Meteorol Atmos Phys (2018) 130:259–264
https://doi.org/10.1007/s00703-017-0508-y

ORIGINAL PAPER

A note on the correlation between circular and linear variables with an application to wind direction and air temperature data in a Mediterranean climate

M. Lototzis[1] · G. K. Papadopoulos[1] · F. Droulia[2] · A. Tseliou[2] · I. X. Tsiros[2]

Received: 13 July 2016 / Accepted: 3 February 2017 / Published online: 14 February 2017

Abstract There are several cases where a circular variable is associated with a linear one. A typical example is wind direction that is often associated with linear quantities such as air temperature and air humidity. The analysis of a statistical relationship of this kind can be tested by the use of parametric and non-parametric methods, each of which has its own advantages and drawbacks. This work deals with correlation analysis using both the parametric and the non-parametric procedure on a small set of meteorological data of air temperature and wind direction during a summer period in a Mediterranean climate. Correlations were examined between hourly, daily and maximum-prevailing values, under typical and non-typical meteorological conditions. Both tests indicated a strong correlation between mean hourly wind directions and mean hourly air temperature, whereas mean daily wind direction and mean daily air temperature do not seem to be correlated. In some cases, however, the two procedures were found to give quite dissimilar levels of significance on the rejection or not of the null hypothesis of no correlation. The simple statistical analysis presented in this study, appropriately extended in large sets of meteorological data, may be a useful tool for estimating effects of wind on local climate studies.

Responsible Editor: J.-F. Miao.

✉ I. X. Tsiros
yohan.tsiros@gmail.com

[1] Mathematics and Statistics Laboratory, Agricultural University of Athens, Athens, Greece

[2] Meteorology Laboratory, Agricultural University of Athens, Athens, Greece

1 Introduction

There are various important environmental problems where a correlation between circular and linear variables is encountered. In general, in the case of a circular-linear correlation, the sample space is the cylinder, and therefore the usual techniques of correlation analysis, requiring a plane sample space, cannot be applied (e.g., Mardia 1976). In such cases, a parametric and a non-parametric test are available. The parametric method requires the fulfillment of some strict conditions, whereas the non-parametric one does not require such assumptions at all (e.g., Mardia and Jupp 2000); the latter, however, does not provide any estimates of the amplitude of the directional effect or the direction of the maximal effect of the considered variable (e.g., Batscelet 1981). Although such parametric and non-parametric tests are available in their general form, their application in meteorological problems is limited. For example, a correlation between wind direction and air temperature data may provide useful information on the effects of wind regime on local climate in rural areas, in urban heat island analysis, climate change related studies, etc. At present, there is a lack of such applications and there are only a few papers focusing in related fields, e.g., air pollution (Solow et al. 1988; Jammalamadaka and Lund 2006; Fernández-Durán 2007; García-Portugués et al. 2013) and solar/wind energy applications (Pérez et al. 2007; SenGupta and Ugwuowo 2006).

The present paper deals with the correlation of a set of wind direction data and air temperature data during summer under Mediterranean climate conditions (Athens, Greece). The selected year is 2007, an exceptionally hot summer year for the whole region of south-eastern Europe. Three periods are analyzed: a 'typical summer' period, a 'heat-wave' period and the aggregated period. The correlation between air temperature and wind direction is examined using both hourly and daily data.

2 Methods

2.1 Parametric method

In the circular-linear correlation using a parametric approach, it is assumed that the linear variable y is normally distributed having its expected value depended on the circular variable θ with constant variance and also that the n pairs of observations are mutually independent. More specifically, the expected value of y is assumed to be a function of the angular distance between the angle θ and the acrophase angle θ_0 (i.e., the direction of maximal effect) (e.g., Batscelet 1981). In this case, the correlation coefficient between the linear variable y and the circular variable θ (denoted by $r_{y\theta}$) is defined as the (non negative) square root of

$$r_{y\theta}^2 = \frac{r_{yC}^2 + r_{yS}^2 - 2r_{yC}r_{yS}r_{CS}}{1 - r_{CS}^2}, \tag{1}$$

where $r_{yC},\ r_{yS},\ r_{CS}$ are partial correlation coefficients with C signifying the cosine and S the sine of θ, respectively.

Note that the correlation coefficient $r_{y\theta}$ lies between 0 and 1 (i.e., for linear-circular correlation there is no negative correlation).

The quantity

$$\frac{(n-3)r_{y\theta}^2}{1 - r_{y\theta}^2} \tag{2}$$

is distributed according to F-distribution with 2 and $n - 3$ degrees of freedom (Mardia and Jupp 2000).

2.2 Non-parametric method

The non-parametric method adopted in the present work does not involve any assumptions about the distribution of the dependent variable y, taken under consideration only that the n pairs of observations contain a random sample (e.g., Mardia 1976). In the context of this method, a circular-linear rank correlation coefficient D_n is calculated in the following way. First, the linear measurements y_i and the circular measurements ϑ_i are ranked in an ascending order independently. For simplicity, we assume that y_i has rank i (meaning,$y_1 \le y_2 \le \cdots \le y_n$). We denote the corresponding ranks of ϑ_i with r_i and we define the uniform scores β_i as:

$$\beta_i = \frac{2\pi r_i}{n}, \quad i = 1, 2, \ldots, n \tag{3}$$

Having then the adapted pairs of observations (1, β_1), (2, β_2)…(n, β_n) (with the scores β_i being evenly spaced directions on the unit circle) leads to the correlation coefficient D_n as follows:

$$D_n = \alpha_n(C^2 + S^2) \tag{4}$$

In the above correlation coefficient, α_n, C and S are appropriate trigonometric transformations of the quantities n and β_i that are not affected by any change in the principle axis of the linear variable y or any axis rotation of the circular variable θ.

$$C = \sum_{i=1}^{n} i\cos\beta_i,\ S = \sum_{i=1}^{n} i\sin\beta_i \quad \text{and}$$

$$\alpha_n = \begin{cases} [1 + 5\cot an^2(\pi/n) + 4\cot an^4(\pi/n)]^{-1} & (n \text{ even}) \\ 2\sin^4(\pi/n)/[1 + \cos(\pi/n)]^3 & (n \text{ odd}) \end{cases} \tag{5}$$

The quantity

$$U_n = \frac{24(C^2 + S^2)}{n^2(n+1)} \tag{6}$$

asymptotically follows the χ^2 distribution with two degrees of freedom. This approximation is adequate for $n \ge 100$ (Mardia 1976).

For the implementation of the rank correlation test, especially for cases that are not covered by the tables of critical values existing in the literature (e.g., Mardia 1976), a program was developed (written in Visual Basic 6.0 programming language) for calculating the value of statistic U_n in Eq. (6) and the corresponding p value as well. The determination of the critical level p value of U_n statistic was based on its empirical distribution where the values were calculated by simulation. Specifically, 10,000 random permutations were applied to the sample (y_i, θ_i | $i = 1, 2, \ldots, n$) by permuting the column of y's while keeping the column of θ's fixed. For the generation of these random permutations three alternative algorithms, based on the work of Maurer and Ralston (2004), were studied of which the most efficient one [order $O(n)$] was integrated into the program. The non-parametric test was implemented into a software tool written in Visual Basic 6.0, available from the authors upon request. The procedure was then applied to a small set of wind and air temperature data for Athens, Greece.

3 Area and data

3.1 Area and climate conditions

Athens (37°58′N, 23°43′E, 107 m a.s.l.) enjoys, in general, a typical Mediterranean climate, with hot dry summers and mild wet winters. More specifically, according to Köppen climatic classification, Athens climate is characterized by a Mediterranean, mild humid climate with dry-warm hot-summer. The average daily sunshine duration varies between 4.5 h in January and 12 h in July. Precipitation averages to be just above 400 mm annually with the rainy season to be between October and March. The average daily winter temperature is 9.4 °C. Summers can be particularly hot with a mean maximum temperature of 32.5 and 32.6 °C for the months of July and August, respectively (Droulia et al. 2009). Heat waves are common during the months of July and August along with high thermal stress conditions, whereas an extended summer period also is not a rare phenomenon in Athens (Shashua-Bar et al. 2010).

The prevailing winds in the Athens basin blow from N and NE in late summer, fall and winter and from SSW and SW in the spring and early summer. In general, in Athens there are two types of summer climate conditions characterized by their wind direction. During most of the summer time the dominant Etesian winds are from the northern sector. Under weak synoptic wind conditions, in general, sea breeze from S–SSW appear before noon or in the afternoon. Those are generally considered as the 'hottest and humid days'. According to the air vapor pressure and its related air temperature, at noon-hours analysis, four types of interchanging summer climate periods can be considered for Athens (Shashua-Bar et al. 2012):

– 'not rare hot-humid periods' (with air partial vapor pressure around 2.67–2.93 kPa),
– 'typical summer periods' (with air partial vapor pressure 1.87–2.0 kPa),
– 'hot-dry periods' (with air partial vapor pressure 1.33–1.60 kPa),
– 'extraordinary very hot and humid periods' with air partial vapor pressure to reach approximately 4.0 kPa.

3.2 Weather conditions and meteorological data

The present analysis makes use of a set of detailed hourly summer meteorological data for the year 2007. The examined summer of 2007 was exceptionally hot for the whole region of south-eastern Europe including Greece. A number of heat waves hit the area at the end of June, in July and in August and Greece experienced an all-time record breaking hot summer with nocturnal air temperatures also to remain at high levels (Founda and Giannakopoulos 2009). During the third week of June the area was affected by the first and more severe summer heat wave ever recorded in Greece. Air temperature then started to rise gradually above the normal values and reached 46.2 °C on June 26 according to the meteorological data recorder at the Elefsis airport close to Athens while on the same day the all-time record of daily maximum air temperature of 44.8 °C was recorded at the National Observatory of Athens station (NOA) exceeding the previous record value observed on June 21, 1916 by almost 2 °C (Tsiros and Hoffman 2014).

In the present study, wind direction and air temperature data of July 2007 are used. During the first 17 days of July 2007, air temperature values were close to normal values (namely, normal for the season levels of temperature) with daily maximum air temperature to be in the range 30.8 and 37.6 °C and with daily average values 26.9–31.2 °C. During the last 14 days of July, however, air temperature values are recorded to be well above normal values with daily air maximum temperature to be in the range of 37.2–41.9 °C and daily average values in the range of 31.2–35.6 °C. Data are arranged in data sets as following: average values of air temperature and wind direction per hour [2 sets, 24 values each (case of an "average day")] and per day (2 sets, 31 values each) along with a set of 31 values of daily maximum air temperature and a set of 31 values of daily prevailing wind direction. We note that the average values of the circular variable of wind direction are vector means, computed by the use of the software Oriana® (Oriana 2004). Further, we clarify that for a sample of n observations $\theta_1, \theta_2, \ldots, \theta_n$, of a circular variable θ (meaning, n directions $\theta_1, \theta_2, \ldots, \theta_n$), the sample mean direction $\bar{\theta}$ is the direction of the resultant vector of the unit vectors $e_i = (\cos\theta_i, \sin\theta_i)$. Therefore, $\bar{\theta}$ is given explicitly by

$$\bar{\theta} = \begin{cases} \arctan(S/C), & \text{if } C \geq 0 \\ \arctan(S/C) + \pi, & \text{if } C < 0\,, \\ \text{Undefined}, & \text{if } S = C = 0 \end{cases} \tag{7}$$

where

$$S = \sum_{i=1}^{n} \sin\vartheta_i, \quad C = \sum_{i=1}^{n} \cos\vartheta_i$$

and arctan takes values in $[-\pi/2, \pi/2]$.

4 Results and discussion

The correlation coefficient $r_{y\theta}$ (Eq. 1) for various air temperature and wind direction data sets was calculated along with the significance controlled by Eq. 2. Subsequently, the corresponding rank correlation test was performed by calculating the values of statistic U_n. The p values were also obtained from the simulation program. We note that the parametric test was performed in cases where the required conditions were fulfilled. More specifically, all air temperature variables examined in this work [actual (raw) hourly, mean hourly, mean daily and daily maximum air temperature variables] were tested for normality by the use of Kolmogorov–Smirnov goodness-of-fit test. All tests indicated that all the above-mentioned air temperature variables can be considered as being normally distributed. Only the actual (raw) values on a daily basis of air temperature are not normally distributed.

Results of the implementation of the parametric and the non-parametric tests are summarized in Tables 1, 2, 3, 4 and 5; those of the former are given in the second and third column, where $r_{y\theta}$ represents the correlation coefficient in (Eq. 1), and of the latter are given in the last three columns, where D_n represents the correlation coefficient in (Eq. 4)

Table 1 Results for average values per hour of air temperature and wind direction

Variables	Parametric test		Non-parametric test		
Mean hourly wind direction and mean hourly air temperature	$r_{y\theta}$	p (prob.)	D_n	U_n	p (prob.)
	0.8410	**0.0000**	0.6808	15.4366	**0.0000**

Variables are strongly correlated

Bold values show significant results ($p \leq 0.05$)

Table 2 Results for average values on a daily basis of air temperature and wind direction

Variables	Parametric test		Non-parametric test		
Mean daily wind direction and mean daily temperature	$r_{y\theta}$	p (prob.)	D_n	U_n	p (prob.)
Aggregated period (31 days)	0.1350	0.6010	0.1654	4.8930	0.0895
Typical summer period (17 days)	0.4630	**0.0480**	0.2502	3.9489	0.1400
Heat wave period (14 days)	0.3640	0.2320	0.0269	0.3459	0.8500

Variables are not correlated

Bold values show significant results ($p \leq 0.05$)

Underlined values show cases of accordance between parametric and non-parametric tests with dissimilar levels of significance

Double-underlined values show cases of disagreement between parametric and non-parametric tests

Table 3 Results for daily wind direction and daily maximum air temperature

Variables	Parametric test		Non-parametric test		
Mean daily wind direction and daily maximum air temperature	$r_{y\theta}$	p (prob.)	D_n	U_n	p (prob.)
Aggregated period (31 days)	0.2700	0.1290	0.2127	6.2915	**0.0300**
Typical summer period (17 days)	0.6500	**0.0020**	0.3482	5.4940	**0.0500**
Heat wave period (14 days)	0.2960	0.3810	0.0269	0.3459	0.4000

Variables are correlated during the 'typical summer' period and not correlated during the 'heat wave' period

Bold values show significant results ($p \leq 0.05$)

Underlined values show cases of accordance between parametric and non-parametric tests with dissimilar levels of significance

Double-underlined values show cases of disagreement between parametric and non-parametric tests

Table 4 Results for daily prevailing wind direction and mean daily temperature

Variables	Parametric test		Non-parametric test		
Daily prevailing wind direction and mean daily temperature	$r_{y\theta}$	p (prob.)	D_n	U_n	p (prob.)
Aggregated period (31 days)	0.1760	0.4200	0.1554	4.5962	0.8000
Typical summer period (17 days)	0.4720	**0.0420**	0.0060	0.0951	0.9887
Heat wave period (14 days)	0.2530	0.4940	0.3591	4.6049	0.5323

Variables are not correlated

Bold values show significant results ($p \leq 0.05$)

Double-underlined values show cases of disagreement between parametric and non-parametric tests

Table 5 Results for daily prevailing wind direction and daily maximum air temperature

Variables	Parametric test		Non-parametric test		
Daily prevailing wind direction and daily maximum air temperature	$r_{y\theta}$	p (prob.)	D_n	U_n	p (prob.)
Aggregated period (31 days)	0.2920	0.0910	0.0693	2.0499	0.9323
Typical summer period (17 days)	0.6600	**0.0010**	0.0411	0.6494	0.9215
Heat wave period (14 days)	0.2690	0.4500	0.2448	3.1402	0.6376

Variables are not correlated

Bold values show significant results ($p \leq 0.05$)

Underlined values show cases of accordance between parametric and non-parametric tests with dissimilar levels of significance

Double-underlined values show cases of disagreement between parametric and non-parametric tests

and U_n represents the statistic in (Eq. 6). The significance of the results, accepting a 5% level of significance, is shown in bold. Cases of disagreement (on the rejection or not of the null hypothesis of no correlation) between the two methods are noted by double-underlining the corresponding results and cases of accordance with dissimilar levels of significance are noted by underlining.

Table 1 shows the results on the average values per hour (1 sample of 24 values, actually the case of an "average day"). Both tests are in total agreement providing very significant results ($p = 0.0000$), indicating thus that average values per hour of air temperature and wind direction are strongly correlated.

In the case of average values per day (1 sample of 31 values—see Table 2) both tests agree that for the aggregated period and the "heat-wave" period the average values on a daily basis of air temperature and wind direction are not correlated ($p \geq 0.0895$). For the "typical summer" period the two methods are in disagreement ($p = 0.0480$ and 0.1400, respectively).

The statistics results for the values of daily maximum air temperature and wind direction on the average per day are shown in Table 3, where both tests agree that these variables are correlated during the "typical summer" period and they are not correlated during the "heat-wave" period. Also, during the aggregated period the two methods are in disagreement ($p = 0.1290$ and 0.0300, respectively).

Table 4 shows that both tests agree that for the aggregated period and the "heat-wave" period air temperature on the average per day and daily prevailing wind direction are not correlated ($p \geq 0.4200$). For the "typical summer" period, the two methods are in disagreement ($p = 0.0420$ and 0.9887, respectively).

Concerning the values of daily maximum air temperature and daily prevailing wind direction (Table 5), both tests agree that the examined variables are not correlated during neither the aggregated nor the "heat-wave" period. One case of disagreement is noted, however, which is related to the "typical summer" period ($p = 0.0010$ and 0.9215, respectively).

Finally, the correlation between actual (raw) hourly values of air temperature and wind direction (24 sets of up to 31 data pairs each) has been tested by the use of both methods. The non-parametric test gave no significant results at all and the parametric test gave significant results in only four cases.

As presented above, several agreements between the results of the two correlation tests studied here were found. There were two marginal and two critical cases of disagreement. It may be argued thus that for the data considered in this work, the two correlation tests are largely in agreement. On the other hand, it should be noted that the two tests, even when they agree on the rejection or not of the null hypothesis of no correlation, they usually give quite dissimilar levels of significance: in five out of nine cases of accordance, p values are in a different order of magnitude. These discrepancies may be attributed to the non-sinusoidal shape of the regression curve (e.g., Solow et al. 1988).

5 Concluding remarks

From the results of the present study on a limited data set of meteorological data on wind direction and air temperature, the following conclusions may be withdrawn:

- Both parametric tests showed that a strong correlation between mean hourly wind direction and mean hourly air temperature data does exist.
- Despite the fact that several agreements between the results of the two correlation tests studied were found they usually gave quite dissimilar levels of significance.
- The simple statistical analysis presented in this study, appropriately extended in large sets of meteorological data, may be a useful tool for estimating effects of wind on local climate studies.

Acknowledgements The authors wish to thank the three anonymous reviewers for their careful review and also for their insightful suggestions, which led to a substantial improvement of the original manuscript.

References

Batscelet E (1981) Circular statistics in biology. Academic Press Inc. Ltd, London

Droulia F, Lykoudis S, Tsiros IX, Alvertos N, Akylas E, Garofalakis I (2009) Ground temperature estimations using simplified analytical and semi-empirical approaches. Sol Energy 83:211–219

Fernández-Durán JJ (2007) Models for circular-linear and circular-circular data constructed from circular distributions based on nonnegative trigonometric sums. Biometrics 63:579–585

Founda D, Giannakopoulos C (2009) The exceptionally hot summer of 2007 in Athens, Greece. A typical summer in the future climate? Glob Planet Change 67:227–236

García-Portugués E, Crujeiras RM, González-Manteiga W (2013) Exploring wind direction and SO_2 concentration by circular-linear density estimation. Stoch Environ Res Risk Assess 27:1055–1067

Jammalamadaka SR, Lund UJ (2006) The effect of wind direction on ozone levels: a case study. Environ Ecol Stat 13:287–298

Mardia K (1976) Linear-circular correlation coefficients and rhythmometry. Biometrika 63:403–405

Mardia K, Jupp P (2000) Directional statistics. Wiley, New Jersey

Maurer SB, Ralston A (2004) Discrete algorithmic mathematics. A K Peters/CRC Press, Boca Raton

Oriana V2.02a (2004) Kovach computing services. http://www.kovcomp.co.uk. Accessed 9 Feb 2017

Pérez IA, Ángeles M, García MA, Sánchez L, de Torre B (2007) Analysis of directional meteorological data by means of cylindrical models. Renew Energy 32:459–473

SenGupta A, Ugwuowo FI (2006) Asymmetric circular-linear multivariate regression models with applications to environmental data. Environ Ecol Stat 13:299–309

Shashua-Bar L, Tsiros I, Hoffman M (2010) A modeling study for evaluating the thermal regime of passive cooling scenarios in urban streets. Case study: Athens, Greece. Build Environ 45:2798–2807

Shashua-Bar L, Tsiros I, Hoffman ME (2012) Passive cooling design options to ameliorate thermal comfort in urban streets of a Mediterranean climate (Athens) under hot summer conditions. Build Environ 57:110–119

Solow A, Bullister J, Nevison S (1988) An application of circular-linear correlation analysis to the relationship between freon concentration and wind direction in Woods Hole Massachusetts. Environ Monit Assess 10:219–228

Tsiros IX, Hoffman ME (2014) Thermal and comfort conditions in a rear wooded garden and its adjacent semi-open spaces in a Mediterranean climate (Athens) during summer. Archit Sci Rev 57:63–82