*Available online at http://www.journalcra.com*

*International Journal of Current Research*
*Vol. 9, Issue, 04, pp.48594-48600, April, 2017*

**INTERNATIONAL JOURNAL OF CURRENT RESEARCH**

**ISSN: 0975-833X**

## RESEARCH ARTICLE

# APPLICATION OF CIRCULAR REGRESSION ANALYSIS ON BIOLOGICAL DATA: CASE STUDY ON MALARIA CASES IN DISTRICT OF VISAKHAPATNAM, INDIA

***Desta Firdu MEKONNEN and Prof. Dr. Ensar BAŞPINAR**

Department of Biometry and Genetics, Ankara University, Turkey

**ARTICLE INFO**

***Article History:***
Received 16th January, 2017
Received in revised form
24th February, 2017
Accepted 22nd March, 2017
Published online 20th April, 2017

***Key words:***

Adjusted frequency,
Circular Distribution,
Circular Regression,
Fourier series,
Malaria.

**ABSTRACT**

One of the priorities of malaria elimination or at least reduction has been prevention methods. To apply these methods, understanding transmission influencing factors is crucial. The majority of driving factor of malaria transmission are environmental and climatic features. Plenty of regression analysis methods has been implemented to assess the relationship of malaria transmission and environmental and/or climatic features. Nonetheless, the majority of regression methods for the case lacked robustness. Recently circular regression analysis is gaining acceptance by many academicians to assess scenarios that have circularity in their nature. There are few circular regression models but their applicability and tractability are not fully assessed as for their linear counterpart. We assessed the applicability of these models on Malaria data collected for time series analysis titled "*Association of climatic variability, vector population and malaria disease in the district of Visakhapatnam, India*": we used Fourier methods with ordinary linear regression. We applied different properties of circular distributions and assumptions. We transformed linear time in months to circular time and convert it to radian form since directional analysis is best suited to radian measures. We used malaria cases in months as a dependent linear variable and months in radian as acircular explanatory variable. We have assessed the correlation between these variables and found it right, so we apply circular regression on them and found avery sensible model with few drawbacks.



**Citation: Desta Firdu MEKONNEN and Prof. Dr. Ensar BAŞPINAR, 2017.** "Application of circular regression analysis on biological data: Case study on malaria cases in district of Visakhapatnam, India", *International Journal of Current Research*, 9, (04), 48594-48600.

## INTRODUCTION

Among vector-borne diseases and infections, there is no other disease that challenged our existence as malaria. Malaria is caused by the protozoan parasite Plasmodium. Human malaria is caused by four different species of Plasmodium: *P. falciparum*, *P. malariae, P. ovale* and P. vivax. There are currently over 100 countries and territories where there is a risk of malaria transmission, and these are visited by more than 125 million international travelers every year US Department of Health and Human Science (2007). As one of vector-borne disease, malaria incidence, and transmission is influenced by geographicaldynamics Caminade *et al*. (2014). From different geographical dynamics, climate and weather factors play big (if not the biggest) role for malaria incidence and transmission (Cairns *et al*., 2015; Grassly and Fraser, 2006; Stuckey Smith and Chitnis, 2014) There have been many types of research to assess the relationship of malaria and geographical dynamics in different parts of the world. Many models have been proposed for transmission and seasonality of malaria. Nevertheless, majority of these researches were/are time series like models where they undermine the cyclical nature of the Environmental and seasonal dynamics. On the other hand, if we agree that vector-borne disease in general and malaria, in particular, are really influenced by environmental factors we have to see intra-annual variation than interannual variations because environmental and climatic features like temperature and precipitation show big variations within months of a single year than different years. Moreover, same months of different years tend to have similar environmental and climatic features. This shows that when we analyze the effects of these features we have to take in account repetitive or cyclical nature of these features. That is where circular analysis arises. There have been attempts to usecircular analysis methods to apply for predicting, assessing and forecasting of the relationship between environmental and climatic features and occurrence, transmission and even epidemiology of vector-borne disease in general and malaria in particular. Different statisticians propose different circular modelsbut applicability, tractability, and interpretability of these models to avariety of scenarios is still in questions. In this paper, we apply circular regression models to assess the applicability of some of these models to malaria data collected in Visakhapatnam district, India from 2005-2011 for time series analysis. Before doing that we introduce some concepts in circular data, analysis methods, distributions related to circular analysis and limitations of circular analysis.

****Corresponding author:*** **Desta Firdu MEKONNEN,**
Department of Biometry and Genetics, Ankara University, Turkey.

## Circular Data

Circular analysis is very different in many ways from customary methods used in the majority of analysis and estimation. The first unique feature of circular analysis is the data type. Circular data arises mainly from two types of measurements that are compass and clock in circular sense (Rao, 2001). Direction of migrations of birds Ozarowska (2013), wind direction, latitude and longitude of the globe and direction of epidemic diseases are few examples of this kind of measurements. There are plenty of examples and researches for theclock in circular sense. Data types like emergency arrival of patients in hospital within 24 hours, number of accidental death in USA within a year Herone (2016) occurrence of any epidemic disease in cyclical time Legrand, t al. (2007) are very few of them. The natural measurement of circular data is degrees or radians. Therefore, circular time measurements have to be changedinto degrees or radians by the following conversion equation. $\theta = \frac{2\pi * x}{y}$ Where $\theta$ is time in radian, $X$ is time to convert in to radian and $Y$ is afull cycle of time. Normally Radian is considered better measurement than degrees because of arbitrariness of degree measures (Kupkova, 2005). The other unique feature of circular data is there is no natural zero point and sense of direction. 45° (1/4π rad) to the east might be zero for mathematician or 60° (2/3π rad) to the north may be zero point for ecologist. Due to this fact, rank-based statistics have to be done with ultimate caution. Circular measurement can be done into two ways. It can be Arc length of $\widehat{p1p2}$ on the circumference of the circle or an angle those points make from the center of the circle. In circular data analysis majority of the objectives are to analyze are directions so knowing magnitudes of circular data has less importance. Due to this fact according to Gill and Hangartner (2010), usually circular data are represented in a unit circle, which means a circle with unit radius.

## Circular distributions

Prior to any circular analysis knowing the distribution of the population of the data in hand is very important. Usually, it is mandatory too since we use Maximum Likelihood Estimation (MLS) or Ordinary Least square Estimation (OLS) methods for inference or predictions. Moreover, knowing the distribution of circular variables helps to assess if there is multimodality. If there is multimodality in distribution the statistical method we use will change, otherwise, we may arrive at completely different result than we supposed to have. Circular distribution is a probability distribution whose total probability concentrated on the circumference of a unit circle (Mardia and Jepp, 1972). Each data point on the circumference of a circle represents directions. Therefore, this is how to assign to probability for different data points. The range of circular random variable measured in radian may represented to be[0,2π) or [-π, π) similar wise in degree measure [0,360°) or [-180°, 180°). In both ways,circular random variable is bound to full circle. Circular continuous distribution has the following basic properties

$$(i).f(\theta) \geq 0;$$

$$(ii).\int_0^{2\pi} f(\theta)d\theta = 1;$$

$$(iii).f(\theta) = f(\theta + k.2\pi)$$

The first and second properties are similar to linear distributions. The third and unique property of circular distribution says random variables θ1, θ2, θ3..,θn have same probability after making afull circle (2π) with k=1,2,3,… a number of rotations. Circular distributions can be made in different ways the most three common ways are by wrapping a linear distribution around the unit circle. Through characterizing properties such as maximum entropyand by transforming bivariate linear random variable to just its directional components, the so-called offset distribution. There are different circular distributions. We present and summarize the three important distributions for this paper that are circular uniform distribution, circular normal distribution and circular bivariate distribution.

## Circular Uniform Distribution

One of the most important distributions of circular analysis is circular uniform distribution. When data values distributed uniformly on the circumference circle there will not be preferred direction (circular mean) and circular variance, since all directions are equally likely, the distribution is also called isotopic or random distribution. Circular uniform distribution shows a constant density.

$$f(\theta) = \frac{1}{2\pi} \text{ ; Where } 0 \leq \theta < 2\pi.$$

## Circular Normal Distribution

Angular random variable θ or a linear variable X,(wrapped on the circumference of a circle)said to have Von mises or circular normal distribution if it has the density function

$$f(\theta, \mu, \kappa) = \frac{1}{2\pi\kappa} e^{k\cos(\theta-\mu)} \text{ ; where } 0 \leq \theta < 2\pi,\ 0 \leq \mu < 2\pi,\ K \geq 0$$

$I_0(k)$ is modified Bessel function of the first kind and order zero.

$$I_0(\kappa) = \frac{1}{2\pi}\int_0^{2\pi} \exp(\kappa\cos\theta)d\theta = \sum_{r=0}^{\infty}\left(\frac{\kappa}{2}\right)^{2r}\left(\frac{1}{r!}\right)^2$$

Properties of circular normal distribution

I. Symmetry: circular distribution is symmetrical distribution about mean direction μ and (μ+π).
II. Mode: - since the cosine function has amaximum value at zero, the circular normal distribution is maximum at θ=μ.

$$f(\mu) = \frac{e^{\kappa}}{2\pi I_0(\kappa)}$$

III. Antimode : - antimode of a circular normal distribution is at μ±π, since cosine of π =-1 which is the minimum value.

$$f(\mu \pm \pi) = \frac{-e^{\kappa}}{2\pi I_0(\kappa)}$$

**IV.The role of k**

When one see the ratio of mode versesantimode the result equation is

$$\frac{f(\mu)}{f(\mu+\pi)} = \frac{\dfrac{e^{\kappa}}{2\pi I_0(\kappa)}}{\dfrac{-e^{\kappa}}{2\pi I_0(\kappa)}} = e^{2\kappa}$$

The above equation shows the effect of ratio of the mode and antimode is only depend on the concentration parameter k. the higher the ratio shows the wider the gap between mode and antimode, the wider the gap results the Peaker the distribution. This is due to the higher concentration of the data towards the mean direction. The concentration parameter k plays the role of variance hence standard deviation in linear normal distribution.

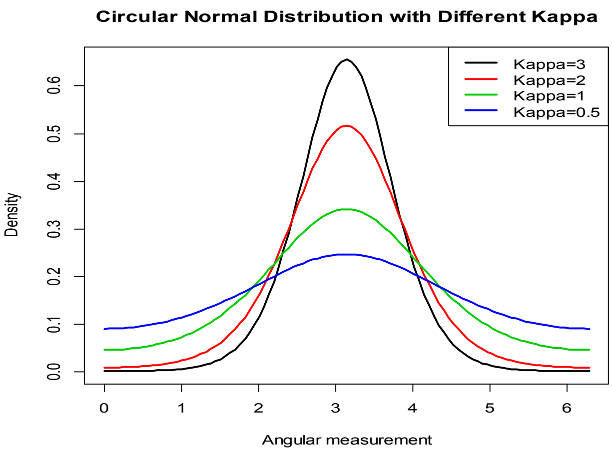



**Figure 1. Circular normal distributions with varying concentration parameters**

### Circular Bivariate distributions

In linearcase, as bivariate is anextension of distribution of variable in line into a plane of two dimensions, bivariate circular distribution is anextension of univariate circular distribution in a complex plane into atoroidal sphere. The probability density function of two circular random variables (linear variables wrapped into acircle) $\theta_1$ and $\theta_2$ Bivariate and multivariate distributions are mostly expressed in Von Mises or normal distribution cases. This is due to the complexity of circular distributions. The importance of bivariate normal distributions in molecular science pinpointed by Mardia, *et al* (2008).

$$f(\theta_1,\theta_2) = \frac{\exp\{\kappa_1\cos(\theta_1-\mu_1)+\kappa_2\cos(\theta_2-\mu_2)+\lambda_{12}\sin(\theta_1-\mu_1)\sin(\theta_2-\mu_2)\}}{T(\kappa_1,\kappa_2\lambda_{12})}$$

For $-\pi \leq \theta_1$ and $\pi \leq \theta_2$, Where K1, $K_2 \geq 0$, $-\infty < \lambda_{12} < \infty$, $-\pi \leq \mu_1$, $\mu_2 < \pi$

The normalizing constant T(.) is given by the equation below

$$T(\kappa_1\kappa_2\lambda_{12}) = 4\pi\sum_{m=0}^{\infty}\binom{2m}{m}\left(\frac{\lambda}{2}\right)^{2m}\kappa_1^{-m}I_m(\kappa_1)\kappa_2^{-m}I_m(\kappa_2)$$

## MATERIALS AND METHODS

The method in this paper is demonstrated on the data that has been used for "association of climatic variability, vector population and malaria disease in district of Visakhapatnam", India: a Modeling and prediction analysis" Chandra, *et al.* (2015) in which data collection methods have been described in detail. Monthly malaria data were collected from the district for the2005-2011 period. Ravi *et al.* in their data analysis since their objective is time series like analysis, they had to specify the years when cases occur separately. On the other hand, in this paper, since the main analysis objective is circular, we only see the years as replication. Which means, it does not matter whether a case occurs in 2005 or 2006 as long as they occur in same months they have to be considered as same measurement group. For this study, we analyzed circular correlation and regression between time of the year and malaria case for the malaria species (*P.falciparum*) in the 28th index study site, which is Visakhapatnam city. We use R and excel software packages to carry out calculations, analysis and plotting process since R has desirable packages "circular" and "CircStat" to carry out the analysis. Fourier series expansion methodisused to get coefficients of the regression at the determined order of polynomial. Calculation Coefficients of Fourier series is clearly presented is Kido (2014). The linear regression analysis method is used to show advantages and drawback of circular regression methods when it is used in such types of observations.

### Data Adjustment

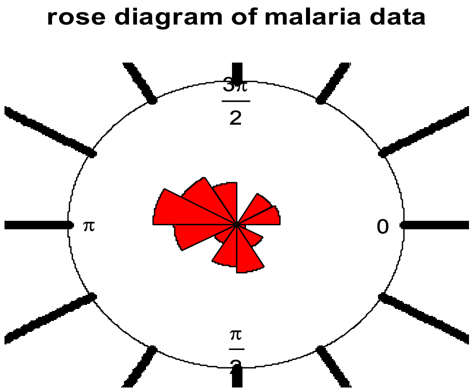
Our first step is to convert months into radian and adjust to the way one day represents 1 degree (0.017453 rad). Since months in a year do not have equal days, we have to adjust each month in the way to contribute equal frequencies. To do so, we have changed 365 days in a year to 360°.

First, we divide 30 by a number of days in each month we have the rate of each month's contribution to 360°.

$Adjustment = \frac{30}{x}$ ;Where$x$is days of each month

## RESULTSAND DISCUSSION

Visualization of the data says too many things about the data type and distribution forms. It also gives some important clue about what kind of analysis methods to follow and what types of results to expect. Due to this fact, it is customary to use agraphical representation of the data in hand. In the this chapter, we plot the relationship of malaria data in Visakhapatnam district over time of the year and see if there is a relationship between time of the years (months) and malaria distribution. According to the Figure (2A), circular plot of the data, the distribution of malaria case is highest in between July and September. In the linear plot as in Figure (2B) it looks the distribution is abit skewed to the left. In circular plots, skewness of any distribution usually depends on the choice of starting point of measurements. If we decide the choice of the starting point is February (1/6π rad), the January distribution would shift to the end and the distribution would look" more" normal. One of the advantages of circular data management is the choice of starting point, which is arbitrary, so we can have a circular normal distribution opposite to when we manage linear data. One who wants to do time series analysis can easily see from the figure (2C) that as time goes by malaria cases increase. Our first step is to convert months into radian and adjust to the way one day represents 1 degree (0.017453 rad). Since months in a year do not have equal days, we have to adjust each month in the way to contribute equal frequencies. To do so, we have changed 365 days in a year to 360°.



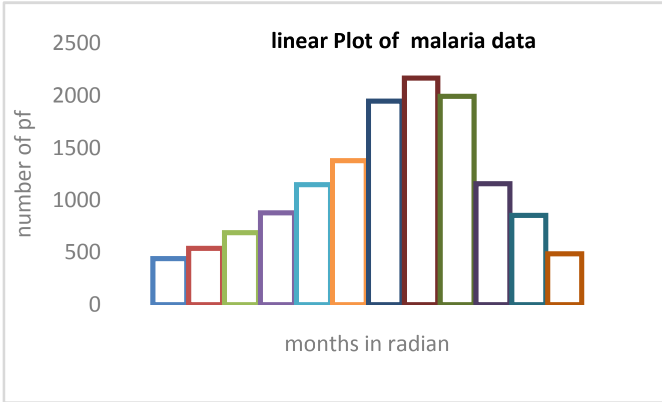




**Figure 2A. Circular plot of dataFigure 2B linear plot of data**

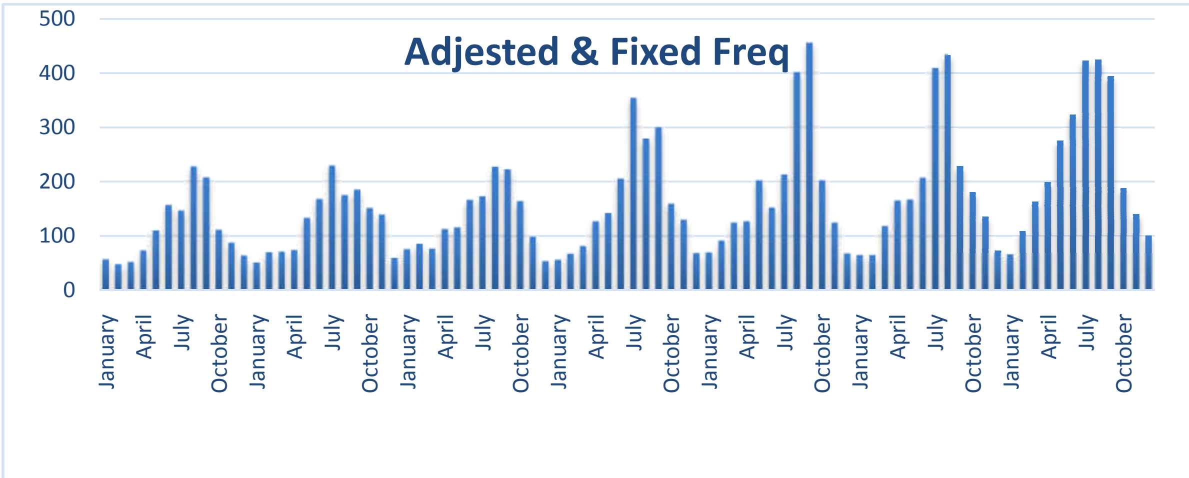



**Figure 2C. Time series plot of Adjusted and fixed frequencies**

**Table 1. Frequency adjustment**

| Index | Months | Days | Adjustment | Month | degree | radian | Pf |
|---|---|---|---|---|---|---|---|
| 1 | January | 31 | 0.967741935 | 1 | 1 | 0.017453293 | 58 |
| 2 | February | 28 | 1.071428571 | 2 | 30 | 0.523598776 | 44 |
| 3 | March | 31 | 0.967741935 | 3 | 60 | 1.047197551 | 53 |
| 4 | April | 30 | 1 | 4 | 90 | 1.570796327 | 72 |
| 5 | May | 31 | 0.967741935 | 5 | 120 | 2.094395102 | 112 |
| 6 | June | 30 | 1 | 6 | 150 | 2.617993878 | 154 |
| 7 | July | 31 | 0.967741935 | 7 | 180 | 3.141592654 | 149 |
| 8 | August | 31 | 0.967741935 | 8 | 210 | 3.665191429 | 231 |
| 9 | September | 30 | 1 | 9 | 240 | 4.188790205 | 204 |
| 10 | October | 31 | 0.967741935 | 10 | 270 | 4.71238898 | 113 |
| 11 | November | 30 | 1 | 10 | 300 | 5.235988 | 138 |
| 12 | December | 31 | 0.967741935 | 11 | 330 | 5.759587 | 102 |

**Table 2. Correlation table of circular regression analysis of 1-4 order of polynomials and linear**

| *Regression statistics* | *1st order* | *2nd order* | *3rd order* | *4th order* | *Linear* |
|---|---|---|---|---|---|
| Multiple R | 0.218446323 | 0.765908098 | 0.813802906 | 0.814898154 | 0.267440806 |
| R square | 0.047718796 | 0.586615215 | 0.66227517 | 0.664059001 | 0.071524584 |
| adjusted R square | 0.02420568 | 0.56568434 | 0.635958949 | 0.628225294 | 0.060201714 |
| Standard deviation | 101.6527865 | 67.81767387 | 62.08909426 | 62.74513589 | 99.76023634 |
| observations | 84 | 84 | 84 | 84 | 84 |

**Table 3. Overall model ANOVA of 1-4 order of polynomial and linear regression analysis**

| *Regression* | *1st order* | *2nd order* | *3rd order* | *4th order* |
|---|---|---|---|---|
| residual *df* | 81 | 79 | 77 | 75 |
| Regression SS | 41941.87678 | 515598.5713 | 582099.0024 | 583666.8797 |
| Error SS | 836996.4088 | 363339.7142 | 296839.2832 | 295271.4059 |
| P-values | 0.138035598 | $1.69552 \times 10^{-14}$ | $2.54673 \times 10^{-16}$ | $5.37109 \times 10^{-15}$ |
| observations | 84 | 84 | 84 | 84 |

First, we divide 30 by a number of days in each month we have the rate of each month's contribution to 360°.

$Adjustment = \frac{30}{x}$ ;Where $x$ is days of each month

We multiplied the adjustment result by the frequency of each month presented in methodology chapter. To retain the total frequency we added the adjusted frequencies and divide by the sum of frequencies before adjustment. Then, we multiplied the result by adjusted frequencies. The resulting frequencies is shown in Table (1)

From the above table, we used the seventh column, which is radian measure of days and radian, and 8th column, which is adjusted and fixed frequencies to our analysis.

**Circular-Linear Association of Time of the Year and Malaria Case**

From Figure (3), one can clearly see that there is one peak within a period (from January-next January). Sinusoidal regression curve can be used to regress cyclical time on malaria incidence. Before the regression, it is appropriate to know whether there is correlation (association) between the dependent linear variable number of malaria cases and a circular variable; months of the year in radian, which is circular using the formula bellow

$$r^2 = \frac{r_{mc}^2 + r_{ms}^2 + 2r_{mc}r_{ms}r_{cs}}{1 - r_{cs}^2}$$

Where
$r_{mc}$=corr(case,cos(months in radian),$r_{ms}$=corr(case,sin(months in radian),
$r_{sc}$=corr(cos(months in radian, sin(months in radian),

$$\frac{(n-3)r^2}{1-r^2} \sim F_{2,(n-3)}$$
$$116.8431 \sim 3.1065$$

**Circular-Linear Regression of Time of the Year on Malaria Case**

From the result of correlation analysis above, we found enough evidence to reject the null hypothesis that claims there is no significant correlation between these variables. After being confident, malaria case and months of the year in radian have circular-linear association, we proceed into sinusoidal regression analysis using least square method and Fourier series expansion. Using series expansion logic and least square methods the equation of circular regression for this data is presented as follows.

$$y_j = A_0 + A_1\cos(\omega\theta - \varphi) + A_2\cos(2\omega\theta - \varphi) + \ldots + A_p\cos(p\omega\theta - \varphi)$$
$$+ B_1\sin(\omega\theta - \varphi) + B_2\sin(2\omega\theta - \varphi) + \ldots + B_p\sin(p\omega\theta - \varphi) + \varepsilon_j$$

Where
$\theta$ is the independent angle and *P=1,2,3…* are the order of polynomial
$A_0$ is a mean level, ω is angular frequency.

$A_{1,}A_{2,\ldots}A_p$ and $B_{1,}B_{2,\ldots}B_p$ are coefficients to be predicted from the model.

$\varphi$ is peak of angular time where the frequency is highest. It is called acrophase.

We used Ordinary Least square method for the above Fourier series equation. The biggest challenge in circular regression analysis is determination of the order of polynomials. One way to tackle this challenge is running the regression with one more order of polynomial and see if the regression coefficient is significant at this order. If regression coefficient at this order of polynomial is also significant, run the regression until it is not significant anymore. Accept the model as "best" model the one with one less order of polynomial of the model with insignificant order. Determination of order of polynomial is

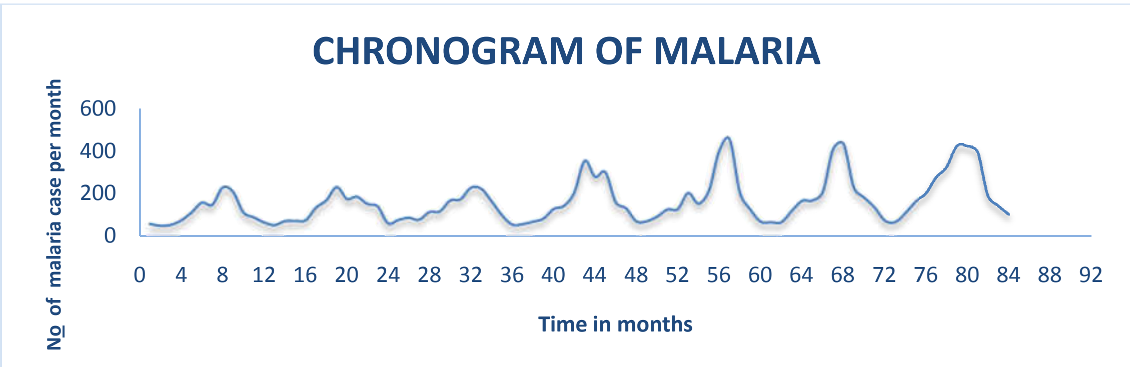



**Figure 3. Chronogram of Malaria distribution**

**Table 4. The Pth order of polynomial p-values**

| *Regression* | *1st order* | *2nd order* | *3rd order* | *4th order* |
|---|---|---|---|---|
| P-cosine | 0.083739976 | $8.54725\text{X}10^{-09}$ | 0.00396409 | 0.848432743 |
| P-sine | 0.487609911 | 0.091924757 | 0.004250338 | 0.679340993 |

Using the above formula, we found that $r^2$ to be 0.59058; we concluded that there is fairly enough association between these two variables. If we suspect the resulting value is not enough to conclude if there is circular-linear association, we can apply F-test used in (Rao, 2001) to reject the null hypothesis claiming there is no circular-linear association.

**Table 5. Regression coefficients table of third order of polynomial**

| | *coefficients* | *Standard dev.* | *t Stat* | *P-values* | *lower %95* | *Upper %95* |
|---|---|---|---|---|---|---|
| *intercept* | 145,1841744 | 55,19465468 | 2,630402804 | 0,010295531 | 35,2775757 | 255,0907732 |
| *cos(ωθ-φ)* | -28,27173067 | 38,76168972 | -0,729373019 | 0,467986426 | -105,4561157 | 48,91265436 |
| *sin(ωθ-φ)* | 5,124309228 | 9,827532009 | 0,521423815 | 0,603568107 | -14,44480708 | 24,69342554 |
| *cos(2ωθ-φ)* | 124,2057774 | 19,43201428 | 6,39181176 | 1,14941E-08 | 85,51169355 | 162,8998613 |
| *sin(2ωθ-φ)* | 114,1844083 | 45,17945049 | 2,527352747 | 0,013541771 | 24,22062893 | 204,1481877 |
| *cos(3ωθ-φ)* | 34,93872006 | 11,76197002 | 2,970481986 | 0,00396409 | 11,51764556 | 58,35979457 |
| *sin(3ωθ-φ)* | -41,17719323 | 13,97489504 | -2,946511805 | 0,004250338 | -69,00476441 | -13,34962205 |

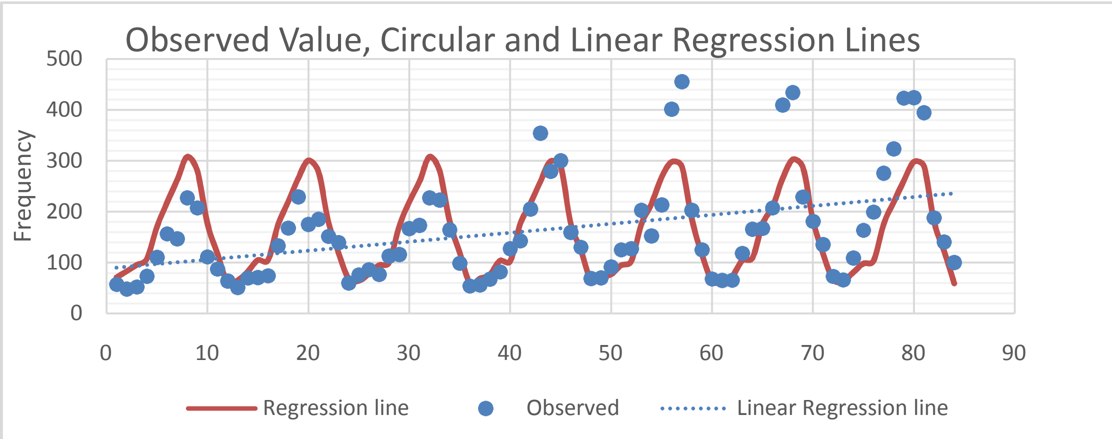



**Figure 4. Plot of observed and predicted values**

related to reduction of error sum of squares. As Rao (2001) presented, it helps to see the proportional reduction of error sum of square as the order of polynomial increases because circular-linear regression as a sinusoidal regression with Fourier series is family of polynomial analysis.

We use excel regression analysis method for Fourier regression equation above.

Table 2shows the dependent variable malaria case and components of the independent variable time of the year in radian measurement have increasing association with decreasing standard deviations when order of polynomial increases. The coefficient of determinations also increasing as order of polynomial increases but on the fourth order of polynomial, the increment of coefficient of polynomial is very small. This might give some clue the fourth order of polynomial is not significant anymore. When we compare circular regression methods based on varying order of polynomials with linear regression,4.77%, 58.66%,66.22%, 66.40% source of variations are explained by our models respectively whereas only 7.15% of source of variation is explained by linear regression models. This implies that circular regression method is better regression method than the linear regression for these scenarios.

P-values in Table 3 shows overall models are significant for all of order of polynomials. The main advantage of increasing of order of polynomial is reduction of the sum of squares of residuals (Nurhab *et al*., 2014). As one can see from the table 3 residuals sum of squares are decreasing from the first order of polynomial through the third order of polynomial. When we see the difference of residual sum of squares between 3$^{rd}$ and 4$^{th}$ order of polynomial in is minimal. This result also gives us another clue inclusion of 4$^{th}$ order of polynomial is not significant. In circular regression analysis, when one order of polynomial is significant we suspect that the next order of polynomial might be significant as well. Due to this fact, we run our regression analysis repeatedly until both components of polynomial order is not significant anymore. When we run the regression analysis for the first order of polynomial, we found both of the components are not significant but it is better to run p+1 order of polynomial and see the result if the components are not significant. So we have run the regression with one more order of polynomial and found that the regression is significant at that level too (second order of polynomial). We run the regression for the third order of polynomial and found significant again. When we run the regression for the fourth order of polynomial we found both cos(ω4θ-φ) and sin(ω4θ-φ) are not significant anymore as shown in Table 4.The inclusion of coefficients of 4$^{th}$order polynomial is not appropriate. Therefore, the regression model with three order of polynomial is our best model. This model with its coefficients, t-stats, and p-values, upper and lower boundaries are presented in Table 5.

$$y = 145.1842 - 28.2717\cos(\omega\theta-\varphi) + 124.2058\cos(2\omega\theta-\varphi) + 34.93872\cos(3\omega\theta-\varphi) + 5.124309\sin(\omega\theta-\varphi) + 114.1844\sin(2\omega\theta-\varphi) - 41.1772\sin(3\omega\theta-\varphi)$$

InFigure4, one can see that the regression wave has similar trend with the frequency wave of observed data. The regression wave looks like having high peak due to high frequency of observed data around month 57,58,80,81. These values look more outliers when we used linear regression line. The reason for the higher peak in these months depends on factors like whether there is anepidemic of malaria in these months or some other environmental and social factors like sudden population size change. Compression between linear and circular model

**Conclusion**

Environmental and climatic features influence malaria transmission. These environmental and climatic features are time dependent variables. Time-dependent variable has circular nature and the best way to see this circular nature is monthly. One of time dependent variable is malaria transmission in different parts of the world especially in tropical areas where climatic variation is higher. One of the countries that has higher malaria is India. Circular regression analysis has been done to regress months of the year on malaria case. Since circular analysis in general and circular regression, in particular in relatively new field of statistics first, we highlighted some basic underlying features and properties of circular statistics. We have seen circular distributions, particularly Von Mises distribution. Our assumption for the regression analysis was our data follows Von Mises distribution. Based on our assumption we used Fourier regression method to determine coefficients of regression. We found asignificant association between radian time and malaria occurrence in the study area. The study has also some limitations. Since we used data collected or other analysis methods suiting the data for our analysis was bit handy. There was also alimitation of statistical packages specific to the analysis especially when one think about multivariate case. This analysis method can be used for any circular time-dependent regression analysis such as Ebola outbreak and time dependency in Tropical countries, Malaria seasonality in African countries… and so on.

**Acknowledgement**

We would like to express our sincere gratitude to Dr. Ravi Chandra Pavan Kumar S.T.P., DrNageswaraRao R.N. and NareshSidagam for their brilliant publication, which we used as a basis for our paper. We would like to thank Indian Ministry of Health and Family welfare for their data. Finally, we would like to pass our acknowledgment to PLOS-one for making the data Public we implemented to verify on our model.

## REFERENCES

Cairns *et al*. 2015. Seasonality in malaria transmission: implication for case-management with long-acting artemisinin combination Therapy in Sub-Saharan Africa. *Malaria Journal*, DOI: 10.1186/s12936-015-0839-4.

Caminade, C.C. *et al*. 2014. Impact of climate change on global malaria distribution. *ProcNatlAcadSci USA*. 111(9), 3286-3291.

Chandra, R. 2015. Association of climatic variability, vector population and malaria in district of Visakhapatnam,India, A modelling and prediction analysis. *PLOS ONE,* 10 (6).

Gill, J. and Hangartner, D. 2010. Circular data in political science and how to handle it. *Oxford Journals*, 18(3), 316-336.

Grassly, N.C. and Fraser, C. 2006. Seasonal infectious diseases epidemiology. *Proc.BiolSci.,* 273 (1600), 2541-2550.

Gubel, E.J., Greenwood, J.A. and Duran, D. 2012. The circular normal distribution: Theory and table. *Journal of American Statistical Association*, 48(261), 131-152.

Herone, M. 2016. Deaths: Leading causes for 2013. *National vital statistics reports*. 65(2).

Jammalamadaka, S.R. 2001. Topics in circular statistics. Singapore, *World scientific publishing Co.Pte.Ltd.*

Kido, K. 2014. Digital Fourier analysis: Fundamentals. *Springer.* Japan.

Kupkova, E. 2005. Radians vs Degrees. *Actadidactica universitstis comenianae mathematics*. Issue (5).

Legrand,J. Understanding the dynamics of Ebola epidemic. *Epidemoil Infect*., 1 34(4), 610-621.

Mardia, K.V. and Jepp, P.E. 1972. Directional statistics. London, *Academic press Inc*.

Mardia, K.V. *et al*. 2008. Multivariate von Misses distribution with application to Bioinformatics. *Statistical society of Canada*, 36(1), 99-109.

Nurhab, M.I.,Kurnia, A. and Sumertajaya, I.M. 2014. Circular-Circular Linear regression analysis of order m in circular variable α and β against linear variable (Y). *IOSR Journal of Mathematics (IOSR-SM).*10(4),49-54

Ozarowska, A. 2013. A new approach to evaluate multimodal orientation behavior of migratory passerine birds recorded in circular orientation cages. *Journal of Experimental Biology*, 216, 4038-4048.

Stuckey, E.M., Smith, T. and Chitnis, N. 2014. Seasonally dependent relationships between indicators of malaria transmission and disease provided by mathematical model simulations. *PLOS, Computational Biology,*10(1371).

U.S Department of Health and Human Science 2007. Understanding Malaria. *National institute of allergy and infectious disease.* No 07-7139.

*******