THE DEVELOPMENT OF NEW CRESCENT MOON VISIBILITY CRITERIA USING CIRCULAR REGRESSION MODEL

ZUHAILI BIN MOHD NASIR

FACULTY OF SCIENCE

UNIVERSITY OF MALAYA

KUALA LUMPUR

2021

THE DEVELOPMENT OF NEW CRESCENT MOON VISIBILITY CRITERIA USING CIRCULAR REGRESSION MODEL

ZUHAILI BIN MOHD NASIR

DISSERTATION SUBMITTED IN FULFILMENT OF

THE REQUIREMENTS FOR THE DEGREE OF MASTER OF SCIENCE

INSTITUTE OF MATHEMATICAL SCIENCES

FACULTY OF SCIENCE

UNIVERSITY OF MALAYA

KUALA LUMPUR

2021

UNIVERSITI MALAYA

ORIGINAL LITERARY WORK DECLARATION

Name of Candidate: **ZUHAILI BIN MOHD NASIR**

Registration/Matric No: **SMA190015**

Name of Degree: **MASTER OF SCIENCE (STATISTICS)**

Title of Thesis ("this Work"): **THE DEVELOPMENT OF NEW CRESCENT MOON VISIBILITY CRITERIA USING CIRCULAR REGRESSION MODEL**

Field of Study: **APPLIED STATISTICS**

I do solemnly and sincerely declare that:

(1) I am the sole author/writer of this Work;
(2) This Work is original;
(3) Any use of any work in which copyright exists was done by way of fair dealing and for permitted purposes and any excerpt or extract from, or reference to or reproduction of any copyright work has been disclosed expressly and sufficiently and the title of the Work and its authorship have been acknowledged in this Work;
(4) I do not have any actual knowledge nor do I ought reasonably to know that the making of this work constitutes an infringement of any copyright work;
(5) I hereby assign all and every rights in the copyright to this Work to the University of Malaya ("UM"), who henceforth shall be owner of the copyright in this Work and that any reproduction or use in any form or by any means whatsoever is prohibited without the written consent of UM having been first had and obtained;
(6) I am fully aware that if in the course of making this Work I have infringed any copyright whether intentionally or otherwise, I may be subject to legal action or any other actionas may be determined by UM.

Candidate's Signature Date

Subscribed and solemnly declared before,

Witness's Signature Date

Name:

Designation:

THE DEVELOPMENT OF NEW CRESCENT MOON VISIBILITY CRITERIA USING CIRCULAR REGRESSION MODEL

ABSTRACT

The lunar crescent moon visibility has a long research history since Babylonian Era. It is important in determining the local Islamic calendar and setting the dates of important Islamic events. The possible visibility of the crescent moon is dependent on a few criteria used during the sighting process. The methods to determine minimum condition or criteria for crescent moon visibility based on the linear statistical theory that have so far been developed are not that suitable for this study because the useful variables are in the circular unit. Thus, new visibility tests are proposed in this research using the circular regression model, which will divide the data into three visibility categories; visible to the unaided eye, may need optical aid and not visible. Using the residuals of the fitted circular regression model, we formulate the method for separating the categories. We apply the model on 254 observations collected at Baitul Hilal Teluk Kemang Malaysia, starting from March 2000 to June 2019. Based on the analysis, the visibility test developed based on the elongation of the moon (dependent variable) and altitude of the moon (independent variable) gives the smallest rate of misclassification. From the statistical analysis, we propose the elongation of 7.28°, the altitude of the moon of 3.39° and the arc of vision of 3.74° at sunset as the new crescent visibility criteria. The new criteria have a significant impact on improving the chance of observing the crescent moon and hence producing a more accurate Islamic calendar in Malaysia.

Keywords: Circular regression; crescent moon; lunar month; q-test; visibility criteria

PEMBANGUNAN KRITERIA KETERLIHATAN BULAN SABIT BARU MENGGUNAKAN MODEL REGRESI BULAT

ABSTRAK

Keterlihatan bulan sabit mempunyai sejarah penyelidikan yang panjang sejak Era Babylon. Ini penting dalam menentukan kalendar Islam tempatan dan menetapkan tarikh peristiwa penting Islam. Kemungkinan terlihatnya bulan sabit bergantung pada beberapa kriteria yang digunakan semasa proses mencerap. Kaedah untuk menentukan keadaan minimum atau kriteria keterlihatan bulan sabit berdasarkan teori statistik linear yang sejauh ini dibangunkan tidak sesuai untuk kajian ini kerana pembolehubah berguna terdapat dalam unit bulat. Oleh itu, ujian penglihatan baru dicadangkan dalam penyelidikan ini menggunakan model regresi bulat, yang akan membahagikan data menjadi tiga kategori penglihatan; dapat dilihat oleh mata tanpa bantuan, mungkin memerlukan bantuan optik dan tidak kelihatan. Dengan menggunakan sisa model regresi bulat yang disuaikan, kami merumuskan kaedah untuk memisahkan kategori. Kami mengaplikasiakan model tersebut ke atas 254 cerapan yang dikumpulkan di Baitul Hilal Teluk Kemang Malaysia, bermula dari Mac 2000 hingga Jun 2019. Berdasarkan analisis, ujian penglihatan yang dibangunkan berdasarkan pemanjangan bulan (pembolehubah bersandar) dan ketinggian bulan (pembolehubah bebas) memberikan kadar salah klasifikasi terkecil. Dari analisis statistik, kami mencadangkan pemanjangan 7.28°, ketinggian bulan 3.39° dan arka penglihatan 3.74° ketika matahari terbenam sebagai kriteria jarak cahaya bulan sabit yang baru. Kriteria baru ini memberi kesan yang besar dalam meningkatkan peluang melihat bulan sabit dan menghasilkan kalendar Islam yang lebih tepat di Malaysia.

Kata Kunci: Regresi bulat; bulan sabit; kalendar lunar; ujian-q; kriteria penglihatan

ACKNOWLEDGEMENTS

Firstly, I would like to express my sincere gratitude to my supervisors, Prof. Dr. Ibrahim Mohamed, Dr. Nazhatulshima Ahmad and Dr. Rossita Mohamad Yunus, for their guidance, support, invaluable help, encouragement and supervision throughout my research. Their understanding, patience and valuable advice have been the keys to the success of this study. With deep sense of gratitude, I would like to thank my family for their love, understanding, support and encouragement at all stage of this study. Without their tremendous love and support, I would not have been able to concentrate on my study and endure some difficult times through all these years. I am grateful to the University of Malaya for providing the funancial aid during this study. I also would like to gratefully acknowledge the support given by the staff members of the Institute of Mathematical Sciences, University of Malaya especially Puan Budiyah. Special thank also goes to my friends; Afif, Fatihah, Syed and others whom have always directly or indirectly motivate me all along this journey. And Alhamdulillah, Praise to Allah for His Blessings and without His Will, this study will never be completed.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS AND ABBREVIATIONS

$I_0(k)$	:	The modified Bessel function of the first kind and order zero
R	:	Resultant length
$\bar{\theta}$	:	Sample mean direction
$\bar{R}$	:	Mean resultant length
κ	:	Concentration parameter
$(\emptyset)$	:	Sample median direction
Q_1	:	First quartile direction
Q_3	:	Third quartile direction
$r_{\theta\beta}$	:	Circular correlation
T_b	:	Best time
VM (μ, κ)	:	The von Mises distribution with mean direction, μ and concentration parameter, κ.
Alt(*M*)	:	The altitude of the moon in degree unit.
Alt(*S*)	:	The altitude of the sun in degree unit.
ARCV	:	The arc of vision in degree unit.
Elon	:	The elongation in degree unit.
JS	:	Sarma and Jammalamadaka circular regression model
Viz	:	The visibility of the crescent moon.

LIST OF APPENDICES

CHAPTER 1: INTRODUCTION

1.1 Background of the Study

Each set of data has its distributional topology which classify whether it is linear or circular type. The statistical techniques that are commonly found are based on linear data. These data sets can be represented on a straight line. However, some variables use angles as the measurement of direction in the range from 0° to 360° or (0,2π]. A radian is "an angle measure related to the arc length of a circle" (Akkoc, 2008). These types of data are represented as data of directional or circular data type. Examples of several circular data angles or measurements over time are time of day, seasonality and point in the lunar cycle. The circular data type represents the data points whose periodic nature supports it on a circle and differs from the linear type. As a result, statistical techniques developed for directional variables are different from those for linear data.

A point on a circumference of a circle or sphere surface represents the data point of a circular type. Circular data analysis is useful in many scientific fields, such as:

(i) Meteorology

Many circular data analysis are arising in meteorological studies, such as a test comparing tropical and temperate phenology (Staggemeier et al., 2020), change in seasonality (Dhakal et al., 2015), environmental applications, including wave direction (Jona-Lasinio et al., 2012) while Kamisan et al. (2011) conducted a case study of Malaysian wind data for 2005.

(ii) Medicine

In medicine, there is also research on circular statistics, for example, the probabilistic model based on circular statistics for quantifying coverage depth dynamics originating from DNA replication (Suzuki & Yamada, 2020), a practical introduction to circular variables & periodic regression (Bell, 2008) and an elliptic spatial scan statistic (Kulldorff et al., 2006).

(iii) Physics

Skew-symmetric circular distributions and their structural properties (Hatami & Alamatsaz, 2019), environmental contours for circular-linear variables based on the direct sampling method (Vanem et al., 2020) and a circular statistical method for extracting rotation measures (Sarala & Jain, 2001) are some examples of recent research that use circular method in physics fields.

In general, circular data are found in a cycle or periodic phenomenon. Circular data treatment using linear techniques can lead to bias or bad results because it does not represent the real situation. For example, let us use the linear and circular techniques to show the difference in the results of the analysis. Consider the observations of four angles 15˚, 25˚, 335 ˚ and 345˚, as shown in Figure 1.1. When analyzing circular data, it is crucial to take into account their periodical nature. By treating the observations as linear observation, the arithmetic mean of the sample is 180, and the mean direction of the four directions/angles is 0˚, respectively. So, the arithmetic mean of these four observations is

distinctly inaccurate, i.e., it is not even near to the observed points, but in the opposite direction, as can be seen in Figure 1.1. The true mean direction is 0˚, which makes more sense because it is closer to the points of observations. Therefore a special statistical method is needed to analyse circular data while taking into account the circular sample space structure.

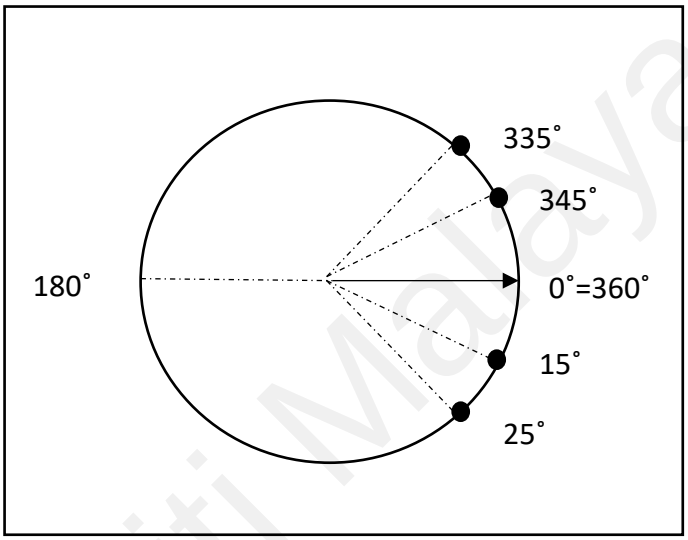

Figure 1.1: A circular map of data on a circle.

One of the most important approaches for calculating the interaction between the variables is regression analysis. For the linear case, linear regression measures the relationship between the linear variables, where both dependent and independent variables are in linear type. In circular regression, the relationship between variables are classified into three categories; (i) circular-circular regression: both the dependent and independent variables are in circular type (ii) circular-linear regression: the linear variable depends on the circular variable and (iii) linear-circular regression: the circular variable depends on the linear variable (Sarma & Jammalamadaka, 1993). In this study, we attempt to use the circular-circular regression in finding the new criteria for crescent moon

sighting activities which is an important aspect in establishing local Islamic calendar in Malaysia.

In recent years, few procedures have been proposed in determining the visibility of the crescent moon. These procedures are developed mostly by setting a minimum condition of some parameters for determining the visibility of the crescent moon during the sighting process. Nonetheless, the values used as minimum conditions in Malaysia are deemed to be too low, resulting in negative visibility decisions (or the moon is not visible from the earth). Then, a new minimum condition in determining the visibility of the crescent moon is seriously needed. This will be the interest of this study.

The classical definition of the new moon is still in use in some cultures. The new moon defines the start of the calendar months or the first day of the month. Specifically, in Malaysia, the beginning or first day of the lunar month depends on the visibility of the crescent moon, which uses two parameters, the elongation (*Elon*) and altitude of the moon (*Alt*(*M*)) to determine the beginning of a month. Many astronomers have studied lunar crescent moon visibility and many methods have been proposed in determining the visibility of the crescent moon (Danjon, 1936; Ilyas, 1994; Yallop, 1997). The importance of studying the visibility of the crescent moon includes determining special events based on the lunar calendar. Hebrew, Muslim, Hindu and some Christian, for example, still use the lunar calendar to determine their ritual day. In determining the first day of the lunar month, different methods and procedures have been used to determine the ritual day. Some of them use directly the observation (sighting of the moon) or the *rukyah* method, the calculation (mathematical calculation) or *hisab* method, and the concept of "a possibility of visibility". For example, a Muslim organization in Indonesia, Muhammadiyyah and Saudi Arabia are using a calculation or 'hisab' method to determine

the first day of Ramadan and Eid festival (Al-Mostafa, 2015). Meanwhile, Brunei Darussalam and Muslim minority in South Thailand use direct observation to determine the first day of fasting month and the Eid festival. As for Sanhedrin, Jewish uses the same method of observation to regulate their religious base on the lunar calendar (Hoffman, 2003).

In Malaysia, Indonesia and Singapore, the Islamic Religious Authorities' Al-Falakh divisions are using the concept of "possibility to visibility" in which a new lunar month begins when the crescent moon could be visible in the clear sky. To determine the new lunar month using the the possibility of visibility method, the history of the observation's records is used to set up a criterion for lunar crescent moon visibility. Then the crescent moon is considered visible if the calculation of the following conditions satisfies (Samad Abu et al., 2001):

> (i) During sunset, the altitude of the moon is not less than 2˚, and the elongation of the moon-sun is not less than 3˚, or
>
> (ii) During the moonset, and the age of the moon is not less than 8 hours.

Thus the new lunar month may begin even though the crescent moon is invisible if the criteria have been fulfilled. These minimum criterion rules are used as a guide for crescent moon visibility. This rule has been used in Malaysia, Indonesia and Singapore to determine the beginning of the lunar months since more than twenty years ago (Nawawi et al., 2015). However, the criteria considered is insignificant in most cases due to the minimum limit of the crescent's moon are below than criteria suggested by the

other studies (Fotheringham, 1910; Maunder, 1911; Bruin, 1977; Yallop, 1997; Odeh, 2004).

1.2 Problem Statement

Statistical analysis is an important consideration in the overall problem-solving process. So far in detecting the crescent moon visibility, there are a few different methods that have been proposed such as *q*-test (Yallop, 1997) and Danjon limit (Danjon, 1936). Various approaches resulting in different minimum criteria to detect the crescent moon visibility. In Malaysia, the criteria used nowadays is based on criteria that have been decided at the unofficial annual meeting of the religious ministers of the country of Brunei Darussalam, the Republic of Indonesia, Malaysia, and the Republic of Singapore (MABIMS) conference which are 2˚ of altitude of moon, 3˚ of elongation and the age of crescent moon must exceed 8 hours. However, based on data from historical records, the values have been found very low and hence do not exceed the minimum criteria for visibility. Moreover, the method used in the development of the crescent moon visibility criteria is conducted using linear statistical theories, whereas most variables consider are of circular types. This may affect the findings. Therefore, we propose in this study the development of a circular regression model using local crescent moon data that have been collected for the past 20 years. Then we suggest a new set of parameters and its minimum condition to increase the probability of visibility of the crescent moon using the circular statistical theories. The performance of the relevant methods will be compared. To summarize, there are two research problems: (i) The historical records show that the moon is visible even when the altitude of the moon value is much lower than the predetermined visibility criteria, and (ii) so far, the linear model has been used to determine the visibility criteria while data are circular in nature.

1.3 Objectives

The main objectives of this study are:

1. To describe local crescent moon data using circular statistics.
2. To propose a new visibility test for crescent moon sighting data using circular regression.
3. To determine the values for the new criteria of crescent moon visibility for local Malaysian crescent moon data.
4. To improve the new criteria of crescent moon visibility using local best time data.

1.4 Research Outline

This study aims to resolve the problem of the invisibility of the crescent moon, by proposing a new visibility test based on circular regression modelling. The outline of this research is as follows:

Chapter two gives the literature review on linear data, circular data and the previous history of the crescent moon visibility criteria. The review on the linear and circular distributions are also presented as well as the regression modelling based on the circular procedure.

Chapter three discusses the descriptive statistics of the circular data. The theory of the Sarma & Jammalamadaka (1993), JS circular regression model is also presented. Then we fit the simple circular regression model to the data and identify the significant parameters that can be used in the determination of the new criteria.

Chapter four presents the development of the new visibility test based on circular regression modelling. We will investigate the performance of several JS regression models with various model variables and will use the best model to obtain the new values of visibility criteria.

Chapter five presents the development of crescent moon visibility criteria using best time data. This chapter aims to observe the performance of the model and procedure given in Chapter 4 using best time data. The moon visibility criteria using the best time data will be compared in performance to the original data, accordingly.

Chapter six presents a review of the analysis work carried out and a summary of the research work. We also provide recommendations for future work.

CHAPTER 2: LITERATURE REVIEW

2.1 Introduction

In this chapter, we review the theories related to the crescent moon, linear data, circular data, and the visibility criteria/test of the crescent moon. First, in the next two sections, we review relevant past studies about the criteria or the minimum condition/value of relevant parameters for the crescent moon to be visible. Then we proceed to the theories on linear and circular data that are relevant for the study. We review some of the topics on the descriptive statistics for circular data. Then we will proceed to circular distributions as well as circular regression models in this chapter.

2.2 Cresent Moon/*Hilal*

Crescent moon or in the Arabic term called "*Hilal*" is defined as a physical structure of the moon that is very difficult to observe with the naked eye. The crescent moon can be observed from the earth a few hours after the event of conjunction or "*ijtimak*". Conjunction (*ijtimak*) states mean that the moon and the sun are at a meeting point or the center of the sun, moon and earth are on a straight line as shown in Figure 2.1. During the conjunction, the moon's light side faces the sun while the moon's dark side faces the world. The reflected light from the moon after several hours of conjunction is called the crescent. The new moon begins when the crescent moon is at the horizon for a short period after sunset without considering whether the crescent moon is visible or not (Pulau Pinang Mufti's Department, 2015).

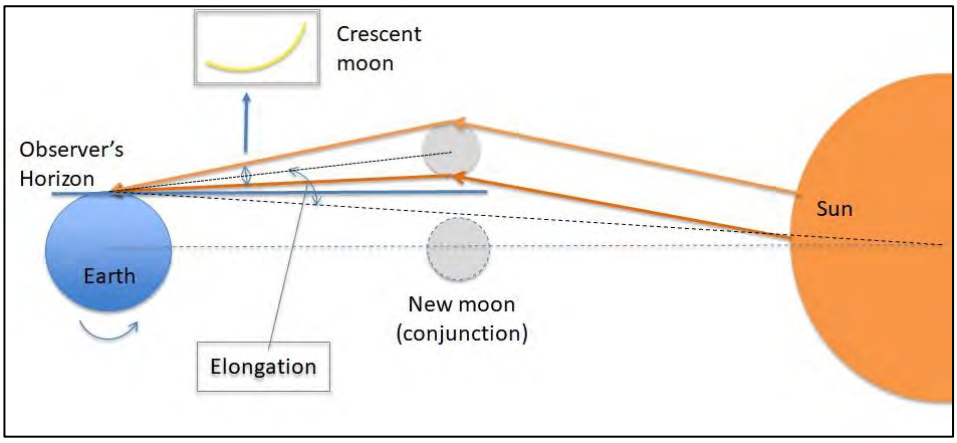

Figure 2.1: Global view of the geometric variables of the sun and moon after a few hours of conjunction.

However, many factors can cause the crescent moon to be invisible despite the best criteria of crescent moon visibility considered in the sighting process. A good weather condition and unobtrusive sky condition will allow the observer to see the crescent moon clearly. In the Islamic calendar, the visibility of the crescent moon is the basis for determining the start of a new month. Consequently, visibility will affect the dates of important events such as the start of the fasting month (Ramadan) and celebrating Eid (Syawal). Methods used by religious authorities in the Islamic countries in determining the first date of the lunar month are largely based on *rukyah* and *hisab* methods.

In the Islamic world, different implementation of either *rukyah* (sighting of the moon) or *hisab* (mathematical calculations) or both to establish the beginning of the Islamic months contribute to the difference of dates of celebrating the Islamic events in various Islamic countries. In Malaysia, *rukyah-hisab* method using the *Imkanur Rukyah* approach is practiced to ease the process of sighting and determining the starting dates of Islamic lunar months, especially Ramadan and Syawal (Nawawi et al., 2015).

Several parameters are useful in predicting the visibility of a new moon. The parameters are the elongation (*Elon*), the arc of vision (*ARCV*), the altitude of the sun

($Alt(S)$), the altitude of the moon ($Alt(M)$), the time between the conjunction of the sun, the moon and sunset in hours (*Age*), and the width of the crescent moon (W). Note that all parameters except *Age* (in hours) are circular variables and measured in degree. The definition of these parameters will be given in Chapter 3.

2.2.1 Cresent Moon Visibility

Visibility of the crescent moon is an important procedure for deciding the beginning of the month dates and important festivals in many religions. Main religions in the world have their own calendar based on lunar month including Islam. The other religions including Jews and Hindu also use crescent moon visibility to determine the important date based on their calendars. As a result, the determination of criteria to indicate the expected visibility of the crescent moon is of utmost importance since the time of the Babylon Era (Ilyas, 1994). The criteria are mainly derived based on the crescent moon data collected at end of the month. The choice of the parameters for the criteria mainly corresponds to the minimum contrast between the brightness of the moon and the sky. That is, we look at certain values such that the moon is bright enough, or the sky is dark enough for the crescent moon to be seen. For example, the Babylonians used both *age* and *lag time* as a measure of the brightness of the moon and the sky, respectively (Bruin, 1977). Here, the lag time is at least 48 minutes after sunset, and this value has changed since Babylonians era.

In the past, different studies reported the criteria based on a different set of variables measured in crescent moon sighting activities. Among the early Arabic astronomers, Al-Tabari utilized the depression angle of the sun in the visibility of the crescent moon. The

crescent would be considered visible at the time of moonset if the altitude of the sun was 9.5° below the horizon (Hogendijk, 1988; Guessoum & Meziane, 2001). In the more recent centuries, most models were built based on the observations made by Julius Schmidt in Athens, Greece, from 1859-1880 (Schaefer, 1988). Based on 76 sets of observations of the crescent moon, Fotheringham (1910) established necessary specific criteria, which include the relative altitude of the moon with respect to the sun's altitude (known as an arc of vision, *ARCV*). By placing a line of separation between the negative and positive crescent moons, Fotheringham (1910) gave a minimum limit of an *ARCV* of 12° and a relative azimuth of 0°. Maunder (1911) formulated a smaller minimum limit than Fotheringham (1910), which is 11° *ARCV* at 0° relative azimuths, for when the crescent moons can be seen, as he suggested there was a technical issue with the negative data reported by Fotheringham (1910). Ilyas (1988) examined the *ACRV* and its relative azimuth criteria and found the criteria proposed by Fotheringham (1910) and Maunder (1911) were limited to a difference in azimuth of 20 degrees. At a larger scale, these criteria cannot be applied. Consequently, to match his criteria of elongation of 10.4°, Ilyas (1988) concluded that the minimum limit of *ARCV* was supposedly 10.5° with a relative azimuth of 0°.

In the early 20th century, discussion on the criteria focused on the Danjon limit (Danjon, 1936). In the year 1931, the French astronomer André Danjon measured 75 moon samples observed using a theoretical approach. He estimated the length of the crescent moon by measuring the parts of the moon illuminated by sunlight. Crescent is expected to be visible if the elongation is more than 7°, hereafter known as the Danjon limit (Fatoohi et al., 1998). Further improvement of the criteria was later published. McNally (1983) suggested that atmospheric seeing causes the crescent to be obscured when it is smaller than the seeing disk. He concluded that the Danjon limit is supposed to

be 5° rather than 7°. However, Schaefer (1991) explained that atmospheric seeing is not the main factor in the deficiency of the arc. He developed a model and suggested 7° as the new Danjon limit for the crescent to be visible. Ilyas (1983) stated that the Danjon limit is intended to be a general guide. However, for the formation of calendar regulation, elongation of 10.5° is the best. Fatoohi et al. (1998) and Odeh (2004) studied the observational reports and respectively concluded that 7.4° and 6.4° as the estimated Danjon limits. Sultan (2007) and Hasanzadeh (2012) developed a photometric model of crescent visibility and re-evaluated the Danjon limit to be 5°.

Different mathematical approaches have been used to arrive at the values of the criteria. McNally (1983) studied mathematically the effect of atmosphere on the shape of the crescent moon and formulated the width as a measure of shortening the crescent moon in terms of θ and ϕ, where $\theta = 180 - x$, x being the elongation of the earth from the sun as viewed from the moon centre and ϕ is the position of the outer terminator near the cusp. He suggested the atmospheric factor should be considered to maximize the length of the outer terminator. The poor seeing condition will cause a shortened terminator of the crescent moon. However, Schaefer (1991) later argued that atmospheric factor is not important by considering the Hapke's lunar surface brightness measure.

Ilyas (1994) reviewed the development of criteria, especially in producing a universal international Islamic calendar amid the challenge for quality crescent moon data. A unified approach of five practical considerations is proposed to come up with a universal international Islamic calendar in the future. Yallop (1997) introduced the q-test as a test of the visibility of the crescent by considering the residuals of the fitted polynomial regression model of *ARCV* on the width of the crescent moon. Several different categories of visibility are proposed. Hoffman (2003) provided a collection of the crescent moon

data set observed in good weather conditions. He used it to update the criteria of the *q*-test and further claimed that the data could be used to validate any visibility tests. Similarly, Odeh (2004) combined data set from different studies and used them to come up with the new values of the existing criteria. Hasanzadeh (2012) used the weighted polynomial function of the arc length of crescent moon against elongation and obtained a new value of the Danjon limit by extrapolating the curves to the case of zero arc length.

The history of visibility crescent moon criteria in Malaysia specifically is based on the minimum condition of the criteria during the sighting process (Nawawi et al., 2015). In Indonesia, there are two methods that are used in determining the first date of the month based on ru'yatul Hilal (moon visibility) which are using the astronomical aid (sighting process) and using *Hisab* method (astronomical calculation) without using any criteria for the crescent moon to be visible (Djamaluddin, 2000).

Visibility criteria for crescent moon sighting that have been used in Malaysia started after the Istanbul International Congress on Crescent Moon back in 1978, which used the criteria 5˚ altitude of the moon and 8˚ of the elongation. Starting from 1983, Malaysia considered a slightly different set of values of the criteria, which is 5.5˚ altitude of the moon and 7.5˚ of the elongation. Furthermore, Malaysia also added other criteria such that the age of the crescent moon must not be less than 8 hours. At the unofficial annual meeting of the religious ministers of the country of Brunei Darussalam, the Republic of Indonesia, Malaysia, and the Republic of Singapore (MABIMS), it was decided that the criteria for crescent moon visibility are; (1) minimum 2˚ altitude of the moon and 3˚ of the elongation when the sunset, or (2) the age of the moon must exceed 8 hours when the moon sets (Nawawi et al., 2015). The inductive analysis has been applied in analyzing the data. Hence, the Djamaluddin (2000) found that while the criteria for altitude 2

degrees and 3 degrees of elongation is formulated based on the data visibility of the crescent in Indonesia, criteria of 8 hours after the conjunction (ijtimak) is derived from the views and ideas of Malaysian astronomers.

In recent years, many astronomers study the visibility of the crescent moon and many criteria have been proposed to determine the visibility of the crescent moon, for example, Utama et al. (2019) and Zainon et al. (2019). However, we expect that the visibility criteria that have been used widely in moon sighting will give us unreliable results and wrong conclusions. Therefore, the visibility criteria to detect the crescent moon needs to be reviewed and other criteria which is more effective in detecting the visibility of the crescent moon need to be proposed. As the technology evolves, the visibility aid tools, such as telescope and image processing algorithms can now be used to improve the detection of the crescent moon during sunset (Fakhar et al., 2014).

So far, the development of the criteria uses only linear statistical theory. However, most of the variables in crescent moon data are measured in degree/radian. Hence, in this study, we consider the circular statistical theory to come up with new criteria for the visibility of local crescent moon data. The proposed criteria follow closely the methodology adopted by Yallop (1997) for the q-test.

2.3 Linear Data

A linear data type refers to data that lie on a straight line. A linear data structure sequentially locates the data elements through which only one data element can be reached directly. It consists of a set of data with the elements are arranged linearly where the data elements are attached to its previous and next adjacent ones as shown in Figure 2.2. The structure represents the linear relationship and involves a single level of data elements. Most of the data available are of a linear data type.

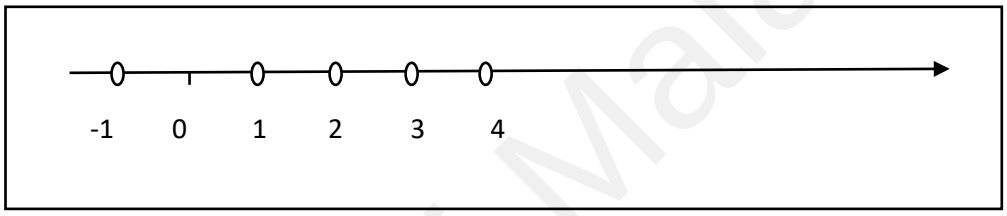

Figure 2.2: Real line number.

2.3.1 Distributions on Real Line

All sample data will form a distribution based on the type of data. Distribution of data is a mathematical feature that provides the likelihood of different potential outcomes happening in the space sample experiment. The distribution represents the density of the data referred to as the probability density function. From a given probability distribution, a calculation can be made to get the probability of any one particular observation in the sample space. There are many probability distributions for the real line type data, including the discrete or continuous distributions and mixed discrete and continuous distribution.

Standard distribution, also known as Gaussian distribution, is the most general continuous probability distribution of a random variable. The probability density function of a normal random variable with mean μ and variance σ^2 is given by equation (2.11).

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}, x \in (-\infty, \infty), \mu \in \text{R}, \sigma > 0. \quad (2.11)$$

The normal distribution is the probability distribution that is symmetrical around the mean and appears like a bell-shaped line. The normal distribution is the most important distribution in statistics and it is very useful as it is widely used in various statistical areas. Not all the data follow the normal distribution but there are many other distributions for example; Poisson distribution, binomial distribution, geometric distribution, and so on.

2.3.2 Multiple Linear Regression

The linear regression model allows us to model mathematically the relationship between two or more variables using simple algebra. Linear regression modelling is a type of predictive analysis, and it examines the significance of the predictor variables (independent variables) towards the model. Simple linear regression is an approach to model a linear relationship between two variables, which is a target variable known as a response (dependent) variable and an explanatory (independent) variable. While for more than one explanatory (independent) variables, it is called multiple linear regression. Formally, the multiple linear regression model is given by:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon, \quad (2.12)$$

where $x_1, x_2, \ldots, x_n$ are the explanatory (independent) variables or regressors and y is response (dependent) variable. In addition, $\beta_1, \beta_2, \ldots, \beta_n$ are the coefficient parameters and the y-intercept is β_0 and ε is the random error. The explanatory (independent) variables explain causal changes in the response variables.

The multiple linear regression model is developed based on the following assumptions:

(i) **Linearity** – There must be a linear relationship between the response variable and the independent variables. Linear or curvilinear relationship can be observed by a scatter plot.

(ii) **Normality** – Multiple regression assumes that with a mean of 0 and variance, σ^2, the residuals are normally distributed.

(iii) **No multicollinearity** – Multiple regression assumes that individual variables are not strongly correlated. Variance Inflation Factor (VIF) values are used to test this statement.

(iv) **Homoscedasticity** – This assumption states that the variance of error terms is similar across the values of the independent variables. A plot of standardized residuals vs. expected values will indicate whether the points are uniformly distributed over all independent variable values.

2.4 Circular data

Circular data are data that refer to variables measured on a circle and a set of observations measured by angles in a unit degree or radian. This is radically different from linear data due to its bounded properties, such that the observations are distributed in the range $(0, 2\pi]$ or $(0°, 360°]$. It is distributed on the perimeter of a unit circle which is in the form of direction on a compass. The measurements of circular data are based on the direction; on a scale from $0°$ to $360°$ North, South, East or West. The observations of $0°$ and $360°$ represent the same direction on a circle but they will be placed at various points on the linear scale. For example, a month of a year is a type of circular data in the form of periodic. January, month 1 is closer to December, month 12, than it is to June, month 6. Statistical techniques developed for linear data do not work for circular data. In such cases, to perform statistical analysis on circular data, specialized statistical techniques are necessary.

There are a few circular distributions that have been proposed for the circular data. The distribution of von Mises is widely used for mathematical analysis of circular data which can be approximated by a normal distribution under large sampling theory. The distribution has two parameters; mean direction, μ, and concentration parameter, κ. Circular data, also known as directional data, occur in various fields including astronomy see example Fakhar et al. (2014), Martins et al. (2015), de Freitas et al. (2018) and Vanem et al. (2020). The next subsection presents some descriptive statistics that can be used to describe circular data.

2.4.1 Descriptive Measure for Circular Data

Descriptive statistics is a summary of statistics that describes quantitatively the basic features of the data. Some measures of location and dispersion parameters are useful properties to describe circular data. Let $\theta_1, \theta_2, \ldots, \theta_n$ be a random circular sample of size n from a circular population. Table 2.1 shows some measures of central tendency and dispersion for two types of data.

Table 2.1: Descriptive statistics for linear and circular variables.

Statistic	**Linear**	**Circular**
Central tendency	Mean, $\overline{\mathrm{X}}$ $$\overline{\mathrm{X}} = \frac{1}{N}\sum_{i=1}^{n} x_i,$$ $\overline{\mathrm{X}}$ = arithmethic mean N = number of values x_i = data set values	Mean Direction
	Median $$\text{Median} = \frac{(N+1)^{th}}{2}\ term$$	Median Direction
Dispersion	Variance, σ^2 $$\sigma^2 = \sum \frac{(x_i - \bar{x})^2}{N}$$	Circular variance
		Concentration parameter
Correlation	Correlation, r $$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$	Circular correlation

(i) Mean direction, μ

Circular data is periodic in nature and most circular data comes in the form of direction (Cremers & Klugkist, 2018). The mean for circular statistics is generally referred to as mean direction. To summarize the data, we use the mean direction as an indicator of a trend that represents the normal direction of a variable in the population. For a given circular random sample, we consider each observation to be a unit vector whose direction is specified by the circular angle and find their resultant vector. The mean direction is defined by the angle made by the resultant vector with a horizontal line. Specifically, we have the resultant length R given by:

$$R = \sqrt{C^2 + S^2}, \tag{2.13}$$

where $C = \sum_{i=1}^{n} \cos\theta_i$ and $S = \sum_{i=1}^{n} \sin\theta_i$. The mean direction, denoted by $\bar{\theta}$, may be obtained by solving the equations, $\cos\bar{\theta} = \frac{C}{R}$ and $\sin\bar{\theta} = \frac{S}{R}$, giving

$$\bar{\theta} = \begin{cases} \tan^{-1}(S/C), & \text{if} \quad S \geq 0, C > 0, \\ \frac{\pi}{2}, & \text{if} \quad S > 0, C = 0, \\ \tan^{-1}(S/C) + \pi, & \text{if} \quad C < 0, \\ \tan^{-1}(S/C) + 2\pi, & \text{if} \quad S < 0, C \geq 0, \\ \text{undefined}, & \text{if} \quad S = 0, C = 0. \end{cases} \tag{2.14}$$

One of the mean direction characteristics is that $\sum_{i=1}^{n} \sin(\theta_i - \bar{\theta}) = 0$, which is analogous to the linear case.

(ii) Concentration parameter, κ

The concentration parameter, denoted by κ, is a significant parameter of the standard measure of dispersion for circular distribution. It shows whether the data set is homogeneously distributed on the circle or shows a concentration in the reference direction. Best & Fisher (1981) gave the maximum likelihood estimates of the concentration parameter κ as follow:

$$\hat{\kappa} = \begin{cases} 2\bar{R} + \bar{R}^3 + 5/6\,\bar{R}^5, & \text{if} \quad \bar{R} < 0.53 \\ -0.4 + 1.39\bar{R} + 0.43/(1-\bar{R}), & \text{if} \quad 0.53 \le \bar{R} < 0.85, \\ \left(\bar{R}^3 - 4\bar{R}^2 + 3\bar{R}\right)^{-1}, & \text{if} \quad \bar{R} \ge 0.85 \end{cases} \tag{2.15}$$

where $\bar{R}$ is mean resultant length and is given by $\bar{R} = \frac{R}{n}$. The larger the value of concentration parameter, the more concentrated the data towards the mean direction.

(iii) Mean resultant length, $\bar{R}$

Mean resultant length $\bar{R}$ is important to measure how concentrated the data towards the center or mean direction, $\bar{\theta}$. It is defined as:

$$\bar{R} = \frac{R}{n}. \tag{2.16}$$

The resultant length $\bar{R}$ is a statistical measure that lies in the range [0, 1] that gives us information on the spread of the circular variables in the population. It is interpreted as a precision measure where 0 means that the spread is large and 1

means that all directions in the data set are the same or concentrated at a single value (Jammalamadaka & SenGupta, 2001).

(iv) Median direction, Quantile and Percentile

Median is denoting to any point which separates the data into two sections or is known as the "middle' value. In circular techniques, the median is known as the median direction. For any circular sample, Mardia and Jupp (1972) defined the median as any point ϕ, where half of the data lie in the arc $[\phi, \phi + \pi)$ and the other points are nearer to ϕ than to $\phi + \pi$. Based on Fisher (1993), the median for any circular sample is referred to the median direction as the observation which minimizes the summation of circular distances to all observations $\theta_1, \theta_2, \dots, \theta_n$, that is,

$$d(\phi) = \pi - \sum_{i=1}^{n} \left| \pi - |\theta_i - \phi| \right|. \qquad (2.17)$$

Fisher's definition is used to obtain the circular median in the Oriana statistical software package.

On the other hand, the first and third quantile directions Q_1 and Q_3 are shown in equation (2.18).

$$Q_1 = \int_{\phi - Q_1}^{\phi} f(\theta) d\theta = 0.25 \quad \text{and} \quad Q_3 = \int_{\phi}^{\phi + Q_3} f(\theta) d\theta = 0.25, \qquad (2.18)$$

Q_1 can be considered as the median of the first half of the ordered data and Q_3 as the median of the second. The percentiles can then be obtained by further splitting the ordered sample.

(v) Sample circular variance and standard deviation

The sample circular variance is given by equation (2.19).

$$V = 1 - \overline{R}\,, 0 \leq V \leq 1. \tag{2.19}$$

The smaller the value of circular variance refers to a more concentrated sample data. The measure of the variability of sample data always refers to the concentration parameter κ, which is interpreting the same information as the circular variance. Then the sample circular standard deviation is given by equation (2.20).

$$v = \sqrt{-2\log\,(1 - V)}\,, 0 < v < \infty, \tag{2.20}$$

where V is the sample circular variance. The reason for defining the circular standard deviation in this way rather than as the square root of the sample circular variance is to obtain some reasonable approximations for the proportion of von Mises distribution provided the distribution is not too dispersed (Fisher, 1993).

(vi) Circular correlation

Special measure of correlation has been developed for any two circular variables. It is given that $(\theta_1, \beta_1), \ldots, (\theta_n, \beta_n)$ is a random sample of observations measured as angles. As for measuring the correlation between two circular variables, we use the sample circular correlation given by equation (2.21).

$$r_{\theta\beta} = \frac{\sum_{i=1}^{n} \sin(\theta_i - \bar{\theta}) \sin(\beta_i - \bar{\beta})}{\sqrt{\sum_{i=1}^{n} \sin^2(\theta_i - \bar{\theta}) \sum_{i=1}^{n} \sin^2(\beta_i - \bar{\beta})}}, \tag{2.21}$$

where $\bar{\theta}$ and $\bar{\beta}$ are sample mean directions. As in the linear case, $r_{\theta\beta}$ takes values in the range [-1, 1] and the closer the value to 1 or -1, the stronger the relationship between the variables.

2.4.2 Circular Distributions

A circular distribution is a probability distribution whose total probability is concentrated on the circumference of a unit circle (Jammalamadaka & SenGupta, 2001). There are a few important circular distributions that have been developed, for example, von Mises (Normal Circular) distribution, wrapped Cauchy distribution and Cardioid distribution (see Jammalamdaka & SenGupta, 2001; Mardia & Jupp, 1972; Fisher, 1993).

The most common circular distribution used is the von Mises distribution. This distribution can be approximated by normal distribution for large concentration parameters. We present some of these distributions below:

i. The von Mises (VM) Distribution

Von Mises (1918) introduced a circular distribution known as von Mises distribution. The von Mises distribution is also known as a circular normal distribution, which is a continuous probability distribution on the circle to emphasize its importance and similarities to the normal distribution on the real line. It is the most common circular distribution considered for unimodal samples of circular data. On the other hand, the von Mises distribution is the maximum entropy distribution for circular data when the real and imaginary parts of the first circular moment are specified. The von Mises distribution has been extensively discussed with many inferential techniques that have been developed. It is denoted as VM (μ, κ), where μ is the mean direction and κ is the concentration parameter. The von Mises probability density function is given by equation (2.22).

$$f(\theta;\mu,\kappa) = \frac{1}{2\pi\, I_0\,(k)} \exp\{k\ \cos(\theta-\mu)\}, \quad 0 < \theta, \mu \leq 2\pi \text{ and } \kappa > 0, \qquad (2.22)$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero and it is given by

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \exp\{\kappa\cos\theta\} d\theta \qquad (2.23)$$

$$= \sum_{r=0}^{\infty} \left(\frac{\kappa}{2}\right)^{2r} \left(\frac{1}{r!}\right)^2.$$

ii. The Wrapped Cauchy (WC) Distribution

Wrapped Cauchy distribution is obtained by wrapping the Cauchy distribution on the real line around the circle. The Wrapped Cauchy distribution is unimodal and symmetric. The probability density function for the Cauchy distribution is given by:

$$f(x) = \left(\frac{1}{\pi}\right)\frac{\sigma}{\sigma^2+(x-\mu)^2}\ , -\infty < x < \infty, \tag{2.24}$$

where σ is scale factor and μ is the peak of the "unwrapped" distribution. Then the probability density function for the wrapped Cauchy distribution is

$$g(\theta) = \frac{1}{2\pi}(1 + 2\sum_{k=1}^{\infty}\rho^k \cos k(\theta-\mu)) \tag{2.25}$$

$$= \frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)},$$

where $0 \leq \theta < 2\pi$ and $\rho = e^{-\sigma}$. The equality of the two expressions above is verified by equating the real parts of the geometric series identify as shown in equation (2.26)

$$\sum_{k=1}^{\infty} a^k = \frac{a}{1-a}, a = \rho e^{-i(\theta-\mu)}. \tag{2.26}$$

iii. Wrapped Normal (WN) Distribution

A wrapped normal distribution is wrapped probability distribution obtained by wrapping a normal distribution denoted by (μ, σ^2) (μ is the mean and σ^2 is variance) around a unit circle. The wrapped normal distribution is denoted by $WN(\mu, \rho)$ where μ is the mean direction and ρ is the mean resultant length parameter. Its probability density function is given by

$$f(\theta; \mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \sum_{k=-\infty}^{\infty} exp\left[\frac{-(\theta-\mu-2k\pi)^2}{2\sigma^2}\right]. \quad (2.27)$$

The alternate and more useful representation of this density can be shown to be

$$f(\theta; \mu, \rho) = \frac{1}{2\pi}\left[1 + 2\sum_{k=-\infty}^{\infty} \rho^{k^2} \cos k(\theta - \mu)\right], \quad (2.28)$$

$$0 < \theta \leq 2\pi, \quad 0 < \rho < 1.$$

The distribution is unimodal and symmetric about value $\theta = \mu$. Unlike the von Mises distribution, the wrapped normal distribution possesses the additive property that is the convolution of two wrapped-normal variables is also wrapped normal. Specifically, if $\theta_1 \sim WN(\mu_1, \rho_1)$, $\theta_2 \sim WN(\mu_2, \rho_2)$ and are independent, then $\theta_1 + \theta_2 \sim WN(\mu_1 + \mu_2, \rho_1 + \rho_2)$ (Whittaker & Watson, 1944).

iv. Circular Uniform (CU) Distribution

A circular uniform distribution is a distribution on the unit circle of uniform density at all angles. This is a unique distribution on a circle which is invariant under rotation and reflection (Mardia & Jupp, 1999). If the total probability is distributed evenly on the circumference, then we have Circular Uniformity (CU) distribution with the constant density function of:

$$t(\theta) = \frac{1}{2\pi}, 0 \leq \theta < 2\pi. \quad (2.29)$$

v. Offset Normal (ON) Distribution

The offset normal (ON) distribution is derived from a bivariate normal distribution $\phi\ (x, y;\ \mu, \Sigma)$ with mean $\mu = (\boldsymbol{\mu}, \boldsymbol{v})'$ and covariance matrix $\sum$. If r denotes the correlation between the variables and σ_1^2, σ_2^2 denote the variances, the probability density function of the offset normal distribution is given by

$$f(\theta) = \frac{1}{C(\theta)}\left\{\phi(\mu, v; 0, \Sigma) + aD(\theta)\Phi[D(\theta)]\phi\left[\frac{a(\mu sin\theta - vcos\theta)}{\sqrt{C(\theta)}}\right]\right\}, \quad (2.30)$$

where

$$a = \frac{1}{\sigma_1\sigma_2\sqrt{1-r^2}},$$

$$C(\theta) = a^2(\sigma_2^2 cos^2\theta - r\sigma_1\sigma_2 sin2\theta + \sigma_1^2 sin^2\theta),$$

$$D(\theta) = \frac{a^2}{\sqrt{C(\theta)}}[\mu\sigma_2(\sigma_2 cos\theta - r\sigma_1 sin\theta) + v\sigma_1(\sigma_1 sin\theta - r\sigma_2 cos\theta)].$$

Here, $\phi(\cdot), \Phi(\cdot)$ are the probability density function and cumulative distribution function of $N(0,1)$, respectively. The particular case when $\mu = 0$ and $\kappa = 0$ leads to the density function

$$f(\theta) = \frac{\sqrt{1-b^2}}{2\pi(1-bcos2\theta)}, \tag{2.31}$$

$$b = \frac{\sigma_1^2 - \sigma_2^2}{\sigma_1^2 + \sigma_2^2}.$$

2.5 Comparison between Linear and Circular Concepts

A data structure is a type of data that needs to be categorized accordingly based on their characteristics. Linear and circular data are examples of data structures that are fundamentally different from each other based on their characteristics. Then different statistical methods are required to treat the different data structures. The most common problem when treating circular data is to estimate a preferred direction and its corresponding distribution. Then the linear statistical method is limited in precisely characterizing the central tendency of circular distributions. For example, if we have observations of angular value on the circle, then the mean of the measures should give an estimate of the true population mean parameter.

There are many problems associated with the use of traditional statistical methods for describing circular data. Classical statistical inference methods do not take into account the circularity scale, so these methods can generate unknown and unrecognized errors. Then it will come out with a result that does not represent the real data set and loss of statistical power in making the best decision. The statistical methods that minimize the interpretational risks associated with circular data when certain distributional assumptions are met (King et al., 1987).

2.6 Circular regression

Regression is a map between two sets of variables that define the relationship between the dependent variable and the independent variable by the slope of the curve that is plotted between them. It is important in statistical analysis to model and investigate the relationship between two or more variables of interest and examine the influence of one or more independent variables on dependent variables. It is also known as the predictive modelling technique which is used for forecasting, time series modelling and finding the causal effect relationship between variables. There are several types of regression modellling including linear regression, logistic regression, multivariate regression and circular regression that have been used widely for various purposes. The basis and simplest form of a regression model is the linear regression model which makes several assumptions, including the measurements of errors that need to be normally distributed. The regression analysis has been applied in numerous and occur in almost every field, including engineering, physical science, economics, management, life and biological sciences and social science.

The regression model may be extended to the circular case which involved circular variables. The study of the circular regression model has been explored for the past 50 years. Discussion on the development of circular regression models began with Gould (1969) who predicted the mean direction of a circular response variable from a vector of linear covariates. Then Johnson & Wehrly (1977) improved the Gould (1969) model by restricting the range of the independent variables to the half-open interval $(0,2\pi]$. Circular regression procedure has been used in various applications including the study of correlations among circadian biological rhythms, where a 24-hour clock is considered as a circle (Downs, 1974; Kronauer et al., 1982; Binkley, 1990).

There are two types of circular regression, namely, simple circular regression which involved only one independent variable and multiple circular regression which involved more than one independent variable. Circular regression can be classified into two categories, where the relationship between circular variables and the relationship between mix variables which is circular and linear variables. The circular regression equation for the data is divided into two types:

i. Linear- circular and circular-linear regression

Linear-circular (linear responses and circular predictors) or circular-linear (circular responses and linear predictors) regression is a mixture of the linear variables and circular variables in the model. Regression model where the independent variables consists of one circular and a set of linear variables was proposed by Lund (1999).

ii. Circular-circular regression model

Circular-circular regression or simple circular regression is a regression model that consists of one independent circular variable and one dependent circular variable. It is very useful for analyzing bivariate circular data.

For the circular response variable denoted as v and circular explanatory variable denoted as u with mean directions α' and β' respectively, Down & Mardia (2002) applied the following mapping to relate u and v such that

$$\tan\frac{1}{2}(v-\beta') = \omega\tan\frac{1}{2}(u-\alpha') \tag{2.32}$$

where ω is a slope parameter in the closed interval [-1,1]. The mapping defines a one-to-one relationship with a unique solution given by

$$v = \beta' + 2tan^{-1}\left\{\omega\tan\frac{1}{2}(u-\alpha')\right\} \tag{2.33}$$

A simple circular-circular regression model consists one independent variable proposed by Hussin et al. (2004) such that

$$v_i = \alpha' + \beta' u_i + \varepsilon_i (\mathrm{mod}\ 2\pi) \tag{2.34}$$

where ε_i is circular random error having a von Mises distribution with circular mean 0 and concentration parameter κ and α' and β' are the coefficients of the

model. The model is useful when we are interested to find a direct relationship between the two circular variables.

Earlier than that, Sarma & Jammalamadaka (1993) proposed an interesting circular model that considered the conditional expectation of the vector $e^{(iv)}$ given u such that

$$E\left(e^{iv}\,|u\right) = \rho(u)e^{i\mu(u)} \tag{2.35}$$
$$= \rho(u)\cos\mu(u) + i\rho(u)\sin\mu(u)$$
$$= g_1(u) + g_2(u),$$

where $\mu(u)$ is the conditional mean direction of v given u and $\rho(u)$ is the conditional concentration parameter for some periodic function $g_1(u)$ and $g_2(u)$. The functions $g_1(u)$ and $g_2(u)$ are expressed in the form of their trigonometric polynomial expansions. In this study, we consider the model proposed by Sarma & Jammalamadaka (1993) and we will describe the model in detail in Chapter Three.

The circular-circular regression model with more than one independent variable is the extended model or generalized model from the simple circular regression model. It is called a multiple circular regression model where more than one independent circular variable is included in the model (Ibrahim, 2013).

2.7 Summary

In this chapter, we have discussed the existence of crescent moon visibility criteria and their minimum condition to sight the moon, as well as the existing methods to detect the crescent moon. The differences between linear and circular data are also highlighted. Due to the particular nature of their properties, we need different approaches for evaluating these data sets. Next, we have reviewed the approach of circular regression method. In the next chapter, we consider the circular regression procedure with the von Mises distributed errors in developing a new statistical test to detect the crescent moon based on Yallop's (1997) procedure.

CHAPTER 3: MODELLING MALAYSIAN CRESCENT MOON DATA USING CIRCULAR REGRESSION MODEL

3.1 Introduction

In this chapter, we study the properties of Malaysian crescent moon data using some circular descriptive statistics. We review the theory of the circular regression model, given in Sarma & Jammalamadaka (1993). After that, we apply the model to the local crescent moon data. Then, we study the correlation between the predictor variables relevant to crescent moon visibility and will use the results to determine the variables required in developing the visibility test proposed in Chapter 4. In this chapter, we will also illustrate the application of the JS regression model for one variable.

3.2 Background of data

The Astronomy Research Group of the University of Malaya in cooperation with Jabatan Kemajuan Islam Malaysia (JAKIM) are involved in the crescent moon sighting activities since 2000. The sighting activities take place at the Baitul Hilal Teluk Kemang. This location is the first location in Malaysia that was used for sighting the crescent moon in the 1970s. Previously, the observations were conducted only for three important lunar months (29th of Ramadhan, Syawal, and Zulhijah of Hijr) each year. It is to determine the start of the fasting month, Eid-alFitr and Eid al-Adha, respectively. Starting from March 2000, the sighting activity has been carried out consistently on the 29th and 30th of each lunar month. It continues until now, except for a few months due to the COVID-19 pandemic in 2020. Consequently, a total of 254 monthly observations have been collected since 2000 (1420H) with 81 (31.89%) positive results that is, the crescent moon is sighted.

The negative results may due to a few reasons, such as bad weather and severe sky conditions or when the values of important parameters are lower than the specified crescent moon visibility criteria. This will be elaborated in detail in Chapter 4.

There are various types of equipment used in the sighting activities, such as theodolites, a portable telescope of 12-inch reflector and a 76mm refractor. A few measurements are taken and the variables considered in this thesis are listed in Table 3.1. The first five variables are measured in radian/degree while the last variable indicates positive (*Y*) or negative (*N*) outcomes. Figure 3.1 shows the local view of geometric variables of the sun and moon. We can see the first five important parameters, as listed in Table 3.1 are indicated. We expect that all the parameters have a relationship with each other. Hence, we will determine the strength of the relationship using a circular correlation coefficient measure.

Table 3.1: Definition of the variables.

Variables	Definition
Width	The width of the crescent moon in arcminutes.
Alt(*S*)	The altitude of the sun in degree unit.
Alt(*M*)	The altitude of the moon in degree unit.
ARCV	The arc of vision in degree unit, i.e., the geocentric difference in the altitude between the centre of the sun and the centre of the moon for a given latitude and longitude with taking into account the effects of refraction.
Elon	The elongation, in degree unit, which refers to the angle between the centre of the sun and the centre of the moon, as viewed from earth.
Viz	The visibility of the crescent moon marked either yes (*Y*) or no (*N*).

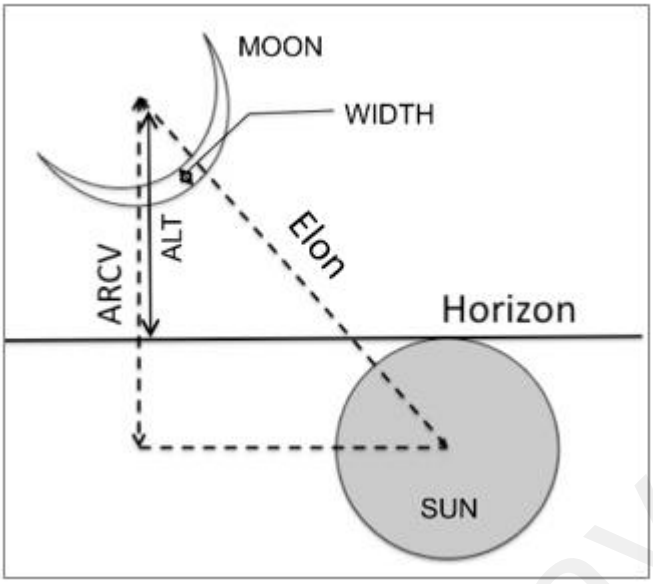

Figure 3.1: Local view of geometric variables of the sun and moon: *ARCV*, relative altitude; Alt, altitude of the center of the crescent moon above the horizon.

Since the explanatory variables are circular, we use circular descriptive statistics to provide basic features about the data. In addition, the plot of the circular variables can be used to identify the correlation between the variables. We then apply the JS circular regression procedure to the data set.

3.3 Descriptive Statistics

Table 3.2 lists the descriptive statistics on the local crescent moon variables. The distribution of the the circular variables in the crescent moon data can be described by a rose diagram, as shown in Figure 3.2. The rose diagrams show the distribution of each circular variable and the angles are measured in degrees clockwise from the east (0°).

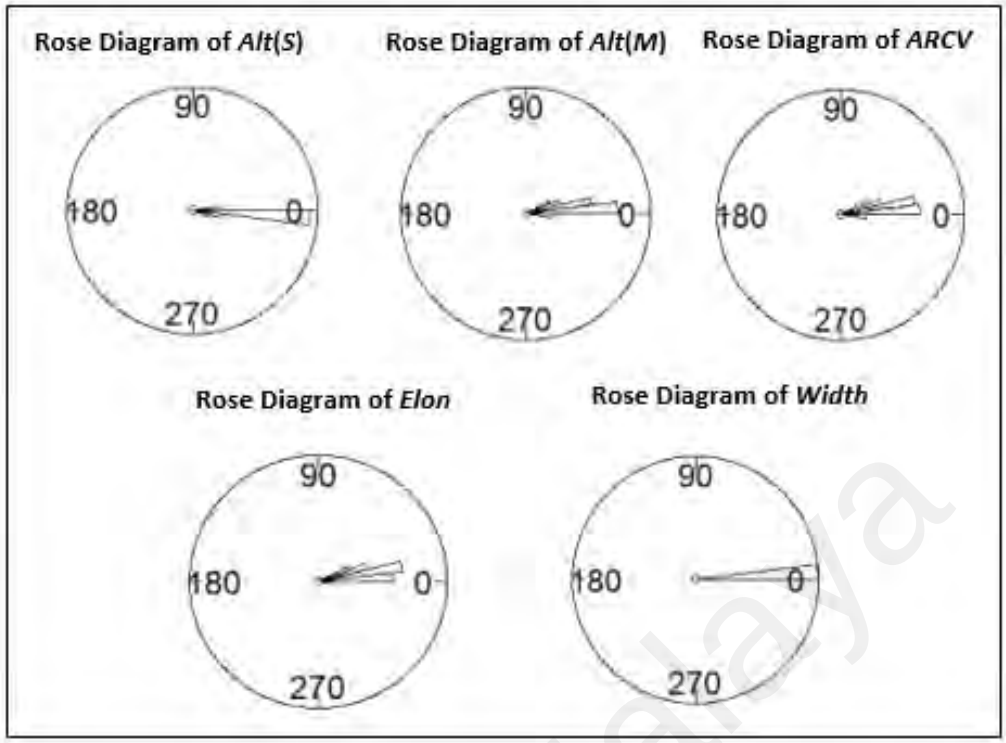

Figure 3.2: Rose diagram for the circular variables in the crescent moon dataset.

Figure 3.2 shows the rose diagram for each circular variable. We can see that most of the variables are concentrated around zero. These show the condition of an early phase of the moon after sunset and the data collected after sunset. As we expected, the *Alt*(*S*) must be negative since the position of the sun is below the horizon while the *Alt*(*M*) must be positive where the position is above the horizon. It is clearly shown from the figure that the values of *Alt*(*S*) and *Alt*(*M*) are as we expected. *ARCV* is the difference between the *Alt*(*S*) and *Alt*(*M*) and the values of *ARCV* are concentrated around zero as shown in Figure 3.2. While for *Elon*, the values are always positive since the values of *Elon* refer to the angles between the centre of the sun and the centre of the moon, as viewed from earth.

Table 3.2 provides the mean direction, concentration, minimum/maximum, and the 95% confidence interval (CI) of mean direction for the circular variables. As expected, mean direction for the *Width* of the crescent moon during sighting is generally small with a large concentration value. The observation for *Alt*(*M*) and *Alt*(*S*) must be positive and negative due to above and below the horizon respectively, though *ARCV* is more substantial than the *Alt*(*M*) because *ARCV* takes into account the position of the sun below the horizon. We can see the mean direction for all variables gives positive value except for *Alt*(*S*). However, the minimum value of *Alt*(*M*) is negative due to the data collection occur before ijtimak or location of the moon still below the horizon. All the variables give the same and large concentration values, which means that the data are very concentrated in one direction. This is expected as sigthing is carried at the same time during sunset and the values will be very close to each other.

Table 3.2: Summary statistics for circular variables in the crescent moon dataset.

Variable	Mean direction (degree)	Concentration, κ	(min, max) (degree)	95% CI of mean direction (degree)
Width	0.006	1×10^5	(0,0. 0.032)	(0.005, 0.007)
Alt(*S*)	–1.652	1×10^5	(–19.934, 4.019)	(–2.094, –1.274)
Alt(*M*)	7.498	1×10^5	(–5.460, 27.835)	(6.791, 8.204)
ARCV	9.159	1×10^5	(–5.165, 26.405)	(8.440, 9.957)
Elon	10.779	1×10^5	(0.563, 28.244)	(10.099, 11.527)

Table 3.3: Circular correlation between circular variables.

Variables	*Width*	*Alt*(*S*)	*Alt*(*M*)	*ARCV*	*Elon*
Width					
Alt(*S*)	-0.233				
Alt(*M*)	0.827	0.212			
ARCV	0.924	-0.334	0.850		
Elon	0.962	-0.287	0.841	0.966	

The correlation values between variables are also calculated, as tabulated in Table 3.3. We found that *Alt*(*M*)-*Width*, *ARCV*-*Width*, *Elon*-*Width*, *ARCV*-*Alt*(*M*), *Elon*-*Alt*(*M*), and *Elon*-*ARCV* have high correlation values than that of the other possible combinations of two variables. Nevertheless, not all the highly correlated variables are suitable to be used as the parameters for crescent moon visibility. Based on the description of variables as given in Figure 3.1, *ARCV*-*Alt*(*M*), *ARCV*-*Alt*(*S*), *Alt*(*S*)-*Alt*(*M*), and *Elon*-*Width* are collinear variables, which mean they have the same corresponding angle of geometry. To study the relationship between two variables using a simple circular regression model, we consider *ARCV* and *Alt*(*M*) only as the model variables. We refuse to consider models with the other two combinations *ARCV*-*Alt*(*S*), and *Alt*(*S*) - *Alt*(*M*) because the *ARCV*, *Alt*(*M*), *Alt*(*S*) are highly correlated with each other. As for *Width*, the variation of the values is too small and might affect the relationship with other variables (Hoffman, 2003). Hence, in this study, we focus on the combination of *Elon*-*Alt*(*M*), *Elon*-*ARCV* and *ARCV*-*Alt(M)* only. The results will be used in the new criteria developed in Chapter 4.

Figures 3.3-3.5 give the scatter plots of *Elon* against *Alt*(*M*), *Elon* against *ARCV* and *ARCV* against *Alt*(*M*), respectively. The *Alt*(*M*) for *Y* visibility is recorded at the time of the moon being sighted, which occurs a few minutes after the sunset. In this case, the altitude of the sun is several degrees below the horizon. Whereas for *N* visibility (invisible), the altitude of the moon is calculated at sunset. Therefore, the distribution of data for *Y* is expected to be more scattered than *N* cases. From both plots, we observed that the larger the values of *Elon, ARCV,* and *Alt*(*M*), the higher the possibility of sighting the crescent moons. For comparison purposes, Figure 3.6 gives the plot of *ARCV* against *Width*. Though the correlation between the two variables is high, the relationship is not linear.

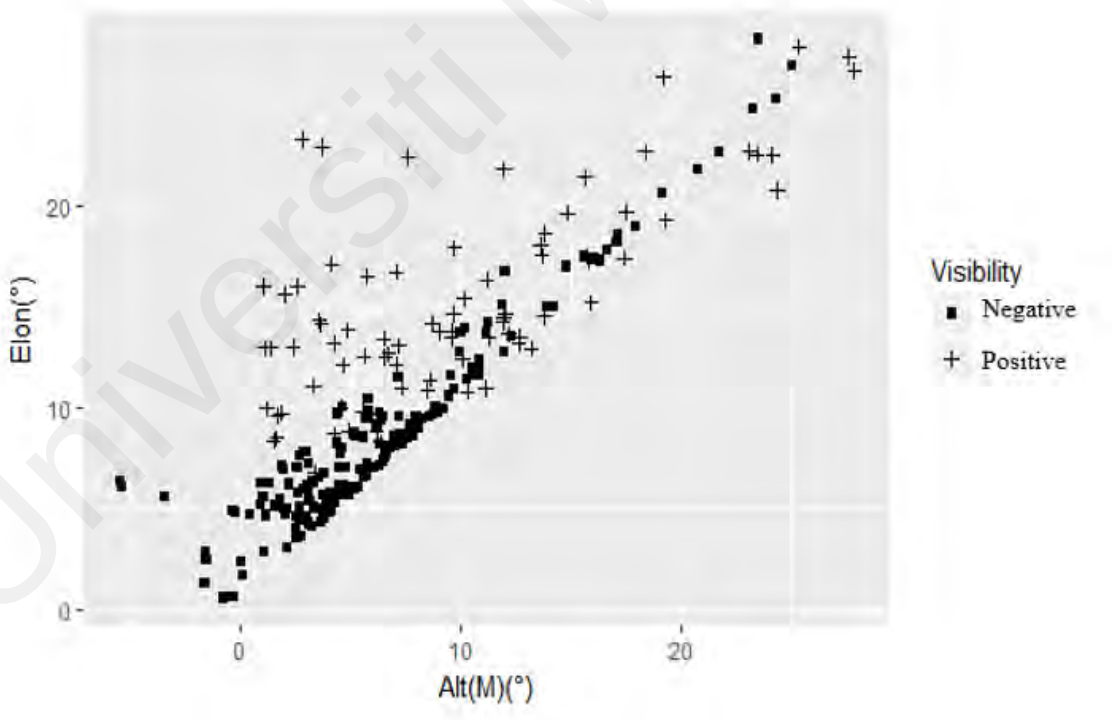

Figure 3.3: Scatter plot of *Elon* versus the *Alt*(*M*).

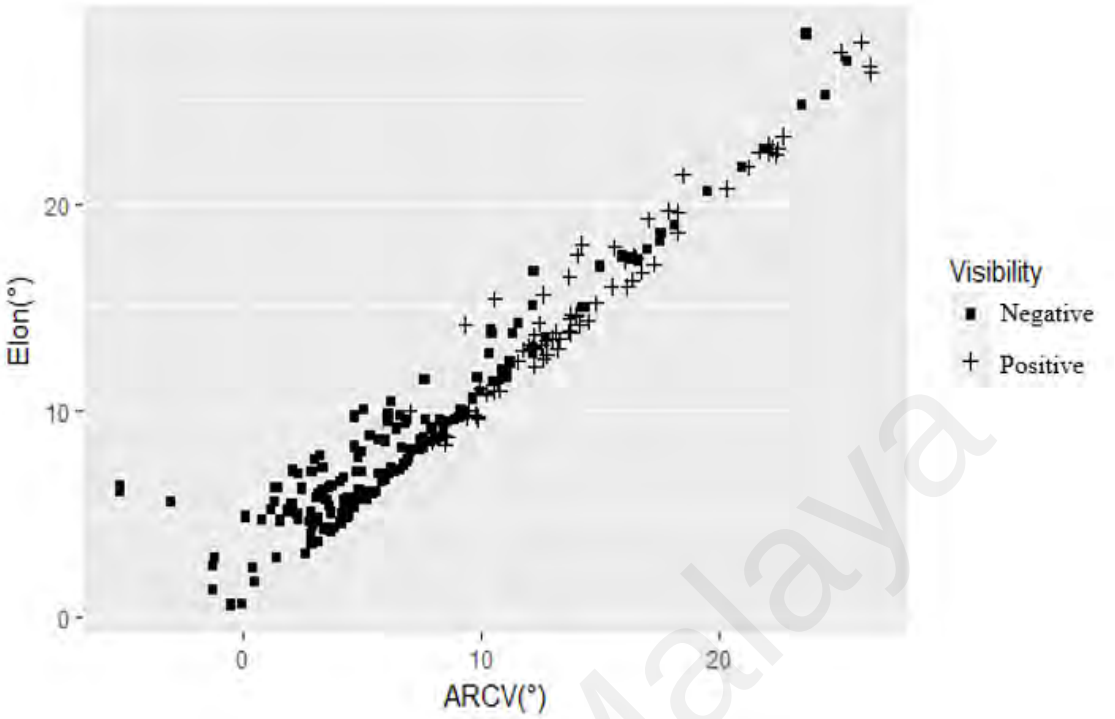

Figure 3.4: Scatter plot of *Elon* versus the *ARCV*.

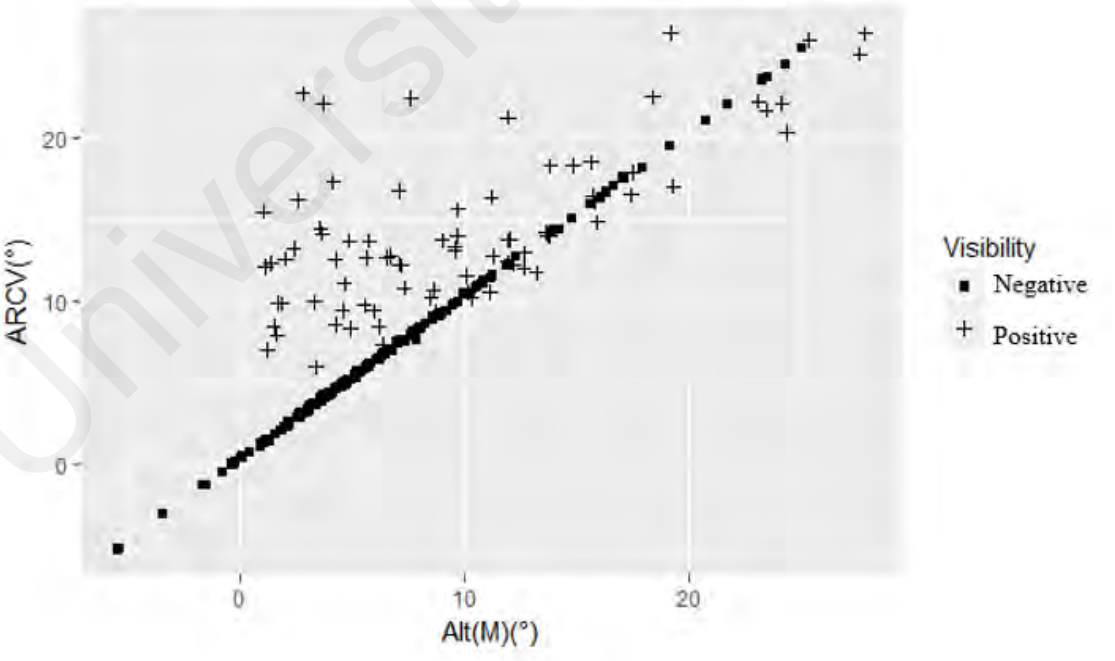

Figure 3.5: Scatter plot of the *ARCV* versus *Alt*(*M*).

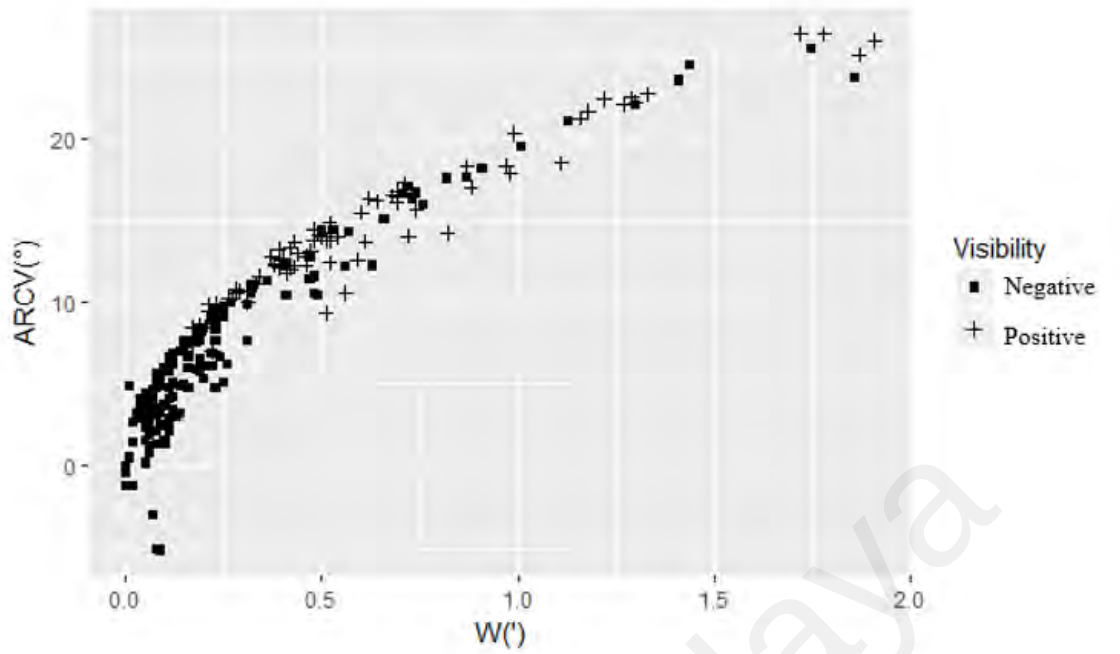

Figure 3.6: Scatter plot of *ARCV* versus *Width.*

3.4 JS Circular Regression Model

We now look the theory of JS Circular Regression model which will be use in this study.

3.4.1 Theory

The Sarma & Jammalamadaka (1993) circular regression model with two circular random variables $\boldsymbol{U}$ and $\boldsymbol{V}$ is considered. The model for two circular random variables $\boldsymbol{U}$ and $\boldsymbol{V}$ in terms of the conditional expectation of the vector $e^{(iv)}$ given u such that

$$E\left(e^{iv}\,|u\right) = \rho(u)e^{i\mu(u)}$$

$$= \rho(u)\cos\mu(u) + i\rho(u)\sin\mu(u)$$

$$= g_1(u) + g_2(u), \tag{3.11}$$

where $e^{iv} = \cos v + i \sin v$, $\mu(u)$ is the conditional mean direction of v given u and $\rho(u)$ is the conditional concentration parameter for some periodic function $g_1(u)$ and $g_2(u)$. Equivalently, we can write

$$E(\cos v \mid u) = g_1(u)$$

$$E(\sin v \mid u) = g_2(u). \tag{3.12}$$

Then we may predict *v* such that

$$\mu(u) = \hat{v} = \arctan^* \frac{g_2(u)}{g_1(u)} = \begin{cases} \tan^{-1} \dfrac{g_2(u)}{g_1(u)} & \text{if } g_1(u) > 0 \\ \pi + \tan^{-1} \dfrac{g_2(u)}{g_1(u)} & \text{if } g_1(u) \le 0 \\ \text{undefined} & \text{if } g_1(u) = g_2(u) = 0. \end{cases} \tag{3.13}$$

The difficulty of non-parametrically estimating $g_1(u)$ and $g_2(u)$ leads us to approximate them by using suitable functions, taking into account the fact that they are both periodic with period 2π. Sarma & Jammalamadaka (1993) considered the trigonometric polynomials of a function of one variable to approximate $g_1(u)$ and $g_2(u)$ (Kufner & Kadlec, 1971). The approximations used are the trigonometric polynomials of suitable degree *m* of the form

$$g_1(u) \approx \sum_{k=0}^{m} (A_k \cos ku + B_k \sin ku)$$
$$g_2(u) \approx \sum_{k=0}^{m} (C_k \cos ku + D_k \sin ku). \tag{3.14}$$

Therefore, we have the following two observational models:

$$\cos \mathrm{v} = \sum_{k=0}^{m} (\mathrm{A_k} \cos \mathrm{ku} + \mathrm{B_k} \sin \mathrm{ku}) + \varepsilon_1$$
$$\sin \mathrm{v} = \sum_{k=0}^{m} (\mathrm{C_k} \cos \mathrm{ku} + \mathrm{D_k} \sin \mathrm{ku}) + \varepsilon_2 \tag{3.15}$$

where $\varepsilon = (\varepsilon_1, \varepsilon_2)$ is the vector of random errors following the normal distribution with mean vector **0** and unknown dispersion matrix $\boldsymbol{\Sigma}$. The parameters A_k, B_k, C_k, and D_k, where $k = 0, 1, \ldots, m$, the standard errors as well as the dispersion matrix $\boldsymbol{\Sigma}$ can then be estimated using the generalized least squares estimation method.

3.4.2 Estimation of JS Circular Regression Parameters

The least square procedure is considered in the calculation of the JS circular regression parameters.

(i) Least Squares Method

Sarma & Jammalamadaka (1993) defined the estimation framework for the JS circular regression model based on the generalized least square (LS) methodology. Let

$(u_1, v_1), (u_2, v_2), \ldots, (u_n, v_n)$ be a random circular sample of size n. From (3.15), we now have the observational regression-like equations that have been given by

$$V_{1j} = \cos v_j = \sum_{k=0}^{m}(A_k \cos ku_j + \mathrm{B_k} \sin ku_j) + \varepsilon_{1j},$$
$$V_{2j} = \sin v_j = \sum_{k=0}^{m}(C_k \cos ku_j + D_k \sin ku_j) + \varepsilon_{2j}, \tag{3.16}$$

for $j = 1, \ldots, n$. Assume that $B_0 = D_0 = 0$ to ensure identifiability. The observational equations (3.16) can then be summarized as follows

$$V^{(1)} = (V_{11}, \ldots, V_{1n})',$$
$$V^{(2)} = (V_{21}, \ldots, V_{2n})',$$
$$\varepsilon^{(1)} = (\varepsilon_{11}, \ldots, \varepsilon_{1n})', \tag{3.17}$$
$$\varepsilon^{(2)} = (\varepsilon_{21}, \ldots, \varepsilon_{2n})',$$

$$\underset{n\times(2m+1)}{U} = \begin{bmatrix} 1 & \cos u_1 & \cdots & \cos mu_1 & \sin u_1 & \cdots & \sin mu_1 \\ 1 & \cos u_2 & \cdots & \cos mu_2 & \sin u_2 & \cdots & \sin mu_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & \cos u_n & \cdots & \cos mu_n & \sin u_n & & \sin mu_n \end{bmatrix}, \tag{3.18}$$

and

$$\gamma^{(1)} = (A_0, A_1, \ldots, A_m, B_1, \ldots, B_m)',$$
$$\gamma^{(2)} = (C_0, C_1, \ldots, C_m, D_1, \ldots, D_m)'. \tag{3.19}$$

Equations (3.16) can be written in a matrix form

$$V^{(1)} = U\gamma^{(1)} + \varepsilon^{(1)},$$

$$V^{(2)} = U\gamma^{(2)} + \varepsilon^{(2)}. \tag{3.20}$$

The least squares estimates turn out to be

$$\hat{\gamma}^{(1)} = (U'U)^{-1}U'V^{(1)},$$

$$\hat{\gamma}^{(2)} = (U'U)^{-1}U'V^{(2)}. \tag{3.21}$$

Then, resulting from equation (3.15), covariance matrix Σ can be estimated using the least square theory. Let

$$R_0(p,q) = V^{(p)\prime}V^{(q)} - V^{(p)\prime}U(U'U)^{-1}U'V^{(q)} \tag{3.22}$$

where $R_0 = \left(R_0(p,q)\right)_{p,q=1,2}$, then

$$\hat{\Sigma} = [n - 2(2m+1)]^{-1}R_0 \tag{3.23}$$

is unbiased estimate of Σ, and the standard error of estimators can then be found. Then we can also estimate $\mu(u)$ by using equation (3.13) and ρ using the equation (3.24) below:

$$\rho(u) = \sqrt{\frac{1}{n}\sum_{j=1}^{n}\rho^2(u_j)} = \sqrt{\frac{1}{n}\sum_{j=1}^{n}[g_1{}^2(u_j) + g_2{}^2(u_j)]}, \tag{3.24}$$

where $0 \leq \rho(u) \leq 1$.

3.5 Application on Malaysian Crescent Moon data

The JS circular regression model is now being extended to the real data set; the local crescent moon data. We choose three combination variables with two circular variables *Elon, E* and *Alt*(*M*), *A*, denoted as *EA-model, Elon,E* and *ARCV,V,* denoted as *EV-model,* and *ARCV,V,* and *Alt*(*M*), A, denoted as *VA-model.* In this study, codes are written in R-studio software to produce results of the model below.

3.5.1 *EA-model*

The *EA-model*, represents the JS circular regression model with circular model variables *Elon, E* and *Alt*(*M*), *A*. First, we must determine the order of the trigonometric polynomials in the circular-circular regression. The appropriate order m of trigonometric polynomial is determined by testing the significance of the $(m + 1)^{th}$ terms in both regression models. If the $(m + 1)^{th}$ terms do not contribute significantly to the models, the m^{th} order models are used.

Table 3.4 displays the regression coefficients of the first order models and the p values for evaluating the need for higher order models. We may claim that models of the second order are not required (p = 0.3729 and 0.1189). The best fitted JS circular regression model for *Elon* and *Alt(M)* is the model of first-order (m = 1) terms. The rho, ρ in circular regression is a measure of explained variance, indeed with interpretation analogous to r-squared. This model shows that the high value of ρ indicates that the model is good enough to fit the data. The scatter plot *Elon* versus *Alt*(*M*) with the fitted line is presented in Figure 3.7. It shows that the fitted line is quite good since the variation of data scatters along the fitted line. It makes sense that the higher-order model is not necessary.

Table 3.4: Coefficient of circular-circular regression of *Elon* and *Alt(M)*.

	Response	
Coefficient	cos(*E*)	sin(*E*)
Intercept	0.1832	1.0108
1st Order terms		
cos(*A*)	0.8104	-0.9178
sin(*A*)	-0.0372	0.6172
p-value for 2nd Order	0.3729	0.1189
ρ (r-squared)	0.9985469	

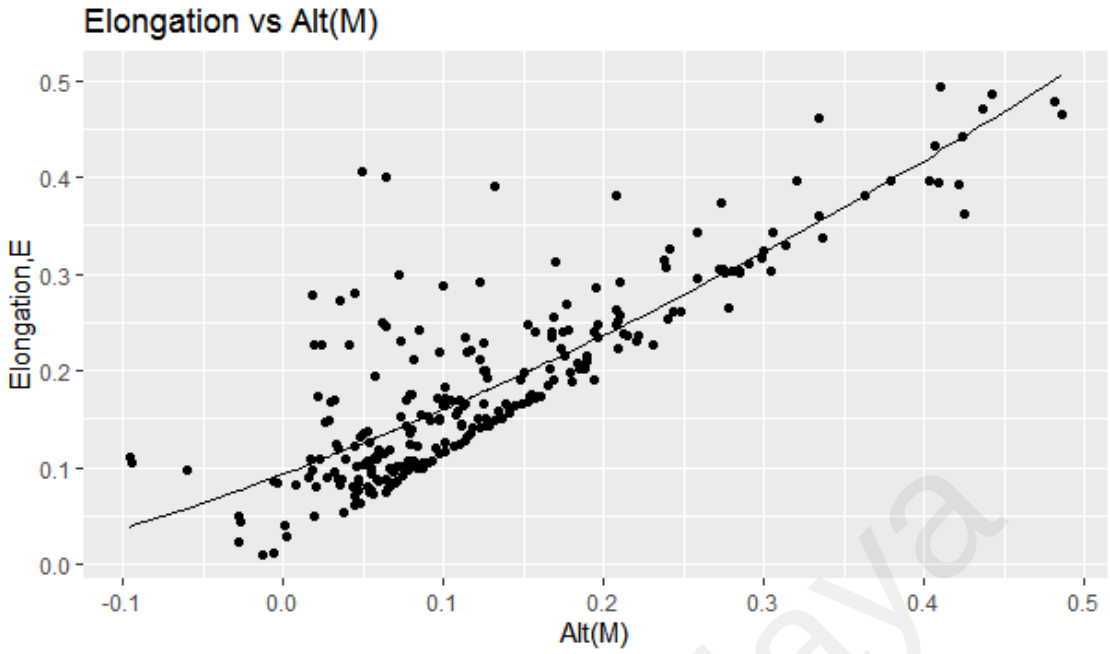

Figure 3.7: Plot of fitted *Elon* versus the *Alt(M).*

3.5.2 *EV*-model

We repeat the process using the *Elon, E* and *ARCV*, *V,* denoted as *EV*-model.

Using the same approach as for *EA-model*, the regression coefficient of the first-order models with *p* values for evaluating the need for higher-order models as shown in Table 3.5. It shows that the higher-order terms are not significant at the 0.05 level. We can say that the first order model is good enough to explain the relationship between variables (p = 0.6038 and 5.1174e-10). Although the *p*-values for $\sin(E)$ of the model are close to zero, in general, to increase the order model is not necessary.

Table 3.5: Coefficient of circular-circular regression of *Elon* and *ARCV*.

	Response	
Coefficient	cos(*E*)	sin(*E*)
Intercept	-0.0480	1.5536
1st Order terms		
cos(*V*)	1.0452	-1.4884
sin(*V*)	-0.0021	0.5877
p-value for 2nd Order	0.6038	5.1174e-10
ρ (r-squared)	0.9997308	

The best fitted JS circular regression model for *Elon* and *ARCV* is a model of first-order ($m = 1$) terms. The scatter plot *Elon* versus *ARCV* with the fitted line is presented in Figure 3.8. It shows that most of the data are close to the fitted line, and we can say that the model fitted with the first-order term is good enough to explain the relationship between these two variables. The value of ρ also gives a high value indicates that the the data are concentrated and gives us that the data fits the model very well.

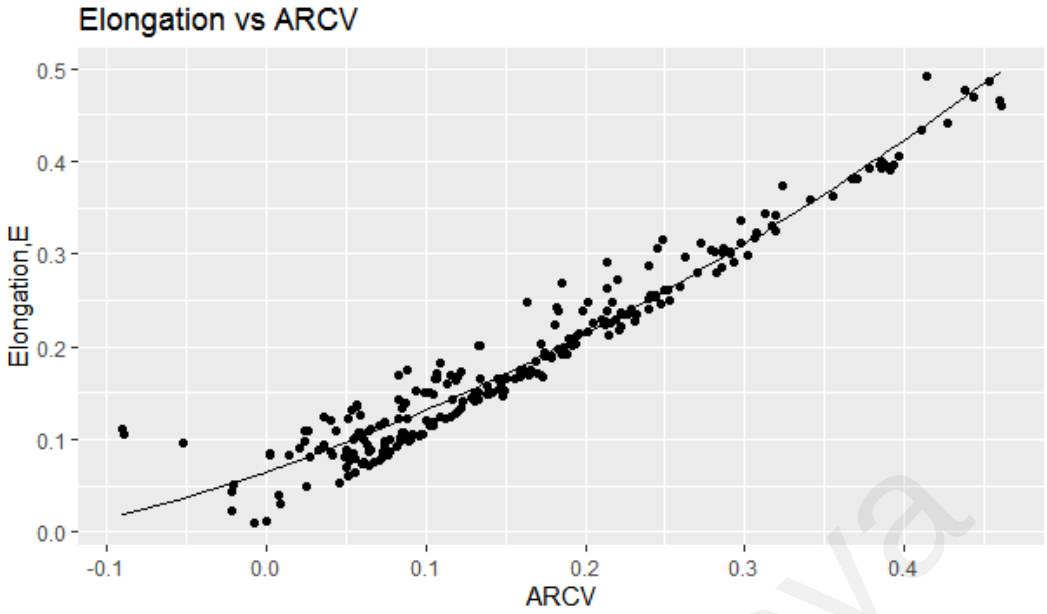

Figure 3.8: Plot of fitted *Elon* versus the *ARCV*.

3.5.3 *VA-model*

The next model *ARCV* versus *Alt(M)*. We repeat the process as *EA-model* and *EV-model* using the *ARCV*, *V,* and *Alt(M)*, *A* denoted as *VA-model*.

The best fitted JS regression model for *ARCV* and *Alt(M)* is the first-order model (m = 1). The p values for the need for higher-order models are seen in Table 3.6 where the higher model is not significant (p = 0.3117 and 0.2039). The value for rho, ρ shows a high value indicates that the model is a good model that fits the data very well. Based on the scatter plot in Figure 3.9, most of the data are close enough to the fitted line however, some observations are far from the fitted line. But in general, the model is still considered a good model.

Table 3.6: Coefficient of circular-circular regression of *ARCV* and *Alt(M)*.

	Response	
Coefficient	cos (V)	sin(V)
Intercept	0.2235	-0.3794
1st Order terms		
cos(A)	0.7724	0.4189
sin(A)	-0.0288	0.9606
p-value for 2nd Order	0.3117	0.2039
ρ (r-squared)	0.9983459	

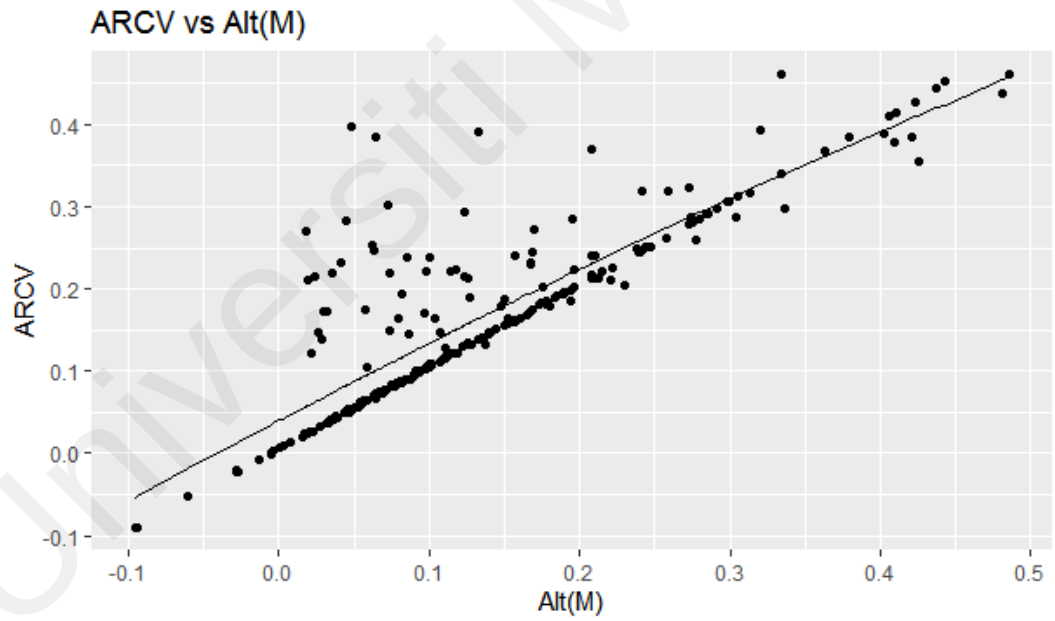

Figure 3.9: Plot of fitted *ARCV* versus the *Alt(M)*.

3.5.4 Discussion

From the analysis of the local crescent data, we can see the important parameters that give a significant contribution to the model. The result for the mean direction, concentration, minimum/maximum, 95% confidence interval (CI) and circular correlation between circular variables are tabulated in Table 3.2 and 3.3, respectively. Then the Tables 3.4 – 3.6 show the coefficients for the circular-circular regression model and Figures 3.7 – 3.9 show the plot of the circular regression model. Several results are observed:

1. For the mean direction, all the circular variables are positive except for the *Alt*(*S*). All the variables have a high value of concentration, indicating the observations are concentrated and close to each other. These are as expected as the sun is always below the horizon after sunset while the sighting of the crescent moon is concluded not long after sunset giving values of the variable to be close.

2. The value of minimum and maximum for all variables should be positive, except for the *Alt*(*S*). However, the minimum value for the *Alt*(*M*) and *ARCV* is negative. This observation corresponds to the case when the moon set occurs before the sunset, that is the conjunction has not happened yet when the crescent moon sighting is carried out.

3. All five variables are correlated with each other. The variables have a strong correlation except for the *Alt*(*S*) – *Width, Alt*(*M*) – *Alt*(*S*), *ARCV* – *Alt*(*S*) and *Elon* – *Alt*(*S*). However, some of the variables are collinear, that is, they contain similar

information about the data. For example, the information on *Alt*(*M*) is explained by *ARCV* and further study on the relationship between the two variables is not informative.

4. The three circular regression models considered show the first-order models regression coefficients are significant with the *p* values for assessing the need for higher-order models. The higher-order model is needed when the first order is not significant.

5. The circular regression modelling with two circular variables, namely, the *EA-model, EV-model,* and *VA-model* give a high value of ρ (r-squared) which is close to one. This means that the independent variable has a significant contribution to the prediction of the response variable.

6. The plot of the regression model (the *EA* model, the *EV* model or the *VA* model) shows the data points of the two variables are close to the fitted line. However, some observations lie far away from the fitted line. Further investigation is needed to treat the data.

In this study, the *Elon*, *Alt*(*M*) and *ARCV* are important variables used for developing the JS circular regression. In the next chapter, the *EA*, *EV* and *VA* models will be used to determine the new visibility criteria of the crescent moon.

3.6 Summary

In this chapter, we have studied the Malaysian crescent moon data, which includes the background of data, the descriptive circular statistics and the correlation between the circular variables in the dataset. We also have reviewed the JS circular regression model theory and have studied the JS regression parameter estimation. Then we have applied the theory of the JS circular regression method to the data of the Malaysian crescent moon. We have obtained three JS circular regression models, the *EA*, *EV* and *VA* models that will be used in the next chapter to derive the visibility tests.

CHAPTER 4: NEW VISIBILITY TEST USING LINEAR JS CIRCULAR REGRESSION MODEL

4.1 Introduction

The problem of obtaining the visibility criteria of the new crescent moon has been explored. Although there are a few new moon visibility criteria procedures proposed in the literature, none uses circular statistics to obtain the criteria. In this chapter, we follow closely the theory developed by Yallop (1997), who uses a linear model to construct the crescent moon visibility test. In this study, we develop a new crescent moon visibility test to determine the visibility criteria of the crescent moon using the JS circular regression model (Sarma & Jammalamadaka, 1993). Then, we find the percentages of correct classification of the crescent moon visibility and invisibility, for all the proposed visibility tests.

4.2 Yallop Method

This chapter reviews the theory on the derivation of the q-test by Yallop (1997) in detecting the crescent moon visibility. In 1997, Yallop introduced a crescent moon visibility test based on the topocentric crescent width, W, and geocentric $ARCV$. He analysed a sample of 295 first sightings of the crescent moons compiled by Schaefer (1996) that covered the period of 1859 to 1996 (Yallop, 1997). Yallop's algorithm computes crescent moon visibility based on residuals obtained by fitting a polynomial regression, $ARCV = f(W)$ where f is a polynomial function with argument Width, W.

The polynomial regression is given by

$$\widehat{ARCV} = 11.8371 - 6.3226W + 0.7319\,W^2 - 0.1018W^3 \qquad (4.1)$$

Yallop (1997) developed a mathematical relation between the visibility test parameter q and the fitted $\widehat{ARCV}$ based on the Indian method (Schoch, 1930).

The residuals $ARCV - \widehat{ARCV}$ are divided by 10 giving the q-statistics in equation (4.2):

$$q = \frac{[ARCV - \widehat{ARCV}]}{10}. \qquad (4.2)$$

He further defined six different categories by type A to F depending on the visibility of crescent moon using various instruments and non-visible categories, as described in Table 4.1.

Table 4.1: The q-test types by Yallop (1997).

Types	q-test value	Justification
A	$> +0.216$	easily visible to the unaided eye ($\geq$ 12ARCV)
B	$-0.014 < q < +0.216$	visible under certain atmospheric conditions
C	$-0.160 < q < -0.014$	may need optical aid to find the thin crescent moon before it can be seen with the unaided eye.
D	$-0.232 < q < -0.160$	can only be seen with binoculars or a telescope.
E	$-0.293 < q < -0.232$	below the normal limit for detection with a telescope.
F	$q < -0.293$	not visible below the Danjon limit.

Hoffman (2003) investigated the validity of the Yallop (1997) criteria using the results of 539 observations of the moon made over several years by many experienced observers in good weather conditions. The data were selected from 1047 reports. He suggested a three-category type of visibility, that is, if q is greater than 0.43, the crescent moon is visible and not visible if q is less than −0.06 and the otherwise need some optical aid to observe the crescent moon. These suggest that different data set may give different ranges of the categories.

Yallop (1997) considered circular variables $ARCV$ and W as the basis to develop the q-test for crescent moon detection as in equation (4.1) but using linear techniques. The q-statistic is developed by the following steps below:

Step 1: Consider a function f such that $ARCV = f(W)$

Step 2: Choose $f(W) = a + bW + cW^2 + dW^3$ as a cubic polynomial function.

Step 3: Fit $ARCV = \hat{a} + \widehat{bW} + \widehat{cW^2} + \widehat{dW^3}$ using the multiple linear regression model setup.

Step 4: Define the q-statistic as $q = \frac{ARCV - (\hat{a} + \widehat{bW} + \widehat{cW^2} + \widehat{dW^3})}{10}$

Step 5: Categorize the q-statistic into 6 groups according to the visibility of the new crescent moon as shown in table 4.1.

4.3 The new cresent moon visibility test

The main interest of this study is to develop an alternative crescent moon visibility test besides the q-test. In developing new crescent moon visibility tests using the circular regression model, we closely follow the procedure q-test of Yallop (1997). As in the q-test, we use the resulting residuals from the estimated JS circular regression model (Sarma & Jammalamadaka, 1993) with two circular variables to categorize the visibility of the crescent moon.

The new tests are developed by generalizing the derivation of the q-test by Yallop (1997) for the case of circular variables. The new crescent moon visibility test, say the *YX*-test, utilizes two circular variables Y and X only. Here, we use the JS circular regression with Y as the dependent variable and X as the independent variable. The resulting residuals are obtained using Equation (4.3). The *YX*-statistic then is derived as follows:

Step 1: Consider a function g such that $Y = g(\mathrm{X})$

Step 2: Fit the two circular variables using the JS circular regression model.

Step 3: Obtain the fitted values of Y, say $\hat{Y}$.

Step 4: Define the *YX*-statistic which is the residuals of the fitted model as given in equation (4.3).

Step 5: Categorise the YX -statistics into three groups according to the visibility of the moon.

As for the error of the model, we use the definition of circular distance as given Jammalamadaka & SenGupta (2001) such that

$$e = \pi - \left|\pi - \left|Y - \hat{Y}\right|\right|, \tag{4.3}$$

where $\hat{Y}$ is the estimated value of Y.

The ranges of categories of the visibility tests are obtained based on the percentage of correct classification of the visibility types. Three visibility tests are proposed based on the residuals of three fitted simple JS circular regression model, each with different response and independent variables. In this study, we use 90%, 95%, and 99% confidence intervals of the residuals, namely [L99, U99], [L95, U95], and [L90, U90] respectively, to indicate the percentage of scores that fall in the interval. For example, category A (visible to unaaided eye) may refer to test values that fall in [U99, ∞), category B (may need optical aid) refers to test values that are in [L90, U99), and category C (not visible) refers to test values that fall in (-∞, L90). For these new visibility tests, we use the combination of variables, that are, *Elon-Alt*(*M*), *Elon-ARCV and ARCV-Alt(M)*, as discussed in Chapter 3. The performance comparison of three new visibility tests is measured using the percentage of correct classification. The higher the percentage of correct classification, the better the test. The percentage of correct classification is calaculated using the following formula:

$$H = \frac{No.of\ (Y) in\ group\ A + No.of\ (N) in\ group\ C}{total\ number\ of\ observations}\ \text{X}\ 100, \tag{4.31}$$

where *H* percentage of correct classification, (*N*) is crescent moon non-visibility and (*Y*) is crescent moon visibility.

4.3.1 *EA*-test

The crescent moon visibility test, called *EA*-test, utilizes two circular variables *Elon, E* and *Alt*(*M*), *A*. The best fitted JS circular regression model with first order ($m = 1$) is given by equation (4.4).

$$\begin{aligned} \cos\hat{E} &= 0.1832 + 0.8104\cos A - 0.0372\sin A \\ \sin\hat{E} &= 1.0108 - 0.9178\cos A + 0.6172\sin A. \end{aligned} \tag{4.4}$$

Using the approach adopted by Yallop (1997), we define the *EA*-test, which takes values of residuals of the fitted JS circular regression model. We then attempt to categorize *EA* using the procedure described above and tabulated in Table 4.2. The second column gives the intervals of the categories based on the *EA*-test; for example, Category A consists of observations with *EA* greater than 0.0086. The third and fourth columns give the frequency of crescent moon non-visibility (*N*) and visibility (*Y*). For Category A, the number of *Y* is greater than *N*, while for Category C, more *N* compared to *Y*. Hence, we label Category A as "Visible to the unaided eye" while Category C as "Not visible." As for Group B, the number of *N* and *Y* are fewer with smaller *Y* compared to *N*. This might be due to many reasons including the condition of the sky, and hence labelled as "May need optical aid". The percentage of correct classification for the *EA*-test is 70.10% using Equation (4.31).

Table 4.2: Distribution of moon visibility based on three categories for *EA*-test.

Category	***EA*-test values**	***N***	***Y***	**Total (**percentage of data, **%)**	**Interpretation**
A	[0.0086, ∞)	21	52	73 (29)	Visible to the unaided eye
B	[-0.00516, 0.0086)	26	9	35 (14)	May need optical aid
C	(-∞, -0.0052)	126	20	146 (57)	Not visible

Figure 4.1 gives the plot of *EA* versus *Alt*(*M*). It shows that the residuals separate the *Y*/*N* values quite well. Observations with low residuals and small *Alt*(*M*) are largely categorized as non-visible, which is below −0.00156. For *EA* above 0.0086, the moon can be observed by unaided eyes. Otherwise, an optical aid may be needed in the sightings.

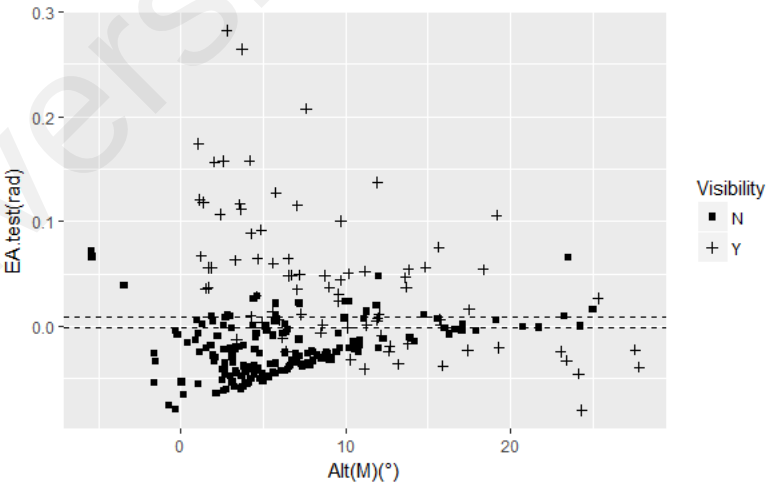

Figure 4.1: *EA*-test versus the *Alt*(*M*).

4.3.2 *EV*-test

We repeat the process using the *Elon, E* and *ARCV*, *V,* denoted as *EV*-test. The best fitted JS regression model with first order ($m = 1$) is given by equation (4.5).

$$\cos \hat{E} = -0.0480 + 1.0452 \cos V - 0.0021 \sin V$$

$$\sin \hat{E} = 1.5536 - 1.4884 \cos V + 0.5877 \sin V \tag{4.5}$$

Using the same approach as the *EA*-test, the distribution of moon visibility based on the three categories are given in Table 4.3. The *EV*-test does not give a good result, with the percentage of correct classification is only 43.7% using Equation (4.31). This low performance is supported by the plot of *EV* versus *ARCV*, as given in Figure 4.2. The distribution of the residual values of fitted model denoted as visible, *Y* and nonvisible, *N* data are more scattered than that of the *EA*-test; thus, it fails to separate the *Y* and *N* data very well.

Table 4.3: Distribution of moon visibility based on three categories for *EV*-test.

Category	***EV*-test values**	**N**	**Y**	**Total (**percentage of data, **%)**	**Interpretation**
A	[0.0039, ∞)	59	21	80(31)	Visible to the unaided eye
B	[-0.0022, 0.0039)	24	16	40(16)	May need optical aid
C	(-∞, -0.0022)	90	44	135(53)	Not visible

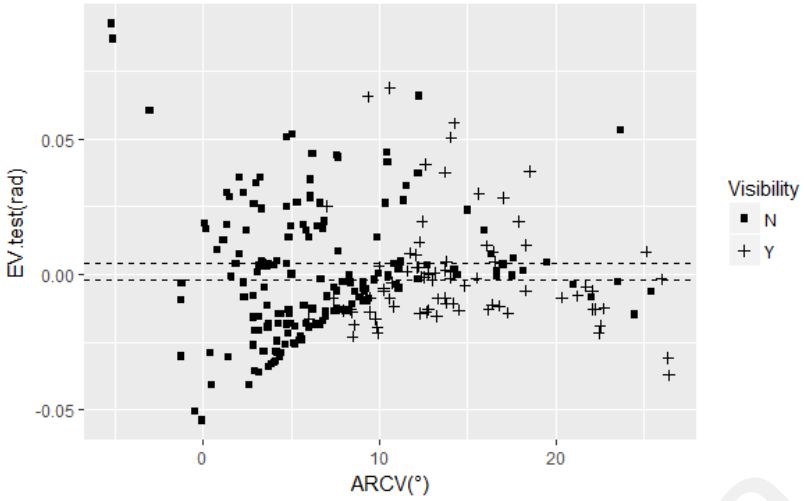

Figure 4.2: *EV*-test versus the *ARCV*.

4.3.3 *VA*-test

We repeat the process using the *ARCV*, *V,* and *Alt*(*M*) denoted as *ArcvA*-test. The best fitted JS regression model with $m = 1$ is given by equation (4.6).

$$\cos\widehat{Arcv} = 0.2235 + 0.7724\cos A - 0.0288\sin A$$

$$\sin\widehat{Arcv} = -0.3794 + 0.4189\cos A + 0.9605\sin A \tag{4.6}$$

The result of the categories is shown in Table 4.4. From the result, category A with visible, *Y* is 49 and category C with nonvisible, *N* is 161. Hence, the percentage of correct classification is 82% using Equation (4.31). This means the model is significant to detect the new crescent moon. Unfortunately, although the percentage of correct classification is high, the model with the combination *VA* is not preferred as the variables *ARCV* and *Alt*(*M*) are collinear, as explained in Chapter 3. Hence we do not proceed using this *VA*-test in the next sections.

In the next section, we will focus on category B of *EA*-test as the best indicator in this study to determine the crescent moon visibility criteria.

Table 4.4: Distribution of moon visibility based on three categories.

Category	***VA*-test values**	**N**	**Y**	**Total (**percentage of data, **%)**	**Interpretation**
A	[0.0085, ∞)	5	49	54(21)	Visible to the unaided eye
B	[-0.0060, 0.0085)	7	11	18(7)	May need optical aid
C	(-∞, -0.0060)	161	21	182(72)	Not visible

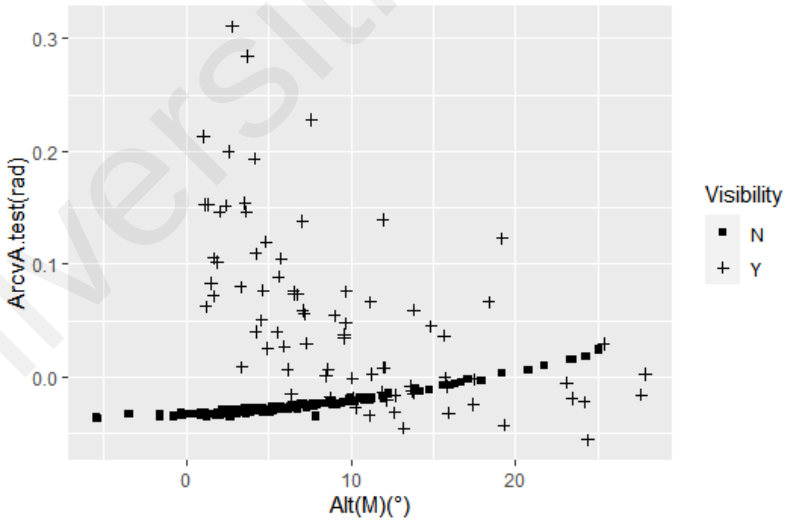

Figure 4.3: *VA*-test versus the *Alt(M)*.

4.4 Discussion

The results in Section 4.3 indicate that *EA*-test provides the best indicator of visibility of the crescent moon because of the higher percentage of correct detection. Hence, we attempt to come up with the new visibility criteria based on the *EA*-test. The lower limit of the *EA* test value is then used as the criteria of crescent moon visibility to set the date of important event in religion. In this work, the criteria will be based on category B of Table 4.2 as nowadays, telescopes or other optical aid systems are used in the observations. As *Elon* is taken as the dependent variable in the *EA*-test, we first estimate the criteria value of *Elon* by its percentile values of category B. To obtain the percentile value of the *Elon*, first we arrange the values of *Elon* for all observations in Category B in ascending order. The 5^{th}, 10^{th}, and 15^{th} percentile mean 5%, 10%, and 15% of the ordered observations will be smaller than the percentile values, respectively. That corresponds to 1, 3, and 5 ordered observations and, hence, the choice of 15% percentile seems to be adequate for this data. This choice is due to the sample size of Group B is small, which is 35. So, taking the 15% percentile of data that refers to the first five ordered observations seems reasonable. From the astronomical point of view, we can only see the moon when its altitude is greater than 2 degrees, and this corresponds to when the percentile of *Elon* in Group B is 15%. The five observations are listed in Table 4.5. Most of them have rather low values of *Alt*(*M*) and *ARCV*, which makes it rather difficult to sight the crescent moon after the sunset. Hence, the value of the criteria for *Elon* is estimated at 7.28° (i.e. the 15% percentile).

We then estimate the corresponding values of *Alt*(*M*) and *ARCV* given that *Elon* = 7.28° using the linear regression model. Hence, the values of *Alt*(*M*) and *ARCV* are 3.03° and 3.74°, respectively. Figure 4.4 shows plot of *Elon* vs *Alt*(M) for observations in

Category B. Consequently, by definition, the estimated *Alt*(*S*) is taken as the difference between *ARCV* and *Alt*(*M*), that is -0.71°. As for Width(*W)*, the observed values are consistently small and we use the 15th percentile as its estimate, which is 0.1°. We note that the sun's altitude of 0.71° below the horizon has considered the effect of refraction near the horizon and semi-diameter of the sun. During sunset, the centre of the sun is estimated at 0.35° below the horizon, and hence the estimated time taken for the sun to the altitude -0.71° is 1.4 minutes after it sets.

Table 4.5: Observations with *Elon* less than the 15th percentile value.

Date of moon sighting (Masehi)	Date of moon sighting (Hijr)	*Elon* (°)	*Alt* (*M*) (°)	*ARCV* (°)	*Alt*(*S*) (°)	*Width* (°)	Visibility (Y/N)
27.07.2014	29 Ramadan 1435	7.042	2.577	2.912	-0.335	0.11	N
10.11.2007	29 Syawal 1428	6.898	1.963	2.314	-0.351	0.11	N
16.09.2012	29 Syawal 1433	6.286	0.945	1.361	-0.416	0.1	N
25.04.2009	29 Rabiulakhir 1430	6.276	1.286	1.495	-0.209	0.1	N
27.06.2014	29 Syaaban 1435	4.888	-0.319	0.114	-0.433	0.05	N

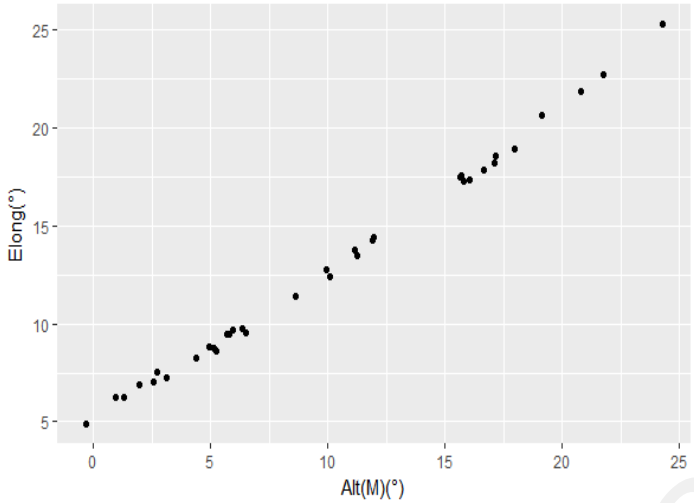

Figure 4.4: *Elon* vs *Alt*(*M*) for observations Category B of the *EA*-test.

In determining the crescent visibility criteria, we use the *Elon* and *Alt*(*M*) at sunset. We note that the duration of 1.4 minutes after sunset is considered negligible to elongation as the average rotation rate of the moon surrounding the earth takes about 0.008°/minute. Hence, the adjusted values of criteria for *Elon* and *Alt*(*M*) are 7.28° and 3.39°, respectively, and the corresponding values for *ARCV* = 3.74° and *Alt*(*S*) = –0.35°. The proposed new values of the crescent moon visibility criteria are as listed in Table 4.6.

Table 4.6: The values of new criteria of variables for Category B of the *EA*-test at sunset.

Variables	Value of criteria (°)
Elon	7.28
Alt(*M*)	3.39
ARCV	3.74
Alt(*S*)	-0.35
Width	0.10

4.5 Summary

In this chapter, we have discussed new visibility test based on Yallop (1997) procedure. We have reviewed the theory of Yallop (1997) procedure in developing the crescent moon visibility test. Then we have applied the theory of Yallop (1997) procedure to the local crescent moon data using circular regression set up, calculated the percentage correct detection of the new crescent moon and identified the significance of the model in detecting the crescent moon. The *EA*-model is the significant model in this chapter and so we use the *EA*-test is the best indicator to determine the new crescent moon visibility criteria using percentile value of observations in Category B. The new crescent moon visibility criteria as shown in Table 4.6 give *Elon* is 7.28° and *Alt*(*M*) is 3.39°. In this study, we consider the definition of circular distance from Jammalamadaka & SenGupta (2001) as the model error for the circular regression procedure used in developing new statistical test to detect the crescent moon based on Yallop (1997) procedure.

CHAPTER 5: APPLICATION BASED ON LOCAL BEST TIME CRESCENT MOON DATA

5.1 Introduction

In this chapter, we review the concept of best time data (Yallop, 1997) that was proposed to improve the criteria for crescent moon visibility. In this study, we transform the original dataset into best time data by applying the transformation procedure given in Yallop (1997) to the local crescent moon data. Then, we propose new criteria of crescent moon visibility based on the best time data obtained and using the procedures developed in Chapter 4.

5.2 The Concept of Best Time

We review the concept of "best time" developed by Bruin (1977), Schaefer (1991) and Yallop (1997) for obtaining the earliest visibility time of the new crescent moon. Crescent moon best time data refers to the observations collected at the best time to sight the moon such that the probability of sighting the moon is higher than when we sight at the time of sunset. It gives an empirical estimate of the time that gives the observers with the best chance to see the new crescent moon. The best time is directly proportional to site elevation and inversely proportional to moon altitude. Generally, it is taken at the time point that is 4/9 of the lag time after sunset.

When observing the crescent moon, there is a point on earth after sunset that the sun and the moon are in a perfect geometrical condition where the different of azimuth is 0 degrees between the sun and the moon. The crescent moon should be observed as early as possible after the conjunction occurs. Although in practice, the sighting of the crescent moon is possible after the sunset, there is the best time to observe the moon which gives the highest chance to observe the moon. When the observation is made too early after sunset then the horizon can be too vivid for the small crescent moon to pick out. The observer needs to wait until there has been enough rise in the contrast between the crescent moon and the twilight sky for the moon to be seen. As the twilight sky gets darker, the brightness of the crescent moon will decrease as the altitude of the moon decreases due to atmospheric extinction, so there is an optimal time to make an observation. In critical condition, the visibility of crescent moon is only possible during the right moment within a limited period of time.

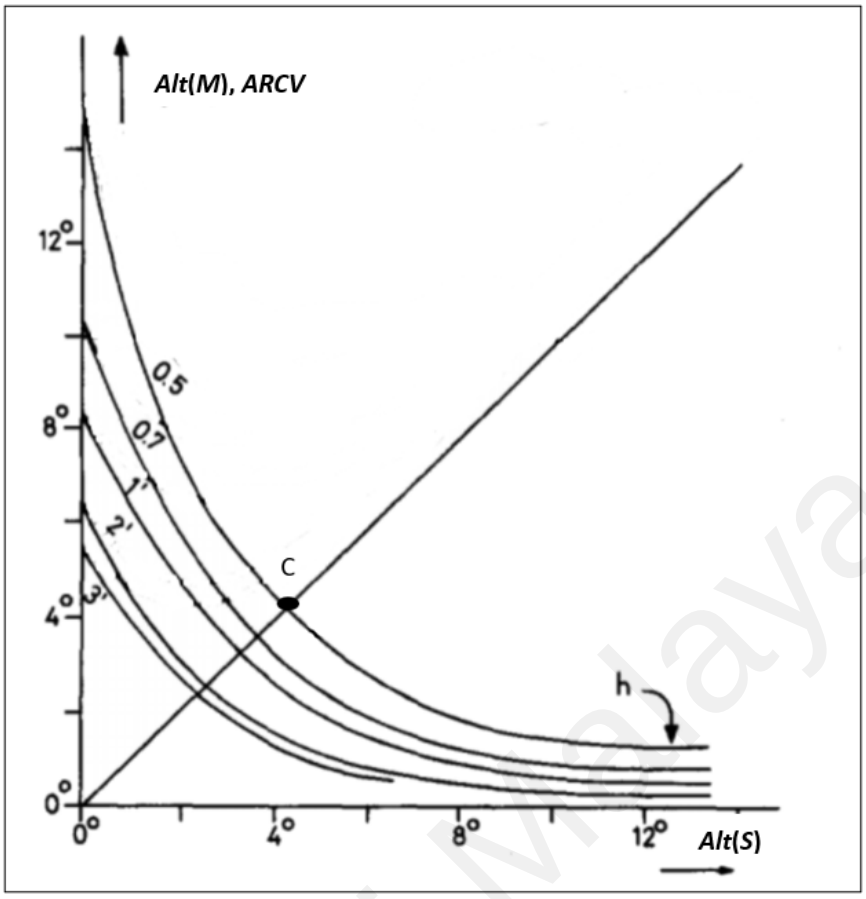

Figure 5.1: **Lunar visibility curves (Taken from Bruin, 1977).**

Schaefer (1988) determined the best time from the logarithm of the real total brightness of the moon divided by the total brightness of the moon needed for visibility for the state of observation. Bruin (1977) gave a simple method for estimating the probability of viewing the moon at some moment after sunset which is sufficient for most purposes. The derivation formula for best time based on the Bruin (1977) comes from the series of curves of visibility as shown in Figure 5.1. The plot shows series of visibility curves of $Alt(M) + Alt(S)$ ($ARCV$) against $Alt(S)$ for W = 0.5', 0.7', 1', 2', and 3'. Hence $Alt(S)$ is the depression of the sun below the horizon and $Alt(M) + Alt(S)$ = ARCV. Bruin (1977) claimed that when the condition is optimal, every of those curves has a minimum. He marks the minimum as point C on the curve for W = 0.5'. If we draw a straight line through the origin (at $Alt(M) + Alt(S) = 0$ ° and $Alt(S) = 0$ °) and through point C (at $Alt(M) + Alt(S) = 9$ ° and $Alt(S) = 4$ °), then we notice that the line goes straight through the minima

of different *W* curve sequences. Hence, at the best time, the equation of the line is 4 *Alt*(*M*) = 5 *Alt*(*S*). Therefore, the best time is given by

$$T_b = (5T_s + 4T_m)/9 = T_s + (4/9)Lag, \tag{5.1}$$

where $T_b = \text{Best time},$

$T_s = \text{Sunset time},$

$T_m = \text{Moonset time},$

$Lag = \text{Moon's lag time}.$

5.2.1 Data Transformation to Best Time

In this section, we consider data transformation from the original data to the new best time data. Generally, the objectives of data transformation are improving the data quality and giving a more reliable result. For linear data, feature scaling or data normalization is one of the methods used to normalize the data, hence to obtain a better analysis of the data. In this study, the transformation of data is made based on the theory or concept of the best time data (Yallop, 1997) as described in section 5.2. The concept of best time is used to determine the new value for each parameter. Together with this concept, we also use the theory that the earth does one revolution (360 degrees) on its axis every 24 hours (Rawlins, 2008). Its means that it will take 4 minutes on average for every one-degree movement of the earth.

The transformation of the original dataset into a new dataset is necessary because adjusting the time using the best time concept will improve the new moon visibility, as discussed in the previous section. In practice, we need to observe the moon right after the conjunction until the crescent moon is visible. However, for the best time concept, the sighting is done when the time is optimal to observe the crescent moon. The best time dataset is expected to give a higher percentage of visibility of the crescent moon.

To transform the original data to best time data, we use the formula proposed by Yallop (1977) and it is given by

$$A_b = 4/9(A_m) + 5/9(A_s) \tag{5.2}$$

where $A_b =$ New transformed value of a parameter

$A_m = Parameter\ value\ at\ moonset$

$A_s = Parameter\ value\ at\ sunset$

In this study, we transform each of the following parameters, *Alt*(*M*), *Alt*(*S*), *ARCV*, *Elon*, and *Width* using the formula given in equation (5.2). After transforming the original data to the best time data, we apply the circular regression to the transformed dataset (best time data) and construct the visbility tests based on the Yallop's procedure. This procedure is expected to improve the visibility criteria compared to using the original data.

5.3 Application on Local Best Time Crescent Moon Data

In this chapter, we apply the best time concept and tranform the original dataset with 254 observations to the local best time data. Based on the procedure developed by generalising the Yallop's procedure to the case of circular variables as described in Chapter 4, we reapply the procedures on the local best time data. In the procedure, first we fit the JS circular regression model to the transformed values of *Elon* and *Alt*(*M*) obtained using equation (5.2).

Table 5.1: Coefficient of circular-circular regression of *Elon* and *Alt*(*M*) in the local best time data.

	Response	
Coefficient	cos(*E*)	sin(*E*)
Intercept	0.2833	0.0114
1st Order terms		
cos(*A*)	0.7155	0.0614
sin(*A*)	-0.1041	0.9237
p-value for 2nd Order	0.6646	0.00149
ρ (r-squared)	0.9983828	

Table 5.1 shows the first order models regression estimated coefficients and with the *p* values for assessing the need for higher order models. We find that the second order models are not neccessary (p = 0.6646 and 0.00149).

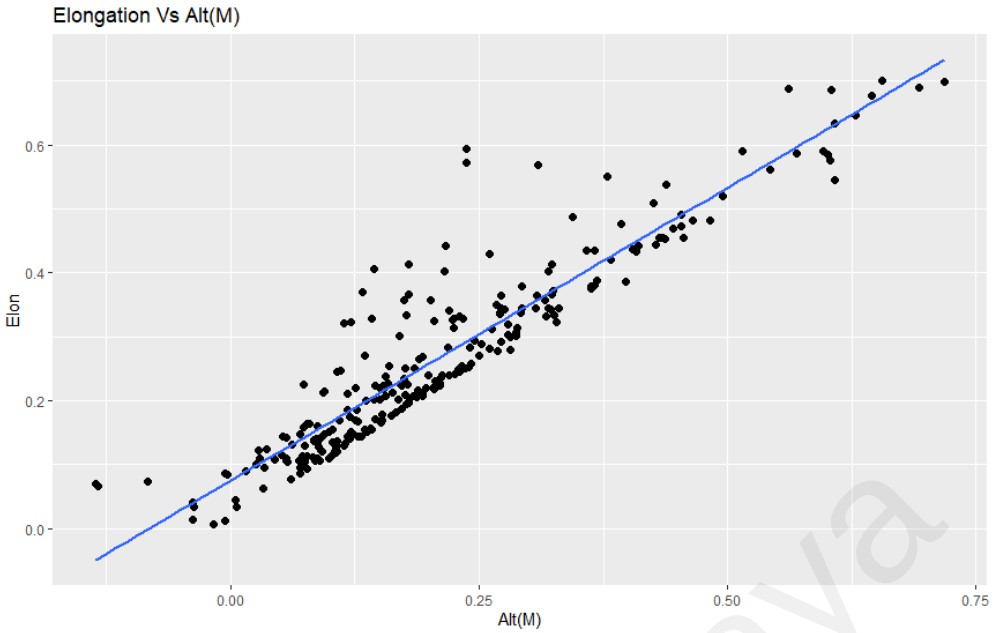

Figure 5.2: Plot of fitted *Elon* versus the *Alt(M)*.

Figure 5.2 presents the scatter plot of *Elon* against *Alt*(*M*) with their fitted JS circular regression line. It shows that the fitted line is quite good where the variation of the data is small, and the points lie close to the fitted line. The first order circular regression model is significant in this case since the *p*-value is small.

Table 5.2: The percentage for correct detection.

Model	Percentage Correct detection
Elon – *Alt*(*M*) (*EA*-test)	70.5%
Elon – *ARCV* (*EV*-test)	46.5%
ARCV – *Alt*(*M*) (*VA*-test)	82.2%

Table 5.2 shows the percentages of correct detection for the three models using the local best time dataset. From the table, we can see that the percentage of correct detection for the $ARCV - Alt(M)$ model is the highest with the percentage of correct detection is 82.2%. Although the percentage is high, collinearity occurs between *ARCV* amd $Alt(M)$. The *ARCV* is the summation of the altitude of moon ($Alt(M)$) and altitude of sun. Due to this relationship, this model is not taken into account for further investigatation. As a result, the *Elon* and $Alt(M)$ model with the percentage of correct detection 70.5% is preferred over the other models.

Table 5.3 shows the crescent moon visibility criteria obtained using the proposed method. We use the circular regression model with elongation (*Elon*) and altitude of the moon ($Alt(M)$) as the dependent and independent variables, respectively, to determine the value for crescent moon visibility criteria at sunset. As *Elon* is taken as the dependent variable in the *EA*-test, we first estimate the criteria value of *Elon* by its percentile values. The 5^{th}, 10^{th}, and 15^{th} percentile mean 5%, 10%, and 15% of the ordered observations will be smaller than the percentile values, respectively. That corresponds to 1, 3, and 4 observations and, hence, the choice of 15% percentile seems to be adequate for this data. Table 5.4 shows four observations based on the 15^{th} percentiles over all data in group B. Most of them have rather low values of $Alt(M)$ and *ARCV*, which makes it rather difficult to sight the crescent moon after the sunset. Hence, the value of the criteria for *Elon* is estimated at 5.58°.

As for $Alt(M)$ and *ARCV*, we consider the linear regression model of *Elon* - $Alt(M)$ and *Elon* - *ARCV*. We then estimate the corresponding values of $Alt(M)$ and *ARCV* given that *Elon* = 5.58°. Hence, the values for $Alt(M)$and *ARCV* are 1.33° and 2.80°, respectively.

Consequently, by definition, the estimated *Alt*(*S*) is taken as the difference between *ARCV* and *Alt*(*M*), that is -1.47°. As for Width (W), the observed values are consistently small and we use the 15th percentile as its estimate, which is 0.06°.

We note that the sun's altitude of -1.47°. below the horizon has considered the effect of refraction near the horizon and semi-diameter of the sun. During sunset, the centre of the sun is estimated at 0.35° below the horizon, and hence the estimated time taken for the sun to the altitude -1.47° is 4.5 min after it sets. In determining the final value for the crescent visibility criteria, we use the elongation and altitude of the crescent moon at sunset. We note that the duration of 4.5 min after sunset is considered negligible to elongation as the average rotation rate of the moon surrounding the earth takes about 0.008°/min. Hence, the adjusted values of criteria for *Elon* and *Alt*(*M*) are 5.58° and 2.45°, respectively, and the corresponding values for *ARCV* = 2.80° and *Alt*(*S*) = -0.35. The final new values of the crescent moon visibility criteria are as listed in Table 5.3.

Table 5.3: The values of new criteria of the *EA*-test at sunset.

Variables	Value of criteria (best time data) (°)
Elon	5.58
Alt(*M*)	2.45
ARCV	2.80
Alt(*S*)	-0.35
Width	0.06

Table 5.4: **Observations with *Elon* less than the 15th percentile value.**

Date of moon sighting (Masehi)	Date of moon sighting (Hijr)	*Elon* (°)	*Alt* (*M*) (°)	*ARCV* (°)	*Alt*(*S*) (°)	*Width* (°)	Visibility (Y/N)
25.10.2003	29 Syaaban 1424	1.92	-2.10	-2.36	0.26	0.02	N
20.08.2009	29 Syaaban 1430	2.32	-2.16	-2.31	0.15	0.02	N
06.11.2010	29 Zulkaedah 1431	5.18	0.87	1.69	-0.82	0.06	N
19.07.2012	29 Syaaban 1433	5.72	1.46	2.28	-0.82	0.06	N

5.4 Discussion

The results for minimum or best criteria for the original and best time crescent moon data for each parameter estimates are tabulated in Table 5.5. Several results are observed:

1. For best time data, the estimated value of criteria for *Elon, Alt*(*M*) and *ARCV* are slightly lower compared to those of the original data. As for the best time data, it gives higher probability to observe the new moon. Using data transformation from

original data to best time data, the percentage of correct detection is higher than the original data as shown in table . It shows that, the criteria value of crescent moon visibility will give a better criterion for sighting the new crescent moon as shown in table 5.2.

Table 5.41: Comparison of the percentage of correct detection for best time data and original data.

Model	Percentage of correct detection original data, %	Percentage of correct detection original data, %
Elon- Alt (*M*) (*EA-test*)	70.1	70.5
Elon-ARCV (*EV-test*)	43.7	46.5
ARCV-Alt(*M*) (*VA-test*)	82.0	82.2

2. Based on the *Elon-Alt*(*M*) model, the percentages of detection of a crescent moon for best time data are slightly higher than the original data. It shows that, the accuracy of crescent moon visibility criteria using best time data is better compared to the criteria using original data due to higher percentage of correct detection. The procedure using the best time is slightly better as it gives higher probability of observing crescent moon and hence give better estimation of the crescent moon visibility criteria.

3. The fit of the circular simple regression model to the best time data with *Elon* and *Alt*(*M*) as the variables of the model is better than the fit of the other models since

the two variables have a high correlation between each other and provides a higher percentage of correct detection.

4. The crescent moon visibility criteria using best time data can be used during sighting process. The possibility of observing the new moon is higher with the new criteria shown in table 5.5.

Applying Yallop's procedure to the circular variables in the best time data using the JS circular regression model also provides almost the same results as the original data. We also find that the *Elon-Alt*(*M*) model performs well in estimating the criteria for all the parameters compared to the other regression models. Based on the results obtained, the percentages of correct detection are slightly higher when we apply the procedure to best time data than the original data. We suspect that, by increasing the number of observations, the percentage of visibility will increase too. A data cleaning process may improve the quality of the data and the criteria of crescent moon visibility.

Table 5.5: Comparison of the output best time data and original data.

Variables	Value of criteria (original data) (°)	Value of criteria (best time data) (°)
Elon	7.28	5.58
Alt(*M*)	3.33	2.45
ARCV	3.74	2.80
Alt(*S*)	-0.35	-0.35
Width	0.10	0.06

5.5 Summary

In this chapter, we have reviewed the theory of 'best time' for observing the crescent moon. We have transformed the original data into the best time data and have applied the procedure that we have developed in Chapter 4 to the local best time data. We have observed a higher probability of observing the crescent moon, or a higher percentage of correct detection of the crescent moon obtained using the best time data compared to the original data.

CHAPTER 6: CONCLUSION

6.1 Summary of the study

This research has achieved all four objectives. We have described the local crescent moon data using circular statistics. A new visibility test based on crescent moon sighting data is developed using circular regression in this study. We have determined the new criteria of moon visibility based on the Malaysian crescent moon data. We also have shown that the local best time improves the visibility criteria in this thesis.

Although several variations of observing the new moon are described in the literature, no existing procedure determines the visibility criteria based on the circular statistical theory. The variables useful for the crescent moon visibility are circular types in nature. Thus, the use of circular descriptive statistics is more appropriate to describe the Malaysian crescent moon data. One of the circular descriptive statistics is the circular correlation statistics. Based on the circular correlation statistics, we have identified several combinations of variables with a strong correlation, this includes *Elon – Alt*(*M*), *Elon – ARCV* and *ARCV – Alt*(*M*). This information is useful for building the circular regression model.

First, we have used circular regression based on JS circular regression (Jammalamadaka & Sarma, 1993) method to obtain a new visibility test for crescent moon. We have chosen three circular regression models, which are the *Elon-Alt*(*M*), *Elon-ARCV* and *ARCV-Alt*(*M*) models, due to the strong correlation between the variables. We have shown that all three models are significantly important using the significance test carried out on each polynomial term in the circular model.

Following Yallop (1979) that has developed the visibility test based on the linear model, we have constructed a visibility test based on residuals in the circular regression method. The three circular regression models *Elon-Alt*(*M*), *Elon-ARCV* and *ARCV-Alt*(*M*) were used to develop a visibility test as an indicator to determine the criteria for crescent moon visibility. We have developed three visibility tests *Elon-Alt*(*M*), *Elon-ARCV* and *ARCV-Alt*(*M*) tests that correspond to each model. Three visibility categories are considered in this research; (i) Category A (Visible to the unaided eye), (ii) Category B (May need optical aid) and (iii) Category C (Not visible). Then we have calculated the percentage of correct detection for each visibility test to compare the performance of the three visibility tests. The higher the percentage of correct detection, the better the test. We find that the best indicator for the visibility test is the *Elon-Alt*(*M*) test with the percentage of correct detection 70.1%. Then, we have used the *Elon-Alt*(*M*) test to obtain the crescent moon visibility criteria for the Malaysian crescent moon data, which are 5.58° for *Elon*, 2.45° for *Alt*(*M*), 2.80° for *ARCV*, -0.35° for *Alt*(*S*) and 0.06° for *Width*(*W*).

In the concept of best time data, the original data are transformed into the best time data. The possibility of viewing the moon is higher during the best time data (Yallop, 1974). We have applied the procedures of obtaining the visibility criteria developed using the circular statistical approach to the local best time data. The visibility criteria obtained using the best time data is better than the original data because the percentage of correct detection for the *Elon-Alt*(*M*) test using the best time data is slightly larger than that of the original data. The crescent moon visibility criteria for best time data are 7.28° for *Elon*, 3.39° for *Alt*(*M*), 3.74° for *ARCV*, -0.35° for *Alt*(*S*) and 0.10° for *Width*(*W*).

In conclusion, we have studied the crescent moon data and proposed new visibility criteria using a circular statistical approach. The work shows that the circular statistical method is beneficial to deal with a real world problem, particularly in the moon sighting process.

6.2 Contributions

The study has contributed to the development of new crescent moon visibility criteria using circular analysis in the following ways:

1. The existing methods used in the development of the crescent moon visibility criteria are conducted using linear statistical theories, whereas most variables considered are of circular types. Using the circular statistical theory, the circular descriptive statistics are used to give simple summary statistics that describe the basic features of the Malaysian crescent moon data.

2. We have proposed a generalized JS circular regression model for accommodating two or more explanatory circular variables in the models. The relevant theory is presented and using the JS circular regression procedure, we developed a model for crescent moon data. The model is used to estimate the minimum value of the parameter or the best criteria for moon visibility using Yallop procedure. The model also proven to perform well in determining the crescent moon visibility criteria which give high percentage of correct detection. Then the new visibility criteria for crescent moon have be determined using the visibility tests developed.

3. We have applied the same procedure (Yallop's procedure) using circular regression to best time crescent moon data. The result shows, based on the model that have been developed, the percentage of correct detection is slightly higher than the original data. It proved that it performs well and gives a better result than using original data. As a result, the improved crescent moon visibility criteria have been determined using best time data.

6.3 Further Research

There are various ways for furthering research in this field. Some suggestions are given as follows:

(i) To extend the simple circular regression into multiple circular regression and the logistic circular regression.

(ii) To extend the work with outlier detection in the data and fit the regression model to the cleaned data.

(iii) To use other models apart from the JS model like Down and Mardia DM circular regression model and make a comparison between the methods developed.

(iv) Use data from other sources or merge all the data collected in Malaysia to develop a better circular regression model.

We are aware that there could be other variables that can affect the determination of the first day of the month and the crescent moon visibility criteria issue can been resolved using more appropriate statistical techniques explored in near future.

REFERENCES

Akkoc, H. (2008). Pre-service mathematics teachers' concept images of radian. *International Journal of Mathematical Education in Science and Technology*, *39*(7), 857-878.

Al-Mostafa, Z. A. (2005). Lunar calendars: the new Saudi Arabian criterion. *The Observatory*, *125*, 25-30.

Bell, K. N. I. (2008). Analysing Cycles in Biology and Medicine—A Practical Introduction To Circular Variables & Periodic Regression. (2th ed.). Razorbill Press, St. John's.

Best, D. J. & Fisher, N. I. (1981). The bias of the maximum likelihood estimators of the von Mises-Fisher concentration parameters. *Communication in Statistics – Simulations and Computations*, *10*(5), 394-502.

Binkley, C. S. (1990). Climate change and forests. *School of Forestry*, *14*(24), 20-23.

Bruin, F. (1977). The first visibility of the lunar crescent. *Vistas in astronomy*, *21*, 331-358.

Cremers, J., & Klugkist, I. (2018). One direction? A tutorial for circular data analysis using R with examples in cognitive psychology. *Frontiers in psychology*, *9*, 2040.

Danjon, A. (1936). Nouvelles recherches sur la photometrie ee la lumiere cendree a 1'albedo de la terre. *Ann, l'obs. Strasbourg*, *8*, 139-181.

de Freitas, V. P., Makler, M., & Dúmet-Montoya, H. S. (2018). Strong lensing cross-sections for isothermal models. I. Finite source effects in the circular case. *Monthly Notices of the Royal Astronomical Society*, *481*(2), 2189-2204.

Dhakal, N., Jain, S., Gray, A., Dandy, M., & Stancioff, E. (2015). Nonstationarity in seasonality of extreme precipitation: A nonparametric circular statistical approach and its application. *Water Resources Research*, *51*(6), 4499-4515.

Djamaluddin, T. (2000). New Crescent Moon Visibility in Indonesia. *Majalah LAPAN*, 2, 137.

Downs, T. D. (1974). Rotational Angluar Correlations. *Biorhythms and Human Reproduction*, 97-104.

Down, T.D. & Mardia, K.V. (2002). Circular regression. *Biometrika*, *89*(3), 683-697.

Fakhar, M., Moalem, P., & Badri, M. A. (2014). Lunar Crescent Detection Based on Image Processing Algorithms. *Earth, Moon, and Planets*, *114*(1-2), 17-34.

Fatoohi, L. J., Stephenson, F. R., & Al-Dargazelli, S. (1998). The Danjon limit of first visibility of the lunar crescent. *The Observatory*, *118*, 65-72.

Fisher, N. I. (1995). Statistical analysis of circular data. cambridge university press.

Fotheringham, J. K. (1910). Moon, on the smallest visible phase of the. *Monthly Notices of the Royal Astronomical Society*, *70*, 527.

Gould, A. L. (1969). A regression technique for angular variates. *Biometrics*, 683-700.

Guessoum, N. & Meziane, K. (2001). Visibility of the thin lunar crescent: The sociology of an astronomical problem (A case study). *Journal of Astronomical History & Heritage*, 4, 1-14.

Hasanzadeh, A. (2012). Study of Danjon limit in moon crescent sighting. *Astrophysics and Space Science, 339*(2), 211-221.

Hatami, M., & Alamatsaz, M. H. (2019). Skew-symmetric circular distributions and their structural properties. *Indian Journal of Pure and Applied Mathematics*, *50*(4), 953-969.

Hoffman, R. E. (2003). Observing the new Moon. *Monthly Notices of the Royal Astronomical Society*, *340*(3), 1039-1051.

Hogendijk, J. P. (1988). New Light on the Lunar Crescent Visibility Table of Yaʿqūb ibn Ṭāriq. *Journal of Near Eastern Studies*, *47*(2), 95-104.

Hussin, A.G., Fieller, N.R.J. & Stillman, E.C. (2004). "Linear regression for circular variables with application to directional data". *Journal of Applied Science and Technology*, 8, 1-6.

Ibrahim, S. (2013). *Some outlier problems in a circular regression model*. (Unpublished doctoral thesis) University of Malaya, Kuala Lumpur, Malaysia.

Ilyas, M. (1994). Lunar crescent visibility criterion and Islamic calendar. *Quarterly Journal of the Royal Astronomical Society*, *35*, 425.

Ilyas, M. (1988). Limiting altitude separation in the new moon's first visibility criterion. *Astronomy and Astrophysics*, *206*, 133-135.

Ilyas, M. (1983). The Danjon limit of lunar visibility: A re-examination. *The Journal of the Royal Astronomical Society of Canada*, 77, 214-219.

Jammalamadaka, S. R., & Sengupta, A. (2001). Topics in circular statistics. Singapore: World Scientific.

Johnson, R. A., & Wehrly, T. (1977). Measures and models for angular correlation and angular–linear correlation. *Journal of the Royal Statistical Society: Series B (Methodological), 39*(2), 222-229.

Jona-Lasinio, G., Gelfand, A., & Jona-Lasinio, M. (2012). Spatial analysis of wave direction data using wrapped Gaussian processes. *The Annals of Applied Statistics*, 6(4), 1478-1498.

Kamisan, N. A. B., Hussin, A. G., Zubairi, Y. Z., & Hassan, S. F. (2011). Distribution of wind direction recorded at maximum wind speed: A case study of Malaysian wind data for 2005. *International Journal of Physical Sciences*, *6*(7), 1840-1850.

King, R. G., Plosser, C. I., Stock, J. H., & Watson, M. W. (1987). Stochastic trends and economic fluctuations. *American Economic Review*, *81*,819-840.

Kronauer, R. E., Czeisler, C. A., Pilato, S. F., Moore-Ede, M. C., & Weitzman, E. D. (1982). Mathematical model of the human circadian system with two interacting oscillators. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology*, *242*(1), R3-R17.

Kufner, A., & Kadlec, J. (1971). Fourier Series (English translation by GA Toombs). Iliffe, London.

Kulldorff, M., Huang, L., Pickle, L., & Duczmal, L. (2006). An elliptic spatial scan statistic. *Statistics in medicine*, *25*(22), 3929-3943.

Lund, U. (1999). Least circular distance regression for directional data. *Journal of Applied Statistics*, *26*(6), 723-733.

Mardia, K. V., & Jupp, P. E. (2009). Directional statistics (Vol. 494). John Wiley & Sons.

Mardia, K. V., & Jupp, P. E. (1999). Chapter 6: Tests of Uniformity and Test of Goodness-of-Fit. *Directional Statistics*, 93-118.

Martins, A., Carvalho, A., & Sousa, J. A. (2015, May). Comparing wind generation profiles: A circular data approach. In *2015 12th International Conference on the European Energy Market (EEM)* (pp. 1-5). IEEE.

Maunder, E. W. (1911). On the smallest visible phase of the Moon. *Journal of the British Astronomical Association*, *21*, 355-362.

McNally, D. (1983). "The Length of the Lunar Crescent", *Quarterly Journal of the Royal Astronomical Society*, 24, 417-429.

Nawawi, M. S. A. M., Man, S., Zainuddin, M. Z., Wahab, R. A., & Zaki, N. A. (2015). SEJARAH KRITERIA KENAMPAKAN ANAK BULAN DI MALAYSIA (History of the Criteria for Lunar Crescent Visibility in Malaysia). *Journal of Al-Tamaddun*, *10*(2), 61-75.

Odeh, M. S. (2004). New criterion for lunar crescent visibility. *Experimental astronomy*, *18*(1-3), 39-64.

Pulau Pinang Mufti's Department (2015). *Penentuan Awal Bulan Islam Brosur Info. Falak*. Jabatan Mufti Negeri Pulau Pinang.

Rawlins, D. (2008). Aristarchos Unbound: Ancient Vision. The Hellenistic Heliocentrists' Colossal Universe-Scale. *DIO*, *14*, 13-32.

Samad Abu et al. (2001). *Kaedah Penentuan Awal Hijrah.* Putrajaya: Jabatan Kemajuan Islam Malaysia

Sarala, S., & Jain, P. (2001). A circular statistical method for extracting rotation measures. *Monthly Notices of the Royal Astronomical Society*, *328*(2), 623-634.

Sarma, Y., & Jammalamadaka, S. (1993). Circular regression. In Statistical Science and Data Analysis. *Proceedings of the Third Pacific Area Statistical Conference* (pp. 109-128).

Schaefer, B.E. (1988). "Visibility of the Lunar Crescent", *Quarterly Journal of the Royal Astronomical Society*, 29, 511-523

Schaefer, B. E. (1996). Lunar crescent visibility. *Quarterly Journal of the Royal Astronomical Society*, 37, 759.

Schaefer, B. E. (1991). Length of the lunar crescent. *Quarterly Journal of the Royal Astronomical Society,* 32, 265-277.

Schoch, C. (1930). Tafel fur Neulicht. *Ergaenzungsheft zu den Astronomischen Nachrichten* (1930), *8*(2), B17.

Staggemeier, V. G., Camargo, M. G. G., Diniz-Filho, J. A. F., Freckleton, R., Jardim, L., & Morellato, L. P. C. (2020). The circular nature of recurrent life cycle events: a test comparing tropical and temperate phenology. *Journal of Ecology*, *108*(2), 393-404.

Sultan, A. H. (2007). First visibility of the lunar crescent: beyond Danjon's limit. *The Observatory*, *127*, 53-59.

Suzuki, S., & Yamada, T. (2020). Probabilistic model based on circular statistics for quantifying coverage depth dynamics originating from DNA replication. *PeerJ*, *8*, 8722.

Utama, J. A., & Simatupang, F. M. (2019). The new hilaal visibility criterion for tropical region. In *Journal of Physics: Conference Series* (Vol. 1280, No. 2, p. 022073). IOP Publishing.

Vanem, E., Hafver, A., & Nalvarte, G. (2020). Environmental contours for circular-linear variables based on the direct sampling method. *Wind Energy*, *23*(3), 563-574.

Von Mises, R. (1918). Uber die "Ganzzahligkeit" der atomgewichte und vewandte fragen. *Physikal. Z*., 19, 490-500.

Whittaker, E.T. & Watson, G.N. (1944). A Course in Modern Analysis. Cambridge: Cambridge University Press.

Yallop, B. D. (1997). A method for predicting the first sighting of the new Crescent Moon. *Technical Note* No. 69. HM Nautical Almanac Office, Royal Greenwich Observatory, Cambridge, UK. NAO.

Zainon, O., Ali, H. R., & Hussin, M. F. A. (2019). Comparing the New Moon Visibility Criteria for International Islamic Calendar Concept. In *2019 6th International Conference on Space Science and Communication (IconSpace)* (pp. 144-149). IEEE.