Journal of Applied Sciences 5 (1): 138-143, 2005
ISSN 1607-8926

Pseudo-replicates in the Linear Circular Functional Relationship Model

[1]Abdul Ghapor Hussin, [2]Nick Fieller and [2]Eleanor Stillman
[1]Centre for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur, Malaysia
[2]School of Mathematics and Statistics, University of Sheffield, Sheffield S10 2UN, England

Abstract: The relationship between readings on two instruments may be represented by a linear functional relationship with errors of observation in both variables. This study describes a method of fitting the relationship between the two circular variables and the replicate observations are available. Maximum likelihood estimates are used and it is shown that the closed-form expression for the estimators are not available and the estimates may be obtained iteratively by choosing a suitable initial value. The model is illustrated with an application to the analysis of the wave and wind direction data recorded by two different instruments and assuming the pseudo-replicates based on time have been obtained from this unreplicated circular data.

Key words: Circular variables, error-in-variables model, maximum likelihood method, functional relationship model

INTRODUCTION

Circular random variable is one which takes values on the circumference of a circle, i.e. they are angles in the range $(0, 2\pi)$ radians or $(0°, 360°)$. This random variable must analysed by techniques differing from those appropriate for the usual Euclidean type variables because the circumference is a bounded closed space, for which the concept of origin is arbitrary or undefined. A continuous linear variable is a random variable with realisations on the straight line which may be analysed straightforwardly by usual techniques. The linear circular functional relationship model refers to the case when both variables are circular and as analogy to the linear functional relationship model, we assume both observation of circular variables X and Y are observed with errors. It is also assumed that the errors are independently distributed and follow the von Mises distribution with mean zero and concentration parameters κ and v, respectively.

Hussin[1] has shown that the full linear circular functional relationship model with unknown X and Y error concentration parameters is inestimable under condition of practical interest, unless replicate observation are available or assumption are made about the ratio of error concentration parameters. This is a similar condition hold for the unreplicated linear functional relationship model, in which we have to know the ratio of error variance[2]. This study proposed the replicated circular functional relationship model, assuming replication can be made available, i.e by identifying groups of pseudo-replicates based on time from the unreplicated circular data. The study presents the model for the replicated linear circular functional relationship, establish the notation and give the estimates of all the parameters. The model is illustrated with an application to the analysis of the wind and wave direction data recorded by two different techniques: the HF radar system and an anchored wave buoy.

THE WIND AND WAVE DIRECTION DATA

The data of particular interest to this study were collected along the Holderness coastline (the Humberside coast of the North Sea, United Kingdom) by using two different techniques which are the HF radar system and anchored wave buoy. The deployment began in October 1994 as part of an experiment studying the transport of sediment away from the coast. There were 129 measurements recorded by HF radar and anchored wave buoy respectively over the period of 22.7 days for wind direction data and 78 measurements over the period of 79.9 days for wave direction data. Figure 1 and 2 show two sets of data that have been collected during the experiment. Measurements of directions are in radians and note that in Fig. 1 and 2 the scale is artificially broken at 0 (or equivalently 2π) radians so the apparent outliers in the top left corner are in fact measurements which are close to those which appear in the bottom left corner or top right corner because of wrap-around observations or measurements from 2π back to 0. For this reason, such

Corresponding Author: Abdul Ghapor Hussin, Centre for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur, Malaysia E-mail: ghapor@um.edu.my

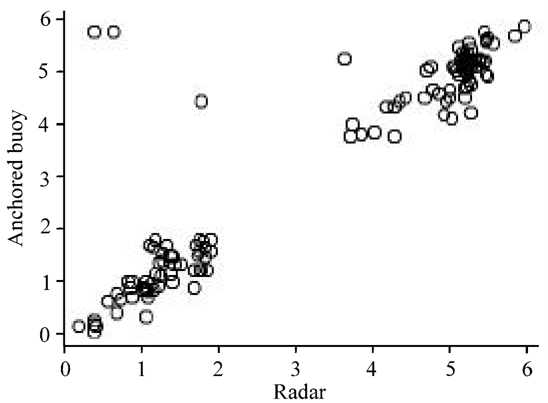

Fig. 1: A scatter plot of wind direction data measured by HF radar system and anchored wave buoy

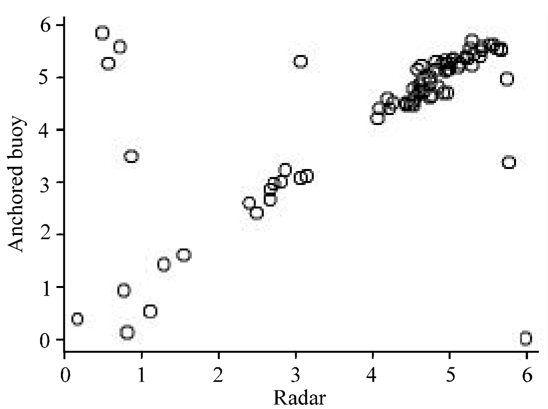

Fig. 2: A scatter plot of wave direction data measured by HF radar system and anchored wave buoy

type of data set cannot be analysed by ordinary linear regression model (which is appropriate if both variables are continuous linear variable).

THE REPLICATED LINEAR CIRCULAR FUNCTIONAL RELATIONSHIP MODEL

In the unreplicated linear circular functional relationship, it iss assumed that the circular or angular observation (x_i, y_i) measures, inexactly, the parameters (X_i, Y_i), where Y_i is a function of X_i, for example, $Y_i = \alpha+\beta X_i \pmod{2\pi}$. However, it is often found that, corresponding to a particular pair (X_i, Y_i) there may be replicated observations of X_i and Y_i occurring in p groups. Measurements y_{ik}, $(j = 1,..., m_i)$ are made on X_i and measurements y_{ik}, $(k = 1,..., n_i)$ on Y_i, where $0 \le x_{ij}, y_{ik} < 2\pi$. Note that there is no inherent pairing of any x_{ij} with y_{ik}, indeed m_i may not equal n_i. We also assume that the observations on X_i and Y_i have been measured with errors δ_i and ε_i, respectively.

The full model for the replicated linear circular functional relationship is therefore:

$$x_{ij} = X_i+\delta_{ij} \text{ and } y_{ik} = Y_i + \varepsilon_i \qquad (1)$$

where, $Y_i = \alpha + \beta X_i \pmod{2\pi}$, for $i = 1,..., p$, $j = 1,..., m$ and $k = 1,..., n_i$

The errors δ_{ij} and ε_{ik} are homogeneous and independently distributed with a von Mises distributions of zero mean circular, i.e. $\delta_{ij} \sim VM(0,\kappa)$ and $\varepsilon_{ik} \sim VM(0, v)$. There are (p+4) parameters to be estimated, i.e. α, β, κ, v and the incidental parameters X_i, $i = 1,..., p$ by the maximum likelihood method. In the replicated case, it may also make different assumptions about the errors, for example the parameter concentration of $(\delta_i, \varepsilon_i)$ may differ for different groups, that is be (κ_i, v_i) or may proportional to each other, that is $(\kappa, c\kappa)$ for some known constant c.

MAXIMUM LIKELIHOOD ESTIMATION OF THE MODEL

Suppose L is the log likelihood function of model (1). Then

$L(\alpha, \beta, \kappa, v, X_i,..., X_p; x_{11},..., x_{pm_p}, y_{11},..., y_{pn_p}) = -NM \log (2\pi) - N \log I_0(\kappa) - M \log I_0(v) + \sum\sum \kappa \cos(x_{ij}-X_i) + \sum\sum v \cos(y_{ik}-\alpha-\beta X_i)$

where, $N = \sum_{i=1}^{p} n_i$, $M = \sum_{i=1}^{p} m_i$ and as usual $I_0(v)$ and $I_0(\kappa)$ are the modified Bessel functions of the first kind and order zero. The estimates of α, β, κ, v and X_i, $i = 1,..., p$ may be solved iteratively given some suitable initial values at the estimate. The first partial derivative of the log likelihood function with respect to α is

$$\frac{\partial L}{\partial \alpha} = \sum \sum v \sin(y_{ik} - \alpha - \beta X_i)$$

Setting this equal to zero and simplifying we get:

$$\hat{\alpha} = \tan^{-1} \left\{ \frac{\sum \sum \sin(y_{ik} - \hat{\beta}\hat{X}_i)}{\sum \sum \cos(y_{ik} - \hat{\beta}\hat{X}_i)} \right\}$$

$$= \tan^{-1}\left(\frac{S}{C}\right), \text{say.}$$

That is:

$$\hat{\alpha} = \begin{cases} \tan^{-1}\left(\frac{S}{C}\right), & S>0, C>0 \\ \tan^{-1}\left(\frac{S}{C}\right) + \pi, & C<0 \\ \tan^{-1}\left(\frac{S}{C}\right) + 2\pi, & S<0, C>0 \end{cases}$$

The first partial derivative with respect to X_i is:

$$\frac{\partial L}{\partial X_i} = \kappa \sum_j \sin(x_{ij} - X_i) + v\beta \sum_k \sin(y_{ik} - \alpha - \beta X_i)$$

An estimate of X_i cannot be obtained analytically and may be estimated numerically by a standard iterative numerical procedure. Set $\frac{\partial L}{\partial X_i} = 0$ and suppose $\hat{X}_{i0}$ is an initial estimate of $\hat{X}_i$, then after simplification an equation is approximately given by:

$$\hat{X}_{i1} \approx \hat{X}_{i0} + \frac{\sum_j \sin(x_{ij} - \hat{X}_{i0}) + \frac{\hat{v}}{\hat{\kappa}}\hat{\beta}\sum_k \sin(y_{ik} - \hat{\alpha} - \hat{\beta}\hat{X}_{i0})}{\sum_j \cos(x_{ij} - \hat{X}_{i0}) + \frac{\hat{v}}{\hat{\kappa}}\hat{\beta}^2\sum_k \cos(y_{ik} - \hat{\alpha} - \hat{\beta}\hat{X}_{i0})} \tag{2}$$

where, $\hat{X}_{i1}$ is an improvement estimate of $\hat{X}_{i0}$.

The first partial derivative with respect to β is

$$\frac{\partial L}{\partial \beta} = \sum \sum vX_i \sin(y_{ik} - \alpha - \beta X_i)$$

As with $\hat{X}_i$, $\hat{\beta}$ may be obtained iteratively. Suppose $\hat{\beta}_0$ is an initial estimate of $\hat{\beta}$, then we have the equation approximately given by:

$$\hat{\beta}_1 \approx \hat{\beta}_0 + \frac{\sum \sum X_i \sin(y_{ik} - \hat{\alpha} - \hat{\beta}\hat{X}_i)}{\sum \sum X_i \cos(y_{ik} - \hat{\alpha} - \hat{\beta}\hat{X}_i)}, \tag{3}$$

where, $\hat{\beta}_1$ is an improvement estimate of $\hat{\beta}_0$.
The first partial derivative with respect to κ is

$$\frac{\partial L}{\partial \kappa} = -N\frac{I_0'(\kappa)}{I_0(\kappa)} + \sum \sum \cos(x_{ij} - X_i)$$

and with respect to v is:

$$\frac{\partial L}{\partial v} = -M\frac{I_0'(v)}{I_0(v)} + \sum \sum \cos(y_{ik} - \hat{\alpha} - \hat{\beta}\hat{X}).$$

It can be shown[3], that the estimates of κ and v are given by:

$$\hat{\kappa} = A^{-1}\left(\frac{1}{N}\sum \sum \cos(x_{ij} - \hat{X}_i)\right) \text{ and } \hat{v} = A^{-1}\left(\frac{1}{M}\sum \sum \cos(y_{ik} - \hat{\alpha} - \hat{\beta}\hat{X})\right),$$

respectively.

Hence $\hat{\alpha}, \hat{\beta}, \hat{\kappa}, \hat{v}, \hat{X}_1, \ldots \hat{X}_p$ can be solved iteratively and possible initial estimates for the iteration are putting $\hat{\beta}_0 = 1.0$ in equation (3) and $\frac{\hat{v}}{\hat{\kappa}} = 1.0$ in equation (2). An initial estimate of X_i in equation (2) can be chosen from the mean direction of x_{ij}, that is:

$$\hat{X}_{i0} = \begin{cases} \tan^{-1}\left(\dfrac{S_i}{C_i}\right), & S>0, C_i>0 \\ \tan^{-1}\left(\dfrac{S_i}{C_i}\right) + \pi, & C_i<0 \\ \tan^{-1}\left(\dfrac{S_i}{C_i}\right) + 2\pi, & S<0, C_i>0 \end{cases}$$

where, $S_i = \sum_{j=1}^{m_i} \sin(x_{ij})$ and $C_i = \sum_{j=1}^{m_i} \cos(x_{ij})$. Further, the estimates of v and κ can be obtain by using the approximation[4], that is:

$$A^{-1}(w) \approx \frac{9 - 8w + 3w^2}{8(1-w)}.$$

Thus for replicated circular functional relationship, when the errors are distributed as a von Mises, all the parameters can be estimated.

In considering the above model, the asymptotic variances of estimators, i.e. $\hat{\alpha}, \hat{\beta}, \hat{\kappa}, \hat{v}, \hat{X}_1, \ldots \hat{X}_p$, of the replicated linear circular functional model can be obtained by inverting the estimated Fisher information matrix. Our main interest is the asymptotic covariance matrix of $\hat{\kappa}, \hat{v}, \hat{\alpha}$ and $\hat{\beta}$. The estimated Fisher information matrix, for $\hat{\kappa}, \hat{v}, \hat{\alpha}$ and $\hat{\beta}$. denoted by G, is given by:

$$G = \begin{bmatrix} (NA'(\hat{k}))^{-1} & 0 & 0 & 0 \\ 0 & (MA'(\hat{v}))^{-1} & 0 & 0 \\ 0 & 0 & H\dfrac{\sum \hat{X}_i^2}{P} & -H\dfrac{\sum \hat{X}_i}{P} \\ 0 & 0 & -H\dfrac{\sum \hat{X}_i}{P} & H1 \end{bmatrix},$$

where:

$$H = \frac{p^2(\hat{\kappa}NA(\hat{\kappa}) + \hat{v}\hat{\beta}^2 MA(\hat{v})}{\hat{v}MA(\hat{v})\hat{\kappa}NA(\hat{\kappa})\left(p\sum \hat{X}_i^2 - (\sum \hat{X}_i)^2\right)}.$$

Therefore, we have the estimated variances of $\hat{\kappa}, \hat{v}, \hat{\alpha}$ and $\hat{\beta}$ given by:

$$\hat{Var}(\hat{\kappa}) = \frac{\hat{\kappa}}{N\left((\hat{\kappa} - \hat{\kappa}A^2(\hat{\kappa}) - A(\hat{\kappa})\right)},$$

$$\hat{Var}(\hat{v}) = \frac{\hat{v}}{M\left((\hat{v} - \hat{v}A^2(\hat{v}) - A(\hat{v})\right)},$$

$$\hat{Var}(\hat{\alpha}) = \frac{p(\hat{\kappa}NA(\hat{\kappa}) + \hat{v}\hat{\beta}^2 MA(\hat{v})\sum \hat{X}_i^2}{\hat{v}MA(\hat{v})\hat{\kappa}NA(\hat{\kappa})\left(p\sum \hat{X}_i^2 - (\sum X_i)^2\right)} \text{ and}$$

$$\hat{Var}(\hat{\beta}) = \frac{p(\hat{\kappa}NA(\hat{\kappa}) + \hat{v}\hat{\beta}^2 MA(\hat{v})}{\hat{v}MA(\hat{v})\hat{\kappa}NA(\hat{\kappa})\left(p\sum \hat{X}_i^2 - (\sum X_i)^2\right)}$$

APPLICATION OF THE MODEL

It is assumed that the replication can be made available based on the time each observation was measured and we estimate the parameters using the maximum likelihood method. For the wind direction data, the times at which data were recorded for radar and anchored buoy are at hourly intervals (usually) from approximately 1.6 to 22.6 days and for the wave direction data times recorded are also at hourly intervals (usually) from approximately 1.7 to 79.7 days. Noting that in a replicated linear circular functional relationship model, an observation from circular variable X is not necessarily paired with any observation from circular variable Y in the same group, the different recording times for x_i and y_i do not matter in this case.

Suppose $T(x_i)$ and $T(y_i)$ be the recorded times for observations x_i and y_i of variables X and Y, respectively. To obtain the replication we divide the data into groups of two consecutive roughly hourly measures, except where $T(x_i)-T(X_{i+1})$, (and/or $T(y_i)-T(y_{i+1})$) is too big (more than 1 hour, say), in which case we keep values separate. As an example for wind direction data, the recorded time difference between $(T(x_1), T(y_1))$ and $(T(x_2), T(y_2))$ was 1 hour, thus take $(T(x_1), T(y_1))$ and $(T(x_2), T(y_2))$ together as our first group. Similarly for $(T(x_3), T(y_3))$ and $(T(x_4), T(y_4))$ as second group, as well as $(T(x_5), T(y_5))$ and $(T(x_6), T(y_6))$ as our third group. However the recorded time difference between observations $(T(x_7), T(y_7))$ and $(T(x_8), T(y_8))$ was more than 13 h, hence observation $(T(x_7), T(y_7))$ alone is the fourth group. Further we consider observations $(T(x_8), T(y_8))$ and $(T(x_9), T(y_9))$. This procedure will be continued until all observations have been assigned. Based on the above procedure we have 71 groups in which 12 have a single observation for the wind direction data and 50 groups in which 20 have a single observation for the wave direction data.

Recall that the model for replicated data that was proposed is given by:

$$x_{ij} = X_i + \delta_{ij} \text{ and } y_{ik} = Y_i + \varepsilon_{ik}$$

where, $Y_i = \alpha + \beta X_i \pmod{2\pi}$, for $i = 1, 2,..., p$, $j = 1, 2,..., m_i$ and $k = 1, 2,..., m_i$ and $k = 1, 2,..., m_i$ and $k = 1, 2,..., n_i$ and also δ_{ij}~$VM(0,\kappa)$ and ε_{ik}~$VM(0, v)$ where, x_{ij} is the measurement for group i by radar with some random error δ_{ij} and y_{ik} is the measurement for group i by an anchored buoy with some random error ε_{ik}. X_i and Y_i are said to be the underlying or real directions measured by radar and anchored buoy, respectively.

Analysis of wind direction data: The estimates of parameters for the wind direction data together with their standard errors (Table 1). The estimated ratio of error concentration parameters is given by $\hat{\lambda} = \frac{\hat{v}}{\hat{\kappa}} = 0.87$

Table 1: Parameter estimates for the wind direction data

Parameter	Estimate	Standard error
α	3.208	0.321
β	0.988	9.247×10^{-2}
κ	1.454	0.141
v	1.259	0.113

Table 2: Parameter estimates for the wind direction data assuming β =1.0

Parameter	Estimate	Standard error
α	3.208	0.323
κ	1.453	0.141
v	1.259	0.113

Table 3: Parameter estimates for the wave direction data

Parameter	Estimate	Standard error
α	3.827	0.442
β	1.039	0.113
κ	1.451	0.181
v	1.361	0.165

Table 4: Parameter estimates for the wave direction data assuming β =1.0

Parameter	Estimate	Standard error
α	3.827	0.435
κ	1.451	0.181
v	1.361	0.163

However, the 95% confidence interval for $\hat{\beta}$ is given by (0.807, 1.169), which suggests that $\beta = 1.0$ is acceptable. Re-estimating the parameters assuming β equal to 1.0, we obtained the estimates as shown in Table 2.

Analysis of wave direction data: The maximum likelihood estimates and standard errors are given in Table 3. The estimated ratio of error concentration parameters is given by $\hat{\lambda} = \frac{\hat{v}}{\hat{\kappa}} = 0.94$. However, the 95% confidence interval for $\hat{\beta}$ is given by (0.818, 1.260), which suggests that $\beta = 1.0$ is a reasonable value, as in the wind direction data. Re-estimating the parameters by assuming β equal to 1.0, we obtain the estimates as shown in Table 4.

It was found that at the 5% significance level, there is no difference from 1.0 in the estimates of the slope parameter, β, for either the wind directions or the wave directions. We found that there is a non-zero intercept for both data sets, i.e $\hat{\alpha} = 3.208$ and 3.827 , respectively. This suggests that there is almost no difference in the relative calibration between the measurements by radar and anchored buoy for both data sets but that an additive correction is required to move from one measurement method to the other.

In both cases it was found that the ratio of error concentration parameters between anchored buoy and radar, λ, is less than 1.0 and also the estimated standard error for error concentration parameters of measurements by anchored buoy is less than the estimated standard error for error concentration parameters by radar which suggest that measurements by anchored buoy seems to be more precise and the ratio of the error concentration parameters which is less than 1.0 in both cases.

CONCLUSION

The motivation of this research is to propose a solution to the practical question of how to look at the relationship between the two circular variables when we have only unreplicated data and in particular how to estimate the slope parameter, β, when both circular variables were measured with error and we have no idea about the ratio of their error concentration parameters. For this two sets of circular data were given (wind direction data and wave direction data), each of which consists of measurements of directions by two different techniques, HF radar and anchored buoy.

This study proposed the linear circular functional relationship model, which is an analogy of the linear functional relationship model for continuous linear variables. We began by proposing the replicated circular functional relationship model, a model where we have a replication or repeated measurements for circular variables. The estimators of parameters by using maximum likelihood estimation are also given. Estimates have been obtained numerically by an iterative method given suitable initial estimates of β i.e. $\hat{\beta}_0$, the ratio of error concentration parameters, $\frac{\hat{v}}{\hat{\kappa}}$ and X_i, i.e. $\hat{X}_{i0}$. In the analysis of the wind and wave direction data we chose $\beta_0 = 1.0$, $\frac{\hat{v}}{\hat{\kappa}} = 1.0$ and initial estimate of X_i from the mean direction of x_{ij}.

Finally, in this study the pseudo-replicates from unreplicated circular data based on the measurements were recorded are obtained and estimates the parameters using the maximum likelihood method. The results shows that the estimates of α, β and λ obtained for wind direction data are very close to the estimates obtained for the wave direction data. This suggest that the radar and anchored buoy technique give a similar results in the measurements of the directions of wind and wave, even though we found that measurement made by radar is less precise than measurements made by anchored buoy by comparing the estimate of error concentration parameters as well as its standard error.

REFERENCES

1. Hussin, A.G., 2001. An approximation technique of MLE for the unreplicated linear circular functional relationship model. Malaysian J. Sci., 20: 121-126.
2. Solari, M.E., 1969. The maximum likelihood solution of the problem of estimating a linear functional relationship. J.R. Statist. Soc. B, 31: 372-375.
3. Hussin, A.G. and Z. Chik, 2003. On estimating error concentration parameter for circular functional model. Bull. Malaysian Math. Soc., 26: 181-188.
4. Dobson, A.J., 1978. Simple approximation for the von Mises concentration statistics. Applied Stat., 27: 345-347.