

A fuzzy clustering method for periodic data, applied for processing turbomachinery beamforming maps

Bence Tóth^{a,*}, János Vad^a

^a*Department of Fluid Mechanics, Faculty of Mechanical Engineering, Budapest University of Technology and Economics. Bertalan Lajos u. 4-6, H-1111 Budapest, Hungary*

Abstract

In the present paper, the fuzzy c-means method is extended, and an algorithm is proposed for fuzzy clustering of data lying in a feature space of arbitrary dimensions, with one of them being periodic. To aid in determining the optimal number of clusters, the Xie-Beni validity index is extended, to account for the periodicity. Furthermore, the relative weights of the dimensions in the calculation of distances are investigated. The method is incorporated into a procedure for processing turbomachinery beamforming maps. Thus, an objective, robust way of identifying the sound sources being present in such machines is obtained. These properties are ensured by selecting the required parameters through parameter studies. Presented through a case study, the method is used to determine the most significant sound source mechanisms in an axial fan.

Keywords: axial flow fan, beamforming, fuzzy clustering, periodicity

2010 MSC: 00-01, 99-00

Nomenclature

Latin letters

*Corresponding author

Email address: tothbence@ara.bme.hu (János Vad)

a	element of the weighting matrix
\mathbf{A}	weighting matrix
B	number of fan blades
c	number of fuzzy clusters
d	distance metric
D	number of data set dimensions
f	frequency
i	index of data points
\mathbf{I}	identity matrix
j	index of clusters
J	clustering objective function
k	period tracking variable
l	running index
m	fuzzifier
N	number of all data points to be clustered
p	dimensionless amplitude
\mathbf{p}	periodic offset vector
P	period
R	dimensionless radius
s	iteration step
w	membership value
\mathbf{x}	generic vector
\mathbf{y}	generic vector
Greek letters	
β	angular position of base source
γ	arbitrary periodic quantity
ϵ	exit criterion residual
μ	cluster centre
σ	cluster amplitude
ξ	data point for clustering
Superscripts and subscripts	

A	weighted using matrix \mathbf{A}
min	minimum
max	maximum
P	periodic distance measure
T	transpose

Abbreviations

AMP & ADC	Amplifier and analogue-to-digital converter
FCM	Fuzzy c-means clustering
FCM4DD	Fuzzy c-means clustering for directional data
LE	Leading edge
PAM	Phased array microphone
PC	Personal computer
ROSI	Rotating source identifier
SPL	Sound pressure level
TE	Trailing edge
XB	Xie-Beni index

10

Others

$\hat{\square}$	normalised to the 0 ... 1 range
-----------------	---------------------------------

1. Introduction

The noise emitted by turbomachinery is a serious issue. To reduce it, the noise generation mechanisms have to be identified, understood, and controlled. Acoustic beamforming can effectively be used for indicating the areas where the strongest sources of noise are present [1, 2, 3]. Some approaches can handle rotating systems [4, 5, 6, 7], enabling the investigation of turbomachinery. Beamforming measurements are often carried out on aero-engines due to noise emission limits becoming more strict. In the recent years, investigation of low-speed axial fans has begun, as they operate in larger numbers and closer to people [8, 9, 10]. Combined with empirical analysis and numerical simulations, as done in [11, 12, 13, 14, 15], the information regarding source locations can aid in understanding noise generation mechanisms. Then an iterative redesign

process can be started, within which, by modifying the blade geometry, the
25 noise emission can be reduced.

The results of beamforming are usually presented in the form of contour
maps over the target, in the present case, over the rotor. These are called
beamforming maps. These maps are usually interpreted visually. This process
however includes subjective elements. Furthermore, such maps are usually gen-
30 erated for several frequency bands, in which case, drawing conclusions based on
them is even more difficult.

The situation is further complicated by the presence of point spread func-
tions in the maps. Due to these, a point-like source appears with a spatial ex-
tension, and close-by sources become difficult to separate. Furthermore, peaks
35 may appear at locations where in reality no sources are present. These effects
are usually treated using deconvolution methods, such as CLEAN-SC [16] or
DAMAS [17], however, in some cases, these algorithms may not be applicable.
Yet, the aforementioned problems have to be tackled in some way.

To do that, a novel algorithmic processing method was proposed in [18,
40 19]. It aimed at the identification of noise generation mechanisms through a
data reduction and clustering procedure. This was illustrated in a case study,
and found adequate, however, still contained a subjective choice due to the
periodicity of the blading.

The main contribution of the present paper is to introduce a method for
45 fuzzy clustering, applicable when one of the dimensions of the data set has
periodic attributes. This is applied for the analysis of turbomachinery beamform
maps as part of the aforementioned algorithm [19], aimed at objective and
reproducible identification of noise source mechanisms. The extended algorithm
is applied onto the beamform maps of an axial fan, to identify the dominant
50 noise generating mechanisms.

In the present application, the novel procedure ensures that clustering takes
the periodicity of the blade passages into account, even if only one blade passage
is used for data analysis. However, it may be applied for any kind of data, where
a periodic dimension is present.

55 The structure of the paper is as follows. First, the complete method for
source identification is described. Second, the original fuzzy clustering algo-
rithm is presented, the meaning and the difficulty of clustering in the presence
of periodic dimensions is explained, and the modified clustering algorithm is
introduced. Third, details about the case study are given. Fourth, the results
60 of applying the new algorithm onto the beamforming data are presented and
compared to those of the previous method. Finally, the conclusions of the work
are summarised.

2. Noise source identification method

To investigate and analyse turbomachinery noise, a novel algorithmic ap-
65 proach was proposed in [18, 19]. It aimed at identifying broadband noise gen-
eration mechanisms in an axial fan, in an objective manner. The method was
designed to decrease the effects of both measurement noise and side lobes, with-
out deconvolution. The reason for omitting deconvolution methods was that a
microphone array with generic microphone placement was used in the measure-
70 ments, and thus the beamform maps were created using the Rotating Source
Identifier (ROSI) [4]. As this method does not provide a cross spectral ma-
trix that would describe the source effects, the widely applied deconvolution
methods were not applicable.

The approach was illustrated on beamforming data taken from the tip radius
75 of an axial fan. The significant features of the beamforming maps were repre-
sented as points in the circumferential angle (β) – frequency (f) plane $[\beta, f]$,
thus the identified *base sources* were described by their location around the ro-
tor, and by frequency of their emitted sound. Those data points that are likely
signatures of the same noise generation mechanism, were grouped through a
80 clustering procedure. The identified clusters were then assigned to known noise
generation mechanisms based on previous literature.

In the previous paper [19], the algorithm was applied to one blade passage
only, as the periodicity of the noise sources was assumed and enforced in a filter-

ing process. This was done to reduce sidelobe, large-scale turbulence ingestion,
85 inflow irregularity, local geometric irregularity, and measurement uncertainty
effects.

The steps of the method were the following.

1. For each investigated frequency band, a circumferential source strength
distribution was extracted from the beamforming maps around the rotor
90 at a selected radius.
2. The distributions obtained in the previous step were expanded into spatial
Fourier series.
3. From each Fourier series, only those modes were considered, whose cir-
cumferential wave number conformed to the periodicity of the blading.
95 This means that the number of periods of the harmonic wave in one blade
passage is a positive integer. Through this step, noise sources not directly
related to the rotor are filtered out. This way, the effects of those peaks
in the maps are reduced, that do not have spatial periodicity: e.g. side
lobes, large-scale turbulence ingestion noise, and other artefacts.
- 100 4. Important components were selected using a significance criterion based on
their p modal amplitudes. The significance limit was the largest amplitude
in the spectrum found at a circular wave number that does not agree with
the periodicity of the blades, i.e. the largest amplitude of those modes,
that have non-integer periods inside one blade passage. This reduces the
105 number of data points, and further reduces measurement uncertainties.
5. Each significant mode was considered as a *base source*. A base source is a
point in the circumferential angle – frequency plane $[\beta, f]$.
6. Clusters were created from the base sources, based on the distances be-
tween them, using a fuzzy clustering procedure allowing for overlapping.
- 110 7. The source mechanisms corresponding to the clusters were identified.

Application details and further explanation for the individual steps can be found
in [19].

As mentioned above, only those sources are considered important, that conform to the periodicity of the blade passages. This periodicity was enforced through the Fourier analysis and filtering. Therefore, choosing one blade passage for analysis and clustering was sufficient, as they all contain the same base sources. However, one had to choose the limits of the *analysis window*, i.e. the part of the $[\beta, f]$ plane in which clusters are sought for. Physical arguments were followed in the f direction, but in the angular β direction, due to the repeating blade passages, no objective decision could be made as to where a blade passage periodic noise pattern begins. After a visual inspection, the beginning of the analysis window, $\beta = 0$ was assigned to the blade trailing edge. This choice however could have significant effects on the results of clustering, which should be avoided.

A further problem regards the calculation of distances required by the clustering method. Even though the quantities are normalised before clustering, the dimensions of β and f are different. Therefore no physical reasoning can be given as to how to calculate the distance between two points in this plane.

To overcome these problems, a modified fuzzy clustering method is proposed, being able to take periodicity in the data set into account. This way, the beginning of the analysis window, $\beta = 0$, can be assigned arbitrarily, as the resulting clusters will not depend on that. Besides the present case, the method will suit any application, where one of the dimensions is periodic. Furthermore, the effects of different dimensions in the distance calculations will be investigated in a case study. In this paper, the method is applied not only onto the tip radius, but at other radii in the rotor annulus area, as well.

3. Clustering

Clustering is a process of assigning data points to groups, based on their similar attributes. The general task is the following. Knowing the ξ_i , $i = 1 \dots N$ data points, lying in the so-called *feature space* of D dimensions, determine the prescribed c number of cluster centres μ_j , $j = 1 \dots c$, such that an objective

function, the error regarding the replacement of the data points with the cluster centres, is minimised. For the sake of simplicity, the data points are assumed to be normalised onto the range 0, ..., 1 in each dimension.

145 This section summarizes the original fuzzy clustering method, the properties of periodic data, and introduces the modified clustering algorithm that accounts for a periodic dimension.

3.1. Fuzzy c-means clustering

Several clustering methods have been proposed. Possibly the simplest algorithm is the *k-means* method [20]. In this case, all data points are assigned to the closest cluster, while the cluster centres are determined as the weighted average of the assigned data points. Since one data point belongs to one cluster only, this clustering method is termed *crisp*.

155 Fuzzy c-means clustering (FCM) [21] is a method similar to k-means, but in this case, the clusters are considered fuzzy. This means that one data point may belong to more clusters, and therefore the clusters may overlap.

As noted before, clusters are identified by minimising an objective function. In FCM, this is the sum of the weighted squared distances between the cluster centres and the data points. The objective function is shown in Eq. (1). The aim is to find cluster centres $\boldsymbol{\mu}_j$, $j = 1 \dots c$ that minimise this expression.

$$J = \sum_{i=1}^N \sum_{j=1}^c w_{i,j}^m d^2(\boldsymbol{\xi}_i, \boldsymbol{\mu}_j) \quad (1)$$

In Eq. (1), d denotes the distance measure between any two points in the feature space. This will be investigated in detail later. $w_{i,j} \in [0, 1]$ values are the so-called *membership values*, quantifying how well a data point $\boldsymbol{\xi}_i$ fits into the j -th cluster. The sum of these for one data point over the c clusters is unity, as shown in Eq. (2). The application of membership values gives the possibility for overlapping clusters.

$$\sum_{j=1}^c w_{i,j} = 1 \quad \forall i \quad (2)$$

The parameter $m > 1$ is the *fuzzifier* that controls the level of separation of the clusters. Its behaviour will be discussed later.

The objective function J can be interpreted as follows, based on [21]. $d^2(\boldsymbol{\xi}_i, \boldsymbol{\mu}_j)$ is the squared distance between data point $\boldsymbol{\xi}_i$ and cluster centre $\boldsymbol{\mu}_j$. The term $w_{i,j} d^2(\boldsymbol{\xi}_i, \boldsymbol{\mu}_j)$ is the error – weighted by the membership raised to a certain power – that arises when $\boldsymbol{\mu}_j$ is taken to represent $\boldsymbol{\xi}_i$. The two summations mean that this error regarding one data point and one cluster centre has to be added up considering each data point – cluster centre combination. Thus a total error is obtained. Minimising this total error means finding cluster centres representing the data points at the best. This is achieved by the FCM algorithm by iteratively calculating the cluster centres and the membership values.

The location of each cluster centre is determined by averaging the data points, weighted by their membership values regarding that cluster. This is shown in Eq. (3).

$$\boldsymbol{\mu}_j = \frac{\sum_{i=1}^N w_{i,j}^m \boldsymbol{\xi}_i}{\sum_{i=1}^N w_{i,j}^m} \quad (3)$$

For the calculation of the $w_{i,j}$ membership values, Eq. (4) is used.

$$w_{i,j} = \left(\sum_{l=1}^c \left(\frac{d(\boldsymbol{\xi}_i, \boldsymbol{\mu}_j)}{d(\boldsymbol{\xi}_i, \boldsymbol{\mu}_l)} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (4)$$

The role of $w_{i,j}$ can better be understood by rewriting Eq. (4) in another form, shown in Eq. (5).

$$w_{i,j} = \frac{\left(\frac{1}{d^2(\boldsymbol{\xi}_i, \boldsymbol{\mu}_j)} \right)^{\frac{1}{m-1}}}{\sum_{l=1}^c \left(\frac{1}{d^2(\boldsymbol{\xi}_i, \boldsymbol{\mu}_l)} \right)^{\frac{1}{m-1}}} \quad (5)$$

By looking at Eq. (5), and considering the reciprocal of the squared distances, $\frac{1}{d^2}$ as a measure of closeness, one can see how $w_{i,j}$ behaves. This formula assigns a large membership value to a cluster – data point pair if they are close to each other (their “closeness” is large), but a small one, if they are remote. It also ensures that the condition in Eq. (2) is fulfilled, since the numerator is a

normalisation by the sum of all membership values regarding the chosen data
180 point.

How steeply the $w_{i,j}$ membership values changes with the distance is defined
by the fuzzifier. As $m \rightarrow 1$, the exponents in Eq. (5) grow, therefore the cluster
centre being closest to the data point will dominate the denominator, thus that
membership value will be close to unity, and necessarily, the other ones will be
185 small. Therefore the overlap between the clusters decreases, and they become
crisper. On the other hand, as $m \rightarrow \infty$, the exponents tend to zero, therefore
the effects of different distances vanish, and all the terms in the denominator
will be similar in magnitude. Thus all $w_{i,j} \rightarrow 1/c$, meaning that the clusters
spread out, the overlap grows, and the ability to differentiate between clusters
190 disappears [21].

In FCM, the aforementioned expressions are used until the results converge
to final $w_{i,j}$ and μ_j values. Details on convergence can be found in [21]. The
algorithm is the following.

1. $w_{i,j}$ values are initialised randomly, but fulfilling the constraint in Eq. (2)
- 195 2. Cluster centres are calculated using Eq. (3)
3. Membership values are updated using Eq. (4)
4. The process is continued from Step 2 using the new values until conver-
gence is reached.

Convergence is checked by comparing the matrices composed of the $w_{i,j}$
200 values obtained in the present and the previous step. If the norm of the difference
of these matrices is smaller than a pre-specified limit ϵ , the algorithm terminates.

This method has proven its capability to identify clusters in several cases,
some relevant ones being the following. FCM was applied for microphone array
measurements in reverberant environments [22, 23], to aid in source localisa-
205 tion. The vortex field behind an oscillating cylinder was analysed in [24], while
turbulent flow patterns were investigated in [25].

Two effects require further analysis regarding the application of FCM in the
present case. First, the relative importance of frequency and angular data in

identifying clusters should be investigated. This is done in the next section in the
 210 form of the distance metric. Second is the periodicity of the fan blading, causing
 the noise patterns to be periodic. Due to this, it is sufficient to investigate the
 noise pattern in one blade passage only. To do clustering in one blade passage,
 one has to assign the passage a beginning $\beta = 0$ angular position, which can not
 be done in an objective way. Furthermore, this choice may change the result
 215 of clustering if periodicity is not taken into account. To avoid the influence of
 this arbitrary choice, the FCM clustering algorithm was extended to handle a
 periodic dimension. This way the result of clustering will be the same for a data
 set, regardless of where $\beta = 0$ was chosen. This problem and its solution will
 further be discussed in later subsections.

220 3.2. Distance weights

FCM relies on the d distance between any two points in the feature space,
 here denoted by \mathbf{x} and \mathbf{y} . This may be the Euclidean norm in the simplest case,
 e.g. $d(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\| = \sqrt{(\mathbf{x} - \mathbf{y})^T (\mathbf{x} - \mathbf{y})}$. However, the dimensions of the
 feature space before non-dimensionalisation may have different physical units,
 225 or have significantly different ranges. In this cases, one may have to alter the
 way the distances are calculated even after the data was made dimensionless,
 to modify the importances of the different dimensions in clustering. This is
 achieved by using weights in the calculation of distances. These weights are
 incorporated into the matrix \mathbf{A} .

The weighting is introduced through the following scalar product definition:
 $\langle \mathbf{x}, \mathbf{y} \rangle = \mathbf{x}^T \mathbf{A} \mathbf{y}$ [21]. With this, the distance metric induced by matrix \mathbf{A} is
 that shown in Eq. (6).

$$d_A(\mathbf{x}, \mathbf{y}) = \|\mathbf{x} - \mathbf{y}\|_A = \sqrt{(\mathbf{x} - \mathbf{y})^T \mathbf{A} (\mathbf{x} - \mathbf{y})} \quad (6)$$

230 In the simplest approach, $\mathbf{A} = \mathbf{I}$, the identity matrix. In this case, the
 Euclidean norm is obtained. Some other approaches are described in [21]. In
 the present case, a diagonal matrix is applied, shown in Eq. (7), with $a_1 = 1$
 as the weight for the first dimension, and adjustable a_k , $k = 2 \dots D$ values for

the other dimensions. In this case, the a_k parameter controls the weight of the
235 k -th dimension compared to the first one.

$$\mathbf{A} = \text{diag}(1, a_2, a_3, \dots, a_D) \quad (7)$$

This is a flexible approach, allowing to set how much influence one dimension has in clustering. However, there are no generally valid guidelines about the choice of the a_k weights. This problem is especially relevant, when the data consists of quantities that are not physically comparable. To this end, a parameter sensitivity study is proposed to determine the a_k values that provide
240 an acceptable clustering result, being in agreement with the properties of the clustering validity method, detailed later.

3.3. Periodicity

When periodic and non-periodic dimensions are both present, the data set is
245 termed *hybrid*, following ref. [26]. The present subset, with exactly one periodic and one non-periodic dimension, is termed *cylindrical* [27]. FCM has been widely used for clustering different types of data, however, it cannot directly be applied onto data with periodic attributes, such as the noise pattern according to the periodicity of the blades in the present case.

250 To apply FCM onto cylindrical data, one has to consider only one period of the infinitely long periodic dimension. This is termed the *analysis window*. Where this analysis window begins in the periodic dimension, is a subjective choice, that can seriously influence the result of clustering. This effect is shown on an illustrative hybrid data set in Fig. 1. There, a hybrid data set is shown,
255 with the horizontal axis being periodic with a repeating data pattern, while the vertical dimension is non-periodic. The crosses indicate the data points. In order to do clustering, one has to define the position of the analysis window, i.e. choose the origin of the β coordinate. This cannot be done based on a physical argument, due to the periodic nature of the dimension. However, this choice
260 may determine the result of partition.

Windows A and B show two possible choices for $\beta = 0$ and thus the location of the analysis window. Note however, that just by visual observation, one can conclude that in window A, two clusters exist, while window B contains three. This illustrates the possible problems of clustering periodic data.

265 One may decide that window A is beneficial, as it does not cut a cluster into two pieces. This would however be a subjective choice, which may not even be applicable when the number of data points N is high. Furthermore, this approach disregards the periodic nature of the data set, leading to incorrect membership values. To avoid this, the clustering method is to be modified

270 to handle cylindrical data. The task is then to cluster data sets lying in a periodic/hybrid domain, in a way that the resulting clusters are invariant to the choice of the analysis window. To reach this, in case of a purely periodic or hybrid data set, distances required for clustering have to be calculated by taking the periodicity into account.

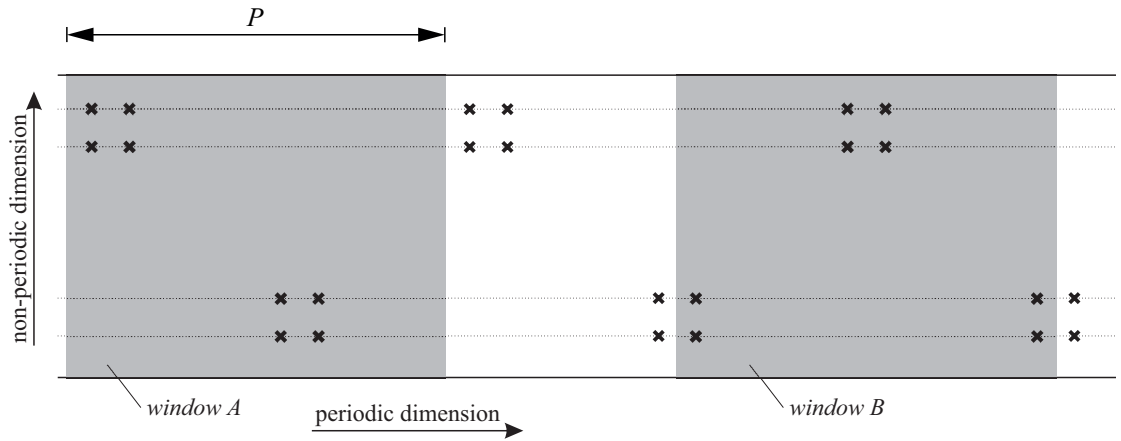


Figure 1: Illustration of the effects of arbitrarily assigned analysis windows in case of hybrid data. The horizontal axis is a periodic dimension, while the vertical is non-periodic. Shaded areas indicate two possible analysis windows. If periodicity is not considered, two clusters are observed in window A, while in B, three.

275 An approach for clustering periodic data using the k-means algorithm is treated in [26]. Besides some other methods, the authors propose the following

metric, defined along a periodic dimension γ with period P , shown in Eq. (8).

$$d_P(\gamma_1, \gamma_2) = \min \{|\gamma_1 - \gamma_2|, P - |\gamma_1 - \gamma_2|\} \quad (8)$$

Furthermore, they discuss the treatment of hybrid data, as well. The method is suitable for k-means clustering, however, it was not extended for fuzzy clustering.
 280

For FCM, the fuzzy c-means algorithm for directional data (FCM4DD) was proposed in [28]. The method was compared against other methods, reported in [29] and [30], and performed favourably, giving better results, while requiring less computational effort. However, in this algorithm, only periodic dimensions
 285 can be taken into account, e.g. FCM4DD is not suitable for hybrid data.

Thus, solutions exist for non-fuzzy clustering of hybrid data, and fuzzy clustering of purely periodic data. The novelty of this paper is an approach to create *fuzzy* clusters in *hybrid* data sets. The idea behind the method is illustrated by Fig. 2. It shows the same points as in Fig. 1, however, now the data set
 290 is wrapped around a cylinder. If the surface of the cylinder is considered for calculating distances and for clustering, periodicity can successfully be taken into account. While Fig. 2 illustrates a two-dimensional hybrid case, the same idea can be applied for more dimensions, as well.

To implement this idea, the distance norm is modified similarly to Eq. (8).
 295 This way, distance can successfully be calculated in a periodic/hybrid data set, and the results are not dependent upon the arbitrary selection of the analysis window.

3.4. Method for fuzzy clustering of hybrid data

The proposed method is able to create fuzzy clusters similarly to FCM, with
 300 a hybrid data set, where *exactly one* dimension is periodic. In the following, without loss of generality, the first dimension is considered as being periodic, with period P .

All dimensions are to be normalised to the range $0 \dots 1$. A D -dimensional periodic offset vector is introduced, shown in Eq. (9), with the period being

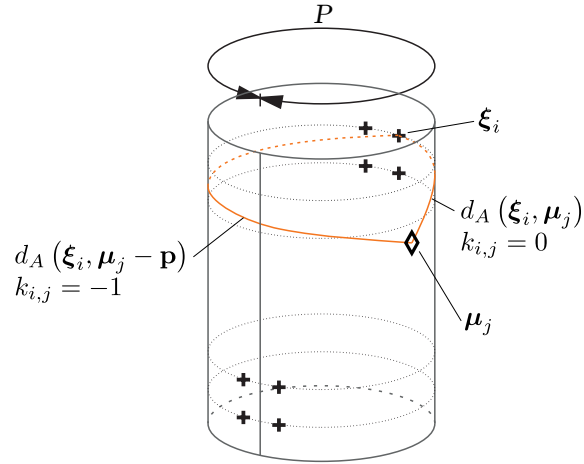


Figure 2: Interpretation of the periodic distance norm. The analysis window is wrapped onto a cylinder. \diamond : cluster centre, $+$: data point. The two possible lines connecting cluster centre μ_j and ξ_i are shown.

305 unity after normalisation. Therefore \mathbf{p} is the unit vector in the periodic dimension, regardless of D .

$$\mathbf{p} = [1, 0, \dots, 0]^T \quad (9)$$

The periodic, weighted distance metric $d_{A,P}$ is to be applied, shown in Eq. (10). It is based on Eq. (6), but was extended for a hybrid case using \mathbf{p} . This includes the diagonal weighting matrix shown in Eq. (7).

$$d_{A,P} = \min\{d_A(\xi_i, \mu_j - \mathbf{p}), d_A(\xi_i, \mu_j)\} \quad (10)$$

This means that every time the distance between data point ξ_i and centre μ_j is required, it has to be considered that a counterpart of μ_j exists in the previous data window as well. The position of this counterpart is $\mu_j - \mathbf{p}$. For
 310 the algorithm, the smallest of the two distances are to be considered. This is analogous to finding the distance on the cylinder shown in Fig. 2, where two points on the surface can be connected from two directions, but the length of the shorter path is defined as their distance.

This distance norm however will often impede convergence. The reason is
 315 that when the new cluster centres are calculated, in Eq. (3), the averaging
 of the data points will again place the $\boldsymbol{\mu}_j$ somewhere in between them. To
 reach convergence even in this case, it is necessary to track, whether the data
 point $\boldsymbol{\xi}_i$ is closer to centre $\boldsymbol{\mu}_j$ from the current analysis window, or to its periodic
 counterpart $\boldsymbol{\mu}_j - \mathbf{p}$ in the previous analysis frame. In the analogy of the cylinder,
 320 this means considering, in which direction were the two points closer to each
 other. Keeping track of this is done by a new variable $k_{i,j} \in \{-1, 0\}$. Its possible
 values are shown in Eq. (11).

$$k_{i,j} = \begin{cases} -1 & \text{if } d_A(\boldsymbol{\xi}_i, \boldsymbol{\mu}_j - \mathbf{p}) < d_A(\boldsymbol{\xi}_i, \boldsymbol{\mu}_j) \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

$k_{i,j}$ is used to indicate if point $\boldsymbol{\xi}_i$ is closer to $\boldsymbol{\mu}_j$ in the present analysis window
 ($k_{i,j} = 0$), or to its periodic counterpart in the previous window ($k_{i,j} = -1$).
 325 The k values are shown in Fig. 2 together with the two paths. $k_{i,j} = -1$
 belongs to the case, when the line connecting $\boldsymbol{\xi}_i$ and $\boldsymbol{\mu}_j$ crosses the generatrix
 corresponding to the 0 value of the periodic dimension. By using $k_{i,j}$, the
 cluster centres can successfully be calculated following Eq. (12). This way,
 the algorithm takes into account, which periodic counterpart of the data point
 330 is closer to the centre, and will use that for the calculation of the new centre
 location.

$$\boldsymbol{\mu}_j = \frac{\sum_{i=1}^N w_{i,j}^m (\boldsymbol{\xi}_i + k_{i,j} \mathbf{p})}{\sum_{i=1}^N w_{i,j}^m} \quad (12)$$

When these are obtained, the membership values can be updated. This is
 reached by using Eq. (4), but by taking the periodic weighted distance norm
 $d_{A,P}$ into account, as shown in Eq. (13).

$$w_{i,j} = \left(\sum_{k=1}^c \left(\frac{d_{A,P}(\boldsymbol{\xi}_i, \boldsymbol{\mu}_j)}{d_{A,P}(\boldsymbol{\xi}_i, \boldsymbol{\mu}_k)} \right)^{\frac{2}{m-1}} \right)^{-1} \quad (13)$$

335 The algorithm is the following.

1. Set c , m , \mathbf{A} , \mathbf{p} , ϵ , $k_{i,j} = 0$, and initial $w_{i,j}$ values. Then repeat the following steps for $s = 1, 2 \dots s_{\max}$
2. Compute centres $\boldsymbol{\mu}_j^{(s)}$, $j = 1, 2 \dots c$ using Eq. (12). If the periodic coordinate of $\boldsymbol{\mu}_j$ leaves the domain $[0, 1)$, reintroduce it to the other side by taking its fractional part.
3. Compute $k_{i,j}^{(s+1)}$ using Eq. (11).
4. Compute membership values $w_{i,j}^{(s+1)}$ using Eq. (13).
5. Compare the matrices constructed from the $w_{i,j}$ membership values at steps s and $s + 1$. If the norm of their difference is less than ϵ , exit. Otherwise, increment s and return to step 2.

3.5. Xie-Beni index

A limitation of FCM is that the number of clusters c is to be known in advance [31]. This information is, however, generally not available. Therefore the suggested procedure is to employ a parameter sensitivity study by varying c . During that, a measure of clustering goodness is to be investigated. These measures are called *validity indices*. Many kinds exist, some of them being summarised in [31]. In the original method in [19], the Xie-Beni (XB) index [32] is used, as it is easy to compute, yet able to quantify the two most important aspects of good clustering:

- compact clusters are sought for,
- being well-separated from each other.

The XB index is defined in Eq. (14).

$$\text{XB} = \frac{\sum_{i=1}^N \sum_{j=1}^c w_{i,j}^m d^2(\boldsymbol{\xi}_i, \boldsymbol{\mu}_j)}{N \min_{j \neq l} d^2(\boldsymbol{\mu}_j, \boldsymbol{\mu}_l)} \quad (14)$$

In Eq. (14), the numerator is in relation to the error introduced by replacing the data points with the cluster centres, therefore the lower this value, the more compact the cluster. The denominator contains the distance of the two clusters lying closest to each other, being a measure of cluster separation. In this case,

the higher its value, the better the clustering. Based on these, the minimum XB value indicates the optimal clustering, and the optimal choice of c .

The periodicity in the data set has to be accounted for in the calculation of XB, as well. This is straightforward, as only the distance measures have to be changed to the new, weighted and periodic ones. The periodic XB index is shown in Eq. (15).

$$XB_P = \frac{\sum_{i=1}^N \sum_{j=1}^c w_{i,j}^m d_{A,P}^2(\boldsymbol{\xi}_i, \boldsymbol{\mu}_j)}{N \min_{j \neq l} d_{A,P}^2(\boldsymbol{\mu}_j, \boldsymbol{\mu}_l)} \quad (15)$$

3.6. Summary of the proposed clustering method

The outline of the method for fuzzy clustering of cylindrical data is the following. First, an arbitrary analysis window is to be chosen, and the normalised $\boldsymbol{\xi}_i$ values are to be calculated. Then hybrid fuzzy clustering is to be applied, using the algorithm above. This is to be repeated with all combinations of the investigated c , m and a_k values. Following that, parameter studies are to be carried out to determine the optimal settings for these variables. The optimal one, based on the XB_P index and physical considerations, is to be chosen and accepted. Finally, clustering results can be used and interpreted corresponding to the selected settings.

4. Case study

The present investigation concerns the same axial fan and the same measurement data as in [19]. Relevant details of the measurement setup are as follows. The investigated fan was an axial flow industrial cooling fan, in a free-inlet free-exhaust rotor-only configuration. It was mounted in a short duct with a short inlet cone. The tip diameter of the fan was 0.3 m, the hub-to-tip ratio was 0.35, and the number of blades was 5. The rotor speed was measured by means of an optical probe, and found to be 1430 revolutions per minute. The tip gap size was 7 mm. The blades were manufactured from steel plates with a circular arc camber profile. The blade tip, the trailing, and the leading edges were rounded.

An OptiNav, Inc. Array 24 phased array microphone was used for measuring the noise. This unit has 24 omnidirectional electret microphones mounted flush against the array surface. They are arranged along logarithmic spiral arms. The diameter of the circumscribed circle is 0.95 m [3]. The array was placed perpendicularly to the fan axis of revolution, while the distance between the array and the fan inlet plane was 1.83 tip diameters. This distance was chosen to improve the spatial resolution of the beamforming maps, while avoiding inflow disturbances. The latter was confirmed in [14] for the same configuration. The inlet velocity profile with and without the array was measured, and no significant difference was found. Measurements taken on the suction side are processed; in this case, the flow field has negligible impact on the microphones, as proved by vane anemometer measurements. A sketch of the measurement setup is shown in Fig. 3.

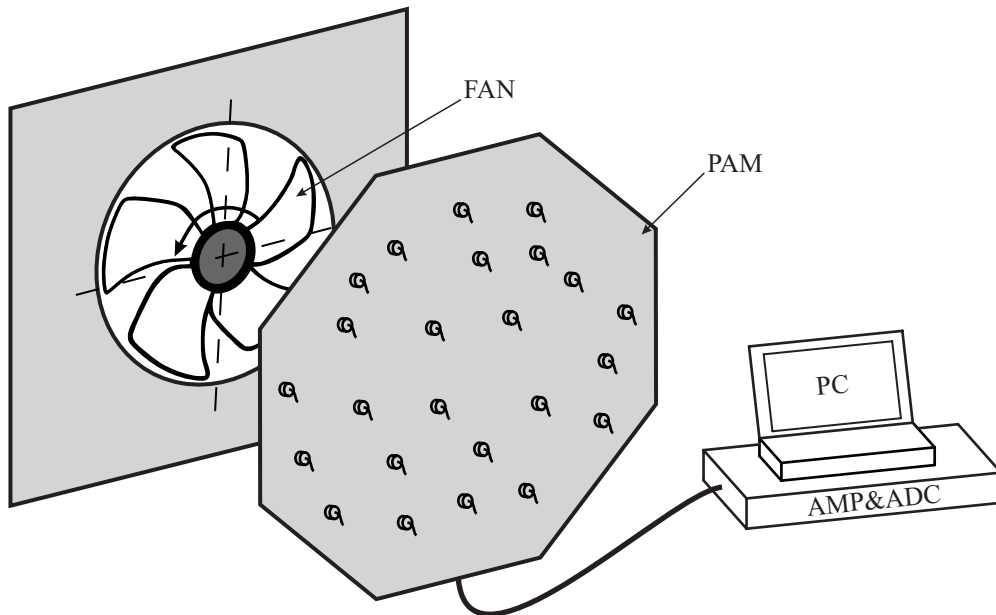


Figure 3: Sketch of the measurement setup. PAM: phased array microphone. AMP&ADC: amplifier and analogue-to-digital converter. PC: personal computer. The microphone positions are for illustrative purposes only; they do not correspond to those of the actual array.

The electric signals were amplified, digitised, and recorded by the accompa-

nying signal processing unit. A sampling frequency of 44.1 kHz was applied and data was acquired for 20 s, corresponding to more than 2000 blade passages. The measurements were carried out in a laboratory with no special acoustical
405 treatment. To remove the effects of rotation and create source maps in a system rotating together with the rotor, the ROSI method was used [4], as the geometry of the array did not allow the application of frequency domain interpolation methods, e.g. [33]. Frequency-domain information was obtained by cutting the data into blocks of 1024 overlapping by 50%, windowing them using
410 a von Hann function, applying a fast Fourier transform on each block, then averaging the spectra. Diagonal removal was applied to reduce the effects of uncorrelated noise between the microphones [4]. The narrow-band source maps were summed to create third-octave band maps centred on 2000, 2500, 3150, 4000, 5000, and 6300 Hz. Lower frequencies were omitted due to limitations in
415 spatial resolution.

The sound pressure spectrum of the noise emitted by the fan, as measured by one microphone of the array, is shown in Fig. 4. It can be seen that in the investigated frequency range, the generated noise has broadband characteristics. Therefore tonal noise was not investigated, and only broadband source mechanisms were considered. Furthermore, this is why analysis with third-octave
420 band frequency resolution was considered appropriate.

The originally obtained beamform maps are shown in Fig. 5. Here one can see that the most significant noise sources are often in the vicinity of the rotor tip radius. However, the features are not completely periodic, which might
425 be the result of some surface inhomogeneity, inflow effects, background noise, measurement uncertainty, or side lobes from beamforming. These effects are however to be filtered out, as at this stage, the basic behaviour of the fan is to be investigated. The reason for applying the data reduction method is to remove these effects, to be able to concentrate on the remaining phenomena.
430 A further aim is to reduce these maps, that can only be inspected visually, into a data set, that can easier be interpreted, in an algorithmic, reproducible manner. This is done by constructing the base sources, clustering them to find

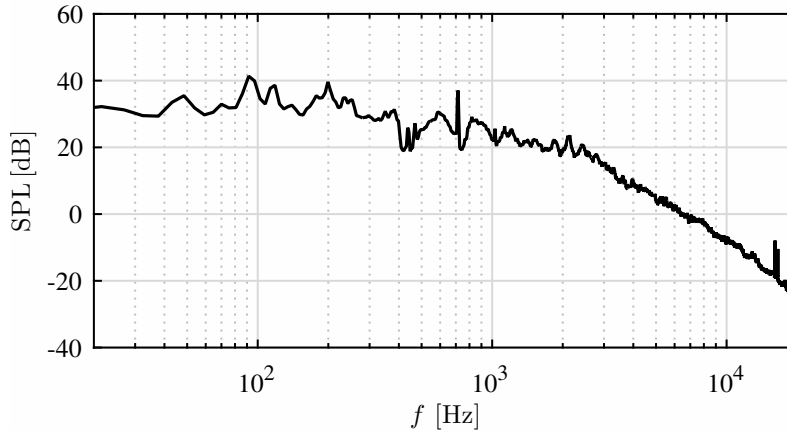


Figure 4: Noise sound pressure level spectra in dB [19]

those that are the result of the same mechanism, and then finding the relevant source mechanisms.

435 Possible broadband noise source mechanisms that may appear in low speed axial fans in general are the following [19].

- *Small scale turbulence ingestion noise.* Noise originating from the turbulence of upstream flow, e.g. caused by a protection grid, impinging on the leading edge.
- 440 • *Noise due to ingestion of large scale turbulence.* This source is due to the ingestion of large eddies, e.g. those shed by the housing, etc.
- *Turbulent boundary layer noise.* This noise originates from wall pressure fluctuations due to turbulence in the boundary layer.
- *Separated flow noise.* It is a result of the pronounced flow unsteadiness in
445 case of flow separation.
- *Profile vortex-shedding noise.* This originates from the vortices being shed over the blade profile.
- *Turbulent boundary layer – trailing edge interaction noise.* This noise is the result of scattering turbulence at the trailing edge in case of an

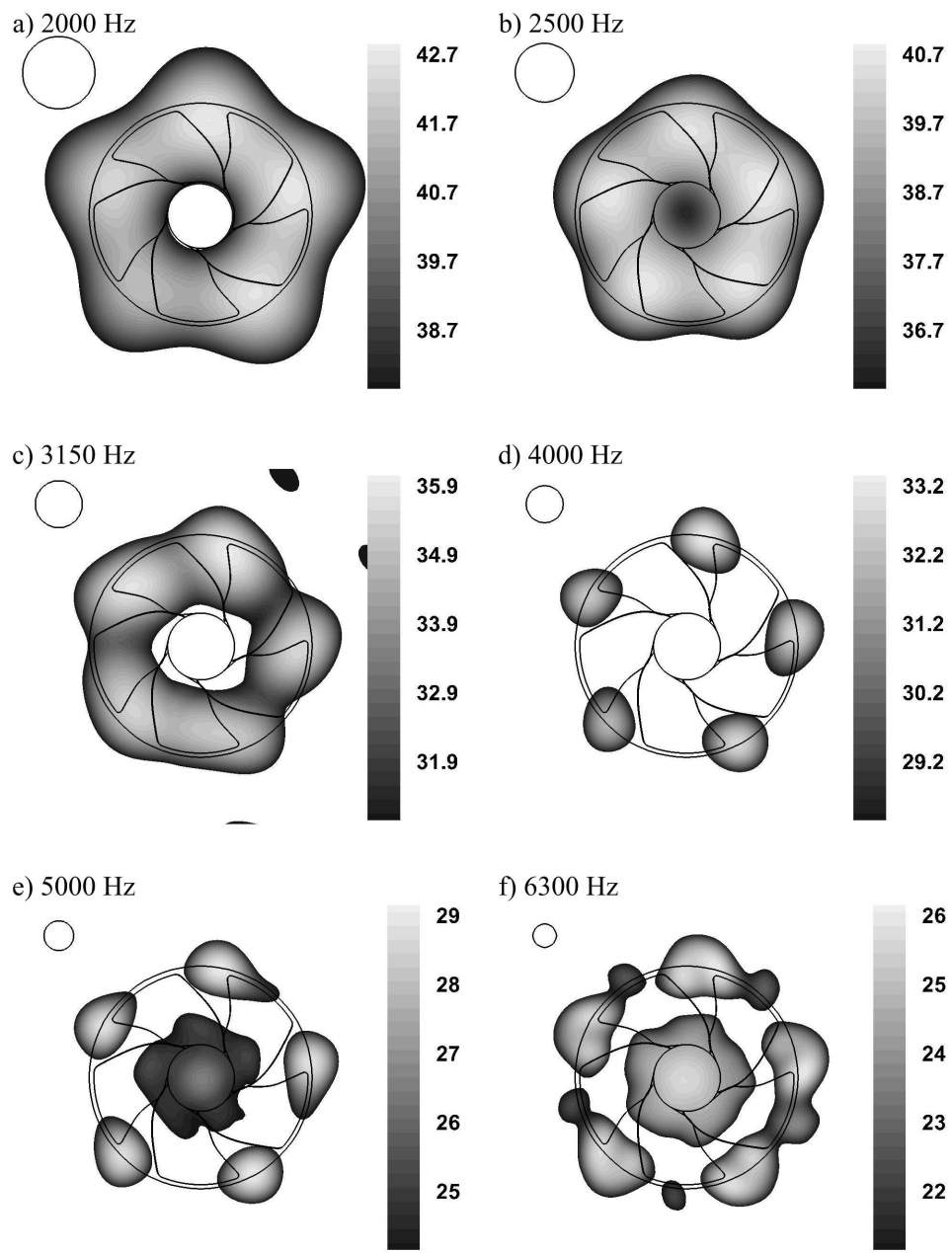


Figure 5: Third-octave band beamforming maps [19] at mid-frequencies between a) 2 kHz and e) 6.3 kHz. The colour bar shows the beamform levels in dB, while the circle indicates the Rayleigh resolution limit at the mid-frequency of the third-octave band.

450

attached boundary layer.

- *Tip leakage flow noise.* This noise is generated by the highly turbulent tip leakage flow coming from the preceding blade.
- *Double leakage flow noise.* Similar to the previous noise, but the flow is coming from the second preceding blade.
- 455 • *Blunt trailing edge vortex shedding noise.* This noise source is the result of vortices being shed at the finite thickness of the trailing edge.

Out of these, the noise sources identified using the clustering method will be presented in the Results section. The original beamform maps are treated in detail and compared to computational fluid dynamics studies in [13, 14, 15].

460

Since third-octave band beamforming maps were filtered and evaluated, the base sources refer to the mid-frequencies of these bands. However, these represent the whole frequency range between the lower and the upper limit frequencies of that third-octave band. Due to this, the normalisation procedure in Eq. (16) was applied. Here (f_{\min}) indicates the lower limit frequency of the lowest third-octave band, while (f_{\max}) is the higher limit of the highest third-octave band.

465

$$\hat{f} = \frac{f - f_{\min}}{f_{\max} - f_{\min}} \quad (16)$$

For the angular dimension, approach in Eq. (17) was followed. Here the number of blades is $B = 5$, the angular extension of one blade passage is therefore $2\pi/B$, and the values of the angle are $\beta \in [0, 2\pi/B)$.

$$\hat{\beta} = \frac{\beta}{2\pi/B} \quad (17)$$

470

The constructed data set consists of $\xi_i = [\hat{\beta}_i, \hat{f}_i]$ base sources. Since the β dimension is periodic, the proposed hybrid clustering method was used. This way, the results are independent of where the $\beta = 0$ angle was assigned to. For the sake of comparability of the results to those published in [19], $\beta = 0$ was assigned to the trailing edge also herein.

475 **5. Results**

The introduced method was applied onto the data set investigated in [19], obtained after the procedure described above.

Data was extracted from the beamforming maps in Fig. 5 at four different radii, covering the annulus area. These were the following values, normalised
480 by the tip radius: $R = 0.4, 0.6, 0.8, 1.0$. The Fourier-based filtering procedure was carried out for each of these radii. At $R = 0.4$, no base sources were found, meaning that no significant blade-passage periodic sources exist. This can be attributed to the vicinity of the hub at $R = 0.35$ and the shortness of the circumferential path relative to the resolution of beamforming. At each higher
485 R , some base sources were identified. At $R = 0.6, 0.8$ and 1.0 , their number was $N = 5, 7$ and 11 , respectively. This is in agreement with the visual observation of Fig. 5, in which blade passage periodic patterns are most spectacular around the tip radius, $R = 1.0$. The m and a parameter studies were carried out for the $R = 1.0$ case, as that contains most information, and its results can be
490 compared to those in [19].

5.1. Results of $R = 1.0$

For the \mathbf{A} distance weight matrix, shown in Eq. (7), the following parameter sensitivity study was carried out at $R = 1.0$. Since $D = 2$, only one element of \mathbf{A} , denoted simply by a , is sought for. The effect of parameter a was investigated in
495 the range $a \in \{0.1, 0.4, 0.7, 1, 1.3, 1.6, 1.9\}$, as it includes the non-weighted case ($a = 1$), and two extremes from that: a high dependence on β ($a = 0.1$), and a high dependence on f ($a = 1.9$). The possible number of clusters was $c \in \{2, 3\}$, following the guidelines in [34]. Specifically, at least two clusters are required for XB_P to give meaningful results, as it contains the distance between the closest
500 clusters, while the recommendation for the maximum c is that $c \approx \sqrt{N}$ [34], giving $\sqrt{11} \approx 3$. To be able to show the trend in the values of XB_P , this was increased to 5, however, $c = 4$ and 5 are not real candidates for the optimum c . The fuzzifier was investigated in the range $m \in \{1.4, 1.6, 1.8, 2.0, 2.2\}$ as well.

The range was chosen since it includes $m = 2$, the recommended setting in
 505 absence of a priori information [34], and five different values were expected to
 give sufficient information on the behaviour of clustering. The XB index was
 evaluated using these settings. The results are shown in Fig. 6, where the three
 diagrams show XB_P as a function of c with varying a weight at three different m
 values. Subfigure a) was obtained with the smallest, $m = 1.4$ value, b) with the
 510 standard $m = 2$, and c) with the highest, $m = 2.2$ value. These were selected
 only for illustration, as all the intermediate values show similar trends.

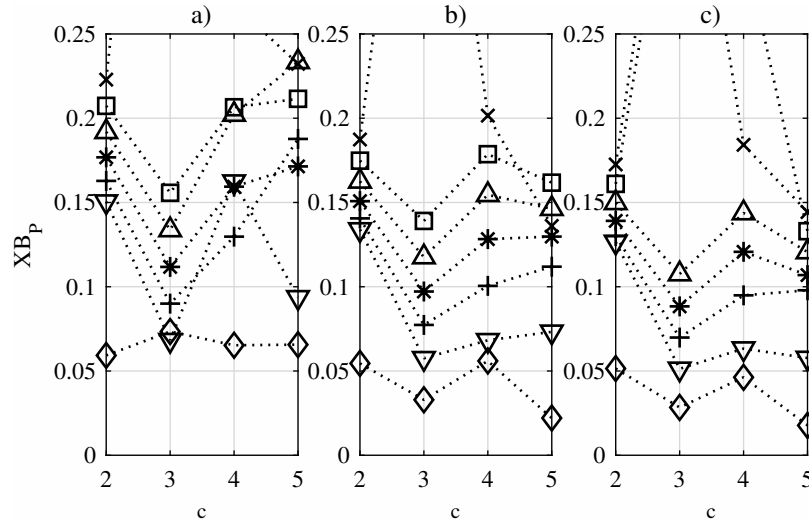


Figure 6: XB_P as a function of c and a , at varied m values. a) $m = 1.4$ b) $m = 2.0$ c) $m = 2.2$
 $\diamond : a = 0.1$, $\nabla : a = 0.4$, $+$: $a = 0.7$, $*$: $a = 1.0$, $\triangle : a = 1.3$, $\square : a = 1.6$, $\times : a = 1.9$

Comparing the three subfigures in Fig. 6, one can see that in most cases,
 the effect of the fuzzifier is small, as the XB_P values follow similar trends in
 case of each m . This is in agreement with the findings of the previous study
 515 [19], and therefore the standard value of $m = 2$ is accepted and applied in the
 following.

By looking for the minimum of XB_P , one can see that for moderate values
 of a around unity, ranging from 0.4 to 1.3, the optimal cluster number is $c = 3$.
 These results are similar to those obtained with a non-periodic FCM in [19],

520 reproduced here in part a) of Fig. 8. For $a = 0.1$, the minimum of XB_P is
 found at $c = 5$ for each m , however, this should not be accepted as the optimum
 c , since XB has a tendency to have decreasing values as c increases, regardless
 of the goodness of clustering [32]. Furthermore, this is in contrast with the
 previously mentioned recommendation regarding \sqrt{N} being the largest number
 525 of clusters to be considered. Therefore such high c values should be treated
 with criticism. For values of $a = 1.6$ and higher, XB_P again often reaches
 its minimum at $c = 5$, but similarly to the previous case, this should not be
 considered as the optimum c . In these cases, the real optimal cluster number
 is probably $c = 2$. These a are significantly larger than unity though, and the
 530 clustering results with these settings differ substantially from those obtained
 with $a \approx 1$, therefore they should not be accepted.

The way a transforms the results is shown in Fig. 7. There the cluster
 maps can be seen for various $a = 0.1$, $a = 1$, and $a = 1.9$ values, with $c = 3$
 and $m = 2$ fixed. For lower values of a , the clusters are elongated in the \hat{f}
 535 dimension, while for higher values, the opposite occurs, and the cluster extend
 in the $\hat{\beta}$ dimension.

Based on Figs. 6 and 7, the optimal weight is $a = 1$. This is because this
 presents a good compromise between the elongation of the clusters along the two
 dimensions. Furthermore, around this value, in a fairly wide range of $a = 0.4$
 540 to $a = 1.3$, the value of XB_P is stabilised and shows the theoretically expected
 tendency, with a global minimum inside the range, e.g. not at the limits of
 the investigated c values, being independent of the choice of m . This $a = 1$
 result means that $\hat{\beta}$ and \hat{f} are to be taken into account with equal weights,
 indicating the equal importance of spatial and frequency-wise extension of the
 545 noise sources.

The optimal cluster number at $R = 1.0$ belonging to $a = 1$ is $c = 3$, according
 to Fig. 6. Therefore the cluster map with these parameters, $m = 2$, $a = 1$ and
 $c = 3$, was evaluated in detail, and compared to the cluster map obtained with
 the standard FCM algorithm in [19]. In the specific case study reported herein,
 550 the periodic and non-periodic methods have both led to $c = 3$ as the optimal

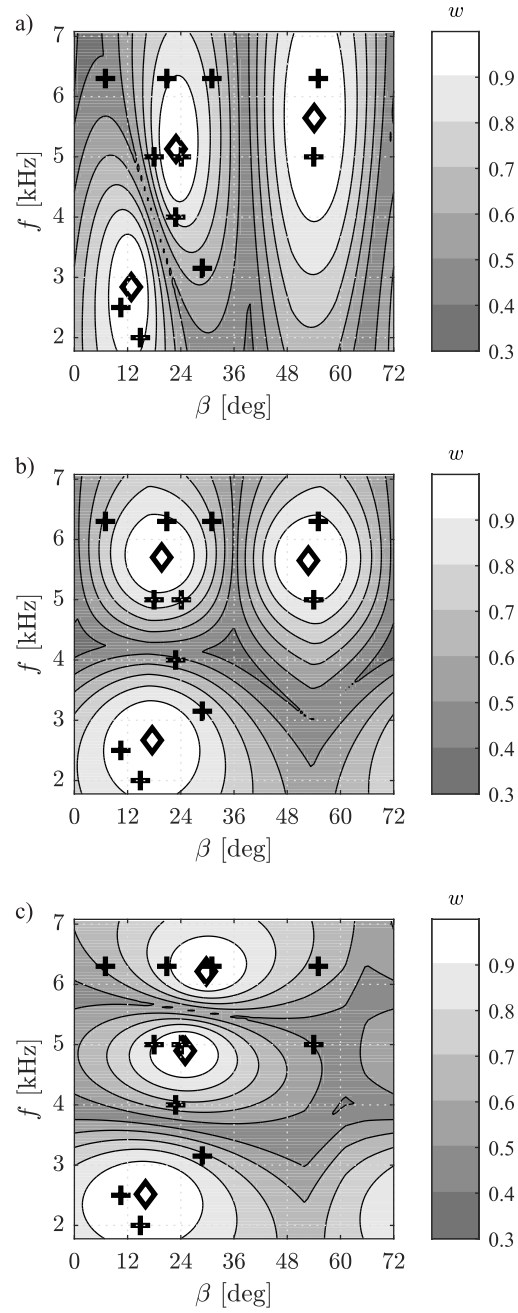


Figure 7: Cluster maps for $c = 3$ and $m = 2$ with a varied. $+$: data points, \diamond : cluster centres. a) $a = 0.1$, b) $a = 1$, c) $a = 1.9$. The contour map shows the maximum membership value at each point.

cluster number. In general however, the introduction of periodic clustering may give results that are substantially different from those of the original FCM, as illustrated by Fig. 1.

At $R = 1.0$, the periodic and non-periodic cluster maps are visually similar, shown in Fig. 8 as parts a) and b), respectively. There is some difference due to the periodic clustering though. The membership contour-lines of the bottom-left cluster, termed *Cluster 1* in [19], extend through the edge of the diagram in case of periodic clustering. The membership iso-lines of the upper two clusters have a modified shape, as well, being compressed in the β direction in the periodic case, compared to the non-periodic one. This is the result of the periodicity being taken into account, as now these clusters influence each other both from in between them, and from the limits of the angular dimension, $\beta = 0^\circ$ and $\beta = 72^\circ$. The contour map shows the maximum membership value at each point. These w values have changed, too, and their contours have become continuous at the vertical edges of the diagram, due to periodicity being accounted for.

The similarity between the two figures shows that intuitively assigning the trailing edge of the blade as $\beta = 0$ was an appropriate choice. However, this was previously done based on a subjective visual evaluation, which is not required any more, when applying the periodic clustering method, since in the present case, the results do not depend on the analysis window.

As the clusters have not changed significantly, their interpretation, given previously in [19], is still considered valid. Here, due to the nature of the spectrum shown in Fig. 4, only broadband source mechanisms were considered. The identified mechanisms are summarised below. The cluster numbers refer to those in Fig. 8.

- Cluster 1 represents the *turbulent boundary layer – trailing edge interaction noise*. There are three reasons leading to that conclusion [19]. First, Cluster 1 is in the vicinity of the trailing edge of the previous blade. Second, the frequency of this cluster centre, being about 2.8 kHz, is in fair

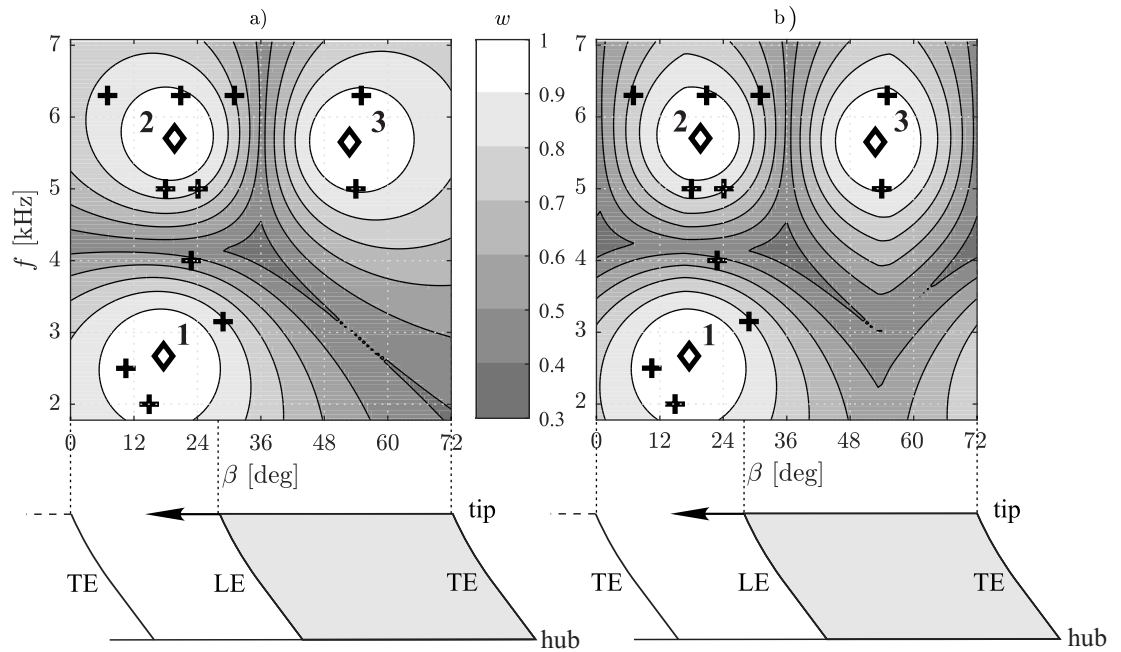


Figure 8: Clusters for $c = 3$, $m = 2$ and $a = 1$. +: data points, \diamond : cluster centres. a) standard FCM clustering b) the proposed hybrid clustering method. The contour map shows the maximum membership value at each point. The numbers indicate the cluster numbers. At the bottom, the lines indicate the trailing edges (TE) and leading edges (LE) of the fan blades, together with the tip radius.

agreement with the peak frequency of this noise source, being about 1.2 kHz [35], especially considering that third octave bands below 2 kHz were not processed here. Third, the derivative of the beamforming level with respect to the frequency, which is about 7 dB/kHz, agrees fairly well to the 8.6 dB/kHz obtained from models related to this noise generation mechanism [35, 36, 37].

- Cluster 2 represents the *tip leakage flow noise*. The reasons behind that are as follows. First, the location, being positioned between two blades. Second, the frequency of the cluster, being about 5.8 kHz, agrees well with the frequency of 5.2 kHz reported for this source type in [35]. Furthermore, the frequency derivative of the beamforming level, being 2.3 dB/kHz, and the shape of the beamforming level spectrum, is again in agreement with references [35, 38] that report a value of 2.2 dB/kHz for tip leakage flow.
- Cluster 3 represents the *double leakage flow noise*. This type of noise source originates from a tip leakage flow that reaches the pressure side of the adjacent blade around the tip region, where it leaks again, reaching the third blade [39]. The frequency of this cluster, and the frequency derivative of the beamforming level in this cluster is similar to those of Cluster 2, suggesting that the two sources are related. However, this is located further downstream from the trailing edge, when compared to Cluster 2. This leads to the conclusion that Cluster 3 is related to the double leakage flow noise. Similar observations, supported by numerical and experimental fluid dynamics analyses, are reported in [13, 14, 15].

For more details on the analysis process, and for the reasons behind excluding further generation mechanisms, the Reader is referred to [19].

5.2. Comparative evaluation of multiple radii

The same approach was applied onto the two remaining radii. As noted before, $N = 5$ and 7 base sources were identified after filtering at $R = 0.6$ and 0.8, respectively. Due to the small number of data points, only $c = 2$

610 clusters were considered meaningful, following the recommendation regarding
 $c_{max} \approx \sqrt{N}$ given in [34]. Therefore the XB_P index was not applied, and for
these angles, maps with $c = 2$ clusters are shown in Fig. 9. There, Cluster 1
is due to the noise of the turbulent boundary layer – trailing edge interaction,
Cluster 2 is due to tip leakage flow, while Cluster 3 is due to double leakage
615 flow noise.

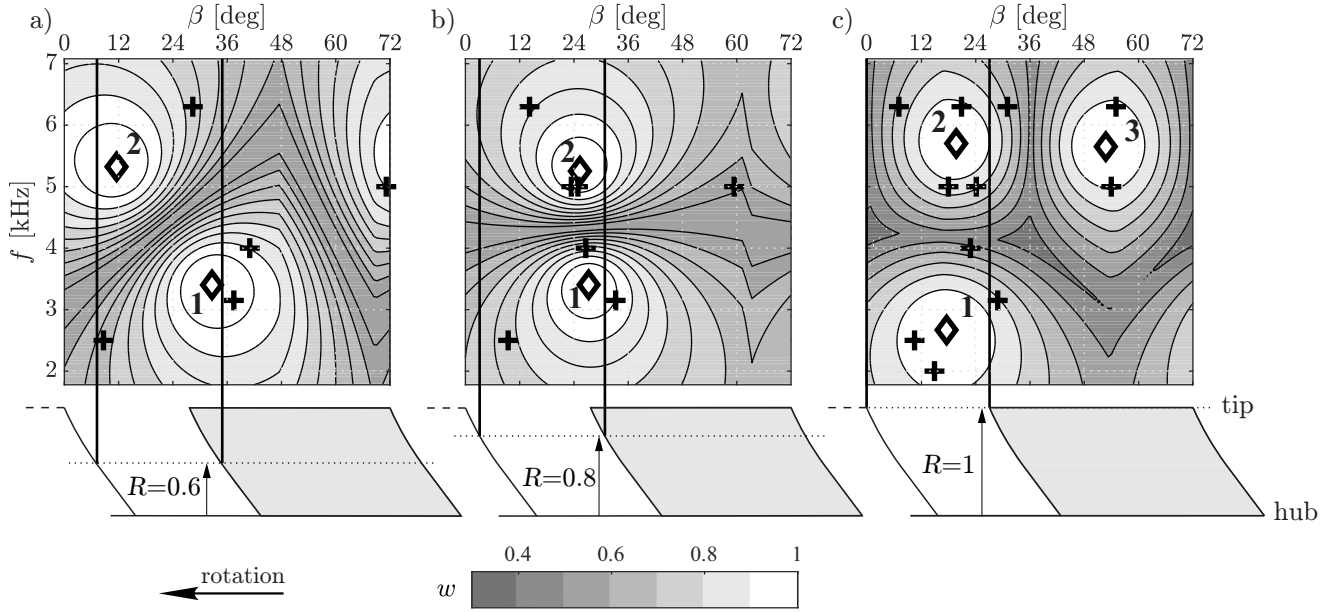


Figure 9: Cluster maps. +: data points, \diamond : cluster centres. a) $R = 0.6$, b) $R = 0.8$, c) $R = 1.0$. $\beta = 0$ indicates the location of the trailing edge of the preceding blade at $R = 1.0$. The shaded areas indicate the fan blades, which are not to scale, however, the lines showing the trailing and leading edges are placed according to their physical positions. w indicates the maximum membership value at each point. Cluster 1: turbulent boundary layer – trailing edge interaction noise. Cluster 2: Tip leakage flow noise. Cluster 3: double leakage flow noise.

To provide further information relating to the maps at different radii, *cluster amplitudes* σ were introduced. The cluster amplitude σ_i represents the strength of the source attributed to the i -th cluster. This is calculated as shown in Eq. (18), using the data point amplitudes p_i , which were obtained from the
620 Fourier expansion of the circumferential source strength distributions. As the

aim is to qualitatively compare the maps and not to quantify them, the exact values of cluster amplitudes are not reported, only their relation to each other, therefore the p_i values are used in a dimensionless form. Cluster amplitudes were determined separately for each cluster and each radius.

$$\sigma_i = \sum_{j=1}^N w_{i,j} p_j \quad (18)$$

625 As seen in Fig. 9, Cluster 1, connected to turbulent boundary layer – trailing edge interaction noise, was identified at all radii. Its location is similar in all maps: the cluster centre is located about 20 degrees downstream of the trailing edge, following its inclination due to blade skew, and its frequency is around 3 kHz. This trend, i.e. that base sources positions follow the skewed trailing
630 edge, is apparent in other maps, as well, from $R = 0.5$ throughout $R = 1.0$, which are not shown here. The amplitude of Cluster 1, σ_1 was found to increase monotonically with increasing radii: at $R = 0.6$, it reached 27% of the amplitude at the tip radius, while at $R = 0.8$, this value was 62%. This may be connected to the spanwise increase of circumferential velocity and flow velocity. Therefore,
635 the sets of Cluster 1 match the spanwise changes in trailing edge geometry and in flow conditions.

Cluster 2 was is also found in Fig. 9, where its frequency is around 5.5 kHz in all three maps. Its location is consistently about 20 degrees downstream of the trailing edge at $R = 0.8$ and 1.0. This cluster was connected to the noise
640 due to tip leakage ($R = 1.0$). Therefore, Clusters 2 of $R = 0.8$ and $R = 0.6$ are considered as the weakened signatures of the tip leakage flow appearing at these lower radii, as well, due to the limited spatial resolution of the beamforming technique. This view is supported by the fact that the σ amplitude of Cluster 2 decreases monotonically towards the hub: at $R = 0.8$, 54% of the amplitude
645 was found, while at $R = 0.6$, it was about 19%. The limit in spatial resolution causes a variance in the angular position of Clusters 2, as well.

At $R = 1.0$, Cluster 3 was found, as well, and it was connected to the noise of double leakage flow. At lower radii, this cluster was not found. This is

attributed to the limited spatial extent and strength of the double leakage flow.

650 In summary, three clusters were localised in the rotor annulus. These are in good agreement with those found earlier in [19], which represent turbulent boundary layer – trailing edge interaction, tip leakage, and double leakage noise. The clusters at the different radii are physically related, are found at similar frequencies, and mostly follow the shape of the skewed trailing edge.

655 6. Conclusions

A new method was presented for the fuzzy clustering of data points in a hybrid data set. Its advantage is being able to identify clusters in a periodic feature space, regardless of the number of total dimensions. It can only be applied however, if there is exactly one periodic dimension.

660 Application of the algorithm was illustrated in a case study to investigate noise sources in axial fans, by analysing beamforming maps. For the sake of robustness, sensitivity studies were carried out to determine the parameters. The number of clusters was found with the help of the extended XB index, taking periodicity into account. The effect of fuzzifier was investigated, and in 665 the present case study, was found to have little influence on the results in the usually suggested range. The relative importance of the different dimensions was evaluated as well, again through a parameter study, to find the optimal weights for the distance measure. These sensitivity studies on m and a_k , also in determining c via the XB index, are recommended on a case-specific basis in 670 every case for which the method is applied.

The procedure was carried out for multiple radii on the blade surface, in order to obtain spanwise information about the noise sources. The clusters of turbulent boundary layer – trailing edge interaction noise were found to be in accordance with the spanwise change in blade geometry and in velocity condi- 675 tions. The clusters of tip leakage flow were noise also appeared at radii further away from the tip, drawing attention to limitations in spatial resolution of the applied beamforming technique.

The results show that the proposed beamforming map analysis method is robust, and the parameter sensitivity study provides objective choices for the variables. Thus, it allows the analysis of beamforming maps in a concerted and objective way, allowing researchers to identify the dominant noise sources being present in an axial fan. The periodic clustering method is however not restricted to beamforming or fan data, and could see a wide range of applications.

Acknowledgements

The authors thank Tamás Kalmár-Nagy for his useful suggestions regarding the distance metric.

This work has been supported by the Hungarian National Research, Development and Innovation Centre under contract No. K 112277.

The research reported in this paper was supported by the Higher Education Excellence Program of the Ministry of Human Capacities in the frame of Water science & Disaster Prevention research area of Budapest University of Technology and Economics (BME FIKP-VÍZ).

References

- [1] R. P. Dougherty, Beamforming in acoustic testing, in: T. J. Mueller (Ed.), Experimental Aeroacoustics, Springer, Berlin, 2002, pp. 62–97. doi:10.1007/978-3-662-05058-3.
- [2] L. Koop, Beamforming in acoustic testing, in: M. L. Riethmuller, M. R. Lena (Eds.), VKI Experimental Aeroacoustics, von Kármán Institute for Fluid Dynamics, Rhode Saint Genese, 2006. doi:10.1007/978-3-662-05058-3.
- [3] E. Balla, J. Vad, Establishment of a beamforming dataset on basic models of low-speed axial fan blade sections, Periodica Polytechnica, Mechanical Engineering 61 (2) (2017) 122–129. doi:10.3311/PPme.9548.

- [4] P. Sijtsma, S. Oerlemans, H. Holthusen, Location of rotating sources
705 by phased array measurements, in: Proceedings of the 7th AIAA/CEAS
Aeroacoustics Conference, Maastricht, 2001. doi:10.2514/6.2001-2167.
- [5] G. Herold, E. Sarradj, Microphone array method for the characterization
of rotating sound sources in axial fans, Noise Control Engineering Journal
63 (6) (2015) 546–551. doi:10.3397/1/376348.
- [6] R. Dougherty, B. Walker, Virtual rotating microphone imaging of broad-
710 band fan noise, in: 15th AIAA/CEAS Aeroacoustics Conference (30th
AIAA Aeroacoustics Conference), 2009, p. 3121. doi:10.2514/6.
2009-3121.
- [7] W. Pannert, C. Maier, Rotating beamforming–motion-compensation in
715 the frequency domain and application of high-resolution beamforming al-
gorithms, Journal of Sound and Vibration 333 (7) (2014) 1899–1912.
doi:10.1016/j.jsv.2013.11.031.
- [8] G. Herold, F. Zenger, E. Sarradj, Influence of blade skew on axial fan
component noise, International Journal of Aeroacoustics 16 (4-5) (2017)
720 418–430. doi:10.1177/1475472X17718740.
URL <https://doi.org/10.1177/1475472X17718740>
- [9] F. Zenger, G. Herold, E. Sarradj, Acoustic characterization of forward-
and backward-skewed axial fans under increased inflow turbulence, AIAA
Journal 55 (4) (2014) 1241–1250. doi:10.2514/1.J0555383.
- [10] F. Krömer, J. Müller, S. Becker, Limitations of phased array beamforming
725 in open rotor noise source imaging, AIAA Journal 56 (4) (2018) 1507–1518.
doi:10.2514/1.J056324.
- [11] O. Minck, N. Binder, O. Cherrier, L. Lamotte, V. Budinger, Fan noise
730 analysis using a microphone array, in: International Conference on Fan
Noise, Technology and Numerical Methods (Fan 2012), Senlis, 2012.

- [12] F. J. Zenger, G. Herold, S. Becker, E. Sarradj, Sound source localization on an axial fan at different operating points, *Experiments in Fluids* 57 (8) (2016) 136. doi:10.1007/s00348-016-2223-8.
- [13] T. Benedek, J. Vad, Study on the effect of inlet geometry on the noise of an axial fan, with involvement of the phased array microphone technique, Paper ID GT2016-57772, in: *Proceedings of ASME Turbo Expo 2016, Seoul, 2016*. doi:10.1115/GT2016-57772.
- [14] T. Benedek, J. Vad, An industrial onsite methodology for combined acoustic-aerodynamic diagnostics of axial fans, involving the phased array microphone technique, *International Journal of Aeroacoustics* 15 (1-2) (2016) 81–102. doi:10.1177/1475472X16630849.
- [15] T. Benedek, J. Vad, Spatially resolved acoustic and aerodynamic studies upstream and downstream of an industrial axial fan with involvement of the phased array microphone technique, in: *Proc. 11th European Conference on Turbomachinery Fluid Dynamics and Thermodynamics, Madrid, Spain, 2015*.
- [16] P. Sijtsma, CLEAN based on spatial source coherence, *International Journal of Aeroacoustics* 6 (4) (2007) 357–374. doi:10.1260/147547207783359459.
- [17] T. F. Brooks, W. M. Humphreys, A deconvolution approach for the mapping of acoustic sources (DAMAS) determined from phased microphone arrays, *Journal of Sound and Vibration* 294 (4) (2006) 856–879. doi:10.1016/j.jsv.2005.12.046.
- [18] B. Tóth, J. Vad, Fourier analysis of beamforming data at the tip of an axial fan rotor, *Periodica Polytechnica, Mechanical Engineering* 60 (3) (2016) 152–158. doi:10.3311/PPme.8935.
- [19] B. Tóth, J. Vad, Algorithmic localisation of noise sources in the tip region

- of a low-speed axial flow fan, *Journal of Sound and Vibration* 393 (2017) 425–441. doi:10.1016/j.jsv.2017.01.011.
- 760 [20] J. A. Hartigan, M. A. Wong, Algorithm AS 136: A k-means clustering algorithm, *Journal of the Royal Statistical Society. Series C (Applied Statistics)* 28 (1) (1979) 100–108. doi:10.2307/2346830.
- [21] J. C. Bezdek, R. Ehrlich, W. Full, FCM: The fuzzy c-means clustering algorithm, *Computers & Geosciences* 10 (2-3) (1984) 191–203. doi:10.1016/0098-3004(84)90020-7.
- 765 [22] M. Kühne, R. Togneri, S. Nordholm, Robust source localization in reverberant environments based on weighted fuzzy clustering, *IEEE Signal Processing Letters* 16 (2) (2009) 85–88. doi:10.1109/LSP.2008.2009833.
- [23] E. D. Di Claudio, R. Parisi, G. Orlandi, Multi-source localization in reverberant environments, in: *Proceedings of the 10th European Signal Processing Conference, IEEE, Tampere, Finland, 2000.*
- 770 [24] F. J. Huera-Huarte, A. Vernet, Vortex modes in the wake of an oscillating long flexible cylinder combining POD and fuzzy clustering, *Experiments in Fluids* 48 (6) (2010) 999–1013. doi:10.1007/s00348-009-0786-3.
- 775 [25] A. Vernet, G. A. Kopp, Classification of turbulent flow patterns with fuzzy clustering, *Engineering Applications of Artificial Intelligence* 15 (3) (2002) 315–326. doi:10.1016/S0952-1976(02)00037-4.
- [26] M. Vejmelka, P. Musilek, M. Paluš, E. Pelikán, K-means clustering for problems with periodic attributes, *International Journal of Pattern Recognition and Artificial Intelligence* 23 (04) (2009) 721–743. doi:10.1142/S0218001409007338.
- 780 [27] C. M. Anderson-cook, B. Sango Otieno, *Cylindrical Data*, John Wiley & Sons, Ltd, 2006. doi:10.1002/9780470057339.vnn159.
URL <http://dx.doi.org/10.1002/9780470057339.vnn159>

- 785 [28] O. Kesemen, Ö. Tezel, E. Özkul, Fuzzy c-means clustering algorithm for directional data (FCM4DD), *Expert Systems with Applications* 58 (2016) 76–82. doi:10.1016/j.eswa.2016.03.034.
- [29] A. P. Dempster, N. M. Laird, D. B. Rubin, Maximum likelihood from incomplete data via the EM algorithm, *Journal of the Royal Statistical Society. Series B (methodological)* 39 (1) (1977) 1–38.
- 790 [30] M.-S. Yang, J.-A. Pan, On fuzzy clustering of directional data, *Fuzzy Sets and Systems* 91 (3) (1997) 319–326.
- [31] W. Wang, Y. Zhang, On fuzzy cluster validity indices, *Fuzzy Sets and Systems* 158 (19) (2007) 2095–2117. doi:10.1016/j.fss.2007.03.004.
- 795 [32] X. L. Xie, G. Beni, A validity measure for fuzzy clustering, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 13 (8) (1991) 841–847. doi:10.1109/34.85677.
- [33] G. Herold, E. Sarradj, Frequency domain deconvolution for rotating sources on an axial fan (ID: BeBeC-2016-D18), in: *Proceedings of the 6th Berlin Beamforming Conference, Berlin-Adlershof, 2016*.
- 800 [34] N. R. Pal, J. C. Bezdek, On cluster validity for the Fuzzy c-Means model, *IEEE Transactions on Fuzzy Systems* 3 (3) (1995) 370–379. doi:10.1109/91.413225.
- [35] T. F. Brooks, D. S. Pope, M. A. Marcolini, Airfoil self-noise and prediction, *Tech. Rep. 1218, NASA Reference Publication* (1989).
- 805 [36] M. Roger, S. Moreau, Extensions and limitations of analytical airfoil broadband noise models, *International Journal of Aeroacoustics* 9 (3) (2010) 273–305. doi:10.1260/1475-472X.9.3.273.
- [37] T. Carolus, *Ventilatoren, Vieweg+Teubner Verlag, Viesbaden, 2003*. doi:10.1007/978-3-8348-2472-1.
- 810

- [38] B. Bizjan, M. Milavec, B. Širok, F. Trenc, M. Hočevár, Energy dissipation in the blade tip region of an axial fan, *Journal of Sound and Vibration* 382 (2016) 63–72. doi:10.1016/j.jsv.2016.06.036.
- [39] S. A. Khalid, A. S. Khalsa, I. A. Waitz, C. S. Tan, E. M. Greitzer, J. J. Cumpsty, N. A. amd Adamczyk, F. E. Marble, Endwall blockage in axial compressors, *Journal of Turbomachinery* 121 (1999) 499–509. doi:10.1115/1.2841344.