

An approximation technique of MLE for the unreplicated linear circular functional relationship model

Abdul Ghapor Hussin

Centre for Foundation Studies in Science, University of Malaya, 50603 Kuala Lumpur, MALAYSIA

ABSTRACT The maximum likelihood estimation (MLE) of parameters for the unreplicated linear circular functional relationship model is discussed in detail. Explanations are given for the difficulty of estimating parameters with no restrictions on the ratio of error concentration parameters. An approximation technique is proposed for the case when the ratio of error concentration parameters is known. The parameter estimates may be obtained iteratively since the closed-form expressions for the maximum likelihood estimates are not available.

ABSTRAK Kaedah penganggaran kebolehjadian maksimum bagi menganggar parameter didalam model hubungan fungsian membulat linear tanpa replika dibincangkan secara terperinci. Penerangan diberikan tentang kesukaran untuk mendapatkan penganggar dimana tiada syarat dikenakan keatas nisbah ralat ketumpatan parameter. Masalah ini boleh diatasi dengan menggunakan kaedah penghampiran disebabkan bentuk tertutup bagi penganggar kebolehjadian maksimum tidak boleh diperolehi.

(maximum likelihood estimation, linear functional relationship model, circular variables, ratio of error concentration parameters, linear circular functional relationship model.)

INTRODUCTION

The problem of estimating the parameters of a *linear functional relationship model* when both variables are observed subject to error has received a good deal of attention. It was shown that the estimation in the linear functional relationship model requires that we estimate the fixed but unobservable true explanatory variables ([1], [2]), most often treating them as nuisance parameters. For maximum likelihood estimation of unreplicated case, consistent estimators of all the parameters in the model do not exist, since the number of parameters increases with increasing sample size n (see Fuller [1]), which most often leads to an unbounded likelihood function. Lindley [2] resolved this difficulty by assuming that the ratio of the error variances is known; particularly, that the error variances are equal.

The *linear circular functional relationship model* refers to the case when both variables are circular, subject to errors and the relationship itself is linear. What we mean by a circular 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 $(0^\circ, 360^\circ)$. An example is wind direction data measured by two

different methods, the anchored wave buoy and HF radar system and shown in Figure 1. We would anticipate an ideal model $y = x$ as appropriate for the data. The scatter plot of one direction measurement against the other (measured in radians anticlockwise from North) gives a cluster of points along the $X=Y$ diagonal and then a few in the top left corner (1° against 359° , etc.). If we think of these data in the context of an ordinary linear regression model, we would regard those points at the top left as outliers, but this is not so, because the measurements are on the circle or circumference, not a straight line. Since 1° is only 2° from 359° , the point $(1^\circ, 359^\circ)$ on the simple scatter plot should not really be far from the ideal model $y = x$. In this respect, the simple scatter plot is misleading but it illustrates that the model which ignores the circularity of the data is equally misleading. Perhaps such scatter plots should be drawn on a torus which maintains the "wrapping" of the measurements scales. This shows the chief problem with ordinary linear regression or linear functional relationship model when applied to circular variables and below we will propose the *linear circular functional relationship model* which is more suited to this form of data.

As an analogy to the linear functional relationship model, we assume both observations for circular variables X and Y , that is radar and anchored buoy in our example data set, are observed with errors. We also assume that the errors are independently distributed and follow the von Mises distribution with mean zero and concentration parameters κ and ν respectively in which a circular random variable θ is said to have a von Mises distribution if its p.d.f is given by

$$g(\theta; \mu_0, \kappa) = \frac{1}{2\pi I_0(\kappa)} \exp\{\kappa \cos(\theta - \mu_0)\},$$

where $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero. The parameter μ_0 is the mean direction while the parameter κ is described as the concentration parameter.

As usual, our main interest is in estimating the intercept (α), slope (β) of a linear functional relationship and the two concentration parameters by using maximum likelihood methods. We use the concentration parameter, κ , to describe the measure of dispersion for circular variable in a von Mises distribution, whereas we use the variance, σ^2 , to describe the measure of dispersion for continuous variables having a

normal distribution. The term $\frac{1}{\kappa}$ describes the spread of the von Mises distribution in the same way as σ^2 describes the spread of the normal distribution. As an analogy to the "ratio of error variances" in the linear functional relationship model, we will use the term "ratio of error concentration parameters" to define the ratio between the two concentration parameters in the linear circular functional relationship model.

In the next section we propose the model for the unreplicated linear circular functional relationship and establish notation. Maximum likelihood estimation of the parameters is discussed. We show that in the full maximum likelihood formulation based on von Mises distributions with unreplicated data, all the parameters are estimable if we know the ratio of the error concentration parameters. Note that, for the linear functional relationship model, we also need to assume knowledge the ratio of error variances in maximum likelihood estimation, if replicates are not available [2].

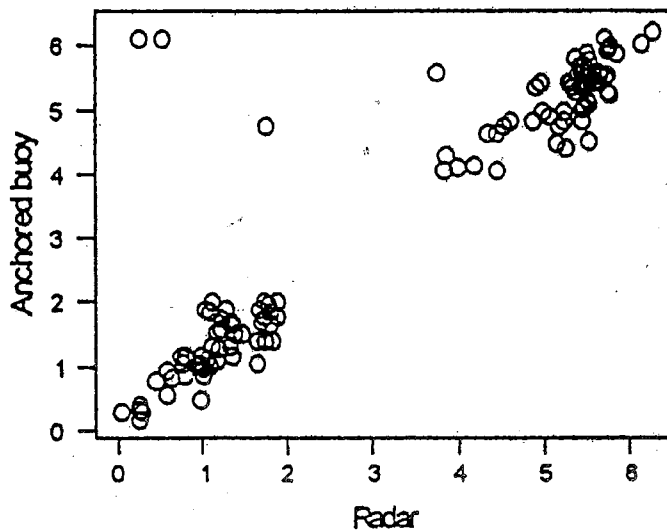


Figure 1. Scatter plot of wind direction data (in radians).

THE UNREPLICATED LINEAR CIRCULAR FUNCTIONAL RELATIONSHIP MODEL

Suppose x_i and y_i are observed values of the circular variables X and Y respectively, thus $0 \leq x_i, y_i < 2\pi$, for $i = 1, \dots, n$. For any fixed X_i , we assume that the observations x_i and y_i have been measured with errors δ_i and ε_i respectively and thus the full model can be written as

$$x_i = X_i + \delta_i \text{ and } y_i = Y_i + \varepsilon_i \text{ where } Y_i = \alpha + \beta X_i \pmod{2\pi}, \text{ for } i = 1, 2, \dots, n \quad (1)$$

We also assumed δ_i and ε_i are independently distributed with (potentially different) von Mises distributions, that is $\delta_i \sim VM(0, \kappa)$ and $\varepsilon_i \sim VM(0, \nu)$. There are $(n+4)$ parameters to be estimated, i.e., $\alpha, \beta, \kappa, \nu$ and the incidental parameters $X_1 \dots X_n$, by maximum likelihood estimation. We also define

$$\frac{\nu}{\kappa} = \lambda, \text{ as a ratio of error concentration}$$

parameters for the circular functional relationship model. We seek an optimal value of β close to 1 since we expect, for example, the values given by the two instruments to be numerically close and a model with a high value of β does not have a practical intuitive interpretation. We also restrict the model to a neighbourhood of $\beta = 1$, though we note that the likelihood for a finite set of points may have a higher maximum at a large value of β well away from the neighbourhood of 1. Clearly as the number of points increases this potential global maximum will be for larger and larger values of β . But restricting attention to the neighbourhood of 1 in the practical situation of a finite number of points ensures that our estimate taken from the local maximum will converge to the true value, thus enabling us to use the familiar results for approximate standard errors etc. In the following section we highlight some problems in maximum likelihood estimation (MLE) of model (1) when no further assumption is made about the error concentration parameters.

MLE WITHOUT RESTRICTION ON THE ERROR CONCENTRATION PARAMETERS

When there is no restriction on the error concentration parameters, κ and ν , the log likelihood function for model (1) is given by

$$\text{Log } L(\alpha, \beta, \kappa, \nu, X_1 \dots X_n; x_1 \dots x_n, y_1, \dots, y_n) = -2n \log(2\pi) - n \log I_0(\kappa) - n \log I_0(\nu) + \kappa \sum \cos(x_i - X_i) + \nu \sum \cos(y_i - \alpha - \beta X_i)$$

Differentiating $\log L$ with respect to parameters $\alpha, \beta, \kappa, \nu$ and X_i , we can obtain $\hat{\alpha}, \hat{\beta}, \hat{\kappa}, \hat{\nu}$, and \hat{X}_i . The first partial derivative of the log likelihood function with respect to α is

$$\frac{\partial \log L}{\partial \alpha} = \sum \sin(y_i - \alpha - \beta X_i)$$

Setting this equal to zero and simplifying we get

$$\sum \sin(y_i - \hat{\beta} \hat{X}_i) \cos \hat{\alpha} - \sum \cos(y_i - \hat{\beta} \hat{X}_i) \sin \hat{\alpha} = 0$$

This gives,

$$\begin{aligned} \tan \hat{\alpha} &= \frac{\sum \sin(y_i - \hat{\beta} \hat{X}_i)}{\sum \cos(y_i - \hat{\beta} \hat{X}_i)} \\ \hat{\alpha} &= \tan^{-1} \left\{ \frac{\sum \sin(y_i - \hat{\beta} \hat{X}_i)}{\sum \cos(y_i - \hat{\beta} \hat{X}_i)} \right\} \\ &= \tan^{-1} \left\{ \frac{S}{C} \right\}, \text{ say.} \end{aligned}$$

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

$$\frac{\partial \log L}{\partial X_i} = \kappa \sin(x_i - X_i) + \nu \beta \sin(y_i - \alpha - \beta X_i)$$

If we set this equal to zero, we may solve iteratively for X_i given some "initial guesses".

Suppose \hat{X}_{i0} is an initial estimate for \hat{X}_i . We write (as suggested by Mardia [3]), $x_i - \hat{X}_i = x_i - \hat{X}_{i0} + \hat{X}_{i0} - \hat{X}_i = (x_i - \hat{X}_{i0}) + \Delta_i$, where $\Delta_i = \hat{X}_{i0} - \hat{X}_i$, and also we have $y_i - \hat{\alpha} - \hat{\beta} \hat{X}_i = (y_i - \hat{\alpha} - \hat{\beta} \hat{X}_{i0}) + \hat{\beta} \Delta_i$.

Hence the above equation becomes,

$$\sin(x_i - \hat{X}_{i0} + \Delta_i) + \frac{\hat{\nu}}{\hat{\kappa}} \hat{\beta} \sin(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_{i0} + \hat{\beta} \Delta_i) = 0$$

For small Δ_i , we have

$$\cos \Delta_i \approx 1, \cos \hat{\beta} \Delta_i \approx 1, \sin \Delta_i \approx \Delta_i \text{ and } \sin \hat{\beta} \Delta_i \approx \hat{\beta} \Delta_i$$

Hence the equation is simplified (approximately) to

$$\hat{X}_{i1} \approx \hat{X}_{i0} + \frac{\sin(x_i - \hat{X}_{i0}) + \frac{\hat{\nu}}{\hat{\kappa}} \hat{\beta} \sin(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_{i0})}{\cos(x_i - \hat{X}_{i0}) + \frac{\hat{\nu}}{\hat{\kappa}} \hat{\beta}^2 \cos(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_{i0})} \quad (2)$$

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

The first partial derivative with respect to β is

$$\frac{\partial \log L}{\partial \beta} = \sum X_i \sin(y_i - \alpha - \beta X_i)$$

$\hat{\beta}$ may also be obtained iteratively. Suppose $\hat{\beta}_0$ is an initial estimate of $\hat{\beta}$. Then $y_i - \hat{\alpha} - \hat{\beta} \hat{X}_i = (y_i - \hat{\alpha} - \hat{\beta}_0 \hat{X}_i) + \Delta \hat{X}_i$, where $\Delta = \hat{\beta}_0 - \hat{\beta}$. For small Δ , setting the partial derivative to zero gives

$$\hat{\beta}_1 \approx \hat{\beta}_0 + \frac{\sum \hat{X}_i \sin(y_i - \hat{\alpha} - \hat{\beta}_0 \hat{X}_i)}{\sum \hat{X}_i^2 \cos(y_i - \hat{\alpha} - \hat{\beta}_0 \hat{X}_i)} \quad (3)$$

where $\hat{\beta}_1$ is an improvement of $\hat{\beta}_0$.

The first partial derivative with respect to κ is

$$\frac{\partial \log L}{\partial \kappa} = -n \frac{I'_0(\kappa)}{I_0(\kappa)} + \sum \cos(x_i - X_i)$$

This give

$$\sum \cos(x_i - \hat{X}_i) = n \frac{I'_0(\hat{\kappa})}{I_0(\hat{\kappa})} = n \frac{I_1(\hat{\kappa})}{I_0(\hat{\kappa})} = nA(\hat{\kappa})$$

where the function A is ratio of the modified Bessel function for the first kind of order one and the first kind of order zero. Thus

$$A(\kappa) = \frac{1}{n} \sum \cos(x_i - \hat{X}_i)$$

Hence

$$\hat{\kappa} = A^{-1}\left(\frac{1}{n} \sum \cos(x_i - \hat{X}_i)\right)$$

The estimate of κ can be obtained by using various simple approximations for $A^{-1}(w)$ given by Dobson [4]. For example, for w near 1, the approximation is

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

Finally, the first partial derivative with respect to ν is $\frac{\partial \log L}{\partial \nu} = \sum \cos(y_i - \alpha - \beta X_i) - nA(\nu)$, and hence

$$\hat{\nu} = A^{-1}\left(\frac{1}{n} \sum \cos(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_i)\right) \quad \nu \text{ may also be}$$

estimated by the above approximations.

Possible initial estimates for iteration are $\hat{X}_{i0} = x_i$ in (2) for $i = 1, \dots, n$ and $\hat{\beta}_0 = 1.0$ in (3). This is sensible since both the X and Y are estimates of the same quantity (i.e. direction), so 1.0 would be a logical initial guess of β , and x_i is a direct measurement of X_i . By using (2) and (3) we can update $\alpha, \beta, X_i, \kappa$ and ν , and proceed iteratively. This iteration procedure will continue until the convergence criterion is satisfied.

However, there is a problem with above iteration procedure. The calculation of \hat{X}_{i1} in (2) to estimate X_i , that is

$$\hat{X}_{i1} \approx \hat{X}_{i0} + \frac{\sin(x_i - \hat{X}_{i0}) + \frac{\hat{\nu}}{\hat{\kappa}} \hat{\beta} \sin(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_{i0})}{\cos(x_i - \hat{X}_{i0}) + \frac{\hat{\nu}}{\hat{\kappa}} \hat{\beta}^2 \cos(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_{i0})}$$

depends on the values of $\hat{\nu}$ and $\hat{\kappa}$, where

$$\hat{\nu} = A^{-1}\left(\frac{1}{n} \sum \cos(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_i)\right)$$

and

$$\hat{\kappa} = A^{-1}\left(\frac{1}{n} \sum \cos(x_i - \hat{X}_i)\right)$$

We found that in our iteration the value of

$$w = \frac{1}{n} \sum \cos(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_i) \quad \text{or} \quad \frac{1}{n} \sum \cos(x_i - \hat{X}_i)$$

is equal, or very close, to 1.0. This has been verified experimentally by using small simulated data sets as well as the wind direction data. For this reason, $\hat{\nu}$ and $\hat{\kappa}$ will be very large and tend to infinity, since the approximation function of $A^{-1}(w)$ is given by

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

which shows that $A^{-1}(w)$ will tend to infinity when w approaches 1.0.

Further, for large $\hat{\nu}$ and $\hat{\kappa}$ we are not be able to compute $I_0(\hat{\nu})$ and $I_0(\hat{\kappa})$ in the log likelihood function, since $I_0(\hat{\nu})$ is given by

$$I_0(\hat{\nu}) = 1 + \frac{\hat{\nu}^2}{2^2} + \frac{\hat{\nu}^4}{2^2 4^2} + \frac{\hat{\nu}^6}{2^2 4^2 6^2} + \dots \quad (4)$$

and a similar definition holds for $I_0(\hat{\kappa})$. Bowman [5] suggests that these series rarely require more than 25 terms, but it is clear that (4), as well as the log likelihood function will tend to infinity, for large value of $\hat{\nu}$ and $\hat{\kappa}$.

It is shown in the next section that this instability problem to estimate the parameters numerically, may be overcome by fixing the ratio of the error

concentration parameters, i.e. assuming it is known. Note that estimation of parameters in the unreplicated linear functional relationship model required us to know the ratio of error variances, and without this additional constraint, the likelihood function is unbounded, that is theoretically the parameters are not estimable [1], [2].

MLE ASSUMING THE RATIO OF ERROR CONCENTRATION PARAMETERS IS KNOWN

Suppose we assume that the ratio of the error concentration parameters, that is $\frac{v}{\kappa} = \lambda$ is known. Then the log likelihood function is given by

$$\begin{aligned} \text{Log}L(\alpha, \beta, \kappa, X_1, \dots, X_m; \lambda, x_1, \dots, x_m, y_1, \dots, y_n) = \\ -2n \log(2\pi) - n \log I_0(\kappa) - n \log I_0(\lambda\kappa) \\ + \kappa \sum \cos(x_i - X_i) + \lambda\kappa \sum \cos(y_i - \alpha - \beta X_i) \end{aligned}$$

Differentiating logL with respect to α, β, κ and X_i we obtain the likelihood equations for the parameters, which may be solved iteratively. The estimates $\hat{\alpha}, \hat{\beta}$ and \hat{X}_i are similar to the estimates given above. The equation for κ may be solved numerically based on the asymptotic properties of the Bessel function.

By setting $\frac{\partial \log L}{\partial \kappa} = 0$, we get the equation

$$A(\kappa) + \lambda A(\kappa) = \frac{1}{n} \left\{ \sum \cos(x_i - \hat{X}_i) + \lambda \sum \cos(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_i) \right\} \quad (5)$$

In the following sections we will consider two cases, first when the ratio of error concentration parameters is equal to one and the case when the ratio is not equal to one.

(a) The case when $\lambda = 1$.

When $\lambda = 1$, equation (5) reduces to

$$A(\kappa) = \frac{1}{2n} \left\{ \sum \cos(x_i - \hat{X}_i) + \sum \cos(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_i) \right\} \quad (6)$$

The approximation given by Dobson [4]

$$A^{-1}(w) = \frac{9 - 8w + 3w^2}{8(1 - w)} \quad (7)$$

to estimate κ in (5) can only be used for this case. Hence, (6) can be solved by using (7) to estimate κ when the ratio of error concentration parameters is equal to one.

(b) The case when $\lambda \neq 1$.

In this section we use the asymptotic properties of the Bessel function so that we can find an estimate of κ for any value of λ . From the asymptotic power series for the Bessel functions $I_0(r)$ and $I_1(r)$, (see Abramowitz & Stegun [6]) we have,

$$A(r) = \frac{I_1(r)}{I_0(r)} = 1 - \frac{1}{2r} - \frac{1}{8r^2} - \frac{1}{8r^3} + O(r^{-4}) \quad (8)$$

Simplifying equation (5) using (8) we have the expression approximately given by

$$8(1 + \lambda - c)\kappa^3 - 8\kappa^2 - (1 + \frac{1}{\lambda})\kappa - (1 + \frac{1}{\lambda^2}) = 0, \quad (9)$$

where $c = \frac{1}{n} \left\{ \sum \cos(x_i - \hat{X}_i) + \lambda \sum \cos(y_i - \hat{\alpha} - \hat{\beta} \hat{X}_i) \right\}$.

It can be shown (see Rade & Westergren [7]) that the above cubic equation in κ , equation (9), has only one positive real root and two complex roots, giving $\hat{\kappa}$ as the positive real root. In this study we solved equation (9) by using a FORTRAN subroutine.

In the above cases, the asymptotic estimate standard deviation of the estimated parameters can be obtained from the inverse of the estimated Fisher information matrix.

NUMERICAL EXAMPLE

The wind direction data (shown in Figure 1) will be used to illustrate the above model. Table 1 below gives a comparison of $\hat{\kappa}$ and its standard error obtained by **Method 1** (approximation given by Dobson [4], i.e. equations (6) and (7)) and **Method 2** (asymptotic power series for Bessel function, i.e. the solution of $\hat{\kappa}$ from equation (9)), in which we assumed the ratio of error concentration parameters is equal to one. From Table 1, we can see that both methods give almost similar estimates of κ and its standard error at three decimal places.

Table 2 gives the estimates of parameters α, β and κ obtained by using various values of the ratio of parameter concentrations, λ . These show that by using the asymptotic power series for Bessel function, i.e. the solution of $\hat{\kappa}$ from equation (9), we can obtain the estimate of parameters for any value of error concentration parameters.

Table 1. Comparison of estimating $\hat{\kappa}$ and its standard error by two different methods.

	Method 1	Method 2
$\hat{\kappa}$	22.9842	22.9839
s.e. ($\hat{\kappa}$)	2.0014	2.0020

Table 2. Estimate of parameters for various value of λ .

λ	$\hat{\alpha}$ (s.d. of $\hat{\alpha}$)	$\hat{\beta}$ (s.d. of $\hat{\beta}$)	$\hat{\kappa}$ (s.d. of $\hat{\kappa}$)
0.5	0.1337 (4.8466x10 ⁻²)	0.9827 (1.1939x10 ⁻²)	37.0550 (3.2301)
1.0	0.1270 (5.0392x10 ⁻²)	0.9881 (1.2417x10 ⁻²)	22.9839 (2.0020)
1.5	0.1140 (5.0267x10 ⁻²)	0.9891 (1.2235x10 ⁻²)	19.8573 (1.7303)
2.0	0.1212 (4.9391x10 ⁻²)	0.9867 (1.2009x10 ⁻²)	18.4175 (1.6054)

CONCLUSION

In this paper we present a linear circular functional relationship model, which is a statistical method for fitting a straight-line relationship when both circular variables are subject to errors. The model that we proposed is analogous to the linear functional relationship model for continuous variables, the so called errors-in-variables model, which has been discussed by many authors. In this model we assumed the circular random errors had von Mises distributions. The maximum likelihood estimates have been obtain numerically by an iterative method, not analytically. However the iterative method fails to give estimates of parameters because of an instability problem in estimation of κ and ν , and we found that this instability problem can be overcome by assuming the ratio of concentration parameters, i.e., $\lambda = \frac{\nu}{\kappa}$ is known. The approximation of asymptotic power series for Bessel function have been proposed to obtained these estimates.

We give estimates of the parameters for various values of λ (Table 2), showing that in the full maximum likelihood formulation of the unreplicated linear circular functional relationship model based on a von Mises

distribution all the parameters are estimable if we know the ratio of the error concentration parameters. This is a similar condition to that which holds for the unreplicated linear functional relationship model, in which we have to know the ratio of error variances.

REFERENCES

1. Fuller, W. A. (1987). *Measurement Error Models*. John Wiley, New York.
2. Lindley, V. D. (1947). Regression lines and the linear functional relationship. *J. R. Statist. Soc. Supp.*, 9:218-244.
3. Mardia, K. V. (1972). *Statistics of Directional Data*. Academic Press, London.
4. Dobson, A. J. (1978). Simple approximations for the von Mises concentration statistic. *Applied Statistics*, 27: 345-347.
5. Bowman, F. (1958). *Introduction to Bessel functions*. Dover Publications Inc., New York.
6. Abramowitz, M. and Stegun, I. G. (1965). *Handbook of mathematical functions*. Dover Publications Inc., New York.
7. Rades, L. and Westergren, B. (1988). *Beta mathematics handbook*. Chartwell-Bratt.