

# POLITECNICO DI TORINO

Corso di Laurea Magistrale

in Ingegneria Elettrica

Tesi di Laurea Magistrale

**ANALISI COMPARATIVA  
TRA METODI DI CLASSIFICAZIONE  
APPLICATI A PATTERN DI CARICO  
NON RESIDENZIALI**



Relatori:  
Prof. Gianfranco Chicco  
Ing. Andrea Mazza

Candidato:  
Matteo Gentileschi  
Matricola: S256653

Anno Accademico 2020/2021

# Indice

<b>1</b>	<b>Introduzione</b>	<b>5</b>
1.1	Carichi aggregati . . . . .	7
1.2	Relazione tra valori medi e varianze . . . . .	8
1.3	Centri di attrazione . . . . .	10
1.4	Distanze tra cluster . . . . .	12
1.5	Riduzione di Sammon . . . . .	13
<b>2</b>	<b>Similitudine</b>	<b>15</b>
2.1	Metriche . . . . .	15
2.2	Norma di un vettore . . . . .	16
2.3	Metriche note in letteratura . . . . .	16
2.4	Metriche in Matlab . . . . .	18
<b>3</b>	<b>Entropia</b>	<b>19</b>
3.1	Entropia nel clustering (Renyi Entropy) . . . . .	20
3.2	Divergenza di Kullback–Leibler . . . . .	22
<b>4</b>	<b>Indicatori di validità</b>	<b>23</b>
4.1	Davies-Bouldin Index (DBI) . . . . .	24
4.2	Scatter index (SI) . . . . .	25
4.3	Errore di Sammon . . . . .	25
<b>5</b>	<b>Interpretazione dei risultati</b>	<b>26</b>
<b>6</b>	<b>Indicatore OPISD</b>	<b>28</b>
6.1	Concetto di dominanza stocastica . . . . .	28
6.2	Definizione dell’indicatore OPISD . . . . .	30
<b>7</b>	<b>Esempio</b>	<b>31</b>
<b>8</b>	<b>Analisi nel dominio della frequenza</b>	<b>33</b>
8.1	Trasformata di Fourier . . . . .	33
8.2	Harmonic order ranking . . . . .	34
8.3	Clustering . . . . .	36
<b>9</b>	<b>Between-Cluster Entropy-Based Clustering (BCEBC)</b>	<b>37</b>
<b>10</b>	<b>Follow The Leader</b>	<b>39</b>
10.1	Impostazione del metodo . . . . .	40
10.2	Numero di soluzioni . . . . .	42

10.3 Risultati FTL . . . . .	44
<b>11 Kmeans</b>	<b>62</b>
11.1 Risultati Kmeans . . . . .	63
<b>12 Confronto tra FTL e KMEANS</b>	<b>80</b>
<b>13 Conclusioni</b>	<b>81</b>
<b>14 Bibliografia</b>	<b>82</b>

*Sperare non è necessario per intraprendere.*

# 1 Introduzione

Per definizione, nel contesto degli impianti elettrici, la potenza (attiva) viene così definita:

$$P = \frac{1}{nT} \int_0^{nT} v(t)i(t)dt$$

ove  $T$  è il periodo e  $n$  è un numero intero (positivo): si tratta quindi del valor medio della potenza istantanea nel periodo considerato. La stessa, comunque, può presentare una varianza considerevole durante il funzionamento del carico. Accade quindi che la grandezza in questione viene filtrata: la scelta del periodo di campionamento risulta critica. Inoltre, l'operatore "media" (di qualsiasi natura) rappresenta già un filtro (passa basso). Due profili di carico identici, considerando solo la potenza attiva, potrebbero descrivere carichi differenti. Si può quindi affermare che:

- **Alta frequenza di campionamento:** La serie temporale che ne risulta presenta dimensione elevata impattando sull'onere computazionale senza necessariamente portare a vantaggi. Se la banda dello strumento non è adeguata, si avrà una mole di dati importante ma di bassa qualità.
- **Bassa frequenza di campionamento:** La serie temporale ha dimensioni ridotte e presenta minor onere computazionale, ma il "contenuto informativo" rischia di non essere sufficiente.

I valori misurati vengono organizzati in vettori riga (osservazioni) di dimensione pari al numero di campioni nel periodo considerato (variabili). Ne risulta una matrice (o una lista) di serie temporali, inquadrabili quindi in uno spazio metrico. Dati  $M$  profili temporali di dimensione  $t$ , l'  $m$ -esimo vettore è così strutturato:

$$P^{(m)} = \{p_1^{(m)}, \dots, p_t^{(m)}\}, m = 1, \dots, M$$

ove  $p_j^{(m)} \in R$  rappresenta l'osservazione alla posizione  $j$ -esima (moltiplicando  $j$  per il passo di campionamento si ottiene il tempo trascorso dall'istante iniziale).

Essendo, in questa sede, interessati al pattern (non all'energia assorbita che coinciderebbe con l'area sottesa alla serie temporale), si esegue una normalizzazione per il relativo valore di picco:

$$P_{max}^{(m)} = \max\{p_1^{(m)}, \dots, p_t^{(m)}\} \rightarrow x^{(m)} = \frac{P^{(m)}}{P_{max}^{(m)}}$$

La struttura del database è pertanto la seguente:

		1	2	3	...	...	...	t
x	$x^{(1)}$	$x^{(1)}_{(1)}$	$x^{(1)}_{(2)}$	$x^{(1)}_{(3)}$	...	...	...	$x^{(1)}_{(t)}$
	$x^{(2)}$	$x^{(2)}_{(1)}$	$x^{(2)}_{(2)}$	$x^{(2)}_{(3)}$	...	...	...	$x^{(2)}_{(t)}$
	$x^{(3)}$	$x^{(3)}_{(1)}$	$x^{(3)}_{(2)}$	$x^{(3)}_{(3)}$	...	...	...	$x^{(3)}_{(t)}$
	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...
	...	...	...	...	...	...	...	...
	$x^{(M)}$	$x^{(M)}_{(1)}$	$x^{(M)}_{(2)}$	$x^{(M)}_{(3)}$	...	...	...	$x^{(M)}_{(t)}$

Figura 1: Struttura dataset

A seguire un plot del database (Figura 2):

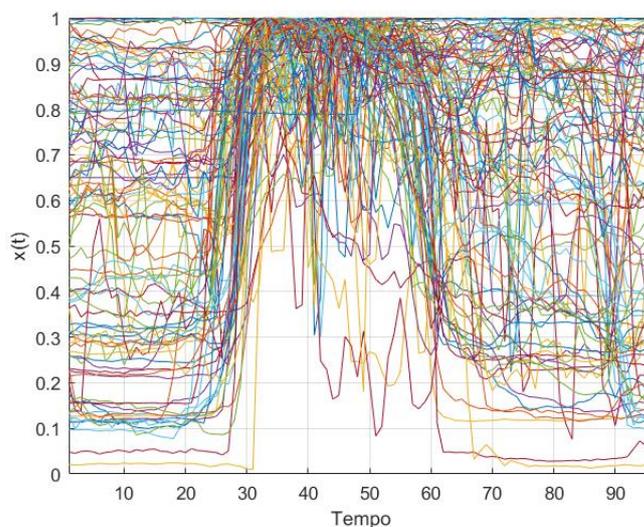


Figura 2: Profili di carico non residenziali quartorari (giornalieri)

I dati così organizzati non sono utilizzabili per fini pratici. Ad esempio, qualora si renda necessario il dimensionamento di una linea di distribuzione, la portata viene dimensionata grazie a metodi quali il Metodo Monte Carlo, che necessitano delle cumulative (C.D.F) di carichi, costruite partendo proprio da dati empirici.

Ci si propone, quindi, l'individuazione di profili di carico (artificiali o reali) che siano **rappresentativi** dell'intero database e che possano pertanto essere utilizzati per future elaborazioni senza eccessiva perdita di informazione (e conoscenza).

## 1.1 Carichi aggregati

L'approccio "classico" consiste in una prima suddivisione in "classi" dei carichi (tipo, livello di tensione ecc) per poi eseguire una vera e propria media (aggregazione).

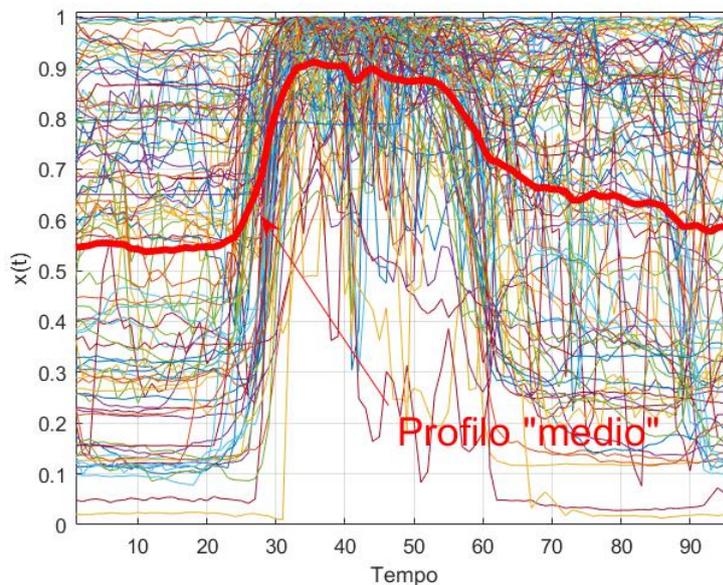


Figura 3: Aggregato

Si manifesta quello che può essere definito il "centro di attrazione" (baricentro) dell'intero database, che presenta il pattern tipico dell'insieme. Trattandosi di carichi non residenziali, era prevedibile (nell'aggregato) osservare un aumento di consumi durante le ore lavorative (dal campione 30 al campione 60, quindi dalle ore 7:50 alle ore 16:00 circa), un lieve abbassamento intorno al campione 45 (le ore 12:00 circa, ora di pranzo) e consumi ridotti nelle ore notturne. Il fatto che non si annulli mai si può interpretare ipotizzando che non si tratti di uffici ma di carichi industriali. Si può affermare che il profilo è coerente con la classe di carico considerata.

## 1.2 Relazione tra valori medi e varianze

Date due variabili aleatorie (aventi stessa distribuzione) X e Y, indipendenti tra loro (scorrelate), si può scrivere:

$$\text{var}[X + Y] = \text{var}[X] + \text{var}[Y]$$

Generalizzando, con  $n$  variabili ( $X_1, \dots, X_n$ ) si ha:

$$\sigma_Y^2 = \sum_{i=1}^n \sigma_{X_i}^2$$

Analogamente, il valor medio della somma delle variabili vale:

$$\mu_Y = \sum_{i=1}^n \mu_{X_i}$$

Interessante è il rapporto tra la media delle varianze e la "media delle medie"

$$\frac{\sqrt{\frac{1}{n} \sum_{i=1}^n \sigma_{X_i}^2}}{\frac{1}{n} \sum_{i=1}^n \mu_{X_i}} = \sqrt{n} \frac{\sigma_Y}{\mu_Y} \rightarrow \frac{\sigma_Y}{\mu_Y} = \frac{1}{\sqrt{n}} \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n \sigma_{X_i}^2}}{\frac{1}{n} \sum_{i=1}^n \mu_{X_i}}$$

Tale rapporto decresce all'aumentare del numero di variabili. Aggregando sempre più carichi, presi arbitrariamente senza considerare un particolare ordinamento, si osserva una diminuzione del rapporto tra varianza e valor medio:

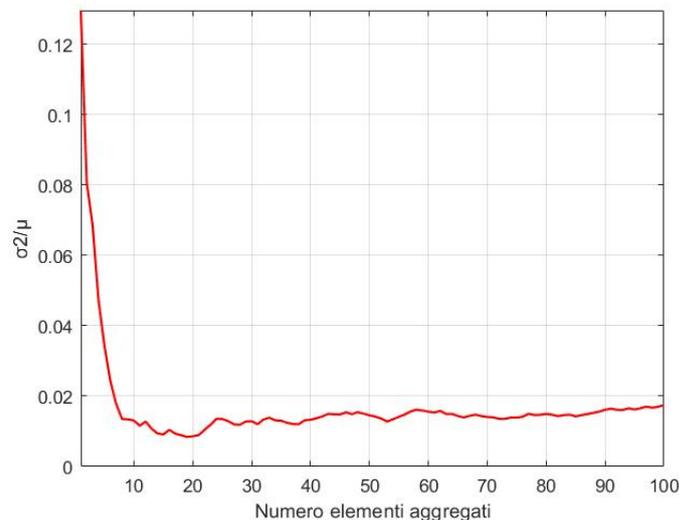


Figura 4: Rapporto tra varianza e valor medio

Il limite di tale approccio sta nel fatto che il profilo risultante non è pienamente rappresentativo del dataset:

- Si tratta di un'entità artificiale (escludendo il caso in cui si considera il "medoide").
- Essendo una media non raggiunge i valori "estremi" (massimo e minimo) che tutti gli altri raggiungono (per definizione, essendo normalizzati, toccano il picco).

Per far fronte a tale problema, si può percorrere la strada della suddivisione in cluster: ogni elemento (profilo di carico, osservazione) viene aggregato ad un insieme di altri elementi "simili" (più avanti verrà analizzato tale concetto). Di ogni cluster si considera il centro di attrazione, eletto a rappresentante. Questo si traduce in una riduzione dell'errore di Sammon (esposto più avanti) rispetto all'aggregazione semplice.

Si può parlare di riduzione degli scenari, ovvero della riduzione di cardinalità dell'insieme di partenza mantenendo invariata l'informazione necessaria. Tra gli aspetti critici, a cui si cercherà di far fronte, vi è la non conoscenza del numero di cluster "ottimale" e dei processi che portano alla relativa determinazione.

La letteratura scientifica offre virtualmente infinite soluzioni, con relative varianti, ognuna applicabile a un determinato problema.

### 1.3 Centri di attrazione

Si rende a questo punto necessario un chiarimento sul concetto di "punto di attrazione" o "baricentro".

Dato un insieme di elementi, il baricentro (in analogia con la fisica classica) è il punto "medio" o "centrale". Il punto medio ha natura "artificiale" ovvero non fa parte dell'insieme di partenza, viceversa il "medoide" è l'elemento (reale) che occupa la posizione "più centrale".

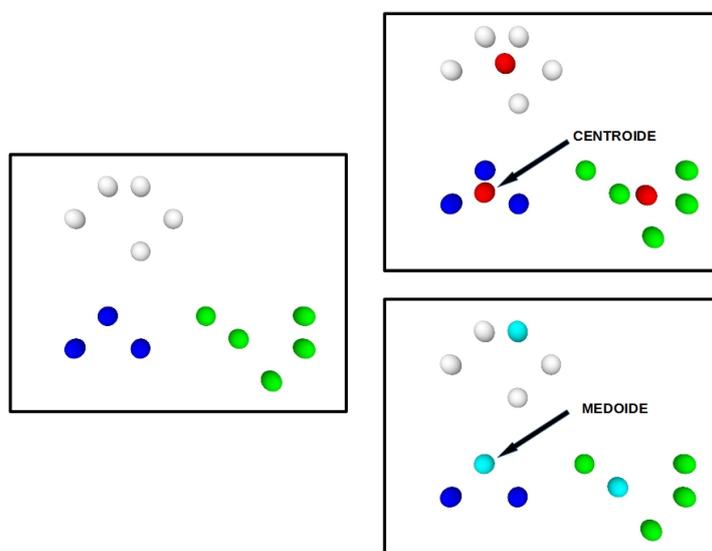


Figura 5: Centroidi e medoidi

Alcuni metodi di clustering, basano il proprio funzionamento sull'individuazione di  $K$  centroidi o medoidi (se si vuole suddividere in  $K$  cluster) per poi associare ad essi gli elementi dai quali sono maggiormente attratti (punto di attrazione).

Variando la natura dei centri di attrazione considerati si ha una variazione del risultato: non è noto a priori, di fatto, quale sia il centro di attrazione che porti ad avere una suddivisione ottimale.

Per questo studio sono stati considerati:

- **centroidi**: si tratta della media degli elementi del cluster;

- **Medoidi:** Elemento del cluster maggiormente vicino al centro. Per ottenerlo si è partiti dal centroide per poi individuare l'elemento (del cluster) che presenta norma euclidea minore rispetto al centroide stesso.
- Media Armonica:** Si tratta della media armonica di tutti gli elementi del cluster. A differenza del centroide (media aritmetica) essa presenta una maggiore variabilità. In questa sede, è stata utilizzata la funzione MatLab "harmmean(X)".

Considerando quindi un gruppo di  $n$  serie temporali, stoccate come righe (vettori riga) di una matrice, il risultato è una serie temporale in cui il generico elemento è la media armonica degli elementi contenuti nella relativa colonna (della matrice).

$$HM = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Tali definizioni, ovviamente, restano valide anche per l'intero dataset (e non solo per i cluster). Il grafico che segue (Figura 6) rappresenta i tre centri di attrazione appena introdotti, calcolati a partire dall'intero database:

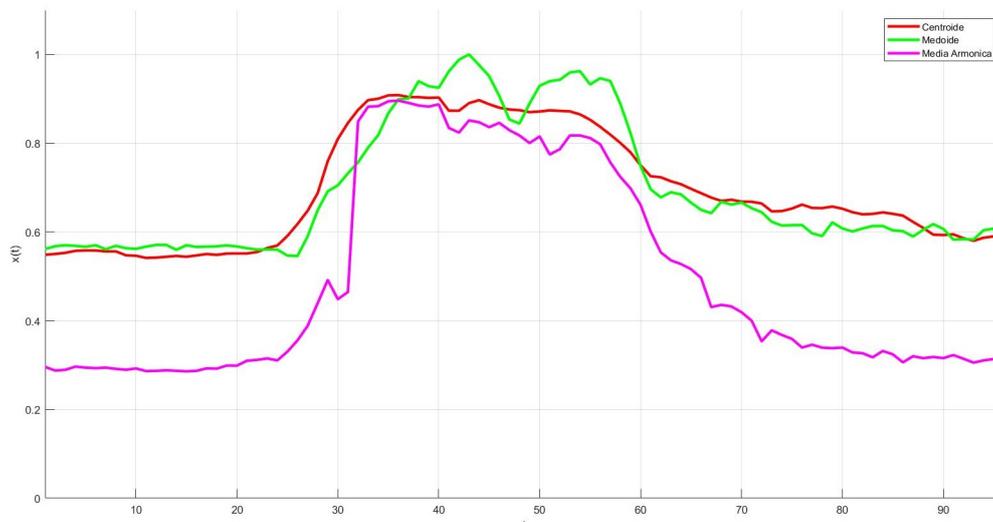


Figura 6: Medoide, Centroide, Media Armonica

Si evince, anche a colpo d'occhio, che si tratta di entità differenti ma non è possibile stabilire a priori quale dei due sia maggiormente adeguato al tipo di studio di interesse.

## 1.4 Distanze tra cluster

La distanza tra due cluster, quindi, può essere calcolata come la distanza tra i centri di attrazione:

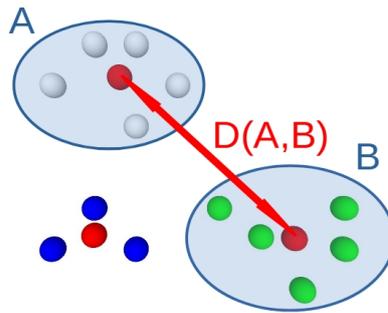


Figura 7: Centroidi e medoidi

L'alternativa è calcolare la "distanza media" ("legame medio) tra gli elementi, ovvero la media delle distanze tra gli elementi di due cluster.

## 1.5 Riduzione di Sammon

Se gli elementi contenuti in un dataset presentano dimensionalità elevata (maggiore di 3), la visualizzazione e l'elaborazione risultano complesse.

La riduzione della cardinalità dell'insieme di partenza (in analogia con la riduzione degli scenari) non è però inerte ai fini del calcolo. Si ha però la possibilità di "quantizzare" l'errore che si commette introducendo l' **errore di Sammon**:

$$E = \frac{1}{\sum_{i<j} d_{ij}^*} \sum_{i<j} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*}$$

ove  $d_{ij}^*$  è la distanza tra il  $j$ -esimo e l' $i$ -esimo elemento nello spazio originale, mentre  $d_{ij}$  è la distanza delle relative proiezioni.

Minimizzando questa funzione si ottiene la miglior riduzione possibile.

Estraendo un solo profilo di carico come "rappresentante" di tutti il database, che sia il medoide o il centroide (media) l'errore di Sammon sarebbe comunque unitario (massimo). L'individuazione del numero e della natura degli elementi rappresentanti ottimale è il punto cardine della presente trattazione.

Si può evincere come l'errore di Sammon presenti un trend discendente all'aumentare del numero di elementi eletti a rappresentare il database complessivo:

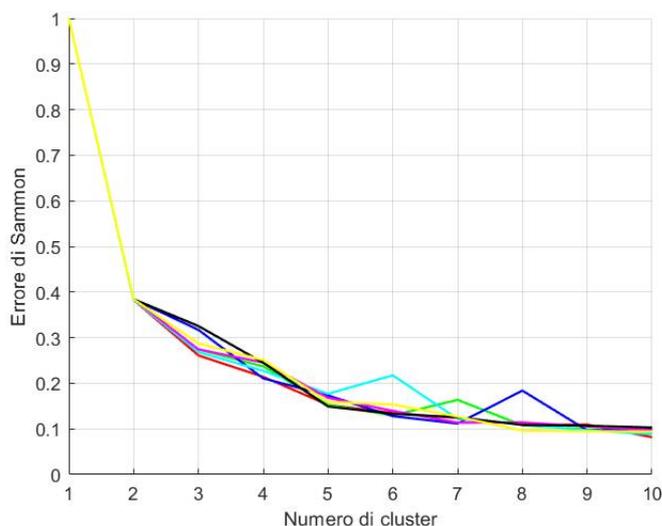


Figura 8: Andamento dell'errore di Sammon

Il grafico sopra riportato mostra l'andamento dell'errore di Sammon al variare del numero di cluster richiesti all'algoritmo kmeans in Matlab (si entrerà nel dettaglio in seguito).

Ogni curva è il risultato di una ripetizione all'interno della quale si fa variare il numero di cluster (i cui centroidi vengono eletti a rappresentanti).

Nonostante non sia una curva monotona, è evidente la tendenza alla riduzione all'aumentare del "numero di rappresentanti".

Dall'ispezione visiva del suddetto grafico, risulta evidente un'abbattimento dell'errore dal passaggio da uno a due rappresentanti.

Si potrebbe pensare di utilizzare tale parametro come funzione obiettivo (da minimizzare) ma il risultato sarebbe fuorviante dal momento che, idealmente, tende a zero quando il numero di rappresentanti è pari al numero di elementi dell'insieme di partenza.

## 2 Similitudine

La classificazione di più elementi passa attraverso il concetto di similitudine. Si rende necessario introdurre dei parametri che consentano di quantificare tale similitudine per consentire una classificazione (elaborazione numerica). Un elemento viene associato ad un cluster quando presenta con esso una similitudine maggiore rispetto a quella che presenta nei confronti degli altri cluster (similitudine tra un elemento e un gruppo).

Non è noto a priori quale sia il "metro" migliore per un determinato database, costituito da elementi con una determinata natura, in un determinato campo di studi.

In questa sede, verranno analizzati differenti "metri" (chiamati "tipo di distanza" per non appesantire l'argomentazione) cercando di individuarne il migliore (tra quelli analizzati) per quanto riguarda la classificazione di curve di carico giornaliera (pattern quartorari) non residenziali. Pertanto, tale documento non è da intendersi valido per altri tipi di dati.

### 2.1 Metriche

In spazi topologici (spazio metrico), si può confondere il concetto di similitudine con il concetto di distanza. Se un elemento è vicino al centro di attrazione di un determinato cluster, allora vi sarà associato.

Volendo portare un esempio, si può pensare al tiro al bersaglio: i punti contenuti all'interno di un settore offrono lo stesso punteggio (quindi sono "simili"), ovvero la distanza dal centro di ognuno di essi è compresa tra due soglie. Altro esempio è rappresentato dai diagrammi di Voronoi.

In uno spazio unidimensionale, una metrica potrebbe essere il semplice operatore "-", ovvero la distanza tra due numeri (relazione di confronto tra due numeri).

I profili di carico descritti sono, di fatto, delle serie temporali quindi inquadrabili come elementi di uno spazio metrico consentendo l'applicazione delle metriche come "parametro di confronto".

## 2.2 Norma di un vettore

Dato un vettore  $x \in R^n$ , la norma è una funzione che assegna una lunghezza positiva:

$$\|x\|_p = (\sum_{i=1}^n x(i)^p)^{\frac{1}{p}}$$

Con  $p \in [1, +\infty)$ . Se  $p = 2$  si ha la **Norma euclidea** (norma L2). Va osservato che applicando tale norma a vettori distinti, si potrebbe ottenere lo stesso risultato.

Questo aspetto lo si ritrova nell'elettrotecnica elementare: due segnali differenti potrebbero presentare lo stesso valore R.M.S.:

$$x_{RMS} = \sqrt{\frac{1}{n} \sum_{i=1}^n x(i)^2}$$

Pertanto due serie temporali con andamento e picchi differenti potrebbero essere associate allo stesso cluster se si utilizzasse tale norma. Questo spiega il perchè, in presenza di distorsione, anche una misura con uno strumento a "True R.M.S. Value" potrebbe non rilevare anomalie.

Si vuole insistere su tale aspetto dal momento che, in determinate condizioni, due elementi potrebbero risultare estremamente simili in caso di applicazione di una metrica, mentre totalmente dissimili nel caso di un'altra.

## 2.3 Metriche note in letteratura

- **Norma Minkowski**

Dati due vettori  $x, y \in R^n$ , la distanza tra gli stessi può essere calcolata tramite la Norma Minkowski:

$$d_M(x, y) = (\sum_{i=1}^n |x(i) - y(i)|^p)^{\frac{1}{p}}$$

Ove  $p$  può essere considerato un indice di dissimilarità, il cui valore porta ad avere specifiche norme:

- **Norma Manhattan** ( $p = 1$ )

Tale norma prende il nome dal quartiere della città di New York per la sua struttura urbanistica a scacchiera. Viene anche definita "la norma dei Taxi":

$$d(x, y) = (\sum_{i=1}^n |x(i) - y(i)|)$$

Di fatto, è la somma delle differenze tra gli elementi.

– **Norma Euclidea** ( $p = 2$ )

$$d(x, y) = (\sum_{i=1}^n |x(i) - y(i)|^2)^{\frac{1}{2}}$$

– **Norma sup** ( $p \rightarrow \infty$ )

Massima differenza tra le componenti dei vettori.

$$d(x, y) = \max\{|x(i) - y(i)|, i = 1, \dots, n\}$$

• **Norma di Canberra**

$$d(x, y) = \sum_{i=1}^n \frac{|x(i) - y(i)|}{|x(i) + y(i)|}$$

• **Coseno di similitudine**

Coseno dell'angolo  $\theta$  compreso tra due vettori: se essi sono coincidenti, il coseno di similitudine assume valore unitario. Di particolare interesse è osservare che viene considerato anche l'ordinamento degli elementi nei vettori. Di fatto, se due vettori presentano gli stessi elementi ma disposti diversamente (stessa area), il coseno di similitudine è inferiore a 1.

$$\cos(\theta) = \frac{\sum_{i=1}^t x(i)y(i)}{\sqrt{\sum_{i=1}^t x(i)}\sqrt{\sum_{i=1}^t y(i)}}$$

Dal momento che, in questa sede, si tratterà il problema del clustering come un problema di minimizzazione, si vuole un parametro che restituisca zero in caso di totale similarità. Per tale motivo, verrà considerato il complemento a 1 del coseno di similitudine.

## 2.4 Metriche in Matlab

In MatLab vengono messe a disposizione delle funzioni (non solo in ambito clustering) basate su differenti metriche. Dal momento che se ne farà utilizzo, in questa sede, risulta necessario richiamarle adottando la notazione (vettoriale) utilizzata nella relativa documentazione.

- Squared Euclidean distance ('sqeuclidean')

$$d(x, y) = (x - y)(x - y)'$$

Si tratta della norma euclidea "al quadrato" (omettendo direttamente l'estrazione della radice quadrata presente nella definizione).

- Sum of absolute differences ('cityblock')  
Non è altro che la norma Minkowski con esponente 1 (Norma Manhattan);
- Complemento a 1 del coseno di similitudine ('cosine')  
Due vettori  $x, y \in R^p$ , un vettore riga di dimensione  $p$  con elementi unitari:

$$\vec{I}_p = \{1, 1, 1, \dots, 1\} \in R^p$$

Definendo vettori medi:

$$\vec{x} = \frac{1}{p}(\sum_{j=1}^p x_j)\vec{I}_p$$

$$\vec{y} = \frac{1}{p}(\sum_{j=1}^p y_j)\vec{I}_p$$

La correlazione vale:

$$d(\vec{x}, \vec{y}) = 1 - \frac{(\vec{x} - \vec{x})(\vec{y} - \vec{y})}{\sqrt{(\vec{x} - \vec{x})(\vec{x} - \vec{x})'}\sqrt{(\vec{y} - \vec{y})(\vec{y} - \vec{y})'}}$$

### 3 Entropia

Il concetto di entropia, in ambito della fisica classica, sta ad indicare il "grado di disordine" in un sistema.

Essa fissa la "freccia del tempo": un sistema isolato evolve verso l'entropia massima.

Associando quindi una probabilità  $W$  (formalmente un rapporto tra probabilità di due stati) ad uno stato di tale sistema, l'entropia è data dalla formula di Boltzmann:

$$S = K \log(W)$$

ove  $K$  è la costante di Boltzmann. Approdando alla Teoria dell'informazione (TI), si può studiare matematicamente la trasmissione e l'elaborazione delle informazioni.

Volendo trasmettere  $n$  simboli che possono assumere  $S$  valori, l'informazione trasmessa vale:

$$H = n \log(S)$$

Si tratta quindi di una quantificazione della capacità del ricevente di distinguere tra più sequenze. Affascinante è il constatare la similitudine con la formulazione proposta da Boltzmann in ambito della Termodinamica.

Se ad esempio si trasmettesse un carattere binario ( $n=1$   $s=2$ ), l'entropia varrebbe:

$$H = 1 \log_2 2 = 1$$

Quindi il numero di bit necessari a trasmettere l'informazione.

Considerando un fenomeno aleatorio, la variabile aleatoria associata  $X \in R^n$  che può assumere il valore  $x_i$ , l'entropia vale:

$$H(X) = - \sum_{i=1}^n P(x_i) \log P(x_i)$$

ove il termine relativo al logaritmo prende il nome di autoinformazione, ovvero l'informazione contenuta nel generico evento  $x_i$ :

$$I(x_i) = -\log P(x_i)$$

Utilizzando logaritmi in base due si ottengono i bit necessari a descrivere l'informazione, mentre utilizzando il logaritmo naturale si passa ai "nat" (1 bit = 0,69315 nat).

Un riscontro si ha applicando concetti elementari di calcolo binario.

Nota la probabilità che l'evento si manifesti, si risale al quantitativo informativo estrapolabile: maggiore è la probabilità che si manifesti maggiore è il contenuto estrapolabile.

Il valore massimo che può assumere l'entropia vale:

$$H(X) \leq \log_b(n)$$

che si ottiene qualora gli eventi siano equiprobabili.

Nel contesto della suddivisione in cluster, l'entropia può essere applicata in funzione di "metro" della similitudine tra elementi.

Si può inoltre applicare come "peso" per le varie metriche.

Coerentemente con con la definizione di "clustering", si può traslare il concetto di entropia in tale ambito. Se, di fatto, per cluster si intende un sottoinsieme di elementi "simili", si assume che il miglior risultato ottenuto è quello che:

- Minimizza l'entropia inter cluster;
- Massimizza l'entropia intra cluster;

### 3.1 Entropia nel clustering (Renyi Entropy)

Considerando  $K$  cluster, l'entropia complessiva (insieme di cluster), può essere calcolata come segue:

$$H_{tot} = \sum_{c \in C} H_c \frac{N_c}{N}$$

ove  $H_c$  è l'entropia del cluster  $c \in C$ ,  $N_c$  è la sua cardinalità e  $N$  è il numero complessivo di elementi.

Considerato un database di  $n$  elementi (vettori di dimensione  $M$ ):

$$X = (X_1, \dots, X_n)^t$$

Stoccati come righe di un array bidimensionale (quindi la  $i$ -esima riga contiene il vettore  $X_i \in R^M$ , con  $i = 1, \dots, n$ ), viene definita una matrice diagonale in cui il generico elemento contiene la varianza relativa alla relativa colonna (avente valore medio  $\mu$ ):

$$\Sigma_X = \begin{bmatrix} \sigma_{11}^2 & 0 & \dots & \dots & 0 & 0 \\ 0 & \sigma_{22}^2 & \dots & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & \dots & \dots & \sigma_{M-1M-1}^2 & 0 \\ 0 & 0 & \dots & \dots & 0 & \sigma_{MM}^2 \end{bmatrix}$$

La similarità tra due elementi viene calcolata basandosi sul concetto di distribuzione di Gauss multivariata:

$$Gauss(X_i, X_j, \Sigma_X) = \frac{1}{\sqrt{(2\pi)^M \det(\Sigma_X)}} e^{-0,5(X_i - X_j)^t \Sigma_X^{-1} (X_i - X_j)}$$

Grazie alla matrice diagonale delle varianze, si ha una contestualizzazione: uno degli argomenti della suddetta funzione tiene conto dell'intero database. Si può applicare tale definizione anche tra due cluster:

$$S(C_q, C_z) = \frac{1}{N(q)N(z)} \sum_{x^{(i)} \in C_q} \sum_{x^{(i)} \in C_z} Gauss(X_i, X_j, \sum_X)$$

Essendo diagonale, la matrice delle varianze è invertibile solo se tutti gli elementi sulla diagonale sono diversi da zero, condizione non sempre verificata. Qualora, in determinati momenti, tutti i carichi risultassero disconnessi (es. Black Out) o se l'assorbimento complessivo (nel sottoperiodo) fosse nullo (presenza di carichi attivi), non sarebbe possibile eseguire il calcolo (determinante nullo)

La soluzione potrebbe essere l'individuazione della matrice "pseudoinversa" oppure escludere le variabili che portano a tale condizione.

### 3.2 Divergenza di Kullback–Leibler

Considerando i profili di carico come distribuzioni probabilistiche, una misura dell'informazione perduta utilizzando la distribuzione  $P$  piuttosto che la distribuzione  $Q$ , è data dalla Divergenza di Kullback–Leibler:

$$D_{KL}(P||Q) = \sum_i P(i) \log_2\left(\frac{P(i)}{Q(i)}\right)$$

A differenza delle metriche, si tratta di una misura asimmetrica:

$$D_{KL}(P||Q) \neq D_{KL}(Q||P)$$

L'applicazione di tale divergenza ai profili di carico, richiede delle accortezze. Per questioni di calcolo numerico, di fatto, si rende necessario escludere le singolarità:

$$P(i) = 0 \rightarrow \text{L'argomento del logaritmo non può essere negativo}$$

$$Q(i) = 0 \rightarrow \text{singolarità nell'argomento del logaritmo}$$

Qualora si manifestasse questo scenario, non è possibile proseguire il calcolo analitico. Si rende necessario quindi forzare la divergenza a zero (come proposto in parte della letteratura di dominio):

$$P(i) = 0 \text{ or } Q(i) = 0 \rightarrow D_{KL}(P||Q) = 0$$

La letteratura propone una versione di tale divergenza resa simmetrica che prende il nome di "divergenza di Jeffreys":

$$D_{KL}(P, Q) = D_{KL}(P||Q) + D_{KL}(Q||P)$$

## 4 Indicatori di validità

In [1] vengono proposte delle procedure per la valutazione delle prestazioni dei metodi di clustering applicati a profili di carico.

I risultati di un clustering possono essere "validati" prendendo in considerazione i dati noti a priori o i risultati stessi. In termini pratici, vengono calcolati alcuni "indici" che possono stabilire se un insieme è più o meno coeso rispetto agli altri.

Questo traduce lo studio in un problema di ottimizzazione in cui le funzioni obiettivo sono proprio i vari indicatori di validità. Si osservi che emerge uno dei aspetti più criptici di tale materia: gli indici di validità possono basarsi su "criteri interni", ovvero non vi è alcun modo di supervisionare la soluzione mentre viene costruita dall'algoritmo.

Una volta eseguita la suddivisione in  $K$  cluster, si ha (di conseguenza) a disposizione l'insieme  $R$  dei pattern "rappresentativi":

$$R = \{r^{(i)}, i = 1, 2, ..K\}$$

ai quali l'algoritmo di clustering associa un insieme di elementi  $L^{(i)}$  (singolo cluster).

La distanza tra un elemento ( $r^{(i)}$ ) e un insieme ( $L^{(i)}$ ) può essere calcolata come la media geometrica tra le distanze tra il singolo elemento e ogni elemento dell'insieme:

$$d^2(r^{(i)}, L^{(i)})$$

Mentre la distanza infra-gruppo, può essere calcolata come la media geometrica tra le distanze tra elementi del gruppo:

$$d(L^{(i)})$$

L'indice più semplice che può essere concepito per un "insieme partizionato" è la sparsità delle matrici quadrate (le cui dimensioni sono pari alla cardinalità dell'insieme di partenza) in cui la generica "cella" contiene il valore 1 se i due elementi fanno parte dello stesso cluster, zero altrimenti.

Ovviamente si utilizzano indici più complessi, elencati a seguire.

Il singolo valore, comunque, non ha validità statistica. Potrebbe trattarsi di un outlier (elemento anomalo rispetto agli altri) oppure non rappresentare la soluzione migliore (o peggiore). Per tale motivo, si rende necessario uno studio statistico prendendo in considerazione intere distribuzioni di soluzioni (per ogni metodo)

## 4.1 Davies-Bouldin Index (DBI)

Si consideri l' $i$ -esimo cluster contenente  $N$  elementi (di dimensione  $n$ ) e sia  $A_j$  il relativo centroide .

Sia  $X_j$  il generico elemento ( $j = 1, \dots, N$ ) e la dispersione intra-cluster è così definita:

$$S_i = \left\{ \frac{1}{N_i} \sum_{j=1}^{N_i} |X_i - A_i|^q \right\}^{\frac{1}{q}}$$

Sovente si utilizza  $q = 2$ . Nel 1979 Davies e Bouldin introdussero la seguente definizione di Separazione tra cluster:

$$M_{ij} = \left\{ \sum_{h=1}^n |a_{hi} - a_{hj}|^p \right\}^{\frac{1}{p}} = \|A_i - A_j\|_p$$

Quindi si tratta semplicemente della norma Minkowski tra due centroidi (in genere di utilizza la norma Euclidea), pertanto sarà verificata la seguente condizione (simmetria):

$$M_{ij} = M_{ji}$$

Di conseguenza la similarità tra due cluster viene definita come:

$$R_{ij} = \frac{S_i + S_j}{M_{ij}}$$

Tanto più due cluster sono "simili" tanto minore sarà l'indice  $R$ . Va osservato che, qualora i centroidi fossero perfettamente coincidenti (ad esempio, se il prodotto scalare fosse unitario), la relativa distanza sarebbe nulla e ciò comporterebbe divergenza numerica: il termine  $M_{ij}$  (al denominatore) sarebbe nullo.

Ad ogni modo, si ha simmetria:

$$R_{ij} = R_{ji}$$

Una volta calcolati tali coefficienti (notare che non ha senso comparare un cluster a se stesso), si associa ad ogni cluster il massimo valore di similarità tra se stesso e tutti gli altri cluster, ad esempio per l' $i$ -esimo cluster:

$$R_i = \max_{i \neq j} \{R_{ij}\}$$

Quindi si può procedere calcolando l'indicatore di Davies Bouldin vero e proprio per un generico numeri di cluster  $NC$ :

$$DBI(NC) = \frac{1}{NC} \sum_{i=1}^{NC} R_i$$

Nel calcolare la similarità tra due cluster ( $R_{ij}$ ) si potrebbe incorrere in una singolarità, ovvero se la distanza tra due baricentri fosse nulla ( $M_{ij} = 0$ ). Questo accadrebbe qualora due o più elementi del database siano coincidenti e vengano "estratti" come medoidi: risulta evidente che la distanza tra cluster intesa come distanza tra i baricentri sarebbe nulla.

L'associazione ai cluster presenterebbe quindi una perturbazione dal momento che qualsiasi elemento dovrebbe essere assegnato "casualmente" all'uno o all'altro cluster (perchè presenterebbero gli stessi baricentri).

## 4.2 Scatter index (SI)

Partendo da un database composto da  $M$  elementi, si definisce pooled scatter la media globale:

$$p = \frac{1}{M} \sum_{m=1}^M x^{(m)}$$

Il rapporto tra la somma delle distanze (al quadrato) tra ogni elemento e il pool scatter e della somma delle distanze (al quadrato) tra i centroidi e il pool scatter.

$$SI(NC) = \frac{\sum_{m=1}^M d^2(x^{(m)}, p)}{\sum_{k=1}^{NC} d^2(c^{(m)}, p)}$$

## 4.3 Errore di Sammon

L'errore di Sammon (definito nei precedenti capitoli) può essere utilizzato come indicatore di validità.

## 5 Interpretazione dei risultati

Prelevando, tra i risultati ottenuti, quello che porta ad avere il minimo valore dell'indice di validità considerato (in questo caso il DBI), si applica la logica elitista. Risulta necessario rammentare che non si è in ambito matematico, quindi il risultato "numerico" non rappresenta la conclusione dello studio, ma l'esperto di dominio deve trarre le adeguate conclusioni. Come esempio, viene riportata la "soluzione migliore" di un'elaborazione:

Tipo distanza: Camberra--Centro attrazione: Media Armonica--Tolleranza: 1.3795

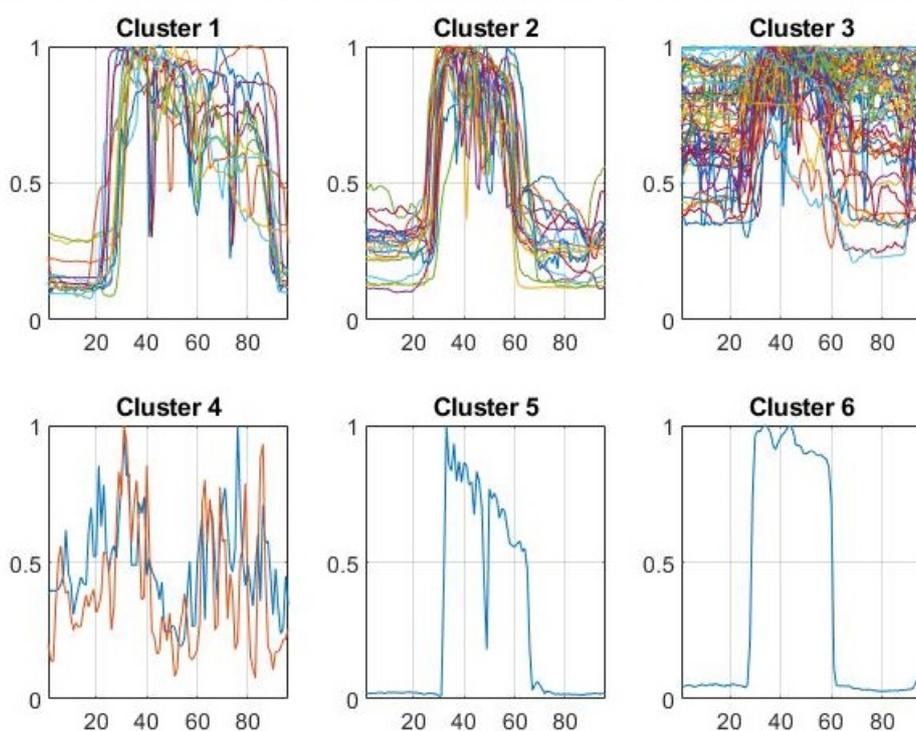


Figura 9: Risultato ottimale (FTL)

La bontà di tale suddivisione, quantificata grazie all'indicatore di validità in oggetto, può rafforzata da una valutazione "visiva":

- **Cluster 1** : In esso l'algorithmo ha allocato utenze che presentano una brusca variazione di assorbimento in determinate coordinate temporali (es. ora di pranzo e ora di cena).

- **Cluster 2:** Sono presenti utenze con netta distinzione di assorbimento tra le ore diurne e notturne (durante tali sottoperiodi i carichi presentano variabilità modesta).
- **Cluster 3:** Utenze che non scendono mai, nelle 24 ore, sotto un certo livello di assorbimento;
- **Cluster 4:** Carichi con notevole variabilità (brusca) nell'arco della giornata.
- **Cluster 5:** Cluster con un solo elemento. L'utenza presenta un picco di assorbimento durante le prime ore del giorno per poi assorbire sempre meno (decrecita marcata) fino alle ore serali. Assorbimenti modesti nelle ore notturne.
- **Cluster 6:** Cluster con un solo elemento. Il pattern mostra la massima escursione tra massimo e minimo rispetto al resto del database.

Si badi che questa interpretazione è propria di chi scrive, ulteriori valutazioni dovrebbero essere implementate per la verifica di quanto detto (non oggetto di questo lavoro).

Il punto è che, considerando lo stato dell'arte, non è possibile stabilire a priori quale sia la classificazione migliore.

Un'idea potrebbe essere l'individuazione di pattern di riferimento con i quali tentare l'accostamento, o quantomeno stabilire a priori quale "proprietà" si intende attenzionare. Se ad esempio si volesse classificare una curva di carico su base energetica (area del profilo di carico), classificare sulla base di profili normalizzati porterà a risultati fuorvianti e totalmente inesatti.

## 6 Indicatore OPISD

La natura aleatoria delle soluzioni rende necessario confrontare i metodi con modalità tipiche dell'analisi statistica. Si tratta di funzioni di ottimizzazione in cui la funzione obiettivo è rappresentata da un indicatore di validità. Applicando una logica elitistica, si può prelevare e conservare la soluzione migliore (controllo a valle), ma non sarebbe possibile avanzare delle ipotesi sulla bontà del metodo che ha condotto alla suddetta soluzione. Si rende quindi necessario considerare delle distribuzioni di soluzioni, entrando in un dominio stocastico.

Il metodo  $m_1$  potrebbe portare ad una soluzione migliore del metodo  $m_2$ , ma questo non è sufficiente a garantire che  $m_1$  sia generalmente migliore di  $m_2$ .

### 6.1 Concetto di dominanza stocastica

Osservando una famiglia di distribuzioni (e quindi le relative cumulative), si può stabilire quale sia la migliore in quel gruppo. Se il problema di ottimizzazione richiede una minimizzazione (si vuole individuare il metodo che produce bassi valori di indici), allora risulta evidente che il metodo migliore è quello che presenta una cumulativa maggiormente spostata a sinistra rispetto alle altre: quella che "domina", in maniera stocastica, sulle altre.

Dati due metodi  $m_i$  e  $m_j$  le cui uscite portano alle cumulative  $F_{m_i}^H$  e  $F_{m_j}^H$ , le cui dimensioni sono funzioni del numero di ripetizioni  $H$ , si può affermare che il metodo  $m_i$  presenta dominanza stocastica sul metodo  $m_j$  se:

- Dominanza stocastica del primo ordine:

$$F_{m_1}^H(y) > F_{m_2}^H(y) \rightarrow \forall y$$

- Dominanza stocastica del secondo ordine:

$$A_{m_1, m_2}^H(y) = \int_{z=0}^y (F_{m_1}^H(z) - F_{m_2}^H(z)) > 0$$

Area differenziale tra le due C.D.F

Notare che, se  $m_1$  è dominante al primo ordine, lo sarà anche al secondo mentre non è verificato l'inverso.

In altri termini, si tratta di valutare la mutua posizione delle due cumulative. Maggiore è il numero di ripetizioni ( $H$ ), migliore sarà il risultato in accordo con uno dei concetti elementari della statistica: l'elevato numero di campioni filtra l'eccezione.

In [4] viene proposto l'utilizzo del concetto di dominanza stocastica nell'ottimizzazione delle soluzioni per metodi basati su variabili aleatorie.

Si costruisce la curva cumulativa di riferimento  $CDF_{ref}$ :

1. Si concatenano gli  $n$  vettori contenenti le soluzioni ottenute dai vari metodi:
  - La dimensione  $q$  dei vettori coincide con il numero di ripetizioni;
  - Il numero di vettori  $n$  è pari al numero di metodi considerati;
2. I vettori vengono concatenati così da formare un singolo vettore la cui dimensione è pari al prodotto tra numero di ripetizioni per il numero di metodi ( $\in R^{n,q}$ );
3. Tale vettore viene "ordinato", ovvero si riorganizzano in ordine crescente gli elementi;
4. Si prelevano  $H$  elementi:

$$H < q$$

- Se si tratta di un problema di minimizzazione, si prelevano i primi  $H$  elementi;
  - Di contro, se si ha un problema di massimizzazione, si prelevano gli ultimi  $H$  elementi;
5. La distribuzione così ottenuta sarà associata ad una curva (cumulativa) costruita da  $H$  gradini aventi altezza  $\Delta c = \frac{1}{H}$ ;

Per costruzione, la cumulativa di riferimento presenterà una totale dominanza stocastica su ognuna delle curve in esame.

## 6.2 Definizione dell'indicatore OPISD

L'OPISD (Optimization Performance indicator based on first-order Stochastic Dominance), relativo al metodo  $m_i$  una volta nota la cumulativa di riferimento, è definito come segue:

$$OPISD_{m_i}^H = \frac{1}{1+A_{m_i}^H}$$

ove:

$$A_m^H = \int_{z=0}^y (F_{m_1}^H(z) - F_{ref}^H(z))$$

ovvero l'area compresa tra le due curve. Tale parametro decresce al discostarsi della cumulativa ottenuta con il metodo in esame da quella di riferimento (se sono coincidenti, l'OPISD è pari a 1). Pertanto, il metodo che nella simulazione in esame ha presentato prestazioni migliori sarà quello che porta ad un OPISD maggiore. Nota la famiglia di distribuzioni (risultati), tutti gli elementi appartenenti ad esse vengono organizzati in un unico vettore (ordine crescente), formando la distribuzione di riferimento  $y_{rif}$ . La relativa cumulativa viene costruita prelevando il valore massimo, tra tutte le cumulative, associate al valore in esame. Sostanzialmente, si esegue un involucro: la cumulativa di riferimento  $CDF_{rif}$  poggia sulle altre cumulative (costruzione geometrica). A seguire, un esempio:

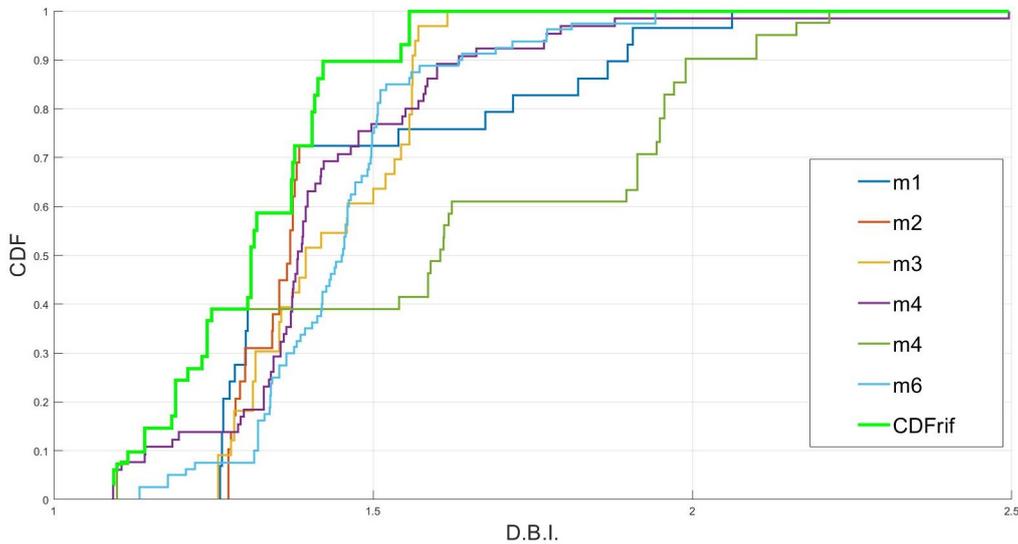


Figura 10: Famiglia di cumulative

## 7 Esempio

A seguire un esempio dell'applicazione dell'indicatore OPISD per questo studio per i metodi basati sul KMEANS. La funzione `kmeans()` in MatLab, che prevede la seguente sintassi:

$$[\text{idx}, \text{C}] = \text{kmeans}(\text{dataset}, \text{Nc}, \text{'Distance'}, \text{'tipo metrica'})$$

- **idx**: Vettore in cui il  $j$ -esimo elemento è il cluster al quale è stato associato il  $j$ -esimo elemento del dataset;
- **C**: Matrice le cui righe contengono i centroidi ottenuti a seguito dell'elaborazione;
- **dataset**: Matrice le cui righe contengono le serie temporali;
- **Nc**: numero di cluster desiderati (classe del risultato);
- **tipo metrica**: il tipo di distanza utilizzato dall'algorithm;

Per ogni classe, vengono tracciate le cumulative degli indicatori di validità (ad esempio, l'errore di Sammon) per ogni metodo:

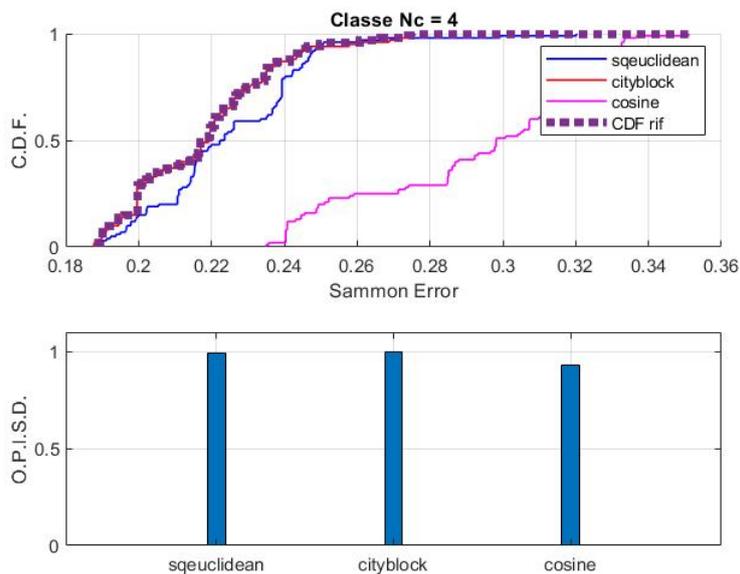


Figura 11: OPISD per la classe "Nc=4"

Calcolando gli indicatori (uno per metodo) per ogni classe di risultato (numero di cluster), si ha il seguente risultato:

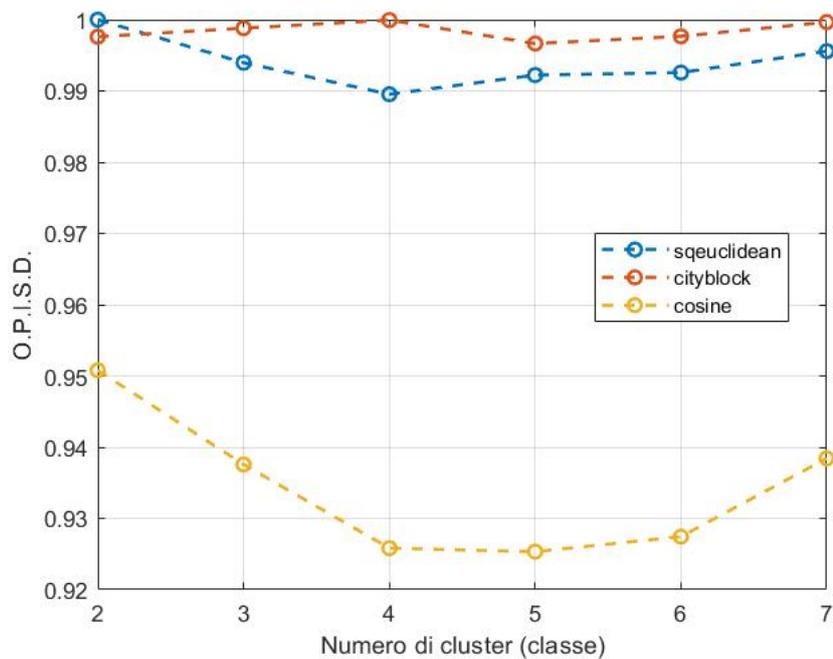


Figura 12: Andamento degli indicatori OPISD al variare della classe del risultato

Nel caso in esame, si evince come il metodo "rosso" tende a ottimizzare (minimizzare) la funzione obiettivo (indicatore di validità) in maniera più marcata rispetto agli altri cluster. Il valore ottimale, ad esempio, potrebbe rappresentare un "quartile" mentre il valor medio potrebbe essere assai più ampio.

## 8 Analisi nel dominio della frequenza

Il concetto di similitudine tra elementi (metro di raggruppamento) può essere esteso alla relative proprietà.

Il discriminare tra una proprietà e l'altra è l'aspetto che si intende attenzionare.

Se si è interessati a raggruppare dei carichi in funzione della loro variabilità nel tempo, ad esempio, è possibile calcolarne la varianza. Se, invece, si intende valutare il "tipo" (residenziale, industriale ecc) è possibile fare delle valutazioni sui pattern (che, come già evidenziato, sono "tipici").

Applicando la trasformata di Fourier ai profili temporali, si possono fare delle considerazioni sulle armoniche: si passa al dominio della frequenza. In questa sede si eseguirà un'analisi nel dominio della frequenza, passando quindi dalle serie temporali ai coefficienti complessi di Fourier (serie di Fourier).

### 8.1 Trasformata di Fourier

Risulta necessario assumere che i profili di carico siano periodici: si tratta di un'assunzione piuttosto forzata ma che potrebbe trovare dei riscontri reali (ad esempio un carico temporizzato).

Assumendo due carichi equivalenti ma sfasati nel tempo, si osserva che a variare sono solo gli argomenti dei coefficienti complessi (di Fourier), quindi in essi risiedono le informazioni temporali. Di contro, le ampiezze delle armoniche non variano.

Considerando gli spettri come appartenenti ad uno spazio metrico, la distanza (es. norma euclidea) tra i vettori contenenti le ampiezze rimane nulla al variare dello sfasamento mentre per quanto riguarda le fasi si ha una variabilità non nulla.

Considerando un segnale  $x(t)$  (in questo caso la serie temporale) di cui sono stati prelevati  $N$  campioni sul periodo (assumendo rispettata la condizione di Shannon) si ha che il generico coefficiente complesso rappresentante l' $h$ -esima armonica, vale:

$$X_h = \sum_{n=0}^{N-1} x_n e^{-jh \frac{2\pi}{N} n}$$

Dando per scontato che il numero di campioni sia sufficiente (altra assunzione pesante), non ha senso utilizzare tutte le armoniche disponibili. Potrebbero verificarsi delle variazioni anomale dei carichi, che rappresentano quindi degli outlier, impattanti sugli spettri (moduli e fasi) portando a delle differenze (o similitudini) errate.

Inoltre, le fasi sono periodiche per antonomasia: presi due fasori (considerando la medesima origine delle fasi) uno con fase  $1^\circ$  e l'altro con fase  $359^\circ$ , risulteranno "vicini" e la loro somma vettoriale potrebbe essere approssimata con la somma algebrica.

Le curve di carico giornaliere (quartorarie) presentano dimensione 96. Dal momento che si tratta di profili non residenziali, ci si aspetta che le armoniche prevalenti non siano di ordine elevato, per tale motivo verranno considerate solo le prime 40 armoniche (decisione che, considerando un numero di campioni pari a 96, porta ulteriormente verso il soddisfacimento della condizione di Shannon per il campionamento).

Inoltre, le frequenze elevate sono associate a fenomeni la cui variazione è considerevole: potrebbe trattarsi di rumore o un degli overshoot nella potenza assorbita causati da atteggiamenti anomali del carico, non interessanti per fini statistici (outlier).

## 8.2 Harmonic order ranking

Una volta eseguita la DFT si avrà un vettore di coefficienti complessi che rappresenteranno la serie temporale di partenza.

In [2] viene proposto un metodo per la discriminazione degli ordini armonici da considerare, assumendo un determinato dataset di profili di carico costituito da  $M$  elementi.

Considerando il  $j$ -esimo ordine armonico viene calcolato l' "Harmonic order ranking" (HOR):

$$\delta(j) = \frac{1}{M} \sum_{i=1}^M |A_{ji} - \bar{A}_j|$$

Si tratta di un indicatore che esprime il pooled scatter intorno al valore medio: tanto minore sarà l'HOR, tanto più saranno concentrate le armoniche.

Si considerano, quindi, gli  $n$  ordini armonici ordinati secondo l'HOR.

Considerando il singolo ordine armonico, se le ampiezze sono "piccole" le fasi non danno informazioni rilevanti, viceversa, se le ampiezze sono notevoli, una piccola variazione di fase tra due numeri complessi si traduce in un elevato contenuto informativo (nella differenza).

Non avrebbe senso, pertanto, considerare un gruppo armonico eccessivamente compatto perchè discriminarne le proprietà implicherebbe l'utilizzo di metodi di notevole sensibilità.

L'HOR si traduce, per l'approccio adottato, si comporta da filtro per l'algoritmo di clustering.

Applicando tale indicatore al database in esame, considerando i primi 6 ordini armonici (sulla base di un ordinamento discendente dell'HOR), si ha il seguente risultato:

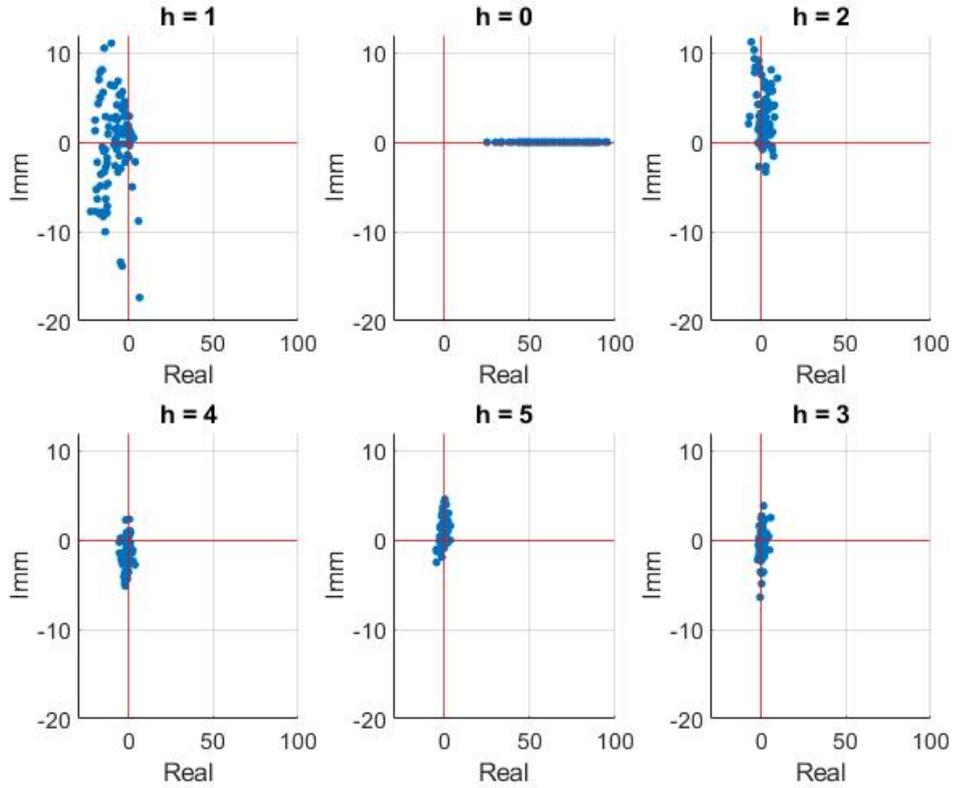


Figura 13: Armoniche ordinate secondo l'HOR

Considerando  $n$  armoniche che costituiranno l'insieme  $\theta_n$ , il cui  $h$ -esimo elemento è costituito dall'ampiezza  $A_h$  e dalla fase  $\alpha_h$ , si possono dedurre due parametri:

$$\psi_h^1 = \frac{1 - \cos(\alpha_h)}{2} \frac{A_h}{\sqrt{\sum_{j \in \theta_n} A_j^2}}$$

$$\psi_h^2 = \frac{1 - \sin(\alpha_h)}{2} \frac{A_h}{\sqrt{\sum_{j \in \theta_n} A_j^2}}$$

Tale parametro varia tra zero e uno. Ogni armonica di ogni elemento del dataset sarà quindi rappresentata da tre elementi: le serie temporali fornite agli algoritmi, dunque, avranno dimensione  $3 \times n$ .

### 8.3 Clustering

Utilizzare le armoniche come feature piuttosto che le serie temporali, ha come scopo quello di rendere più sensibili gli algoritmi. Essendo curve di carico di medesima natura, infatti, presentano andamenti analoghi (tipici del tipo di carico) ma rumorosità differente (tipica dell'utenza). Si vuole, pertanto, discriminare i carichi su basi di caratteristiche globali e non individuali. Due aziende gemelle, ad esempio, dovrebbero ricadere nello stesso cluster ma la tipicità individuale (es. rumore) tende ad inibire la capacità dell'algoritmo di classificarle come tali.

Ci si aspetta che le armoniche di ordine basso giochino un ruolo importante: di contro, in profili di carico residenziali, si potrebbe affermare il contrario. Si procederà classificando le curve di carico sulla base delle armoniche (ordinate in base all'HOR) al variare del numero delle stesse considerato, per poi calcolare gli indicatori sul database originale (serie temporali). Se il numero di armoniche considerato è troppo basso, non si riusciranno a classificare adeguatamente.

Viceversa, se il numero di armoniche considerare è alto, si ha una classificazione analoga a quella della serie temporale.

Come primo step, si ha la determinazione del parametro  $\delta$  per ogni ordine armonico:

$$\delta(0), \dots, \delta(H)$$

Ove  $H$  è il numero massimo di armoniche considerate. L'indicatore HOR viene poi ordinato in maniera decrescente:

$$\delta(a) > \delta(b) > \delta(b) > \dots$$

Quindi, gli ordini armonici considerati, ordinati, saranno i seguenti:

$$h = \{a, b, c, \dots\}$$

Partendo dal database (serie temporali) si ha il seguente database la cui  $i$ -esima riga ( $i$ -esimo elemento) sarà:

$$x'_i = \{A_{a_i}, \psi_{a_i}^1, \psi_{a_i}^2, A_{b_i}, \psi_{b_i}^1, \psi_{b_i}^2, A_{c_i}, \psi_{c_i}^1, \psi_{c_i}^2, \dots\}$$

Notare che, se tra le armoniche considerate compare componente continua i parametri  $\psi_{0_i}^1, \psi_{0_i}^2$  vengono omessi. Essa sarà dunque rappresentata dal solo modulo.

## 9 Between-Cluster Entropy-Based Clustering (BCEBC)

In [3] viene proposto un algoritmo gerarchico aggregativo basato sull'entropia di Renyi.

Considerando la suddivisione già eseguita, avendo ottenuto  $K$  cluster, si può calcolare l'entropia dell'intero set.

L'appartenenza di due elementi al medesimo cluster viene dedotta da un apposito coefficiente:

$$b(x^{(i)}, x^{(j)}) = \begin{cases} 1 & \rightarrow \text{Elementi in cluster differenti} \\ 0 & \rightarrow \text{Elementi di uno stesso cluster} \end{cases}$$

Supponendo ora che ogni cluster è composto da  $N^{(k)}$  elementi, l'entropia complessiva (a seguito della suddivisione) vale:

$$H(C^{(1)}, \dots, C^{(K)}) = -\log\left(\frac{\sum_{i=1}^n \sum_{j=1}^n b(x^{(i)}, x^{(j)}) Gauss(X_i, X_j, \Sigma_X)}{2 \prod_{k=1}^K N^{(k)}}\right)$$

Una volta escluso il  $v$ -esimo cluster dal set, viene calcolata l'entropia complessiva:

$$H^{(v)} = H(C^{(1)}, \dots, C^{(v-1)}, C^{(v+1)}, \dots, C^{(K)})$$

Viene individuato (per iterazione) il cluster la cui esclusione porta ad avere entropia minore:

$$C^{(w)} = \arg[\max_{v=1, \dots, k} \{H^{(v)}\}]$$

A seguire, tale cluster, viene accorpato al cluster, tra quelli rimanenti, con cui presenta la massima similarità.

$$C^{(h)} = \arg[\min_{i=1, \dots, k, i \neq w} \{S(C^{(i)}, C^{(w)})\}] \rightarrow C^{(h)} = C^{(h)} \cup C^{(w)}$$

Si tratta di un metodo prettamente deterministico, insensibile all'ordinamento degli elementi.

Plottando l'andamento dell'entropia complessiva durante l'evoluzione dei cluster (nei cicli dell'algoritmo) si ha il seguente risultato in termini di entropia totale:

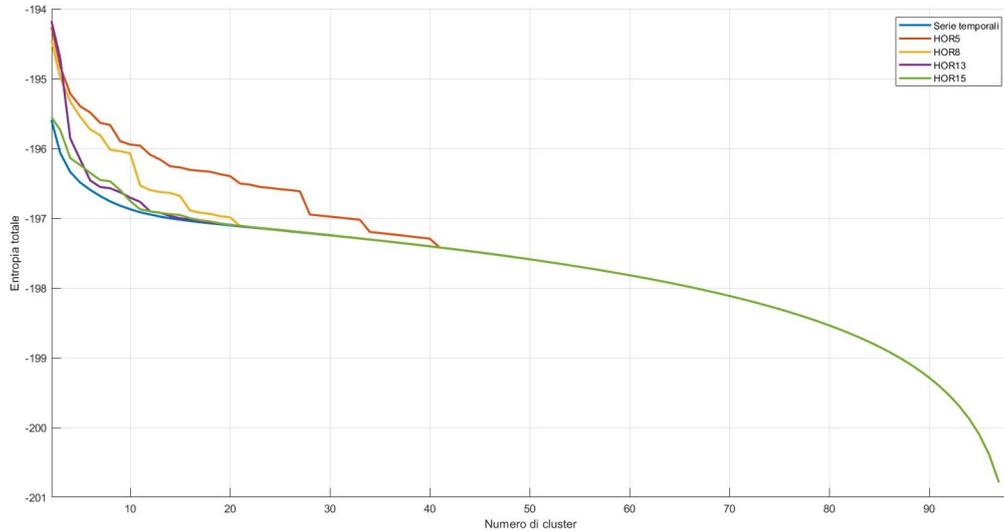


Figura 14: Andamento entropia al variare del numero di cluster

Le curve sono state tracciate partendo dalla suddivisione in cluster da parte del BCEDC: in ingresso le armoniche (ordinate secondo l'HOR) e l'entropia totale calcolata sull database di serie temporali.

Si osserva che considerando solo le prime 15 armoniche (secondo l'HOR), il risultato tende ad essere analogo. Va rammentato che si è interessati solo alle classi 2-10.

Questo grafico mostra come non si abbiano vantaggi ulteriori considerando ordini armonici superiori: la curva ottenuta direttamente dalla serie temporale e quella ottenuta con i coefficienti di Fourier coinciderebbero. Dal momento che si è interessati ad un basso numero di cluster in uscita, si "filtra" la particolarità della serie temporale eliminandone le armoniche di ordine superiore. Ottenere pochi cluster, infatti, implica cercare poche proprietà dal momento che è statisticamente improbabile che molti elementi presentino molte proprietà analoghe. L'operazione di filtraggio, dunque, migliorerà le prestazioni degli algoritmi (e quindi dei metodi) nella ricerca di soluzioni con basso numero di cluster.

## 10 Follow The Leader

L'algoritmo esegue una scansione degli elementi del database (che sarà ordinato arbitrariamente) misurando la distanza tra il singolo elemento e i cluster formati fino a quel momento. Il singolo elemento viene associato al cluster con il quale presenta distanza minore se essa è minore di una tolleranza  $\epsilon$  impostata: In caso contrario, viene creato un nuovo cluster in cui associare tale elemento.

$i$ -esima iterazione:

- Calcolo minima distanza tra l' $i$ -esimo elemento e i centri di attrazione dei  $K$  cluster esistenti all'iterazione  $i-1$  :

$$d = \min\{d(i, k) = d(x(i), c(k)), k = 1, 2, 3, \dots, K\}$$

- Associazione al cluster:

$$d < \epsilon \rightarrow x(i) \in C_k$$

$$d > \epsilon \rightarrow x(i) \in C_{K+1}$$

Il cluster di partenza è rappresentato dall'elemento in testa al database.

Una volta che tutti gli elementi sono assegnati ad un cluster (l'iterazione viene eseguita un numero di volte pari al numero di elementi nel database), si procede con l'ottimizzazione. Si calcolano i centri di attrazione dei cluster ottenuti e, a questo punto, si procede associando ogni elemento al centro di attrazione più vicino, quindi si ottengono nuovi cluster. Questo procedimento viene ripetuto fino a che non si ha convergenza della soluzione, ovvero fino a che due soluzioni successive non sono coincidenti.

Utilizzando i medoidi come centri di attrazione, si possono creare situazioni in cui non si ha convergenza: essendo "discreti", si può giungere a una condizione in cui due soluzioni si rincorrono, scambiandosi in maniera alternata. Dato che, generalmente, quattro o cinque iterazioni sono sufficienti per la convergenza, forzare l'uscita dopo dieci iterazioni (di ottimizzazione) può essere un buon metodo per evitare che il software rimanga intrappolato (never ending cycle). La natura di tale algoritmo è prettamente deterministica, ma è fortemente sensibile all'ordinamento degli elementi del database: il parametro aleatorio, in questo caso, è l'ordinamento dello stesso. Mantenendo quindi invariata la metrica e il centro di attrazione considerato ma variando casualmente l'ordinamento in input, si otterrà una distribuzione probabilistica ad esempio di indicatori di validità (metodo).

Va osservato che il numero di cluster (che dipende fortemente dalla tolleranza) non è noto a priori, ma rappresenta un output: questo rende necessario classificare i risultati a valle dell'elaborazione.

I tempi di calcolo possono essere considerevoli, pertanto è utile implementare degli accorgimenti quali interrompere l'iterazione qualora l'evoluzione (o la costruzione) della soluzione la porti in una direzione non di interesse, ad esempio se il numero di cluster che si stanno formando supera quello impostato (nel caso in esame non si è interessati a soluzioni formate da più di 10 cluster).

## 10.1 Impostazione del metodo

Ulteriore incognita è la tolleranza: il range di tolleranze (oltre che la relativa distribuzione) non è noto a priori. Il metodo in esame (variante FTL) deve tenere conto del database in input:

- **Tolleranza inferiore alla minima distanza tra elementi:** Potenzialmente, il numero di cluster che si viene a formare è pari alla numerosità dell'insieme di partenza.
- **Tolleranza superiore alla massima distanza tra elementi:** Potenzialmente, si ha la formazione di un solo cluster.

Pertanto, un metodo è definito al netto della definizione del range di tolleranze. In questa sede, si ha che il generico metodo è definito come:

- **Modalità di definizione del range di tolleranze;**
- **Tipo di distanza;**
- **Tipo di centro di attrazione;**

Dato un generico set di dati:

$$X = \{x_1, \dots, x_M\}^t$$

Si definisce la matrice delle distanze il cui generico elemento  $(ij)$  è definito:

$$D(i, j) = d(x_i, x_j)$$

Gli elementi del triangolo inferiore (o superiore, data la simmetria) vengono disposti in un vettore riga: la diagonale viene esclusa (contiene le distanze tra un elemento e se stesso, pertanto risulta contenere elementi nulli).

La tolleranza sarà quindi definita nel range:

$$\epsilon \in (\min(d_{ij}), \max(d_{ij}), d_{ij} \neq 0)$$

Tale range viene suddiviso in step (es. 20) formando, quindi, il vettore delle tolleranze.

Il seguente schema (Figura 15) mostra lo schema di principio del metodo.

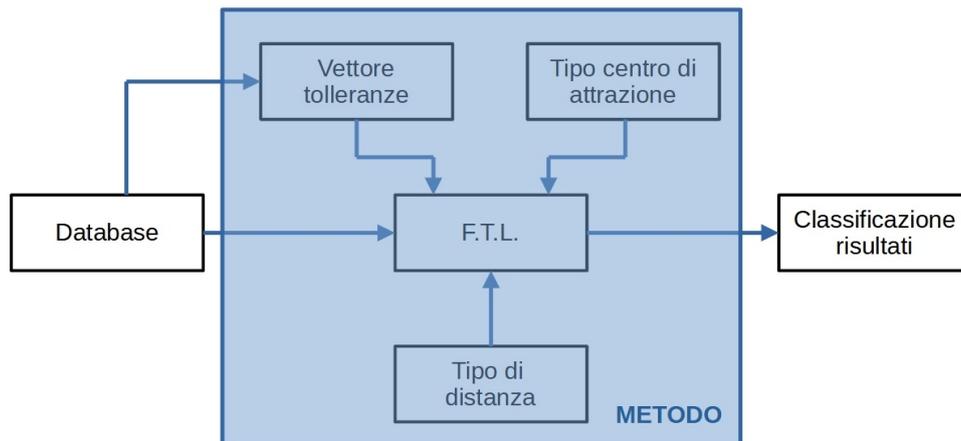


Figura 15: Impostazione del metodo

## 10.2 Numero di soluzioni

A seguire vengono riportate le "numerosità" delle soluzioni per le classi (Numero di cluster) 2-4.

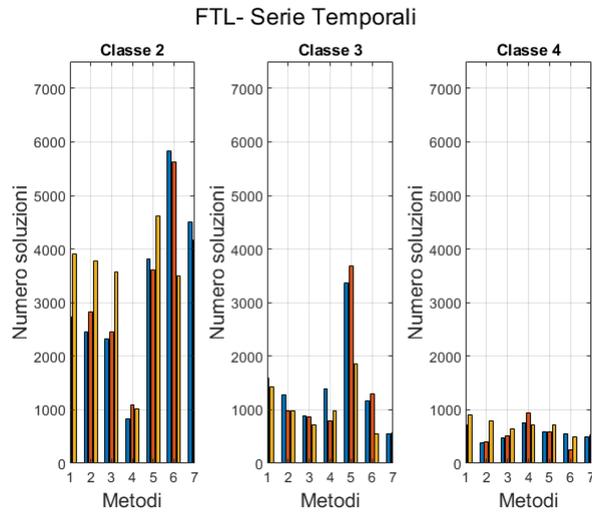


Figura 16: Numerosità soluzioni (Serie Temporal)

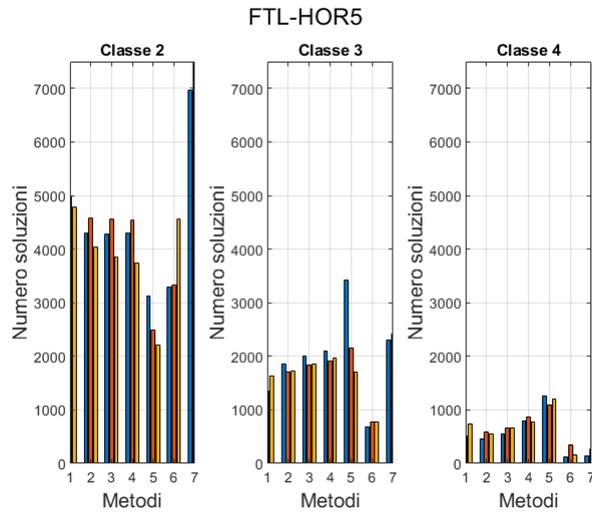


Figura 17: Numerosità soluzioni (HOR5)

I metodi, di fatto, non garantiscono risultati di specifiche classi: alcuni metodi, ad esempio, non portano a nessuna soluzione per le classi in esame. Si osserva come, utilizzando un basso numero di armoniche (eseguendo la classificazione nel dominio della frequenza), il numero di soluzioni per le classi inferiore tende ad essere inversamente proporzionale all'ordine armonico massimo considerato. Questo accade perché, l'utilizzo di bassi ordini armonici, è (di fatto) un filtro. Le differenze tra elementi vengono dunque ridotte e i metodi tendono ad essere meno sensibili.

Questo non comporta un miglioramento delle soluzioni. Si osserva, di fatto, la presenza di soluzioni "gemelle" utilizzando le prime 5 armoniche ordinate secondo l'HOR (HOR5). I profili di carico in esame presentano la stessa natura: per tale motivo presentano spettri inferiori "analoghi".

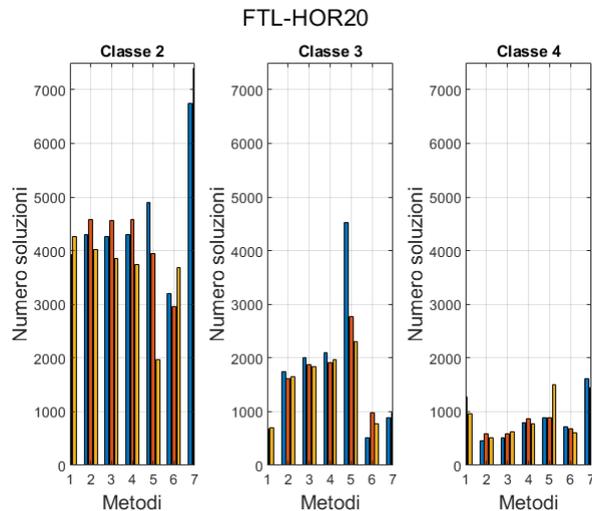


Figura 18: Numerosità soluzioni (HOR20)

Aumentando il contenuto armonico, si tende ad avere risultati analoghi al caso delle serie temporali.

### 10.3 Risultati FTL

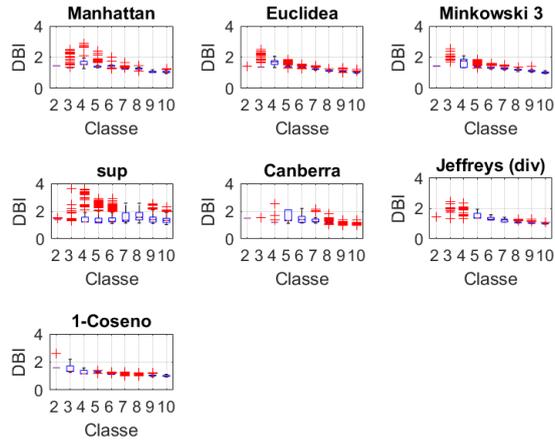


Figura 19: Distribuzioni risultati (DBI) - Serie temporali - centroidi

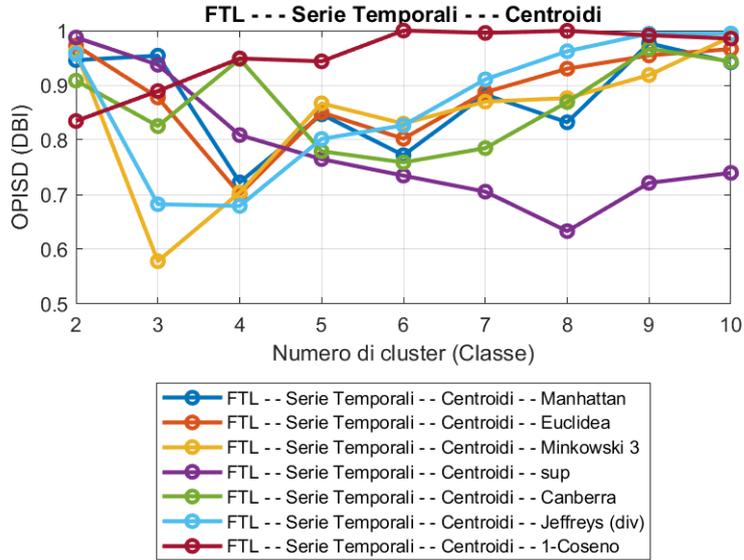


Figura 20: Distribuzioni risultati (DBI) - Serie temporali - centroidi

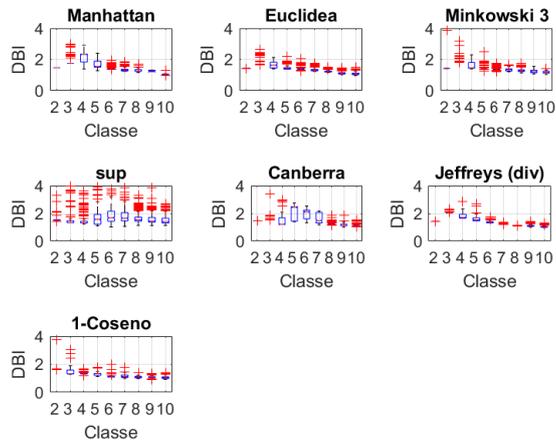


Figura 21: Distribuzioni risultati (DBI) - Serie temporali - medoidi

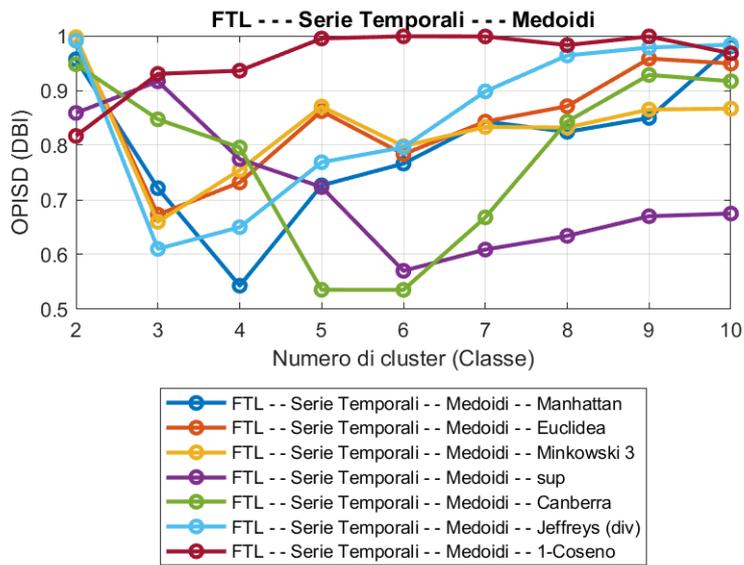


Figura 22: Distribuzioni risultati (DBI) - Serie temporali - medoidi

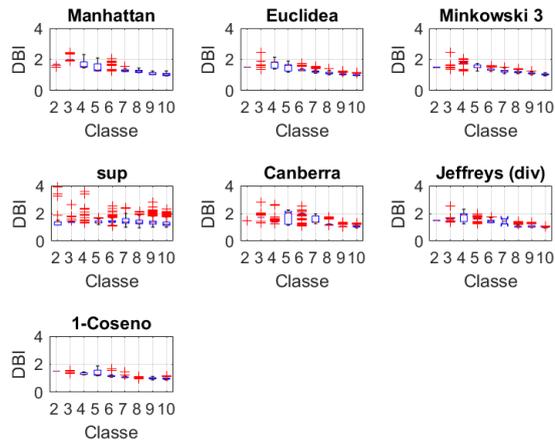


Figura 23: Distribuzioni risultati (DBI) - Serie temporali - media armonica

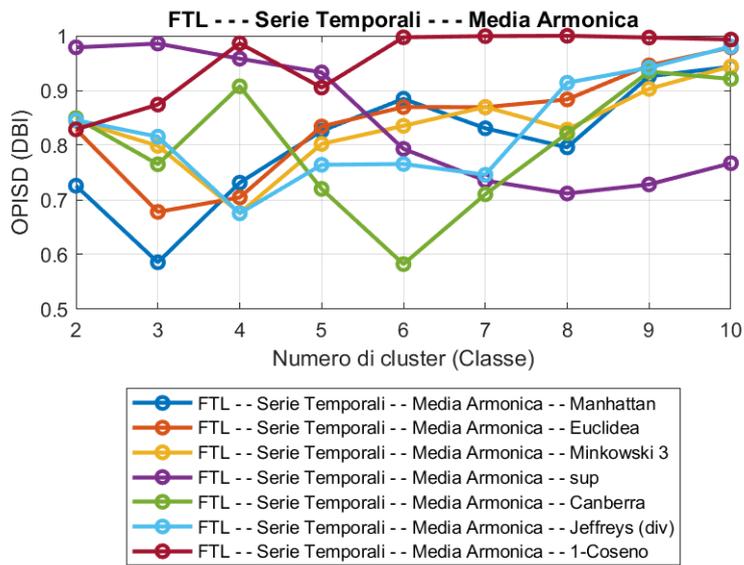


Figura 24: Distribuzioni risultati (DBI) - Serie temporali - media armonica

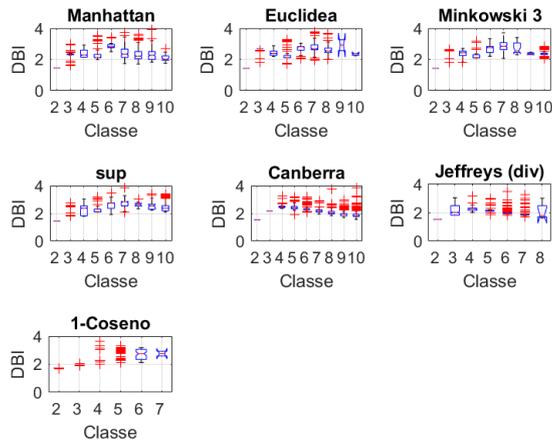


Figura 25: Distribuzioni risultati (DBI) - HOR5 - centroidi

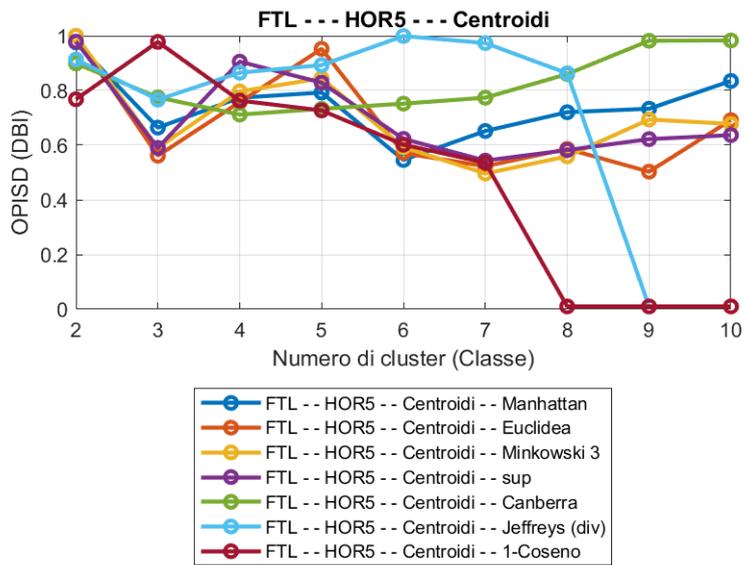


Figura 26: Distribuzioni risultati (DBI) - HOR5 - centroidi

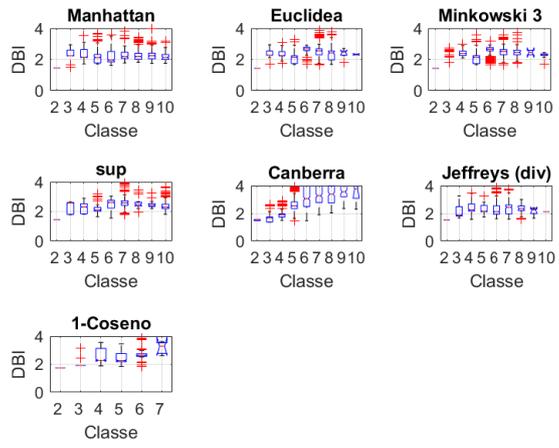


Figura 27: Distribuzioni risultati (DBI) - HOR5 - medoidi

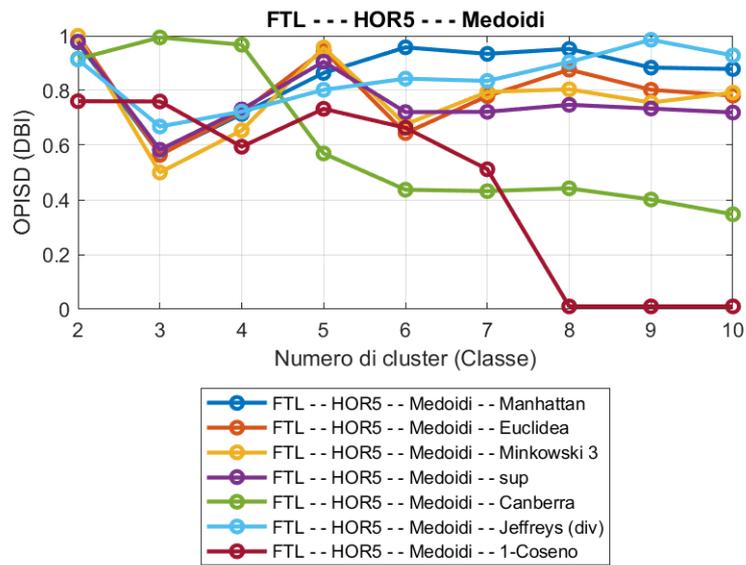


Figura 28: Distribuzioni risultati (DBI) - HOR5 - medoidi

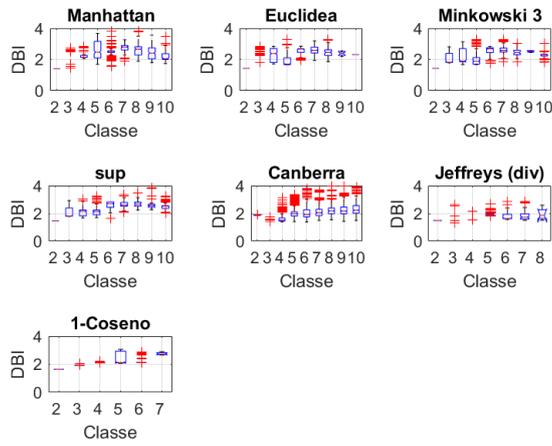


Figura 29: Distribuzioni risultati (DBI) - HOR5 - media armonica

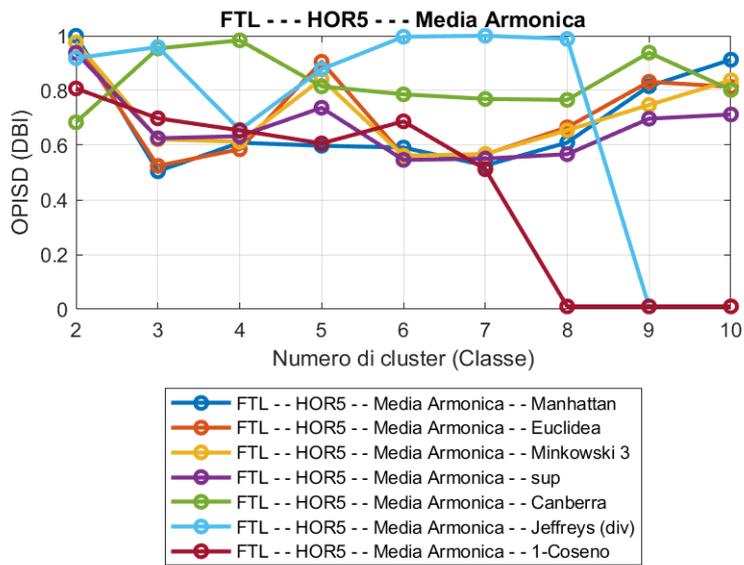


Figura 30: Distribuzioni risultati (DBI) - HOR5 - media armonica

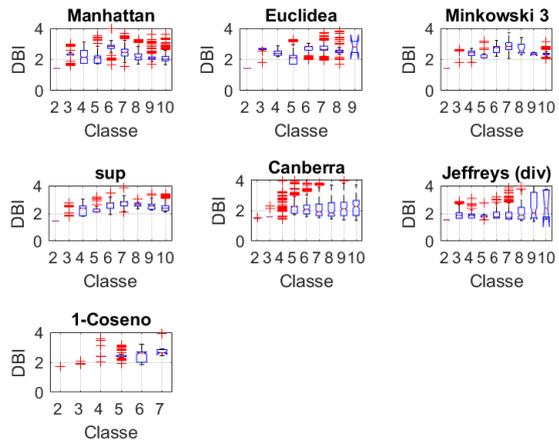


Figura 31: Distribuzioni risultati (DBI) - HOR8 - centroidi

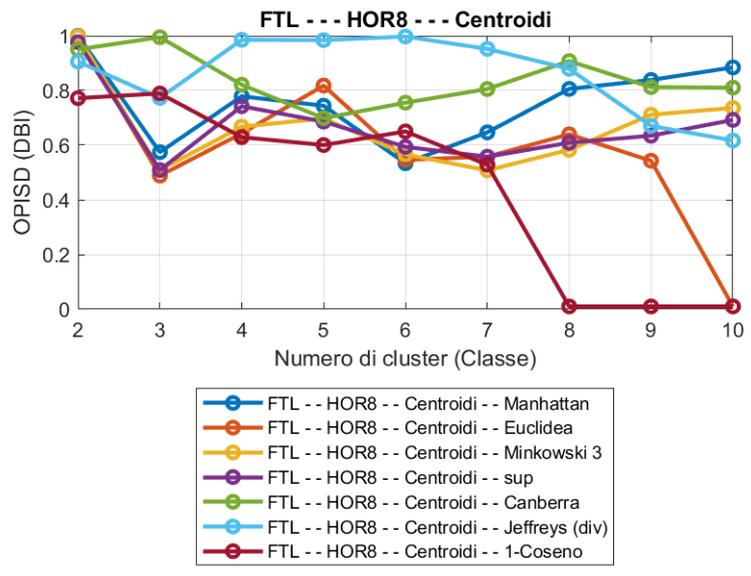


Figura 32: Distribuzioni risultati (DBI) - HOR8 - centroidi

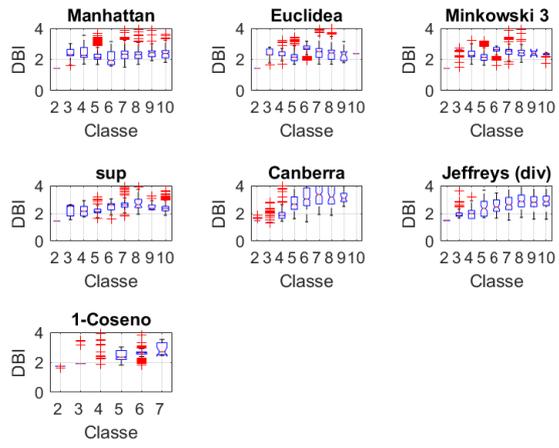


Figura 33: Distribuzioni risultati (DBI) - HOR8 - medoidi

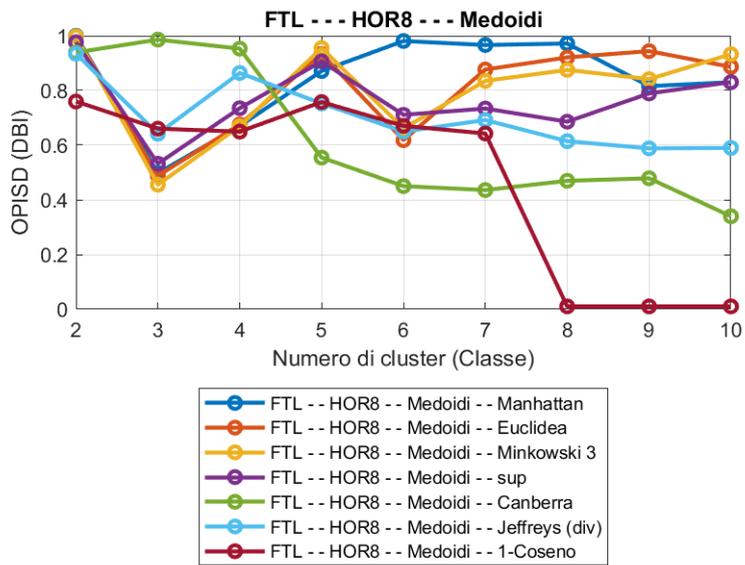


Figura 34: Distribuzioni risultati (DBI) - HOR8 - medoidi

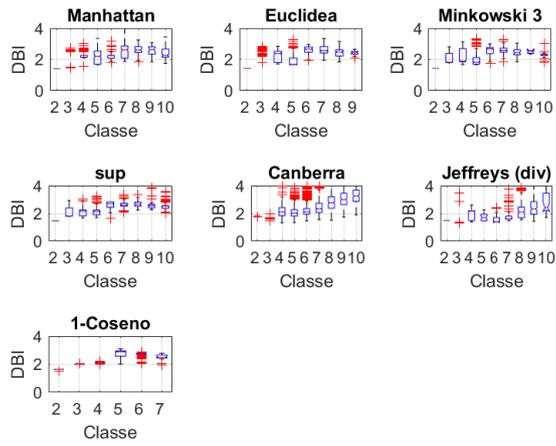


Figura 35: Distribuzioni risultati (DBI) - HOR8 - media armonica

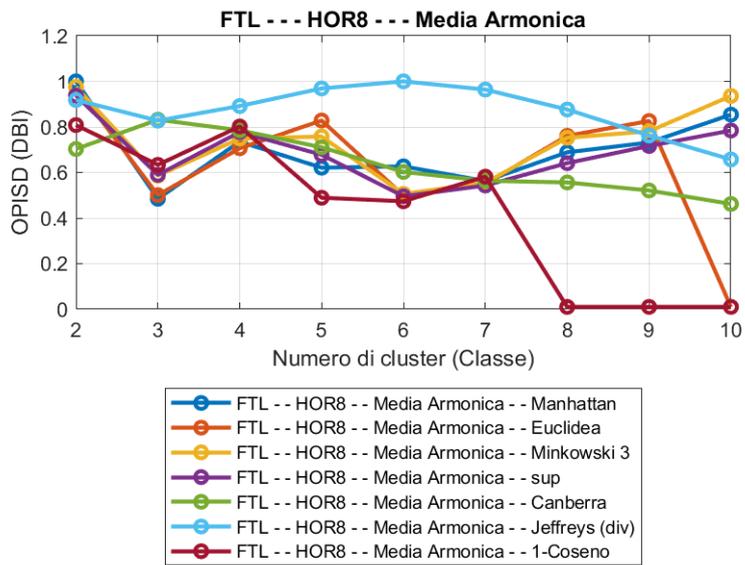


Figura 36: Distribuzioni risultati (DBI) - HOR8 - media armonica

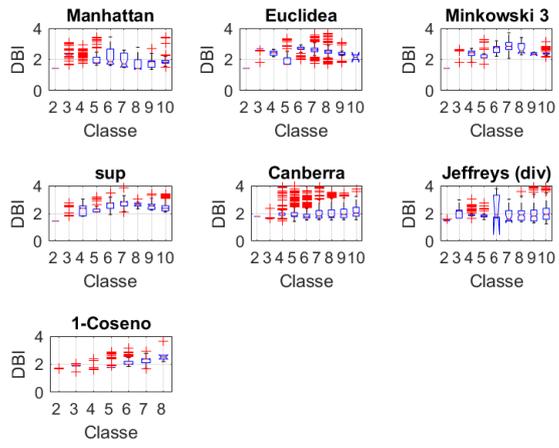


Figura 37: Distribuzioni risultati (DBI) - HOR13 - centroidi

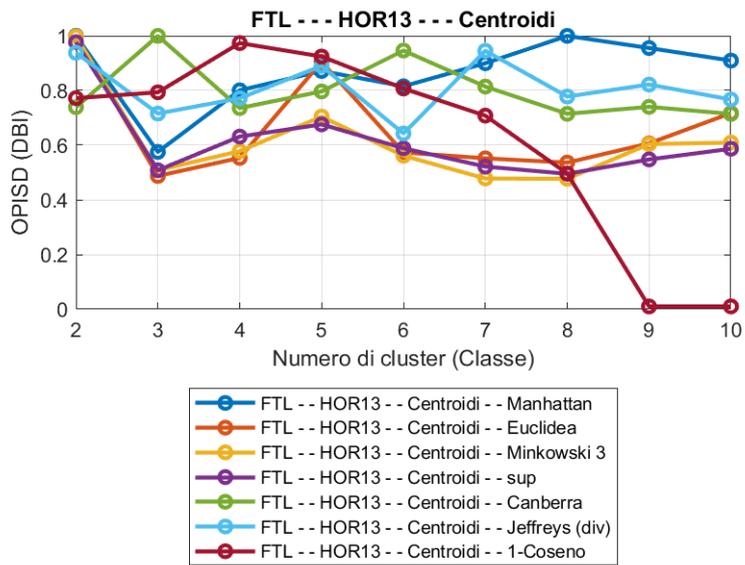


Figura 38: Distribuzioni risultati (DBI) - HOR13 - centroidi

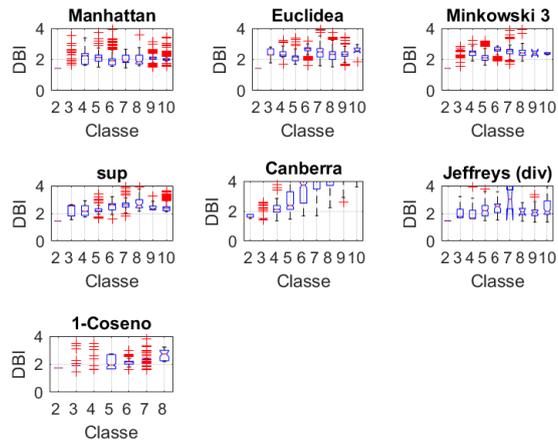


Figura 39: Distribuzioni risultati (DBI) - HOR13 - medoidi

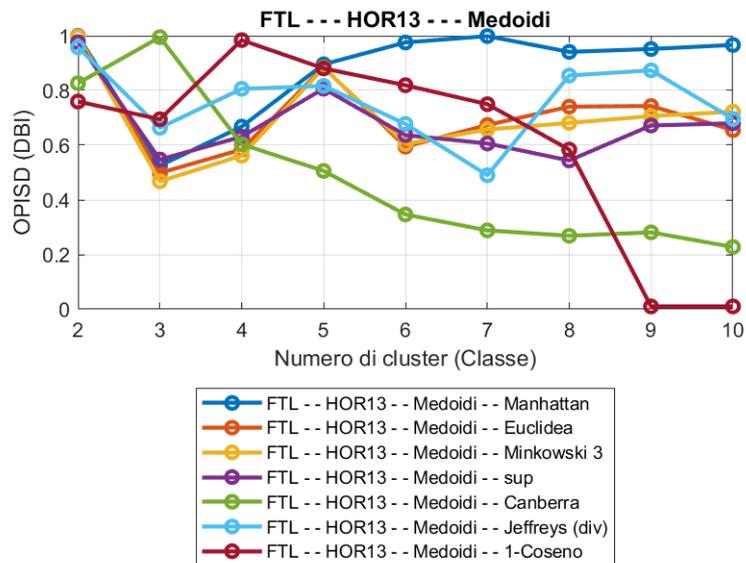


Figura 40: Distribuzioni risultati (DBI) - HOR13 - medoidi

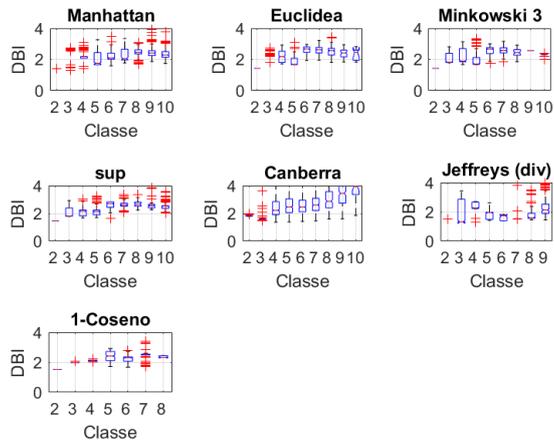


Figura 41: Distribuzioni risultati (DBI) - HOR13 - media armonica

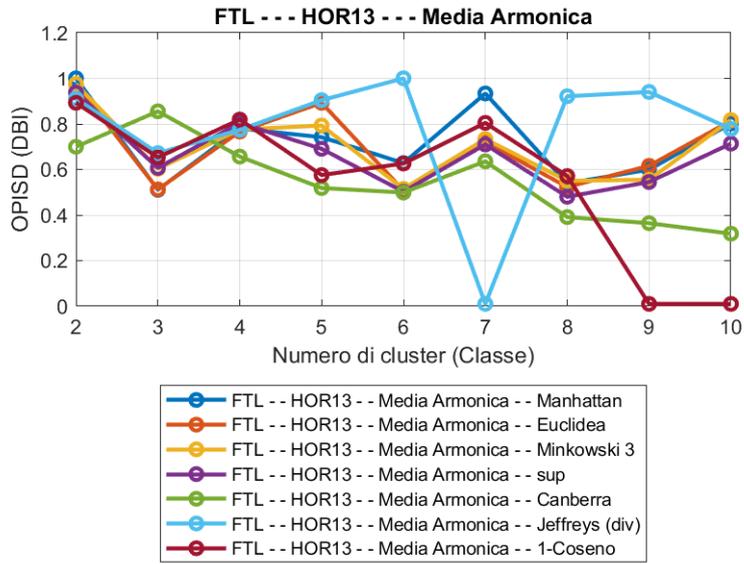


Figura 42: Distribuzioni risultati (DBI) - HOR13 - media armonica

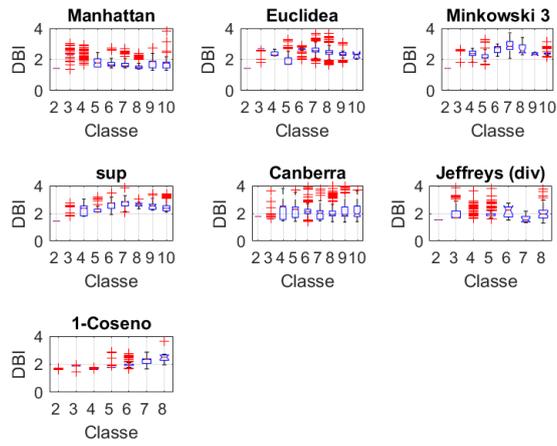


Figura 43: Distribuzioni risultati (DBI) - HOR20 - centroidi

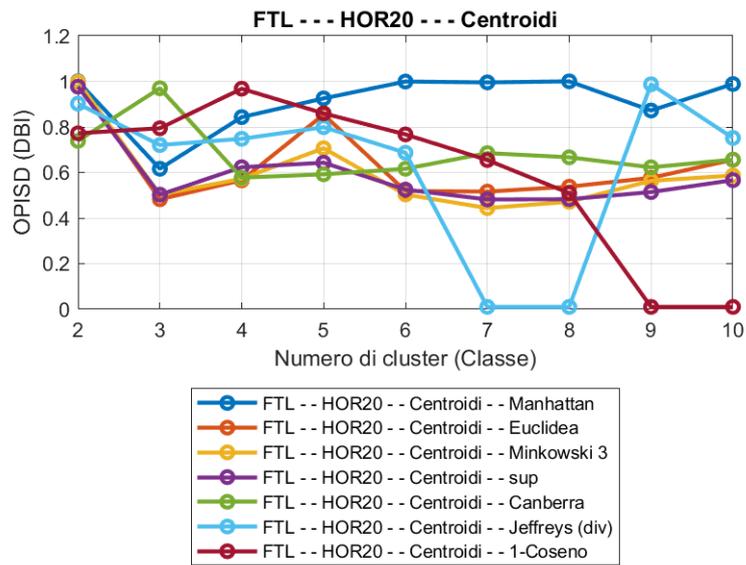


Figura 44: Distribuzioni risultati (DBI) - HOR13 - centroidi

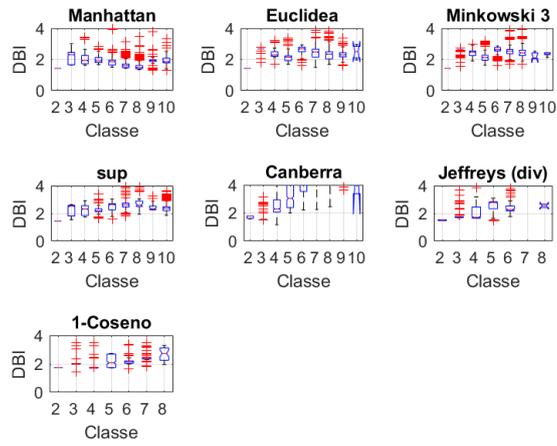


Figura 45: Distribuzioni risultati (DBI) - HOR20 - medoidi

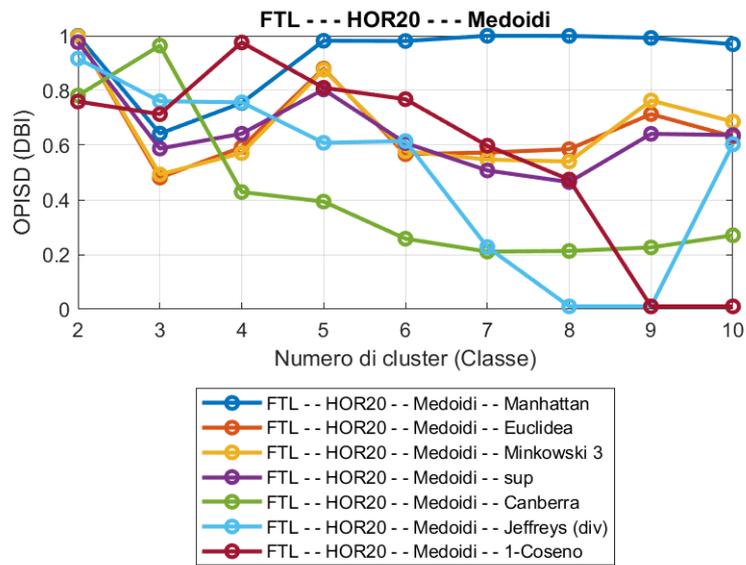


Figura 46: Distribuzioni risultati (DBI) - HOR20 - medoidi

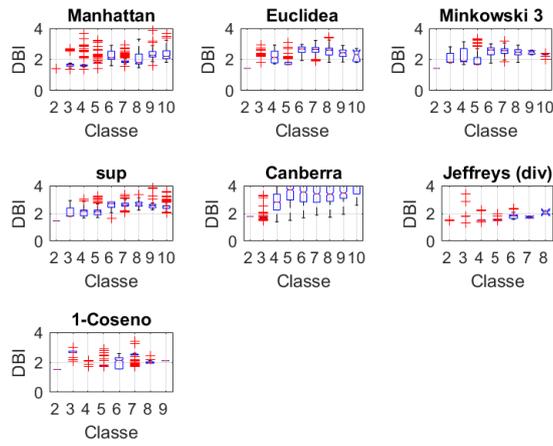


Figura 47: Distribuzioni risultati (DBI) - HOR20 - media armonica

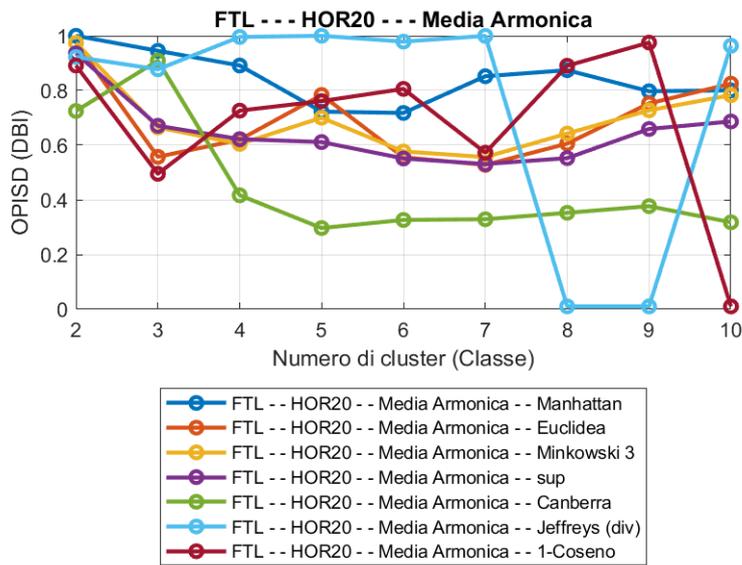


Figura 48: Distribuzioni risultati (DBI) - HOR20 - media armonica

Dai confronti interni (ordine armonico considerato e centro di attrazione), si evince come per tutte le metriche adottate, in caso di serie temporali, si ha un trend discendente dei valori medi delle distribuzioni di risultati all'aumento delle classi di soluzione considerate. Si ha un'inversione di tendenza al diminuire dell'ordine armonico (HOR) considerato nella fase di clustering (si veda il caso di HOR5, dove si hanno trend crescenti): evidentemente i metodi trovano con più facilità soluzioni nelle classi inferiori. Si hanno dei casi in cui non vi sono soluzioni per le classi superiori alla 5 (5 cluster). In questo caso, l'indicatore di OPISD viene forzato a zero (per quella classe) in modo da escluderla (per quella classe). Non necessariamente, però, si hanno soluzioni migliori. Per tutti i tipi di centri di attrazione, la divergenza di Jeffreys e il complemento a uno del coseno di similitudine presentano un elevato indice OPISD (statisticamente, dunque, portano a risultati migliori). Eseguendo un confronto totale tra tutti i metodi e tutte le proprietà considerate si ha quanto segue (sono stati estratti i migliori 7 metodi):

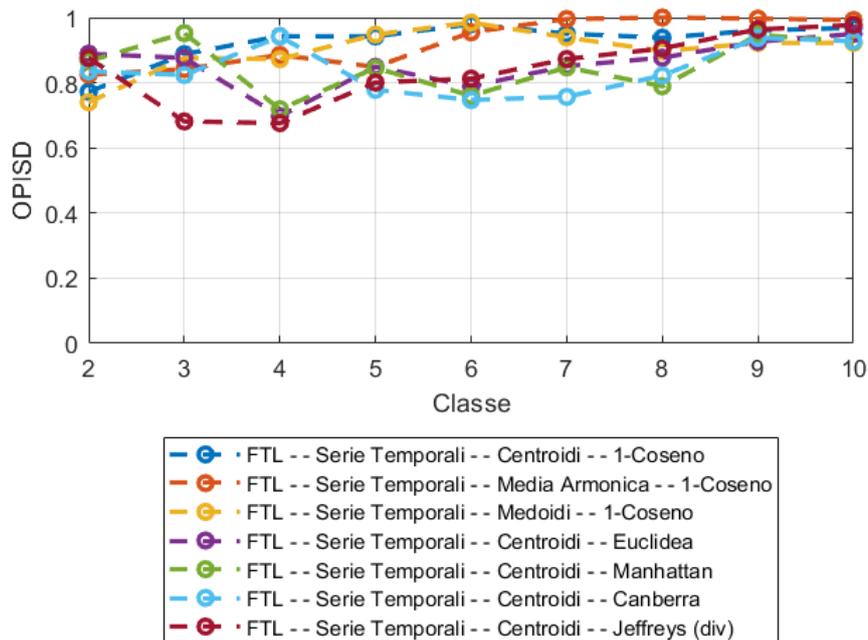


Figura 49: I migliori sette metodi basati sul FTL

Pertanto, alla luce del grafico di cui sopra, per il dataset utilizzato, si può affermare che utilizzando il Follow The Leader, considerando la media ar-

monica come centro di attrazione e la divergenza di Kullback–Leibler resa simmetrica (divergenza di Jeffreys) o il complemento a uno del coseno di similitudine, si ha una buona probabilità di avere un buon risultato (tali metodi portano ad una distribuzione di risultati vicini alla curva di riferimento).

Confrontando i migliori risultati dei metodi che presentano OPISD maggiore per le classi di risultato inferiore, si ha quanto segue:

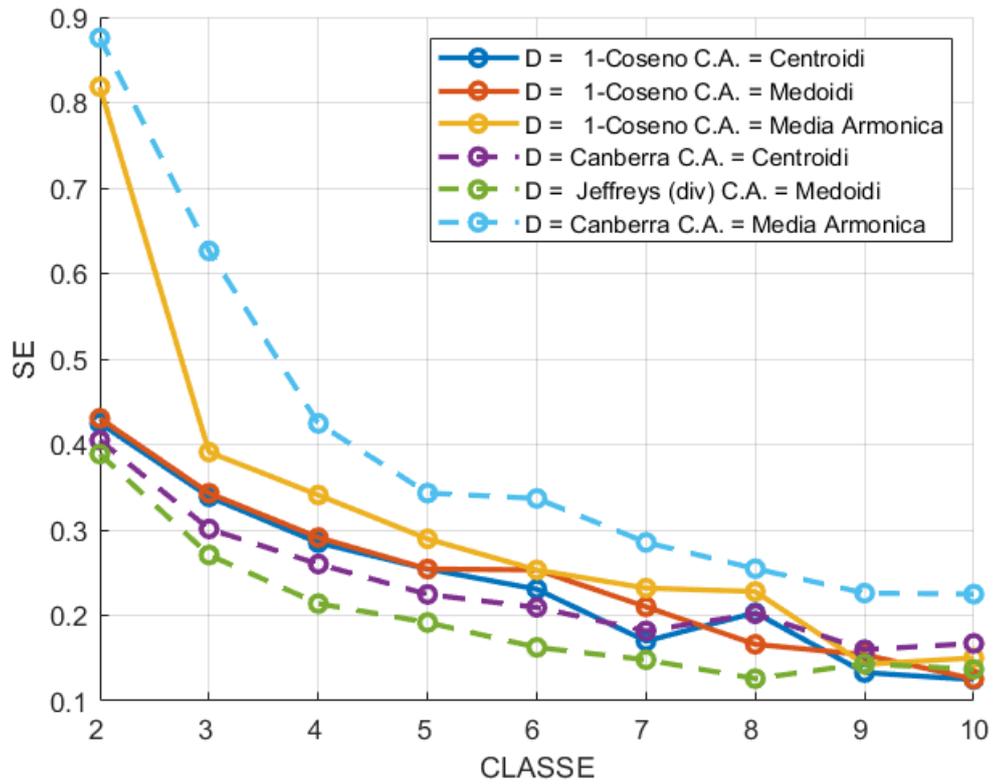


Figura 50: Linea continua: Serie temporali - Linea tratteggiata: HOR5

Il grafico riporta l'andamento dell'errore di Sammon "minore" ottenuto dai metodi elencati (basati sul FTL). Si evince come considerando le prime 5 armoniche (linee tratteggiate) si hanno, tendenzialmente, migliori risultati per basse classi.

## 11 Kmeans

Nel Kmeans, il numero di cluster  $N_c$  rappresenta un parametro fornito dall'utente a monte. Vengono estratti  $N_c$  elementi, senza ripetizioni, dal database da suddividere (eletti a centri di attrazione provvisori) intorno ai quali avviene la suddivisione in cluster (il criterio di associazione è la minima distanza dai baricentri). A seguire, l'iterazione di ottimizzazione calcolerà i nuovi centroidi intorno ai quali avviene la suddivisione. Si avrà convergenza quando non vi sarà variazione tra due iterazioni successive.

Si può manifestare la formazione di cluster vuoti (a cui non viene associato nessun elemento). In questo caso la soluzione va declassata, scartando tali cluster e aggiornando il numero di cluster della soluzione. Per tale motivo, si necessita di una classificazione della soluzione a valle dell'elaborazione principale (anche se la classe viene impostata a priori).

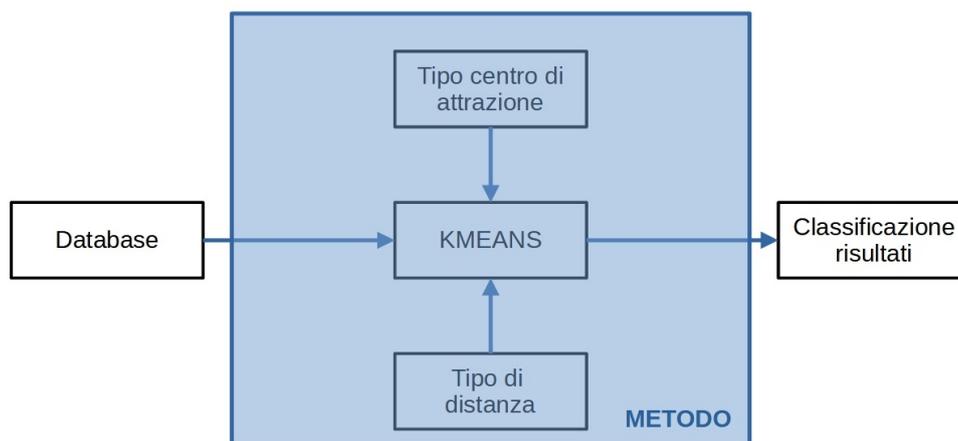


Figura 51: Metodo basato sul Kmeans

A rigore (come si evince dal nome dell'algoritmo) il centro di attrazione dovrebbe essere il centroide (media) ma, anche se in questa sede se ne analizzeranno delle varianti (basate sui medoidi e sulle medie armoniche) mantenendo comunque la stessa nomenclatura.

## 11.1 Risultati Kmeans

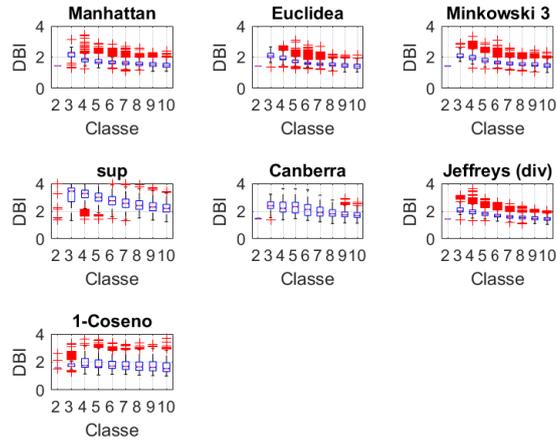


Figura 52: Distribuzioni risultati (DBI) - Serie temporali - centroidi

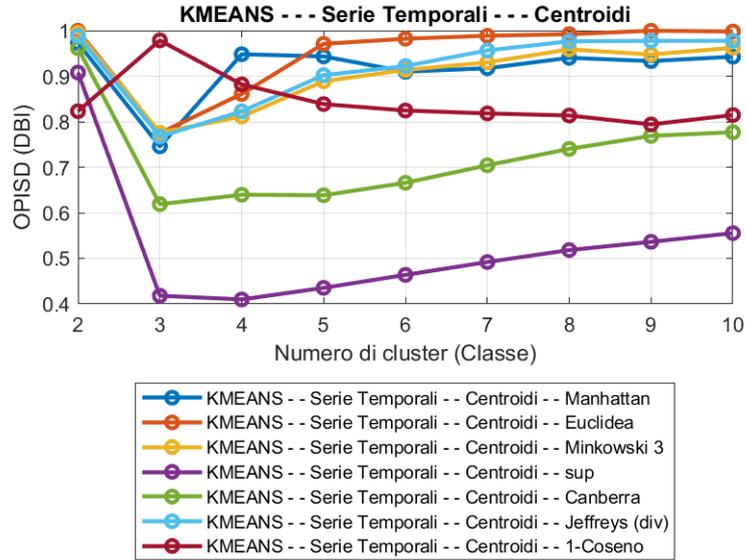


Figura 53: Distribuzioni risultati (DBI) - Serie temporali - centroidi

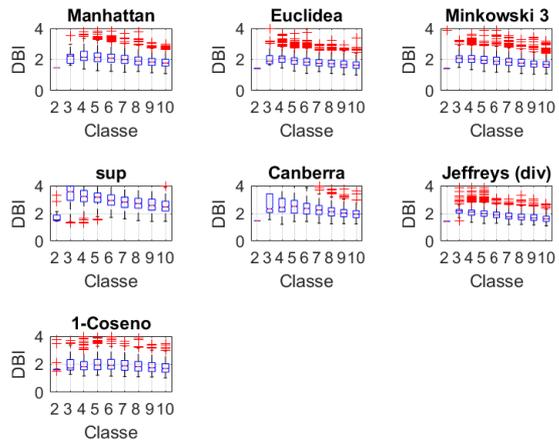


Figura 54: Distribuzioni risultati (DBI) - Serie temporali - medoidi

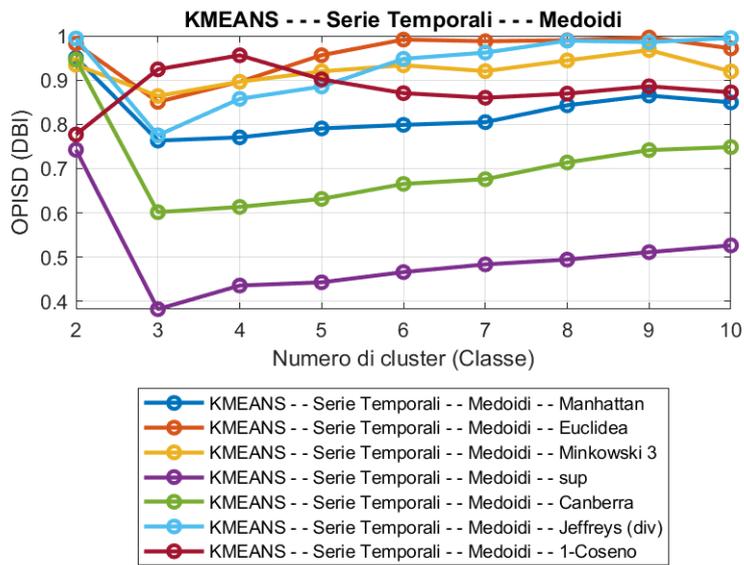


Figura 55: Distribuzioni risultati (DBI) - Serie temporali - medoidi

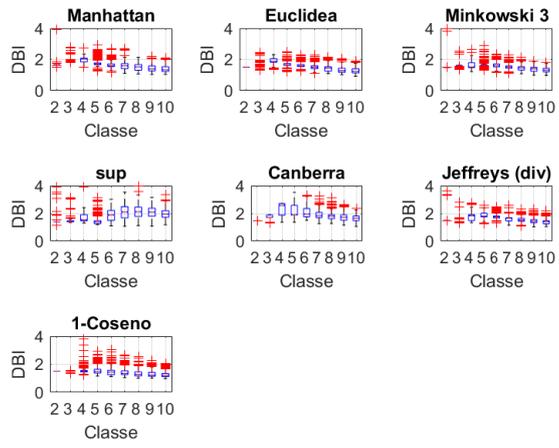


Figura 56: Distribuzioni risultati (DBI) - Serie temporali - media armonica

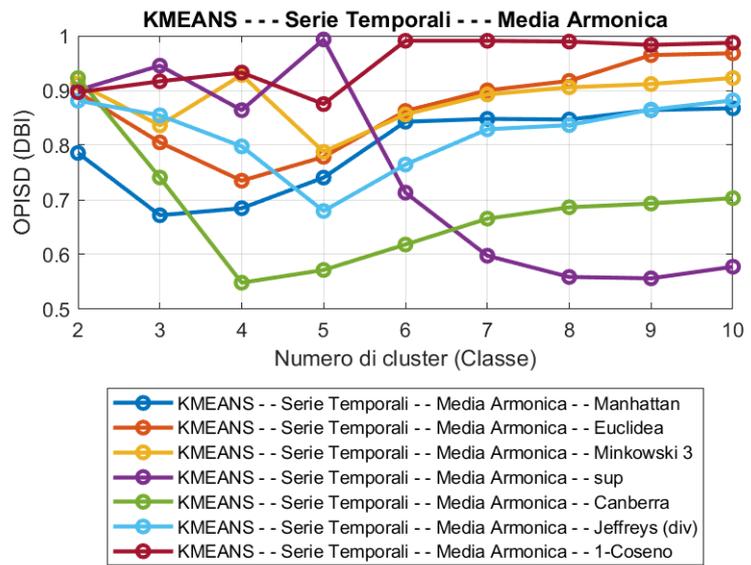


Figura 57: Distribuzioni risultati (DBI) - Serie temporali - media armonica

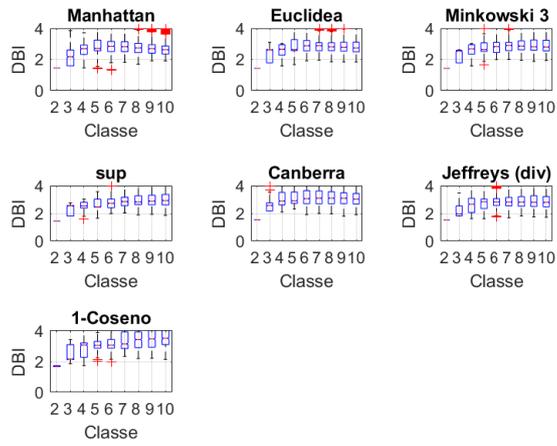


Figura 58: Distribuzioni risultati (DBI) - HOR5 - centroidi

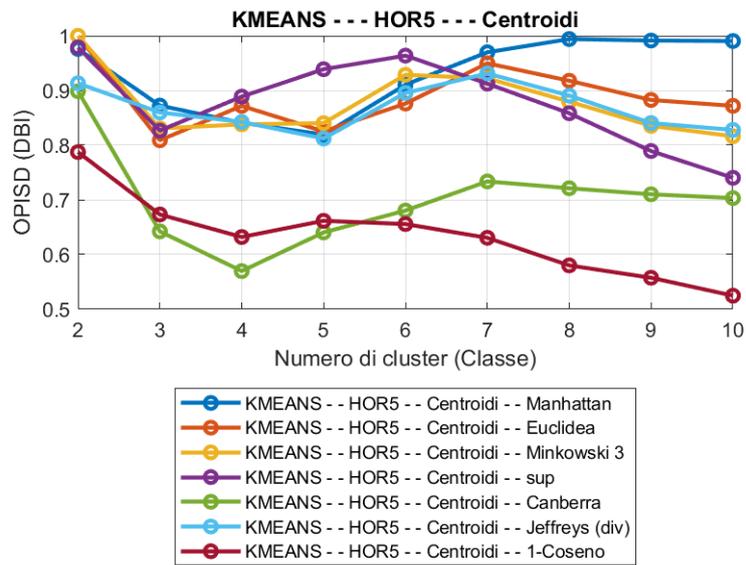


Figura 59: Distribuzioni risultati (DBI) - HOR5 - centroidi

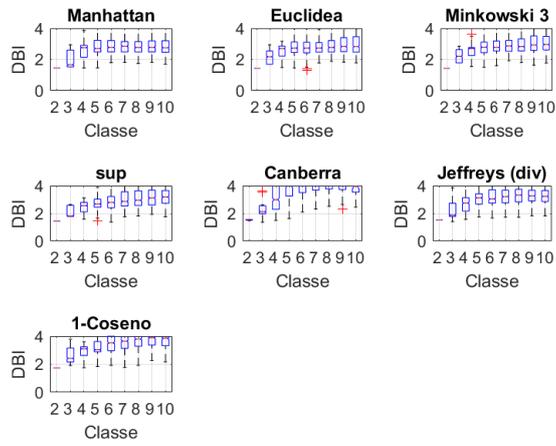


Figura 60: Distribuzioni risultati (DBI) - HOR5 - medoidi

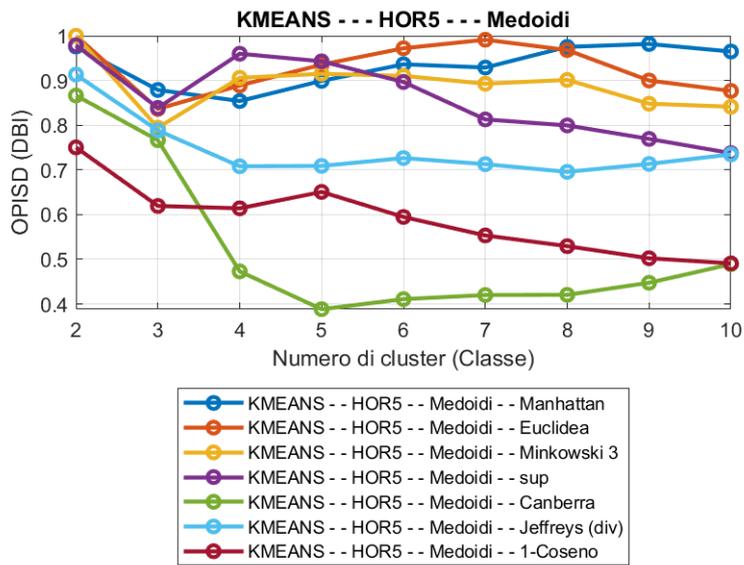


Figura 61: Distribuzioni risultati (DBI) - HOR5 - medoidi

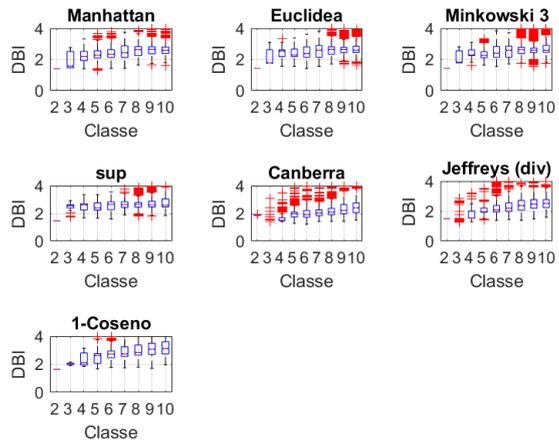


Figura 62: Distribuzioni risultati (DBI) - HOR5 - media armonica

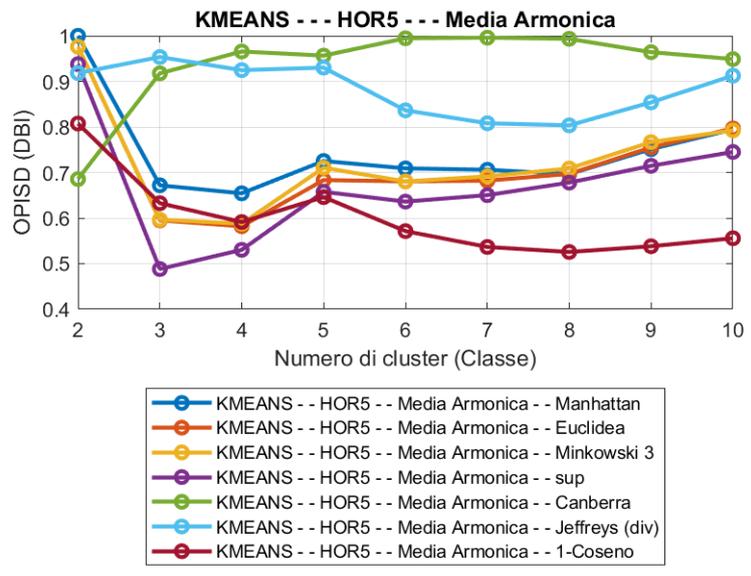


Figura 63: Distribuzioni risultati (DBI) - HOR5 - media armonica

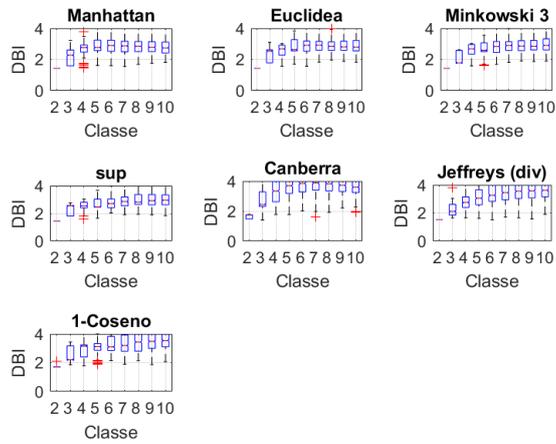


Figura 64: Distribuzioni risultati (DBI) - HOR8 - centroidi

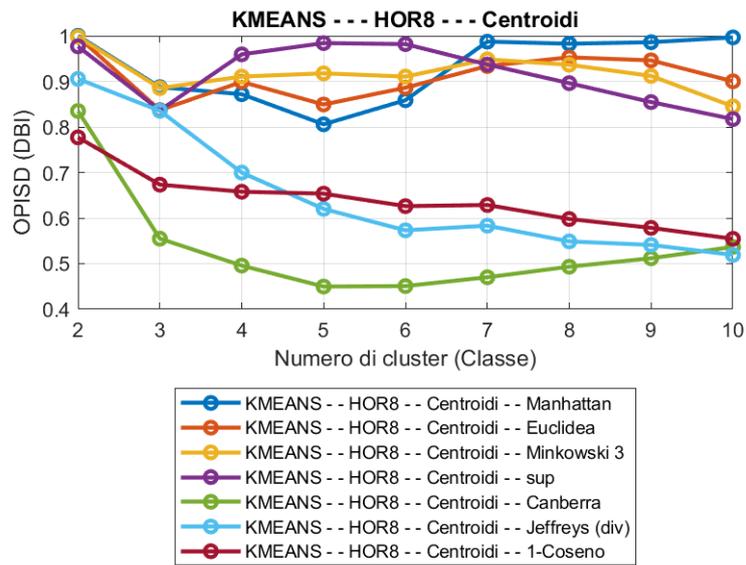


Figura 65: Distribuzioni risultati (DBI) - HOR8 - centroidi

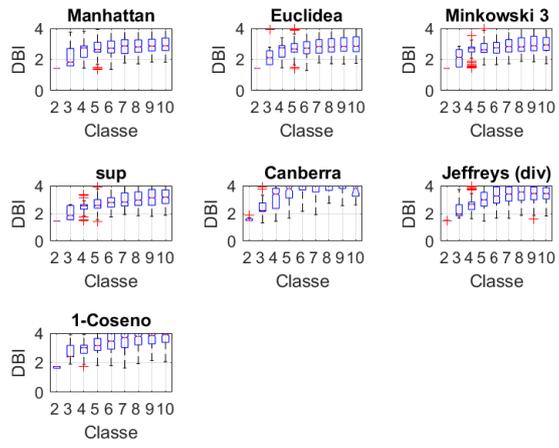


Figura 66: Distribuzioni risultati (DBI) - HOR8 - medoidi

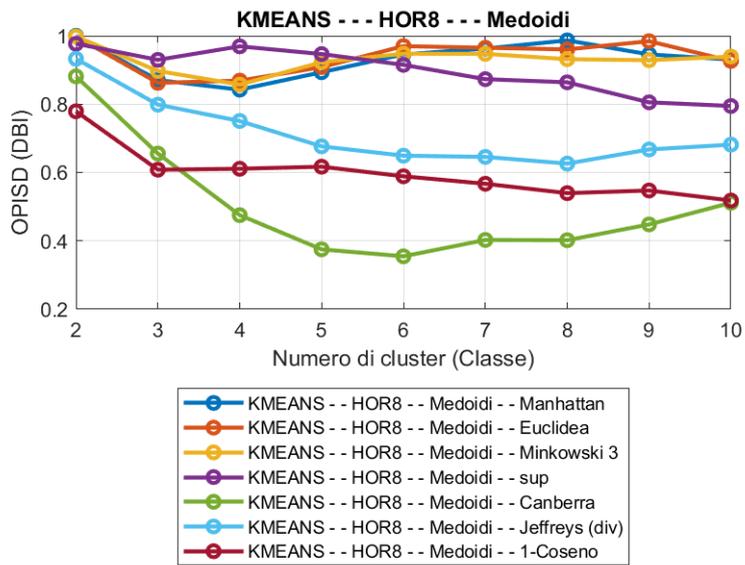


Figura 67: Distribuzioni risultati (DBI) - HOR8 - medoidi

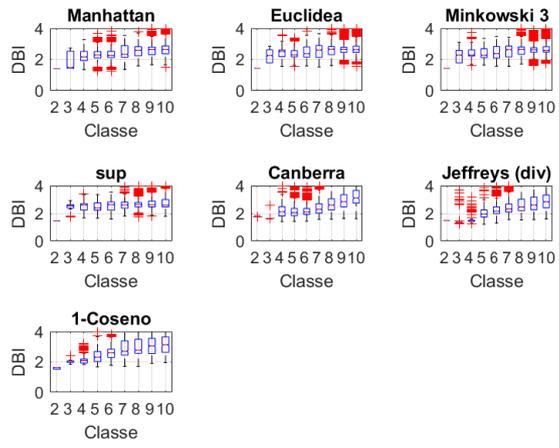


Figura 68: Distribuzioni risultati (DBI) - HOR8 - media armonica

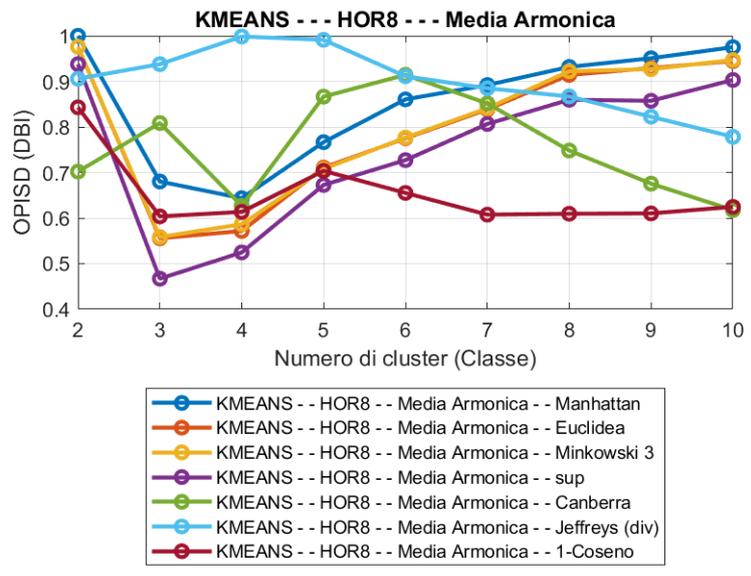


Figura 69: Distribuzioni risultati (DBI) - HOR8 - media armonica

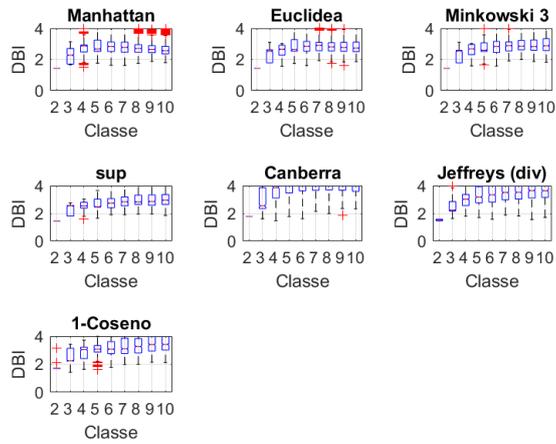


Figura 70: Distribuzioni risultati (DBI) - HOR13 - centroidi

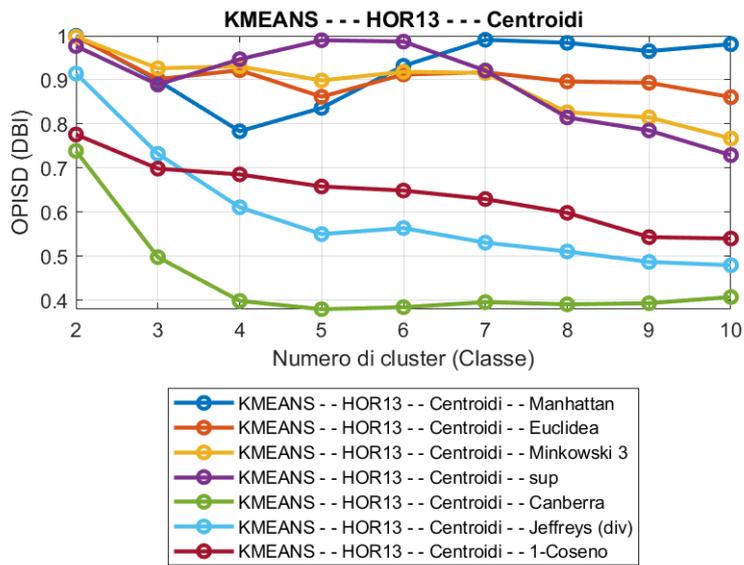


Figura 71: Distribuzioni risultati (DBI) - HOR13 - centroidi

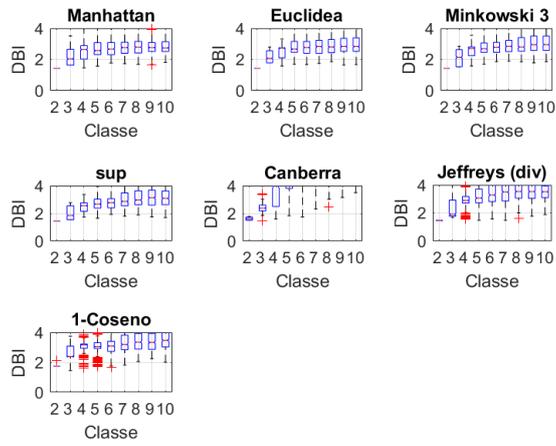


Figura 72: Distribuzioni risultati (DBI) - HOR13 - medoidi

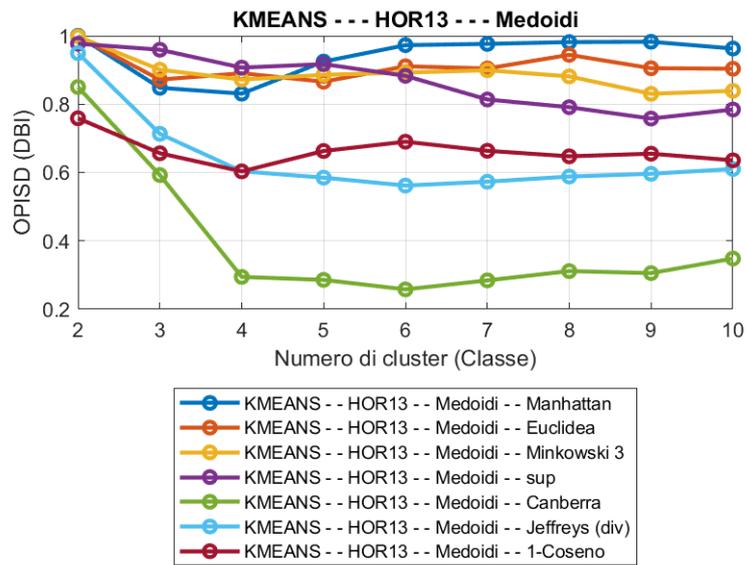


Figura 73: Distribuzioni risultati (DBI) - HOR13 - medoidi

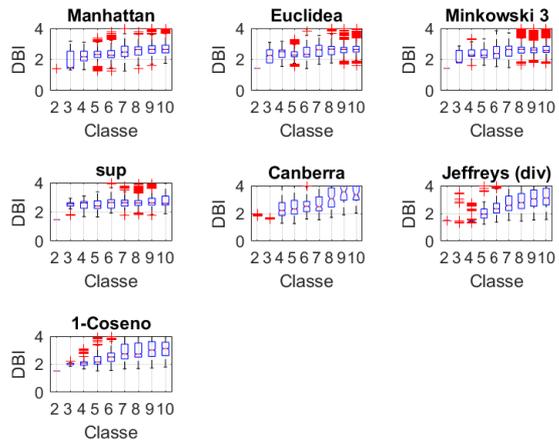


Figura 74: Distribuzioni risultati (DBI) - HOR13 - media armonica

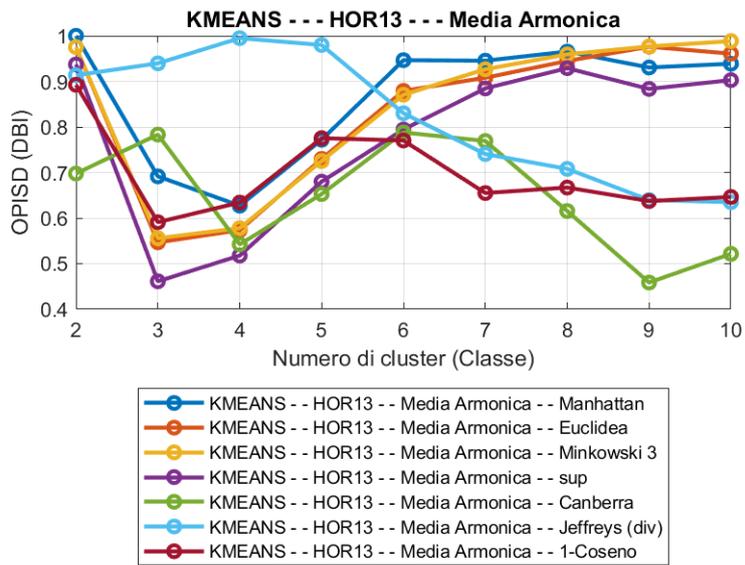


Figura 75: Distribuzioni risultati (DBI) - HOR13 - media armonica

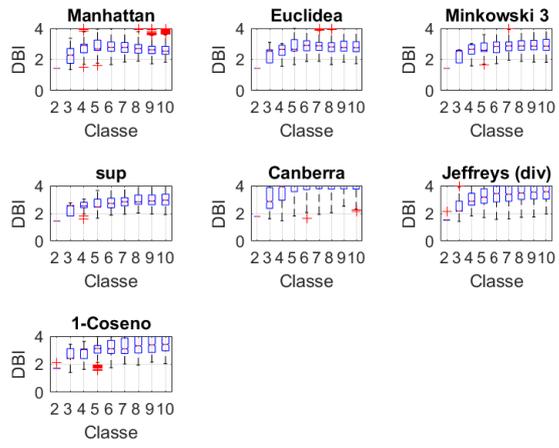


Figura 76: Distribuzioni risultati (DBI) - HOR15 - centroidi

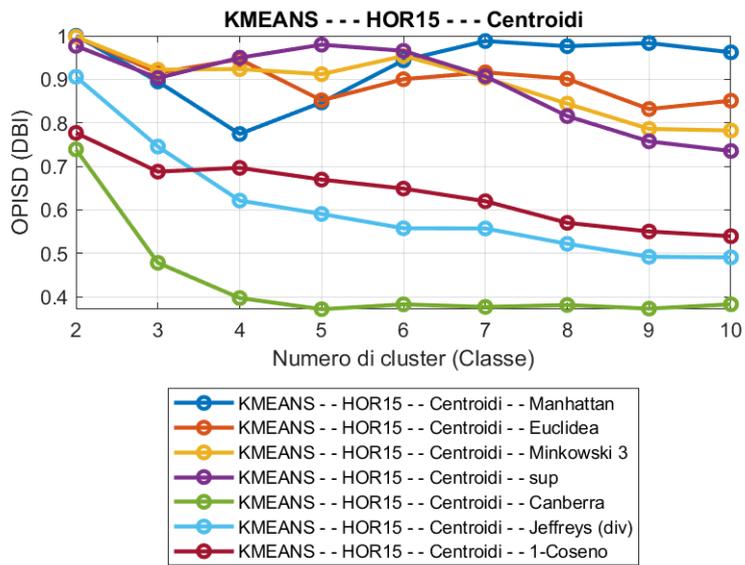


Figura 77: Distribuzioni risultati (DBI) - HOR15 - centroidi

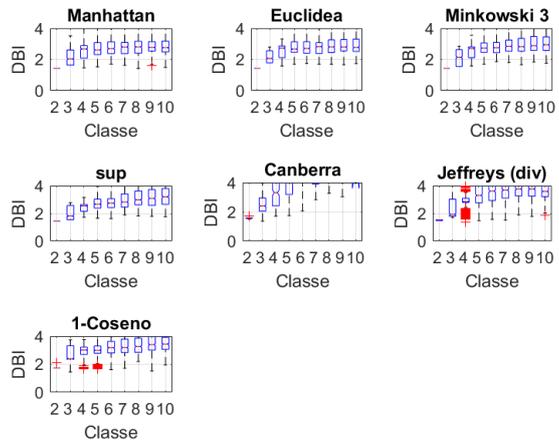


Figura 78: Distribuzioni risultati (DBI) - HOR15 - medoidi

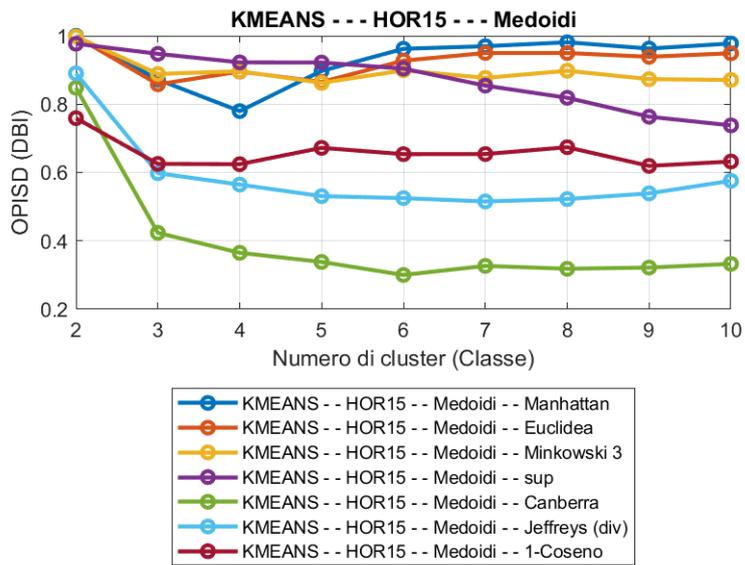


Figura 79: Distribuzioni risultati (DBI) - HOR15 - medoidi

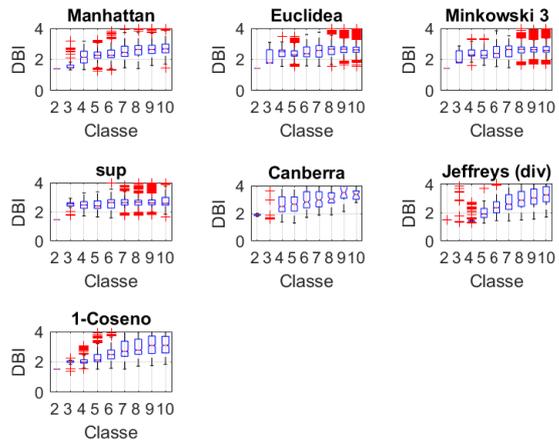


Figura 80: Distribuzioni risultati (DBI) - HOR15 - media armonica

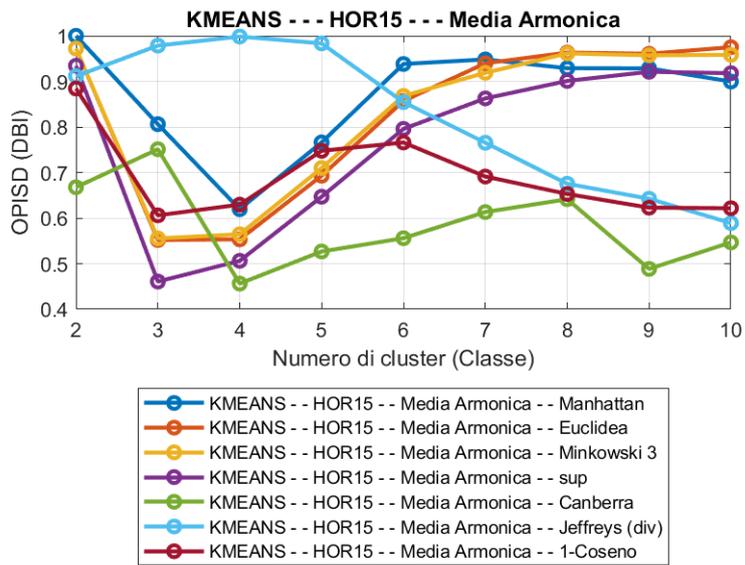


Figura 81: Distribuzioni risultati (DBI) - HOR15 - media armonica

A differenza dei metodi basati sul Follow The Leader, si osserva che l'indicatore OPISD tende ad essere più regolare al variare della classe di risultato, anche se (tra i vari metodi) esso raggiunge valori inferiori. Questo si traduce nel fatto che, il kmeans, produce distribuzioni più distanti rispetto al FTL. Lo si nota, infatti, dai vari box plot (distribuzione di soluzioni). Il coseno di similitudine risulta meno prestante, al contrario delle norma Minkowski (con esponente 1,2 e 3) che sembra portare a risultati migliori (tra i metodi basati sul kmeans).

Tale metrica, sembra portare risultati migliori anche nel dominio della frequenza.

Anche in questo caso, si esegue un confronto interno individuando i metodi migliori in termini di OPISD (confronto interno kmeans):

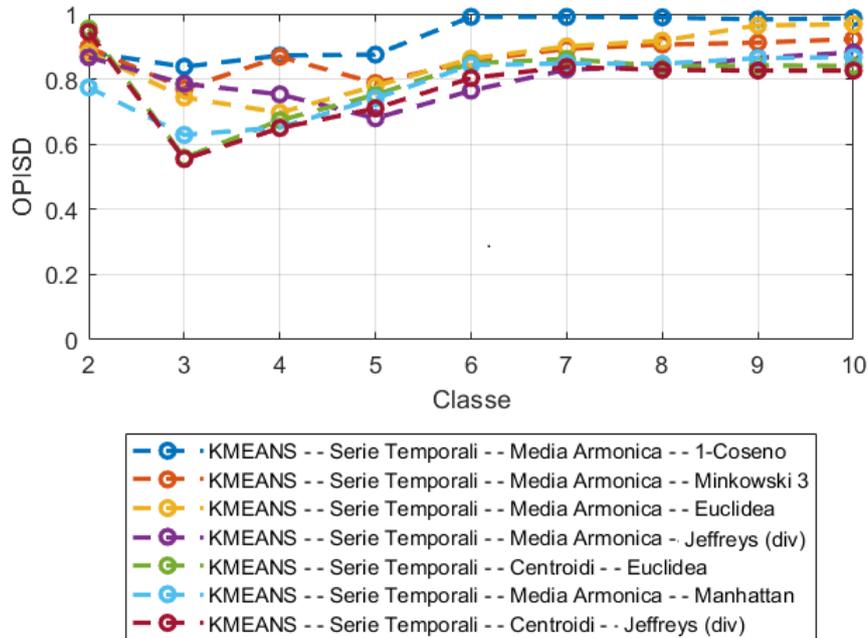


Figura 82: I migliori sette metodi basati sul KMEANS

Tra i metodi basati sul kmeans, la media armonica degli elementi come centro di attrazione e il coseno di similitudine come metrica, portano ad avere migliori distribuzioni di risultati.

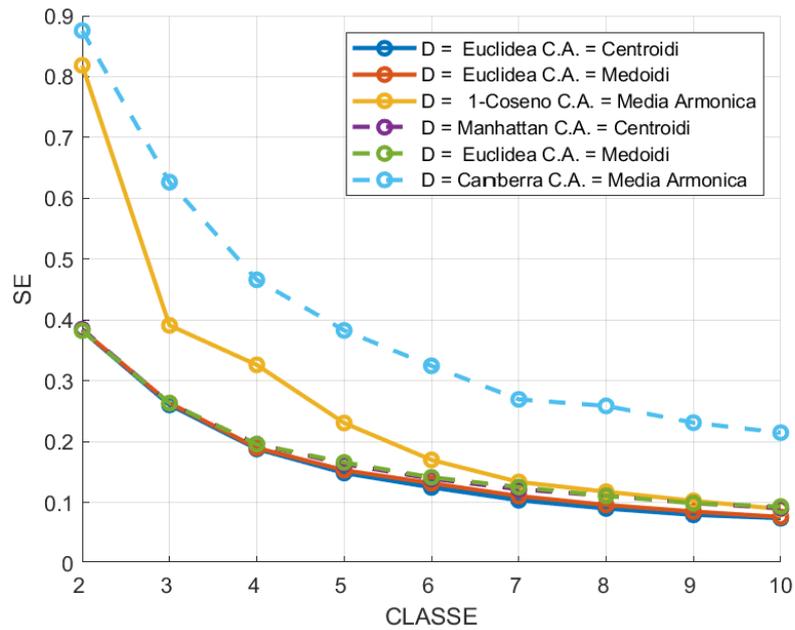


Figura 83: Linea continua: Serie temporali - Linea tratteggiata: HOR5

## 12 Confronto tra FTL e KMEANS

Ottenute le cumulative di soluzioni in uscita da ogni metodo (sia per il FTL che per il Kmeans), possono essere confrontate in termini globali:

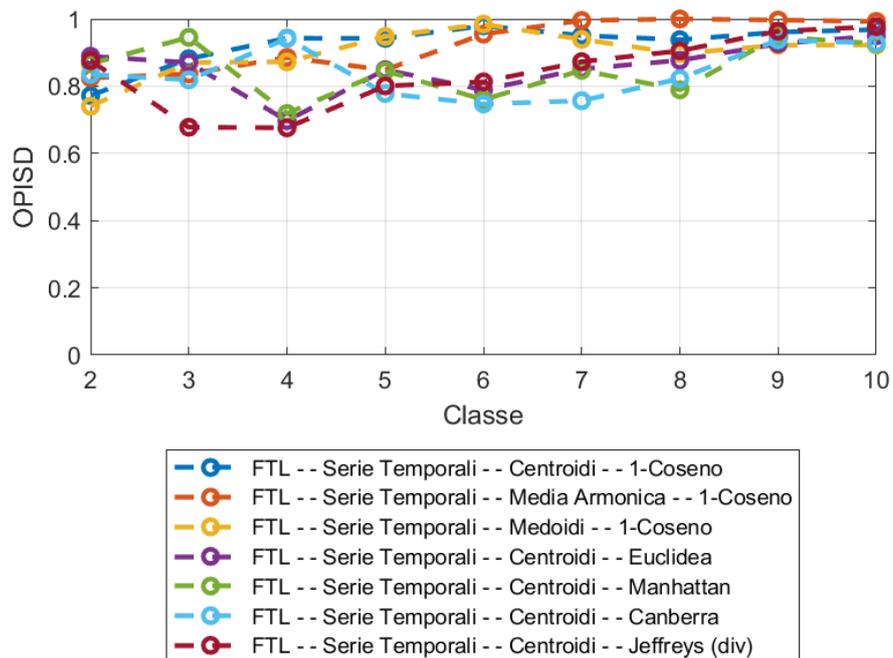


Figura 84: Migliori sette metodi tra tutti quelli analizzati

I metodi basati sul Follow The Leader, considerando i centroidi come centri di attrazione, portano a risultati i cui distribuzioni sono più vicine alla distribuzione di riferimento (le cumulative sono vicine a quella di riferimento).

## 13 Conclusioni

I risultati ottenuti possono essere riassunti come segue:

- Lo studio effettuato mostra che, partendo da una serie temporale di 96 elementi (pattern di carico quartorari non residenziali), i metodi migliori (confrontati grazie al concetto di dominanza stocastica) sono quelli basati sul metodo Follow The Leader.
- L'utilizzo del complemento a uno del coseno di similitudine nell'individuare la similarità tra elementi, porta a prestazioni prevalenti (rispetto alle altre metriche) indipendentemente dal tipo di centro di attrazione considerato.
- Tra i tre tipi di centri di attrazione considerati, la media armonica e i centroidi tendono a migliorare le prestazioni.
- I metodi basati sul kmeans, di contro, presentano una maggiore stabilità, ovvero i metodi presentano prestazioni analoghe al variare della classe delle soluzioni.
- L'utilizzo di un numero ridotto di armoniche (ordinate secondo l'HOR) porta ad una riduzione di sensibilità degli algoritmi (metodi) considerati ai rumori. Questo si traduce in un maggiore sensibilità rispetto al tipo di classificazione desiderato: si vuole un numero estremamente ridotto di "rappresentanti".
- Raggruppare gli elementi riducendone il contenuto armonico fa sì che la numerosità delle soluzioni delle soluzioni di classe inferiore aumenti, senza però migliorare la qualità del raggruppamento in termini statistici (il singolo risultato potrebbe essere migliore).
- Passando ad una logica elitista, ovvero prelevando la soluzione migliore, sembrerebbe che sia sufficiente considerare le prime 5 armoniche (ordinate secondo l'HOR) per ottenere buoni risultati.

## 14 Bibliografia

I grafici e gli schemi, come ogni script, presenti nel presente manoscritto sono frutto del lavoro dell'autore.

Le fonti possono essere rintracciate in:

- Documentazione Python (<https://www.python.org/>);
- Documentazione Matlab (<https://it.mathworks.com/help/index.html>);

## Riferimenti bibliografici

- [1] Gianfranco Chicco. Overview and performance assessment of the clustering methods for electrical load pattern grouping. pages 68–80, 1999.
- [2] R.Napoli M.Scutariu E.Carpaneto, G.Chicco. Electricity customer classification using frequency-domain load pattern data. *Electrical Power and Energy Systems 28 (2006)*, pages 13–20, 2005.
- [3] J. Sumaili Akilimali G. Chicco. Renyi entropy-based classification of daily electrical load patterns. *IET Gener. Transm. Distrib.*, pages 736–745, 2010.
- [4] Andrea Mazza Gianfranco Chicco. Assessment of optimal distribution network reconfiguration results using stochastic dominance nts. *Sustainable Energy, Grids and Networks 9 (2017)*, pages 75–79, 2016.