Proceedings of the 7th International Conference on Statistics: Theory and Applications (ICSTA'25)
Paris, France - August, 2025
Paper No. 145
DOI: 10.11159/icsta25.145

Circular-Circular Regression Models for Wind Directions in Thailand

Orathai Polsen, Pianpool Kamoljitprapa
Department of Applied Statistics, Faculty of Applied Science, King Mongkut's University of Technology North Bangkok
1518 Pracharat 1 Road, Wongsawang, Bangsue, Bangkok, Thailand
orathai.p@sci.kmutnb.ac.th; pianpool.k@sci.kmutnb.ac.th

Abstract **-** In the applications of many scientific fields, circular data is collected and analysed. In meteorology, the meteorological data is collected in order to monitor the weather and publicise the natural disaster warnings. Wind direction measured in angles is a circular data and is one of the important factors having numerous impacts on weather patterns, climate change, and quality of the atmospheric environment. The knowledge of wind direction is useful for several sectors, including construction and agriculture. In this paper, the wind directions at Don Mueang station, Thailand, collected by the Thai Meteorological Department were analysed for regressing the wind directions at 1 p.m. on that at 7 a.m. and regressing the wind directions at 4 p.m. on that at 10 a.m. Taylor's models–circular-circular regression model and polynomial circular model–in which the angular error is distributed as a von Mises distribution and a wrapped Cauchy distribution were considered in this study. The findings show that the circular-circular regression model of order one in which the angular error is distributed as a von Mises distribution is the best model for fitting the relationship between the wind directions in this study.

Keywords**:** Circular-Circular Regression, Polynomial Circular Model, Von Mises Distribution, Wind Direction, Wrapped Cauchy Distribution.

1. Introduction

In the field of meteorology, weather data–such as air temperature, relative humidity, air pressure, rainfall, wind speed, wind direction–are collected for monitoring and analysing in order to understand the weather and climate patterns [1]. Wind direction is one of the important factors affecting atmospheric phenomena. Wind direction has an impact on humidity and temperature. In addition, wind direction impacts on air pollutant concentrations and wildfires [2-3]. Accordingly, wind direction has a substantial influence on the lives of human beings. Knowing wind direction is important in a number of ways, including weather forecasting, environmental protection and control, agricultural activities, wind energy operation, construction design, aviation and maritime operations, and some sport activities [4-5].

Wind direction is a circular or directional data, which is different from the linear domain. Circular data is defined as angles θ where $-\pi \le \theta < \pi$ or points on the circumference of a unit circle. A useful method for analysing circular data to construct a relationship model is a circular regression [6]. In the literature, a circular regression has been applied in many diverse fields, for example, in marine biology study, a relationship between spawning time and time of low tide is of interest [7]. In meteorology, it is useful to investigate a relationship between wind directions [8-10]. In medical study, it is often of interest to know a relationship between the peak times for two successive measurements of diastolic blood pressure [11]. In environmental science, it is interesting to know whether the direction of ground movement is related to the direction of steepest descent [12]. The circular regression in which both covariate and response are circular variables is called a circular-circular regression.

In the literature, there are various studies on a circular-circular regression in attempt to introduce a model and the models were widely utilised in many areas. Taylor [13] proposed a four-parameter model in which the angular error is distributed as a von Mises distribution and studied some properties of the model including parameter estimation. The extension of model, a polynomial regression model and a multiple regression model were also introduced. Moreover, the models were applied to protein structure data. Polsen and Taylor [10] examined the properties of the existing circular-circular regression models including Taylor's model in order to know the similarities and the differences. The models considered in the study can be written as a general form of the tangent link function of two trigonometric polynomial functions, however, the model error distributions are varied. Furthermore, the study showed that Taylor's model has the desired properties which the others do not have. In addition, Taylor's model is easy to handle and comparatively simple to execute. Polsen [14] introduced the modified version of Taylor's model in which the angular error is assumed to follow a wrapped Cauchy distribution and extended to a polynomial circular model. The parameter estimations were presented and the performances of estimators were also examined using the simulation study. In addition, the application to wind direction showed that the models represent the relationship reasonably well.

This paper aims to construct the circular-circular regression models for wind directions in Thailand using Taylor's models. The model accuracy is examined based on common criteria for determining the best-fitting model. The remaining sections of the paper are organised as follows. In section 2, the datasets, the circular-circular regression models, and parameter estimation are presented. The study results are given in Section 3. Finally, Section 4 provides discussion and conclusion.

2. Materials and Models

2.1. Study Area and Datasets

Don Mueang meteorological station located in the northern area of Bangkok, Thailand is considered in this study. This station is one of the stations that monitors the weather and climate, and collects the meteorological data in every time interval as defined by the World Meteorological Organization (WMO). The data is used not only for weather and climate analysis and warning but also provided to other organizations. The wind directions at 7 a.m., 10 a.m., 1 p.m., and 4 p.m. measured each day at Don Mueang station for 184 consecutive days during the period from July to December 2024 are considered in this study. The data was obtained from the data-service system of Thai Meteorological Department (TMD).

2.2. Circular-Circular Regression Models

Circular-circular regression is a powerful statistical method for investigating and modelling the relationship of the bivariate circular data. This approach has existed for a long time and applied in many fields. A range of circular-circular regression models has been introduced and investigated in the available literature, including Taylor's model.

Let Y be a circular response variable and X be a circular explanatory variable, where both X and Y take values on the unit circle, represented as the value in the interval $[-\pi,\pi)$ or as the real numbers mod 2π. Taylor's model is given by

$$\begin{aligned} y_i &= \mu(x_i;\alpha,\beta,a,b)+\varepsilon_i, \quad i=1,...,n \\ &= \beta+\operatorname{atan2}(a\sin(x_i-\alpha), b\cos(x_i-\alpha)+1)+\varepsilon_i, \end{aligned} \tag{1}$$

where a and b are real numbers, α and β are angular location parameters. The mean function is centred on (α,β), ε_i is the angular error, and the function $\operatorname{atan2}(v,u)$ gives the angle between the positive x-axis and the vector from the original to the point (u,v). Graphs of Taylor's model in Eq. (1) for various values of parameters are presented in Fig. 1. (a) - (b).

A polynomial circular model of order k is an extension of the model in Eq. (1) proposed by Taylor [13]. The model is given by

$$\begin{aligned} y_i &= \mu(x_i;\alpha,\beta,a_1,...,a_k,b)+\varepsilon_i, \quad i=1,...,n \\ &= \beta+\operatorname{atan2}(a_1\sin(x_i-\alpha)+\cdots+a_k\sin^k(x_i-\alpha), b\cos(x_i-\alpha)+k)+\varepsilon_i, \end{aligned} \tag{2}$$

where $a_1,...,a_k$ and b are real numbers. Fig. 1. (c) displays examples of the model in Eq. (2) for selected values of parameters.

In the work of Taylor [13], the angular error is distributed as a von Mises distribution shown in Eq. (3) while the error is assumed to be a wrapped Cauchy distribution in the study of Polsen [14] as in Eq. (4). The probability density function of a von Mises distribution is given by

$$f(\theta;\mu,\kappa)=\frac{1}{2\pi I_0(\kappa)}e^{\kappa\cos(\theta-\mu)},\quad -\pi\le\theta<\pi, \tag{3}$$

where $-\pi\le\mu<\pi$ is a location parameter and is the mean direction, $\kappa\ge 0$ is a scale parameter which measures the concentration, and $I_0(\kappa)$ is the modified Bessel function of the first kind and order zero. The probability density function of a wrapped Cauchy distribution is given by

$$f(\theta;\mu,\rho)=\frac{1}{2\pi}\frac{1-\rho^2}{1+\rho^2-2\rho\cos(\theta-\mu)},\quad -\pi\le\theta<\pi, \tag{4}$$

where $-\pi\le\mu<\pi$ is a location parameter and $0\le\rho<1$ is a scale parameter which controls the concentration.

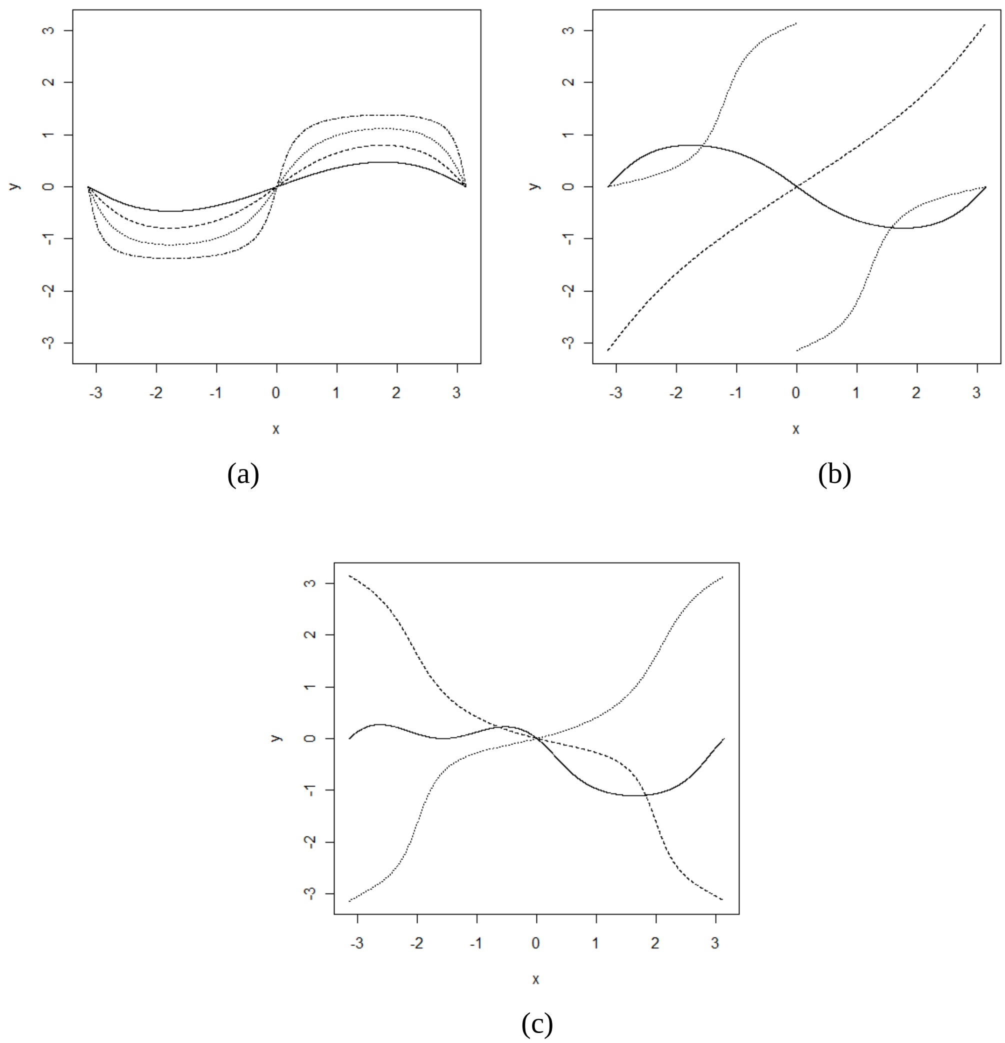


Fig. 1: Graphs of Taylor's model: (a) $a=0.5$ and $b=0.2$ (solid line), $a=1$ and $b=0.2$ (dashed line), $a=2$ and $b=0.2$ (dotted line), $a=5$ and $b=0.2$ (dot-dashed line), (b) $a=-1$ and $b=0.2$ (solid line), $a=3$ and $b=3$ (dashed line), $a=-1$ and $b=-3$ (dotted line), and (c) $a_1=-2$, $a_2=-2$, and $b=0.2$ (solid line), $a_1=-2$, $a_2=0.5$, and $b=5$ (dashed line), $a_1=2$, $a_2=0.5$, and $b=5$ (dotted line).

2.3. Parameter Estimation

Let $\boldsymbol{\varphi}$ be a vector of model parameters and can be estimated by the method of maximum likelihood. In case of the model in Eq. (1) when the angular error is distributed as a von Mises distribution with zero mean and concentration $\kappa,$ the log-likelihood function can be expressed as

$$\log L(\boldsymbol{\varphi}) = -n\log I_0(\kappa) + \kappa\sum_{i=1}^{n}\frac{\cos(y_i-\beta)(b\cos(x_i-\alpha)+1)+a\sin(y_i-\beta)\sin(x_i-\alpha)}{v_i} + const. \tag{5}$$

As the error of the model in Eq. (1) is assumed to be a wrapped Cauchy distribution with zero mean and scale parameter $\rho,$ the log-likelihood function is given by

$$\log L(\boldsymbol{\varphi}) = \sum_{i=1}^{n}\log\left\{\frac{1-\rho^2}{1+\rho^2-2\rho\dfrac{\cos(y_i-\beta)(b\cos(x_i-\alpha)+1)+a\sin(y_i-\beta)\sin(x_i-\alpha)}{v_i}}\right\} + const, \tag{6}$$

where $v_i = \sqrt{a^2\sin^2(x_i-\alpha)+(b\cos(x_i-\alpha)+1)^2}.$

For a polynomial circular model in Eq. (2) when the error is distributed as a von Mises distribution with zero mean and concentration $\kappa,$ the log-likelihood function can be written as

$$\log L(\boldsymbol{\varphi}) = -n\log I_0(\kappa) + \kappa\sum_{i=1}^{n}\{\cos(y_i-\beta)(b\cos(x_i-\alpha)+1)+\sin(y_i-\beta) (a_1\sin(x_i-\alpha)+\cdots+a_k\sin^k(x_i-\alpha))\}/w_i + const. \tag{7}$$

As the error of the model in Eq. (2) is assumed to be a wrapped Cauchy distribution with zero mean and scale parameter $\rho,$ the log-likelihood function is given by

$$\log L(\boldsymbol{\varphi}) = \sum_{i=1}^{n}\log\left\{\frac{1-\rho^2}{1+\rho^2-2\rho\dfrac{\cos(y_i-\beta)(b\cos(x_i-\alpha)+1)+B}{w_i}}\right\} + const, \tag{8}$$

where $w_i = \sqrt{(a_1\sin(x_i-\alpha)+\cdots+a_k\sin^k(x_i-\alpha))^2+(b\cos(x_i-\alpha)+k)^2},$

$B = \sin(y_i-\beta)(a_1\sin(x_i-\alpha)+\cdots+a_k\sin^k(x_i-\alpha)),$

and *const* represents the constant term that does not depend on parameters.

In order to obtain the model estimates, taking the derivative of the log-likelihood functions in Eqs. (5) - (8) with respect to the unknown parameters and equating the derivatives to the zero. Nevertheless, the closed-form solutions for the estimates cannot be solved analytically. Therefore, the parameter estimation must be done by iterative numerical methods. The numerical estimates can be acquired using the optimisation functions in R called *nlm* and *optim*. Remark that, these recursive functions can sometimes yield a locally optimal solution rather than a globally optimal solution. A simple way is to try several initial values for the parameters. An alternative strategy that appears to work well in practice was studied and proposed by Polsen and Taylor [10].

3. Results

The wind directions at 7 a.m., 10 a.m., 1 p.m., and 4 p.m. measured in degrees $(0°, 360°)$ at Don Mueang station were first converted to the values in the interval $[-\pi, \pi)$. The models in Eq. (1) and Eq. (2) where the angular error is distributed as a von Mises distribution and a wrapped Cauchy distribution were applied for regressing the wind directions at 1 p.m. on that at 7 a.m. and regressing the wind direction at 4 p.m. on that at 10 a.m. In addition, how well the models fit were evaluated and compared using the common criteria for model selection called Akaike information criterion (AIC) and Bayesian information criterion (BIC). Furthermore, the residual analysis using graphical tools was used to evaluate the validity and adequacy of the model. The analyses involved in this paper were carried out using the statistical software, R programming language [15].

Consider the case of regressing the wind direction at 1 p.m. on that at 7 a.m., the maximum likelihood estimates, the maximum log-likelihood, AIC, and BIC values for Taylor's models where the angular error is distributed as a von Mises distribution and a wrapped Cauchy distribution are given in Table 1 and Table 2 respectively.

Table 1: Maximum likelihood estimates (and SEs), maximum log-likelihood, AIC, and BIC of Taylor's models (von Mises distribution) for wind directions at 7 a.m. and 1 p.m.

k	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\kappa}$	LL	AIC	BIC
1	-5.565	-	-1.260	1.825	-0.655	1.481	-266.284	542.568	558.643
	(2.238)		(0.423)	(0.095)	(0.110)	(0.147)			
2	-11.941	3.870	-2.556	1.822	-0.702	1.424	-266.192	544.384	563.674
	(5.710)	(7.953)	(0.914)	(0.094)	(0.139)	(0.144)			

Table 2: Maximum likelihood estimates (and SEs), maximum log-likelihood, AIC, and BIC of Taylor's models (wrapped Cauchy distribution) for wind directions at 7 a.m. and 1 p.m.

k	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	LL	AIC	BIC
1	-5.779	-	-1.404	1.799	-0.645	0.542	-267.332	544.664	560.739
	(2.055)		(0.333)	(0.066)	(0.099)	(0.035)			
2	-11.660	0.484	-2.808	1.798	-0.651	0.542	-267.329	546.658	565.948
	(4.429)	(6.555)	(0.671)	(0.068)	(0.126)	(0.035)			

According to the AIC and BIC criteria, Taylor's model of order one $(k = 1)$ has lower values than the model of order two $(k = 2)$ in both cases of the distribution of the angular error. In addition, the model of order one in which the error is distributed as a von Mises distribution provides the best fit and Fig. 2. (a) shows that the fitted regression model represents the relationship between the wind directions at 7 a.m. and 1 p.m. reasonably well and the residual plots in Fig. 2. (b) - (d) indicate that the fitted model is valid.

In case of regressing the wind direction at 4 p.m. on that at 10 a.m., the maximum likelihood estimates, the maximum log-likelihood, AIC, and BIC values for Taylor's models where the angular error is distributed as a von Mises distribution and a wrapped Cauchy distribution are presented in Table 3 and Table 4 respectively.

Table 3: Maximum likelihood estimates (and SEs), maximum log-likelihood, AIC, and BIC of Taylor's models (von Mises distribution) for wind directions at 10 a.m. and 4 p.m.

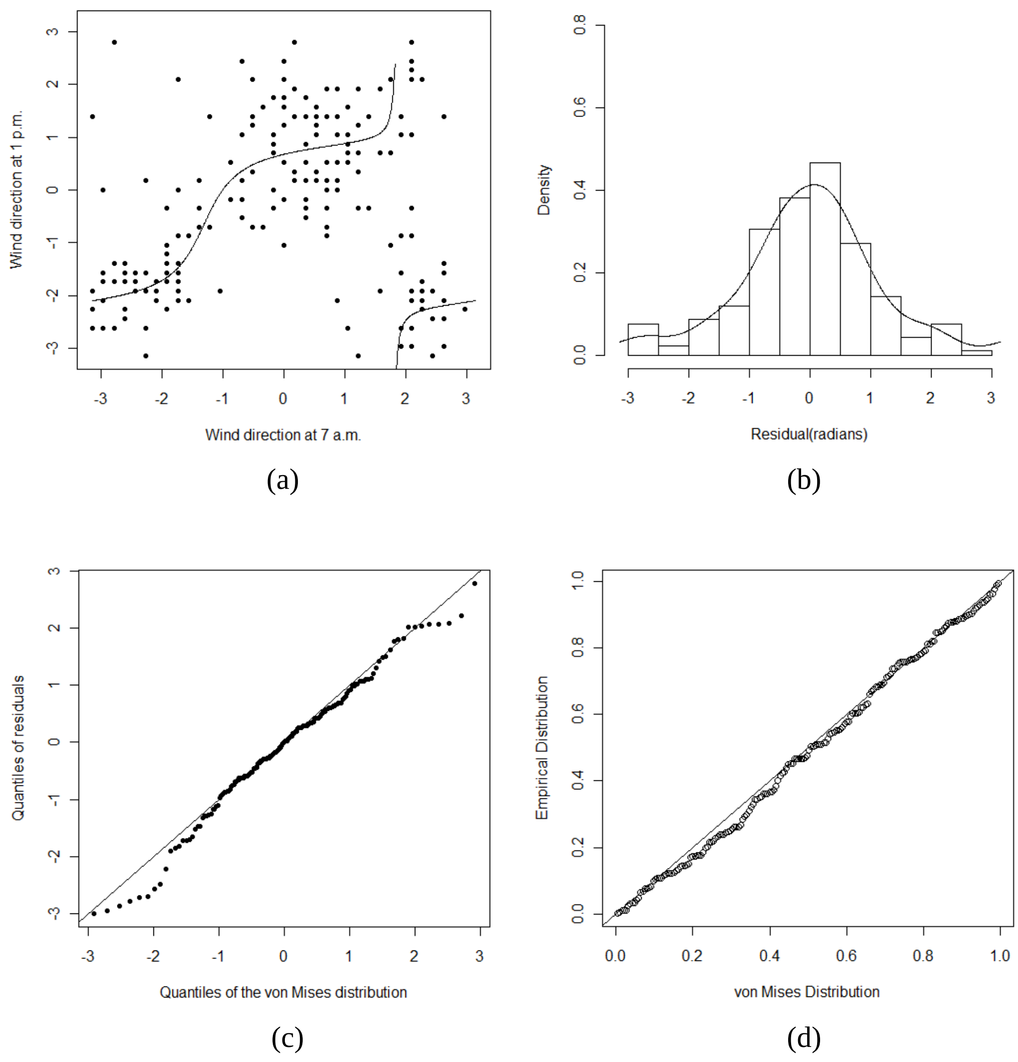
k	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\kappa}$	LL	AIC	BIC
1	-3.438	-	-1.412	2.498	-0.708	1.459	-267.747	545.494	561.569
	(1.119)		(0.329)	(0.095)	(0.123)	(0.146)			
2	-8.701	5.150	-3.142	2.487	-0.853	1.363	-266.597	545.194	564.484

	(4.211)	(5.087)	(1.043)	(0.095)	(0.156)	(0.141)			

Table 4. Maximum likelihood estimates (and SEs), maximum log-likelihood, AIC, and BIC of Taylor's models (wrapped Cauchy distribution) for wind directions at 10 a.m. and 4 p.m.

k	$\hat{a}_1$	$\hat{a}_2$	$\hat{b}$	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\rho}$	LL	AIC	BIC
1	-3.404	-	-1.332	2.450	-0.690	0.532	-270.919	551.838	567.913
	(0.856)		(0.213)	(0.045)	(0.091)	(0.036)			
2	-8.677	4.702	-2.838	2.434	-0.834	0.534	-269.782	551.564	570.854
	(3.480)	(4.213)	(0.639)	(0.044)	(0.130)	(0.036)			

Taylor's model of order one provides a slightly better fit than the one of order two in both cases of the distribution of the angular error according to the BIC criterion. Moreover, Taylor's model of order one in which the angular error is assumed to follow a von Mises distribution is the best-fitting model. Fig. 3. (a) displays that the estimated regression curve gives a good fit for the wind directions at 10 a.m. and 4 p.m. and the residual plots in Fig. 3. (b) - (d) show that the estimated curve is adequate.



(a) (b) (c) (d)

Fig. 2: (a) the fitted regression curve for wind directions at 7 a.m. and 1 p.m., (b) histogram of the residuals and the kernel density estimate, (c) Q-Q plot, and (d) P-P plot.

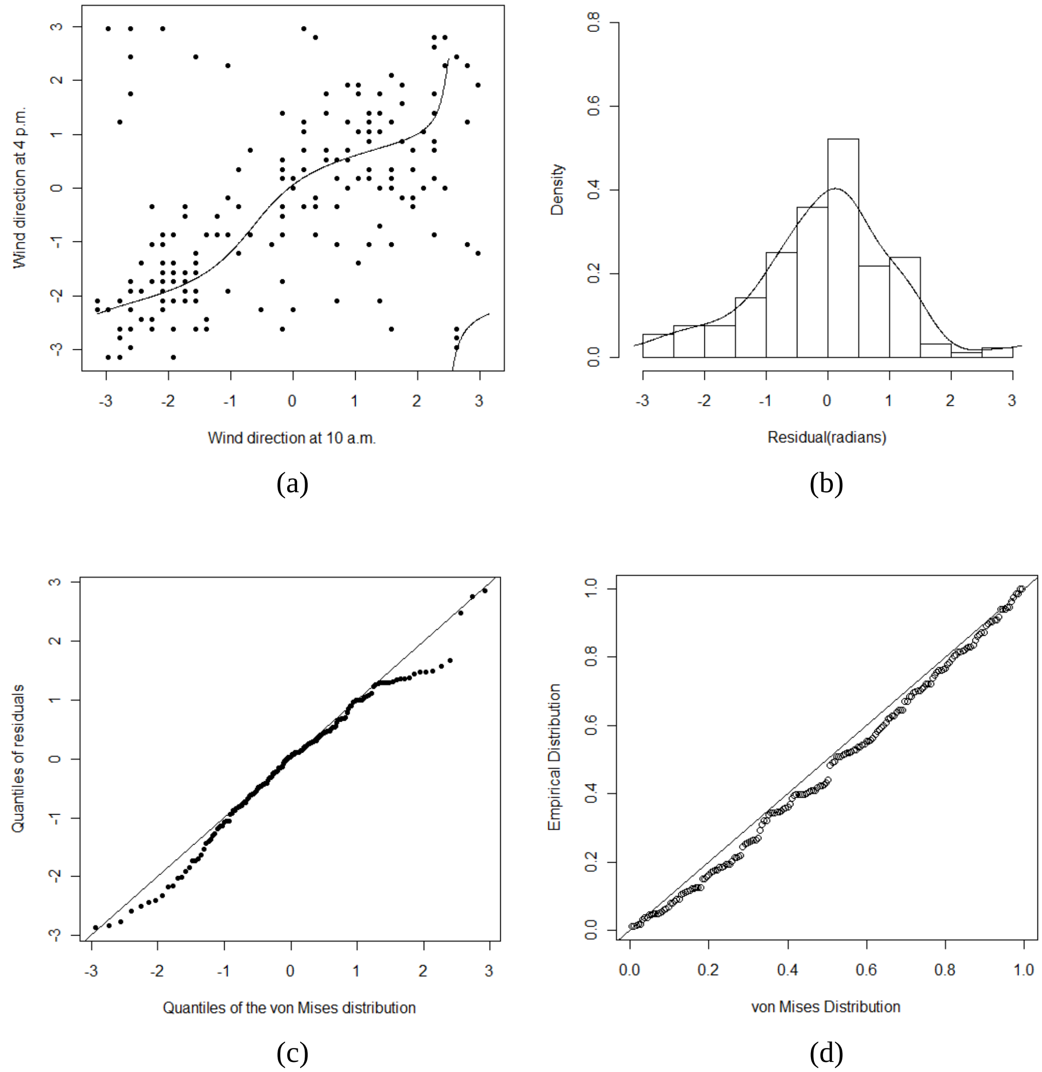


Fig. 3: (a) the fitted regression curve for wind directions at 10 a.m. and 4 p.m., (b) histogram of the residuals and the kernel density estimate, (c) Q-Q plot, and (d) P-P plot.

4. Conclusion

In this paper, the wind directions measured at Don Mueang meteorological station by the Thai Meteorological Department were investigated for regressing the wind directions at 1 p.m. on that at 7 a.m. and regressing the wind direction at 4 p.m. on that at 10 a.m. The circular-circular regression and polynomial circular models of Taylor in which the angular error is distributed as a von Mises distribution and a wrapped Cauchy distribution were applied to obtain the estimated regression models. In this study, the statistical computing was carried out with the help of the free software environment R. The analyses show that Taylor's model of order one ($k=1$) provides the best fit for the relationship between the wind directions considered in this study. In addition, the findings indicate that the models in which the angular error is distributed as a von Mises distribution and a wrapped Cauchy distribution provide a similar fit and are comparable. Furthermore, the models can represent the relationship between the wind directions reasonably well.

For further research, the extension of the study to wind directions of other stations and other circular data could be interesting. The multiple circular regression could also be possible topic for research in this area of application.

Acknowledgements

We are grateful to the Faculty of Applied Science, King Mongkut's University of Technology North Bangkok for funding and resources on this research (fiscal year 2025, contract no. 680038).

References

[1] Thai Meteorological Department. (2025, February 18). [Online]. Available: https://www.tmd.go.th

[2] X. Li, W. Liu, Z. Chen, G. Zeng, C. Hu, T. León, J. Liang, G. Huang, Z. Gao, Z. Li, W. Yan, X. He, M. Lai, and Y. He, "The application of semicircular-buffer-based land use regression models incorporating wind direction in predicting quarterly NO2 and PM10 concentrations," *Atmospheric Environment*, vol. 103, pp. 18-24, 2015. doi: 10.1016/j.atmosenv.2014.12.004

[3] A. Harvey, S. Hurn, D. Palumbo, and S. Thiele, "Modelling circular time series," *Journal of Econometrics*, vol. 239, pp. 1-15, 2024.

[4] N. Masseran, A. M. Razali, K. Ibrahim, and M. T. Latif, "Fitting a mixture of von Mises distributions in order to model data on wind direction in Peninsular Malaysia," *Energy Conversion and Management*, vol. 72, pp. 94-102, 2013.

[5] Arabia Weather. (2025, February 18). What is the benefit of knowing the wind direction? [Online]. Available: https://www.arabiaweather.com/en/content/what-is-the-benefit-of-knowing-the-wind-direction

[6] K. V. Mardia and P. E. Jupp, *Directional Statistics*. John Wiley & Son, Chichester, 2002.

[7] U. Lund, "Least circular distance regression for directional data," *Journal of Applied Statistics*, vol. 26, no. 6, pp. 723-733, 1999.

[8] S. Kato and M. C. Jones, "A family of distributions on the circle with links to, and applications arising from Möbius Transformation," *Journal of the American Statistical Association*, vol. 105, no. 489, pp. 249-262, 2010.

[9] S. Kato, K. Shimizu, and G. S. Shieh, "A circular-circular regression model," *Statistica Sinica*, vol. 18, pp. 633-645, 2008.

[10] O. Polsen and C. C. Taylor, "Parametric circular-circular regression and diagnostic analysis," *Geometry Driven Statistics*, John Wiley & Son, London, pp. 115-128, 2015.

[11] T. Downs, "Rotation angular correlation," *In Biorhythms and Human reproduction, Ed. M. Ferin, F. Halberg and L. van der Wiele*, John Wiley & Son, New York, 1974.

[12] L. P. Rivest, "A decentred predictor for circular-circular regression," *Biometrika*, vol. 84, no. 3, pp. 717-726, 1997.

[13] C. C. Taylor, "Directional data on the torus, with applications to protein structure," in *Proceedings of the SIS 2009 Statistical Conference on Statistical Methods for the Analysis of Large Data-Sets*, pp. 105-108, 2009.

[14] O. Polsen, "Circular-circular regression model, with application to meteorological data," *The Journal of KMUTNB*, vol. 34, no. 2, pp. 1-11, 2024.

[15] R Core Team, "R: A language and environment for statistical computing," *R Foundation for Statistical Computing*, Vienna, Austria, 2022. http://www.R-project.org/