, daytime values should be measured *more frequently* - about eleven or twelve times during daytime period. Nighttime values should be measured *frequently* – about six or seven times during nighttime period.
2. Variogram of humidity is periodic with constant sill and range. Sill is represented by amplitude while range is represented by wavelength [66]. Considering symmetrical nature of the variogram half-waveform, both daytime and nighttime values have same variance dependency on lag. That is, at lags before and after amplitude, values are correlated with significant percentage change in variances. Thus, both daytime and nighttime values should be measured *more frequently* - about eleven or twelve times during each daytime and nighttime period.
3. Variogram of soil moisture is also periodic with constant sill and range. Due to symmetric nature of the variogram half-waveform, both daytime and nighttime values have same variance dependency on lag. At lags before and after amplitude, unlike humidity variogram, soil moisture values are correlated with lower percentage change in variances. Thus, both daytime and nighttime values should be measured *frequently* - about six or seven times during each daytime and nighttime period.

It should be recalled that periods representing daytime and nighttime are as explained earlier. That is, daytime is from 06:13:08 to 17:43:08 while nighttime is from 18:13:08 to 05:43:08 the following day.

Monthly observations (datasets) were investigated using variogram. The variograms examined and compared observations of variables on monthly basis from 2008 to 2011. In order to simplify analysis, months of January to April were selected. The datasets were logarithmic-transformed twice in order to secure stationarity. During first transformation, ten was added to each data before it was log-transformed. This aligns with additive rule of log-transformation performed when data being transformed is close to zero or below zero. Figures below show 2-day variogram of PAR observations for months of January, February, March and April in 2008 – 2011.

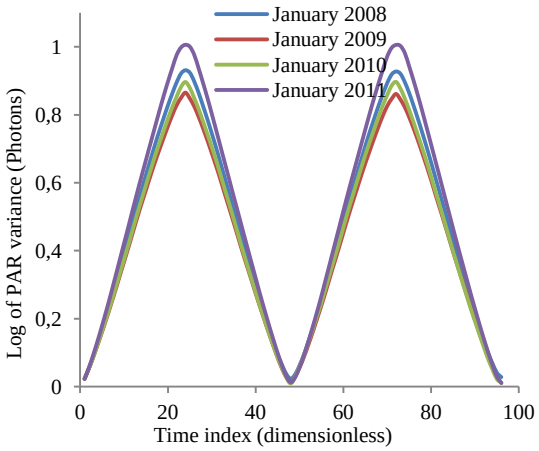


Figure 44a: 2-day variogram of PAR for January

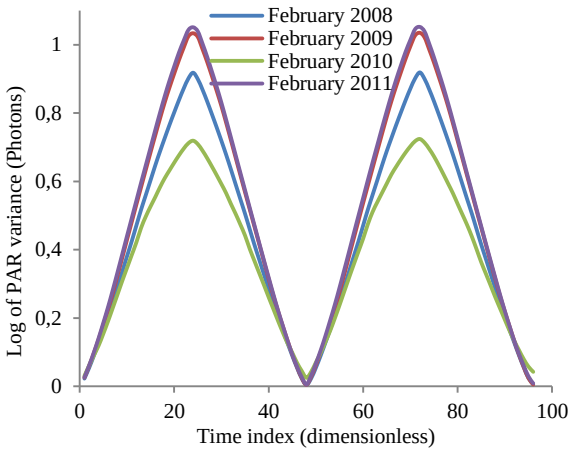


Figure 44b: 2-day variogram of PAR for February

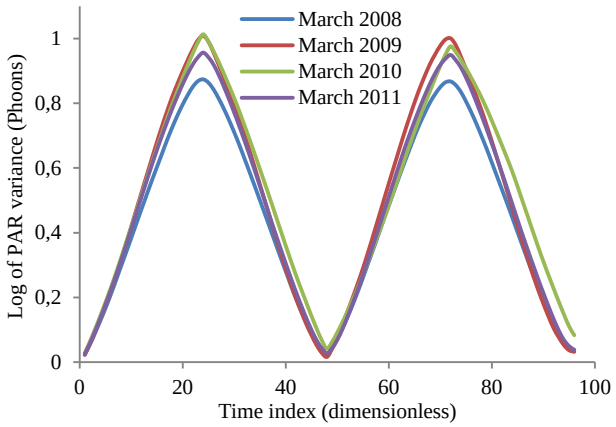


Figure 44c: 2-day variogram of PAR for March

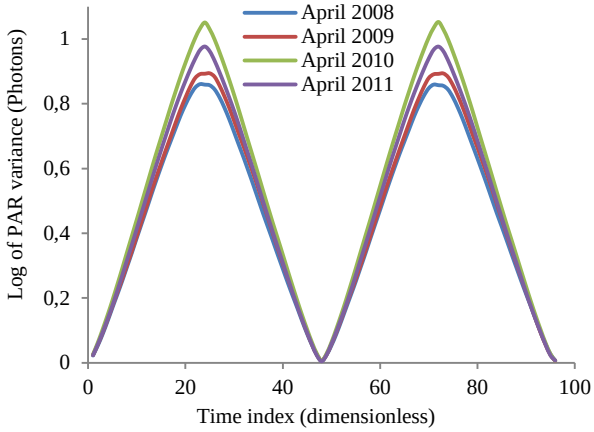


Figure 44d: 2-day variogram of PAR for April

From above figures, it would be seen that all variograms are periodic. Their sills and range are deduced from variograms since according to Funk [66], sill of periodic variogram is represented by its amplitude while range is represented by wavelength. Comparing the variograms, it would be seen that all months investigated with exception of February have approximately the same variogram for each of the years. That is, variograms of January 2008, 2009, 2010 and 2011 were close to one another. The same applies to March and April variograms. Their amplitudes (sill) did not differ significantly like those in February variograms. Significant variabilities among yearly observations were noticeable in February variograms. This may have been caused by physical factors such as changes in urbanization or natural factors such as change in weather. This was not investigated since it is not within the scope of this work.

For air temperature, figures below show its 2-day variogram for months of January, February, March and April in 2008 – 2011.

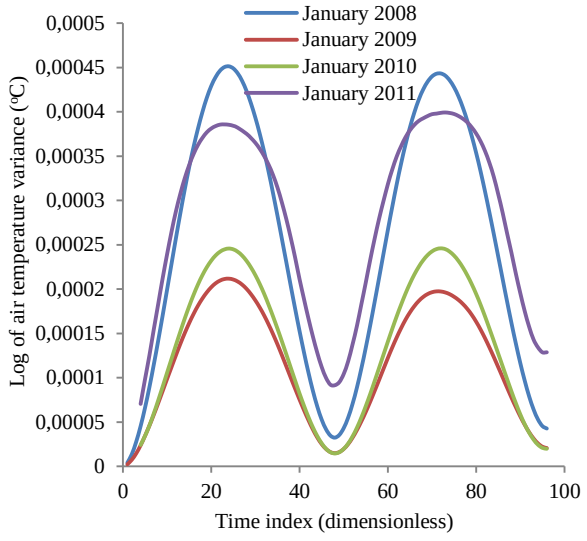


Figure 44e: 2-day variogram of air temperature for January

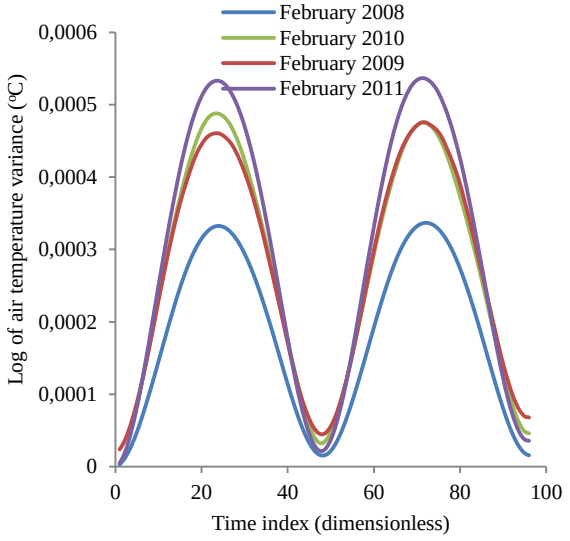


Figure 44f: 2-day variogram of air temperature for February

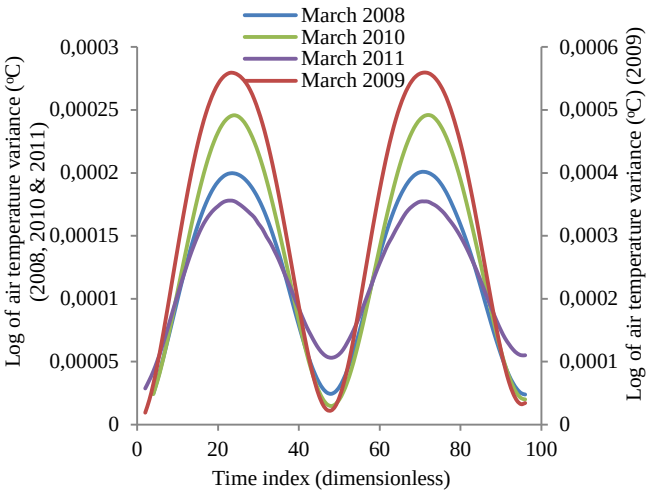


Figure 44g: 2-day variogram of air temperature for March

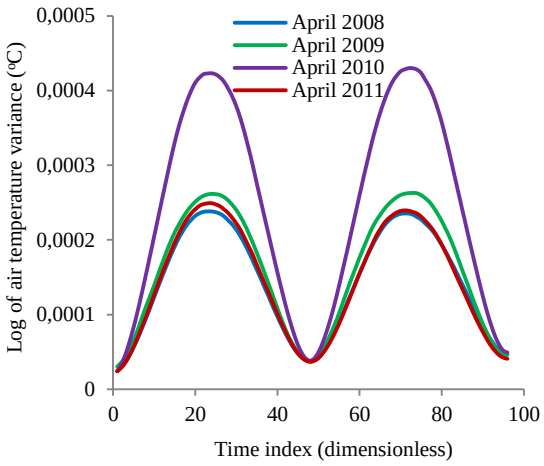


Figure 44h: 2-day variogram of air temperature for April

From above figures, it would be seen that variograms are periodic. It would also been seen that they vary significantly unlike those for PAR variable that were closely related. In some months, they formed groups of homogeneous pair with variability reduced within groups.

Figures below show 2-day variograms of humidity for months of January, February, March and April in 2008 – 2011.

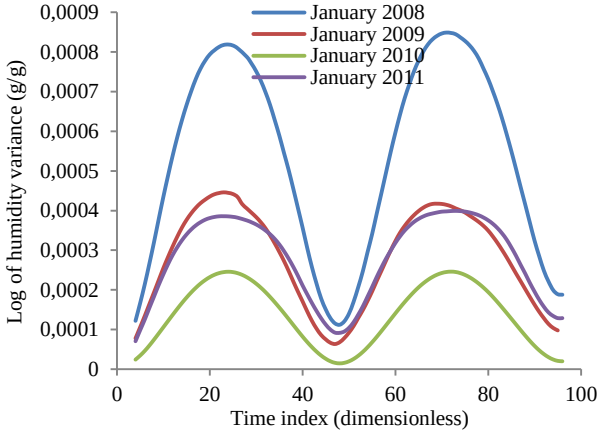


Figure 44i: 2-day variogram of humidity for January

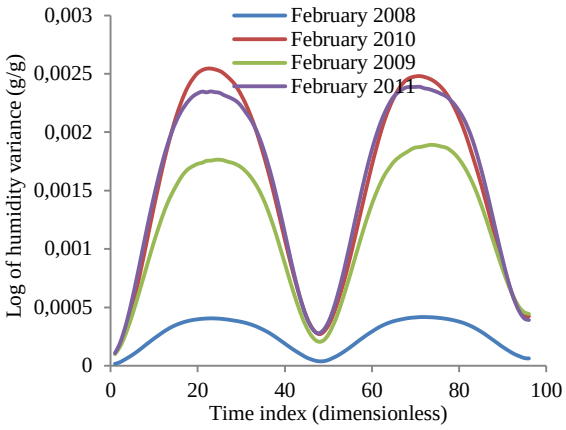


Figure 44j: 2-day variogram of humidity for February

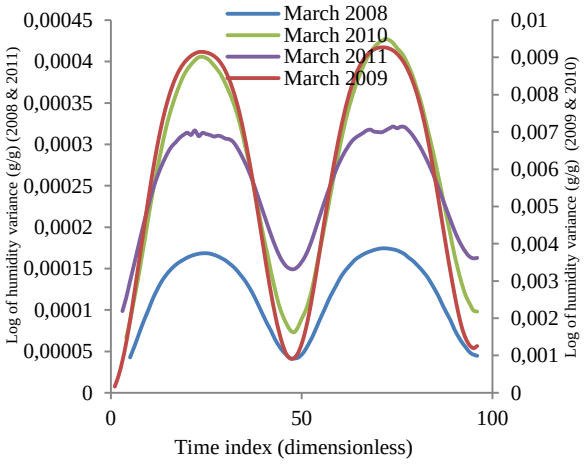


Figure 44k: 2-day variogram of humidity for March

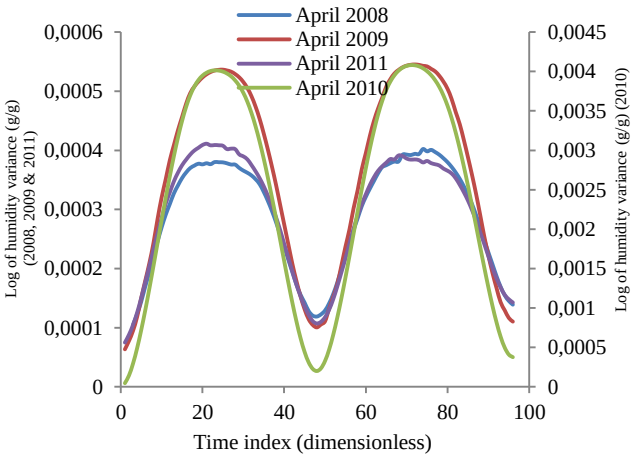


Figure 44l: 2-day variogram of humidity for April

From above figures, it would be seen that variograms are periodic. It would also been seen that they vary significantly more than those for PAR and air temperature variables. In some months, they formed groups of homogeneous pair with variability reduced within groups.

For soil moisture, figures below show its 2-day variogram for each months of January, February, March and April in 2008 – 2011.

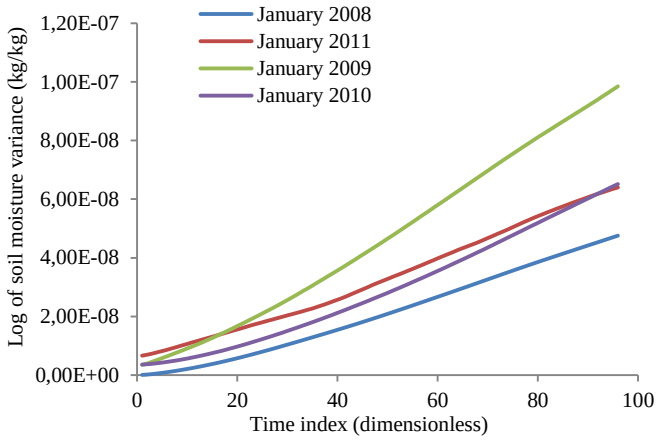


Figure 44m: 2-day variogram of soil moisture for January

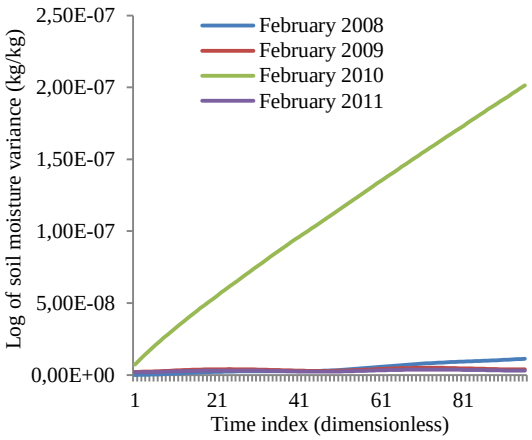


Figure 44n: 2-day variogram of soil moisture for February

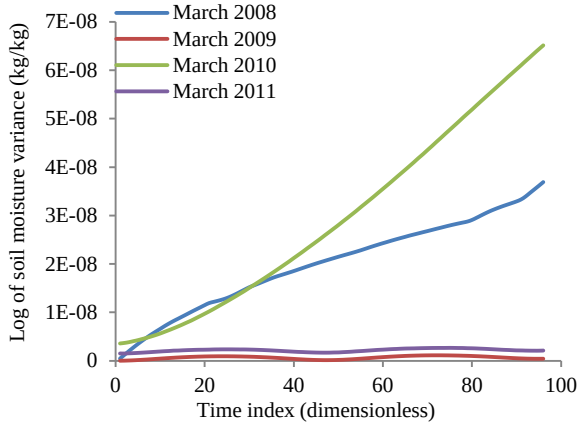


Figure 44o: 2-day variogram of soil moisture for March

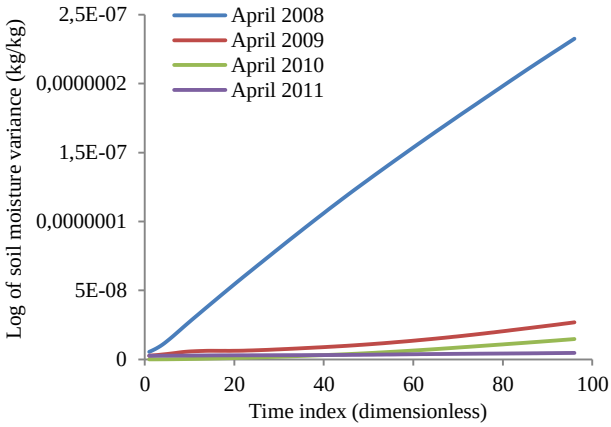


Figure 44p: 2-day variogram of soil moisture for April

From the above figures, it would be seen that soil moisture variograms are linear with some gentle slopes. Some slopes were almost flat (horizontal), indicating a near constant variance. This represents sill and variability among dataset is reduced. In other cases where numerical value of slope is significant (e.g. January variograms), sill was not available since such linear variograms have no sill. Since values of variance on ordinate (vertical) axis are low, it is acceptable to cut off variogram at its population variance; otherwise dataset is modeled using other methods.

4.3 Box-Jenkins Analysis

Box-Jenkins stochastic method was used to investigate seasonal relationship existing between observations of same months of successive years. For this analysis, two variables – air temperature and PAR were selected. Measures for assessing goodness-of-fit in Box-

Jenkins models discussed earlier were used to assess if fitted models could be used to predict same month of other years.

Monthly observations of variables for given year were used to estimate orders of model and their coefficients. These constitute the calibrated models and their model equations called calibrated model equations. Calibrated models are used to test same month observations of other years. Testing same month observations of another year means using some observations of the month and calibrated model equation of another year to predict values for the remaining period of the month. For example, calibrated model equation of January 2008 air temperature was obtained by fitting ARIMA model to air temperature observations of January 2008. Computed calibrated model equation of January 2008 together with air temperature observations of January 2009 (that are useable in January 2008 model equation) were used to predict values for remaining period of January 2009. Similar analyses were done for January 2010 and January 2011. Performance results obtained were compared with those obtained using calibrated models. They are summarized below.

Table 22: ARIMA results summary for air temperature - January observations

		Jan 2008 model	Jan 2009 model	Jan 2010 model	Jan 2011 model
With Jan 2008 data	Model orders	(2,1,2)	(1,1,1)	(2,1,2)	(2,2,2)
	Coefficients (AR; MA)	(0.13, 0.58; 1.12, -0.52)	(0.846; -0.408)	(0.83, 0.2; -0.32, -0.11)	(0.04, 0.67; 0.4, -0.24)
	AIC value	2496.62	2507.79	2726.01	3129.68
	R-squared value	0.981	0.663	0.609	0.488
	Residuals	Uncorrelated with mean of 0 and variance of 0.3	Uncorrelated with mean of 0.2 and variance of 5.3	Uncorrelated with mean of -0.6 and variance of 5.9	Uncorrelated with mean of 0.1 and variance of 8.1
With Jan 2009 data	Model orders	(2,1,2)	(1,1,1)	(2,1,2)	(2,2,2)
	Coefficients (AR; MA)	(0.13, 0.58; 1.12, -0.52)	(0.846,-0.408)	(0.83, 0.2; -0.32, -0.11)	(0.04, 0.67; 0.4, -0.24)
	AIC value	2569.08	1974.54	2297.02	2350.87
	R-squared value	0.481	0.982	0.567	0.552
	Residuals	Uncorrelated with mean of -0.2 and variance of 5.6	Uncorrelated with mean of 0 and variance of 0.8	Uncorrelated with mean of -0.8 and variance of 4.1	Uncorrelated with mean of -0.1 and variance of 4.8
With Jan 2010 data	Model orders	(2,1,2)	(1,1,1)	(2,1,2)	(2,2,2)
	Coefficients (AR; MA)	(0.13, 0.58; 1.12, -0.52)	(0.846; -0.408)	(0.83, 0.2; -0.32, -0.11)	(0.04, 0.67; 0.4, -0.24)
	AIC value	2787.42	2416.40	2410.43	2415.09
	R-squared value	0.572	0.687	0.981	0.733
	Residuals	Uncorrelated with mean of 0.6 and variance of 6.1	Uncorrelated with mean of 0.8 and variance of 4.2	Uncorrelated with mean of 0 and variance of 0.3	Uncorrelated with mean of 0.7 and variance of 3.6
With Jan 2011 data	Model orders	(2,1,2)	(1,1,1)	(2,1,2)	(2,2,2)
	Coefficients (AR; MA)	(0.13, 0.58; 1.12, -0.52)	(0.846; -0.408)	(0.83, 0.2; -0.32, -0.11)	(0.04, 0.67; 0.4, -0.24)
	AIC value	3104.58	2418.57	2408.03	2406.18
	R-squared value	0.52	0.723	0.782	0.982

With Jan 2011 data	Residuals	Uncorrelated with mean of -0.1 and variance of 8.0	Uncorrelated with mean of 0.08 and variance of 4.6	Uncorrelated with mean of -0.7 and variance of 3.1	Uncorrelated with mean of 0 and variance of 0.3

Key

	Represents calibrated model

Table 23: ARIMA results summary for PAR - January observations

		Jan 2008 model	Jan 2009 model	Jan 2010 model	Jan 2011 model
With Jan 2008 data	Model orders	(1,0,2)	(1,0,2)	(1,0,2)	(1,0,2)
	Coefficients (AR; MA)	(0.94; 0.25, 0.16)	(0.94; 0.16, 0.15)	(0.94; 0.24, 0.10)	(0.94; 0.24, 0.20)
	AIC value	-93,37	-93,37	-93,37	-93,37
	R-squared value	0.998	0.998	0.998	0.998
	Residuals	Uncorrelated with mean of 0 and variance of 0.1	Uncorrelated with mean of 0 and variance of 0.1	Uncorrelated with mean of 0 and variance of 0.1	Uncorrelated with mean of 0 and variance of 0.1
With Jan 2009 data	Model orders	(1,0,2)	(1, 0, 2)	(1,0,2)	(1,0,2)
	Coefficients (AR; MA)	(0.94; 0.25,0.16)	(0.94; 0.16, 0.15)	(0.94; 0.24, 0.1)	(0.94; 0.24, 0.20)
	AIC value	-47.01	-47.01	-47.01	-47.01
	R-squared value	0.998	0.998	0.998	0.998
	Residuals	Uncorrelated with mean of -0.1 and variance of 0.1	Uncorrelated with mean of 0 and variance of 0.1	Uncorrelated with mean of -0.1 and variance of 0.1	Uncorrelated with mean of -0.1 and variance of 0.1
With Jan 2010 data	Model orders	(1,0,2)	(1,0,2)	(1,0,2)	(1,0,2)
	Coefficients (AR; MA)	(0.94;0.25,0.16)	(0.94; 0.16, 0.15)	(0.94; 0.24, 0.10)	(0.94; 0.24, 0.20)
	AIC value	-18.91	-18.91	-18.91	-18.91
	R-squared value	0.998	0.998	0.998	0.998
	Residuals	Uncorrelated with mean of 0 and variance of 0.1	Uncorrelated with mean of 0 and variance of 0.1	Uncorrelated with mean of 0 and variance of 0.1	Uncorrelated with mean of 0 and variance of 0.1
With Jan 2011 data	Model orders	(1,0,2)	(1,0,2)	(1,0,2)	(1,0,2)
	Coefficients (AR; MA)	(0.94; 0.25,0.16)	(0.94; 0.16, 0.15)	(0.94; 0.24, 0.1)	(0.94; 0.24, 0.20)
	AIC value	434.29	434.29	434.29	434.29
	R-squared value	0.995	0.995	0.995	0.995
	Residuals	Uncorrelated with mean of 0 and variance of 0.1	Uncorrelated with mean of 0 and variance of 0.1	Uncorrelated with mean of 0 and variance of 0.1	Uncorrelated with mean of 0 and variance of 0.1

Key

	Represents calibrated model

From Tables 22 and 23, it would be seen that calibrated models showed best performance for monthly observations they were trained and calibrated with. They gave lowest AIC values and maximum R-squared values. Their AIC values were in line with Box-Jenkins statement

on parsimony of models which demands that the lowest possible model be selected in order to reduce variability among residuals.

From Table 22, it would be seen that January 2008 air temperature model when tested with January 2009 air temperature data did not give good results. It gave an R-squared value of 0.481. Variability in residuals was high with variance of 5.6. Similar results were obtained when the model was tested with January 2010 and January 2011 observations. Furthermore, air temperature models of January 2009, 2010 and 2011 did not give good results when tested with January 2008 air temperature observations. Variability in residuals was high with variance reaching highest value of 8.1. Consequently, none of the monthly model equations could give good prediction of monthly values of the variable in other years.

For PAR, it would be seen from Table 23 that both calibrated models and non-calibrated models performed well. R-squared values were high with low variability among residuals. The residuals approximated to strict white noise - a true innovation that does not require further modeling or investigation. Consequently, each monthly model equations could give good prediction of monthly values of the variable in other years. These results are in agreement with those obtained using variogram method where all yearly January variograms were found to be approximately the same for PAR whereas for air temperature, they differ.

CHAPTER FIVE
Case Study II

5.1 Site Area and Sampling Design

Field data used in this case study was obtained through measurements taken within a residential area in the city of Edmonton, Alberta. Three sensors were deployed at different locations within the site. The sensors measured photosynthetically active radiation (PAR). One PAR sensor (identified as PAR1) was installed near fence so that partial shading of the fence cast over it. Another sensor (PAR2) was installed under a canopy formed by trees. The third sensor (PAR3) was installed so that it receives solar radiation directly with no shadow or canopy over it. An additional sensor was installed and used for air temperature measurement. Data were collected from May 3, 2015 to July 2, 2015 with intervals between measurements changing from 5 to 30 seconds in several steps (5, 10, 15, and 30).

The sensors were powered by solar harvester supplemented by supercapacitor and backup primary batteries. They were mounted such that sufficient distance exists between them. This is required so as to fulfill sensing range specification which helps avoid electromagnetic interferences. Summary of measurement (sampling) plan used is shown below.

Table 24: Sampling plan

Sampling interval used (seconds)	**Start date (yy-mm-dd)/ Time (hh:mm:ss)**	**Finish date (yy-mm-dd)/ Time (hh:mm:ss)**
5	2005-05-03/16:00:00	2005-05-10/22:20:30
10	2005-05-10/22:42:05	2005-05-13/22:12:45
15	2005-05-13/22:21:15	2005-05-23/22:48:00
30	2005-05-23/22:53:15	2005-07-02/22:58:15

5.2 Descriptive and Inferential Statistical Analysis

Concepts of one-way analysis of variance (ANOVA) and homogeneity of distribution were used to investigate the observations. This is essential since it helps determine if observations have the same distribution, among others. Measurements were not repeated at the same time in successive years. Therefore, the observations constitute a special type of longitudinal data since they were taken progressively with different time intervals.

Observations referred in this context are measurements recorded at 5, 10, 15 and 30 seconds. The variables are PAR1, PAR2, PAR3 and air temperature. 25744 observations were obtained while sampling at 10 seconds interval. This is lowest number of observations obtained during measurements. Consequently in this analysis and for each variable, first 25744 observations were drawn from all observations recorded while sampling at 5, 10, 15 and 30 seconds. Results obtained are shown in the tables below.

Table 25a: ANOVA results for air temperature

		Mean values			**Standard error of mean**	**Standard deviation**	**Coefficient of variation**
Sampling interval (s)	**No of dataset**	**1**	**2**	**3**			
15	25744	16.4125			0.0808	12.9602	10.2341
5	25744	17.4867			0.0709	11.3710	7.3942
10	25744		19.2160		0.0961	15.4151	12.3661
30	25744			23.3507	0.0691	11.0902	5.2672

Table 25b: ANOVA results for PAR1

		Mean values			Standard error of mean	Standard deviation	Coefficient of variation
Sampling interval (s)	**No of dataset**	**1**	**2**	**3**			
5	25744	222.8274			2.2643	363.2834	592.2738
10	25744	210.7321			2.0861	334.6825	531.5394
15	25744		150.026		1.6552	265.5555	470.0500
30	25744			102.9498	1.2391	198.8049	383.9094

Table 25c: ANOVA results for PAR2

		Mean values			Standard error of mean	Standard deviation	Coefficient of variation
Sampling interval (s)	**No of dataset**	**1**	**2**	**3**			
5	25744	273.2411			2.5530	409.6017	614.0129
10	25744	281.1414			2.4352	390.7058	542.9688
15	25744		202.3197		1.9836	318.2485	500.6045
30	25744			152.5282	1.6106	258.4038	437.7716

Table 25d: ANOVA results for PAR3

		Mean values			Standard error of mean	Standard deviation	Coefficient of variation
Sampling interval (s)	**No of dataset**	**1**	**2**	**3**			
5	25744	253.6795			2.2696	364.1408	522.701
10	25744		345.0743		2.7629	443.2721	569.4141
15	25744	236.7537			2.2390	359.2229	545.0437
30	25744			206.5156	2.0510	329.0581	524.3148

From Tables 25a – 25d, it would be seen that for each variable, only observations from two different sampling intervals could be approximated as being homogeneous. Such observations possess the same probability distribution and consequently same distribution parameters such as mean. It would also be seen that air temperature and PAR3 each has a fairly constant standard error of mean. This indicates that mean values of air temperature and PAR3 did not deviate significantly from their respective observations. This suggests that variability of observations with respect to mean value is approximately constant. Thus, air temperature observations obtained using each of the sampling intervals have the same type of probability distribution.

Furthermore, it would be seen that lowest estimate of standard error of mean occurred at sampling interval of 30 seconds. Standard deviation and coefficient of variation were also at lowest values when sampling interval was 30s. This is in agreement with results published in Åstrom [67] where optimal sampling interval occurred when variance of sample mean was minimal.

Finally, figure 45 and its corresponding time series revealed that battery voltage lasted for 168,471 seconds when sampling interval was 5 seconds. When sampling interval was increased to 10 seconds, battery voltage lasted for 204,720 seconds. With further increase of sampling interval to 15 seconds, battery voltage lasted till 1,521,800 seconds. Increasing sampling interval further to 30 seconds resulted in battery voltage lasting for 2,849,000 seconds. This implies battery lifetime was extended by more than 87% when sampling interval was increased from 15 to 30 seconds. Battery was recharged and restored to full-charge capacity before change in sampling interval was made.

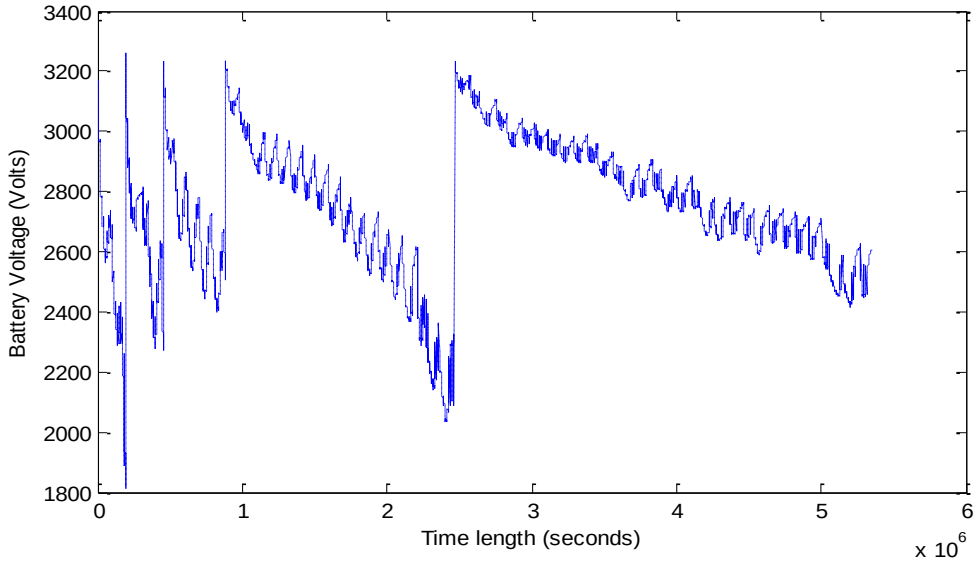


Figure 45: Time series of battery voltage for different sampling frequencies

5.3 Box-Jenkins Analysis

Box-Jenkins method was used to analyze field datasets. The analysis investigated autocorrelation nature of the datasets. This is required in order to fit stochastic models to datasets. With accurate stochastic model, missed values as well as future values can be predicted. It also serves as gateway to determining optimal sampling interval. Steps involved in using Box-Jenkins method and how they are implemented have been demonstrated in section 3.2.3.4 using a typical time series. Following the steps, air temperature data obtained using 30 seconds sampling interval (from 2015-05-23 to 2015-07-02) were analyzed below.

Stationarity for time series was established by differencing the dataset once. ACF plot of differenced series indicates a non-zero autocorrelation value at lag 1. PACF plot also indicates non-zero partial autocorrelation values at lag 1. This indicates AR component is of order 1 while MA component is of order 1. The ACF and PACF plots are shown below.

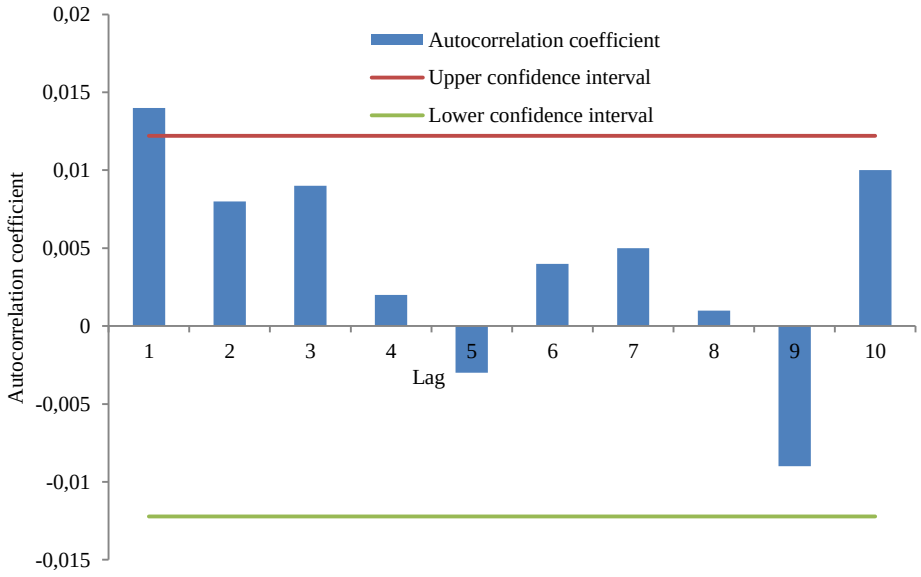


Figure 46a: Autocorrelation function plot of stationarized air temperature series

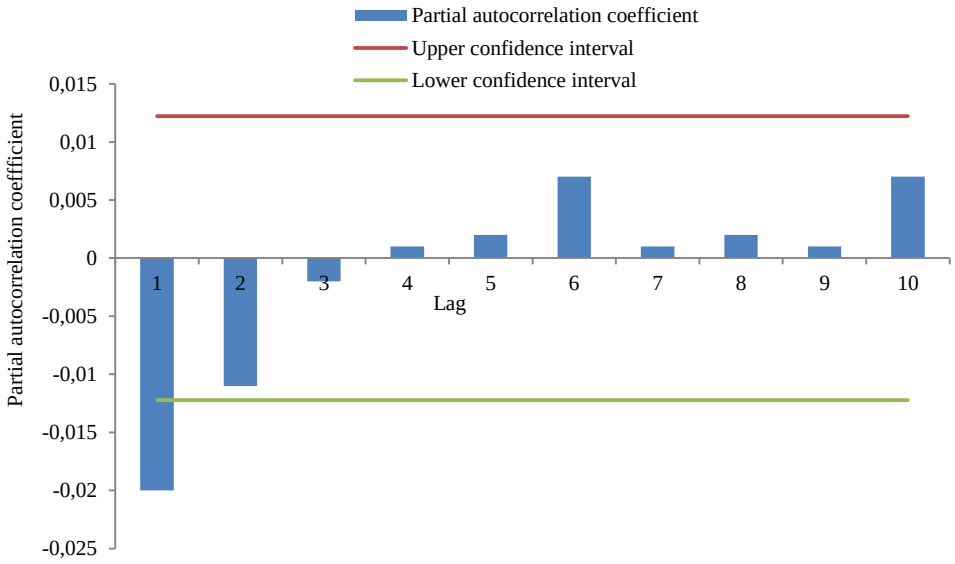


Figure 46b: Partial autocorrelation function plot for stationarized air temperature series

Model coefficients were estimated. Results obtained are summarized below.

Table 26: ARIMA model orders and parameters - air temperature

ARIMA Type/Order = ARIMA(1,1,1)		
ARIMA Parameters	**Estimates**	**Standard Error**
Constant term	0.000	0.000
AR Lag 1	-0.091	0.003
Difference	1	-
MA Lag 1	0.474	0.003

Therefore, model equation becomes:

$$(1-\Phi_1 B)(1-B)^1 X_t = (1-\theta_1 B)\epsilon_t => (1-B-\Phi_1 B+\Phi_1 B^2)X_t = \epsilon_t - (\theta_1 B)\epsilon_t \quad (59a)$$

$$X_t - X_{t-1} - (-0.091)X_{t-1} + (-0.091)X_{t-2} = \epsilon_t - (0.474)\epsilon_{t-1} \quad (59b)$$

$$=> X_t - 0.909X_{t-1} - 0.091X_{t-2} = \epsilon_t - 0.474\epsilon_{t-1} \quad (59c)$$

$$=> X_t = 0.909X_{t-1} + 0.091X_{t-2} + \epsilon_t - 0.474\epsilon_{t-1} \quad (59d)$$

For simplicity purpose and also to avoid cloudy graph due to large number of observations investigated (25744 observations), graph of the model is provided below showing first 1500 observations and their corresponding predicted values. It would be seen that the model is a good fit of the actual series. R-squared value was computed as 99%.

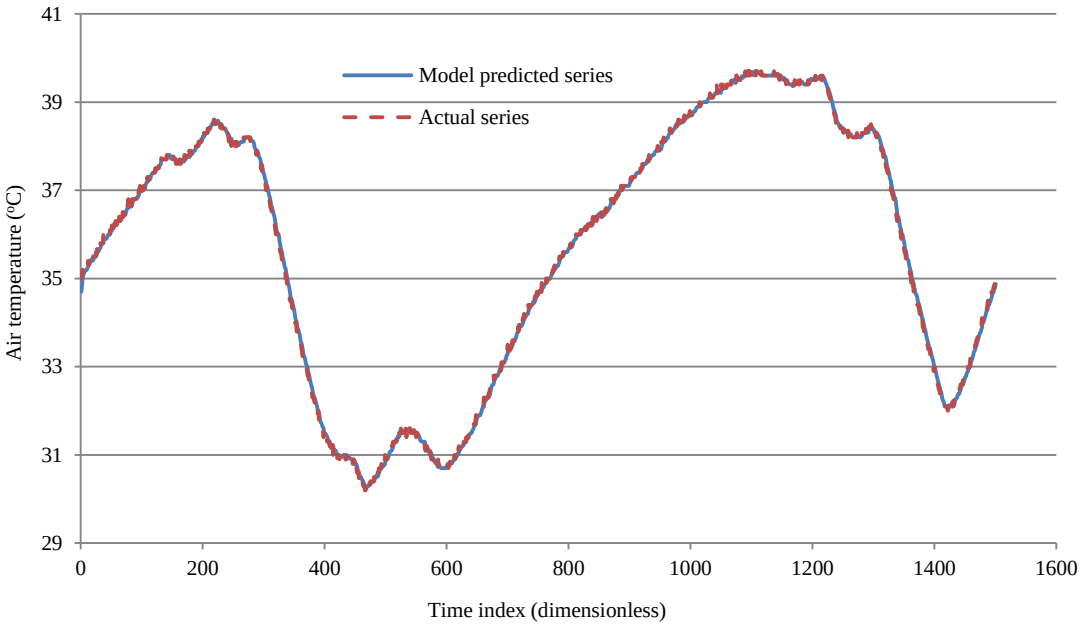


Figure 46c: Graph of first 1500 actual observations of air temperature against their predicted values

Resulting residuals were investigated. Figure below shows its histogram while table below provides estimates of its descriptive statistics.

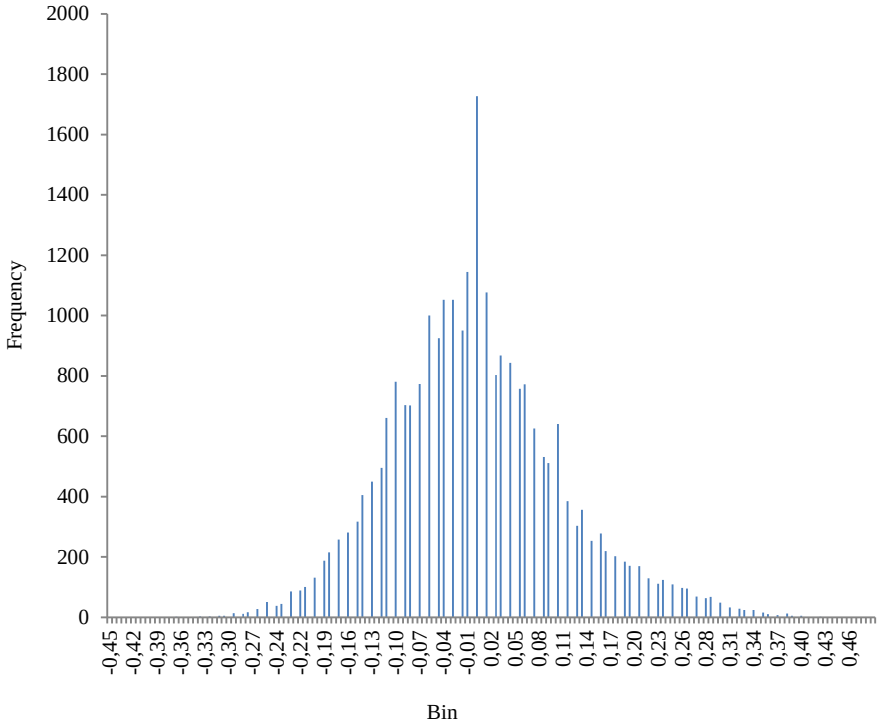


Figure 46d: Distribution of air temperature residuals obtained after model fitting

Table 27: Descriptive statistics estimates of air temperature residuals

Statistics	Estimate
Mean	0.00
Median	0.00
Mode	0.00
Variance	0.01
Standard deviation	0.11
Kurtosis	0.57
Skewness	0.42

From figure 46d and Table 27, it would be seen that resulting residuals are normal distributed. However, it should be noted that in Box Jenkins methodology, it is not mandatory that residuals be normal distributed. Residuals can assume any probability distribution but they must be stationary and independent with zero mean and finite variance [68].

For PAR1, PAR2 and PAR3, ten was added to their observations before being logarithmic-transformed. This is required in order to achieve stationarity. It is in line with additive rule of logarithmic-transformation which requires the addition of positive integers in cases where values being transformed are zero or near zero. However, when absolute values are required, back-transformation of values is performed.

Using Box-Jenkins methodology, model orders and their coefficients for PAR3, PAR2 and PAR1 were estimated. Their results are summarized in Tables 28, 29 and 30 respectively. Similarly, their model equations are shown in equation (60) for PAR3, equation (61) for PAR2 and equation (62) for PAR1.

Table 28: ARIMA model orders and parameters – PAR3

ARIMA Type/Order = ARIMA(1,0,1)		
ARIMA Parameters	**Estimates**	**Standard Error**
Constant	1.910	0.116000
AR Lag 1	0.999	0.000074
Difference	0.000	0.000000
MA Lag 1	0.282	0.002000

Table 29: ARIMA model orders and parameters – PAR2

ARIMA Type/Order = ARIMA(1,0,1)		
ARIMA Parameters	**Estimates**	**Standard Error**
Constant	1.909	0.105000
AR Lag 1	0.999	0.000077
Difference	0.000	0.000000
MA Lag 1	0.403	0.002000

Table 30: ARIMA model orders and parameters – PAR1

ARIMA Type/Order = ARIMA(1,0,1)		
ARIMA Parameters	**Estimates**	**Standard Error**
Constant	1.846	0.098000
AR Lag 1	0.999	0.000076
Difference	0.000	0.000000
MA Lag 1	0.365	0.002000

$$X_t = 1.910 + 0.999X_{t-1} + \epsilon_t - 0.282\epsilon_{t-1} \quad (60)$$

$$X_t = 1.909 + 0.999X_{t-1} + \epsilon_t - 0.403\epsilon_{t-1} \quad (61)$$

$$X_t = 1.846 + 0.999X_{t-1} + \epsilon_t - 0.365\epsilon_{t-1} \quad (62)$$

Histogram plots of PAR3, PAR2 and PAR1 residuals are shown in figures 47a, 47b and 47c respectively. Their descriptive statistics estimates are presented in Table 31.

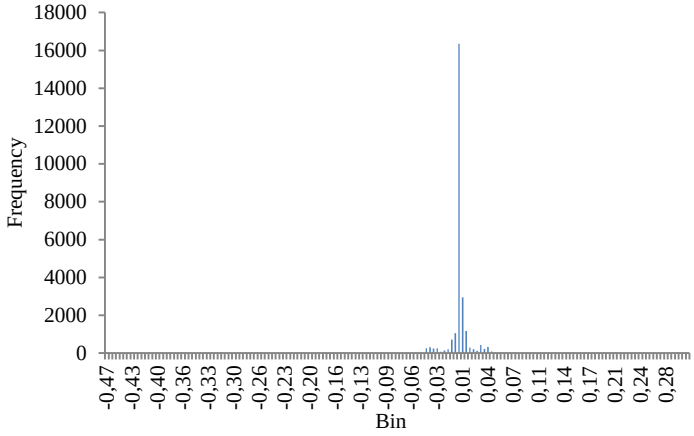


Figure 47a: Histogram plot of PAR3 residuals

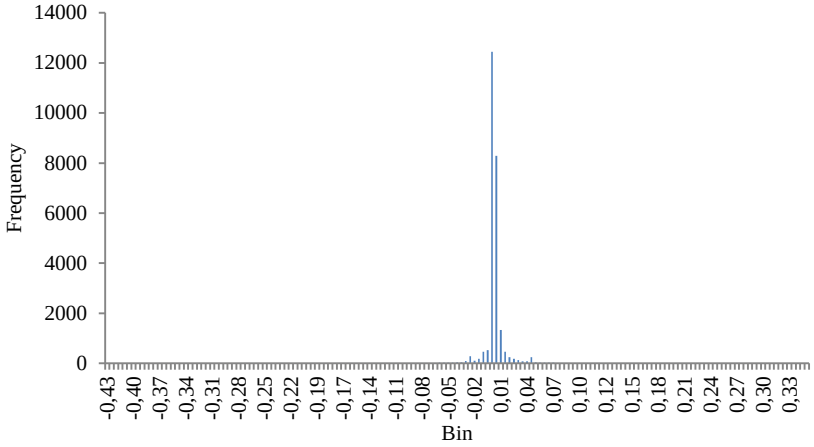


Figure 47b: Histogram plot of PAR2 residuals

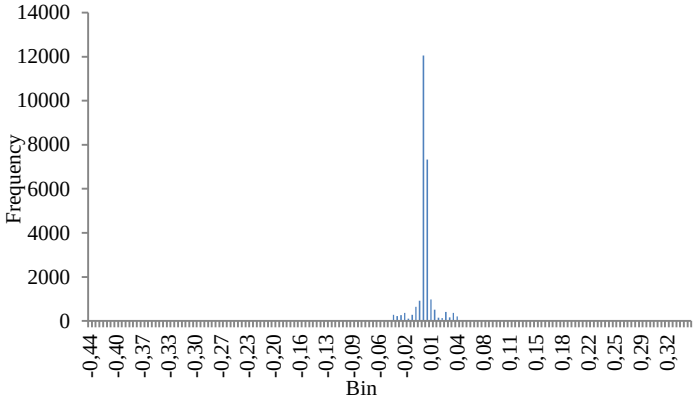


Figure 47c: Histogram plot of PAR1 residuals

Table 31: Descriptive statistics estimates of PAR3, PAR2 and PAR1 residuals

PAR3		**PAR2**		**PAR1**	
Statistics	**Estimate**	**Statistics**	**Estimate**	**Statistics**	**Estimate**
Mean	-0.0002	Mean	-0.0003	Mean	-0.0002
Median	-0.0010	Median	-0.0016	Median	-0.0009
Variance	0.0003	Variance	0.0003	Variance	0.0004
Standard deviation	0.0168	Standard deviation	0.0174	Standard deviation	0.0184
Kurtosis	91.4797	Kurtosis	113.8473	Kurtosis	99.2863
Skewness	-2.1378	Skewness	-2.0748	Skewness	-1.9228

It would be seen from figures 47a, 47b and 47c that residuals of PAR3, PAR2 and PAR1 are not normal distributed. Their computed values of kurtosis shown in Table 31 further support this since they exceed 1.5. As noted earlier, variables are reasonably close to normal distribution, and considered being normal distributed if both skewness and kurtosis lie between -1.5 and +1.5 [61, 62]. It is acceptable in Box Jenkins methodology for the residuals to assume any distribution provided they are stationary and independent with zero mean and finite variance. Therefore, the residuals do not need to be further analyzed or modeled using GARCH model, etc. However, their probability distribution needs to be characterized. A few paragraphs below have been denoted to this. The need for this is due to the fact that most symmetrical probability distributions look-alike. This would be seen in figure below.

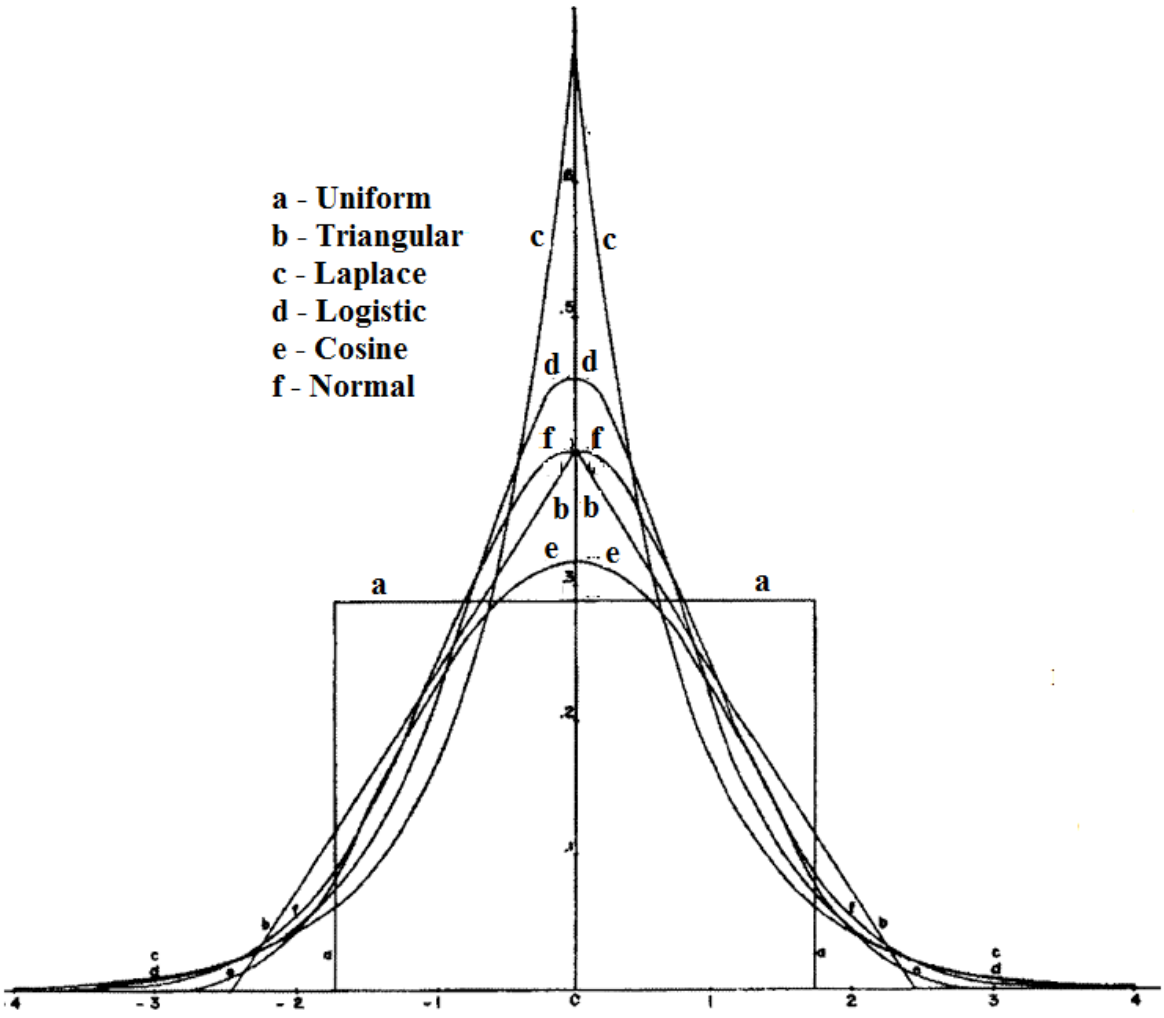


Figure 48: Graph of density function of some distributions (Culled from Wilson [69])

Additionally, Wilson [69] did emphasize the need for accurate analysis of data and residuals, and role they play in determining their statistical distributions. According to Wilson, it is unsafe to rush ahead and apply the second law of Laplace (i.e. normal distribution) or any of its various extensions without proper analysis of data.

Therefore, if S_1, S_2, S_3, S_4 represent mean deviation, mean squared deviation, mean cubed deviation and fourth-power of mean deviation respectively, and ratio of S_1: S_2: S_3: S_4 nearer to the ratio 1.000: 1.414: 1.817: 2.213 than the ratio 1.000: 1.253: 1.465: 1.645, then the

distribution is nearer to Laplace first law (i.e. Laplace distribution) than Laplace second law (i.e. normal distribution) [69]. Residuals of PAR3, PAR2 and PAR1 were found to be nearer to Laplace first law based on above statement.

Furthermore, ratio of mean deviation of PAR3 residuals to standard deviation of PAR3 residuals was computed as 0.5. Those for PAR2 and PAR1 were computed as 0.45 and 0.6 respectively. They were all nearer to 0.707 than 0.798. This further confirms that residuals are Laplace distributed since Laplace distributions have ratio of mean deviation to standard deviation close to 0.707 while normal distributions have it close to 0.798 [69].

Having established that residuals are Laplace distributed, scale and location parameters of the Laplace distributions were estimated. According to Norton [70], location parameter of Laplace distributed iid series is its median while the scale parameter is mean deviation of the series. Consequently, scale parameter was determined using equation (63a) while equation (63b) shows probability density function of a Laplace distribution.

$$\beta = \frac{1}{N} * \sum_{t=1}^{N} \left| X_t - \overline{x} \right| \quad (63a)$$

$$P(x) = \frac{1}{2\beta} e^{\frac{-(x-\mu)}{\beta}} \quad (63b)$$

where:

β represents scale parameter,

N represents number of residuals,

X_t represents residual at time *t*,

$\overline{x}$ represents mean of residual distribution,

μ represents location parameter,

$P(x)$ represents probability density function.

The table below shows values calculated for location and scale parameters of the residuals.

Table 32: Location and scale parameter estimates of the residuals

Variable whose residuals distribution is being characterized	**Median**	**Estimated location parameter**	**Estimated scale parameter**
PAR3 ($\mu molm^{-2}s^{-1}$)	-0.0010	-0.0010	0.00839
PAR2 ($\mu molm^{-2}s^{-1}$)	-0.0016	-0.0016	0.00781
PAR1 ($\mu molm^{-2}s^{-1}$)	-0.0009	-0.0009	0.0110

Having fitted models to the variables (air temperature, PAR3, PAR2 and PAR1), they are validated and used in estimation of missed and future values. How this is implemented is discussed in section 6.1.2.

5.4 Artificial Neural Network

Artificial neural network was discussed in detail in section 3.2.3.6 were all steps involved were presented. An example was also provided to demonstrate how it is used to model time series data.

In this section, artificial neural network is used to train, test and validate field data discussed in section 5.1. Since four variables (air temperature, PAR1, PAR2 and PAR3) are involved, four separate networks are therefore required. Each network consists of an input of one neutron, one hidden layer and an output of one neutron. Number of neutrons for hidden layer was set to standard default value of 10. Considering equation (64) which is usually used in determining number of weights in neural network, it follows that each network has 31 weights.

$$N_w = (n * h) + h + (h * m) + m \qquad (64)$$

where:

N_w represents number of weights,

n represents number of input nodes,

h represents number of hidden nodes,

m represents number of output nodes.

Levenberg-Marquardt algorithm was considered and used as training function. Training process helped assign values to weights so that residuals are minimized.

Data selection method was in line with rule of thumb for training-validating and testing in neural network. It requires that 60 percent of datasets be used to train the network and remaining datasets shared between validations and testing. This selection method is called test-technique. On the other hand, all-train technique uses all datasets for training while train-test technique uses 80 percent of dataset for training and the remaining for testing.

Structure of the network is presented below while performance results obtained with sampling interval of 30 seconds are summarized in Table below. Similarly, figure 50 depicts testing and validation results for PAR1. Plots for PAR2, PAR3 and air temperature are shown in figures 51 – 53 respectively.

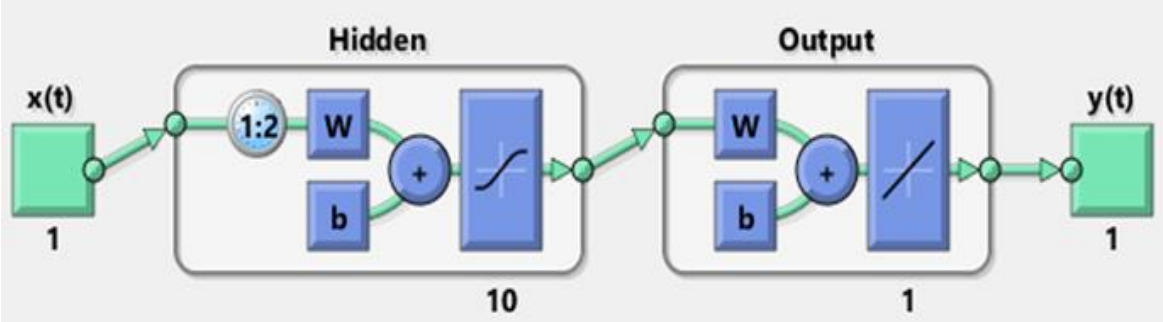


Figure 49**:** Network structure used for each variable

Table 33: Performance summary table

Parameters	**PAR1 (µmolm^{-2}s^{-1})**	**PAR2 (µmolm^{-2}s^{-1})**	**PAR3 (µmolm^{-2}s^{-1})**	**Air Temperature (^{o}C)**
Epoch (iterations)	7	8	8	12
Performance (MSE)	0.0146	0.0139	0.0145	0.0115
Gradient	0.00210	0.00178	0.00110	0.00341
Validation checks	6	6	6	6
Overall MSE performance	0.0225	0.0202	0.0199	0.0190

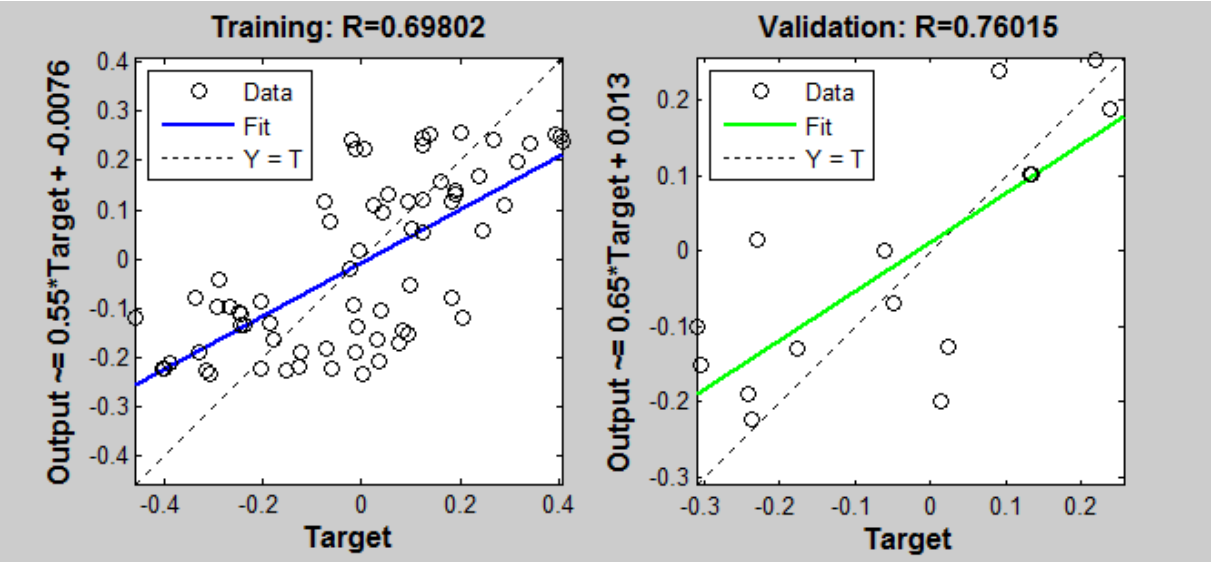


Figure 50: Training and validation of PAR1 in (µmolm^{-2}s^{-1}).

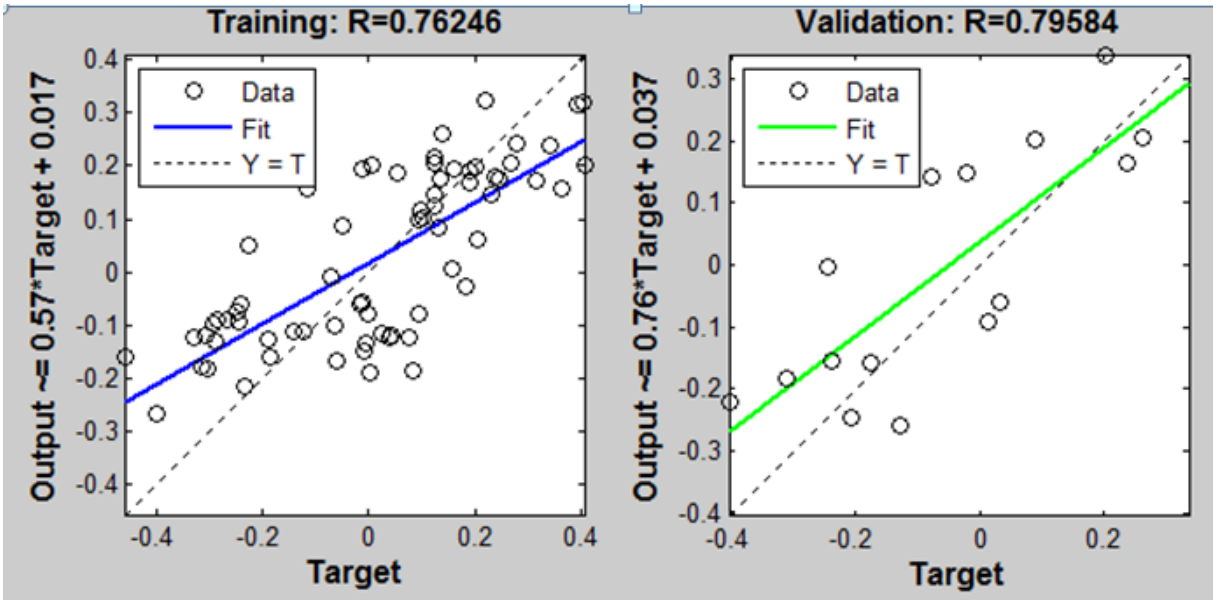


Figure 51: Training and validation of PAR2 in (µmolm^{-2}s^{-1}).

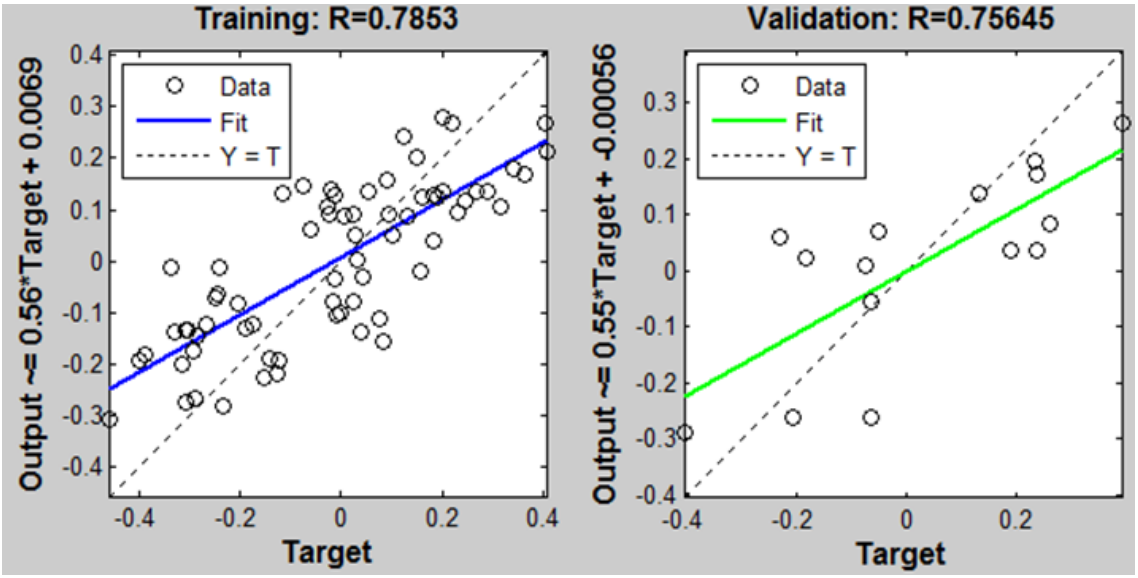


Figure 52: Training and validation of PAR3 in (µmolm^{-2}s^{-1}).

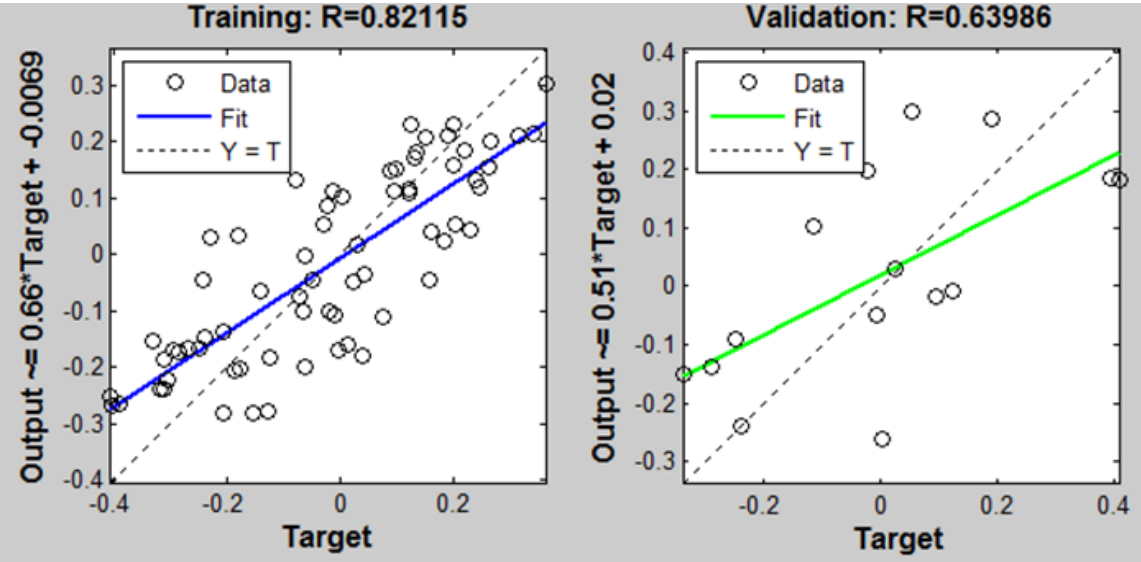


Figure 53: Training and validation of air temperature in (°C)

From Table 31, it would be seen that PAR1 iteration stopped after 7 training epochs. It resulted in mean square normalized error (MSE) less than 0.015. Also its overall MSE performance was 0.0225. Similar results apply to other variables. None of the variables gave an MSE greater than 0.015. Also, their overall MSE performance never exceeded 0.025. All these indicate good performance in computation and that of the model.

Furthermore, regression plots showed good R-values which indicate good fit for each of the variables. Also the regression lines did not deviate significantly from the dotted center line which is the ideal (perfect) regression line where output equals target with R-value as 1. Relationship between outputs of the network and targets was deduced for each of the variables. This helps in determination of outputs for given targets. Targets were normalized between -0.4 and 0.4 so as to secure all input data within range.

CHAPTER SIX
Main Contributions of the Thesis

6.1 New Adaptive Sampling Method

6.1.1 Description of method

A new adaptive sampling method for measurement of environmental variables was proposed and evaluated using real environmental field data. The new method allows sampling with high sampling interval (low sampling frequency) so that energy consumed is reduced. To balance trade-off on data quality, the new method requires accurate stochastic model of variable be estimated so that forward and backward predictions are performed accurately. It also requires that its stochastic model be continually updated to ensure alignment with changes in input data.

Stochastic models obtained while using this new method are used in determining optimal sampling interval. This is discussed in section 6.2. On top level, the new adaptive sampling method is composed of following steps:

1. Estimate stochastic model that fits measured dataset - Few measurements are taken using high sampling interval and appropriate ARIMA model fitted to dataset.
2. Compute residuals - Residuals are computed using predicted and measured values. Ensure that residuals are stationary and independent so that fitted ARIMA model could be used to estimate missed and future values.
3. Update model – This is achieved by re-estimating model parameters while keeping stochastic orders of model constant. This holds for cases where time series and its stochastic model represent a defined period such as summer period or winter period, etc. However, if residuals deviate significantly from threshold interval (i.e. if it deviates by more than four multiples of threshold interval), both stochastic orders and model parameters are updated.

Explicitly, the method allows skip of sensor sampling and use of model-predicted values when an adaptive sampling condition is met. On a lower level, the new adaptive sampling method is guided as follows:

1. Take few measurements of the variable.
2. Determine the stochastic model and compute residuals. Numerical value of the threshold interval is also computed. This is value representing 95^{th} percentile of residuals distribution. Thus, if 95^{th} percentile of residuals distribution is determined as, say $k1_{inital}$, threshold interval is then set as $[-k1_{initial}, +k1_{initial}]$.
3. Let sensor continue sampling. The program computes new $k1$ value (and new threshold interval) each time sensor samples. The new value ($k1_{new}$) is compared with initial (stored) value computed in step 2. Result from the comparison is noted and stored (i.e., is $-k1_{initial} \leq k1_{new} \leq +k1_{initial}$).
4. Step 3 is repeated four more times, each time comparison performed and result stored.
5. If at least 3 cases out of the 5 cases (times) of step 3 gave threshold intervals that are within predefined range computed in step 2, sensor stops sampling and resumes

sampling after time for 20 samplings has elapsed. The program returns to step 3 after this time has elapsed.

6. However, if not up to 3 cases out of the 5 cases (times) of step 3 gave threshold intervals that are within predefined range computed in step 2, sensor continues sampling. The program only stores sampled data. It stops computing new *k1* values and performing comparisons, until after time corresponding to 20 samplings has elapsed. The program returns to step 3 after this time has elapsed. However, before program returns to step 3, it uses stored sampled data to update model and threshold interval (that is, step 2 is performed again).
7. During time period described in step 5 when sensor stopped sampling, if need for missed data is required, program computes and releases model-predicted value.

The above is described in flow chart shown below.

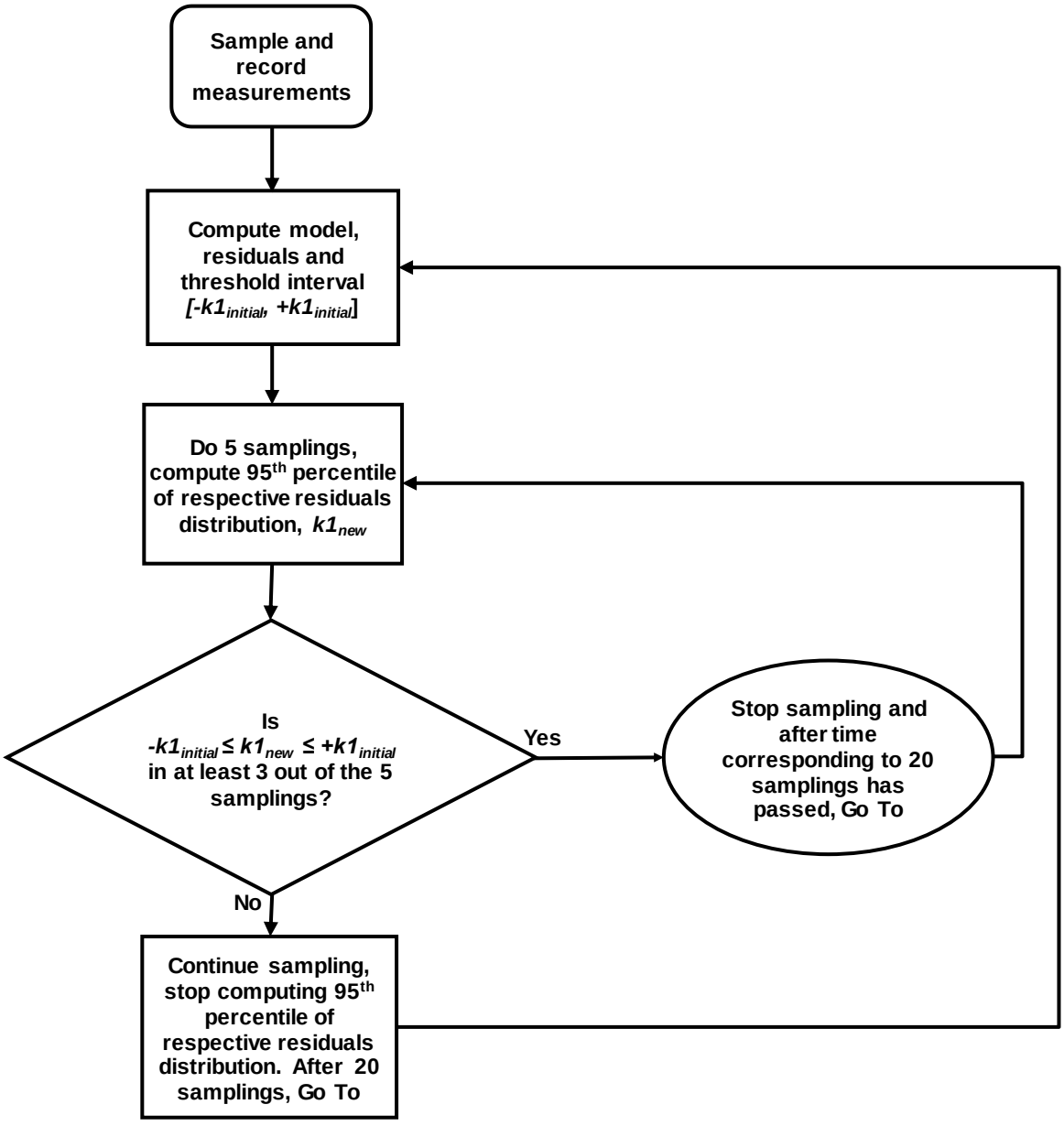


Figure 54: Flow chart describing the new adaptive sampling method

The new adaptive sampling method when commercialized shall be useful in numerous engineering applications and industrial processes, among others. Of particular interest is the automotive industry where current drive to reduce energy consumption and limit sensing activities of some variables in autonomous vehicles is in focus. Further details relating to the new adaptive sampling method are available in Ezeora et al [71].

6.1.2 Estimation of Missed Values

A stationary time series looks the same if moved forward or backward in time. Consequently, the same model used for forward predictions can be used for backward predictions. In order to estimate missed values, two cases were considered. The first case considered values missed due to use of high sampling interval (for example, sampling every 30 seconds instead of 15 seconds). For such cases, Box Jenkins models such as ARIMA cannot be used. This is because they require time index to be a positive or negative integer. If time index is an integer, data required in their model equation is available.

Illustrating further, assume a sampling interval of 30 seconds. At time index 1 (i.e. start of sampling), sampled (measured) value is known. At time index 2 (i.e. 30 seconds away from start), measured value is also known. For time index 1.5 (15 seconds away from start), ARIMA model equation cannot be used since data are not available for non-integer time indexes. For this type of case, techniques such as use of numerical interpolation methods are considered.

On the other hand, the second case concerns values missed at time indexes that are integer multiples of sampling interval. Illustrating further, assume the sampling interval is 30 seconds. At time index 1 (i.e. start of sampling), sampled (measured) value is known. Assuming that at time index 2 (30 seconds after starting), measured value was missed due to reasons such as adaptive sampling. Assuming also that at time index 3 (60 seconds after starting), measured value was taken. Thus, Box Jenkins models can be used to estimate missed value at time index 2. This is because time index 2 is an integer multiple of the sampling interval.

An excerpt for first case discussed above is shown in Table 34. It shows results obtained using numerical interpolation methods. Similarly, an excerpt for the second case is shown in Table 35.

Table 34: Estimation of missed values due to high sampling interval

Date (ddmmyy)/ Time (hh:mm:ss)	Measured Temp. (^oC)	Measured PAR1 ($\mu molm^{-2}s^{-1}$)	Linear Inter. Temp. (^oC)	Linear Inter. PAR1 ($\mu molm^{-2}s^{-1}$)	Cubic Inter. Temp. (^oC)	Cubic Inter. PAR1 ($\mu molm^{-2}s^{-1}$)
10-06-2015 13:09:15	30.7	195.0	-	-	-	-
10-06-2015 13:09:30	Missed value	Missed value	30.65	182.50	30.65	179.69
10-06-2015 13:09:45	30.6	170.0	-	-	-	-
10-06-2015 13:10:00	Missed value	Missed value	30.60	166.00	30.60	167.24
10-06-2015 13:10:15	30.6	162.0	-	-	-	-
10-06-2015 13:10:30	Missed value	Missed value	30.60	161.00	30.60	158.21

10-06-2015 13:10:45	30.6	160.0	-	-	-	-
10-06-2015 13:11:00	Missed value	Missed value	30.60	178	30.60	171.12
10-06-2015 13:11:15	30.6	196.0	-	-	-	-
10-06-2015 13:11:30	Missed value	Missed value	30.55	355	30.57	238
10-06-2015 13:11:45	30.5	514.0	-	-	-	-

Key: Temp. = Air temperature; Cubic Inter = Cubic spline interpolation; Inter = Interpolation

Table 35: Estimation of missed values due to adaptive sampling

Date (dd-mm-yy)/ Time (hh:mm:ss)	Measured Temp. (^{o}C)	ARIMA Estimated Temp. (^{o}C)	Linear Inter. Temp. (^{o}C)	Cubic Inter. Temp. (^{o}C)
10-06-2015 13:09:15	30.7	-	-	-
10-06-2015 13:09:45	AM (30.6)	30.65	30.65	30.65
10-06-2015 13:10:15	30.6	-	-	-
10-06-2015 13:10:45	AM (30.6)	30.60	30.60	30.60
10-06-2015 13:11:15	30.6	-	-	-
10-06-2015 13:11:45	AM (30.5)	30.60	30.60	30.61
10-06-2015 13:12:15	30.6	-	-	-
10-06-2015 13:12:45	AM (30.5)	30.56	30.60	30.60
10-06-2015 13:13:15	30.6	-	-	-

Legend: Temp. = Air temperature; Cubic Inter = Cubic spline interpolation; AM (XX) = Assumed missing and XX represents the measured value if it were not assumed missing.

For case due to high sampling interval, it would be seen in Table 34 that linear interpolation gives better estimation results. For case due to adaptive sampling, it would be seen in Table 35 that ARIMA performs better.

6.1.3 Validations

The new adaptive sampling method was validated through a step-by-step approach that establishes validity of constituents. The models were validated based on:

- Evaluation of performance parameters of models and their residuals. Results were compared with those obtained using other methods. Performance parameters used include standard error, coefficient of determination, mean and variances, etc. Mean and variance estimates computed using the new method were compared with those computed using bootstrap technique. They were found to differ by less than 5%. Similarly, standard error estimated using the new method was found to be close to those estimated using bootstrap method.
- Checking with command syntax in order to identify how many measured values lie outside the lower and upper confidence interval limits of model predicted values. It was found that more than 95 percent of measured values lie within the limits.

- Analysis discussed and shown in Table 35 above contributes in validating the method.

6.2 Determination of Optimal Sampling interval

In addition to deterministic method of finding optimal sampling interval discussed in section 4.2.1, a new stochastic method was proposed and discussed. The new method requires that some measured values be available. This is needed in order to determine the stochastic model of the variable. The new stochastic method takes into account dynamics associated with the variable being measured. It considered properties of the variable in relation to process and system used in measuring the variable.

These are dynamics which vary from system to system, and from variable to variable. Such dynamics are often difficult to be represented accurately in mathematical form. This is one of the reasons why the new method is data-driven instead of being based completely on first principle approach.

With the new method, continuous measurement is not required after stochastic model has been determined. Future values are estimated using model equation. Only few measurements at certain time instants are taken so that model is checked and updated if necessary. This may also require update of optimal sampling interval. Thus, using estimated stochastic model, optimal sampling interval is determined. This is achieved by finding a local maximum time index that allows new measured values to fall within the acceptable tolerance of previously measured value. The acceptable tolerance is obtained using maximum permissible error (MPE). Based on well-known 1:3 ratio of MPE - uncertainty relation, maximum permissible error (MPE) is expected to be three times more than uncertainty. This relation is based on international standards of ISO 10012:1998, ISO 22514-7:2012, OMIL-R and EA-4/02 M: 2013 [72 - 75].

Uncertainty, which is related to accuracy of measurement, is deduced from technical datasheet of measuring device. Consequently, maximum permissible error (i.e. tolerance) of measured value can be computed. It then follows that all values of a variable that are within tolerance range of its measured value are acceptable based on international standard. Therefore, optimal sampling interval is that maximum length of time for which measured values are still within the acceptable tolerance range of previously measured value. For example, considering data on technical datasheet for Hobo sensor S-THB-M002, its maximum permissible error should be $(\pm 0.2)*(3) = \pm 0.6\ ^{o}C$. Hence, measured temperature value of say, $20^{o}C$ has temperature values within $20 \pm 0.6^{o}C$ acceptable as accurate and close enough to nominal value of $20^{o}C$.

The question becomes: "why set sampling interval of air temperature involving this instrument to perform several samplings (measurements) within this range $(20 \pm 0.6^{o}C)$?" Considering above discussion and need to minimize energy consumption, one sampling (measurement) suffices for the range $20 \pm 0.6^{o}C$. Air temperature values that lie within the range $(20 \pm 0.6^{o}C)$ are considered as good as the nominal value $(20^{o}C)$. Any value within the tolerance range can stand in for others. This contributes in reducing energy consumption of the system.

However, in systems where responsiveness to small temperature changes is mandatory, repeated temperature measurements are usually required within the tolerance range. For such systems, smaller instrument with lower uncertainties and energy consumption should be

considered. Alternatively, temperature values could be predicted from measured values. This obviously requires a model. Model development methods and steps discussed in this work are recommended. The table below shows an excerpt analysis for air temperature measurement taken on 2015-05-31 at 22:53:15.

Table 36: Optimal sampling interval – air temperature

Date, Time	Measured value (^{o}C)	Acceptable tolerance range (^{o}C)	Predicted value (^{o}C)
2015-05-31 22:53:15	16.5	15.9 - 17.1	16.5
2015-05-31 23:24:45			15.9

From Table 36, it would be seen that for a measured value of 16.5^{o}C, length of time corresponding to its lower limit of acceptable tolerance range was estimated using stochastic model. It was found to be 1890 seconds (31.5 minutes). The upper limit value could as well be used and backward prediction performed. This is because air temperature values are decreasing with time during the period investigated.

The analysis was repeated using other measured values of the variable. All results obtained were in agreement. Model used in this analysis was based on data obtained using an initial 30 seconds sampling interval. Therefore using above described system, values obtained when air temperature is measured at 31 minutes time interval are within the acceptable tolerance range of immediate past measured value. Therefore, 31 minutes is considered as the optimal sampling interval. It gives reduction in energy consumption while guaranteeing data quality.

The above analysis is repeated after some time using measured data. This is required in order to update optimum sampling interval in event significant changes occur in measured values. For cases involving big data (large and varied datasets), even though stationarity is achievable during model fitting, it is recommended that datasets be classified based on homogeneity before being modeled. This will reduce variability and simplify models. It will also result in several optimal sampling intervals. Impact of various sampling intervals on energy consumption has been discussed in section 5.2. For extended reading on determination of optimal sampling interval, please see Ezeora et al [76].

6.3 New Event-Triggered Sampling Method

6.3.1 Description of method

A new event-triggered sampling method was proposed and its algorithm developed. Event-triggered sampling just like adaptive sampling belongs to the same group known as energy-aware data acquisition method (see figure 1 in chapter one). Both methods are event-driven. They take into account past information. Unlike the conventional (periodic) sampling method which is open-loop process, event-triggered and adaptive sampling methods are closed-loop processes. This is because data stored in memory are compared against input data appearing at sampler or sensing unit. Rate at which this is performed is among the distinguishing factors between event-triggered and adaptive sampling methods. Event-triggered sampling is quite popular and widely used in applications where quality of data is of main focus.

In traditional event-triggered sampling method, comparison is made once there is data at sensor input. The trigger mechanism is activated and it compares data retrieved from memory against data appearing at sensor input. Based on outcome of the comparison, decision is made

whether data appearing at sensor input should be sampled or not. In certain designs, feedback circuits comprising of hardware components such as digital-to-analog converters are involved. They help restore data retrieved from memory to be the same form as the analog data appearing at sensor input.

In this work, new event-triggered sampling method was proposed and discussed. Unlike the conventional event-triggered sampling method, adjustments were made in the new method in order to improve energy-efficiency. One key adjustment is use of software-only solution working with limited hardware. Also, whenever sampling interval is increased, length of time for sensor subunit to hibernate (be inactive) is increased since it must make one sensing within the adjusted sampling interval. It takes only a few milliseconds for the subunit to sense data. Therefore, the new event-triggered sampling method works to reduce total energy consumption while maintaining data quality.

Linear time algorithms that implement the new method were developed and demonstrated using PAR and air temperature data obtained from Brazil. The algorithms use statistical changepoint detection methods to determine changepoints. Changepoints represent states within a series where numerical values of certain statistical properties of the series change. The statistical properties may be mean, variance, etc. With fixed sampling rate, data are analyzed and compared with estimate. This is to determine if changepoint has occurred. If established that changepoint has occurred, the algorithm activates a signal which controls ADC clock so that clock frequency is adjusted. Sampling interval is also adapted accordingly and sensor senses within the adjusted sampling interval. This adjusts the duty cycle. As a result, energy is efficiently managed.

6.3.2 The Algorithms

Program and data used in computing and comparing changepoints are saved in a memory array. As a result, time required to retrieve and update data from the array is linear. This reduces run time of the algorithm. In addition, the algorithm uses only one "*for*" loop, further contributing to linear run time. The table below shows algorithms for PAR and air temperature.

Table 37: Algorithms for PAR and air temperature

Algorithm for PAR	**Algorithm for air temperature**
1: ADC_{cycle} = 30 {Initialize ADC total cycle to 30s}	1: ADC_{cycle} = 30 {Initialize ADC total cycle to 30s}
2: let *Sleep* = Inactive {initialize sleep}	2: let *Sleep* = Inactive {initialize sleep}
3: let *S* = [] {initialize empty array to store values}	3: let *S* = [] {initialize empty array to store values}
4: **for *i*** = 1 to *n* **do**	4: **for** *i* = 1 to *n* **do**
5: **if** *S[i]%2* = 0 ∧ *S[i] - S[i-1]* = 2 then	5: **if** *S[i]%2* = 0 ∧ *S[i] - S[i-1]* = (-2…+2) ∧ *S[i]* <18 **then**
6: ADC_{cycle} = 300 {let ADC total cycle = 300s}	6: ADC_{cycle} = 300 {let ADC total cycle = 300s}
7: while $ADC_{cycle} \leq 300$ ∧ Data Register = TRUE **do**	7: **while** $ADC_{cycle} \leq 300$ ∧ Data Register = TRUE **do**
8: let *Sleep* = Active	8: let *Sleep* = Active
9: **end while**	9: **end while**
10: **else**	10: **else**
11: *S = S + 1* {update S[]}	11: *S = S + 1* {update S[]}
12: **end if**	12: **end if**
13: *increment* - ADC_{cycle}	
14: **end for**	

	13: *increment - ADC_{cycle}* 14: **end for**

Similarly, Table 38 provides an overview of estimated daily energy consumption of a typical ADC (AD7980) using field data from Brazil discussed in case study I.

Table 38: Estimated 1-day energy consumption of an ADC without (w/o) and with (w/) proposed algorithm.

Sampled period (every 30 seconds)	**PAR (W/m^2)**	**Energy consumed (mJ)**	
		Without algorithm	**With algorithm**
Night 09-09-02 18:00:30 09-09-03 06:00:00	Constant at 1.2 (discussed in section 4.2.1)	*10.06* *(A1)*	*1.03* *(A2 + A3 + A4)*
Day 09-09-03 06:00:30 09-09-03 18:00:00	Normal distributed (discussed in section 4.2.1)	*10.06* *(A1)*	*10.07* *(A1)*
	Total	*20.12*	*11.10*

where:

$$A1 = 1439 * 1000 * (10^{-9}) * 30s * (10^{6}) * (0.000233 * 10^{-3}W) = 10.06\, mJ \quad (65)$$

$$A2 = 4 * 1000 * (10^{-9}) * 30s * (10^{6}) * (0.000233 * 10^{-3}W) = 0.03\, m\, J \quad (66)$$

$$A3 = 72 * 1000 * (10^{-9}) * 300s * (10^{6}) * (0.0000233 * 10^{-3}W) = 0.50\, mJ \quad (67)$$

$$A4 = 71.5 * 1000 * (10^{-9}) * 300s * (10^{6}) * (0.0000233 * 10^{-3}W) = 0.50\, mJ \quad (68)$$

From the above, it would be seen that daily savings up to 45% of ADC energy consumption would be achieved while using the proposed algorithm. Performance of the algorithms was evaluated with favorable results. Further details are available in Ezeora et al [77].

6.4 Energy Analysis

In order to understand better how energy is consumed within sensing unit, energy analysis was performed. This is vital since it reveals components whose energy consumption is affected most by changes in sampling interval. This is helpful since it could allow the use reverse engineering methodology in determining optimum sampling interval. That is, energy consumption being a function of sampling interval could be treated as a minimization problem. Once solutions to constraint optimization problem are known, optimum sampling interval could be deduced. Constraint minimization of energy consumption is not performed in this work. It is outside the scope of this work.

Sensing unit shown in figure 4 of section 1.4 was considered for the energy analysis. Its energy model is sum of energy consumption of various modules that constitute the unit. This is shown in equation (69).

$$E_{tot} = E_{sen} + E_{sam} + E_{com} + E_{cap} + E_{con} \quad (69)$$

where:

E_{tot} represents total energy consumed per conversion,

E_{sen} represents energy consumed by sensor per conversion,

E_{sam} represents energy consumed by charged sampling capacitor per conversion,

E_{com} represents energy consumed by ADC comparator per conversion,
E_{cap} represents energy consumed by ADC capacitor array per conversion,
E_{con} represents energy consumed by ADC control logic per conversion.

E_{sen} constitutes resistive losses dissipated during the sampling (charging) mode. It is usually due to sensor output resistance and ADC input resistance. E_{sen} is expressed in equation below.

$$E_{sen} = \frac{(V_{ref} - V_{nom})^2}{(R_o + R_i)} * T_{sample} \qquad (70)$$

where:
V_{ref} represents the reference voltage in volt,
R_o represents sensor output resistance in ohms,
R_i represents ADC input resistance in ohm,
V_{nom} represents voltage at ADC input. This considers voltage drop due to R_o and R_i,
f_{clock} represents ADC clock frequency in Hertz,
N_{sample} represents number of clock periods or cycles made during the sampling mode when sampling capacitor is been charged,
$T_{sample} = \frac{N_{sample}}{f_{clock}}$. It represents time required to charge the sampling capacitor. This is the same as the acquisition time of the ADC.

An illustration to help understand the above is provided in figure 55. t_p shown in figure 55 represents time required to complete one clock cycle.

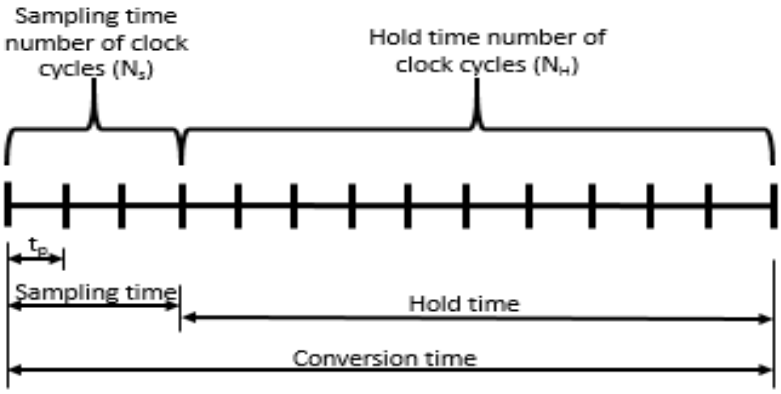


Figure 55: SAR ADC - Sampling time and hold time

Similarly, E_{sam} is the energy used to charge sampling capacitor. This is twice the energy stored in the charged sampling capacitor. This is due to the fact that energy in form of heat, is always lost during charging of capacitors due to parasitic resistance effect (capacitor equivalent series resistance). Therefore energy stored in a charged capacitor is always half of the energy utilized in charging it. E_{sam} is represented in equation below.

$$E_{sam} = C_{sample} * (V_{nom})^2 \qquad (71)$$

where:
C_{sample} represents sampling capacitor in Farad,
V_{nom} represents the analog voltage at ADC input.

Furthermore, E_{com} is energy dissipated by the comparator. Dissipated energy did not consider bubble errors at comparator output. It also did not consider losses due to non-idealities such as jitter, switching and flicker noise. According to Fredenburg and Flynn [29], E_{com} is represented in equation below.

$$E_{com} = S_{F*} C_o * (V_{dd})^2 \qquad (72)$$

where:

C_o represents ADC output capacitance in Farad,

V_{dd} represents ADC digital voltage supply in volt,

N represents ADC resolution,

$S_F = \frac{4}{23}(N^2 + 12N + 2)$. This represents scaling parameter.

Similarly, E_{cap} is energy consumed by DAC block. DAC block is a set of binary-scaled array of capacitors which converts most significant bits to analog voltage references. This is required so that comparator can successively compare analog voltage references with input voltage. According to Ginsburg [28], E_{cap} is represented by equation (73).

$$E_{cap} = 2^{b+1} * \eta * C_u * V_{dd} * V_{fs} \qquad (73)$$

where:

η represents efficiency; usually set to 0.7,

b represents number of binary-scaled capacitors,

C_u represents unit capacitance,

V_{fs} represents full-scale input voltage,

V_{dd} represents digital voltage supply.

Additionally, E_{con} is energy consumed due to logical operations of ADC. It did not contain energy consumed due to fanout. Fanout was not considered because capacitive load of capacitor switches are low. According to Fredenburg, J., and Flynn [29], E_{con} is represented by equation (74).

$$E_{con} = b_w * C_{sw} * (V_{dd})^2 \qquad (74)$$

where:

C_{sw} represents the switched capacitance,

V_{dd} represents digital voltage supply,

b_w represents width of shift registers. Control logic unit is assumed to be of shift register type.

Therefore, total energy consumed becomes sum of all energy consumption components deduced above as guided by equation (69).

CHAPTER SEVEN
Conclusions

7.1 General Conclusions

In this dissertation, both stochastic and deterministic methods of investigating field data have been presented and discussed. Stochastic methods such as Box-Jenkins method, hidden Markov model (HMM) and artificial neural network were thoroughly reviewed.

Both predictive and empirical approaches were applied in studying and analyzing field data. With predictive approach, a new stochastic method that gives optimized solutions was developed. The new method gives models that accurately predict missed and future values with minimal energy. The new method is also useful in determining optimal sampling interval. This was demonstrated using real environmental field data. Results obtained were validated and satisfactory results obtained. Using the new stochastic method gave estimates that were within the acceptable tolerance of missed data. Therefore with the new method, missed data are adequately covered since their estimated values are within the acceptable range. Data sampling using the new method results in reduced energy consumption with no loss to data quality. This fulfills the objective of this work.

On the other hand, empirical part of this work focused on sampled data. Consequently, new type of event-triggered mechanism was proposed and discussed. The mechanism investigates sampled data to determine if statistical changepoint has occurred. Based on the outcome, it evaluates and decides if sampling interval should be adjusted. Algorithm for the mechanism was developed. The algorithm activates a signal which controls ADC clock so that clock frequency is adjusted. Sampling interval is also adapted accordingly and sensor senses within the adjusted sampling interval. The algorithm has a linear run time and also uses one "for" loop. This was demonstrated using real environmental field data. The amount of energy that would be saved while using the algorithm in a typical system was estimated. Results obtained for PAR indicate a daily savings of 45% in ADC energy consumption.

Artificial neural network was also used to investigate the field data. Overall mean square normalized error performance was good with good R-square values. They indicate good fit of model to data. It was also found that datasets of PAR obtained directly under solar radiation were more homogenous than those obtained under canopy and in shadow area. This was supported by results obtained using the datasets in a descriptive statistical analysis. Descriptive statistical results revealed that datasets of PAR obtained directly under solar radiation have reduced variability than those within canopy and in shadow area.

Finally, energy savings in form of extended battery-lifetime was realized when sampling interval increased in steps. Results obtained showed that battery lifetime was extended by 87% when sampling interval was increased from 15 to 30 seconds.

7.2 Future Work

One part of the future work is software implementation of developed algorithms and its associated test on field. The algorithms should be written in low-level language such as C programming language. Thereafter compiled using any complier or with Atmel AVR studio. The resulting machine code should be programmed in a microcontroller chip such as AVR

microcontroller. The final system should be field-tested. The above is required for validation data needed for patent to be obtained. Dongmei et al [78] who developed new compressive sampling method while working within the same focus area, has already sought for US Patent. Such footstep is recommended.

In addition, proposed algorithms may have to be standardized using standard changepoint detection mathematical equations such as those of Mann-Whitney or Moody. This shall broaden scope of its application. However, caution should be exercised so that linear run-time order of the algorithm is not altered. As worst case, the algorithms may take quadratic run-time order. This is important because energy-savings effectiveness of an algorithm depends on its run-time order.

Finally, emerging deep learning methods for time series analysis such as conditional restricted Boltzmann machine have to be investigated. At present, information and tools relating to them are limited.

REFERENCES

1. Abelaal, M., et al, Improving energy-efficiency in QoS-constrained wireless sensor networks, International Journal of Distributed Sensor Network, 2016, 12 (5), p. 1-28. ISSN 1550-1477.
2. Deloitte, Energy market reform in Europe: European energy and climate policy – Achievements and challenges to 2020 and beyond, Deloitte Consell, Zurich, Switzerland, 2015.
3. Enami, N., et al, Neural network based energy efficiency in wireless sensor networks: A survey, In International Journal of Computer Science & Engineering Survey, IJCSES 2010, India: IJCSES (International Journal of Computer Science & Engineering Survey), 2010, 1(1), p. 39-55. ISSN 09763252.
4. Artiola, J., Pepper, I., and Brusseau, M., Environmental Monitoring and Characterization, 5th edition, San Diego, USA: Elsevier Academic Press, 2006. ISBN 13: 9788131200889.
5. Mottus, M. et al, Photosynthetically active radiation: measurement and modeling, in Encyclopedia of Sustainability Science and Technology, Robert A. Meyers (ed.), New York: Springer Science Business Media, 2012, p. 7902-7932. ISBN 978-0-387-89469-0.
6. Arnold, J., Laymon, C., and Samuelson, D., Soil moisture, Marshall Space Flight Center Earth Science Office Publication, National Aeronautics and Space Administration, NASA, USA, 2017. [Online]. Available: https://weather.msfc.nasa.gov/landprocess/ [Accessed: 28-March-2017].
7. Lin, L., et al, Enabling cyber physical systems with wireless sensor networking technologies, International Journal of Distributed Sensor Networks, 2012, article id - 489794, p. 21. ISSN 1550-1329.
8. Lacanette, K., Using analog temperatures with ADCs, Application Note 571, Maxim Integrated, USA: Maxim Integrated, April 1, 2001.
9. AN1636, Understanding and minimizing ADC conversion errors, Application Note, ST Microcontroller Division Applications, March 2003.
10. Poshala, P., Why oversampling when undersampling can do the job, In TEXAS Instruments Applications Report - SLAA594A, Dallas: July, 2013.
11. Wescott, T., Sampling: What Nyquist did not say and what to do about it, Wescott Seminar, Wescott Design Services, June 20, 2016.
12. Wu, M., Data prediction, compression and recovery in clustered wireless sensor networks for environmental monitoring applications, In Journal of Information Sciences, JIS 2016, United Kingdom (UK): JIS (Journal of Information Sciences), 2016, v. 329, p. 800-818. ISSN 0165-5515.
13. Logan, S., These power-saving techniques promote longer ADC system life, In Electronic Design Magazine, USA: March 24, 2011.
14. Watts, A., Musilek, P., and Wyard-Scott, L., Managing the energy-for-data exchange in remote monitoring systems, In IEEE International Conference on Electric Power, EPEC 2013: Conference Proceedings. New York: IEEE (Institute of Electrical and Electronics Engineers), 2013, p. 1-4, August 21-23. ISBN 978-1-4799-0105-0.
15. Langhe, R., Steps in developing research methodology in a degree project, Degree Thesis Structure in Project Management and Operational Development, Royal Institute of

Technology, KTH, 2010, Södertälje, Sweden: KTH (Royal Institute of Technology), 2010.
16. Al-Hoqani, N., and Yang, S., Adaptive sampling for wireless household water consumption monitoring, Procedia Engineering Journal, USA: Procedia Engineering, 2015, v. 119, p. 1356-1365. ISSN 1877-7058.
17. Padhy, P., Dash, R., Martinez, K., and Jennings, N., A utility-based adaptive sensing and multihop communication protocol for wireless sensor networks, ACM Transactions on Sensor Networks, TOSN, New York: TOSN (ACM Transactions on Sensor Networks), 2010, vol. 6, no. 3, p. 1-39. ISSN 1550-4859.
18. Solving the wireless mesh multi-hop dilemma, White Paper from Access One Network, Available:http://www.strixsystems.com/products/datasheets/strixwhitepaper_multihop.pdf, Strix Systems Inc, 2005, [Online], [Accessed: 10-January-2018].
19. Chatterjea, S., and Havinga, P., An adaptive and autonomous sensor sampling frequency control scheme for energy-efficient data acquisition in wireless sensor networks, In Proceedings of 4th IEEE International Conference on Distributed Computing in Sensor Systems, DCOSS 2008, New York: IEEE (Institute of Electrical and Electronics Engineers), 2008, p. 60-78, June 11-14. ISSN 9783540691693.
20. Law, Y., et al, Energy-efficient data acquisition by adaptive sampling for wireless sensor networks, In Proceedings of International Conference on Wireless Communications and Mobile Computing, IWCMC 2009, New York: IWCMC (International Conference on Wireless Communications and Mobile Computing), 2009, p. 1146-1151, June 21-24. ISSN 1937-9412.
21. O'Driscoll, S., Shenoy, K., and Meng, T., Adaptive resolution ADC array for an implantable neural sensor, IEEE Transactions on Biomedical Circuits and Systems, New York: IEEE (Institute of Electrical and Electronics Engineers), 2011, 5(2), p. 120-130, April. ISSN: 1932-4545.
22. Schroeder, D., Adaptive low-power analog-to-digital converters for wireless sensor networks, In Proceedings of International Workshop on Intelligent Solutions in Embedded Systems, WISES 2005: Conference Proceedings, New York: IEEE (Institute of Electrical and Electronics Engineers), 2005, p. 70 -78, ISBN 3902463031.
23. Chen, S., A power-efficient adaptive fuzzy resolution control system for wireless body sensor networks, IEEE Access Journal, New York: (Institute of Electrical and Electronics Engineers), 2015, v. 3, p. 743-751. ISBN 2169-3536.
24. Artan, N., Xu, X., Shi, W., and Chao, J., Optimizing analog-to-digital converters for sampling extracellular potentials, In Proceedings of 34th IEEE International Conference on Engineering in Medicine and Biology, EMBC 2012, New York: IEEE (Institute of Electrical and Electronics Engineers), 2012, p. 1663-1666. ISBN 9781424441198.
25. Hanfoug, S., Bouguechal, N., and Barra, S., Behavioral non-ideal model of 8-bit current-mode successive approximation registers ADC by using Simulink, International Journal of u-and e-Service, Science and Technology, IJUNESST 2014, Australia: IJUNESST (International Journal of u-and e-Services, Science and Technology), 2014, 7(3), p. 85-102. ISSN 2005-4246.

26. Halgamuge, M., Zukerman, M., Ramamohanarao, K., and Vu, H., An estimation of sensor energy consumption, In Progress in Electromagnetics Research B, PIERB 17, Cambridge, Massachusetts - USA: PIERB (Progress in Electromagnetic Research B), 2009, 12(12), p. 259–295. ISSN 1937-6472.
27. Razzaque, M., and Dobson, S., Energy-efficient sensing in wireless sensor networks using compressed sensing, Journal of Sensors, JoS, New York: JoS (Journal of Sensors), 2014, 14(2), p. 2822-2859. ISSN 1687-7268.
28. Ginsburg, B., Energy-efficient analog-to-digital conversion for ultra-wideband radio, PhD Thesis Dissertation Submitted to Department of Electrical and Computer Science, Massachusetts Institute of Technology (MIT), Massachusetts, USA: 2007, July 25.
29. Fredenburg, J., and Flynn, M., ADC trends and impacts on SAR ADC architecture and analysis, Invited paper in IEEE Custom Integrated Circuits Conference, CICC 2015, New York: IEEE (Institute of Electrical and Electronics Engineers), 2015, p. 1-8, September 28-30. ISBN 9781479986828.
30. Commandeur, J., and Koopman, S., An Introduction to State Space Time Series Analysis, Oxford: Oxford University Press, 2007. ISBN 9780199228874.
31. Adhikari, R., and Agrawal, R., An Introductory Study on Time Series Modeling and Forecasting, Saarbrucken, Germany: Lambert Academic Publishing, 2013. ISBN 9783659335082.
32. Spiegel, M., and Stephens, L., Theory and Problems of Statistics, Third Edition, Schaum‘s Outline Series McGraw-Hill Companies Inc, 1999, New York, USA. ISBN 0-07-060281-6.
33. Brockwell, P., and Davis, R., Introduction to Time Series and Forecasting, 2nd edition, New York: Springer-Verlag Publishing, 2002. ISBN 387953515.
34. Bradley, E., and Kantz, H., Nonlinear time series analysis, An Interdisciplinary Journal of Nonlinear Science (AIP Chaos), 2015, (25), p. 097610. ISSN: 1054-1500.
35. Fan, J., and Yao, Q., Nonlinear Time Series, New York: Springer Science Publications, 2003. ISBN 9780387261423.
36. Zivot, E., Non-linear Time Series, in Modeling Financial Time Series with S-PLUS, Springer Science and Business Media Inc, New York, pp. 653-712, 2006, ISBN 978-0-387-27965-7.
37. Deng, K., Moore, A., and Nechyba, M., Learning to recognize time series: combining ARMA models with memory-based learning, In Proceedings of 1997 International Symposium on Computational Intelligence in Robotics and Automation, CIRA 1997 Conference Proceedings, New York: IEEE (Institute of Electrical and Electronics Engineers), 1997, p. 246-251, June 10-11, ISBN 0818681381.
38. Hardle, W., Lutkepohl, H., and Chen, R., A review of nonparametric time series analysis, International Statistical Review, 1997, 65(1), p. 49-72. ISSN 1751-5823.
39. Harvey, A., and Oryshchenko, V., Kernel density estimation for time series, International Journal of Forecasting, March 2012, 1(28). ISSN: 0169-2070.
40. Zucchini, W., Applied Smoothing Techniques, Part 1: Kernel Density Estimation, October 2003, [Online]. Available: www.stateok.wismo.unigoettingen.de/ [Accessed: 29-March-2017].

41. Freedman, D., and Diaconis, P., On the histogram as a density estimator, Springer Verlag Publishing, 1981, 57(4), pp. 453 – 476, ISSN: 0178-8051.
42. Gelfan, A., Dynamic-stochastic models of river run-off generation, Hydrological Systems Modeling - Encyclopedia of Life Sciences, Oxford: EOLSS Publishing, 2009, 2, p. 49-69. ISBN 978-1-84826-648-3.
43. Helsel, D., and Hirsch, R., Statistical methods in water resources, in Techniques of Water-Resources Investigations of the United States Geological Survey, Book 4 – Hydrologic Analysis and Interpretation, United States Department of the Interior, Washington:, September 2002.
44. Ratkowsky, D., Nonlinear Regression Modeling: A Unified Practical Approach, New York: Marcel Dekker Publishing, 1983.
45. Haslett, J., On the sample variogram and the sample autocovariance for non-stationary time series, Journal of the Royal Statistical Society (The Statistician), Series D, 1997, 46(4), p. 475-485. ISSN 1467-9884.
46. Trauth, M., Matlab Recipes for Earth Sciences, New York: Springer Publishing Company, 2006. ISBN: 978-3540-27983-9.
47. Burgess, T., and Webster, R., Optimal interpolation and isarithmic mapping of soil properties - I. The Semivariogram and punctual Kriging. American Journal of Soil Science, 1980, 31, p. 315-331. ISSN 1435-0661.
48. Webster, R., and Oliver, M., Geostatistics for Environmental Scientists, Second Edition, West Sussex, UK: John Wiley and Sons, 2007. ISBN: 9780470028582.
49. Durbin, J., and Koopman, S., Time Series Analysis by State Space Methods, 2nd edition, Oxford: Oxford University Press, 2012, ISBN 978-0-19-964117-8.
50. Lucas, J. M. and Saccucci, M. S. (1990). "Exponentially weighted moving average control schemes: Properties and enhancements", *Technometrics 32*, 1-29.
51. Chen, Z., et al, Discrete-and continuous-time probabilistic models and algorithms for inferring neuronal up and down states, International Journal of Neural Computation, 2009, 21(7), p. 1797-1862. ISSN 0899-7667.
52. Chen, P., Pedersen, T., Bak-Jensen, B., and Chen, Z., ARIMA-based time series model of stochastic wind power generation, IEEE Transactions on Power Systems, 2010, 25(2), p. 667 –676. ISSN 0885-8950.
53. Box, G., Jenkins, G., and Reinsel, G., Time Series Analysis: Forecasting and Control, 4th edition, New Jersey: John Wiley and Sons, 2011. ISBN 9780470272848.
54. Time series and forecasting methods, NCSS 12 Data Analysis, 2018, [Online], Available: https://www.ncss.com/software/ncss/time-series-and-forecasting-in-ncss/ [Accessed: 10-February-2018].
55. Hipel, K., and McLeod, A., Time Series Modeling of Water Resources and Environmental Systems, Amsterdam, the Netherlands: Elsevier Science B.V., 1994. ISBN 9780444892706.
56. Reider, Volatility Forecasting I: GARCH Models, October 19, 2009, [Online], Available: http://cims.nyu.edu/~almgren/timeseries/Vol_Forecast1.pdf [Accessed: 29-March-2017].
57. Pham, L., Time Series Analysis with ARIMA – ARCH/GARCH model in R, L-Stern Group, 2013.

58. Blitzstein, J., Hwang, J., Introduction to Probability, Boca Raton, Florida: Chapman and Hall/CRC Press, 2015. ISBN: 9781466575578
59. Graves, A., Supervised sequence labeling with recurrent neural networks, Studies for Computational Intelligence, Berlin Heidelberg, Germany: Springer-Verlag GmbH, 2012. ISBN: 1860-949X
60. Choong, J., Build Neural Network With MS Excel, Kuala Lumpur, Malaysia: XLPert Enterprise Publishing Co., 2009.
61. Tabachnick, B., and Fidell, L., Using Multivariate Statistics, Essex, UK: Pearson Higher Education Publishing, 2014, ISBN: 978-0205849574.
62. Pituch, K., and Stevens, J., Applied Multivariate Statistics for the Social Sciences: Analyses with SAS and IBM's SPSS, 6th edition, New York, USA: Taylor and Francis Publishing Group, 2016, ISBN: 978-0415836661.
63. Wilks, D., Statistical Methods in the Atmospheric Sciences, 3rd edition, Oxford, UK: Elsevier Inc., 2011. ISBN: 9780123850225.
64. Warner, R., Applied Statistics: From Bivariate through Multivariate Techniques, 2nd edition, Thousand Oaks, California: Sage Publications, 2013. ISBN: 9781412991346.
65. Kish, L., Survey Sampling, New York: John Wiley and Sons Inc., 1965. ISBN: 978-0471109495.
66. Funk, C., Overview of statistical analysis of spatial data, Geog. 210C, Modeling Semivariograms, Lecture 13, Department of Geography, University of California, Santa Barbara, Spring, 2011.
67. Åstrom, K., On the choice of sampling rates in parametric identification of time series, Journal of Information Sciences, 3(1), July 1969, p. 273-278.
68. Lothian, J. (ed.), The ARIMA Procedure, SAS/ETS Documentation, SAS Institute Inc., Cary, North Carolina, USA, 2014.
69. Wilson E. B., and Hilferty, M. M., A device that allows allows the use of a normal approximation for chi-square probabilities, Proceedings of National Academia of Sciences, vol. 17, in , S. Kotz et al, The Laplace Distribution and Generalization: A revisit with applications to communications, economics, engineering, and finance, Birkhäuser Publishing Boston, 2001, ISBN: 978-1-4612-0173-1.
70. Norton, R., M., The double exponential distribution: Using calculus to find a maximum likelihood estimator, The American Statistician, American Statistical Association, 1984, 38 (2), p. 135–136. ISSN: 0003-1305.
71. Ezeora, O., Heckenbergerova, J., and Musilek, P., A new adaptive sampling method for energy-efficient measurement of environmental parameters, In 16th IEEE International Conference on Environment and Electrical Engineering, EEEIC 2016: Conference Proceedings, New York: IEEE (Institute of Electrical and Electronics Engineers), 2016, Florence, Italy, June 7-10.
72. ISO 10012-1, Quality Assurance Requirements for Measuring Equipment - Part 1, International Organization for Standardization, Geneva, Switzerland, 1998. ISBN 0 580 41654 2.

73. ISO 22514-7, Statistical Methods in Process Management – Capability and Performance, Part 7: Capability of Measurement Process, Geneva, Switzerland, 2012. ISBN 978 0 580 71185 5.
74. OIML R 111, International Organization for Legal Metrology, Paris, France, 2004.
75. European Accreditation, Evaluation of uncertainty of measurement in calibrations, EA-4/02 M: 2013, September 2013, revision 2.
76. Ezeora, O., Heckenbergerova, J., and Musilek, P., Stochastic model and optimum sampling interval of variable for environmental measurement systems, In 18th IEEE International Conference on Electric Power Engineering: Conference Proceedings, New York: IEEE (Institute of Electrical and Electronics Engineers), 2017, Kouty nad Desnou, Czech Republic, May 17-19, 2017.
77. Ezeora, O., Heckenbergerova, J., and Musilek, P., Improving energy-efficiency of analog-to-digital conversion in environmental monitoring systems, In 17th IEEE International Conference on Electric Power Engineering: Conference Proceedings, New York: IEEE (Institute of Electrical and Electronics Engineers), 2016, p. 1-4, May 16-18.
78. Dongmei, L., et al, Method for collecting signal with sampling frequency lower than Nyquist frequency, United States Patent Application Publication, Pub. No: US 2015/0326246 A1, USA: Patent Application, 2015, 12 November.

APPENDIX: List of Published Work

O. Ezeora, J. Heckenbergerova, P. Musilek, Stochastic model and optimum sampling interval of variable for environmental measurement systems, Proceedings of 18th International Scientific Conference on Electric Power Engineering, Kouty nad Desnou, Czech Republic, May 17-19, 2017.

O. Ezeora, J. Heckenbergerova and P. Musilek, A new adaptive sampling method for energy-efficient measurement of environmental parameters, Proceedings of 16th IEEE International Conference on Environment and Electrical Engineering, Florence, Italy, June 7-10, 2016.

O. Ezeora, J. Heckenbergerova, and P. Musilek, Improving energy-efficiency of analog-to-digital conversion in environmental monitoring systems, Proceedings of 17th International Scientific Conference on Electric Power Engineering, Prague, Czech Republic, May 16-18, 2016.

O. Ezeora, J. Heckenbergerova, P. Musilek, and J. Rodway, Sampling control in environmental monitoring systems using recurrent neural networks, Proceedings of 29th IEEE Canadian Conference on Electrical and Computer Engineering, Vancouver, British Columbia, Canada, May 15-18, 2016.

O. Ezeora, J. Heckenbergerova, P. Musilek, and P. Kromer, Evaluation of errors caused by lowering frequency of measurements of an environmental sensor station, Proceedings of 7th IEEE International Conference on Intelligent Networking and Collaborative Systems, Taipei, Taiwan, September 2-4, 2015.

O. Ezeora, J. Heckenbergerova and P. Musilek, Statistical analysis of environmental measurements for design of energy-efficient monitoring systems, Proceedings of 15th IEEE International Conference on Environment and Electrical Engineering, Rome, Italy, June 10-13, 2015.

O. Ezeora, J. Heckenbergerova and P. Musilek, Analysis of environmental measurements using variograms, Proceedings of 16th International Scientific Conference on Electric Power Engineering, Kouty nad Desnou, Czech Republic, pages 63-66, May 20-22, 2015.