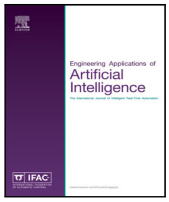
Engineering Applications of Artificial Intelligence 124 (2023) 106635

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Uncertainty clustering internal validity assessment using Fréchet distance for unsupervised learning

Nestor Rendon [a,*], Jhony H. Giraldo [b], Thierry Bouwmans [c], Susana Rodríguez-Buritica [d], Edison Ramirez [a], Claudia Isaza [a]

[a] SISTEMIC, Engineering Deparment, Universidad de Antioquia, Medellín, Colombia
[b] LTCI, Télécom Paris, Institut Polytechnique de Paris, Palaiseau, France
[c] Laboratoire MIA, La Rochelle Université, La Rochelle, France
[d] Alexander Von Humboldt Institute, Cl. 28a 15-09, Bogota, Colombia

ARTICLE INFO

Keywords:
Unsupervised learning
Clustering validity
Fréchet distance
Type-2 fuzzy sets

ABSTRACT

Knowing the number of clusters a priori is one of the most challenging aspects of unsupervised learning. Clustering Internal Validity Indices (CIVIs) evaluate partitions in unsupervised algorithms based on metrics like compactness, separation, and density. However, specialized CIVIs for specific applications have been designed, and there is no general CIVI that works in all scenarios. The absence of CIVIs based on crisp uncertainty metrics is especially critical in decision-making processes that involve ambiguity, non-convex distributions, outliers, and overlapping data. To address this problem, we propose a novel Uncertainty Fréchet (UF) CIVI that assesses the certainty of a well-defined partition. UF leverages uncertainty fingerprints based on Type-2 fuzzy Gaussian Mixture Models (T2FGMM) and the Fréchet distance between clusters to introduce a metric that evaluates partition quality. We integrate UF into a merging methodology that combines similar clusters within a partition, allowing us to determine the number of clusters without the need to run the clustering algorithms iteratively as other CIVIs require. We undertake a comprehensive evaluation of our proposal on 5,250 convex, 36 non-convex synthetic datasets, and five benchmark real datasets. In addition, we apply UF in a real-world scenario that involves high uncertainty: Passive Acoustic Monitoring (PAM) of ecosystems, which aims to study ecological transformations through acoustic recordings. The results show that UF exhibits notable performance in synthetic and real-world scenarios, obtaining an Adjusted Mutual Information (AMI) score higher than 0.88 for normal, uniform, gamma, and triangular distribution datasets. In the PAM application, UF identifies the transformation of ecosystems through sound using clustering algorithms and UF, achieving an F1 score of 0.84. Therefore, results show that the UF index is a suitable tool for researchers and practitioners working with highly uncertain data.

1. Introduction

Unsupervised learning algorithms are used to discover groups and identify distributions with underlying patterns in data without predefined classes (Sakai and Imiya, 2009). Clustering techniques have been shown to be efficient even on high dimensional data (Han et al., 2017). However, the characteristics of multivariate data raise questions about the uncertainty of what constitutes a group, and most clustering algorithms require the number of clusters as a hyperparameter. This leads to the use of Clustering Validation Indices (CVI) that can be used to evaluate the quality of partitions helping to identify the number of clusters. Depending on prior information about data, these indices can be classified according to internal and external criteria. Clustering Internal Validity Indices (CIVIs) are based on cohesion and separation metrics and do not require prior information. External indices compare the clustering results with prior data labels or cluster prototypes (Halkidi et al., 2002b).

CIVIs are used to measure different characteristics of data and are based on different assumptions. Numerous studies have evaluated CIVIs for crisp (Arbelaitz et al., 2013; Halkidi et al., 2002a,b), density (Agrawal et al., 2015a), and fuzzy clustering (Wang and Zhang, 2007). It is widely accepted that no single CIVI can perform well in all applications (Iglesias et al., 2020). Instead, the validation paradigm has shifted towards using specific CIVIs for particular applications based on the assumptions of each index. For instance, CIVIs have been developed for social networks (Campo et al., 2016), pharmacological datasets (Rivera-Borroto et al., 2012), bioinformatics (Handl et al., 2005), genome data (Bolshakova and Azuaje, 2003), medical

* Corresponding author.
E-mail address: nestor.rendon@udea.edu.co (N. Rendon).

https://doi.org/10.1016/j.engappai.2023.106635
Received 30 March 2023; Received in revised form 29 May 2023; Accepted 8 June 2023
Available online xxxx

images (Ouchicha et al., 2018), and other fields. In this context, Passive Acoustic Monitoring (PAM) appears as a specific application of CIVIs that studies endangered ecosystems through acoustic recordings. Research on PAM has highlighted the need for a new CIVI that can measure uncertainty in defining clusters (Rendon et al., 2022b). Although plenty of CIVIs exist, just two CIVIs proposed by Sirmen and Üstöndag (2022) and Ozkan and Türkşen (2012) consider data uncertainty. However, these CIVIs have the constraint of not working for crisp clustering. Wang and Zhang (2007) have demonstrated that incorporating uncertainty metrics in clustering can be beneficial in cases where the distribution shape is unknown, there are outliers present, or the data is overlapping.

In this work, we propose a novel CIVI, named Uncertainty Fréchet index (UF), which uses a modified Wang's Type 2 fuzzy Gaussian Mixture Models (T2FGMM) uncertainty function (Wang and Zhang, 2007) and a Fréchet Distance inspired by the research field of optimal transport (Panaretos and Zemel, 2019), which has been never used in CVI. To evaluate UF's performance, we apply a CVI assessment methodology (Gurrutxaga et al., 2011) that employs various metrics to compare UF with other established CIVIs. We extensively evaluate UF using 5,250 synthetic datasets with different convex distributions, such as normal, uniform, logistic, gamma, and 36 non-convex datasets such as triangular, moon shape, and ring shape, as well as real-world benchmark datasets (Iris, Wine, Breast cancer, Liver, Digits, Haberman, Ionosphere SwedishLeaf, Wafer, ArrowHead, BeetleFly, Car) from the UCI repository (Dua and Graff, 2017) and UCR (Chen et al., 2015) repository and the PAM application dataset (Rendon et al., 2022a). Although UF was initially designed for Gaussian non-isotropic data, our evaluations demonstrate that it achieves remarkable results for uniform, triangular, and non-convex distributions. Our findings indicate that UF outperforms or performs similarly to well-established CIVIs in terms of accuracy, robustness, and stability across a range of datasets. The proposed UF index stands out from other CIVIs in the literature due to two notable features: First, A Footprint Of Uncertainty (FOU) of partitions is quantified, measuring the uncertainty associated with each cluster; Secondly, our proposal measured inter-cluster distances through the Fréchet distance, considering the relationships among variables. Both metrics incorporate the underlying geometry of the space, which has not been explored before in cluster validation research. We integrated these metrics to define the UF CIVI, used in a proposed merging methodology that identifies and joins closely related clusters to estimate the number of clusters. The input of the merging approach is the initial partition obtained through any clustering algorithm. Subsequently, similar clusters are merged using the FOU and Frechet distance metrics. The resulting partition's quality is assessed using the UF to estimate the recommended number of clusters. Our proposal avoids the computational overhead that results from performing the clustering algorithm on each iteration. This approach allows to find the number of clusters maintaining a lower computational efficiency.

The rest of the paper is organized as follows. In Section 2, we review the most commonly used CIVIs in the literature. Our proposal is introduced in Section 3, and the testing framework is described in Section 4. The results and discussion about the conducted tests on both synthetic and real-world datasets are presented in Section 5. Finally, we present conclusions and perspectives in Section 6.

2. Related works and motivations

2.1. Previous works

Several CIVIs estimate the intra-similarity and inter-separation among the clusters in the literature. Cohesion and separation-based indices examine the relationship between within-group scatter/dispersion and between-group scatter/dispersion distances. These metrics are based on cluster centers and typically do not consider the shape of the cluster, thereby failing to capture non-convex clusters. Some authors have addressed this limitation with graphical models such as ideogram methodologies (Iglesias Vázquez et al., 2021), but these methods introduce additional hyperparameters that increase the algorithm's complexity. Cohesion and separation-based CIVIs, such as Silhouette (SI), Davies Bouldin (DB), Gamma (G), and Calinski–Harabasz (CH), have demonstrated the best results for normally distributed datasets. However, SI and CH tend to produce clusters formed by noise (Arbelaitz et al., 2013). Similarly, G, DB, and Dunn Index (DI) penalize noisy cluster behavior at the expense of struggling with non-globular densities. Dispersion-based CIVIs that rely on the dispersion of data are not necessarily the optimal choice for finding convex partitions. Nevertheless, these CIVIs are commonly used with dense, sparse, and arbitrarily shaped datasets. Some cohesion and separation CIVIs, including Partition Coefficient (PC), Partition Entropy (PE), and Interclass Contrast Coefficient (ICC), use only fuzzy membership degrees, considering those clusters with high degrees of membership for data close to the cluster prototype and low degrees for data far from the prototype. These indices are useful when it is not feasible to work with the data directly but with their degrees of belonging to each class (Isaza, 2007). Only Two CIVIs, Relative Uncertainty (RU) (Sirmen and Üstöndag, 2022) and ϵ-stable (Ozkan and Türkşen, 2012), have been proposed in terms of Type-2 fuzzy clustering, which allows working with a higher level of uncertainty regarding the variables. However, these indexes are only proposed for fuzzy partitions.

The performance of CIVIs generally depends on dataset properties. For instance, as the degree of overlapping increases, the performance of most CIVIs tends to decrease, except for Dunn and DB indices. Noise level is the factor that has the greatest impact on CIVIs, degrading their performance by an average of three times (Arbelaitz et al., 2013; Iglesias et al., 2020). Furthermore, non-convex clusters usually pose greater challenges for most CIVIs (Lee et al., 2018). Although CIVIs can be adjusted by varying the number of clusters within a lower and an upper bound, alternative methodologies have been proposed, specialized in dissolving irrelevant and unrepresentative clusters commonly generated by noisy data (Gurrutxaga et al., 2011). In many cases, a lack of certainty regarding data assumptions leads to two problems: clustering algorithms fail due to incorrect tuning or assumptions are violated (Iglesias Vázquez et al., 2021). Consequently, specific clustering applications require data exploration to select a particular CIVI based on data characteristics and assumptions.

In Tables 1 and 2, we summarize and classify CIVIs for crisp and fuzzy clustering-based metrics such as cohesion, separation, density, and uncertainty.

2.2. Motivations

Currently, there exist several advanced machine learning algorithms such as federated learning (FL) (Xiao et al., 2021b; Xing et al., 2022), improved K-means clustering algorithms (Borlea et al., 2022), semi-supervised fault detections (Zhang et al., 2023), Learning Algorithms for Multivariate Data Analysis (LAMDA), multiplicative fuzzy clustering methods (Yapıcı Pehlivan and Turksen, 2021), as well as traditional clustering algorithms for label assignment (Guo et al., 2017) among others. In particular, clustering algorithms are widely used in pipelines to discover patterns in new domains (Guo et al., 2017; Xie et al., 2023a) such as natural language processing applications (Wolf et al., 2020), Eco-acoustics applications (Guerrero et al., 2023), and damage detection and diagnosis for offshore wind foundations (Puruncajas et al., 2020), etc. However, in these algorithms, the number of clusters K is typically assumed to be known as a priori in which there exists a high variety of evaluation metrics (Jaskowiak and Costa, 2023; Anand and Kumar, 2022). In cases where there is no assumption about the data, CIVIs are necessary to determine the number of clusters (Guo et al., 2017; Muranishi et al., 2014; Ma et al., 2021).

CIVIs are highly influenced by the underlying data. Utilizing CIVIs entails performing parameter sweeps and iterative runs of the algorithms, which can be computationally demanding. Furthermore, the

Table 1
Classification of most commonly used CIVIs in crisp cases considering metrics: cohesion & separation, dispersion, and uncertainty.

Crisp metrics				
Cohesion & Separation	Dispersion			Uncertainty
CH (Caliński and Harabasz, 1974)	TW (Friedman and Rubin, 1967)	SCV (Anon, 2020)	BH (Ball and Hall, 1965)	UF (Ours)
DB (Davies and Bouldin, 1979)	W-G (Wemmert et al., 2000)	SDbw (Halkidi et al., 2002a)	KsqDetW (Agrawal et al., 2015a)	
C (Hubert and Schultz, 1976)	GOI (Iglesias et al., 2020)	Si (Jajuga et al., 2020)	RL (Ratkowsky and Lance, 1978)	
MR (Zhang et al., 2018)	BR (Banfield and Raftery, 1993)	Dunn (Dunn, 1974)	RT (Ray and Turi, 2000)	
PBM (Towsey et al., 2014a)	CVNN (Liu et al., 2012)	GSCV (Xu et al., 2020)	TW (Cureton et al., 2010)	
PB (Hands and Everitt, 2010)	DCVI (Xie et al., 2019)	NDR (Agrawal et al., 2015a)	Connectivity (Handl et al., 2005)	
SD (Liang et al., 2020)	LCCV (Cheng et al., 2018)	GD (Dunn, 1974)	SSDD (Liang et al., 2020)	
AHC (Zhou et al., 2017)	CVTAB (Fu and Wu, 2017)	α-elbow (Shi et al., 2021)		
Gamma (Jain and Dubes, 1988)				

Table 2
Fuzzy metrics used in Internal Clustering Validity Indices (CIVIs) considering metrics: cohesion & separation, dispersion, and uncertainty.

Fuzzy metrics		
Cohesion & Separation		Uncertainty
YDI (Kim and Ramakrishna, 2005)	Voc (Rizman Žalik, 2010)	BM (Liu et al., 2019a)
MFCM (Sadeghi and Etemadfard, 2022)	PC (Bezdek et al., 1984)	MPC and MPE (Silva et al., 2015)
Gplus (Iglesias et al., 2020)	XB (Xie et al., 2020)	RU (Sirmen and Üstöndag, 2022)
PBMF (Towsey et al., 2014a)	PE (Bezdek et al., 1984)	ϵ-Stable (Ozkan and Türkşen, 2012)
VIMI (Liu et al., 2021)	VHY (Wang et al., 2021)	
MPC (Liu et al., 2019a)	CV (Isaza, 2007)	
CSBM (Liu et al., 2019c)	ICC (Franco et al., 2002)	
VECS (Ouchicha et al., 2020)	CRITIC (Wang et al., 2023)	

distribution of data can be uncertain, posing challenges to the performance of CIVIs. To alleviate these challenges, it is crucial to develop a CIVI computationally lightweight, capable of handling uncertainty in unsupervised applications, and offering interpretability, allowing for the understanding of the resulting partitions. Measuring uncertainty based on the intrinsic characteristics of data can provide insight into the number of clusters present, particularly for applications where information about cluster characteristics such as noise level, overlap, and distribution is unavailable (Wang and Zhang, 2007). Therefore, it is essential to incorporate information on uncertainty into unsupervised models (Sirmen and Üstöndag, 2022), since data characteristics such as labels, density, and noise are not typically accounted for and are essential for each group to establish the extent of their reliability. Fuzzy Type-2 has emerged as a promising solution for handling uncertainty in data, which has shown potential in clustering applications by measuring the uncertainty of primary membership functions (Wang and Zhang, 2007). Achieving these characteristics is critical for the effective deployment of CIVIs in real-world scenarios. Thus, in the proposed methodology, we incorporate properties of the Fuzzy Type-2 theory to obtain information for measuring the quality of partitions.

Previous work stresses the importance of developing new methodologies and evaluating the performance of new CIVIs across various applications (Arbelaitz et al., 2013; Xu et al., 2020). To address these challenges, we propose the Uncertainty Fréchet (UF) index, which can identify both convex and non-convex clusters under uncertain data assumptions. UF is based on Type-2 fuzzy sets, but it is designed to handle a large number of data points and clusters without relying on the membership degrees of the data. To find the number of clusters, we propose a merging methodology based on UF, as a metric to establish the number of clusters. The following section presents the UF index and how we use it within the merging methodology to find the number of clusters.

3. Uncertainty Fréchet index

3.1. Preliminaries

In this work, uppercase boldface letters such as $\mathbf{A}$ represent matrices, calligraphic letters like $\mathcal{X}$ represent sets, lowercase boldface letters such as $\mathbf{x}$ denote vectors, and $(\cdot)^{\mathsf{T}}$ represents transposition. To facilitate reading, Table 3 shows a list with the variables and abbreviations used in this paper.

Table 3
Variables and abbreviations.

Variable	Meaning
$\mathbf{A}$	Uppercase boldface letters represent matrices
$\mathcal{X}$	Calligraphic letters represent sets
$\mathbf{x}$	Lowercase boldface letters denote vectors
$(\cdot)^{\mathsf{T}}$	Transposition
n	Number of standardized elements in the set $\mathcal{X}$
k	Ground truth number of clusters
nc	Number of clusters elected to make the partition
$\mathbf{U}$	Clustering partition matrix of size $nc \times n$
C_i	Cluster i
u_{ij}	Membership of the sample $\mathbf{x}_j$ to the cluster C_i
$\boldsymbol{\mu}_i$	Mean of the ith Gaussian distribution
$\boldsymbol{\Sigma}$	Covariance matrix
M	Number of Gaussian distributions of the dataset
$\mathbf{w}_i$	Weights of each Gaussian density component
k_{uf}	Best number of clusters
m	Number of partitions
S	Set of partitions
$I(C_i)$	Value of the evaluated CIVI for partition Ci
C^I	Best partition, where $C^I = \operatorname{argmax} C_i \in S(I(C_i))$
CIVIs	Clustering Internal Validity Indices
T2FGMM	Type-2 fuzzy Gaussian Mixture Models
UF	Uncertainty Fréchet
AMI	Adjusted Mutual Information
PAM	Passive Acoustic Monitoring

The purpose of clustering algorithms is to classify a data set $\mathcal{X} = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n\}$ composed of n standardized elements, where $\mathbf{x}_i$ is the ith sample and k is the ground truth number of clusters. The clustering partition matrix $\mathbf{U}(A)$ of size $nc \times n$, is represented by $\mathbf{U} = [u_{ij}]_{nc \times n}$, where nc is the number of clusters elected to make the partition, $i = 1, \ldots, nc$ and $j = 1, \ldots, n$. u_{ij} is the membership of the sample $\mathbf{x}_j$ to the cluster C_i. In crisp partition $u_{ij} = 1$ if $\mathbf{x}_j \in C_i$, otherwise $u_{ij} = 0$. In practice, two kinds of clustering exist as follows:

- For crisp clustering, the purpose is to classify a dataset $\mathcal{X}$ such that: $C_i \neq \varnothing$, $C_i \cap C_k = \varnothing$ for $i, k = 1, \ldots, nc$, and $i \neq k$, $\bigcup_{i=1}^{nc} C_i = \mathcal{X}$.
- For fuzzy clustering, $\mathbf{U} = [u_{ij}]_{nc \times n}$ where $u_{ij} \in [0, 1]$ denotes the membership degree of j_{th} element to i_{th} cluster, with the next

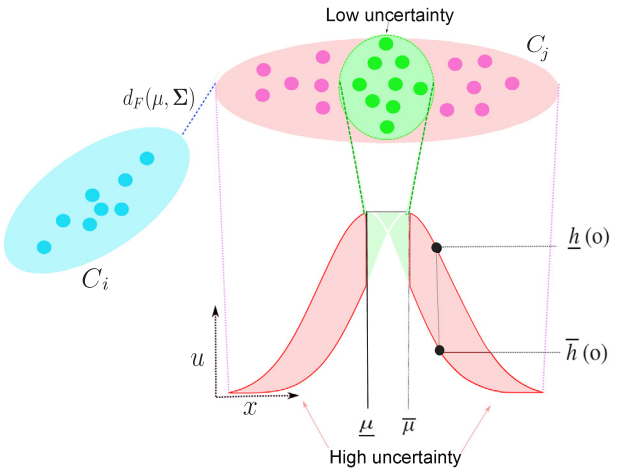


Fig. 1. Uncertainty metrics. Considering two clusters C_i and C_j, FOU can be calculated through the degrees of membership. The green area represents high degrees of intra-class membership or low uncertainty. The pink area represents high degrees of uncertainty. The Fréchet distance between clusters is represented by $d_F(\mu, \Sigma)$. The distribution shows the Gaussian primary membership function with uncertain mean as an interval $[\underline{h}(x), \overline{h}(x)]$, where the FOU is the green and red shaded regions.

conditions: $0 < \sum_{j=1}^{n} u_{ij} < n$ for $j = 1, \ldots, nc$, $\sum_{i=1}^{nc} u_{ij} = 1$ for $j = 1, \ldots, n$, $\sum_{i=1}^{n} \sum_{j=1}^{n} u_{ij} = n$.

UF assumes the low uncertainty hypothesis: high cluster memberships and low for other clusters guarantee better partitions through a merging methodology. Fig. 1 shows the properties of our proposal, the Footprint of Uncertainty (FOU) to measure the cluster's uncertainty, and the Fréchet distance to measure distance among clusters. We introduce the proposed metrics and UF in the following subsections.

3.2. Modified footprint of uncertainty as compaction metric

Our proposal analyzes the Footprint of Uncertainty (FOU) (Zeng et al., 2008) which provides a theoretical well-funded approach for handling the uncertainty of Gaussian distributions. FOU is based on T2 membership functions of Gaussian Mixture Models (GMM) which is represented as the sum of weights of Gaussian density components as follows:

$$P(\mathbf{x}_i) = \sum_{i=1}^{M} \frac{\mathbf{w}_i}{(2\pi)^{D/2}|\mathbf{\Sigma}_i|} \exp\left[-\frac{1}{2}(\mathbf{x}_i - \boldsymbol{\mu}_i)^{\mathsf{T}} \mathbf{\Sigma}_i^{-1} (\mathbf{x}_i - \boldsymbol{\mu}_i)\right], \tag{1}$$

where D is the number of dimensions of each object, $\boldsymbol{\mu}_i$ is the mean, $\mathbf{\Sigma}$ the covariance matrix, M is the number of Gaussian distributions of the data set, and $\mathbf{w}_i$ is the weight of each Gaussian density component (Reynolds et al., 2000).

The hyperparameters of the GMM algorithm are denoted as $\lambda = \mathbf{w}_i, \boldsymbol{\mu}_i, \mathbf{\Sigma}_i$. Each parameter is estimated through the Expectation-Maximization algorithm, according to the Maximum-Likelihood (ML) criterion. However, the GMM cannot accurately reflect the underlying distribution of observations. Then, introducing descriptions of uncertain hyperparameters can help the clustering algorithms' performance. Type-2 fuzzy GMM (T2FGMM) considers interval likelihoods of GMMs to describe the observation uncertainty whose computations involve only arithmetic intervals of the data which can provide additional information for pattern classification.

Fig. 1 shows the likelihood of the T2 FGMM becoming an interval with uniform possibilities where $[\underline{h}(x), \overline{h}(x)]$ are the two interval primary membership grades. Zeng et al. (2008) uses the length between two log-likelihood bounds $H(x) = [ln\,(\underline{h}(x)) - \ln(\overline{h}(x))]$, to define the T2 FGMM with an uncertain mean vector. Then, given a one-dimensional observation $\mathbf{x} = x_1, x_2, \ldots, x_D$, and the mean vector $\boldsymbol{\mu}$, the T2 FGMM is defined by:

$$H(x_{iD}) = \begin{cases} \frac{2k_m|x_{iD}-\mu_j|}{\sigma_j} & \text{if } 0 \le \mu - k_m\sigma \text{ or } 0 \ge \mu + k_m\sigma, \\ \frac{|x_{iD}-\mu_j|^2}{\sigma_j} + \frac{k_m|x_{iD}-\mu_j|}{\sigma_j} + \frac{k_m^2}{2} & \text{if } \mu - k_m\sigma < 0 < \mu + k_m\sigma, \end{cases} \tag{2}$$

where μ and σ are the mean and standard deviation of the original certain T1 membership function without uncertainty. k_m constants control the hyperparameters of the FOU. One dimensional Gaussian has 99.7% of its probability mass in the range of $[k_m - 3\sigma, k_m + 3\sigma]$, consequently, we constrain $k_m = 0.3$. Bigger values imply high degrees of freedom to account for uncertainty. Then our proposed method considers the information about uncertainty without using fuzzy hyperparameters. With the FOU we propose to quantify the $H(x)$ in all the clusters. However, this metric is calculated on each dimension of $\mathbf{x}$, so it can lose information about variables covariance which is a drawback in almost all CIVIs. Due to that, $H(x)$ is a metric defined with the terms of $Zscore = \frac{|x-\mu_j|^2}{\sigma_j}$ for one dimension, we define the Partition Footprint Of Uncertainty (PFOU) that quantify the uncertainty for multidimensional real space:

Definition 1 (*PFOU*)**.** The partition generalization for $\mathbb{R}^N$ of quantification of uncertainty $H(x_{iD})$ (Zeng et al., 2008) is defined as:

$$PFOU = \begin{cases} \sum_j \sum_D 2k_m d_M(x, \mathbf{\Sigma}) & \text{if } 0 \le \mu - k_m\sigma \text{ or } 0 \ge \mu + k_m\sigma, \\ \sum_j \sum_D d_M(x, \mathbf{\Sigma}) + k_m d_M(x, \mathbf{\Sigma}) + \frac{k_m^2}{2} & \text{if } \mu - k_m\sigma < 0 < \mu + k_m\sigma \end{cases}, \tag{3}$$

where $d_M(\mathbf{x}, \Sigma)$ is the Mahalanobis distance defined by:

$$d_M(\mathbf{x}, \Sigma) = \sqrt{(\mathbf{x} - \boldsymbol{\mu})^{\mathsf{T}} \mathbf{\Sigma} (\mathbf{x} - \boldsymbol{\mu})}, \tag{4}$$

The definition of PFOU is based on replacing each $Zscore$ with the Mahalanobis distance, which is the equivalent of $Zscore$ for more than one dimension. As PFOU is a fuzzy-oriented measure, it is considered useful for high-uncertainty problems such as PAM. A comparison analysis using PFOU and the Silhouette index compaction metric (intracluster pairwise distance) is included in Appendix A. According to the analysis, it was decided to use PFOU due to its potential to quantify uncertainty.

3.3. Fréchet distance as a separation metric

We observe that the problem of the distance between clusters can be framed as an optimal transport problem (Panaretos and Zemel, 2019). This ubiquitous problem seeks the minimum effort required to reconfigure the probability mass of one distribution CD_i in order to recover another distribution CD_j. One of the distances inspired by the optimal transport problem is the Wasserstein distance, defined in the p-moment by:

$$W_p(\boldsymbol{\mu}, \upsilon) = \inf_{X \sim \mu, Y \sim \upsilon} \left(\mathbb{E}\|CD_i - CD_j\|^p\right)^{1/p}. \tag{5}$$

Particularly, in the closed-form formulae case where there is a separable real infinite dimensional Hilbert space, with $\boldsymbol{\mu}$ and υ tending to be Gaussian, the Wasserstein distance between two probability distributions $(j, \hat{j})$ it is normally called the Fréchet distance and is defined by:

$$d_F(\boldsymbol{\mu}, \mathbf{\Sigma}) = |\mu_j - \mu_{\hat{j}}|^2 + \mathrm{tr}\,|\mathbf{\Sigma}_j - \mathbf{\Sigma}_{\hat{j}} - 2(\mathbf{\Sigma}_j \mathbf{\Sigma}_{\hat{j}})|. \tag{6}$$

The Fréchet distance considers the relationship between the variables through $\mathbf{\Sigma}$, and for families of distributions that are not necessarily Gaussian, including the underlying structure of the data, which is beneficial in what clustering methods refer to Parsa et al. (2020). In terms of CIVIs, any of the inter-cluster distance metrics proposed in the

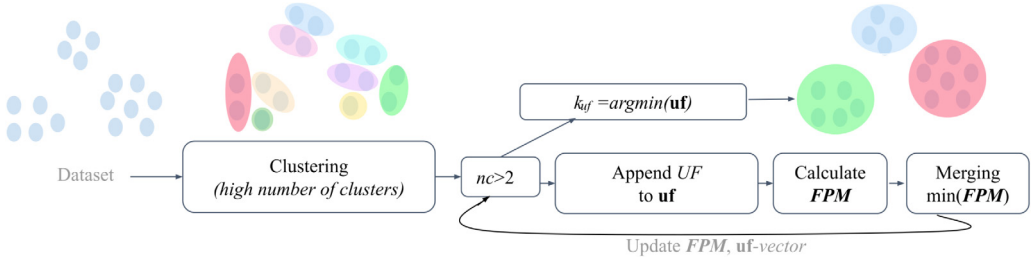


Fig. 2. The UF merging methodology begins by clustering the data into a large number of clusters. In each subsequent iteration, the Fréchet-Pairwise Matrix (FPM) is used to identify the closest pair of clusters, which are then merged using the UF criterion. The minimum value of UF obtained during the iterations determines the optimal number of clusters.

literature capture the underlying structure of the cluster space. For this reason, we derive the Fréchet inception distance for all pair distances of all clusters as follows:

$$d_F(\mu, \Sigma) = \sum_{j} \sum_{\hat{j}} |\mu_j - \mu_{\hat{j}}|^2 + \mathrm{tr}\,|\mathbf{\Sigma}_j - \mathbf{\Sigma}_{\hat{j}} - 2(\mathbf{\Sigma}_j \mathbf{\Sigma}_{\hat{j}})|, \tag{7}$$

where $j \neq \hat{j}$.

3.4. Merging methodology to calculate the clusters number k

We search for the minimization of Intra-cluster Uncertainty ($IU(C_i)$) of the cluster C_i $\forall\ 0 \leq i \leq nc$, and the maximization of the Separation between Groups (SG) for clusters C_i, C_j with $i \neq k$. Then, our objective function is:

$$\min_{k_{nc}} \frac{IU(C_i)}{SG(C_i, C_j)} \quad subject\ to:$$

$$C_i \neq \emptyset, C_i \cap C_j = \emptyset\ for\ i, j = 1, \ldots, nc,\ i \neq j, \quad \bigcup_{i=1}^{nc} C_i = \mathcal{X},$$

Replacing $IU(C_i)$ with $PFOU$, $SG(C_i, C_k)$ with $d_F(\mu, \Sigma)$, dividing by minimum Fréchet distance ($d_f min$) and multiplying by the maximum Fréchet distance ($d_f max$) for standardization, we define the Uncertainty Fréchet (UF) index as our objective function:

$$UF = \frac{PFOU d_f max}{d_F(x, \mathbf{\Sigma}) d_f min}, \tag{8}$$

To find the best number of clusters we need to find the argument that minimizes UF:

$$k_{uf} = \arg\min_{k_{nc}} UF(k_{nc}) \quad subject\ to:$$

$$C_i \neq \emptyset, C_i \cap C_j = \emptyset\ for\ i, j = 1, \ldots, nc,\ i \neq j, \quad \bigcup_{i=1}^{nc} C_i = \mathcal{X}, \tag{9}$$

as we search to minimize UF, we introduce the heuristic merging methodology presented in Fig. 2.

The methodology is agnostic to the clustering method, does not require hyperparameters tuning, adds information related to non-convex clustering, and reduces computational time. The agglomerate heuristic dissolves clusters with low-density distances and chooses the number of clusters based on UF. The first step is to apply clustering on $\mathcal{X}$ with a large number of clusters and calculate $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$ for each one. The next step is the calculation of UF. While nc is greater than two, then calculate the Fréchet-Pairwise Matrix (**FPM**) in which each element is the Fréchet distance ($d_F(\mu, \mathbf{\Sigma})$) for each pair of clusters. The next step is to choose those of the lower element of **FPM**(h, l), which represent the closest clusters. The next step is to update the **FPM** by merging the two closest clusters. To update the matrix, the next steps should be followed: first, the value of UF is added to the slope vector **uf**. Second, update of $\boldsymbol{\mu}$ and $\mathbf{\Sigma}$. Update the FPM: delete the columns and rows of the small pairs clusters. Finally, when $nc = 2$, check for the minimum value of **uf** which corresponds to the recommended number of clusters k_{uf}.

The number of iterations depends solely on the initial number of clusters. As clusters are merged in pairs until only one cluster remains, the number of iterations will be equal to the initial number of clusters. The clustering algorithm is only executed once during the first iteration, during which the pairwise Frechet matrix is calculated. In each subsequent iteration, only the columns and rows of merged clusters are updated and the UF is calculated. Our approach differs from other CIVIs, which require the clustering algorithm to be run in each iteration. We summarize the proposed methodology in the algorithm 1.

Algorithm 1 UF Merging Algorithm

Require: **X**
1: Perform clustering with a large number of nc and calculate $\mathbf{\Sigma}, \mu$ for each cluster
2: **while** $nc > 2$ **do**
3: Calculate UF and append in **UF**,
4: Compute the $\mathbf{FPM}_{(nc \times nc)}$
5: Identify the most closest clusters represented by h and l positions of **FPM**,
6: Merge the clusters of positions k, l in to a single element
7: Update the $\mathbf{\Sigma}$, μ, labels, $\mathbf{FPM}_{(nc-1 \times nc-1)}$
8: **end while**,
9: **return** k_{uf} =argmin (**UF**)

4. Experimental framework

This section details the testing of UF on multiple datasets with varying data distributions, including both convex and non-convex types. To evaluate UF's effectiveness, we conduct tests on different data features, such as noise, compactness, and overlapping. By manipulating these features, we sought to thoroughly evaluate the performance of UF across a range of scenarios. In summary, we test our proposal with three case studies: synthetic datasets, the real-world case study of Passive Acoustic monitoring datasets, and the case study with publicly available real-world data. We explain each of these case studies in the following subsections.

4.1. Synthetic datasets

We test UF with synthetic convex and non-convex datasets. We generate the data using combinations of the parameters presented in Table 4 and Table 5 to generate 5,250 convex and 36 non-convex synthetic datasets respectively.

Each generated dataset has 5,000 samples each with different distributions uniform, normal, logistic, gamma, triangular, ring-shape, and moon-shape. For data generation, we used the MDCenpy framework (Iglesias et al., 2019) and Scikit-learn with make-moons and make rings functions (Pedregosa et al., 2011).

4.2. Benchmark datasets

We test our methodology in common real-world datasets as benchmarks commonly used to test CIVIs. We use Iris, Wine, Breast cancer,

Table 4
Combination parameters to generate convex datasets.

	Parameters
Samples	5,000
Number of features	2, 10, 20, 40, 100, 200
K	2, 3...40
Distributions	Normal, Gamma, Logistic, triangular
Compactness Factor	0.1, 0.3
Alpha_n	1, 5
Noisy Samples	0 (only for normal), 500

Table 5
Values of parameters to generate non-convex datasets.

	Moon	Circles
N of Clusters	2 to 20	2
Sigma	0.1	0.01
Radius	0	–
Noise	–	0.05, 0.15, 0.25
Factor	–	0.04: 0.04: 0.24
Dimensions	2	–
Total number of test	18	36

Table 6
Characteristics of real-world datasets.

Dataset	Nsamples	Features	Classes
Wine	178	13	3 or 2
Iris	150	4	3
Cancer	569	31	2
Liver	566	11	3
Digits	1797	65	10
Haberman	306	3	2
Ionosphere	351	34	3
Spambase	4601	57	2
SwedishLeaf	500	128	15
Wafer	1000	152	2
Arrowhead	36	251	3
BeetleFly	40	512	2
Car	577	60	4

Digits, Haberman, Ionosphere, and Liver dataset from the UCI repository (Dua and Graff, 2017). Also, Inspired by Xiao et al. (2021a), Guo et al. (2017), and Ma et al. (2021), we include the following time series datasets: Swedish Leaf, Wafer, Arrowhead, Beet Fly, and Car datasets of UCR 2018 which is a widely used time series repository (Xing et al., 2022; Guo et al., 2017)

See Table 6 to see the principal characteristics of each one.

In the case of time-series data, there are usually local and global patterns. New algorithms, such as Deep Temporal Clustering Representation (DTCR) (Guo et al., 2017), and Robust Temporal Feature Network (RTFN) for feature extraction related to time series classification (Xiao et al., 2021a), have been able to deal with this type of data. However, one of the challenges for recent functional clustering algorithms for time series (Guo et al., 2017) is to increase the performance and deal with missing data. Therefore, we employ the recently proposed Clustering Method Representation Learning on Incomplete time-series data (Guo et al., 2017; Ma et al., 2021) in the time-series evaluation, as UF is capable of being agnostic to the clustering algorithm.

4.3. Passive acoustic monitoring application

Passive Acoustic Monitoring (PAM) is a particular application that needs to consider uncertainty metrics to evaluate the obtained partitions. PAM uses recordings of soundscapes collected at ecosystem sites that are analyzed through clustering algorithms for analyzing community dynamics and landscape transformation. The recorded soundscape is composed of an amalgam of sounds such as anthrophony: human-generated sources, geophony: earth sounds such as rain, wind, and

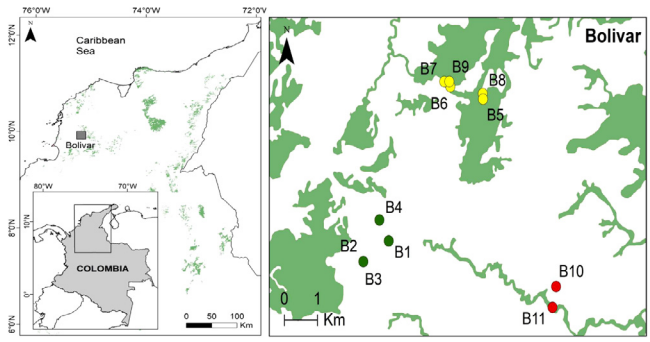


Fig. 3. PAM study area situated in the Caribbean region of Colombia. Sites B1,.., B11, tagged with red, yellow, and green points represent high, medium, and low transformation.

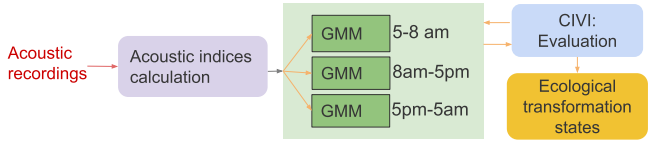


Fig. 4. PAM methodology: The first step is calculating the acoustic indices with the acoustic recordings. Then we train GMM of recordings for each stage of the day (5–8 am, 8 am–5 pm, and 5 pm–5 am) obtaining 3 GMM models. For each model, a clustering evaluation is performed to obtain the number of clusters. Finally, an analysis of the ecological transformation is made.

thunders, and biophony: biologically generated sounds. Particularly, there are data for biological phenomena, that clustering could confuse with outliers, noise, or vice versa, as well as high uncertainty for ecologists to identify the type of transformation of a point. Therefore, in this application a CIVI based on uncertainty is crucial. Some CIVIs were tested on PAM studies (Rendon et al., 2022b), finding that none has shown good performance to define a good evaluation. Consequently, we decided to test UF in a PAM dataset of 23,540 acoustic recordings supplied by the Alexander von Humboldt Biological Resources Research Institute of Colombia. The acoustic recordings were collected between 2015 and 2017 in the Bolivar department of Colombia's Caribbean region as part of the Global Environment Facility (GEF) project (Rendon et al., 2022a). The study locations are located in a tropical dry forest ecosystem, which is situated below 1,000 meters above sea level, highly seasonal in its rainfall with dry periods of at least three months. Twelve sites were sampled over a landscape transformation gradient in the Arroyo Grande (Bolivar) area as shown the image 3. The recordings were obtained using Wildlife Acoustics' SM2 and SM3 recorders, which were set up to record for five days nonstop at a rate of 5 minutes every 10 minutes. Prior to field campaigns and after a landscape transformation analysis was conducted, each site was classified as high, medium, and low transformation. High transformation corresponds to sites with a low proportion of retained forest and a high proportion of lost forest between 2000 and 2016. Low transformation sites have a high proportion of retained forest and a low proportion of lost forest. The medium transformation corresponds to sites between these two extremes. Although discrete states are commonly used in ecology for site evaluation, they are not actually present in nature. In this regard, finding transition states could help to a better understanding of ecosystem health.

These ecological transformation labels were used to compare the results of unsupervised algorithms and internal validation indices through the F1 score. In PAM applications it is expected to find new groups based on the ground truth, but without falling into monotonicity or lack of performance. For the experiment, we follow the methodology in Rendon et al. (2022a) that is described in Fig. 4.

We use 15 acoustic indices as features that describe the soundscape complexity as variables: 'acift' (Farina et al., 2016), 'beta' (Boelman

et al., 2007), 'ndsi', 'p', 'm' (Depraetere et al., 2012), 'np' (Depraetere et al., 2012), 'mid' (Depraetere et al., 2012), 'bnf' (Towsey, 2013), 'md' (Coensel, 2007), 'fm' (Towsey et al., 2014b), 'sf' (Ellis et al., 2015), 'rms' (Towsey et al., 2014b), 'sc' (Ellis et al., 2015), 'tonnets' (Ellis et al., 2015). We use 3-day periods due to the sound variability of day hours. Then, we use 3 GMM clustering models, one for each period.

4.4. Evaluation metrics

The evaluation of CIVIs is usually done concerning expected labels. That is, given a set of different values for parameter nc, m partitions $S = C_1, C_2, \ldots, C_m$ are obtained where only one has the best number of clusters (k_{uf}). Generally, in the evaluation, the CVIs are computed for each partition where the maximization or minimization of the CIVI is sought to choose the 'best' partition. For example, for the maximization case, the function $I(C_i)$ computes the value of the evaluated CIVI for the partition C_i, where $C^I = argmax_{C_i \in S}(I(C_i))$ is the best partition. This evaluation approach works through the "correctness assumption" (Gurrutxaga et al., 2011) which establishes that clustering algorithms work correctly. In other words, of m partitions the best partition is the one closest to k. However, there has been evidence that clustering algorithms do not always determine the correct partitions (complex environments such as noisy datasets, and over-sampled clusters). In 2011, Gurrutxaga et al. (2011) proposed a new evaluation methodology, where the best partition must be chosen using different similarity metrics depending on the criteria evaluation. In this work, we use the three following metrics:

1. The Adjusted Mutual Information (AMI) (Vinh et al., 2010) measures the degree of similarity of two labels for the same data set independently of the label values. While other metrics like the Rand index score are available, some studies like Romano et al. (2016), Guo et al. (2017) have shown that such measures can be adversely affected by imbalanced data or a high number of clusters. Therefore, we use AMI to evaluate synthetic convex cluster quality (see Fig. 5). We aim to assess the partitions by considering the CIVI performance relative to the clustering algorithm and penalizing an incorrect number of clusters evaluation.
2. The "number of clusters score" is a metric that measures how well the clustering results align with the ground truth labels in terms of the number of clusters. In our evaluation of both benchmarks and nonconvex datasets, we use this metric to determine the best partition, considering the shapes and capacity of CIVIs, regardless of the clustering algorithm used. Fig. 6 illustrates this approach. Once we have identified the optimal number of clusters, we evaluate the partition by looking at the number of cluster hits.
3. The F1 score is based on the harmonic mean of precision and recall (Sokolova et al., 2006). The cluster labels do not necessarily correspond to real labels, so we use a contingency matrix to evaluate the F1 score.

In PAM, the ground truth gives an idea of clusters, but it may not necessarily correspond to the ground truth number of partitions. For example, suppose that experts label some sites with discrete states according to the ecological transformation: high, medium, and low. Clustering methods help to find transitions among these states and discover more clusters than the ground truth. Therefore, the evaluation must consider the ground truth but aim to analyze better new clusters, as shown in Fig. 7. Nevertheless, Metrics like AMI do not serve this purpose. Then, to evaluate PAM, we use the cluster best-fit combination F1 score as the evaluation metric. We also consider the number of clusters for tracking monotonicity.

There are two recent approaches related to measuring uncertainty for partition evaluation RU (Sirmen and Üstöndag, 2022), and ϵ-Stable (Ozkan and Türkşen, 2012), both CIVIs have the limitation that they require fuzzy membership degree hyperparameters. Therefore, we compare our proposal with the best-performance CIVIS in the literature for synthetic and real-world datasets. The CIVIs are the following: SI (Jajuga et al., 2020), RTL (Ratkowsky and Lance, 1978), BH (Ball and Hall, 1965), SD (Dudek, 2020), DB (Dudek, 2020), XB (Muranishi et al., 2014), CH (Caliński and Harabasz, 1974), Dnn (Dunn, 1974), MCR (Zhang et al., 2018), PB (Hands and Everitt, 2010), RT (Ray and Turi, 2000), WG (Wemmert et al., 2000), C (Hubert and Schultz, 1976). We did not consider other recent CIVIs in the comparison due to their limitations concerning their underlying assumptions. However, we acknowledge their potential under specific usage conditions. For instance, the absolute GOI indices (Iglesias et al., 2020) capture the essence of geometric measurements and provide a coherent interpretation of the data structure. However, their effectiveness depends on the input context, self-tuning strategies, and the interpretation of the output indices. The GSCV (Xu et al., 2020) demonstrated great potential but only for hierarchical clustering algorithms. Finally, we aim to achieve independence from the type of clustering algorithm used. Therefore, we do not consider CIVIs that are not agnostic to clustering algorithms, such as the CIVI α-elbow (Shi et al., 2021), or CIVIs that rely on fuzzy-type algorithms, such as CRITIC (Wang et al., 2023), RU (Sirmen and Üstöndag, 2022) and VHY (Wang et al., 2021).

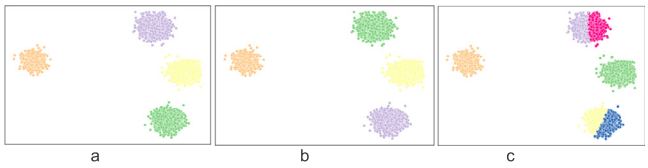


Fig. 5. Comparison of evaluation metrics for convex datasets. (a) Real clusters. (b) Clustering with k: 4, with AMI: 1, F1 score: 1, and a number of clusters score:1. All evaluation metrics assume good-quality clusters. (c) Clustering with k: 4, with AMI: 0.84, F1 score: 1, number of clusters score:0. AMI performance is affected, since k does not correspond to the real number of clusters, in contrast to F1 score, and n of clusters score that are not affected.

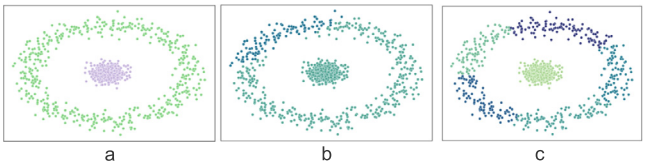


Fig. 6. Comparison of evaluation metrics for convex datasets. (a) Real clusters. (b) Clustering with k: 4, with AMI: 1, F1 score: 1, and a number of clusters score of 1. All evaluation metrics assume good-quality clusters. (c) Clustering with k: 4, with AMI: 0.84, F1 score: 1, number of clusters score:0. AMI performance is affected, since k does not correspond to the real number of clusters, in contrast to F1 score, and n of clusters score that is not affected.

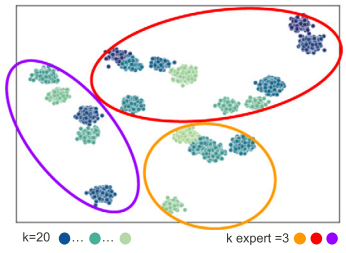


Fig. 7. Evaluation metrics for PAM datasets. PAM datasets consist of many sound elements with different kinds of distributions. Experts provide a general idea of the present groups, but each group can contain subgroups. In this example, yellow, red, and purple circles represent the groups provided by the experts, while the real number of clusters is 20 (shown in green shades). Due to the ground truth, metrics such as AMI and the number of clusters score may not accurately evaluate the partition. In contrast, the F1 score enables intracluster evaluation.

Table 7
The summarized results for the tested datasets are presented below. The best scores are highlighted in bold.

CIVI	Normal (not noise) AMI	Noise convex AMI	Real benchmarks AMI	PAM F1score	Non-convex hits	Non-convex AMI
SI (Jajuga et al., 2020)	0.980	0.965	0.323	0.647	0	0.652
BH (Ball and Hall, 1965)	**0.980**	**0.965**	0.284	0.647	0	0.652
RTL (Ratkowsky and Lance, 1978)	0.944	0.944	0.275	0.803	0	0.566
SD (Dudek, 2020)	0.955	0.954	0.333	0.640	0	0.644
XB (Muranishi et al., 2014)	0.955	0.953	0.287	0.787	0	0.557
MCR (Zhang et al., 2018)	**0.980**	0.945	0.294	0.800	4	0.785
WG (Wemmert et al., 2000)	0.956	0.956	0.277	**0.810**	0	0.557
PB (Hands and Everitt, 2010)	**0.980**	0.959	0.273	0.560	5	0.783
UF (Ours)	**0.960**	**0.864**	**0.495**	**0.803**	35	**0.815**

We use Gaussian Mixture Models GMM (Expectation Maximization) to test the performance on real and synthetic datasets as it showed a proficiency for datasets with high uncertainty (Rendon et al., 2022b). To ensure a fair comparison of the CIVIs, we evaluate them using identical hyperparameters across all clustering algorithms and datasets. For the GMM, we used the following hyperparameters: covariance type: full, convergence threshold: 1e-3, and K-means initialization method to define means and covariance matrices. In the case of time-series data, we applied Clustering Representation Learning on the Incomplete time-series (CRLI) algorithm (Xie et al., 2023a) with the following hyperparameters, n generator layers=2, RNN hidden size=64 and epochs=5.

5. Results

We designed a GUI Python application that calculates the mentioned CIVIs (including our proposal) to run the proposed tests. The code is available in: https://github.com/David9203/CVI-Frechet-Mahalanobis.[1] The link includes the real-world PAM dataset and the code to generate the synthetic data. The real-world data from benchmarks (Iris, Wine, Breast cancer, Digits, Haberman, Ionosphere, Liver, Swedish Leaf, Wafer, Arrowhead, Beet Fly, and Car) are available at UCI repository[2] and UCR repository.[3]

The computing system used was a GenuineIntel CPU family 64-bit Byte, specifically an Intel(R) Core(TM) i7-6800K CPU running at a clock speed of 3.40 GHz, 32 GB of RAM, 12 CPU(s) with two thread(s) per core. In the following subsections, we describe and analyze each one of the tests (see Table 7).

5.1. Synthetic datasets

To evaluate the performance of CIVIs on synthetic data, as described in Section 4.1, we use the number of clusters score for non-convex data and AMI for convex data, as outlined in Section 4.4. We evaluate a total of 5,286 synthetic datasets using the aforementioned metrics, and the summarized results are presented in the following sections.

5.1.1. Convex clusters

Fig. 8 shows normal convex data without noise. As the graph shows UF, SI, BH, WG, MR and XB obtain AMI performance higher than 0.99.

Fig. 9 shows best-performed indices in convex-noisy datasets. For normal, uniform, and triangular distributions with noise, each one with 900 datasets generated. We have evidence that there is an AMI higher than 0.90 for UF, SI, BH, SD, XB WG, and MR. UF's performance obtained an AMI score of less than 0.75 for gamma and logistic distributions.

Regarding the noisy-convex distributions, the factors which reduce performance are the overlapping, the compaction factor, and the higher

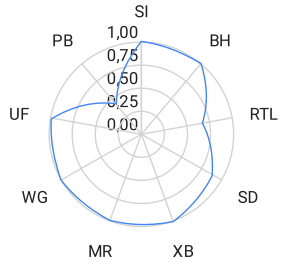


Fig. 8. CIVIs AMI performance for convex data broken down by normal distribution.

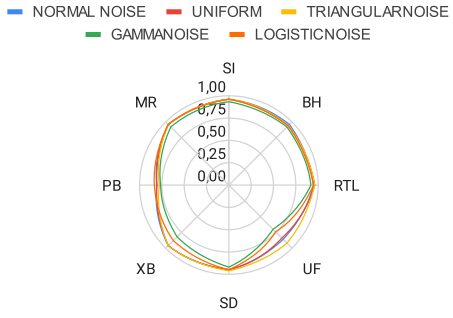


Fig. 9. CIVIs AMI performance for noisy convex data.

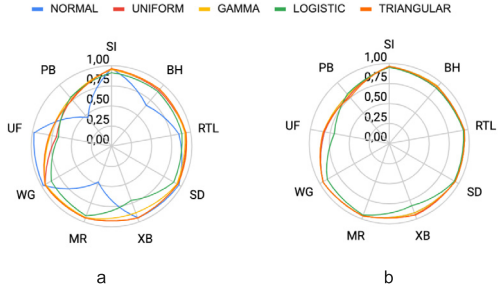


Fig. 10. CIVIs AMI performance of noisy convex datasets broken down by compactness factor. (a) compactness factor= 0.3, (b) compactness factor= 0.1.

dimensions. SI obtain the best performance for the noisy convex distribution, similar to what Arbelaitz et al. (2013) found in 2013. We find that a high compactness factor (0.3) affects PB, BH, and XB for normal-noisy distributions as shown in Fig. 10. Gamma distribution presents the most difficulties in performance, which affect all indices. UF performed well (AMI $\geq$ 0.98) for Normal distribution. However, UF decreases the performance when the datasets have a high degree of compactness factor, as shown in Fig. 10.

The degree of overlap is the most significant factor influencing the performance of CIVIS on noisy-convex distributions. Specifically, UF and PB are highly affected by the level of overlap, as illustrated

[1] All codes are available here https://github.com/David9203/CVI-Frechet-Mahalanobis.

[2] UCI repository: https://archive.ics.uci.edu/ml/index.php.

[3] UCR repository: https://www.cs.ucr.edu/~eamonn/time_series_data/.

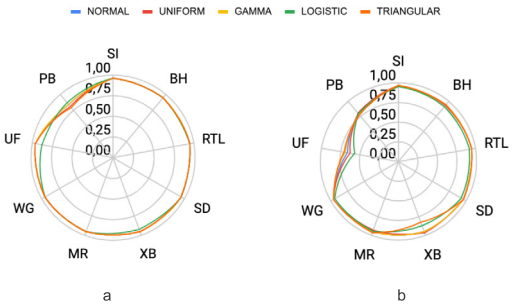


Fig. 11. CIVIs AMI performance of noisy convex datasets broken down by (a) low overlap, (b) high overlap.

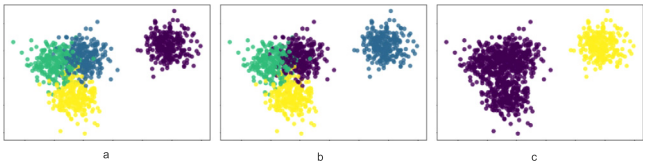


Fig. 12. Example of failure of UF. Dataset with high overlapping and high compaction. (a) real-world dataset (b) SI recommended partition (c) UF recommended partition.

in Fig. 11. This behavior of UF can be attributed to the merging methodology.

UF is prone to merge most closed clusters while not other indices. For example, Fig. 12.a shows a dataset with four clusters, in which SI recommends the same partitions as it is shown in Fig. 12.b, while UF merges the most closes indices recommending Fig. 12.c partition with two clusters. This relative behavior can be an advantage or disadvantage according to the criteria to decide what a cluster is.

Regarding the dimensions, Fig. 13 shows the performance of the CIVIs with different numbers of features.

UF methodology exhibits superior performance ($AMI > 0.88$) when applied to datasets characterized by high dimensions and following Gamma, Triangular, and Uniform distributions. Nevertheless, UF tends to decrease in performance in logistic noisy environments. We find that UF obtained the best performance for high dimensions ($AMI = 0.96$), and even the performance decreases for smaller dimensions (more uncertainty). The PB CIVI performed poorly for all noisy convex distributions with high dimensions. On the other hand, XB under-performed (AMI $\leq$0.80) in the case of noisy-normal datasets.

5.1.2. Non-convex datasets

We show the performance for non-convex data in Fig. 15. Only UF and PB are the CIVIs that discovered non-convex shapes. Most frequently used CIVIs (C, DR, WG, DNN, XB, and DB) showed low performance for both of the distribution shapes used in the text: interleaving half circles (moon) and circles containing a smaller circle (circles) in the two shape compositions used, interleaving half circles (moon) and the circle surrounding a smaller circle (circles). The good results for UF are due to its ability to join small neighboring clusters based on the Fréchet distance proving the potential of uncertainty and merging methodologies used in non-convex data to select the correct number of clusters (see Fig. 14).

5.2. Real-world benchmark datasets

Overall, Table 8 demonstrates that UF shows the potential to recommend the correct number of clusters for nearly all of the real-world benchmark datasets tested. The table provides detailed information about each dataset in the first four columns, followed by the number of clusters recommended by each CIVI in the subsequent columns. The table's last row summarizes the number of hits for each CIVI in the respective datasets. In these datasets, UF accurately predicted the number of clusters in 8 out of 13 datasets. Regarding the AMI performance, Fig. 15 presents the performance behavior that CIVIs have according to each real-world dataset.

Fig. 15 indicates that the performance of clustering algorithms varies significantly across datasets and CIVIs. For example, the IRIs dataset exhibits higher average values (0.71), compared to the Haberman dataset, in which the clustering algorithm did perform poorly despite the CIVIs recommending the correct number of clusters (see Table 8). Additionally, while CIVIs such as Silhouette yield impressive high performance for synthetic datasets, they show low performance for nearly all non-time-series datasets, except for the Wine dataset. Then, we have evidence that not necessarily the number of clusters equal to the ground truth corresponds to the real performance of the dataset. For example, UF recommends the correct number of clusters of ground truth for the Liver dataset but obtains a low performance compared to other CIVIs such as RTL, BH, XB, AND WG. However, UF is the best-performing CIVI in almost all cases of non-time series, except for the Liver dataset. We highlight that UF is also the best AMI for the Digits dataset, recognizing the graphic patterns of this dataset better than other CIVIs.

Regarding the number of clusters in the case of time series datasets, UF, WG, and Silhouette are the ones that behaved better in terms of predicting the number of partitions correctly (see Table 8). However, regarding AMI, the indices obtained a performance lower than 0.70, as shown Fig. 16, demonstrating the significant limitation of these indices suggesting the number of clusters for time series datasets.

5.3. PAM dataset

As PAM seeks to understand the dynamics of ecosystems through sound, it is critical to partition interpretability in which the clusters represent different sound behaviors of anthropophony, geophony, and biophony. Table 9 shows the best-performed CIVIs with their F1 score performance and the number of clusters for the PAM application in each day period. The results show that it is possible to identify the ecological transformation of the tropical forest ecosystem through the CIVI UF.

We found that UF, MR, BH, and SD have a performance greater than 0.77 for each stage. However, UF recommends a low number of clusters (bold and italic values in Table 9) with high performance compared to

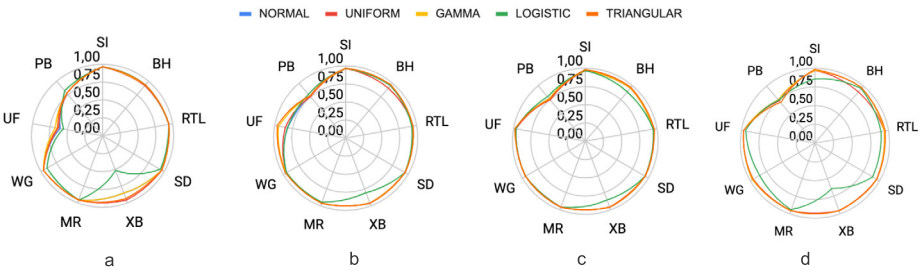


Fig. 13. CIVIs AMI performance of noisy convex datasets broken down by a number of dimensions. (a) <20, (b) 40, (c) 100, (d)200.

Table 8
Results of UF for real-world datasets, each cell of CIVIs columns represents the recommended number of clusters for each cluster. The best results are marked in bold.

Dataset	Nsamples	Features	Classes	SI (Jajuga et al., 2020)	RTL (Ratkowsky and Lance, 1978)	BH (Ball and Hall, 1965)	SD (Dudek, 2020)	XB (Muranishi et al., 2014)	MCR (Zhang et al., 2018)	PB (Hands and Everitt, 2010)	WG (Wemmert et al., 2000)	UF (ours)
WINE	178	13	3 or 2	**3**	2	9	7	16	**2**	9	9	**2**
Iris	150	4	3 or 2	**2**	9	8	**3**	7	8	9	**2**	**3**
Cancer	569	31	2	**2**	19	**2**	5	14	19	18	**2**	**2**
Liver	566	11	3	2	9	2	5	9	9	6	2	**3**
Digits	1797	65	10	8	8	2	9	8	8	2	2	19
Haberman	306	3	2	**2**	**2**	9	3	3	8	8	5	**2**
Ionosphere	351	34	3	7	7	2	8	10	7	2	2	**3**
SwedishLeaf	500	128	15	2	13	5	2	2	5	2	12	19
Wafer	1000	152	2	**2**	14	5	3	**2**	14	**2**	**2**	**2**
ArrowHead	36	251	3	2	14	13	26	8	14	14	2	2
BeetleFly	40	512	2	**2**	14	11	26	8	14	13	**2**	**2**
Car	577	60	4	2	13	13	24	9	11	14	2	3
			n hits	**6**	1	1	1	1	1	1	4	**8**

Table 9
CIVIs performance for PAM data considering the recommended cluster numbers and F1 performance score that each CIVI had in each day period. Best result considering both metrics are marked in bold.

Index	*N* clusters (5-8)	*N* clusters (8-17)	N_CLUSTERS (17-5)	F1 score (5-8)	F1 score (8-17)	F1 score (17-5)
CH (Caliński and Harabasz, 1974)	2	6	4	0.461	0.604	0.487
BH (Ball and Hall, 1965)	67	**84**	**59**	0.773	**0.806**	**0.845**
C (Hubert and Schultz, 1976)	97	***97***	***97***	0.790	**0.804**	**0.847**
DB (Dudek, 2020)	**100**	6	**99**	**0.802**	0.602	**0.858**
SI (Jajuga et al., 2020)	81	6	6	0.785	0.602	0.561
PB (Hands and Everitt, 2010)	99	2	**100**	0.797	0.334	**0.852**
SD (Dudek, 2020)	37	65	60	0.740	0.797	0.833
XB (Muranishi et al., 2014)	65	**84**	60	0.771	**0.806**	0.833
MCR (Zhang et al., 2018)	99	***97***	***97***	0.797	**0.806**	**0.843**
UF (ours)	***44***	***42***	***45***	**0.771**	**0.806**	**0.843**

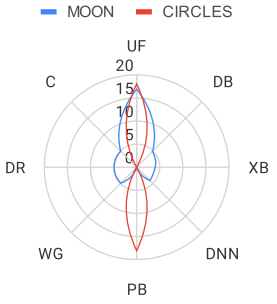


Fig. 14. CIVIs number of clusters performance for nonconvex data, axis represent the number of hits, of best performed CIVIs.

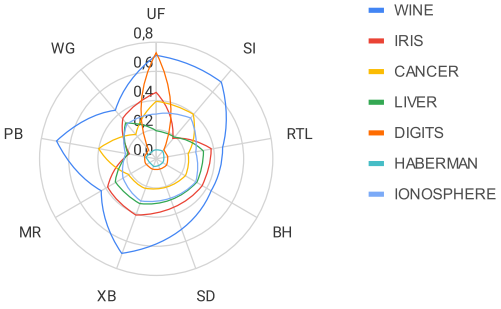


Fig. 15. AMI CIVIs performance for real-world non-time-series data: Wine, Iris, Cancer, Liver, Digits, Haberman, Ionosphere, and time-series SwedishLeaf, Wafer, ArrowHead, BeetleFly, Car datasets.

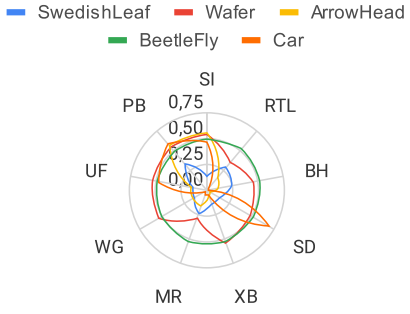


Fig. 16. AMI CIVIs performance for time series data: SwedishLeaf, Wafer, ArrowHead, BeetleFly Car.

the other indices, which ensures better interpretability with the same score. On the other hand, other indices as XB, DB, BR, BH, MR, and G show monotonicity. UF obtain the best results considering both: the low number of clusters and the performance. The 17–5 period acquires the highest F1 score performance using almost all CIVIs (BH, C, DB, PB, SD, XB, UF, MR). This behavior could be related to the variability and distribution of sounds that occur at night and their intra-similarity.

In PAM there are different distributions of data which are not necessarily normal since each recording varies according to the sounds of diverse sources. Results show that UF is capable of grasping these acoustic behaviors. We demonstrate that using UF and GMM distributions, making it possible to find intermediate value transformations and correlate locations with comparable transformations for the distinct hours that offer diverse patterns of sound: morning (5–8), day (8–17), and night (17–0). In a completely unsupervised manner, our methodology assesses whether the ecological transformation calculate automatically at a place is equivalent to the transformation labels determined in advance by field staff and remote sensing.

5.4. Discussion

UF demonstrates the capability to handle cases of noisy convex clusters for normal, uniform, triangular distributions, and nonconvex distributions as well as real-world applications. In the PAM application, there may be non-convex forms of the soundscape, not necessarily of biological species in which UF is liable to give an idea of these clusters. The performance of CIVIs such as Silhouette is satisfactory when dealing with convex synthetic datasets. These CIVIs tend to display unexpected clusters when faced with a lack of samples to define a distribution in real-world datasets or when dealing with data that has non-convex distributions. It is clear that there is no CIVI that is superior to others, as each has its advantages and limitations depending on the specific application. Certain indices such as SI, MR, BH, SD, and WG are suitable for handling datasets with convex assumptions. Therefore, selecting the appropriate CIVI for a specific application should be based on understanding the underlying data assumptions. The silhouette CIVI may be a good option for some applications, while UF may be more suitable for others, as shown in Fig. 12.

UF relies on the Frechet distance and the PFOU computation. The computational complexity of the Frechet distance, $d_F(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, repeated nc times is $\mathcal{O}(D^2nc^2)$ since we compute it between all clusters and the number of dimensions. On the other hand, the computational complexity of the PFOU relies on the Mahalanobis distance, which depends on the number of dimensions, D, with the complexity of $\mathcal{O}(D^2n)$. Then, the overall complexity of UF is $\mathcal{O}(D^2nc^2)$. However, if the used clustering algorithms provide the covariance matrices the computational cost is reduced to $\mathcal{O}(nc^2)$.

Although UF has shown to be an efficient methodology in most cases, our proposal has a limitation of reduced performance in the case of datasets with noisy distributions with a high overlap and logistic distributions. Probably, this is because of the method's normal distribution assumption and the merging methodology. Therefore, to improve the performance in datasets with noisy distributions and high overlap in logistic distributions, it would be beneficial to consider statistical assumptions for such data types in future work. The relativity of what a cluster is depends on the application and could be subjective, as well as the choice of the CIVI to use. Nevertheless, we emphasize the capability of UF to identify representative clusters even in the absence of distributional information. Regarding the PAM application, we show

the high potential to identify the health of transformed ecosystems through CIVIs. This approach helps with landscape monitoring, identifying sites with unique characteristics, and creating action plans to halt the ecological decline passively.

In PAM and real benchmarks (Iris, Wine, Breast cancer, Digits, Haberman, Ionosphere, Liver, Swedish Leaf, Wafer, Arrowhead, Beet Fly, and Car datasets), UF is the one that behaved best, concerning performance and the number of clusters obtained when the merging is applied. Convex noisy-overlapped data could affect the performance of UF depending on what would be considered a cluster or not, as we showed in the example of Fig. 12. The importance of our application lies in the component that can aid in our comprehension of natural settings and which can support ecological research initiatives. Besides, UF might influence pertinent domains where clustering validity is essential, like in computer vision of unsupervised domain adaptation for person re-identification (Rami et al., 2022). The use of unsupervised algorithms and CIVIs offers an advantage over supervised algorithms since they allow the discovery of new groups that were not previously defined, providing valuable information to the analyst. For instance, in this study, environmental conservation experts labeled each location with only three transformation states based on satellite images, which did not fully characterize the sites due to their dynamic nature.

6. Conclusion

In this study, we proposed a new CIVI, the Uncertainty Fréchet (UF), that can quantify uncertainty and assess the quality of cluster partitions in unsupervised learning scenarios. UF is a versatile metric that can accommodate both crisp and soft clustering approaches, making it a suitable option for Passive Acoustic Monitoring (PAM) of ecosystems. While other CIVIs such as SI, MR, BH, SD, and WG handled applications with assumptions of convex datasets, UF has shown competitive performance in these scenarios as well. More importantly, UF has demonstrated superior performance in non-convex and real-world datasets, showcasing the potential of uncertainty-based metrics that consider the underlying geometry of the dataset space. The results encourage further work on these types of CIVIs. For example, an interesting future direction would be improving UF for compressing time series applications and datasets with overlapping noisy data. Choosing the most appropriate CIVI for an application requires careful consideration of the data distribution's characteristics and assumptions. Continuing to refine and develop CIVIs can lead to a better understanding of their strengths and limitations, resulting in improved accuracy and reliability of clustering results. We believe that UF and other uncertainty-based metrics may impact on unsupervised learning. Nonetheless, more research and testing on specific applications are needed to validate the effectiveness of these methods.

CRediT authorship contribution statement

Nestor Rendon: Conceptualization, Methodology, Software, Writing – original draft. **Jhony H. Giraldo:** Conceptualization, Methodology, Supervision, Writing – review & editing. **Thierry Bouwmans:** Conceptualization, Supervision, Writing – review & editing, Funding acquisition. **Susana Rodríguez-Buritica:** Data curation, Resources, Review. **Edison Ramirez:** Data curation. **Claudia Isaza:** Conceptualization, Supervision, Methodology, Writing – review & editing, Funding acquisition, Resources, Project administration.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Passive acoustic monitoring data will be made available on request.

Acknowledgments

This work was supported by Universidad de Antioquia, Instituto Tecnologico Metropolitano de Medellín, Alexander von Humboldt Institute for Research on Biological Resources, Colombian National Fund for Science, Technology, and Innovation, Francisco Jose de Caldas, MINCIENCIAS (Colombia): Program No. 111585269779, and ECOS-Nord program: MESRI (France) C22M01 and MINCIENCIAS (Colombia) [Cod. 83477, 807404942021. Cto.494-2021]

Appendix A. Supplementary data

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.engappai.2023.106635.

References

Agrawal, K., Garg, S., Patel, P., 2015a. Performance measures for densed and arbitrary shaped clusters. Int. J. Comput. Sci. Commun. 6, 338–350. http://dx.doi.org/10.090592/IJCSC.2015.637.

Anand, S.K., Kumar, S., 2022. Experimental comparisons of clustering approaches for data representation. ACM Comput. Surv. 55 (3), 1–33. http://dx.doi.org/10.1145/3490384.

Anon, 2020. Efficient synthetical clustering validity indexes for hierarchical clustering. Expert Syst. Appl. 151, 113367. https://doi.org/10.1016/j.eswa.2020.113367. URL: https://www.sciencedirect.com/science/article/pii/S0957417420301925.

Arbelaitz, O., Gurrutxaga, I., Muguerza, J., Pérez, J.M., Perona, I., 2013. An extensive comparative study of cluster validity indices. Pattern Recognit. 46 (1), 243–256. http://dx.doi.org/10.1016/j.patcog.2012.07.021.

Ball, G., Hall, D., 1965. ISODATA: A Novel Method of Data Analysis and Pattern Classification. Technical Report, Stanford Research Institute, Menlo Park.

Banfield, J.D., Raftery, A.E., 1993. Model-based Gaussian and non-Gaussian clustering. Biometrics 49, 803–821.

Bezdek, J., Ehrlich, R., Full, W., 1984. FCM—the fuzzy C-means clustering-algorithm. Comput. Geosci. 10, 191–203. http://dx.doi.org/10.1016/0098-3004(84)90020-7.

Boelman, N.T., Asner, G.P., Hart, P.J., Martin, R.E., 2007. Multi-trophic invasion resistance in Hawaii: Bioacoustics, field surveys, and airborne remote sensing. Ecol. Appl. 17, 2137–2144. http://dx.doi.org/10.1890/07-0004.1.

Bolshakova, N., Azuaje, F., 2003. Cluster validation techniques for genome expression data. Signal Process. 83, 825–833. http://dx.doi.org/10.1016/S0165-1684(02)00475-9.

Borlea, I.D., Precup, R.E., Borlea, A., 2022. Improvement of K-means cluster quality by post processing resulted clusters. Procedia Comput. Sci. 199, 63–70. http://dx.doi.org/10.1016/j.procs.2022.01.009.

Caliński, T., Harabasz, J., 1974. A dendrite method for cluster analysis. Commun. Stat. 3 (1), 1–27. http://dx.doi.org/10.1080/03610927408827101.

Campo, D., Stegmayer, G., Milone, D., 2016. A new index for clustering validation with overlapped clusters. Expert Syst. Appl. 64, 549–556. http://dx.doi.org/10.1016/j.eswa.2016.08.021.

Chen, Y., Keogh, E., Hu, B., Begum, N., Bagnall, A., Mueen, A., Batista, G., 2015. The UCR time series classification archive. www.cs.ucr.edu/~eamonn/time_series_data/.

Cheng, D., Zhu, Q.s., Jinlong, H., Wu, Q., Yang, L., 2018. A novel cluster validity index based on local cores. IEEE Trans. Neural Netw. Learn. Syst. 30(4): 985-999, 1–15. http://dx.doi.org/10.1109/tnnls.2018.2853710.

Coensel, B.D., 2007. Introducing the temporal aspect in environmental soundscape research. Imec Publ. 291.

Cureton, E., Cureton, L., Durfee, R., 2010. A method of cluster analysis. Multivar. Behav. Res. 5, 101–116. http://dx.doi.org/10.1207/s15327906mbr0501_7.

Davies, D.L., Bouldin, D.W., 1979. A cluster separation measure. IEEE Trans. Pattern Anal. Mach. Intell. PAMI-1 (2), 224–227. http://dx.doi.org/10.1109/TPAMI.1979.4766909.

Depraetere, M., Pavoine, S., Jiguet, F., Gasc, A., Duvail, S., Sueur, J., 2012. Ecol. Indic. http://dx.doi.org/10.1016/j.ecolind.2011.05.006.

Dua, D., Graff, C., 2017. UCI Machine Learning Repository. University of California, Irvine, School of Information and Computer Sciences, URL: http://archive.ics.uci.edu/ml.

Dudek, A., 2020. Silhouette index as clustering evaluation tool. In: Jajuga, K., Batóg, J., Walesiak, M. (Eds.), Classification and Data Analysis. Springer International Publishing, Cham, pp. 19–33.

Dunn, J.C., 1974. Well-separated clusters and optimal fuzzy partitions. J. Cybern. 4 (1), 95–104. http://dx.doi.org/10.1080/01969727408546059.

Ellis, D., Nieto, O., McFee, B., Liang, D., McVicar, M., Raffel, C., Battenberg, E., 2015. librosa: Audio and music signal analysis in python. In: Proceedings of the 14th Python in Science Conference. pp. 18–24. http://dx.doi.org/10.25080/majora-7b98e3ed-003.

Farina, A., Pieretti, N., Salutari, P., Tognari, E., Lombardi, A., 2016. The application of the acoustic complexity indices (ACI) to ecoacoustic event detection and identification (EEDI) modeling. Biosemiotics 9, 227–246. http://dx.doi.org/10.1007/s12304-016-9266-3.

Franco, C., Vidal, L., Cruz, A., 2002. A validity measure for hard and fuzzy clustering derived from Fisher's linear discriminant. 2, pp. 1493–1498. http://dx.doi.org/10.1109/FUZZ.2002.1006727,

Friedman, H., Rubin, J., 1967. On some invariant criteria for grouping data. J. Amer. Statist. Assoc. 62, 1159–1178. http://dx.doi.org/10.1080/01621459.1967.10500923.

Fu, L., Wu, S., 2017. An internal clustering validation index for Boolean data. Cybern. Inf. Technol. 16 (6), 232–244. http://dx.doi.org/10.1515/cait-2016-0091.

Guerrero, M.J., Bedoya, C.L., López, J.D., Daza, J.M., Isaza, C., 2023. Acoustic animal identification using unsupervised learning. Methods Ecol. Evol. http://dx.doi.org/10.1111/2041-210X.14103.

Guo, X., Gao, L., Liu, X., Yin, J., 2017. Improved deep embedded clustering with local structure preservation. http://dx.doi.org/10.24963/ijcai.2017/243.

Gurrutxaga, I., Muguerza, J., Arbelaitz, O., Pérez, J.M., Martín, J.I., 2011. Towards a standard methodology to evaluate internal cluster validity indices. Pattern Recognit. Lett. 32 (3), 505–515. http://dx.doi.org/10.1016/j.patrec.2010.11.006.

Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2002a. Cluster validity methods: Part I. SIGMOD Rec. 31, http://dx.doi.org/10.1145/565117.565124.

Halkidi, M., Batistakis, Y., Vazirgiannis, M., 2002b. Clustering validity checking methods: Part II. SIGMOD Rec. 31, 19–27. http://dx.doi.org/10.1145/601858.601862.

Han, X., Quan, L., Xiong, X., Almeter, M., Xiang, J., Lan, Y., 2017. A novel data clustering algorithm based on modified gravitational search algorithm. Eng. Appl. Artif. Intell. 61, 1–7. http://dx.doi.org/10.1016/j.engappai.2016.11.003.

Handl, J., Knowles, J., Kell, D., 2005. Bioinformatics computational cluster validation in post-genomic data analysis. Bioinformatics (Oxford, England) 21, 3201–3212. http://dx.doi.org/10.1093/bioinformatics/bti517.

Hands, S., Everitt, B., 2010. A Monte Carlo study of the recovery of cluster structure in binary data by hierarchical clustering techniques. Multivar. Behav. Res. 22, 235–243. http://dx.doi.org/10.1207/s15327906mbr2202_6.

Hubert, L., Schultz, J., 1976. Quadratic assignment as a general data analysis strategy. Br. J. Math. Stat. Psychol. 29 (2), 190–241. http://dx.doi.org/10.1111/j.2044-8317.1976.tb00714.x.

Iglesias, F., Zseby, T., Ferreira, D., 2019. MDCGen: Multidimensional dataset generator for clustering. J. Classification 36, 599–618, https://link.springer.com/article/10.1007/s00357-019-9312-3.

Iglesias, F., Zseby, T., Zimek, A., 2020. Absolute cluster validity. IEEE Trans. Pattern Anal. Mach. Intell. 42, 2096–2112. http://dx.doi.org/10.1109/TPAMI.2019.2912970.

Iglesias Vázquez, F., Zseby, T., Zimek, A., 2021. Clustering refinement. Int. J. Data Sci. Anal. 12, 1–21. http://dx.doi.org/10.1007/s41060-021-00275-z.

Isaza, C., 2007. Diagnostic par techniques d'apprentissage floues: concept d'une méthode de validation et d'optimisation des partitions Doctoral dissertation. INSA de Toulouse, 2007.

Jain, A.K., Dubes, R.C., 1988. Algorithms for Clustering Data. Prentice-Hall, Inc., USA.

Jajuga, K., Batóg, J., Walesiak, M. (Eds.), 2020. Silhouette index as clustering evaluation tool. In: Classification and Data Analysis. Springer International Publishing, Cham, pp. 19–33.

Jaskowiak, P.A., Costa, I.G., 2023. Clustering validation with the area under precision-recall curves. arXiv:2304.01450.

Kim, M., Ramakrishna, R.S., 2005. New indices for cluster validity assessment. Pattern Recognit. Lett. 26, 2353–2363. http://dx.doi.org/10.1016/j.patrec.2005.04.007.

Lee, S.H., Jeong, Y.S., Kim, J.Y., Jeong, M.K., 2018. A new clustering validity index for arbitrary shape of clusters. Pattern Recognit. Lett. 112, 263–269. http://dx.doi.org/10.1016/j.patrec.2018.08.005.

Liang, S., Han, D., Yang, Y., 2020. Cluster validity index for irregular clustering results. Appl. Soft Comput. 95, 106583. http://dx.doi.org/10.1016/j.asoc.2020.106583.

Liu, Y., Jiang, Y., Hou, T., Liu, F., 2021. A new robust fuzzy clustering validity index for imbalanced data sets. Inform. Sci. 547, 579–591. http://dx.doi.org/10.1016/j.ins.2020.08.041, URL: https://www.sciencedirect.com/science/article/pii/S0020025520308094.

Liu, Y., Li, Z., Xiong, H., Gao, X., Wu, J., Wu, S., 2012. Understanding and enhancement of internal clustering validation measures. IEEE Trans. Syst. Man Cybern. B 43, http://dx.doi.org/10.1109/TSMCB.2012.2220543.

Liu, Y., Zhang, X., Chen, J., Chao, H., 2019a. A validity index for fuzzy clustering based on bipartite modularity. J. Electr. Comput. Eng. 2019, 2719617. http://dx.doi.org/10.1155/2019/2719617.

Liu, Y., Zhang, X., Chen, J., Chao, H., 2019c. A validity index for fuzzy clustering based on bipartite modularity. J. Electr. Comput. Eng. 2019, http://dx.doi.org/10.1155/2019/2719617.

Ma, Q., Chen, C., Li, S., Cottrell, G., 2021. Learning representations for incomplete time series clustering.

Muranishi, M., Honda, K., Notsu, A., 2014. Xie-Beni-Type fuzzy cluster validation in fuzzy co-clustering of documents and keywords. In: Cho, Y.I., Matson, E.T. (Eds.), Soft Computing in Artificial Intelligence. Springer International Publishing, Cham, pp. 29–38.

Ouchicha, C., Ammor, O., Meknassi, M., 2018. Cluster validity index: Comparative study and a new validity index with high performance. ACM Int. Conf. Proc. Ser. 1–6. http://dx.doi.org/10.1145/3230905.3230917.

Ouchicha, C., Ammor, O., Meknassi, M., 2020. A new validity index in overlapping clusters for medical images. Autom. Control Comput. Sci. 54, 238–248. http://dx.doi.org/10.3103/S0146411620030050.

Ozkan, I., Türkşen, I.B., 2012. MiniMax ϵ -stable cluster validity index for type-2 fuzziness. Inform. Sci. 184 (1), 64–74. http://dx.doi.org/10.1016/j.ins.2011.07.036, URL: https://www.sciencedirect.com/science/article/pii/S0020025511003689.

Panaretos, V., Zemel, Y., 2019. Statistical aspects of Wasserstein distances. Annu. Rev. Stat. Appl. 6, 1–20. http://dx.doi.org/10.1146/annurev-statistics-030718-104938.

Parsa, M.G., Zare, H., Ghatee, M., 2020. Unsupervised feature selection based on adaptive similarity learning and subspace clustering. Eng. Appl. Artif. Intell. 95, 103855. http://dx.doi.org/10.1016/j.engappai.2020.103855.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., Duchesnay, E., 2011. Scikit-learn: Machine learning in python. J. Mach. Learn. Res. 12, 2825–2830.

Puruncajas, B., Vidal, Y., Tutivén, C., 2020. Damage detection and diagnosis for offshore wind foundations. In: ICINCO 2020 - Proceedings of the 17th International Conference on Informatics in Control, Automation and Robotics. pp. 181–187, URL: http://hdl.handle.net/2117/329860.

Rami, H., Ospici, M., Lathuilière, S., 2022. Online unsupervised domain adaptation for person re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops. pp. 3830–3839.

Ratkowsky, D., Lance, G., 1978. A criterion for determining the number of groups in a classification. Aust. Comput. J. 3.

Ray, S., Turi, R., 2000. Determination of number of clusters in K-means clustering and application in colour image segmentation. In: Proceedings of the 4th International Conference on Advances in Pattern Recognition and Digital Techniques, Vol. 1. pp. 1–20.

Rendon, N., Rodríguez-Buritica, S., Sanchez-Giraldo, C., Daza, J.M., Isaza, C., 2022a. Automatic acoustic heterogeneity identification in transformed landscapes from Colombian tropical dry forests. Ecol. Indic. 140, 109017. http://dx.doi.org/10.1016/j.ecolind.2022.109017.

Rendon, N., Rodríguez-Buritica, S., Sanchez-Giraldo, C., Daza, J.M., Isaza, C., 2022b. Identification of tropical dry forest transformation in the Colombian caribbean region using acoustic recordings through unsupervised learning. IARIA Annu. Congr. Front. Sci. Technol. Serv. Appl. 32–38.

Reynolds, D.A., Quatieri, T.F., Dunn, R.B., 2000. Speaker verification using adapted Gaussian mixture models. Digit. Signal Process. Rev. J. 10, 19–41. http://dx.doi.org/10.1006/dspr.1999.0361.

Rivera-Borroto, O.M., Rabassa-Gutiérrez, M., del Corazón Grau-Ábalo, R., Marrero-Ponce, Y., de la Vega, J.M.G., 2012. Dunn's index for cluster tendency assessment of pharmacological data sets. Can. J. Physiol. Pharmacol. 90, 425–433. http://dx.doi.org/10.1139/Y2012-002.

Rizman Žalik, K., 2010. Cluster validity index for estimation of fuzzy clusters of different sizes and densities. Pattern Recognit. 43 (10), 3374–3390. http://dx.doi.org/10.1016/j.patcog.2010.04.025, URL: https://www.sciencedirect.com/science/article/pii/S0031320310002013.

Romano, S., Vinh, N.X., Bailey, J., Verspoor, K., 2016. Adjusting for chance clustering comparison measures. J. Mach. Learn. Res. 17 (134), 1–32, URL: http://jmlr.org/papers/v17/15-627.html.

Sadeghi, V., Etemadfard, H., 2022. Optimal cluster number determination of FCM for unsupervised change detection in remote sensing images. Earth Sci. Inform. 15 (2), 1045–1057. http://dx.doi.org/10.1007/s12145-021-00757-5.

Sakai, T., Imiya, A., 2009. Unsupervised cluster discovery using statistics in scale space. Eng. Appl. Artif. Intell. 22 (1), 92–100. http://dx.doi.org/10.1016/j.engappai.2008.04.011.

Shi, C., Wei, B., Wei, S., Wang, W., Liu, H., Liu, J., 2021. A quantitative discriminant method of elbow point for the optimal number of clusters in clustering algorithm. EURASIP J. Wireless Commun. Networking 2021 (1), 31. http://dx.doi.org/10.1186/s13638-021-01910-w.

Silva, L., Moura, R., Canuto, A.M.P., Santiago, R.H.N., Bedregal, B., 2015. An interval-based framework for fuzzy clustering applications. IEEE Trans. Fuzzy Syst. 23 (6), 2174–2187. http://dx.doi.org/10.1109/TFUZZ.2015.2407901.

Sirmen, R.T., Üstöndag, B.B., 2022. Internal validity index for fuzzy clustering based on relative uncertainty. Comput. Mater. Contin. 72, 2909–2926. http://dx.doi.org/10.32604/cmc.2022.023947.

Sokolova, M., Japkowicz, N., Szpakowicz, S., 2006. Beyond accuracy, F-score and ROC: A family of discriminant measures for performance evaluation. In: AI 2006: Advances in Artificial Intelligence, Lecture Notes in Computer Science, Vol. 4304. pp. 1015–1021. http://dx.doi.org/10.1007/11941439_114.

Towsey, M., 2013. Noise removal from waveforms and spectrograms derived from natural recordings of the environment. pp. 1–4, URL: http://eprints.qut.edu.au/61399/.

Towsey, M., Wimmer, J., Williamson, I., Roe, P., 2014a. The use of acoustic indices to determine avian species richness in audio-recordings of the environment. Ecol. Inform. 21, 110–119. http://dx.doi.org/10.1016/j.ecoinf.2013.11.007.

Towsey, M., Zhang, L., Cottman-Fields, M., Wimmer, J., Zhang, J., Roe, P., 2014b. Visualization of long-duration acoustic recordings of the environment. Procedia Comput. Sci. 29, 703–712. http://dx.doi.org/10.1016/j.procs.2014.05.063.

Vinh, N.X., Epps, J., Bailey, J., 2010. Information theoretic measures for clusterings comparison: Variants, properties, normalization and correction for chance. J. Mach. Learn. Res. 11 (95), 2837–2854.

Wang, G., Wang, J.S., Wang, H.Y., Liu, J.X., 2023. Component-wise design method of fuzzy C-means clustering validity function based on CRITIC combination weighting. J. Supercomput. http://dx.doi.org/10.1007/s11227-023-05234-y.

Wang, H.Y., Wang, J.S., Zhu, L.F., 2021. A new validity function of FCM clustering algorithm based on intra-class compactness and inter-class separation. J. Intell. Fuzzy Syst. 40, 12411–12432. http://dx.doi.org/10.3233/JIFS-210555, 6.

Wang, W., Zhang, Y., 2007. On fuzzy cluster validity indices. Fuzzy Sets and Systems 158, 2095–2117. http://dx.doi.org/10.1016/j.fss.2007.03.004.

Wemmert, C., Gancarski, P., Korczak, J., 2000. A collaborative approach to combine multiple learning methods. Int. J. Artif. Intell. Tools 9, 59–78. http://dx.doi.org/10.1142/S0218213000000069.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., Davison, J., Shleifer, S., von Platen, P., Ma, C., Jernite, Y., Plu, J., Xu, C., Le Scao, T., Gugger, S., Drame, M., Lhoest, Q., Rush, A., 2020. Transformers: State-of-the-art natural language processing. In: Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Association for Computational Linguistics, Online, pp. 38–45. http://dx.doi.org/10.18653/v1/2020.emnlp-demos.6, URL: https://aclanthology.org/2020.emnlp-demos.6.

Xiao, Z., Xu, X., Xing, H., Luo, S., Dai, P., Zhan, D., 2021a. RTFN: A robust temporal feature network for time series classification. Inform. Sci. 571, 65–86. http://dx.doi.org/10.1016/j.ins.2021.04.053, URL: https://www.sciencedirect.com/science/article/pii/S0020025521003820.

Xiao, Z., Xu, X., Xing, H., Song, F., Wang, X., Zhao, B., 2021b. A federated learning system with enhanced feature extraction for human activity recognition. IEEE Transactions on Instrumentation and Measurement, 71, 1-12, 2022 229, 107338. http://dx.doi.org/10.1016/j.knosys.2021.107338, URL: https://www.sciencedirect.com/science/article/pii/S0950705121006006.

Xie, J., Hu, K., Zhu, M., Guo, Y., 2020. Data-driven analysis of global research trends in bioacoustics and ecoacoustics from 1991 to 2018. Ecol. Inform. 57, 101068. http://dx.doi.org/10.1016/j.ecoinf.2020.101068.

Xie, J., Xiong, Z.Y., Dai, Q.Z., Wang, X.X., Zhang, Y.F., 2019. A new internal index based on density core for clustering validation. Inform. Sci. 506, http://dx.doi.org/10.1016/j.ins.2019.08.029.

Xie, Y., Zhang, J., Xia, Y., van den Hengel, A., Wu, Q., 2023a. ClusTR: Exploring efficient self-attention via clustering for vision transformers. URL: https://openreview.net/forum?id=CvfiXFOW2n.

Xing, H., Xiao, Z., Qu, R., Zhu, Z., Zhao, B., 2022. An efficient federated distillation learning system for multi-task time series classification. CoRR abs/2201.00011.

Xu, Q., Zhang, Q., Liu, J., Luo, B., 2020. Efficient synthetical clustering validity indexes for hierarchical clustering. Expert Syst. Appl. 151, http://dx.doi.org/10.1016/j.eswa.2020.113367.

Yapıcı Pehlivan, N., Turksen, I., 2021. A novel multiplicative fuzzy regression function with a multiplicative fuzzy clustering algorithm. Romanian J. Inf. Sci. Technol. 24, 79–98.

Zeng, J., Xie, L., Liu, Z.-Q., 2008. Type-2 fuzzy Gaussian mixture models. Pattern Recognit. 41 (12), 3636–3643. http://dx.doi.org/10.1016/j.patcog.2008.06.006, URL: https://www.sciencedirect.com/science/article/pii/S0031320308002380.

Zhang, J., Nguyen, T., Cogill, S., Bhatti, A., Lingkun, L., Yang, S., Nahavandi, S., 2018. A review on cluster estimation methods and their application to neural spike data. J. Neural Eng. 15, http://dx.doi.org/10.1088/1741-2552/aab385.

Zhang, J., Xu, Y., Chen, H., Xing, L., 2023. A novel building heat pump system semi-supervised fault detection and diagnosis method under small and imbalanced data. Eng. Appl. Artif. Intell. 123, 106316. http://dx.doi.org/10.1016/j.engappai.2023.106316, URL: https://www.sciencedirect.com/science/article/pii/S0952197623005006.

Zhou, S., Xu, Z., Liu, F., 2017. Method for determining the optimal number of clusters based on agglomerative hierarchical clustering. IEEE Trans. Neural Netw. Learn. Syst. 28 (12), 3007–3017. http://dx.doi.org/10.1109/TNNLS.2016.2608001.