


RESEARCH

Open Access



# An adaptive kernel width convex combination method for maximum correntropy criterion

Aluisio I. R. Fontes<sup>1\*</sup> , Leandro L. S. Linhares<sup>1</sup>, João P. F. Guimarães<sup>2,3</sup>, Luiz F. Q. Silveira<sup>3</sup> and Allan M. Martins<sup>3</sup>

\*Correspondence:

[aluisio.rego@ifrn.edu.br](mailto:aluisio.rego@ifrn.edu.br)

<sup>1</sup>Federal Institute of Education, Science and Technology of Rio Grande do Norte (IFRN), BR 405, KM 154, S/N, Chico Cajá, CEP 59900-000 Pau dos Ferros, RN, Brazil  
Full list of author information is available at the end of the article

## Abstract

Recently, the maximum correntropy criterion (MCC) has been successfully applied in numerous applications regarding nonGaussian data processing. MCC employs a free parameter called kernel width, which affects the convergence rate, robustness, and steady-state performance of the adaptive filtering. However, determining the optimal value for such parameter is not always a trivial task. Within this context, this paper proposes a novel method called adaptive convex combination maximum correntropy criterion (ACCMCC), which combines an adaptive kernel algorithm with convex combination techniques. ACCMCC takes advantage from a convex combination of two adaptive MCC-based filters, whose kernel widths are adjusted iteratively as a function of the minimum error value obtained in a predefined estimation window. Results obtained in impulsive noise environment have shown that the proposed approach achieves equivalent convergence rates but with increased accuracy and robustness when compared with other similar algorithms reported in literature.

**Keywords:** Adaptive filter, Adaptive kernel width, Maximum correntropy criterion

## Introduction

Correntropy is a similarity measure based on Rényi entropy capable of extracting high-order statistical information from data, thus generalizing the correlation concept [1]. In the past few years, this concept been successfully applied in the solution of various engineering problems, such as the kernel Kalman filtering [2], adaptive time delay estimation [3], automatic modulation classification [4, 5], nonlinear auto-regressive with exogenous (NARX) for identification system [6], compressive sensing problem in an impulsive noise environment [7], signal detector in MIMO systems [8], and fuzzy neural system [9]. Training adaptive systems with cost functions such as the error correntropy criterion requires the selection of a proper kernel width. This parameter essentially controls the nature of the performance surface over which the system parameters are adapted. Therefore, it has important effects on distinct aspects e.g. convergence speed, presence of local optima, and stability of weight tracks [10].

The maximum correntropy criteria (MCC) uses correntropy as a cost function in optimization problems such as adaptive filtering. It has been shown in literature that MCC is able to achieve better performance than second-order methods [11–15]. Clearly, the use of a fixed value for the kernel width to estimate the error correntropy throughout the adaptation process would be suboptimal. However, several works fix this parameter imposing tradeoffs among convergence rate, robustness, and steady-state accuracy[10].

A series of adaptive kernel width algorithms has been proposed in order to properly choose iteratively the value for this parameter in optimization problems. An algorithm called adaptive kernel width MCC (AMCC) is proposed in [16] aiming to improve the learning speed, especially when the initial weight vector is far from being optimal. Another method called switch kernel width method of correntropy (SMCC) in [17] updates the kernel width based on the instantaneous error between the estimation and the desired signal in order to adjust such parameter for each iterations. Recently, it has been shown the technique developed in [18] called variable kernel width (VKW-MCC) is able to search for the best kernel width at each iteration, implying reduced error. This strategy is able to provide fast convergence rate and stable steady-state performance. On the other hand, the work presented in [19] report a novel robust adaptive algorithm associated to the convex combination of maximum correntropy (CCMC) with MCC employing distinct step sizes and fixed kernel width to satisfy the conflicting requirements between convergence rate and the steady-state mean square error.

Within this context, this work proposes an algorithm called adaptive convex combination kernel width method for maximum correntropy criterion (ACCMCC), which combines two MCC-based adaptive algorithms with different step sizes while adjusting the kernel size. The introduced approach is based on the gradient method, as two learning rates are used and then combined in order to keep fast convergence rate and small error. Simulation results are presented to validate the performance of ACCMCC method compared with other similar algorithms previously reported in literature, e.g., MCC, VKW-MCC, SMCC, AMCCC, and CMCC.

The remainder of the paper is organized as follows. the “[Methods](#)” section reviews the correntropy function concepts. The “[Adaptive kernel algorithms](#)” section describes in detail each kernel adaptive algorithm evaluated in this work. The “[Adaptive convex combination kernel width MCC \(ACCMCC\)](#)” section proposes an algorithm called adaptative convex combination kernel width for maximum correntropy criterion. The “[Impulsive noise](#)” section shows the noise model used by the simulations, while the “[Results and discussion](#)” section presents the performance of the proposed method compared with other similar solutions. Finally, the “[Conclusion](#)” section discusses the main contributions and results regarding the study carried out in this work.

## Methods

### Correntropy

Correntropy is a similarity measure between two arbitrary signals or random variables. Given two random variables  $X$  and  $Y$ , correntropy is defined by [1]

$$V_{\sigma}(X, Y) = E[\kappa_{\sigma}(X, Y)], \quad (1)$$

where  $E[\cdot]$  denote the expectation operator,  $k_\sigma(\cdot, \cdot)$  is any positive definite kernel with bandwidth  $\sigma$ . In this work, the kernel function is a Gaussian kernel given by

$$\kappa_\sigma(X, Y) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(X - Y)^2}{2\sigma^2}\right), \tag{2}$$

where  $\sigma$  corresponds to the kernel bandwidth, also referred to as the kernel size or kernel width. Using the Taylor series expansion of the Gaussian function in (1), it is possible to obtain:

$$V_\sigma(X, Y) = \frac{1}{\sqrt{2\pi}\sigma} \sum_{k=0}^{\infty} \frac{(-1)^k}{2^k \sigma^{2k} k!} E[(X - Y)^{2k}]. \tag{3}$$

By using the Gaussian kernel, the correntropy function becomes a weighted sum of all the even moments from a random variable  $(X - Y)$ . It is interesting to observe that the kernel size in Eq. (3) appears as a weighting parameter that controls which moments are effectively used. For sufficiently large values, the second-order moments are dominant and the measure gets closer to the correlation [20].

**Maximum correntropy criterion (MCC)**

Due to the well-know ability of correntropy to deal with outliers, it is often used as a cost function in regression problems [1]. The goal is to maximize the similarity between a given desired signal,  $d_i$ , and an estimated output,  $y_i = \mathbf{w}^T \mathbf{x}_i$ , where  $\mathbf{w}$  is the estimated weight vector and  $\mathbf{x}_i$  the input vector. In many cases, only a finite number of data  $[d_i, y_i]_{i=1}^N$  is available, as it is possible to employ a sample mean estimator of correntropy [20]:

$$\widehat{V}_\sigma = \frac{1}{N} \sum_{i=1}^N \kappa_\sigma(d_i, y_i). \tag{4}$$

By using the Gaussian kernel, a cost function  $J$  can be expressed as:

$$J = \frac{1}{\sqrt{2\pi}\sigma N} \sum_{i=1}^N \exp\left(-\frac{(d_i - y_i)^2}{2\sigma^2}\right). \tag{5}$$

Then, by maximizing (5) associated to a stochastic gradient ascent, it is possible to obtain an update rule for weights  $\mathbf{w}$  in the form:

$$\mathbf{w}_k = \mathbf{w}_{k-1} + \frac{\mu}{\sigma} \exp\left(-\frac{e_k^2}{2\sigma^2}\right) e_k \mathbf{x}_k, \tag{6}$$

where  $e_k = d_k - y_k = d_k - \mathbf{w}^T \mathbf{x}_k$  is the error in the  $k$ th algorithm iteration and  $\mu$  is the step or learning rate parameter obtained from the stochastic gradient ascent method.

**Adaptive kernel algorithms**

In most cases, replacing the mean square error for MCC as a cost function makes the algorithm robust to outliers. On the other hand, another parameter is introduced, i.e., the kernel width, which is an inherent characteristic to correntropy.

The kernel size works as a scale parameter that controls the width of the Gaussian kernel used by correntropy, being directly associated to the steady-state performance, convergence rate, and impulsive noise rejection. Since it is a free parameter, the kernel width must be chosen by the user, whose value changes according to data and application nature. Then, the definition of an optimal value for the kernel width is not an easy task [20].

In this context, many works have been proposed in order to help determining the optimal kernel width, e.g., [16–18, 21]. This work introduces a novel algorithm to address this issue, as the forthcoming section is dedicated to the detailed discussion of some methods that eventually led to the conception of the proposed approach.

**Adaptive kernel width MCC (AMCC)**

According to [16], AMCC consists in selecting the kernel as a combination of a fixed kernel bandwidth, which could be defined using the Silverman rule [22], and the prediction error  $e_k$  as expressed by:

$$\sigma_k^2 = e_k^2 + \sigma^2. \tag{7}$$

The authors in [16] showed that this approach makes the algorithm to converge faster, especially when the initial weight vector is far away from the optimal one. Besides the fast convergence rate, prominent advantages of the method lie in simplicity, no extra computational cost, and no extra free parameters are required.

**Switch kernel width MCC (SMCC)**

In order to improve the convergence rate of the method, the switch kernel width MCC (SMCC) algorithm introduced in [17] proposes a maximum adaptive kernel between the absolute value of the instantaneous error divided by two and a fixed kernel width, i.e.,

$$\sigma_k^2 = \max\left(\frac{e_k^2}{2}, \sigma^2\right), \tag{8}$$

where  $\sigma$  is a predetermined value which can be obtained using the Silverman rule [22] or any similar method or a different method.

This is another example of a simple update rule for the kernel that does not add new free parameters to the traditional MCC algorithm, although robustness is maintained.

**Variable kernel width MCC (VKW-MCC)**

The VKW-MCC algorithm calculates the kernel size at each iteration by maximizing  $\exp(-e^2/2\sigma^2)$  with respect to the kernel width  $\sigma$  [18]. For this purpose, the authors employ a modified cost function to reduce the interference of the kernel size. Instead of making  $J_n = E[G_\sigma(e)]$ , the following statement is considered:

$$J_k = E[\sigma^2 G_\sigma(e)]. \tag{9}$$

Using the gradient ascent approach, the modified MCC algorithm is given as

$$\mathbf{w}_{k+1} = \mathbf{w}_k + \mu \exp\left(-\frac{e_k}{2\sigma^2}\right) e_k \mathbf{x}_k. \tag{10}$$

Assuming that the noise is not impulsive, the work developed in [18] has also shown that the optimal kernel size in the  $k$ th iteration is given by:

$$\sigma_k = k_\sigma |e_k|, \tag{11}$$

where  $k_\sigma$  is a positive constant. In order to ensure a robust response to impulsive noise [23], the VKW-MCC method computes  $E[|e_k|]$  instead of  $|e_k|$  in Eq. (11), i.e.,

$$\bar{e}_k = \tau \bar{e}_{k-1} + (1 - \tau) \min(A_{e,k}). \tag{12}$$

where  $\tau$  is a smoothing factor that can assume any value between 0 and 1, and  $A_{e,k}$  is a set of values  $|e_k|$  in the form:

$$A_{e,k} = [ |e_k| \ |e_{k-1}| \ \dots \ |e_{k-N_w+1}| ]. \tag{13}$$

being  $N_w$  the length of the estimation window. Then, Eq. (11) can be rewritten as:

$$\sigma_k = k_\sigma \bar{e}_k. \tag{14}$$

The authors in [18] also mention that the VKW-MCC algorithm lacks robustness when  $\sigma_k$  is too large. To prevent this from happening, the kernel size must be within an interval  $[0, \sigma_0]$ .

**Adaptive convex combination kernel width MCC (ACCMCC)**

Recently, the authors from [19] defined a novel robust adaptive algorithm by convexly combining two MCC-based adaptive algorithms as in Eq. (6) with different step-sizes parameters  $\mu$ , a larger and a smaller one. The two adaptive filters are then combined based in the error. However, this approach does not use any adaptive kernel technique in order to update the kernel size used by correntropy, which is another free parameter that could be optimized.

Given the recognized importance to properly define the kernel size parameter in algorithms based on correntropy, this work proposes the combination of the kernel size optimization strategy from [18] with the CMCC algorithm described in [19] in order to obtain an new algorithm with improved performance, which we defined as adaptive convex combination kernel width MCC (ACCMCC). This novel method is based on the convex combination of two adaptive filters with distinct convergence rates. This is achieved by setting a larger step  $\mu_1$  for the faster filter, and a smaller step  $\mu_2$  for the slower one. The parameter vectors  $\mathbf{w}_{jk}, j = 1, 2$ , are then updated according to the following rule:

$$\mathbf{w}_{jk} = \mathbf{w}_{j(k-1)} + \mu_j \exp\left(-\frac{e_{jk}^2}{2\sigma_{jk}^2}\right) e_{jk} \mathbf{x}_k. \tag{15}$$

The outputs of the adaptive filters employed in ACCMCC are combined with a so-called mix parameter  $\lambda_k$  to obtain the overall output as:

$$y_k = \lambda_k y_{1k} + (1 - \lambda_k) y_{2k}, \tag{16}$$

where  $y_{1k} = \mathbf{w}_{1k}^T \mathbf{x}_k$ ,  $y_{2k} = \mathbf{w}_{2k}^T \mathbf{x}_k$ , and  $\mathbf{w}_{1k}$  and  $\mathbf{w}_{2k}$  are the weight vectors of the filter with large step size and small step size, respectively. Once the overall output is calculated, the overall filter output error can be obtained as:

$$e_k = d_k - y_k. \tag{17}$$

Analogously to (16), the overall filter weight vector can be determined by combining  $\mathbf{w}_{1k}$  and  $\mathbf{w}_{2k}$  with the mixing parameter, resulting in:

$$\mathbf{w}_k = \lambda_k \mathbf{w}_{1k} + (1 - \lambda_k) \mathbf{w}_{2k}. \tag{18}$$

The desired behavior for the mixing parameter  $\lambda_k$  in (18) lies in keeping it as close to 1 as possible when the algorithm starts, while it should tend to 0 as the ACCMCC begins to

converge. According to [19], this behavior can be obtained by using a sigmoidal function to update parameter  $\lambda_k$  i.e.

$$\lambda_k = \frac{1}{(1 + \exp(-\alpha_k))}. \tag{19}$$

Expression 19, denotes that  $\alpha_k$  controls the update rate of  $\lambda_k$ . In order to maintain robustness to impulsive noise, the adaptation of  $\alpha_k$  is expressed by an adaptive rule, which is obtained from the correntropy maximization with the stochastic ascent gradient method:

$$\begin{aligned} \alpha_{k+1} &= \alpha_k + \mu_\alpha \sigma^2 \frac{\partial \exp\left(-\frac{e_k^2}{2\sigma^2}\right)}{\partial \alpha_k} \\ &= \alpha_k + \eta_k \exp\left(-\frac{e_k^2}{2\sigma^2}\right) e_k, \end{aligned} \tag{20}$$

where  $\mu_\alpha$  is the adaptation step and  $\eta_k$  corresponds to a simplified mathematical notation in the form:

$$\eta_k = \mu_\alpha \lambda_k (1 - \lambda_k) (y_{1k} - y_{2k}). \tag{21}$$

Similarly to the former CMCC method, the range for  $k$  is restricted to a symmetric interval  $[-4, 4]$ .

The proposed ACCMCC is summarized as Algorithm 1, where it is possible to notice that the VKW-MCC mechanism is adopted to adjust the kernel width for each one of the adaptive filters. Therefore, it is possible to state that ACCMCC is a hybrid approach derived from the combination of CMCC and VKW-MCC.

### Impulsive noise

Several statistical models have been proposed so far to describe impulsive noise. One of the most popular statistical distributions for this purpose is the  $\alpha$ -stable distribution [24]. The fundamental principles associated with alpha-stable modeling are based on the generalized central limit theorem [25]. By adjusting its respective free parameters, it is possible to generate various probability distribution functions, such as Gaussian, CauchyLorentz, or Lèvy.

In traditional signal processing, Gaussian noise is commonly used in system modeling due to its simple probability distribution function (PDF). With the rapid development of technology, it is nearly impossible to ignore the impact of nonGaussian noise, thus leading to the adoption of a generalized distribution. It is reported in literature as the so-called alpha-stable distribution, which is a unique solution that satisfies the generalized central limit theorem. Due to the lack of closed formulas for density and distribution functions, the alpha-stable distribution is commonly expressed in terms of a characteristic function [25]. Firstly, let us consider that the characteristic function of a random variable  $X$  has the form described in [26] as:

$$\phi(\omega, \alpha, \beta, \gamma, \delta) = \exp\{\gamma^\alpha (-|\omega|^\alpha + j\omega\Theta(\omega, \alpha, \beta)) + j\delta\omega\}, \tag{22}$$

where

$$\Theta(\omega, \alpha, \beta) = \begin{cases} \beta|\omega|^{\alpha-1} \tan\left(\frac{\pi\alpha}{2}\right) & \alpha \neq 1 \\ -\beta\frac{2}{\pi} \ln|\omega| & \alpha = 1. \end{cases} \tag{23}$$

---

**Algorithm 1** ACCMCC

---

Free parameters:  $\tau = 0.98$ ,  $\mu_1 = 0.05$ ,  $\mu_2 = 0.01$ ,  $\sigma_0 = 20$ ,  $k_\sigma = 200$

Initialization values:  $\alpha_0 = \lambda_0 = 1$ ,  $\mathbf{w}_{j0} = \mathbf{0}$ ,  $\bar{e}_{j0} = |d_1|$

```

1: for each iteration  $k$  do
2:   for  $j = 1$  to 2 do
3:      $y_{jk} = \mathbf{w}_{1k}^T \mathbf{x}_k$ 
4:      $e_{jk} = d_k - y_{jk}$ 
5:      $A_{e_j,k} = [ |e_{jk}| \ |e_{j(k-1)}| \ \dots \ |e_{j(k-N_w+1)}| ]$ 
6:      $\bar{e}_{jk} = \tau \bar{e}_{j(k-1)} + (1 - \tau) \min(A_{e_j,k})$ 
7:     if  $\bar{e}_{jk} \geq \frac{\sigma_0}{k_\sigma}$  then
8:        $\sigma_{jk} = \sigma_0$ 
9:     else
10:       $\sigma_{jk} = k_\sigma \bar{e}_{jk}$ 
11:    end if
12:     $\mathbf{w}_{jk} = \mathbf{w}_{j(k-1)} + \mu_j \exp\left(-\frac{e_{jk}^2}{2\sigma_{jk}^2}\right) e_{jk} \mathbf{x}_k$ 
13:  end for
14:   $y_k = \lambda_k y_{1k} + (1 - \lambda_k) y_{2k}$ 
15:   $e_k = d_k - y_k$ 
16:   $\lambda_k = \frac{1}{(1 + \exp(-\alpha_k))}$ 
17:   $\eta_k = \mu_\alpha \lambda_k (1 - \lambda_k) (y_{1k} - y_{2k})$ 
18:   $\alpha_{k+1} = \alpha_k + \eta_k \exp\left(-\frac{e_k^2}{2\sigma^2}\right) e_k$ 
19:   $\alpha_{k+1} = \min\{\max\{\alpha_{k+1}, -4\}, 4\}$ 
20:   $\mathbf{w}_k = \lambda_k \mathbf{w}_{1k} + (1 - \lambda_k) \mathbf{w}_{2k}$ 
21: end for

```

---

It can be seen that  $\alpha$  is the characteristic exponent in (22), which mainly characterizes the intensity of the impulsive noise ( $0 \leq \alpha \leq 2$ ). This parameter is responsible for controlling the impulsivity of the density function, i.e., the smaller the value of alpha, the longer the distribution tail will be. Particularly for  $\alpha = 2$ , it determines the sign and degree of asymmetry. If  $\beta = 0$ , the distribution is symmetric. Finally,  $\gamma$  is the dispersion parameter, which is similar to the variance  $\sigma^2$  in a Gaussian distribution and capable of measuring the degree of discretization of sample data.

**Results and discussion**

In this section, Monte-Carlo simulations are carried to verify the theoretical analysis and validate the behavior of the ACCMCC algorithm. The novel method is used in a system identification problem, while its performance is compared with those regarding the traditional MCC with fixed kernel width and four other methods reported in literature i.e. AMCC, SMCC, CMCC, and VKW-MCC, which use adaptive kernel widths. The system adopted in the simulation tests has nine taps and its respective vector of weight parameters is defined as in [27]

$$\mathbf{w} = [0.1, 0.2, 0.3, 0.4, 0.5, 0.4, 0.3, 0.2, 0.1] \tag{24}$$

**Table 1** Alpha-stable parameters employed in simulation tests

Distribution parameters	Value
Location ( $\delta$ )	0
Skewness ( $\beta$ )	0
Characteristic exponent ( $\alpha$ )	1.2 to 2
GSNR	12 to 20 dB

The parameter settings are listed in Table 1. Two different nonGaussian noise signals are individually used to contaminate the signal of the system identified by the MCC-based algorithms. Besides, each one of them corresponds to a distinct scenario. The performance of ACCMCC is compared with those regarding other similar methods using the weighted signal-to-noise ratio (WSNR) metric, which is expressed as [13]:

$$WSNR_{db} = 10 \log_{10} \left( \frac{\bar{\mathbf{w}}^T \bar{\mathbf{w}}}{(\bar{\mathbf{w}} - \mathbf{w}_n)^T (\bar{\mathbf{w}} - \mathbf{w}_n)} \right) \tag{25}$$

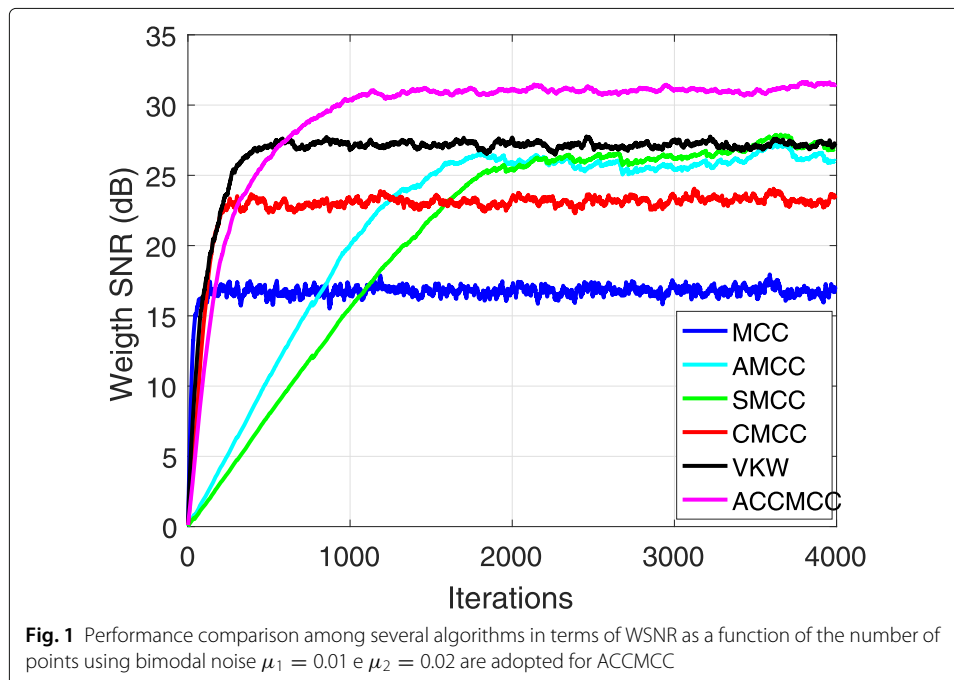
It is worth to mention that the convergence curves for WSNR are usually used to compare the convergence performance of distinct algorithms.

**Scenario 1—System under bimodal noise**

In the first scenario, the system is under a nonGaussian noise described by a bimodal signal given by [1]:

$$0.95 \mathcal{N}(0, 10^{-4}) + 0.05 \mathcal{N}(0, 10) \tag{26}$$

Figure 1 presents a COMPARISON among several methods under this condition. After the 500th sample, ACCMCC achieves improved performance if compared to the other methods. In steady-state, ACCMCC becomes stable when  $WSNR = 33dB$ . It can be also stated that the proposed method presents good convergence rate.





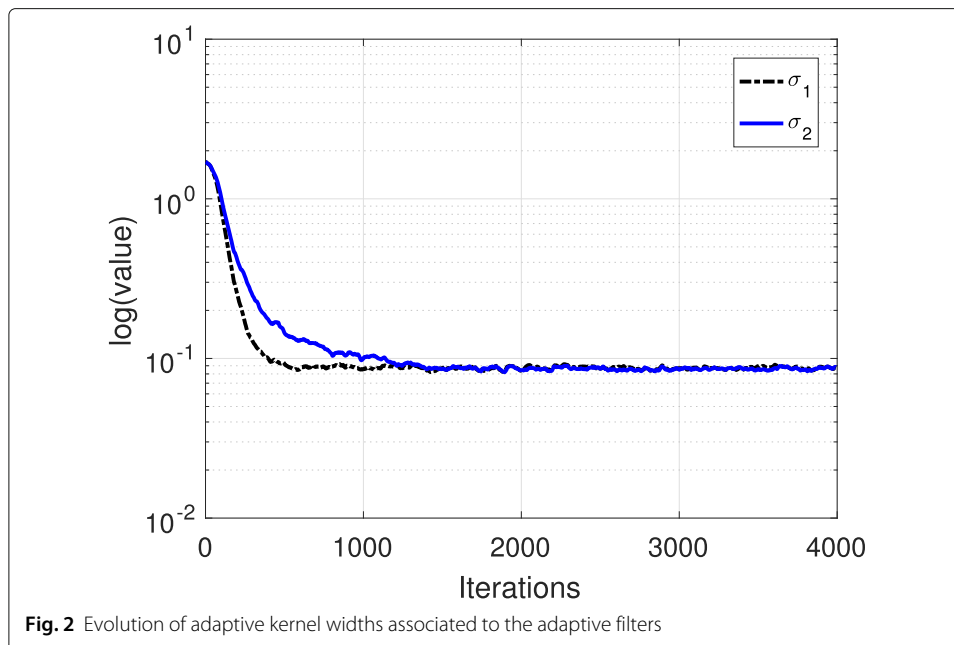


Figure 2 shows the values assumed by the adaptive kernel width for both adaptive filters used in ACCMCC. The initial values are initially large towards the search for a more convex space, i.e., where only few local minimums exist. When the algorithm is close to the solution, the kernel size is adjusted to find a more accurate solution. In particular, ACCMCC comprises a combination of two MCC adaptive filters with distinct kernel widths and steps.

**Scenario 2—System under alpha-stable noise**

The alpha-stable noise defined by 22 is used in the second scenario to contaminate the system under study. Tests using different noise intensities are performed to evaluate the behavior of ACCMCC. In this case, the generalized signal-to-noise ratio (GSNR) is employed and defined as [13]:

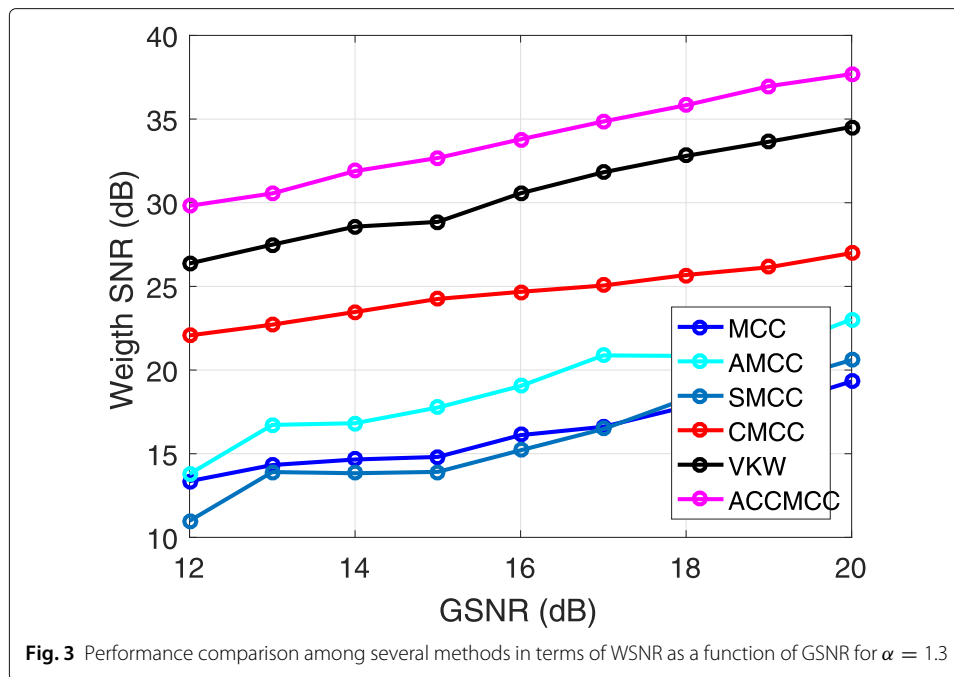
$$GSNR = 10\log_{10} \left( \frac{1}{\gamma M} \sum_{t=1}^M |s(t)^2| \right), \tag{27}$$

where  $s$  is the original signal and recalling that  $\gamma$  is the dispersion parameter from the alpha-stable distribution.

Different values of GSNR are used in the simulation tests, which range from 12 to 20 dB as shown in Fig. 3. All MCC-based algorithms present reduced efficiency as GSNR decreases. However, ACCMCC and CMCC are the only two methods with good performance over the entire range adopted for GSNR. The specifications for the alpha-stable parameters are listed in Table 2.

The performance of ACCMCC is compared with those regarding other MCC-based methods while varying GNSR from 12 dB to 20 dB according to Fig. 3. In particular, the method proposed in this work achieves the best results for all values assumed by GSNR.

Figure 4 represents the performance of ACCMCC when the system is under an alpha-stable noise as  $\alpha = 1.3$  and  $GSRN = 15$ . It is possible to notice that ACCMCC becomes



stable with about  $WSNR = 25\text{dB}$  after the 500th iteration, being this the best result achieved among all analyzed methods.

It worth to mention that the performances of AMCC and SMCC are significantly improved if compared to that regarding ACCMCC as the noise becomes Gaussian, i.e.,  $\alpha$  tends to 2. According to Fig. 5, the proposed algorithm still outclasses MCC, CMCC, AMCC, SMCC, and VKW-MCC in terms of convergence rate, steady-state misalignment, and tracking performance in nonGaussian environments.

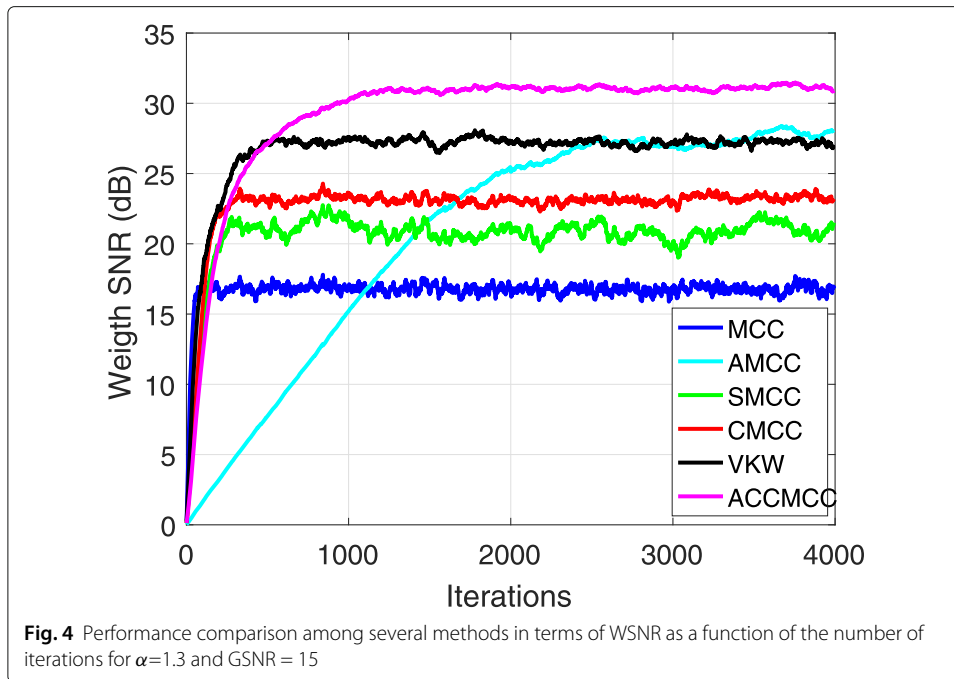
According to the previous analysis, the results demonstrate that the proposed method presents better accuracy and robustness in both noisy scenarios.

**Conclusion**

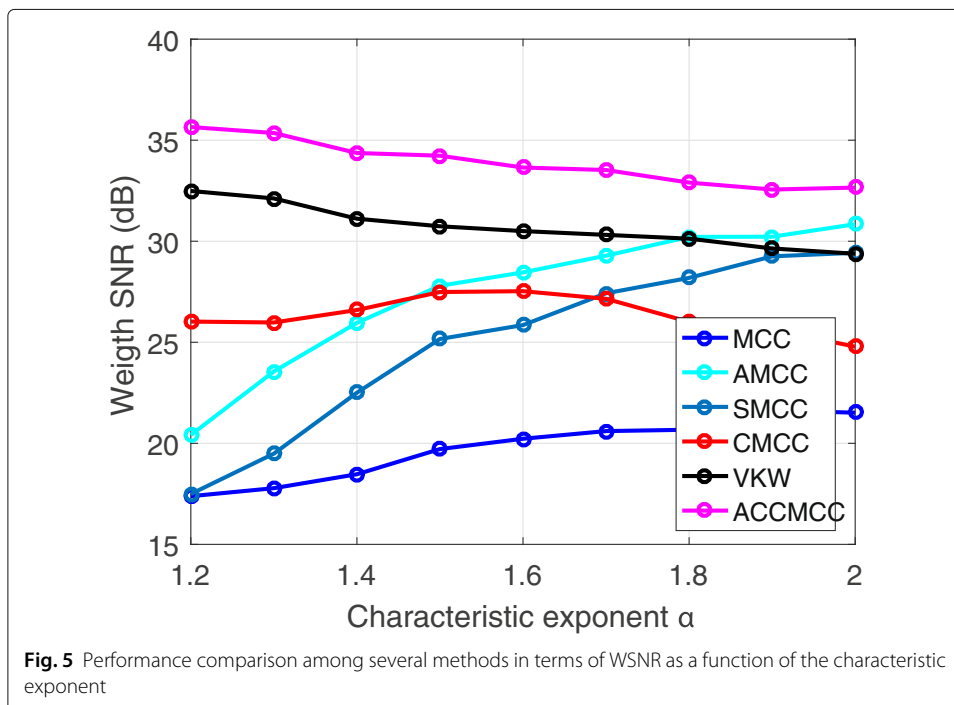
This work has presented a novel algorithm called adaptive convex combination maximum correntropy criterion, which is based on the VKW-MCC and CMCC methods. Since the kernel is used to calculate the value assumed by the cost function, modifying the kernel width implies changing the estimator of the cost function. The proposed method to adapt the kernel size can also be seen as a mechanism to use different cost functions depending on the nature of the data seen by the adaptive system.

**Table 2** Relationship between kernel size and step size

Methods	Scenario 1	Scenario 2
ACCMCC	$\sigma = 3, \mu_1 = 0.01, \mu_2 = 0.02$	$\sigma = 2, \mu_1 = 0.01, \mu_2 = 0.02$
VKW-MCC	$\sigma = 3, \mu_1 = 0.03$	$\sigma = 2, \mu_1 = 0.03$
SMCC	$\sigma = 4, \mu_1 = 0.015$	$\sigma = 2, \mu_1 = 0.015$
AMCC	$\sigma = 1, \mu_1 = 0.0025$	$\sigma = 2, \mu_1 = 0.018$
CMCC	$\sigma = 3, \mu_1 = 0.01, \mu_2 = 0.02$	$\sigma = 2, \mu_1 = 0.01, \mu_2 = 0.02$
MCC	$\sigma = 3, \mu_1 = 0.03$	$\sigma = 1.2, \mu_1 = 0.069$



In order to provide robustness against impulsive interference, the adaptive rule of the mixing factor is derived by maximizing correntropy with the stochastic gradient ascent method with two distinct learning rates. This approach has presented improved performance than those achieved by other existing methods, especially in the presence of highly impulsive noise. Simulation results have also shown that, even including new free parameters to the method, the proposed algorithm was able to effectively eliminate



interference caused by impulsive noise conditions while also improving the estimation accuracy.

#### Abbreviations

ACCMCC: Adaptive convex combination maximum correntropy criterion; AMCC: Adaptive kernel width maximum correntropy criterion; CCMCC: Convex combination maximum correntropy criterion; GSNR: Generalized signal-to-noise ratio; MCC: Maximum correntropy criterion; SMCC: Switch kernel width method of correntropy; SNR: Signal-to-noise ratio; VKW-MCC: Variable kernel width-maximum correntropy criterion; WSNR: Weighted signal-to-noise ratio

#### Acknowledgements

The authors would like to thank the Federal Institute of Rio Grande do Norte and Federal University of Rio Grande do Norte by technical support.

#### Authors' contributions

The authors declare that they all contributed to the manuscript. All authors read and approved the final manuscript.

#### Funding

This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001

#### Availability of data and materials

Data and MATLAB® source code are available from the corresponding author upon request.

## Declarations

#### Competing interests

The authors declare that they have no competing interests.

#### Author details

<sup>1</sup>Federal Institute of Education, Science and Technology of Rio Grande do Norte (IFRN), BR 405, KM 154, S/N, Chico Cajá, CEP 59900-000 Pau dos Ferros, RN, Brazil. <sup>2</sup>Federal Institute of Education, Science and Technology of Rio Grande do Norte (IFRN), BR 406, Km 73, nº 3500, Perímetro Rural, 59550-000 João Câmara, RN, Brazil. <sup>3</sup>Department of Computer Engineering and Automation (DCA), Federal University of Rio Grande do Norte (UFRN), UFRN Campus Universitário Lagoa Nova, 59078-970 Natal, RN, Brazil.

Received: 29 October 2018 Accepted: 28 February 2021

Published online: 23 March 2021

#### References

1. Santamaria I, Pokharel P, Principe J (2006) Generalized correlation function: definition, properties, and application to blind equalization. *IEEE Trans Sig Process* 54(6):2187–2197. <https://doi.org/10.1109/TSP.2006.872524>
2. Dang L, Chen B, Wang S, Gu Y, Principe J (2019) Kernel Kalman filtering with conditional embedding and maximum correntropy criterion. *IEEE Trans Circ Syst I Regular Pap* 66(11):4265–4277
3. Jin F, Qiu T (2019) Adaptive time delay estimation based on the maximum correntropy criterion. *Digit Sig Process* 88:23–32
4. Câmara T, Lima A, Lima B, Fontes A, Martins A, Silveira L (2019) Automatic modulation classification architectures based on cyclostationary features in impulsive environments. *IEEE Access* 7:138512–138527
5. Fontes A, Rego J, Martins AdM, Silveira L, Principe J (2017) Cyclostationary correntropy: definition and applications. *Expert Syst Appl* 69:110–117
6. Araújo IB, Guimarães J, Fontes A, Linhares L, Martins A, Araújo F (2019) Narx model identification using correntropy criterion in the presence of non-Gaussian noise. *J Control Autom Electr Syst* 30(4):453–464
7. Guimarães J, Da Silva F, Fontes A, Von Borries R, Martins A (2019) Complex correntropy applied to a compressive sensing problem in an impulsive noise environment. *IEEE Access* 7:151652–151660
8. De Souza P, Fontes A, De Souza V, Silveira L (2019) A novel signal detector in MIMO systems based on complex correntropy. *IEEE Access* 7:137517–137527
9. Bao R-J, Rong H-J, Angelov P, Chen B, Wong P (2017) Correntropy-based evolving fuzzy neural system. *IEEE Trans Fuzzy Syst* 26(3):1324–1338
10. Singh A, Principe J (2010) Kernel width adaptation in information theoretic cost functions. In: 2010 IEEE International Conference on Acoustics, Speech and Signal Processing. IEEE, Dallas, pp 2062–2065. <https://doi.org/10.1109/ICASSP.2010.5495035>
11. He R, Zheng W, Hu B (2011) Maximum correntropy criterion for robust face recognition. *IEEE Trans Pattern Anal Mach Intell* 33(8):1561–1576. <https://doi.org/10.1109/TPAMI.2010.220>
12. Hakimi S, Abed Hodtani G (2018) Generalized maximum correntropy detector for non-gaussian environments. *Int J Adapt Control Sig Process* 32(1):83–97. <https://doi.org/10.1002/acs.2827>
13. Araújo J, Fontes A, Rego J, de M. Martins A, Principe J (2017) Complex correntropy: probabilistic interpretation and application to complex-valued data. *IEEE Sig Process Lett* 24(1):42–45. <https://doi.org/10.1109/LSP.2016.2634534>
14. Fontes A, Martins A, Silveira L, Principe J (2015) Performance evaluation of the correntropy coefficient in automatic modulation classification. *Expert Syst Appl* 42(1):1–8. <https://doi.org/10.1016/j.eswa.2014.07.023>

15. Wang S, Dang L, Wang W, Qian G, Tse C (2018) Kernel adaptive filters with feedback based on maximum correntropy. *IEEE Access* 6:10540–10552. <https://doi.org/10.1109/ACCESS.2018.2808218>
16. Wang W, Zhao J, Qu H, Chen B, Principe J (2017) Convergence performance analysis of an adaptive kernel width MCC algorithm. *AEU Int J Electron Commun* 76:71–76. <https://doi.org/10.1016/j.aeue.2017.03.028>
17. Wang W, Zhao J, Qu H, Chen B, Principe J (2015) A switch kernel width method of correntropy for channel estimation. In: 2015 International Joint Conference on Neural Networks (IJCNN). IEEE, Killarney, pp 1–7. <https://doi.org/10.1109/IJCNN.2015.7280632>
18. Huang F, Zhang J, Zhang S (2017) Adaptive filtering under a variable kernel width maximum correntropy criterion. *IEEE Trans Circ Syst II Express Briefs* 64(10):1247–1251. <https://doi.org/10.1109/TCSII.2017.2671339>
19. Shi L, Lin Y (2014) Convex combination of adaptive filters under the maximum correntropy criterion in impulsive interference. *IEEE Sig Process Lett* 21(11):1385–1388. <https://doi.org/10.1109/LSP.2014.2337899>
20. Liu W, Pokharel P, Principe J (2007) Correntropy: properties and applications in non-Gaussian signal processing. *IEEE Trans Sig Process* 55(11):5286–5298. <https://doi.org/10.1109/TSP.2007.896065>
21. Wang W, Zhao J, Qu H, Chen B, Principe J (2015) An adaptive kernel width update method of correntropy for channel estimation. In: 2015 IEEE International Conference on Digital Signal Processing (DSP). IEEE, Singapore, pp 916–920. <https://doi.org/10.1109/ICDSP.2015.7252010>
22. Silverman B (1986) *Density Estimation for Statistics and Data Analysis*. Chapman & Hall/CRC, London
23. Huang F, Zhang J, Zhang S (2017) NLMS algorithm based on a variable parameter cost function robust against impulsive interferences. *IEEE Trans Circ Syst II Express Briefs* 64(5):600–604. <https://doi.org/10.1109/TCSII.2016.2594069>
24. Georgiou P, Tsakalides P, Kyriakakis C (1999) Alpha-stable modeling of noise and robust time-delay estimation in the presence of impulsive noise. *IEEE Trans Multimed* 1(3):291–301
25. Shao M, Nikias C (1993) Signal processing with fractional lower order moments: stable processes and their applications. *Proc IEEE* 81(7):986–1010
26. Samorodnitsky G, Taqqu M (1994) *Stable Non-gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, New York
27. Singh A, Principe J (2009) Using correntropy as a cost function in linear adaptive filters. In: 2009 International Joint Conference on Neural Networks. IEEE, Atlanta, pp 2950–2955. <https://doi.org/10.1109/IJCNN.2009.5178823>

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Submit your manuscript to a SpringerOpen<sup>®</sup> journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

---

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)

---