Received April 26, 2022, accepted May 18, 2022, date of publication May 27, 2022, date of current version June 8, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3178423

Circular Correntropy: Definition and Application in Impulsive Noise Environments

MANOEL B. L. AQUINO[1,2], (Student Member, IEEE), JOÃO P. F. GUIMARÃES[1,2], ALUISIO I. R. FONTES[1,2], LEANDRO L. S. LINHARES[3], JOILSON B. A. REGO[2,4], AND ALLAN DE M. MARTINS[2,4], (Member, IEEE)

[1]Department of Information, Federal Institute of Rio Grande do Norte (IFRN), Natal, Pau dos Ferros, Rio Grande do Norte 59015-000, Brazil
[2]Federal University of Rio Grande do Norte (UFRN), Natal, Rio Grande do Norte 59078-970, Brazil
[3]Federal Institute of Education, Science, and Technology of Paraíba, Joao Pessoa, Cajazeiras 58900-000, Brazil
[4]Department of Computer Engineering and Automation, Federal University of Rio Grande do Norte, Natal 59078-900, Brazil

Corresponding author: Manoel B. L. Aquino (bonfim.aquino@ifrn.edu.br)

ABSTRACT Circular statistics has been applied to several areas of knowledge in which the input data is circular or directional. Noisy measurements are still a problem in circular data applications and, like non-circular data, second-order statistics have some limitations to deal with non-Gaussian noise. Recently, a similarity function called correntropy has been successfully employed in applications involving impulsive noise for being capable of extracting more information than second-order methods. However, correntropy has not been studied from the perspective of circular data so far. This paper defines a novel statistical measure called circular correntropy (CC). It uses the von Mises density function in order to redefine correntropy in this domain. In particular, it is proved analytically that the CC contains information regarding second-order and higher-order moments, being a generalization of the circular correlation measure. The performance of this novel similarity measure is evaluated as a cost function in a nonlinear regression problem, where the signals are contaminated with additive impulsive noise. The simulations demonstrate that the CC is more robust than circular correlation in impulsive noise environments.

INDEX TERMS Circular correntropy, circular statistics, correntropy, directional statistics.

I. INTRODUCTION

Circular data can be found in nature or produced artificially by physical devices such as a compass, watch hands, windsock, theodolite, among others. Such signals are commonly represented as angle values in degrees or radians in relation to an arbitrary origin in a clockwise or counterclockwise direction [1]. The fact that the mean direction between 10° and 350° is 0° instead of 180° provides an illustration of how specific statistical methods are required for analyzing angular data.

Circular statistics is an area of statistics that deals with periodic data that assume values in an interval from $-\pi$ to π, which can also be represented as points on the circumference of a unit circle. In this case, the support for circular data is the unit circle like a real line is the support for linear data [2].

Several works have presented specific concepts for the representation and analysis of circular data, e.g., circular and linear histograms, rose diagrams, data distributions such as uniform and von Misses distributions, measures of location, concentration, dispersion and similarity such as mean, variance and correlation, among others [2]–[5]. The aforementioned theories are applied in different areas of science such as meteorology [6], [7], biology [8], animal migration [9], medicine [10], [11], temporal events [12], [13]. However, these types of applications are not free from noise degradation [14], [15].

Correntropy is a kernel-based similarity measure capable of extracting infinite even-order statistical moments from data, being a generalization of the correlation concept [16]. Another important characteristic of this measure is that it provides improved performance when compared with second-order methods when dealing with non-Gaussian noise such as impulsive noise environments [17]–[26]. Owing to such characteristics, correntropy has been successfully applied to many practical problems, e.g., extraction of higher-order temporal characteristics [27], [28], estimation of impulsive noise [29] and active noise control [30], [31]–[33],

The associate editor coordinating the review of this manuscript and approving it for publication was Zhen Ren.

Kalman filter [19], [34], compressive sensing problems in impulsive noise environments [35], [36]. However, the application of correntropy to circular data has not yet been explored or defined.

This paper presents a novel similarity measure applied to circular data defined as circular correntropy (CC), which is based on the conventional correntropy definition, but using the von Mises density function as the kernel function in the Parzen window estimator. Besides that, it is shown that, analogously to correntropy, CC has high-order statistical information, but with a computational burden equivalent to second-order methods. Furthermore, it is shown that the CC generalizes the circular correlation concept. This is a significant result, because it provides an efficient way of analyzing both second-order and higher-order environments with a computational burden equivalent to that of correlations. Simulation results demonstrate the advantages of the proposed measure in non-Gaussian environments when compared with second-order statistical measures, e.g., circular correlation and the mean square error (MSE) method, where the impulsive noise deteriorates the performance of the algorithms such as with non-circular data [37].

The remainder of this paper is organized as follows. Section II introduces circular statistics and highlights important concepts and measures. Section III provides a theoretical background on correntropy. Section IV defines the CC for circular statistics and shows that it generalizes the circular correlation. Section V presents the obtained results in order to analyze the performance of CC in non-Gaussian environments. Section VI summarizes the main conclusions.

II. CIRCULAR STATISTICS

In circular statistics, which is also referred to as directional statistics, measures like mean and variance are not as straightforward as their counterparts associated with non-circular statistics [4]. Since values now can wrap around from $-\pi$ to π, it is not possible to rely on polynomial moments that would be sensitive to limits of integration and wrapping around boundaries. Hence, measures like mean and variance are defined using circular moments in the form [2], [4]:

$$m_n^{\Phi} = E\{e^{jn\Phi}\} = \int_{-\pi}^{\pi} e^{jn\phi} f_{\Phi}(\phi)\, d\phi, \tag{1}$$

where m_n is the n-th circular moment generating function of the random variable Φ, whose distribution is $f_{\Phi}(\phi)$. Besides, it is worth mentioning that m_n is a complex number, while some measures can be defined by such parameter. Based on this concept, it is possible to define the following quantities [4], [38]:

$$\mu_{\Phi} = \angle(m_1^{\Phi}), \tag{2}$$

$$R(\Phi) = |m_1^{\Phi}|, \tag{3}$$

where μ_{Φ} is the mean direction angle and $R(\Phi)$ is the mean resultant length, which denotes how concentrated the data are. Based on this assumption, one can define the variance as $\mathrm{Var}(\Phi) = 1 - R(\Phi)$ [4], [5].

A. CIRCULAR PROBABILITY DENSITY FUNCTION

The data involved associated with probability density functions in circular statistics must be limited to the interval $-\pi$ to π or 0 to 2π. There are several distributions that can be used to meet this criterion. A popular distribution used for this purpose in the literature is the von Mises distribution, i.e., the circular normal distribution [38], which is given by

$$M_{\sigma}(\phi|\mu, \sigma) = \frac{e^{\sigma\cos(\phi-\mu)}}{2\pi I_0(\sigma)}, \tag{4}$$

where parameter μ is the circular mean of the random variable that has a von Mises distribution, and σ is the parameter related to the variance of such random variable; I_0 is the modified zero-order Bessel function of the first kind. The von Mises distribution resembles the Gaussian distribution as σ tends to zero.

For the von Mises distribution, the variance of the random variable can be computed as [2], [4]:

$$\mathrm{Var}(\Phi) = 1 - E\{\cos(\Phi - \mu)\}. \tag{5}$$

B. CIRCULAR CORRELATION

In circular statistics, correlation is a very important measure that has been used in several areas of knowledge, which include biology [39], signal processing [40], neuroscience [41], among others. It has also been used to measure the association of two circular random variables. A very common correlation concept is the Fisher correlation coefficient [42], which is given by

$$\begin{aligned} &Corr(\Phi, \Theta) \\ &= \frac{E\left[\sin(\Phi - \mu_{\Phi})\sin(\theta - \mu_{\theta})\right]}{\sqrt{E\left[\sin(\Phi - \mu_{\Phi})^2\right] E\left[\sin(\theta - \mu_{\theta})^2\right]}} \\ &= \frac{E\left[\cos\left(\frac{(\Phi-\mu_{\Phi})-(\Theta-\mu_{\Theta})}{2}\right)^2 - \cos\left(\frac{(\Phi-\mu_{\Phi})+(\Theta-\mu_{\Theta})}{2}\right)^2\right]}{\sqrt{E\left[\sin(\Phi - \mu_{\Phi})^2\right] E\left[\sin(\theta - \mu_{\theta})^2\right]}}, \end{aligned} \tag{6}$$

where Φ and Θ are random variables defined in the interval $-\pi$ to π and μ_{Φ} and μ_{Θ} are the mean direction angles for Φ and Θ, respectively.

III. CORRENTROPY

Correntropy is a generalized similarity measure defined as [16]

$$V_{\sigma}(X, Y) = E\{\kappa_{\sigma}(X, Y)\}, \tag{7}$$

where X and Y are scalar real-valued random variables, $E[\,.\,]$ is the expected value operator, and $\kappa_{\sigma}(.)$ is any positive-definite kernel with parameter σ. Equation (7) can also be written in the form:

$$V_{\sigma}(X, Y) = \iint \kappa_{\sigma}(X, Y) f_{XY}(x, y)\mathrm{d}x\mathrm{d}y, \tag{8}$$

where $f_{XY}(x, y)$ is the joint probability density function (PDF) obtained from X and Y. Selecting the Gaussian kernel as

$\kappa_\sigma(X, Y) = G_\sigma(X, Y)$ leads to

$$V_\sigma(X, Y) = \iint G_\sigma(X, Y) f_{XY}(x, y)\mathrm{d}x\mathrm{d}y = E\{G_{\sigma'}(X, Y)\}, \tag{9}$$

where $\sigma' = \sqrt{2}\sigma$ [43].

IV. CIRCULAR CORRENTROPY

Based on the definition of conventional correntropy for planar statistics in (7), one can define the CC for circular statistics as:

$$C_\sigma(\Theta, \Phi) = E\{k_\sigma(\Theta, \Phi)\}, \tag{10}$$

where Θ and Φ are random variables defined in the interval $-\pi$ to π and σ is a size parameter for the kernel k.

In equation (10), the kernel must satisfy the same properties as the conventional correntropy. In order to obtain such a kernel, let us proceed as in the case of the conventional correntropy concept and apply the probabilistic interpretation [37], [44]. A detailed description of this procedure can be found in Appendix. In this approach, calculating the correntropy between two random variables Θ and Φ is equivalent to estimating the probability density of the event $\Theta = \Phi = \psi$.

$$C_\sigma(\Theta, \Phi) = \hat{P}(\Theta = \Phi) = \int_{-\pi}^{\pi} \hat{f}_{\sigma\Theta\Phi}(\theta, \phi)|_{\theta=\phi=\psi}\mathrm{d}\psi, \tag{11}$$

where $\hat{P}(\Theta = \Phi)$ is the estimated probability of the event $\Theta = \Phi$.

Then, in order to model the joint PDF $\hat{f}_{\sigma\Theta\Phi}(\theta, \phi)$ the Parzen window method is used with the von Mises density function as a kernel. In most cases, only a finite number of samples is available, resulting in:

$$\hat{f}_{\sigma\Theta\Phi}(\theta, \phi) = \frac{1}{N}\sum_{i=1}^{N} M_\sigma(\theta - \theta_i) M_\sigma(\phi - \phi_i), \tag{12}$$

where N is the number of samples.

Substituting the joint probability estimated in (12) in (11) leads to

$$\begin{aligned}\hat{C}_\sigma(\Theta, \Phi) &= \int_{-\pi}^{\pi} \hat{f}_{\sigma\Theta\Phi}(\psi, \psi)\mathrm{d}\psi \\ &= \int_{-\pi}^{\pi} \frac{1}{N}\sum_{i=0}^{\infty} M_\sigma(\psi - \theta_i) M_\sigma(\psi - \phi_i)\mathrm{d}\psi.\end{aligned} \tag{13}$$

Using equation (4) in (13) gives

$$\hat{C}_\sigma(\Theta, \Phi) = \frac{1}{4N\pi^2 I_0^2(\sigma)}\sum_{i=1}^{N}\int_{-\pi}^{\pi} e^{\sigma(\cos(\psi-\theta_i)+\cos(\psi-\phi_i))}\mathrm{d}\psi. \tag{14}$$

The integral in equation (14) can be solved from equations (15), (16), and (17):

$$\cos(\psi - \theta_i) + \cos(\psi - \phi_i) = A_i \cos(\psi + B_i), \tag{15}$$

$$\begin{aligned}A_i &= \sqrt{(\cos(\theta_i) + \cos(\phi_i))^2 + (\sin(\theta_i) + \sin(\phi_i))^2} \\ &= 2\sqrt{\frac{1 + \cos(\theta_i - \phi_i)}{2}} = 2\cos\left(\frac{\theta_i - \phi_i}{2}\right),\end{aligned} \tag{16}$$

$$B_i = -\tan^{-1}\left(\frac{\sin(\theta_i) + \sin(\phi_i)}{\cos(\theta_i) + \cos(\phi_i)}\right). \tag{17}$$

Then, the integral can then be rewritten as

$$R = \int_{-\pi}^{\pi} e^{\sigma A_i \cos(\psi + B_i)}\mathrm{d}\psi = \int_{-\pi - B_i}^{\pi - B_i} e^{\sigma A_i \cos(\psi)}\mathrm{d}\psi, \tag{18}$$

since the argument of the integrand is periodic with period 2π, equation (18) can be defined as

$$R = \int_{0}^{2\pi} e^{\sigma A_i \cos(\psi)}\mathrm{d}\psi, \tag{19}$$

this term can be identified as the modified Bessel function [4], from which it is possible to write

$$R = 2\pi I_0\left(2\sigma\cos\left(\frac{\theta_i - \phi_i}{2}\right)\right), \tag{20}$$

finally, equation (21) provides an estimator for the CC defined in (10) for finite samples. Under the condition of $N \to \infty$, $\hat{C}_{N,\sigma}(\Theta, \Phi)$ is an unbiased estimator of $C_{N,\sigma}(\Theta, \Phi)$ and consistent in mean square.

$$\hat{C}_\sigma(\Theta, \Phi) = \frac{1}{N}\sum_{i=1}^{N}\frac{I_0\left(2\sigma\cos\left(\frac{\theta_i - \phi_i}{2}\right)\right)}{2\pi I_0^2(\sigma)} \approx E\{\kappa_\sigma(\Theta, \Phi)\}, \tag{21}$$

where $k_\sigma(\Theta, \Phi)$ is defined as:

$$k_\sigma(\Theta, \Phi) = \frac{I_0\left(2\sigma\cos\left(\frac{\Theta - \Phi}{2}\right)\right)}{2\pi I_0^2(\sigma)}. \tag{22}$$

Now, the correntropy concept can be extended to applications involving circular data.

A. MAXIMUM CIRCULAR CORRENTROPY CRITERION

After defining correntropy for circular data, it is possible to extend its application to optimization problems. This section is concerned with the introduction of a novel method for the maximum circular correntropy criterion, which employs the CC as a cost function.

Let $\Theta = \{\theta_i\}_{i=0}^N$ and $\Gamma = \{\boldsymbol{\gamma}_i\}_{i=0}^N$ be samples from the random variables Θ and the random vector Γ, respectively. One can now consider the generalized linear regression problem that can be applied to filtering, prediction, estimation, among other tasks. In this case, Θ and Γ are random variables measured in some process in which the values of Θ depend on the samples of Γ. Thus, the problem is to estimate the values

of Θ from the Γ samples, defined as $\hat{\Theta} = \boldsymbol{w}^{\intercal}\Gamma$, where $\boldsymbol{w}$ is a vector that weighs the values of each element in the random vector Γ.

As previously defined, $C_\sigma(\Theta, \Phi)$ estimates the probability of the variables Θ and Φ of being similar. Applying this concept to the regression problem defined above and replacing Φ with $\hat{\Theta}$ gives

$$J_{MCCC} = E\{\kappa_\sigma(\Theta, \hat{\Theta})\} = \frac{1}{N}\sum_{i=1}^{N} \frac{I_0\left(2\sigma\cos\left(\frac{\theta_i - \hat{\theta}_i}{2}\right)\right)}{2\pi I_0^2(\sigma)}, \tag{23}$$

where J_{MCCC} is the cost function for Maximum Circular Correntropy Criterion

The goal is to determine the optimal values for $\boldsymbol{w}$ that maximize the probability of the estimated sign $\hat{\theta}_i = \boldsymbol{w}^{\intercal}\boldsymbol{\gamma}_i$ to be similar to θ_i.

Analytically, it is hard to find a set of $\boldsymbol{w}$ that maximizes the cost function J_{MCCC}. Thus, one can use the ascending gradient method defined by:

$$\boldsymbol{w}_{n+1} = \boldsymbol{w}_n + \mu\nabla J_{MCCC} = \boldsymbol{w}_n + \mu\frac{\partial\hat{C}_\sigma(\Theta, \hat{\Theta})}{\partial\boldsymbol{w}}, \tag{24}$$

where the subscripts n, $n+1$ represent the iterations of the gradient method, and

$$\frac{\partial\hat{C}_\sigma(\Theta, \hat{\Theta})}{\partial\boldsymbol{w}} = \frac{\sigma}{2\pi N I_0^2(\sigma)}\sum_{i=1}^{N} I_1\left(2\sigma\cos\left(\frac{\theta_i - \hat{\theta}_i}{2}\right)\right) \times \sin\left(\frac{\theta_i - \hat{\theta}_i}{2}\right)\frac{\partial\hat{\Theta}}{\partial\boldsymbol{w}_i}, \tag{25}$$

where I_1 is the modified first-order Bessel function of the first kind.

Now, this new criterion can be applied to optimization problems involving circular data. Similarly to the conventional correntropy, the kernel size must be adjusted for each scenario and the particular characteristics of the kernel size associated with the CC are discussed in the forthcoming section.

B. KERNEL ANALYSIS

Analogously to the usual correntropy with Gaussian kernel, the new kernel function presented in the CC is based on von Mises probability distribution, which under certain conditions resembles a normal distribution. When the CC is estimated, the resulting values depend on the selected kernel size represented by σ.

The possibility of adjusting the kernel size provides an efficient mechanism that allows eliminating outliers or values that are too different from the data set statistics. In order to perform this analysis, let us consider the case for which two random variables Θ and Φ are compared. In particular, the kernel size controls the weight that each sample will have in the underlying estimation of the joint PDF of the data. In order to illustrate such behavior, the random variable $\delta = \frac{\Phi - \Theta}{2}$ is adopted. Thus, it is possible to expand equation (22) in terms of a Taylor series and rewrite it as

$$k_\sigma(\Theta, \Phi) = 1 + \frac{2^2\sigma^2\cos(\delta)^2}{4} + \frac{2^4\sigma^4\cos(\delta)^4}{64} + \frac{2^6\sigma^6\cos(\delta)^6}{2304} + \ldots. \tag{26}$$

When the random variable δ has no outliers, it means that Θ and Φ are more similar. Then, the cosine function remains close to one and the Bessel function will be weighted by higher-order terms in equation (26). In this case, choosing a small kernel size has little impact on the value of the kernel function for each sample and the correntropy presents a second-order behavior. On the other hand, when δ has outliers, it means that Θ and Φ are in different directions on the trigonometric circle, resulting in a longer trigonometric distance given by $d(\alpha, \beta) = 1 - \cos(\alpha - \beta)$ [5]. Thus, it leads the cosine function to assume small values, while the Bessel function has low values. In this case, a large value of the kernel size can be used to minimize the influence of outliers. In order to illustrate this effect, Figure 1 shows the behavior of $k_\sigma(\Theta, \Phi)$ as function of δ for three different values of the kernel size σ (1, 10, and 100).

When σ is high, it can be stated that the outliers, which correspond to values of $|\delta|$ much greater than 0, are weighted out and do not contribute to the final kernel value. The dependence of the kernel size on parameter σ has a numerical behavior opposite to that observed in the conventional correntropy. In other words, as σ increases, the kernel size is reduced in the case of CC.

C. GENERALIZED CORRELATION FUNCTION

Using the series expansion defined in (26) and considering a sufficiently small kernel size σ, the high-order terms can be neglected and the kernel function can be considered proportional to $1 + \cos\left(\frac{\Theta - \Phi}{2}\right)^2$. However, they are not directly proportional because of the constant term equal to 1 in (26). For optimization purposes, this constant has no effect, as one can always bias and rescale the measure to make (26) equal to a second-order measure according to (27). Therefore, one can state that, for small values of σ, despite a bias term, correntropy relates do Θ and Φ approximately like:

$$C_\sigma(\Theta, \Phi) \sim E\left\{\cos\left(\frac{\Theta - \Phi}{2}\right)^2\right\}. \tag{27}$$

Equation (27) is related to the variance of the difference between two variables used in second-order circular statistics as explained in Section II. Hence, the CC can be regarded as the generalization of the conventional second-order circular statistics when the kernel size is small.

In particular, in order to demonstrate that the CC is capable of generalizing the second-order statistics when σ is sufficiently small, equations (6), (27), and (21) can be used to determine the correlation among random variables. Let u consider two variable A and B, where $B = A + N(0, \eta)$.

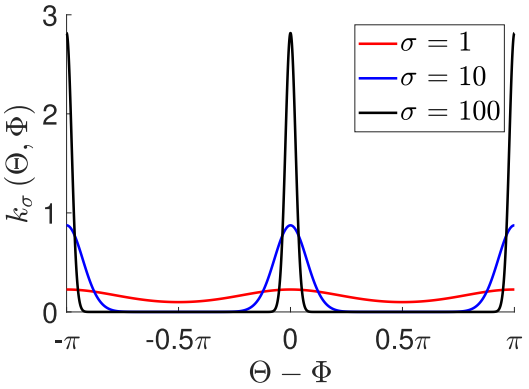

FIGURE 1. Behavior of the kernel function for distinct values of σ.

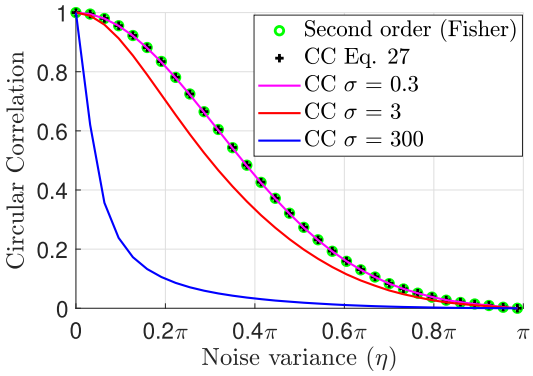

FIGURE 2. Comparison between correlation and correntropy with a small kernel size, demonstrating that circular correntropy generalizes second-order circular statistics. The plots were re-scaled to match the first and the last values of the correlation to 1 and 0, respectively.

Figure 2 shows a comparison among the CC for several values of σ, the simple correlation as defined by Fisher, and the approximation of the CC given by equation (27). The noise variance was modified so as to minimize the correlation between A and B and verify the behavior of the measures. As expected, the approximation for a small value of σ as obtained from equation (27), the result provided by equation (21) for $\sigma = 0.3$, and the Fisher correlation present nearly the same tendency. On the other hand, as σ increases, the correlation presents a different behavior than the second-order measures.

In fact, both equations (27) and (6) are similar to each other when considering only the numerator. As expected, the results presented in Figure 2 show that the CC generalizes the circular correlation concept when the kernel size is small enough.

D. JOINT PROBABILITY SPACE FOR CIRCULAR DATA

As previously mentioned, the correntropy measures the similarity between two random variables while estimating the probability of being equal to each other. This metric can be well illustrated plotting the PDF $\hat{f}_{\Theta\Phi}(\theta, \phi)$. When two variables are similar to each other, the higher values of $\hat{f}_{\Theta\Phi}(\theta, \phi)$ are concentrated around region $\Theta = \Phi$. As previously mentioned, both the CC and conventional correntropy calculate the integral of the density over this region. On the other hand, for two random variables with low similarity, the peaks or high values of $\hat{f}_{\Theta\Phi}(\theta, \phi)$ are scattered outside the region $\Theta = \Phi$. In the specific case of the CC, the space topology is no longer a plane. Both Θ and Φ are restricted to interval $[0,\ 2\pi)$, and, besides that, values equal to 0 are considered neighbors of values equal to 2π, while the space defined by $\Theta \times \Phi$ is a toroid. In this context, $\Theta = \Phi$ now defines a region in the toroid that corresponds to an inner ring. Thus, the CC now estimates the integral of $\hat{f}_{\Theta\Phi}(\theta, \phi)$ along the ring. In order to illustrate this behavior, let us consider the relationship between Θ and Φ as $\Phi = a\Theta + N(0, 0.05)$. Figure 4 shows the joint probability space for two random variables in terms of the planar and toroidal representations. Figs. 3a and 3b correspond to two random variables that are highly similar, where $a = 1$ and the Figs. 4a and 4b correspond to $a = 2$, i.e., the variables have low degree of similarity. The light cyan lines represent the surface $\Theta = \Phi$. The first column denotes that the data are distributed around the line, thus evidencing a high degree of similarity. The data are not coincident in the second column, thus corresponding to a low degree of similarity, i.e., the integral of $\hat{f}_{\Theta\Phi}(\theta, \phi)$ has a low value.

V. RESULTS

This section aims to demonstrate the application of the CC to two distinct approaches. Firstly, a brief interpretation of the CC concept involving real data on wind direction measurements is presented. Secondly, some results on linear system identification are discussed in order to assess the influence of the kernel size, validate the generalization of the proposed metric experimentally, and establish a fair comparison with results provided by second-order statistics.

A. SIMILARITY ANALYSIS

In order to analyze the behavior of the method introduced in this work, wind direction data measured in the city of Porto Alegre, Brazil in 2016 were used, which are available at [45].The data can be represented by a circular random variable Θ, being $\theta_1 \ldots \theta_n \in [0,\ 2\pi)$ and was analyzed in two scenarios: in the presence of outliers and in an environment with impulsive noise.

The performance of the new metric was then compared with that of second-order statistics. The joint probability space was plotted along a toroid following the same methodology employed in Figure 4. The values of Θ were distributed along the circular arc generatrix of the toroid, while the samples of contaminated variable Φ were distributed along the circumference at the central axis.

1) SIMILARITY ANALYSIS IN PRESENCE OF OUTLIERS

In order to investigate the robustness of the assessed methods, the data were contaminated with 40% of outliers in direction $\pi/4$, while the new set of samples was defined as belonging to random variable Φ.

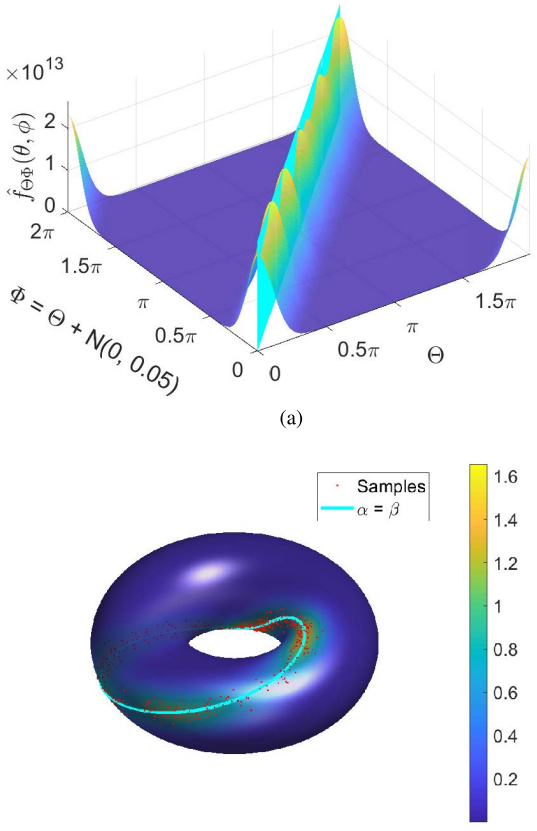

FIGURE 3. Joint probability space for two variables defined as Θ and $\Phi = a\Theta + N(0, 0.05)$ for $a = 1$. The representation is shown in the plane and in the toroid, respectively.

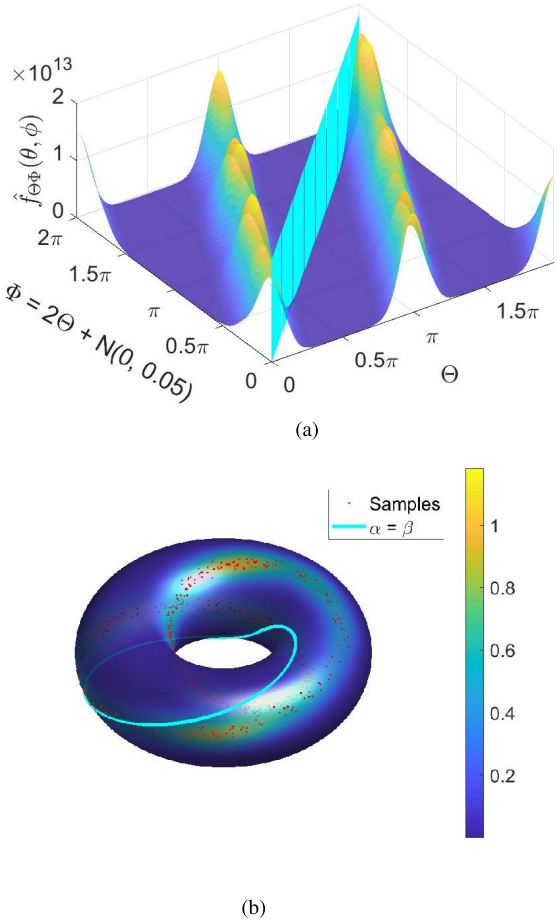

FIGURE 4. Joint probability space for two variables defined as Θ and $\Phi = a\Theta + N(0, 0.05)$ for $a = 2$. The representation is shown in the plane and in the toroid, respectively.

Figure 5(a) comprises the analysis using the second-order statistics. It shows that the probability values are strongly influenced by the presence of outliers since the values at $\phi = \pi/4$ are significantly higher than those along the line $\Theta = \Phi$. On the other hand, Figure 5(b) considers the same scenario for the application of CC with $\sigma = 50$. It is observed that the probability function is not concentrated around the region containing outliers, but distributed along the line $\Theta = \Phi$. Therefore, it is reasonable to state that the CC is more robust to outliers with an appropriate kernel size selection.

2) SIMILARITY ANALYSIS FOR ALPHA-STABLE NOISE ENVIRONMENTS

Moreover, an additional test was performed by contaminating the samples with an alpha-stable noise comprising a wrapped Cauchy (WC) distribution [5].

Figure 6(a) shows the behavior of MSE in an impulsive noise environment, which considers only second-order statistics. One can notice that the probability values are strongly influenced by the samples outside the line $\Theta = \Phi$. On the other hand, Figure 6(b) considers the same scenario using CC with $\sigma = 50$ instead of MSE. It is observed that the probability function is distributed along the line $\Theta = \Phi$. Thus, is reasonable to state that the CC is more robust to noise that follows a long tail distribution such as the alpha-stable.

B. CIRCULAR REGRESSION APPLICATION

In order to evaluate the applicability of the CC, let us consider the example of a regression using different kernel size values. The goal is to fit the following simple model:

$$\Theta = \bar{w}_2\Phi^2 + \bar{w}_1\Phi + \bar{w}_0 + \eta. \tag{28}$$

The observations of variables Θ and Φ are defined by θ_i and ϕ_i, where $i = 1, \ldots, N$ and $\bar{\boldsymbol{w}}_0$, $\bar{\boldsymbol{w}}_1$ and $\bar{\boldsymbol{w}}_2$ are the known parameters of the model; and η is an additive non-Gaussian noise signal. Therefore, this procedure aims at determining values of w_0, w_1, and w_2 for which the estimated model $\hat{\Theta} = w_2\Phi^2 + w_1\Phi + w_0$ is as close as possible to Θ. In order to achieve this goal using the CC approach, it is necessary to maximize this statistical measure between the estimation $\hat{\Theta}$ and the observations of model Θ.

Owing to the oscillatory nature of this cost function, it is expected to contain the local maximum. Therefore, a little improvement associated with the gradient rule is required. In particular, the momentum gradient rule is used in this work for this purpose [46]. Basically, the cost function is used as the

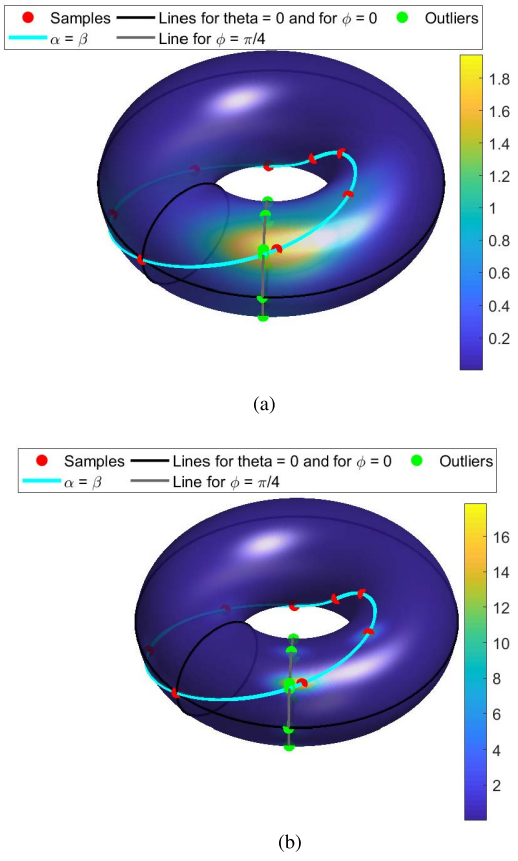

FIGURE 5. Joint probability space for θ and Φ with outliers in $\pi/4$. (a) Second-order statistics (b) CC for $\sigma = 50$.

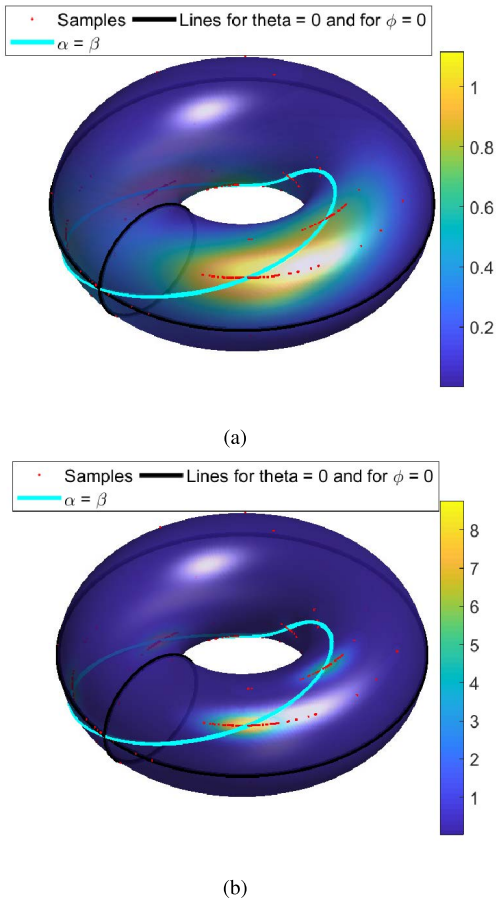

FIGURE 6. Joint probability space for θ and Φ with alpha stable noise. (a) Second-order statistics (b) CC for $\sigma = 50$.

gravity, while a friction coefficient, a mass, and an integration time interval are adopted. Now, it is possible to write the update rule for parameters $\boldsymbol{w}_j$ as:

$$\begin{aligned} v_{j+1} &= v_j + \mu m \nabla J_{MCCC} - \beta v_j, \\ w_{j+1} &= w_j + \mu v_{j+1}, \end{aligned} \tag{29}$$

where v_j and w_j are the velocity and position of the parameters, respectively; ∇J_{MCCC} represents a derived cost function with respect to w_j; m, β, and μ are the mass, friction and integration time interval, respectively as defined in Equation 23.

In this section, simulation results are presented to validate the theoretical analysis and demonstrate the performance of the proposed measure. All results are calculated from the average of 10^3 Monte Carlo trials. The performance is evaluated by the weighted signal-to-noise ratio (WSNR), which is used to quantify the convergence rate properly in decibels [47], as defined by

$$\text{WSNR}_{db} = 10 \log_{10} \left(\frac{\bar{\boldsymbol{w}}^{\mathsf{T}} \bar{\boldsymbol{w}}}{(\bar{\boldsymbol{w}} - \boldsymbol{w})^{\mathsf{T}} (\bar{\boldsymbol{w}} - \boldsymbol{w})} \right), \tag{30}$$

where $\boldsymbol{w}$ are the parameter vectors computed by the aforementioned methods and $\bar{\boldsymbol{w}}$ are the known parameters used in the tests.

The desired signal is formed by the product of the proper weights $\bar{\boldsymbol{w}}$ and the input random variable values as defined in equation (28). Then, an impulsive noise signal is added, whose probability density function is characterized by $0.9N(0, 0.1) + 0.1N(2, 0.5)$.

TABLE 1. Mass and β coefficients used in the simulations for each assessed algorithm.

method	m	β
CC $\sigma = 0.8$	0.231	1×10^{-3}
CC $\sigma = 1.5$	0.9326	0.0825
CC $\sigma = 2$	0.7055	0.1565
MSE	0.2154	1×10^{-3}

Of course, there are proper sizes of m, β, and σ for each system. Thus, the performance of the CC was analyzed while varying m between 0 and 1; varying β between 0 and 0.6; using three distinct kernel sizes; and calculating the resulting MSE. Figures 7(a) to 7(d) show the performance of MSE and CC with kernel sizes of 0.8, 1.5, and 2. All simulations were performed considering the following parameters: $\mu = 0.02$, $\bar{\boldsymbol{w}} = \{0.3, -1, 1.5\}$, and $\boldsymbol{w} = \{0, 0, 0\}$. Table 1 presents the best results obtained in the tests.

In order to evaluate the proposed measure, the regression problem has been considered while using the values of m, β, and k listed in Table 1.

Figures 8 and 9 compare the WSNR performance for the MSE and CC using different values of kernel size in the

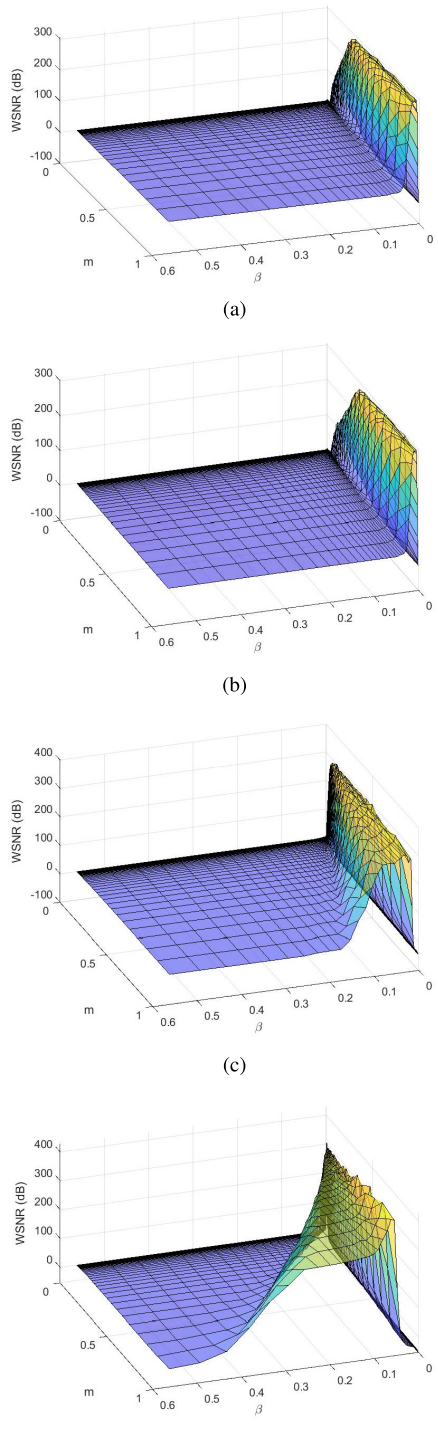

FIGURE 7. WSNR as a function of the m and β coefficients for the following algorithms: (a) MSE, (b), (c) and (d) CC with $\sigma = 0.8$, $\sigma = 1.5$, and $\sigma = 2$, respectively.

regression problem. In this experiment, it is possible to analyze the behavior of the CC. It is observed that the proposed measure performs significantly better than the MSE, while also generalizing second-order statistics for small values of σ.

The results demonstrate that the CC presents a significant improvement in terms of the WSNR when compared with the

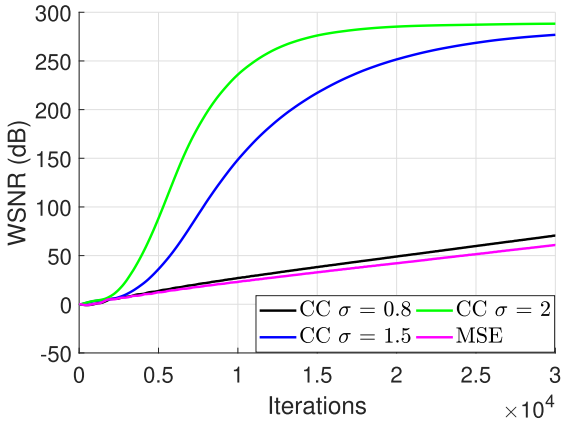

FIGURE 8. Performance comparison of correntropy for different kernel sizes and MSE in terms of the WSNR.

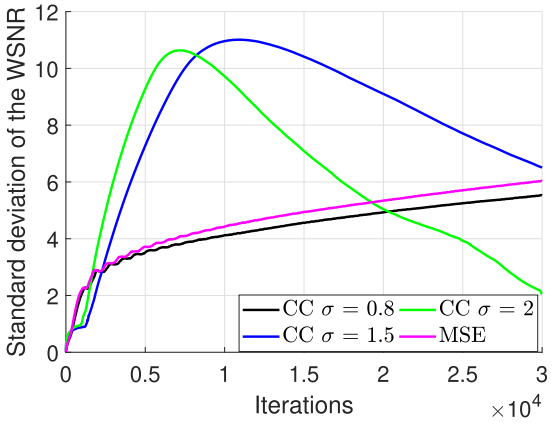

FIGURE 9. Standard deviation for the WSNR in Figure 8.

MSE when $\sigma = 1$ and $\sigma = 1.3$. This is due to the high-order moments that minimize the effects of impulsive noise. In particular, the CC tends to the second-order statistics for small kernel sizes as previously mentioned in Section IV-B.

VI. CONCLUSION

This work has presented a novel extension of the correntropy concept to circular data defined as CC. A significant contribution of this work lies in obtaining the expression for the CC through a probabilistic interpretation, where the von Mises density function was applied in the Parzen Window method to estimate how similar two circular random variables are. An analytical proof was presented to evidence that the CC contains information regarding second-order and higher-order moments, being a generalization of the circular statistics. A detailed analysis on the use of kernel size was presented to adjust high-order information, denoting that the CC tends to the second-order statistics for small values, this being the opposite behavior observed in the case of the conventional correntropy. The CC was used in computer simulations employing real data on the wind direction, showing its robustness to impulsive noise. Besides that, the maximum circular correntropy criterion (MCCC) was introduced and applied to a nonlinear regression problem, also achieving

significant performance improvement when compared with a second-order derived method. Future work includes investigating the application of the introduced measure to other real problems, e.g., direction of movement of icebergs, propagation of cracks, and directional phenomena.

APPENDIX
PROBABILISTIC INTERPRETATION

The probabilistic interpretation establishes that it is possible to estimate the correntropy between two random variables Θ and Φ by calculating the value of the probability density associated with the event $\Theta = \Phi$. Let σ be the parameter that defines the circular kernel width $K_\sigma(\Phi, \theta)$. Thus, making $\sigma \to \infty$ gives the following relation.

$$\lim_{\sigma\to\infty} C_\sigma(\Theta, \Phi) = \int_{-\pi}^{\pi} f_{\Theta,\Phi}(\theta, \phi)|_{\theta=\phi=\psi} \mathrm{d}\psi = P(\Theta = \Phi), \tag{31}$$

where $f_{\Theta,\Phi}(\theta, \phi)$ is the joint probability density function (PDF) obtained from Θ and Φ.

Proof: Starting from the definition of correntropy established in (7) and applying the von Mises distribution as a kernel function, one can obtain:

$$\begin{aligned} C_\sigma(\Theta, \Phi) &= E\{M_\sigma(\Theta - \Phi)\} \\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_{\Theta,\Phi}(\theta, \phi) M_\sigma(\Theta - \Phi)\mathrm{d}\theta\mathrm{d}\phi. \end{aligned} \tag{32}$$

Since the von Misses distribution tends to a Gaussian one when σ tends to infinity, the following equality can be established:

$$\lim_{\sigma\to\infty} M_\sigma(\theta) \equiv \delta(\theta), \tag{33}$$

Substituting (33) in (32) gives:

$$\lim_{\sigma\to\infty} C_\sigma(\Theta, \Phi) = \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} f_{\Theta,\Phi}(\theta, \phi)\delta(\theta - \phi)\mathrm{d}\theta\mathrm{d}\phi, \tag{34}$$

If $\Theta = \Phi = \psi$, one can write the following expression:

$$\lim_{\sigma\to\infty} C_\sigma(\Theta, \Phi) = \int_{-\pi}^{\pi} f_{\Theta,\Phi}(\psi, \psi)\mathrm{d}\psi = P(\Theta = \Phi). \tag{35}$$

Thus, it is reasonable to assume that correntropy can be interpreted as the probability density associated with event $\Theta = \Phi$.

REFERENCES

[1] G. Stienne, S. Reboul, M. Azmani, J. B. Choquel, and M. Benjelloun, "A multi-temporal multi-sensor circular fusion filter," *Inf. Fusion*, vol. 18, pp. 86–100, Jul. 2014.

[2] A. Pewsey, M. Neuhäuser, and G. D. Ruxton, *Circular Statistics in R*. Oxford, U.K.: Oxford Univ. Press, 2013.

[3] A. Pewsey and E. García-Portugués, "Recent advances in directional statistics," *Test*, vol. 30, no. 1, pp. 1–58, Mar. 2021.

[4] K. V. Mardia and P. E. Jupp, *Directional Statistics*. Hoboken, NJ, USA: Wiley, 2009, vol. 494.

[5] S. R. Jammalamadaka and A. Sengupta, *Topics in Circular Statistics*, vol. 5. Singapore: World Scientific, 2001.

[6] S. S. Alshqaq, A. A. Ahmadini, and A. H. Abuzaid, "Some new robust estimators for circular logistic regression model with applications on meteorological and ecological data," *Math. Problems Eng.*, vol. 2021, pp. 1–15, May 2021.

[7] J. A. Carta, C. Bueno, and P. Ramírez, "Statistical modelling of directional wind speeds using mixtures of von Mises distributions: Case study," *Energy Convers. Manage.*, vol. 49, no. 5, pp. 897–907, May 2008.

[8] L. P. C. Morellato, L. Alberti, and I. L. Hudson, "Applications of circular statistics in plant phenology: A case studies approach," in *Phenological Research*. Springer, 2010, pp. 339–359.

[9] R. M. Mutwiri, "Statistical distributions and modelling of GPS-telemetry elephant movement data including the E ECT of covariates," Ph.D. dissertation, Univ. KwaZulu Natal, Durban, South Africa, 2015.

[10] T. Drew and S. Doucet, "Application of circular statistics to the study of neuronal discharge during locomotion," *J. Neurosci. Methods*, vol. 38, nos. 2–3, pp. 171–181, Jul. 1991.

[11] A. Karaibrahimoglu, S. Ayhan, M. Karaagac, and M. Artac, "Circular analyses of dates on patients with gastric carcinoma," *J. Appl. Statist.*, vol. 48, nos. 13–15, pp. 2931–2943, Nov. 2021.

[12] M. G. Leguia, V. R. Rao, J. K. Kleen, and M. O. Baud, "Measuring synchrony in bio-medical timeseries," *Chaos: Interdiscipl. J. Nonlinear Sci.*, vol. 31, no. 1, Jan. 2021, Art. no. 013138.

[13] C. Brunsdon and J. Corcoran, "Using circular statistics to analyse time patterns in crime incidence," *Comput., Environ. Urban Syst.*, vol. 30, no. 3, pp. 300–319, May 2006.

[14] C. Agostinelli, "Robust estimation for circular data," *Comput. Statist. Data Anal.*, vol. 51, no. 12, pp. 5867–5875, Aug. 2007.

[15] E. A. Mahmood, S. Rana, H. Midi, and A. G. Hussin, "Detection of outliers in univariate circular data using robust circular distance," *J. Modern Appl. Stat. Methods*, vol. 16, no. 2, p. 22, 2017.

[16] I. Santamaria, P. P. Pokharel, and J. C. Principe, "Generalized correlation function: Definition, properties, and application to blind equalization," *IEEE Trans. Signal Process.*, vol. 54, no. 6, pp. 2187–2197, Jun. 2006.

[17] A. I. R. Fontes, J. B. A. Rego, A. M. de Martins, L. F. Q. Silveira, and J. C. Principe, "Cyclostationary correntropy: Definition and applications," *Expert Syst. Appl.*, vol. 69, pp. 110–117, Mar. 2017.

[18] T. Liu, T. Qiu, and S. Luan, "Cyclic correntropy: Foundations and theories," *IEEE Access*, vol. 6, pp. 34659–34669, 2018.

[19] H.-S. Jang, M. S. Muhammad, and M.-K. Kang, "Removal of non-Gaussian jitter noise for shape from focus through improved maximum correntropy criterion Kalman filter," *IEEE Access*, vol. 8, pp. 36244–36255, 2020.

[20] P. Yue, H. Qu, J. Zhao, and M. Wang, "An adaptive channel estimation based on fixed-point generalized maximum correntropy criterion," *IEEE Access*, vol. 8, pp. 66281–66290, 2020.

[21] Y. Guo, X. Li, and Q. Meng, "An outlier robust finite impulse response filter with maximum correntropy," *IEEE Access*, vol. 9, pp. 17030–17040, 2021.

[22] J. Zhao, J. A. Zhang, Q. Li, H. Zhang, and X. Wang, "Recursive maximum correntropy algorithms for second-order Volterra filtering," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 69, no. 4, pp. 2336–2340, Apr. 2022.

[23] I. B. Araújo, J. P. Guimarães, A. I. Fontes, L. L. Linhares, A. M. Martins, and F. M. Araújo, "Narx model identification using correntropy criterion in the presence of non-gaussian noise," *J. Control, Autom. Electr. Syst.*, vol. 30, no. 4, pp. 1–12, 2019.

[24] A. I. R. Fontes, A. de M. Martins, L. F. Q. Silveira, and J. C. Principe, "Performance evaluation of the correntropy coefficient in automatic modulation classification," *Expert Syst. Appl.*, vol. 42, pp. 1–8, Jan. 2015.

[25] J. P. Guimarães, A. I. Fontes, J. B. Rego, A. D. M. Martins, and J. C. Principe, "Complex correntropy function: Properties, and application to a channel equalization problem," *Expert Syst. Appl.*, vol. 107, pp. 173–181, Oct. 2018.

[26] E. L. Zhou, B. Y. Xia, E. Li, and T. T. Wang, "An efficient algorithm for impulsive active noise control using maximum correntropy with conjugate gradient," *Appl. Acoust.*, vol. 188, Jan. 2022, Art. no. 108511.

[27] E. Santana, J. C. Principe, E. Santana, and A. K. Barros, "Extraction of signals with higher order temporal structure using correntropy," *Signal Process.*, vol. 92, no. 8, pp. 1844–1852, Aug. 2012.

[28] M. A. Bakhshali, M. Khademi, A. Ebrahimi-Moghadam, and S. Moghimi, "EEG signal classification of imagined speech based on Riemannian distance of correntropy spectral density," *Biomed. Signal Process. Control*, vol. 59, May 2020, Art. no. 101899.

[29] J. Zhang, T. Qiu, A. Song, and H. Tang, "A novel correntropy based DOA estimation algorithm in impulsive noise environments," *Signal Process.*, vol. 104, no. 6, pp. 346–357, Nov. 2014.
[30] F. Jin, T. Qiu, S. Luan, and W. Cui, "Joint estimation of the DOA and the number of sources for wideband signals using cyclic correntropy," *IEEE Access*, vol. 7, pp. 42482–42494, 2019.
[31] Y. Zhu, H. Zhao, X. Zeng, and B. Chen, "Robust generalized maximum correntropy criterion algorithms for active noise control," *IEEE/ACM Trans. Audio, Speech, Language Process.*, vol. 28, pp. 1282–1292, 2020.
[32] L. Lu and H. Zhao, "Active impulsive noise control using maximum correntropy with adaptive kernel size," *Mech. Syst. Signal Process.*, vol. 87, pp. 180–191, Mar. 2017.
[33] E. L. Zhou, B. Y. Xia, E. Li, and T. T. Wang, "An efficient algorithm for impulsive active noise control using maximum correntropy with conjugate gradient," *Appl. Acoust.*, vol. 188, Jan. 2022, Art. no. 108511.
[34] S. K. Nanda, G. Kumar, V. Bhatia, and A. K. Singh, "Kalman filtering with delayed measurements in non-Gaussian environments," *IEEE Access*, vol. 9, pp. 123231–123244, 2021.
[35] J. P. F. Guimaraes, F. B. Da Silva, A. I. R. Fontes, R. Von Borries, and A. D. M. Martins, "Complex correntropy applied to a compressive sensing problem in an impulsive noise environment," *IEEE Access*, vol. 7, pp. 151652–151660, 2019.
[36] Y. He, F. Wang, S. Wang, J. Cao, and B. Chen, "Maximum correntropy adaptation approach for robust compressive sensing reconstruction," *Inf. Sci.*, vol. 480, pp. 381–402, Apr. 2019.
[37] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: Properties and applications in non-Gaussian signal processing," *IEEE Trans. Signal Process.*, vol. 55, no. 11, pp. 5286–5298, Nov. 2007.
[38] N. I. Fisher, T. Lewis, and B. J. Embleton, *Statistical Analysis of Spherical Data*. Cambridge, U.K.: Cambridge Univ. Press, 1993.
[39] K. Kitamura, A. Wakahara, H. Mizuno, Y. Baba, and K. Tomita, "Conformationally'concerted'changes in nucleotide structures. a new description using circular correlation and regression analyses," *J. Amer. Chem. Soc.*, vol. 103, no. 13, pp. 3899–3904, 1981.
[40] L. Pakula and S. Kay, "Detection performance of the circular correlation coefficient receiver," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 3, pp. 399–404, Jun. 1986.
[41] R. Kempter, C. Leibold, G. Buzsáki, K. Diba, and R. Schmidt, "Quantifying circular–linear associations: Hippocampal phase precession," *J. Neurosci. Methods*, vol. 207, no. 1, pp. 113–124, 2012.
[42] N. I. Fisher and A. J. Lee, "A correlation coefficient for circular data," *Oxford Univ. Press Behalf Biometrika Trust*, vol. 70, no. 8, pp. 27–332, Aug. 1983.
[43] J. C. Principe, *Information Theoretic Learning: Renyi's Entropy and Kernel Perspectives* (Information Science and Statistics). New York, NY, USA: Springer, 2010.
[44] W. Liu, P. P. Pokharel, and J. C. Principe, "Correntropy: A localized similarity measure," in *Proc. IEEE Int. Joint Conf. Neural Netw.*, Jul. 2006, pp. 4919–4924.
[45] (2020). INMET. *Weather Stations*. [Online]. Available: https://mapas.inmet.gov.br/
[46] N. Qian, "On the momentum term in gradient descent learning algorithms," *Neural Netw.*, vol. 12, no. 1, pp. 145–151, 1999.
[47] A. Singh and J. C. Principe, "A closed form recursive solution for maximum correntropy training," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Mar. 2010, pp. 2070–2073.

MANOEL B. L. AQUINO (Student Member, IEEE) received the B.Sc. degree in electrical engineering and the M.Sc. degree in electrical engineering and computing from the Federal University of Rio Grande do Norte (UFRN), Rio Grande do Norte, Brazil, in 2006 and 2008, respectively. Currently, he is a Professor with the Instituto Federal de Educação, Ciência e Tecnologia do Rio Grande do Norte (IFRN). His expertises include information theoretic, data processing, communication systems, and artificial intelligence.

JOÃO P. F. GUIMARÃES received the B.Sc. and M.Sc. degrees in computer engineering and electrical engineering from the Federal University of Rio Grande do Norte (UFRN), Natal, Brazil, in 2010 and 2012, respectively, where he is currently pursuing the Ph.D. degree. He is also an Associate Professor with the Federal Institute of Education, Science, and Technology of Rio Grande do Norte (IFRN). He was a Visiting Ph.D. Student with the Department of Electrical and Computer Engineering, University of Texas, El Paso, USA. His research interests include digital signal processing and machine learning.

ALUISIO I. R. FONTES received the B.S. degree in computer engineering and the M.S. and Ph.D. degrees in electrical engineering and computing from the Federal University of Rio Grande do Norte (UFRN), Rio Grande do Norte, Brazil, in 2006, 2012, and 2015, respectively. Currently, he is a Professor with the Federal Institute of Education, Science, and Technology of Rio Grande do Norte (IFRN). His expertises include information theoretic, data processing, communication systems, and artificial intelligence.

LEANDRO L. S. LINHARES received the B.Sc. degree in computer engineering and the M.S. and Ph.D. degrees in electrical engineering and computing from the Federal University of Rio Grande do Norte (UFRN), Rio Grande do Norte, Brazil, in 2003, 2010, and 2015, respectively. Currently, he is a Professor with the Federal Institute of Education, Science, and Technology of Paraiba (IFPB). His expertises include control theory, artificial intelligence, and information theoretic.

JOILSON B. A. REGO received the B.Sc. degree in computer engineering from Potiguar University (UnP), Rio Grande do Norte, Brazil, in 2007, and the M.Sc. and Ph.D. degrees in electrical and computing engineering from the Federal University of Rio Grande do Norte (UFRN), Natal, Brazil, in 2010 and 2014, respectively. He is currently a Professor with UFRN. His research interests include signal processing, differential geometry of manifolds, and machine learning.

ALLAN DE M. MARTINS (Member, IEEE) received the B.Sc., M.Sc., and Ph.D. degrees in electrical engineering from the Universidade Federal do Rio Grande do Norte (UFRN), Natal, Brazil, in 2000, 2001, and 2005, respectively. He is currently an Associate Professor with UFRN. He held a postdoctoral positions with the Computational Neuro-Engineering Laboratory (CNEL), Florida University, FL, USA, in 2006, and the Observatoire du Genéve, Geneva, Switzerland, in 2017. His expertises include digital signal and image processing, machine learning, and control systems and other related areas.