

**POLITECNICO DI MILANO**

Facoltà di Ingegneria dei Sistemi

Corso di Laurea in  
Ingegneria Gestionale



Monitoraggio di Profili Multiscala basato su Empirical Mode  
Decomposition

Relatore: Chiar.ma Prof.ssa Bianca Maria COLOSIMO

Co-relatori: Ing. Marco GRASSO  
Ing. Matteo Francesco MAGGIONI

Tesi di Laurea Magistrale di:

Andrea CARROCCIA Matr. 801837

Anno Accademico 2013 – 2014

## Abstract

Nell'ambito delle lavorazioni meccaniche, si osserva un crescente interesse per il monitoraggio della qualità del processo basato su segnali acquisiti da uno o più sensori durante la lavorazione stessa. I metodi più usati per il monitoraggio in-process includono l'analisi di indicatori sintetici nel dominio temporale, e l'analisi di Fourier in quello delle frequenze. In molti casi applicativi, tuttavia, è necessario monitorare segnali complessi, caratterizzati da contenuti *multiscala*, i.e., segnali che presentano diversi livelli tempo-frequenza, alcuni dei quali localizzati nel tempo. In questi casi le tecniche tradizionali non permettono una corretta caratterizzazione del contenuto informativo ed è necessario studiare metodi alternativi.

L'analisi nel dominio tempo-frequenza permette di distinguere i diversi modi oscillatori di un segnale e di migliorare la capacità di individuare dei guasti nel sistema. Tra questi metodi, quello preponderante in letteratura è la *Wavelet Analysis*, i cui risultati tuttavia dipendono dalla scelta della Wavelet madre e dal numero di livelli di decomposizione selezionati. La *Empirical Mode Decomposition (EMD)* è un metodo alternativo che prescinde dalla definizione di una funzione di base. Questa tecnica adattiva e non-parametrica permette di decomporre il segnale in un numero finito e limitato di sotto-segnali detti *Intrinsic Mode Functions* in modo completamente data-driven.

Nel presente lavoro, la EMD viene utilizzata con lo scopo di estrarre le diverse features di segnali ciclici, i.e. profili, con contenuto multiscala. Il metodo delle Combined Mode Functions (CMFs) ha l'obiettivo di ottenere una decomposizione sintetica attraverso un numero limitato di modi distinti, rappresentati dalle cosiddette CMFs. In particolare, nel presente studio, si propone un metodo innovativo e automatico per la separazione delle CMFs che si basa sull'analisi della distribuzione dei dati in ciascuna IMF. Viene inoltre studiata la possibilità di selezionare in modo automatico il numero di CMFs, e si propone un'analisi comparativa di diversi metodi per tale scopo. In questo modo si rende possibile l'utilizzo della EMD come strumento di pre-processing di segnali complessi in applicazioni di monitoraggio on-line di profili.

Questa tecnica si coniuga con la *Principal Component Analysis (PCA)*, che viene applicata allo scopo di ridurre il numero di variabili per ogni CMF. Il sistema di monitoraggio proposto ha l'obiettivo di segnalare deviazioni di forma in una o più scale del segnale che non possono essere spiegate dalla variabilità naturale del processo.

Le prestazioni del metodo proposto sono analizzate sia attraverso segnali simulati, sia utilizzando dati reali acquisiti in un processo di taglio waterjet. I risultati dimostrano che il metodo proposto permette di ottenere una separazione automatica delle CMFs più efficace rispetto ai metodi tradizionalmente proposti in letteratura. Inoltre, si evidenziano i vantaggi e gli svantaggi di diverse tecniche per la scelta automatica del numero di livelli di decomposizione, e la criticità di una scelta corretta rispetto alle prestazioni di monitoraggio.

# Indice Generale

<b>Abstract</b>	<b>2</b>
<b>Indice Generale</b>	<b>3</b>
<b>Elenco delle Figure e delle Tabelle</b>	<b>5</b>
<b>Capitolo 1 Introduzione</b>	<b>7</b>
<b>Capitolo 2 Il caso di studio: profili di pressione WaterJet</b>	<b>13</b>
<b>Capitolo 3 La Empirical Mode Decomposition (EMD)</b>	<b>18</b>
<i>3.1 La Empirical Mode Decomposition</i>	<i>18</i>
<i>3.2 Un esempio numerico illustrativo</i>	<i>22</i>
<i>3.3 Applicazioni della EMD</i>	<i>23</i>
<i>3.4 Vantaggi e svantaggi della EMD</i>	<i>25</i>
<i>3.5 La Ensemble Empirical Mode Decomposition</i>	<i>28</i>
<i>3.6 La Complete Ensemble EMD With Adaptive Noise</i>	<i>32</i>
<i>3.7 Il metodo delle Combined Mode Functions</i>	<i>36</i>
<b>Capitolo 4 Il metodo proposto</b>	<b>41</b>
<i>4.1 Il calcolo delle CMFs</i>	<i>43</i>
<i>4.2 Determinare il numero ottimale K di CMFs</i>	<i>45</i>
<i>4.2.1 Internal Validity Indexes</i>	<i>46</i>
<i>4.2.2 Sum-of-squares-based indexes</i>	<i>49</i>

<b>Capitolo 5 Analisi dei risultati: simulation study</b>	<b>52</b>
<i>5.1 Il calcolo delle CMFs</i>	52
<i>5.2 Scelta del K ottimale</i>	58
<b>Capitolo 6 Analisi dei risultati: real case study</b>	<b>64</b>
<i>6.1 Inefficacia della EEMD e della CEEMDAN</i>	64
<i>6.2 Il calcolo delle CMFs</i>	69
<i>6.3 Scelta del K ottimale</i>	74
<b>Capitolo 7 Combinazione di EMD e PCA per profile monitoring</b>	<b>75</b>
<i>7.1 La Principal Component Analysis (PCA)</i>	75
<i>7.2 Le carte di controllo</i>	78
<i>7.3 Determinare il numero <math>m_k</math> di Componenti Principali</i>	81
<i>7.4 Tasso di fault detection del sistema di monitoraggio</i>	81
<i>7.5 Confronto con altri metodi</i>	85
<b>Capitolo 8 Conclusioni</b>	<b>88</b>
<b>Bibliografia</b>	<b>90</b>

## Elenco delle Figure e delle Tabelle

Figura 2.1 Schema di funzionamento della testa di taglio di una macchina AWJ .....	15
Figura 2.2 Profili di pressione WJ in controllo .....	16
Figura 2.3 Profilo di pressione WJ IC con transitori .....	16
Figura 2.4 Profili WJ per i quattro scenari di guasto .....	17
Figura 3.1 La Empirical Mode Decomposition .....	21
Figura 3.2 EMD del segnale $x(t)$ in Equazione 3.12 .....	23
Tabella 3.1 Riferimenti bibliografici ai casi di applicazione della EMD .....	24
Figura 3.3 IMFs per il segnale $x(t)$ in Equazione (3.13) .....	27
Figura 3.4 La Ensemble Empirical Mode Decomposition (EEMD) .....	29
Figura 3.5 EEMD del segnale in Equazione (3.13) .....	31
Figura 3.6 CEEMDAN del segnale in Equazione (3.13) .....	35
Figura 3.7 Numero di iterazioni eseguite per il calcolo delle IMFs rispettivamente dalla EEMD (a sinistra) e CEEMDAN (a destra) .....	36
Figura 3.8 Combined Mode Functions per il segnale in Equazione (3.22) .....	38
Figura 4.1 Andamento dei principali indicatori per le IMFs del profilo WJ IC .....	42
Figura 5.1 Insieme dei segnali del dataset simulato .....	52
Figura 5.2 Uno dei segnali simulati $x(t)$ e le sue componenti in frequenza .....	53
Figura 5.3 IMFs di un segnale simulato $x(t)$ .....	54
Figura 5.4 PDFs delle IMFs del segnale $x(t)$ .....	55
Figura 5.5 Distanze tra PDFs contigue per il segnale $x(t)$ .....	56
Figura 5.6 CMFs del segnale simulato $x(t)$ con $K=3$ .....	57
Figura 5.7 Indice di Dunn medio per il dataset simulato .....	59
Figura 5.8 Indice di Dunn per ogni profilo del dataset simulato .....	59
Figura 5.9 Distribuzione del $K$ ottimale secondo l'indice di Dunn .....	60
Figura 5.10 SSW medio per i segnali del dataset simulato .....	61
Figura 5.11 SSW per i segnali del dataset simulato .....	61
Figura 5.12 Indice di Hartigan per i segnali del dataset simulato .....	62
Figura 5.13 Distribuzione del $K$ ottimale secondo l'indice di Hartigan .....	62
Tabella 5.1 Distribuzione del $K$ ottimale per il dataset di test secondo i diversi indici .....	63
Figura 6.1 EMD del profilo WJ IC in Figura 2.3 .....	66
Tabella 6.1 Risultati della EMD sui profili WJ IC .....	67
Tabella 6.2 Risultati della EEMD sui profili WJ IC .....	68
Tabella 6.3 Risultati della CEEMDAN sui profili WJ IC .....	69

Figura 6.2 PDFs delle IMFs del profilo WJ IC di Figura 2.3 .....	71
Figura 6.3 Distanza tra coppie di PDFs contigue per il profilo WJ IC n.5.....	72
Figura 6.4 CMFs per il profilo WJ IC n.5 con K=3 .....	73
Figura 6.5 Fasi del ciclo dei tre gruppi cilindro-pistone della pompa intensificatrice.	73
Tabella 6.4 Distribuzione del K ottimale per i profili WJ IC secondo i diversi indici.	74
Figura 7.1 Schema di monitoraggio EMD-PCA proposto .....	80
Figura 7.2 Esempio di Fase di Training – diagrammi di Pareto.....	83
Figura 7.3 Esempio di Fase di Tuning con limiti teorici (in verde) e limiti empirici (in rosso) .....	84
Tabella 7.1 Tasso di Fault Detection medio per il metodo EMD-PCA proposto.....	84
Tabella 7.2 Tasso di Fault Detection medio con separazione delle IMFs basata sull'Energia .....	85
Tabella 7.3 Tasso di Fault Detection medio con separazione delle IMFs basata sul Coefficiente di Correlazione.....	85
Tabella 7.4 Tasso di Fault Detection medio con separazione delle IMFs basata sull'Ampiezza.....	86
Figura 7.4 Confronto delle performances con N_train=25 e N_tun=106.....	86
Figura 7.5 Confronto delle performances con N_train=50 e N_tun=81.....	87
Figura 7.6 Confronto delle performances con N_train=75 e N_tun=56.....	87

# Capitolo 1

## Introduzione

Il monitoraggio statistico basato su segnali (monitoraggio in-process) è un'area dello *Statistical Process Control (SPC)* che sta guadagnando sempre più attenzione da parte dei ricercatori e del mondo industriale soprattutto a causa del vasto range di applicabilità di queste tecniche in diversi settori. L'SPC può essere definito come un insieme di procedure e metodi statistici (come ad esempio le carte di controllo) che analizzano la variabilità inerente un processo (o il suo output) allo scopo di mantenere il processo stesso sotto controllo. In particolar modo lo schema di SPC classico prevede di misurare a campione una o più grandezze di interesse sull'output del processo (ad esempio i pezzi realizzati tramite un certo processo manifatturiero) e monitorarne l'andamento nel tempo per valutare se esso rimane in uno stato di controllo o meno [1]. Questo approccio dunque opera direttamente a valle del processo.

Con il *monitoraggio in-process* invece il focus viene spostato direttamente sul processo stesso. In tal caso, il monitoraggio si concentra su segnali acquisiti durante la lavorazione, con vari vantaggi in termini di reattività a scenari di guasto, maggiore copertura della produzione e integrazione con logiche di controllo adattativo. Questo lavoro, in particolare, si concentra su una tipologia di segnali che si incontra frequentemente in applicazioni industriali, cioè segnali ciclici in cui un pattern, d'ora in poi chiamato "profilo" si ripete nel tempo. Un profilo non è altro che una funzione che mostra l'andamento di una grandezza fisica (e.g. una forza) in funzione di una o più variabili indipendenti (e.g. il tempo). Le tecniche di SPC che permettono di trattare dati in tale forma prendono il nome di "profile monitoring" (si vedano, e.g., [2] e [3]). Il loro scopo è quello di individuare delle deviazioni fuori-controllo rispetto a un modello di riferimento che caratterizzi il naturale funzionamento del processo. Spesso queste tecniche vengono applicate nell'ambito del Machine Tool Condition Monitoring, al fine di valutare le condizioni di salute di una macchina utensile o delle sue parti. In questi casi dunque i profili che vengono analizzati sono dei segnali registrati tramite dei sensori. Essi mostrano l'andamento nel tempo di una grandezza fisica (e.g. forza, pressione, accelerazione, ...) che viene sfruttata come fonte di informazione circa (1) lo stato di salute dell'utensile o della macchina che esegue la lavorazione, o (2) la qualità/stabilità del processo.

L'analisi *in-process* dei segnali rispetto all'SPC classico permette di godere di molti benefici come la riduzione di materiali di scarto e pezzi da rilavorare, aumento della produttività, riduzione dei costi di produzione, risparmio di tempo, e infine dà la possibilità di intervenire immediatamente sul processo con delle azioni correttive al fine di riportarlo al più presto nelle condizioni di corretto funzionamento, senza

aspettare che esso realizzi un output non conforme alle specifiche e quindi da scartare.

Il presente lavoro si focalizza sul monitoraggio di profili che per loro natura risultano essere *multiscala*, ovvero caratterizzati dalla compresenza di diversi livelli tempo-frequenza. Un profilo multiscala è per esempio costituito da uno o più modi oscillatori a bassa frequenza a cui si sommano delle componenti transitorie (ovvero localizzate nel tempo) ad alta frequenza e ampiezza limitata. Questo tipo di profili sono abbastanza comuni nella pratica industriale e il loro monitoraggio presenta alcune difficoltà che rendono inadatte la maggior parte delle tecniche convenzionalmente utilizzate in quest'ambito.

Le tecniche di analisi dei segnali possono essere suddivise in tre macro-categorie, in base al modo in cui rappresentano il segnale di input: nel dominio del tempo, nel dominio delle frequenze e nel dominio tempo-frequenze.

Le tecniche di monitoraggio di segnali nel dominio del tempo consistono nel calcolare uno o più indicatori sintetici che riassumano il contenuto informativo del segnale stesso. Monitorando l'andamento nel tempo di questi indicatori è possibile dedurre se il profilo è stato generato da un processo in controllo o fuori controllo. Alcuni degli indicatori più usati includono la media, l'rms, la deviazione standard, l'indice di kurtosis o di skewness, etc.

L'utilizzo di queste tecniche nel caso di profili multiscala comporta scarsi risultati in quanto l'uso di indicatori sintetici non permette di rappresentare nel modo opportuno la suddivisione dei diversi livelli tempo-frequenza.

L'analisi di un segnale nel dominio delle frequenze invece cambia la rappresentazione che viene data dei dati. Mentre il grafico nel dominio del tempo mostra come il segnale varia nel tempo, un grafico nel dominio delle frequenze mostra come il segnale si distribuisce su diversi range di frequenza. Un grafico di questo tipo è detto spettro delle componenti in frequenza. Per questo motivo l'analisi di un segnale nel dominio delle frequenze è anche nota come *spectrum analysis* o *spectral density estimation*.

La conversione di un segnale dal dominio del tempo a quello delle frequenze si ottiene tramite degli operatori matematici detti Trasformate. La più famosa Trasformata utilizzata per l'analisi dei segnali è quella di Fourier [4-5]. La Trasformata di Fourier di fatto scompone un segnale nella somma di tanti segnali semplici, ognuno con una frequenza diversa, matematicamente rappresentati da sinusoidi e cosinusoidi. Ognuna delle barre verticali nello spettro delle componenti in frequenza rappresenta l'ampiezza della sinusoide con la frequenza riportata in ascissa. Questi coefficienti possono quindi essere monitorati nel tempo per valutare la stabilità del segnale.

Il problema fondamentale legato all'utilizzo della Trasformata di Fourier come tecnica di analisi dei segnali multiscala sta nell'ipotesi di stazionarietà del segnale che deve essere rispettata. Spesso nella pratica industriale si ha a che fare con dei segnali che non sono assolutamente stazionari, e questo comporta l'incapacità di individuare profili fuori controllo da parte del sistema di monitoraggio.



Le tecniche di analisi dei segnali nel dominio tempo-frequenze sono quelle maggiormente indicate per un sistema di monitoraggio di profili multiscala. Queste permettono di catturare la distribuzione tempo-frequenza del segnale e quindi migliorano l'estrazione delle features di guasto che possono essere localizzate nel tempo. L'analisi tempo-frequenza include tutte quelle tecniche che studiano un segnale simultaneamente nel dominio del tempo e in quello delle frequenze. Dunque invece di dare una rappresentazione mono-dimensionale del segnale, queste tecniche lo rappresentano in due dimensioni, tramite delle opportune trasformate. In questo modo è possibile per esempio valutare come il contenuto in frequenza di un segnale varia nel tempo della sua durata.

La *Short-time Fourier Transform (STFT)* [6], è una trasformata che valuta come variano nel tempo i contributi delle diverse sinusoidi in cui il segnale viene scomposto. Si ottiene così un grafico detto Spettrogramma che mostra come varia nel tempo lo spettro delle componenti in frequenza.

La *Wavelet Analysis* [7-9] è un altro approccio molto importante nell'ambito della rappresentazione dei segnali nel dominio tempo-frequenze. La procedura che si segue per applicare l'analisi Wavelet a un segnale prevede innanzitutto la selezione di una opportuna Wavelet madre, e poi il confronto del segnale in esame con diverse versioni traslate e dilatate della Wavelet madre selezionata. La Wavelet madre è una forma d'onda oscillante di lunghezza finita che deve soddisfare alcune condizioni e che viene traslata e scalata per essere adattata al segnale in ingresso [8,9].

Esistono diverse Wavelet madre che possono essere selezionate per analizzare un segnale. La scelta della Wavelet madre che meglio si adatta al segnale in esame e del numero di livelli di decomposizione possono essere molto critiche in quanto una scelta errata potrebbe comportare dei risultati non soddisfacenti in termini di rappresentazione tempo-frequenza del segnale [10].

Nel presente studio viene applicata una tecnica di analisi dei segnali nel dominio tempo-frequenze nota come Empirical Mode Decomposition (EMD) [11]. Questa tecnica presenta diversi vantaggi rispetto alle altre presentate finora.

La EMD realizza la decomposizione di un segnale nella somma di un numero finito di funzioni dette Intrinsic Mode Functions (IMFs), ognuna delle quali è caratterizzata da un range di frequenze diverso. Ciascuna delle IMFs rappresenta uno dei livelli tempo-frequenza che costituiscono il segnale. La EMD quindi può essere efficacemente sfruttata al fine di separare le features di guasto sulle diverse componenti di un segnale multiscala.

La sua caratteristica principale è quella di essere completamente adattiva e non-parametrica: la EMD, cioè, non richiede la definizione di una funzione base o di altri parametri e non necessita che il segnale di input soddisfi determinate condizioni di stazionarietà o linearità, ma si adatta a tutti i profili, indipendentemente dalle loro caratteristiche. Il fatto di essere una procedura del tutto data-driven le permette di essere applicata all'analisi di pressoché qualsiasi tipo di segnale.

Alcuni autori, come per esempio Peng *et al.* [12] hanno effettuato degli studi che mettono a confronto le performance della EMD e della Trasformata Wavelet nell'ambito dell'individuazione di guasti e hanno mostrato come la EMD possa essere effettivamente un eccellente competitor.

In questo lavoro si prende in considerazione come caso di studio una macchina che impiega la tecnologia WaterJet [13] per eseguire delle lavorazioni di taglio. Questa tecnologia consiste nello sfruttare l'energia cinetica di un getto d'acqua ad altissima pressione per eseguire un vasto range di possibili operazioni su un ampio ventaglio di materiali lavorabili. La pressione di questo getto d'acqua, e in particolar modo il profilo rappresentante il suo andamento nella durata di una lavorazione, vengono utilizzati in questo studio come fonte d'informazione sulle condizioni di salute della macchina. Questi segnali risultano essere multiscala in quanto alle oscillazioni di bassa frequenza si sommano dei transitori di alta frequenza localizzati nel tempo che andrebbero isolati al fine di massimizzare le prestazioni del sistema di monitoraggio.

Uno dei problemi principali riscontrati nell'applicazione della EMD a questi segnali è il fenomeno noto in letteratura come *effetto Mode-Mixing*: esso comporta una sovra decomposizione del segnale dovuta alla presenza di rumore di fondo. In altre parole la EMD decompone il segnale in un numero di IMFs eccessivo, e alcune delle componenti in frequenza vengono ripartite su due o più IMFs. Gran parte del presente lavoro è quindi focalizzato sulla definizione di una strategia per ovviare a questo problema e per ottenere in modo automatico una decomposizione sintetica del segnale.

Alcuni dei metodi proposti in letteratura come la EEMD o la CEEMDAN si sono rivelati, alla luce dei risultati ottenuti, di scarsa utilità pratica: questi infatti non solo non risolvono il problema del Mode-Mixing, che continua a presentarsi, ma risultano anche molto onerosi dal punto di vista dei tempi computazionali e quindi sono inadatti ad un sistema di monitoraggio in-process che dovrebbe segnalare prontamente l'eventuale presenza di anomalie. L'alternativa proposta in letteratura a questi algoritmi è il cosiddetto metodo delle Combined Mode Functions (CMFs) che propone di risolvere il mode-mixing raggruppando e sommando tra di loro quelle IMFs che risultano essere simili tra di loro in quanto, in realtà, non sono altro che parti di uno stesso modo oscillatorio. Queste fusioni generano K segnali che vengono chiamati CMFs e che rappresentano i veri livelli tempo-frequenza che compongono il segnale multiscala.

Nell'applicare questo metodo ci sono due grandi questioni che la letteratura scientifica ha trascurato di affrontare o ha affrontato in maniera superficiale: la prima riguarda il criterio con cui selezionare le IMFs da raggruppare insieme, e la seconda è la scelta del numero K di CMFs.

Per quanto riguarda la prima questione, in letteratura spesso sono stati proposti dei metodi a grado di automazione nulla, basati per esempio sull'ispezione visiva delle IMFs oppure sull'analisi di indicatori sintetici che riassumessero il contenuto informativo delle varie IMFs. Metodi di questo tipo sono ovviamente inadatti a un sistema di monitoraggio on-line dei profili multiscala come quello che si intende

mettere a punto in questo studio. Essi infatti presuppongono la presenza di un operatore che esegua queste ispezioni off-line, il che aumenterebbe di molto i tempi necessari a segnalare un'eventuale guasto, mandando in fumo di fatto i vantaggi derivanti dall'utilizzo di un sistema di profile monitoring al posto di un approccio di SPC classico. In questo lavoro quindi viene proposto un metodo del tutto automatico basato sulle Probability Distribution Functions (PDFs) delle  $n$  IMFs di un segnale e sulla distanza euclidea tra PDFs contigue.

La seconda questione invece non viene mai affrontata nella letteratura scientifica. Nella stragrande maggioranza dei casi infatti, o l'intenzione degli autori è quella di separare semplicemente la componente di rumore da quella del segnale – e quindi necessariamente viene selezionato  $K=2$  – oppure gli autori presuppongono che sia noto il valore di  $K$ . Quest'ultima ipotesi risulta poco fattibile in ambito industriale in quanto un utente di un sistema di monitoraggio non può ragionevolmente conoscere a priori il numero di componenti in cui scomporre un segnale. Nel presente lavoro vengono proposti alcuni indici per la valutazione del valore ottimale di  $K$ . Questi indici sono presi in prestito dal mondo del clustering e della cluster validation, facendo leva sull'idea che valutare la separazione in clusters di  $n$  osservazioni di un dataset non sia molto dissimile dal valutare la separazione in CMFs di  $n$  IMFs.

In questo modo il sistema può operare in maniera del tutto automatica, senza che sia necessaria la presenza di un operatore che supervisioni.

Una volta separate le diverse componenti di un profilo multiscala tramite questa procedura è possibile passare alla fase di monitoraggio vero e proprio dei profili. Ognuna delle CMFs risultanti da ogni profilo infatti risulterà essere non più multiscala, e quindi il monitoraggio può essere effettuato applicando una tecnica classica per il monitoraggio di profili quale la *Principal Component Analysis (PCA)*. Questa tecnica, usata nell'ambito della statistica multivariata, permette di ridurre la complessità dimensionale di un dataset ad un numero limitato di nuove variabili, dette appunto Principal Components (PCs). Una volta addestrato il modello e progettate delle carte di controllo note come “carta  $T^2$ ” e “carta  $Q$ ” con un certo numero di profili In Controllo è possibile passare alla fase di monitoraggio vero e proprio, in cui per ogni nuovo profilo vengono calcolate le statistiche di controllo per ogni CMFs, che vengono poi confrontate con i rispettivi limiti di controllo. La combinazione di EMD e PCA dà vita a un approccio denominato EMD-PCA che è stato già presentato da alcuni studiosi per migliorare le performance della EMD classica nell'analisi dei segnali (e.g. [14]). Nel caso dei segnali di pressione WaterJet i risultati che si ottengono denotano chiaramente la capacità che questa tecnica ha di individuare delle condizioni di guasto sulla macchina.

Il documento è organizzato secondo la seguente struttura:

**Capitolo 2:** viene presentato il caso di studio affrontato come esempio di monitoraggio di profili multiscala. Si illustra quindi la composizione del dataset di

profili di pressione WJ utilizzato per lo svolgimento delle analisi e le caratteristiche tempo-frequenza di questi segnali.

**Capitolo 3:** si affronta nel dettaglio la tecnica EMD, se ne discutono vantaggi e svantaggi e viene presentato un breve stato dell'arte sui casi di studio in cui è stata applicata la EMD in letteratura. Nella seconda parte di questo capitolo vengono poi introdotte le varianti più comunemente indicate in letteratura come soluzioni al problema del mode-mixing con i rispettivi pro e contro.

**Capitolo 4:** viene introdotto il metodo innovativo proposto per la separazione delle IMFs in  $K$  CMFs e per la scelta del valore ottimale di  $K$ .

**Capitolo 5:** vengono mostrati i risultati ottenuti applicando il metodo proposto e i suoi competitor ad un dataset di profili multiscala generati artificialmente. In una prima parte si ipotizza di conoscere il valore di  $K$  e successivamente vengono presentate e confrontate una serie di tecniche per la scelta automatica del  $K$ .

**Capitolo 6:** vengono illustrati i risultati ottenuti relativamente al case-study dei profili di pressione WJ.

**Capitolo 7:** viene presentata la strategia di monitoraggio della stabilità delle CMFs nel tempo basata sulla PCA e si mostrano i risultati ottenuti.

**Capitolo 8:** si presentano le conclusioni del lavoro svolto e i possibili sviluppi futuri.

## Capitolo 2

### Il caso di studio: profili di pressione WaterJet

Un processo industriale in cui si possono ottenere segnali caratterizzati da pattern multiscala è il taglio waterjet.

La tecnologia WaterJet sfrutta l'energia di un getto d'acqua ad altissima pressione per eseguire diverse operazioni (taglio, fresatura, finitura superficiale, etc...) su una gamma molto ampia di materiali lavorabili, con spessori fino a qualche centinaio di millimetri. Il getto può essere costituito da acqua pura, come e.g. quando questa tecnologia viene applicata in ambito alimentare, eventualmente arricchita con particelle di sostanze solide abrasive qualora si avesse la necessità di tagliare materiali più spessi o più duri con velocità maggiore. In quest'ultimo caso si parla di tecnologia Abrasive WaterJet (AWJ). Figura 2.1 mostra uno schema di funzionamento della testa di taglio di una macchina AWJ.

Nel presente lavoro di tesi viene utilizzato un dataset di dati reali acquisiti in processi di taglio waterjet per valutare l'applicabilità della Empirical Mode Decomposition quale tecnica per il monitoraggio on-line di segnali multiscala. Il dataset consiste di segnali acquisiti tramite un trasduttore di pressione sulla linea ad alta pressione dell'acqua, durante il taglio di lamiera di alluminio. I dettagli del caso applicativo e dei segnali acquisiti sono descritti in Grasso *et al.* [30].

Alcuni di questi segnali sono stati campionati in condizioni di normale funzionamento della macchina e quindi rappresentano il modello di riferimento per il sistema di monitoraggio, mentre altri sono stati acquisiti simulando diversi tipi di guasto per alcuni dei componenti più importanti. In questo modo è possibile valutare l'effetto che diversi tipi di guasto hanno sulla pressione del getto d'acqua.

Lo scopo del presente lavoro è quello di realizzare un sistema di monitoraggio on-line di profili multiscala che sia in grado di segnalare in maniera tempestiva, affidabile e totalmente automatica la presenza di guasti ad una delle componenti della macchina WJ (pompa ad altissima pressione o testa di taglio), utilizzando i segnali di pressione come fonte di informazione.

La macchina WaterJet in cui è stato montato il trasduttore di pressione per il campionamento dei dati presenta una pompa intensificatrice costituita da tre sistemi cilindro-pistone paralleli. Per questo, ogni ciclo di lavorazione della macchina è costituito da tre cicli consecutivi, uno per ogni pistone. Il ciclo di ciascun pistone è dato da tre diverse fasi:

- a. La fase di pre-compressione, durante la quale il pistone si muove in avanti fino a quando la pressione dell'acqua all'interno del cilindro eguaglia la pressione dell'acqua nel circuito di scarico. Durante questa fase la valvola di ritegno

dello scarico rimane chiusa e quindi il movimento del pistone non contribuisce attivamente alla formazione del getto d'acqua;

- b. La fase di compressione, in cui la valvola di ritegno si apre e il pistone pressurizza l'acqua che fluisce nella testa di taglio;
- c. La fase di aspirazione, durante la quale l'acqua viene spostata dal circuito a bassa pressione all'interno del cilindro fino a che il pistone non ritorna a inizio corsa, pronto per cominciare un nuovo ciclo.

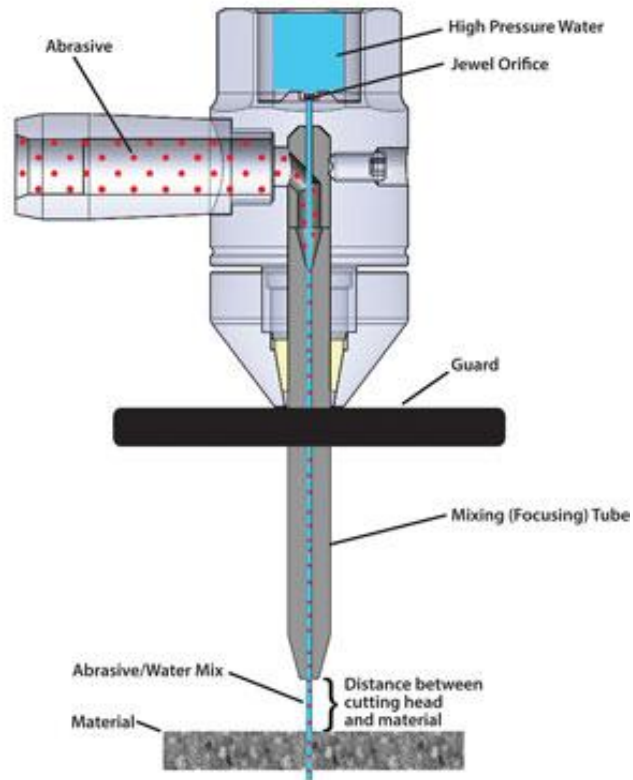
Il dataset che sta alla base delle analisi presentate nel seguito è formato da un totale di 426 profili, così suddivisi:

- 131 profili in controllo (IC): questi rappresentano l'andamento della pressione del getto idrico nel caso di normale funzionamento della macchina WJ. Costituiranno quindi il modello di riferimento per l'individuazione di qualsiasi tipo di guasto;
- 81 profili relativi al funzionamento della macchina WJ con guasto di tipo A, ovvero la presenza di una cricca sulla superficie interna di uno dei cilindri che costituiscono la pompa ad altissima pressione;
- 51 profili riguardanti il guasto di tipo B, ovvero una rottura nella valvola di ritegno dello scarico;
- 76 profili di pressione campionati con presenza di un guasto di tipo C alla macchina, consistente in un livello di usura eccessivo per la sede della valvola di mandata dell'acqua;
- 87 profili di pressione rappresentativi del guasto D, ovvero una rottura dell'orifizio della testa di taglio dal quale fuoriesce il getto idrico.

In [30] è possibile trovare informazioni più dettagliate sul modo in cui sono stati campionati i segnali e simulate le quattro situazioni di guasto.

In figura 2.2 viene riportato l'andamento dei 131 segnali IC, mentre Figura 2.3 ne evidenzia uno in particolare. Figura 2.4, invece, mostra l'andamento di 4 segnali, ognuno rappresentativo di una particolare condizione di guasto.

Dall'analisi di questi grafici è possibile capire come i segnali di pressione che si intende monitorare siano multiscala di natura in quanto presentano un andamento ondulatorio a bassa frequenza a cui però si sommano dei pattern transitori ad alta frequenza localizzati nel tempo.



**Figura 2.1 Schema di funzionamento della testa di taglio di una macchina AWJ**

In particolare esistono due tipi diversi di transitori, che possiamo indicare rispettivamente come principali e secondari e a cui è possibile assegnare significati diversi dal punto di vista della dinamica di funzionamento della macchina:

- i transitori principali, localizzati a circa il 35% e 70% di completamento del ciclo, sono dovuti al passaggio dal funzionamento di un pistone a quello successivo. Questi transitori cioè si verificano nel momento in cui un pistone arriva a fine corsa e il pistone successivo comincia a pressurizzare l'acqua nella testa di taglio. La chiusura di una valvola di ritegno, seguita dall'apertura di quella successiva, causa una discontinuità nella pressione.
- I transitori secondari, visibili in corrispondenza del 15%, 50% e 80% circa di completamento del ciclo, si verificano nel momento in cui un gruppo cilindro-pistone conclude la fase di aspirazione dell'acqua. Queste discontinuità, di minore entità rispetto alle precedenti, sono dovute al fatto che il pistone ritorna all'inizio della sua corsa.

La Empirical Mode Decomposition potrebbe dunque risultare molto utile quale tecnica per separare le features dei segnali su scale diverse e migliorare la capacità del sistema di monitoraggio di individuare guasti sulle componenti della macchina.

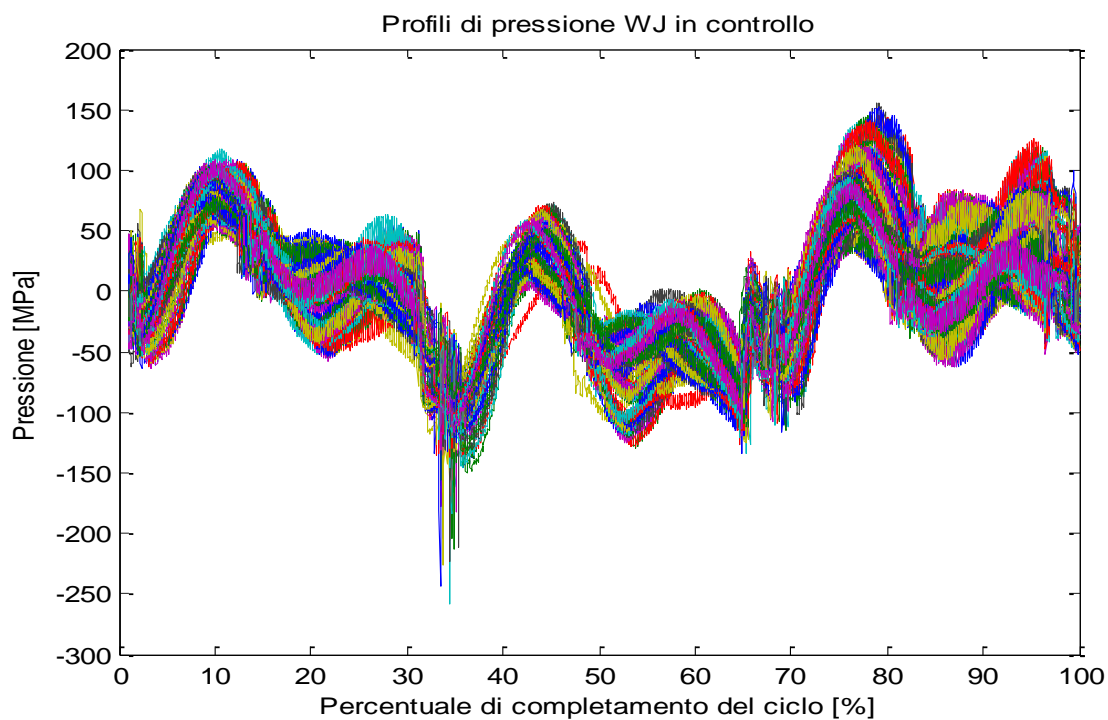


Figura 2.2 Profili di pressione WJ in controllo

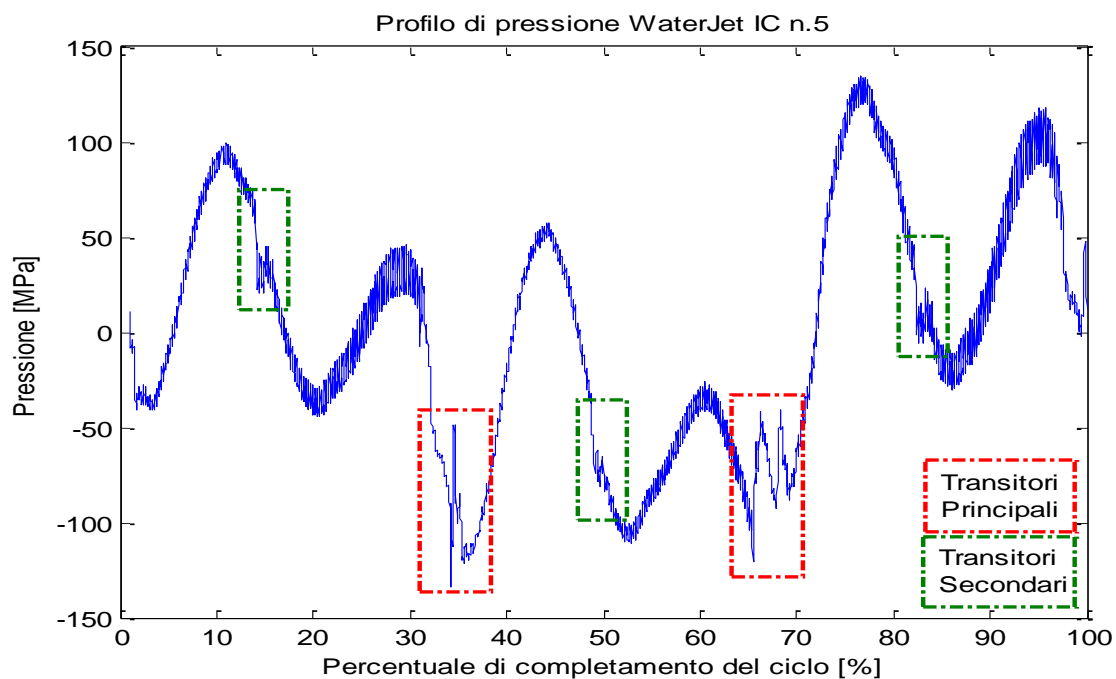


Figura 2.3 Profilo di pressione WJ IC con transitori



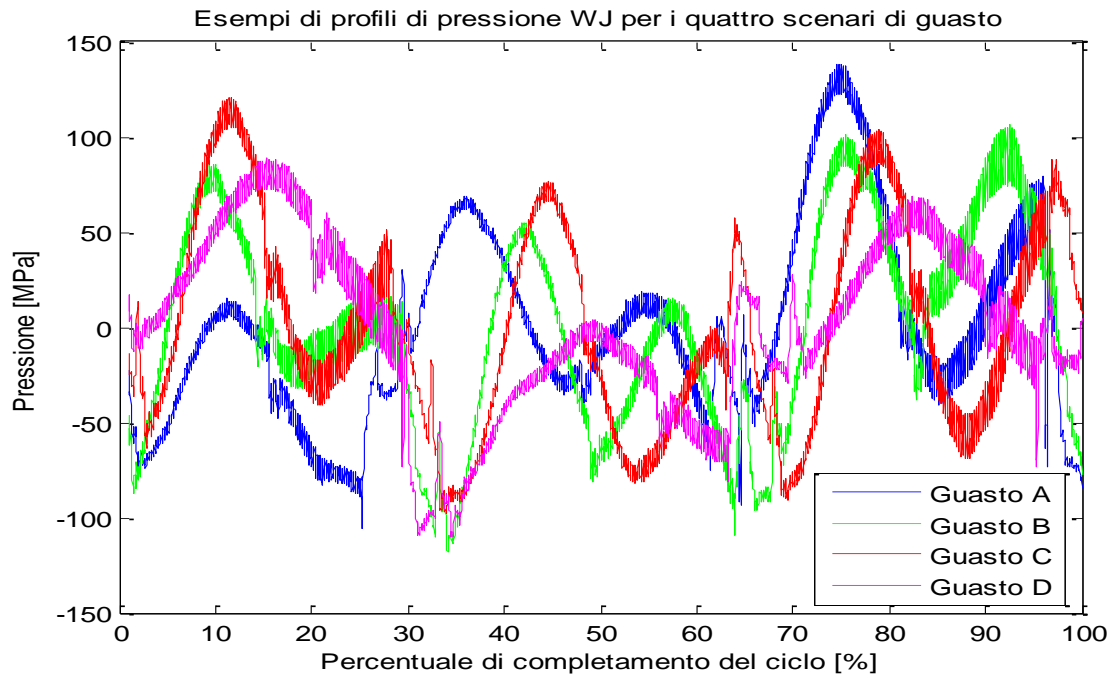


Figura 2.4 Profili WJ per i quattro scenari di guasto

## Capitolo 3

# La Empirical Mode Decomposition (EMD)

### 3.1 La Empirical Mode Decomposition

La Empirical Mode Decomposition (EMD) è una tecnica molto potente per l'analisi tempo-frequenza di un segnale. Essa fu introdotta per la prima volta da Huang *et al.* [11] nel 1998 e da allora è stata ampiamente studiata e applicata in ambiti anche particolarmente diversi gli uni dagli altri, quali ad esempio l'identificazione di guasti di organi rotanti [12,15-20], lo studio di fenomeni naturali [11,21-22], l'analisi di segnali biologici [14,23-28], il monitoraggio delle condizioni di un utensile in lavorazione (o Tool Condition Monitoring, TCM) [29,30], etc...

A partire dalle caratteristiche di un segnale nel dominio del tempo, la EMD permette di ottenere una decomposizione dello stesso in un numero finito (e tipicamente limitato) di funzioni, dette Intrinsic Mode Functions (IMFs). Ogni IMF corrisponde a uno dei naturali modi oscillatori di cui il segnale si compone e quindi ha un certo significato dal punto di vista fisico per il sistema da cui esso deriva [11].

Affinchè una funzione possa essere considerata una Intrinsic Mode Function essa deve soddisfare le seguenti due condizioni:

(1) considerata la funzione nella sua interezza, il numero di estremi (massimi e minimi locali) e il numero di punti in cui essa assume valore nullo sono uguali o al più differiscono di una unità;

(2) in ogni punto della funzione il valor medio dell'involuppo definito dai punti di massimo locale e dell'involuppo definito dai punti di minimo locale è zero.

In realtà, come è stato notato da alcuni autori, per una funzione continua la condizione (1) è ridondante, in quanto è automaticamente soddisfatta qualora venga rispettata la condizione (2).

Nella pratica di solito la EMD viene applicata a dei segnali di tipo discreto, i cui valori sono stati campionati a intervalli regolari da una funzione continua. Per questo nel presentare la tecnica, considereremo un segnale discreto  $x(t) = [x_1, x_2, \dots, x_T]$  con  $t=1,2,\dots,T$ . L'EMD decompone il segnale nelle Intrinsic Mode Functions che lo costituiscono seguendo questa procedura [11]:

1. Si identificano i punti di massimo locali e di minimo locali del segnale  $x(t)$ , dove  $x(t_0)$  viene considerato un punto di massimo locale se i punti ad

esso vicini  $x(t_0-1)$  e  $x(t_0+1)$  sono entrambi più piccoli, mentre  $x(t_0)$  viene considerato un punto di minimo locale se  $x(t_0-1)$  e  $x(t_0+1)$  sono entrambi più grandi.

2. Tutti i punti di massimo locali e tutti quelli di minimo locali vengono connessi da due spline cubiche a formare rispettivamente l'involuppo superiore  $u(t)$  e quello inferiore  $l(t)$ . Essi dovrebbero contenere al loro interno tutti i punti del segnale  $x(t)$ .
3. Si calcola la media degli involuppi superiore e inferiore come segue:

$$m_1(t) = \frac{u(t) + l(t)}{2} \quad (3.1)$$

per ogni  $t=1,2,..T$ .

4. Si calcola la differenza  $h_1(t)$  tra il segnale e la media degli involuppi superiore e inferiore

$$h_1(t) = x(t) - m_1(t) \quad (3.2)$$

Se  $h_1(t)$  soddisfa le due condizioni di cui sopra con cui si definisce una IMF, allora  $h_1(t)$  è la prima IMF, e la si indica con  $c_1(t)$ . Se non soddisfa quei due criteri allora viene eseguito un processo di estrazione (*sifting process*) che consiste nel prendere  $h_1(t)$  come nuovo segnale da decomporre, calcolarne gli involuppi inferiore e superiore e quindi la loro media  $m_{11}(t)$ , per poter infine ricavare la differenza  $h_{11}(t)$  tra il nuovo segnale e la media stessa

$$m_{11}(t) = \frac{l_1(t) + u_1(t)}{2} \quad (3.3)$$

$$h_{11}(t) = h_1(t) - m_{11}(t) \quad (3.4)$$

dove  $l_1(t)$  e  $u_1(t)$  sono rispettivamente l'involuppo inferiore e quello superiore di  $h_1(t)$ .

In generale è necessario eseguire questo processo di estrazione  $k$  volte per generare una Intrinsic Mode Function. La prima IMF ad esempio sarà data da

$$c_1(t) = h_{1k}(t) = h_{1(k-1)}(t) - m_{1k}(t) \quad (3.5)$$

Questo processo di *sifting* viene eseguito fin quando la deviazione standard dei risultati di due iterazioni successive non scende al di sotto di una soglia predeterminata, tipicamente compresa tra 0.2 e 0.3:

$$sd = \sum_{t=1}^T \left[ \frac{|h_{1(k-1)}(t) - h_{1k}(t)|^2}{h_{1(k-1)}^2(t)} \right] \quad (3.6)$$

5. Per identificare le altre IMFs si prende il segnale residuo come il nuovo segnale e si ripetono i punti da 1 a 4:

$$x(t) = r_i(t) = x(t) - c_{i-1}(t) \quad (3.7)$$

6. L'algoritmo si arresta quando l'ennesimo residuo  $r_n(t)$  è una funzione monotona oppure una funzione con meno di due punti di minimo o massimo locale. Da un segnale siffatto infatti non è più possibile estrarre alcuna IMF con l'algoritmo finora descritto.

Il criterio di stopping del processo di estrazione delle IMFs riportato al punto 4 della procedura appena presentata è in realtà solo uno dei tanti criteri che sono stati proposti in letteratura.

Wu e Huang [32], e.g., hanno indicato che il calcolo potrebbe essere arrestato quando il seguente rapporto scende al di sotto di una soglia predefinita:

$$SD_k = \frac{\sum_{t=1}^T |m_{1k}(t)|^2}{\sum_{t=1}^T |h_{1k}(t)|^2} \quad (3.8)$$

Il problema di questi criteri di arresto è che essi non sono in alcun modo relazionati alla definizione di IMF che abbiamo dato, ma si limitano a verificare che i risultati ottenuti da due estrazioni successive siano simili entro un certo limite. Allora Huang et al. [33] hanno proposto un criterio alternativo, in base al quale la procedura di estrazione si arresta nel momento in cui il numero di estremi rimane lo stesso per un numero  $S$  di iterazioni consecutive.

Rato et al. [31], e.g., hanno proposto un criterio di stopping per la procedura di estrazione basato sul calcolo di un fattore di risoluzione. Questo risulta dal rapporto tra l'energia del segnale all'inizio dell'estrazione  $x(t)$  e l'energia della media degli involucri  $m_{ik}(t)$  alla  $k$ -esima iterazione. Qualora questo fattore dovesse superare una soglia prestabilita il calcolo della IMF deve arrestarsi. Secondo Rato et al. [31] l'utilizzo di questo criterio di arresto comporta il vantaggio di poter avere un controllo sul numero di IMFs che verranno estratte, data l'interdipendenza che c'è tra il numero di IMFs e il resolution factor. In altri casi ancora è stato applicato un criterio basato semplicemente sul conteggio del numero di iterazioni eseguite.

Nel presente lavoro la EMD è stata implementata nella sua versione "classica", i.e. quella descritta da Huang *et al.* in [11].

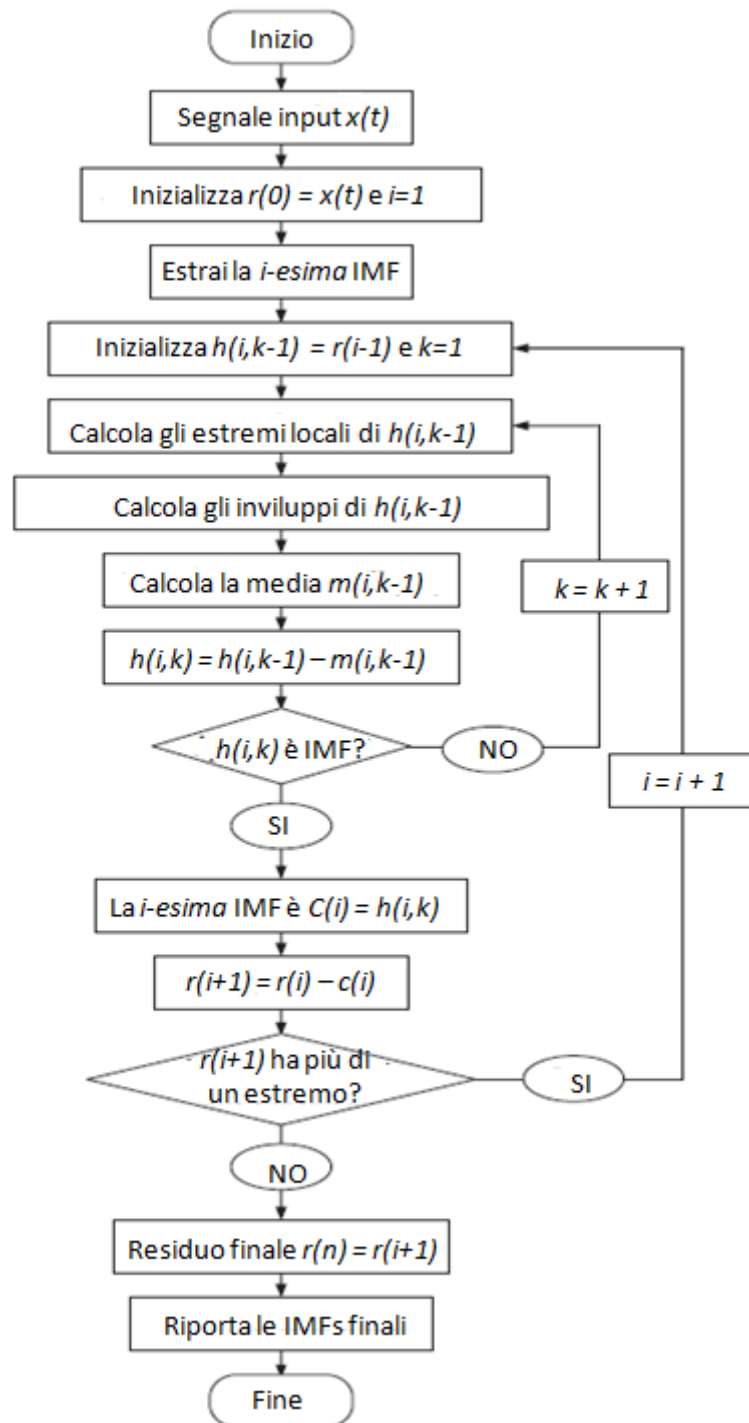


Figura 3.1 La Empirical Mode Decomposition

Alla fine della procedura il segnale originale  $x(t)$  risulterà sempre scomposto in un insieme di  $n$  Intrinsic Mode Function, ognuna rappresentante uno dei modi oscillatori che lo costituiscono, e un residuo  $r_n(t)$  che invece tipicamente rappresenta la componente di trend insita nel segnale :

$$x(t) = \sum_{i=1}^n c_i(t) + r_n(t). \quad (3.9)$$

Per costruzione il numero di estremi delle IMFs decresce dalla prima alla ennesima. Inoltre la prima Intrinsic Mode Function è quella caratterizzata da frequenza maggiore, e infatti tipicamente corrisponde alla componente di rumore insita nel segnale analizzato, mentre le altre risultano avere una frequenza via via minore. Nella maggior parte dei casi pratici, anche segnali estremamente complicati possono essere rappresentati con poche IMF [21].

In figura 3.1 viene riportato uno schema a blocchi che facilita la comprensione di come funziona l'algoritmo EMD [21].

### 3.2 Un esempio numerico illustrativo

Al fine di comprendere meglio il principio funzionamento della EMD risulta utile vederla applicata ad un caso semplice. Le IMFs risultanti dalla decomposizione rivelano la struttura interna del segnale analizzato. La IMF  $c_1$  contiene la componente a frequenza maggiore del segnale, mentre le successive IMF  $c_2, \dots, c_n$  presentano delle frequenze appartenenti a range di valori via via minori.

Consideriamo, e.g., un segnale  $x(t)$  che è dato dalla somma di due componenti oscillatorie di frequenze rispettivamente 100 e 20 Hz [17]:

$$x_1(t) = \cos(2\pi 100t) \quad (3.10)$$

$$x_2(t) = 0.5 \sin(2\pi 20t) \quad (3.11)$$

$$x(t) = x_1(t) + x_2(t) \quad (3.12)$$

Usando la Empirical Mode Decomposition il segnale viene decomposto esattamente in due IMFs e un residuo nullo. Le due IMFs vengono mostrate in Figura 3.2, ed è possibile vedere che la prima IMF  $c_1$  è caratterizzata da un'ampiezza e una frequenza che sono esattamente pari a quelle della armonica a 100 Hz, mentre la seconda IMF  $c_2$  rappresenta la componente armonica a 20 Hz. Questo dimostra che la EMD può estrarre correttamente i modi oscillatori intrinseci di un segnale.

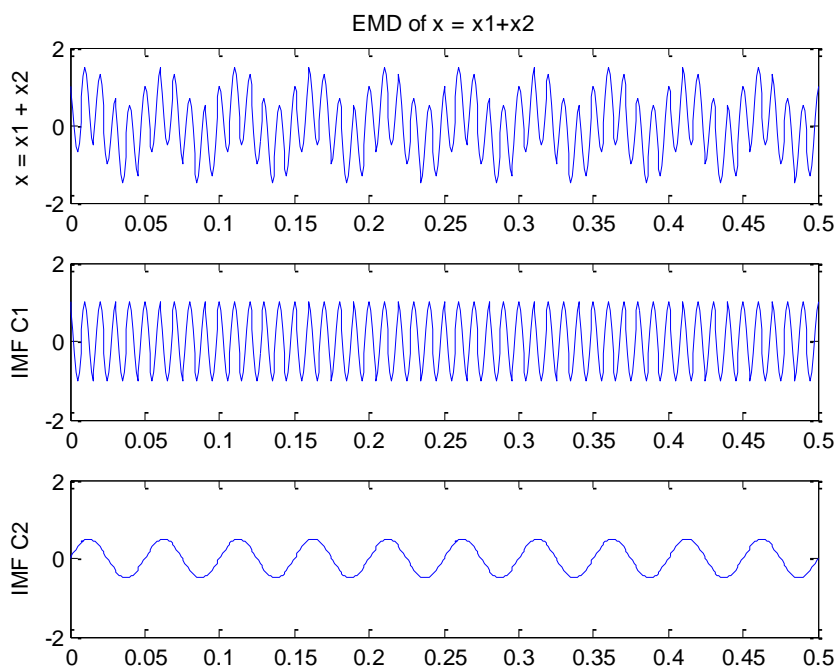


Figura 3.2 EMD del segnale  $x(t)$  in Equazione 3.12

### 3.3 Applicazioni della EMD

La Empirical Mode Decomposition permette di dare una rappresentazione tempo-frequenza di un segnale qualsiasi scomponendolo in un numero finito  $n$  di Intrinsic Mode Functions. La caratteristica principale di questa tecnica è il fatto di essere completamente non-parametrica e adattiva, il che la rende adatta ad essere applicata praticamente a qualsiasi tipo di segnale, indipendentemente dalle sue caratteristiche.

In questo paragrafo cercheremo di dare un'idea della capacità che questa tecnica ha di adattarsi bene anche in contesti molto diversi tra loro e faremo una panoramica sugli ambiti in cui la EMD è stata applicata, senza avere la pretesa di essere esaustivi. In [34] il lettore potrà trovare dei riferimenti più precisi all'applicazione di questa tecnica in molti ambiti, alcuni dei quali non citeremo qui di seguito. In Tabella 3.1 sono riportati alcuni riferimenti bibliografici.

- *Geofisica*: Quando Huang et al. [11] hanno presentato per la prima volta la EMD e la Trasformata di Hilbert come tecniche di analisi dei segnali, ne hanno dimostrato la efficacia applicandole a dei segnali acquisiti durante dei fenomeni naturali quali terremoti o tsunami.

<b>Ambito applicativo</b>	<b>Riferimenti bibliografici</b>
Geofisica	Huang <i>et al.</i> [11] Huang <i>et al.</i> [21] Huang <i>et al.</i> [22]
Medicina e Biologia	Pinheiro <i>et al.</i> [14] Boutana <i>et al.</i> [23] Bu <i>et al.</i> [24] Liang <i>et al.</i> [25] Huang <i>et al.</i> [26] Wang <i>et al.</i> [27] Liang <i>et al.</i> [28]
Structure Health Monitoring	Pines and Salvo [36] Li <i>et al.</i> [37]
Fault Detection per organi rotanti	Lei <i>et al.</i> [15] Loutridis [16] Gao <i>et al.</i> [17] Yan and Gao [18] Junsheng <i>et al.</i> [19] Ricci and Pennacchi [20]
Tool Condition Monitoring	Peng [29] Gratto <i>et al.</i> [31] Roth <i>et al.</i> [38]
Economia e Finanza	Zhang <i>et al.</i> [40]

**Tabella 3.1 Riferimenti bibliografici ai casi di applicazione della EMD**

- *Medicina e biologia*: uno dei campi in cui la EMD ha avuto maggiore successo e in cui ha trovato maggiore applicazione è quello dell'analisi di segnali biomedici come elettrocardiogrammi (ECG), elettroencefalogrammi (EEG), segnali manometrici, etc...

- *Structure Health Monitoring*: si tratta del processo con cui si tracciano e si determinano l'integrità di una struttura quale ad esempio un ponte o un palazzo e con cui si valuta la natura di danni eventualmente riportati a causa dell'invecchiamento e dell'utilizzo [35].

- *Diagnosi di guasti su organi rotanti*: organi rotanti come cuscinetti, rotor e ingranaggi sono una delle classi più comuni di strumenti meccanici e hanno un



ruolo molto importante nelle applicazioni industriali. Generalmente operano in un ambiente di lavoro molto ostile e quindi sono facilmente soggetti a guasti [14].

I segnali di vibrazione che vengono emessi da questi sistemi sono non-stazionari per loro natura. Dato che la Empirical Mode Decomposition è una delle tecniche di analisi dei segnali in tempo-frequenza più potenti e che si adatta benissimo a segnali non-stazionari, questa ha attratto molta attenzione dai ricercatori nel campo della diagnosi dei guasti di organi rotanti.

- *Tool condition monitoring*: con questa espressione ci si riferisce al processo di monitoraggio nel tempo di uno o più parametri rappresentativi delle condizioni di salute di un utensile di taglio in modo tale da poter identificare una variazione significativa che sia indicativa di un guasto imminente. Il tool condition monitoring (TCM) dunque è un aspetto fondamentale per l'adozione di politiche manutentive Condition-Based nell'ambito di un processo manifatturiero. L'adozione di un sistema di TCM efficace ed efficiente può comportare diversi benefici come migliore qualità dei prodotti realizzati, minore downtime delle macchine e quindi riduzione dei costi complessivi di produzione [38].

- *Economia e Finanza*: in alcuni casi la EMD è stata applicata anche per lo studio di profili che costituiscono l'andamento nel tempo di grandezze di tipo economico.

### 3.4 Vantaggi e svantaggi della EMD

La Empirical Mode Decomposition è una tecnica per l'analisi in tempo-frequenza dei segnali che dà buoni risultati in situazioni dove altri metodi, come la trasformata di Fourier o le Wavelet, sono meno adatti oppure più complessi da applicare. In particolare va sottolineato come, almeno da un punto di vista concettuale, questo algoritmo presenta i seguenti vantaggi [31]:

- è semplice, e infatti non richiede alcuna conoscenza preliminare sulla struttura del segnale in esame;
- non si basa su alcun tipo di ipotesi relativa al segnale analizzato, soprattutto non richiede che siano soddisfatte le condizioni di linearità e stazionarietà del segnale;
- può essere applicato a un insieme molto vario di segnali.

Ciononostante, questa tecnica presenta alcuni difetti, che sono stati studiati e discussi da Rato et al. [31], i quali hanno anche successivamente proposto una versione modificata della Empirical Mode Decomposition che sia in grado di

superare alcuni dei problemi evidenziati. In particolar modo in [31] i seguenti step dell'algoritmo vengono indiziati come critici:

- *Individuazione degli estremi*: dato che l'algoritmo nella maggior parte dei casi opera su segnali discretizzati e non nel continuo, questo potrebbe comportare dei problemi nella individuazione degli estremi qualora il campionamento non venisse effettuato con una frequenza adeguata. In tal caso infatti potrebbe succedere che alcuni degli estremi del segnale continuo cadano tra due punti campionati e quindi non vengano correttamente individuati. Per evitare questo problema Rilling et al. [41] hanno proposto di sovracampionare il segnale da sottoporre a EMD.

- *Interpolazione degli estremi*: la qualità della decomposizione dipende dalla scelta della funzione di interpolazione. Nella versione originale dell'algoritmo [11] gli estremi vengono interpolati con una Spline cubica. L'utilizzo di funzioni di interpolazione diverse potrebbe portare a stime diverse delle IMFs. Per esempio Rato et al. [31] hanno provato anche ad utilizzare l'interpolatore Akima, ottenendo comunque dei risultati molto simili.

- *End effect*: con questo termine ci si riferisce all'effetto che ha sul calcolo delle IMFs il modo in cui vengono considerati e trattati il primo e l'ultimo punto del segnale. Questi infatti possono o essere considerati allo stesso tempo sia punti di minimo che di massimo, e questo forza tutte le IMFs ad avere valore nullo in quei punti, oppure possono essere considerati punti di massimo o di minimo a seconda della posizione relativa del punto adiacente.

- *Criterio di arresto della procedura di sifting*: nell'introdurre la tecnica EMD abbiamo già avuto modo di sottolineare come non esista un criterio univoco ma in letteratura ne siano stati proposti diversi.

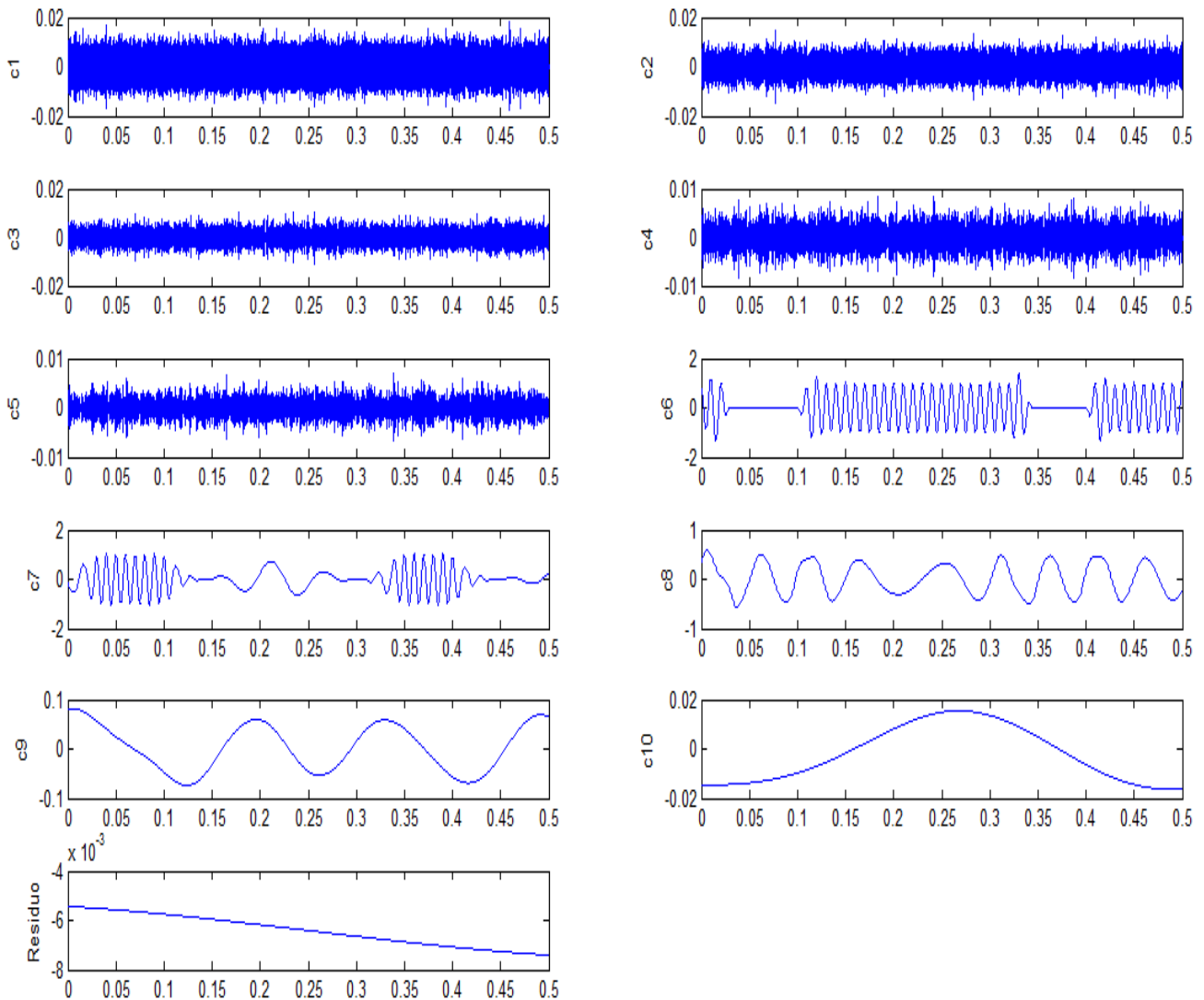
Il problema principale nell'applicazione della Empirical Mode Decomposition è però rappresentato dal possibile manifestarsi del "mode-mixing" [32] quando essa viene applicata a dei segnali affetti da rumore. Questo fenomeno comporta la sovra-decomposizione di un segnale, e cioè porta l'algoritmo a estrarre dal segnale un numero di IMFs superiore al numero di modi oscillatori che effettivamente compongono il segnale. In tal modo uno o più modi oscillatori risultano essere ripartiti in più IMFs. Questo problema limita la affidabilità di uno schema di monitoraggio dei segnali basato sulla EMD [30].

Per illustrare questo problema, consideriamo il segnale in (3.12) e aggiungiamo ad esso una componente di rumore bianco  $e(t)$  di ampiezza 0.0075 [17].

$$x(t) = \cos(2\pi 100t) + 0.5\sin(2\pi 20t) + 0.0075e(t) \quad (3.13)$$

Come è possibile vedere in Figura 3.3 – che mostra la decomposizione ottenuta applicando la EMD al segnale in Equazione 3.13 - invece di ottenere 3 IMFs,

ovvero tante quanti sono i modi oscillatori che compongono il segnale  $x(t)$ , la EMD ne ha estratte 10 più un residuo.



**Figura 3.3** IMFs per il segnale  $x(t)$  in Equazione (3.13)

Questo fenomeno è dovuto al fatto che la EMD è una tecnica del tutto data-driven: l’aggiunta di rumore comporta la nascita di punti di massimo e di minimo “fittizi” che vengono identificati dall’algoritmo e interpolati. Questo fa sì che anche un segnale mono-componente come una semplice sinusoide venga scomposta in più sotto-segnali se affetta da rumore.

In letteratura sono stati proposti diversi metodi per affrontare il problema del mode-mixing. In particolare è possibile individuare due macro filoni: da un lato

si hanno algoritmi come la Ensemble Empirical Mode Decomposition (EEMD) [32] o la Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) [42] che rappresentano delle varianti della EMD [11], mentre dall'altro si ha il metodo delle Combined Mode Functions (CMFs) [17]. In questo caso il problema principale sta nell'identificare un criterio in base al quale raggruppare le IMFs generate dalla Empirical Mode Decomposition e nello scegliere il numero più opportuno di CMFs da realizzare.

Nei prossimi paragrafi queste tecniche vengono discusse più nel dettaglio e vengono anche presentati alcuni dei criteri introdotti in letteratura per combinare le IMFs.

### 3.5 La Ensemble Empirical Mode Decomposition

La Ensemble Empirical Mode Decomposition (EEMD) è stata introdotta per la prima volta da Huang e Wu [32] come tecnica per l'eliminazione del mode mixing effect. Essa viene anche definita come un metodo di analisi dei segnali "noise-assisted". Infatti questo algoritmo definisce le "vere" Intrinsic Mode Functions (che nel seguito indicheremo come  $\overline{IMF}$ ) come la media delle corrispondenti IMFs ottenute applicando la EMD su un insieme di profili di prova, generati aggiungendo diverse realizzazioni di rumore bianco di varianza finita al segnale originale  $x(t)$ . L'aggiunta del rumore al segnale dovrebbe portare alla soluzione del mode mixing andando a cambiare la distribuzione dei punti di massimo e di minimo [43].

Più nel dettaglio gli step da eseguire per applicare la EEMD sono i seguenti [32]:

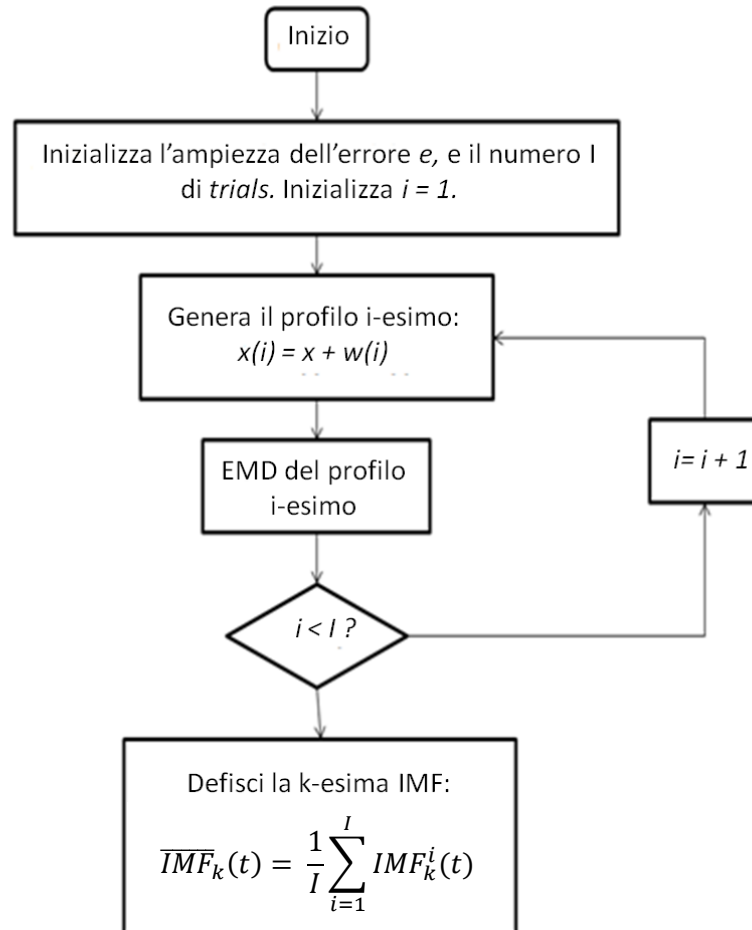
- 1) Generare, a partire dal segnale originale  $x(t)$ , un ensemble di  $I$  profili  $x^i(t)$  dati dalla somma del segnale originale e di una componente di errore gaussiano  $w^i(t)$  ( $i=1,2,\dots,I$ ) di ampiezza  $e$ :

$$x^i(t) = x(t) + w^i(t) \quad (3.14)$$

- 2) Decomporre ciascun profilo  $x^i(t)$  ( $i=1,2,\dots,I$ ) con la EMD e ottenere i rispettivi modi  $IMF_k^i(t)$ , dove  $k=1,2,\dots,K$  indica i modi.
- 3) Definire la  $k$ -esima IMF del segnale  $x(t)$  come la media delle corrispondenti  $IMF_k^i(t)$ :

$$\overline{IMF}_k(t) = \frac{1}{I} \sum_{i=1}^I IMF_k^i(t) \quad (3.15)$$

In figura 3.4 è mostrato uno schema a blocchi che illustra la procedura EEMD.



**Figura 3.4 La Ensemble Empirical Mode Decomposition (EEMD)**

Sebbene la Ensemble Empirical Mode Decomposition sia stata introdotta al fine di ovviare al problema del mode mixing, le performance della EEMD dipendono molto dalle scelte relative ai valori di due parametri, ovvero l'ampiezza  $e$  dell'errore bianco  $w^i$  e il numero di prove  $I$ . Huang e Wu [32] consigliano di utilizzare un rumore con deviazione standard pari a 0.2 volte la deviazione standard del segnale  $x(t)$  che si intende scomporre.

Zhang *et al.* [43] hanno analizzato più approfonditamente l'effetto che questi due parametri hanno sul numero di IMFs generate dalla EMD. In particolare gli autori hanno dimostrato che se l'ampiezza del rumore aggiunto è troppo piccola questo non ha effetto sugli estremi locali del segnale originale e quindi non ha effetto sulla decomposizione, mentre al contrario se è troppo grande comporta

l'estrazione di un numero eccessivo di IMFs, alcune delle quali risultano essere ridondanti.

Applicando la EEMD al segnale in Equazione (3.13), con  $I=100$  e ampiezza del rumore pari a 0.2 volte la deviazione standard di  $x(t)$ , si ottengono i risultati riportati in Figura 3.5. Come è possibile vedere la EEMD applicata con questa particolare combinazione di parametri ha portato alla estrazione di un numero addirittura superiore di IMFs rispetto alla EMD.

Stando a questo esempio, la Ensemble Empirical Mode Decomposition non sembra essere un efficace soluzione al problema del mode mixing. L'applicazione della EEMD comporta inoltre elevati tempi di calcolo se messi a confronto con la EMD tradizionale. Questo è dovuto al fatto che la EEMD sostanzialmente consiste nella ripetizione delle EMD per un numero di volte pari al numero di ensembles selezionato.

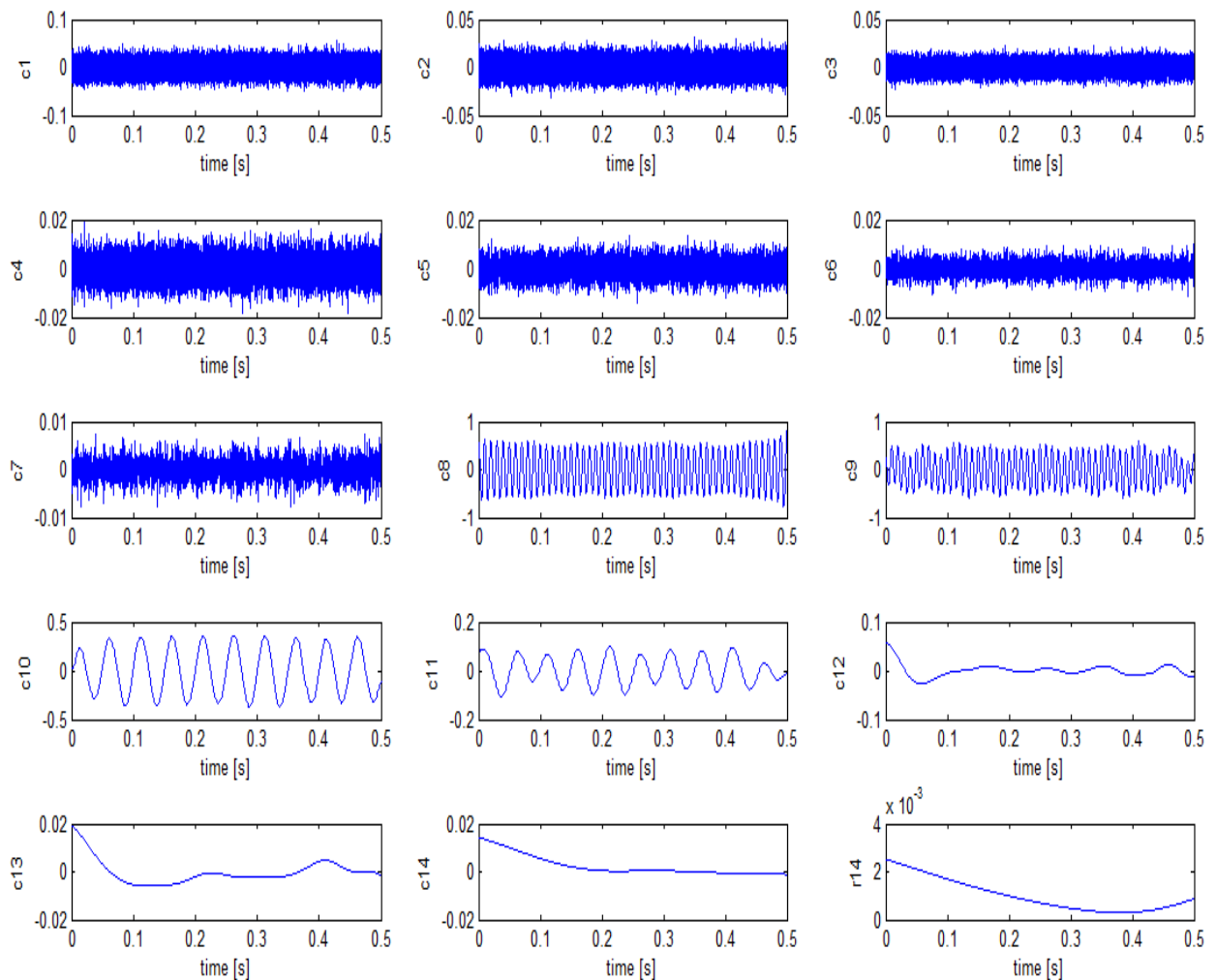
Zhang *et al.* [43] consigliano di adottare un approccio automatico di selezione dell'ampiezza del rumore basato sul calcolo del rapporto segnale-rumore (SNR = Signal to Noise Ratio) tra il segnale da decomporre e il rumore che vi si aggiunge nelle diverse prove:

$$SNR = 10 * \log_{10} \left( \frac{P_{segnale}}{P_{rumore}} \right) \quad (3.16)$$

dove  $P_{segnale}$  e  $P_{rumore}$  rappresentano rispettivamente la potenza del segnale  $x(t)$  e quella del rumore  $w^i(t)$  ad esso aggiunto in ogni prova dell'ensemble.

In [43] un valore di SNR compreso tra 50 e 60 dB viene indicato come un buon valore per applicare l'EEMD di modo che non vi siano né IMFs ridondanti né mode-mixing. Una volta stabilito il valore di SNR desiderato, essendo nota la potenza del segnale, è possibile risalire all'ampiezza opportuna del rumore da aggiungere.

Zhang *et al.* [43] hanno anche proposto una versione leggermente modificata della EEMD al fine di ovviare al suo problema principale, ovvero la scarsa efficienza computazionale, introducendo quella che è stata chiamata la Modified EEMD (MEEMD). Secondo questi ricercatori infatti la parte a bassa frequenza del rumore bianco che si aggiunge al segnale nei diversi trials ha un effetto sui risultati dell'applicazione del processo EEMD (e cioè porta a ridurre il mode mixing effect) a patto che il suo contenuto in frequenza si sovrapponga a quello del segnale originale. Al contrario, la parte ad alta frequenza del rumore bianco aggiunto non ha alcun effetto. Questo suggerisce come un miglioramento dell'efficienza computazionale del processo EEMD possa essere ottenuto sostituendo il rumore bianco con un rumore a banda limitata. Da un punto di vista tecnico questo rumore a banda limitata può essere ottenuto applicando un filtro passa-basso al rumore bianco. La frequenza di cut-off deve essere valutata di caso in caso guardando allo specchio di frequenza del segnale in esame ottenuto mediante Trasformata di Fourier.



**Figura 3.5 EEMD del segnale in Equazione (3.13)**

Secondo Zhang et al. [43] questa sostituzione del rumore bianco con un rumore a bassa frequenza rappresenta una buona soluzione per la riduzione del numero di ensembles richiesti per ottenere delle IMFs significative, e inoltre la MEEMD ridurrebbe il carico computazionale richiesto.

Anche Lei et al. [44] hanno studiato una versione modificata della Ensemble Empirical Mode Decomposition, che hanno chiamato Adaptive EEMD. Questa tecnica, secondo gli autori, migliorerebbe la capacità della EEMD di evitare il problema del mode-mixing.

Nella maggior parte dei casi in cui la EEMD è stata utilizzata come tecnica di signal processing, i parametri di ampiezza dell'errore bianco e di numero di estrazioni sono sempre stati fissati ad un valore costante per tutte le IMFs. In realtà però Lei et al. [44] hanno notato che le componenti a bassa frequenza e

quelle ad alta frequenza del segnale che deve essere decomposto hanno una diversa sensibilità all'ampiezza del rumore aggiunto.

Per questo nell'Adaptive EEMD [44] si procede utilizzando rumore di ampiezza maggiore e un numero maggiore di iterazioni di estrazione per le componenti ad alta frequenza (ovvero le prime IMFs), mentre per estrarre le Intrinsic Mode Functions a bassa frequenza l'ampiezza del rumore così come il numero di estrazioni sono ridotti.

### 3.6 La Complete Ensemble EMD With Adaptive Noise

Il problema del mode mixing nell'EEMD è stato affrontato anche da Torres *et al.* [42]. Applicando l'algoritmo EEMD a due segnali, di cui uno è un segnale delta e l'altro un elettrocardiogramma (ECG), gli autori hanno osservato una grande variabilità nel numero di modi oscillatori ottenuti in ogni trial. Le soluzioni proposte a questo problema sono essenzialmente tre:

1. effettuare uno "zero-padding" dei modi mancanti, ovvero inserire dei segnali costanti a valore nullo al posto dei modi mancanti, così da raggiungere una situazione in cui per ogni trial si ha esattamente lo stesso numero di modi. In questo modo però le IMFs derivanti dal calcolo della media su tutte le prove risulteranno avere una ampiezza ridotta.

2. Fissare il numero di modi da estrarre in ogni prova ad un certo valore. In [34], e.g., viene proposto di usare la formula

$$\text{Numero Modi} = 1 + \log_2(P) \quad (3.17)$$

dove P rappresenta la lunghezza del segnale in termini di numero di punti campionati.

3. Implementare un algoritmo innovativo per il calcolo delle IMFs, leggermente diverso dalla EEMD, detto Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN).

Nella EEMD tradizionale ogni  $x^i(t)$  viene decomposta indipendentemente dalle altre I-1 realizzazioni e quindi per ognuna di esse si ottiene un residuo  $r_k^i(t)$ . Nel metodo presentato in [42], invece, viene proposto di calcolare la prima IMF come si farebbe con la EEMD classica e poi calcolare subito un unico residuo  $r_1(t)$  come:

$$r_1(t) = x(t) - \overline{IMF}_1(t) \quad (3.18)$$



A questo si può applicare la EMD per ottenere la prima IMF su un ensemble di segnali dati dalla somma di  $r_1(t)$  più diverse realizzazioni di errore gaussiano. In questo modo, facendo una media, si ottiene la seconda Intrinsic Mode Function, che sarà in generale diversa da quella che si sarebbe ottenuta con la EEMD, e che viene indicata con  $\widetilde{IMF}_2$ . Il residuo successivo è definito come

$$r_2(t) = r_1(t) - \widetilde{IMF}_2(t) \quad (3.19)$$

Questa procedura va avanti con il resto dei modi fintanto che non viene soddisfatto il criterio di arresto.

Definiamo  $E_j(\cdot)$  come l'operatore che, dato un segnale, restituisce il j-esimo modo ottenuto tramite EMD. Sia poi  $w^i(t)$  un errore bianco distribuito come una normale standard. Il metodo CEEMDAN può essere descritto come segue:

- Decomporre con la EMD le I realizzazioni  $x(t) + \varepsilon_0 * w^i(t)$  in modo da ottenere i primi modi e calcolare

$$\widetilde{IMF}_1(t) = \frac{1}{I} \sum_{i=1}^I IMF_1^i(t) \quad (3.20)$$

- Al primo stadio (k=1) calcolare il primo residuo come in equazione ...

$$r_1(t) = x(t) - \widetilde{IMF}_1(t) \quad (3.21)$$

- Decomporre le realizzazioni  $r_1(t) + \varepsilon_1 E_1(w^i(t))$  fino ad ottenere la loro prima Intrinsic Mode Function, e definire il secondo modo:

$$\widetilde{IMF}_2(t) = \frac{1}{I} \sum_{i=1}^I E_1 \left( r_1(t) + \varepsilon_1 E_1(w^i(t)) \right) \quad (3.22)$$

- Per  $k=2, \dots, K$  calcolare il k-esimo residuo come

$$r_k(t) = r_{k-1}(t) - \widetilde{IMF}_k(t) \quad (3.23)$$

- Definire il (k+1)-esimo modo come

$$\widetilde{IMF}_{(k+1)}(t) = \frac{1}{I} \sum_{i=1}^I E_1 \left( r_k(t) + \varepsilon_k E_k(w^i(t)) \right) \quad (3.24)$$

- Andare al punto 4 per il prossimo valore di k.

Secondo Torres *et al.* [42] l'applicazione del loro metodo comporta diversi vantaggi rispetto all'EEMD tradizionale. Questo algoritmo infatti non solo permetterebbe di ottenere sempre lo stesso numero di modi per ogni trial, ma consentirebbe anche di:

- ridurre i tempi di calcolo riducendo il numero di iterazioni di sifting necessarie per estrarre ogni IMF
- avere una maggiore completezza, ovvero una minore ampiezza dell'errore di ricostruzione del segnale
- migliorare la separazione del contenuto in frequenza dei diversi modi, ovvero avere delle PSD (Power Spectral Density) meno sovrapposte.

In Figura 3.6 viene mostrato il risultato ottenuto applicando la CEEDMAN al profilo in Equazione (3.13). In questo sono stati usati gli stessi valori utilizzati precedentemente per la EEMD, ovvero  $I=100$  trials e un'ampiezza  $e = 0.2$  volte la deviazione standard del segnale  $x(t)$ . Anche nel caso della CEEDMAN le performance dipendono dalla particolare combinazione di valori scelti per questi due parametri. In questo caso particolare, a parità di condizioni, l'algoritmo ha generato un numero ancora più grande di Intrinsic Mode Functions: se ne possono contare 17, con l'aggiunta del residuo finale.

Oltre a non risolvere il problema del mode mixing, l'algoritmo proposto da Torres *et al.* [42] risulta anche essere dotato di scarsa efficienza dal punto di vista computazionale. Questo rende la CEEMDAN una tecnica poco raccomandabile per il monitoraggio in tempo reale basato sull'analisi di segnali.

In Figura 3.7 vengono mostrati i boxplot relativi al numero di iterazioni di sifting eseguite dai due algoritmi EEMD e CEEMDAN per estrarre ciascuna delle IMFs. Stando a questi risultati è possibile affermare che, contrariamente a quanto affermato da Torres *et al.* [42] la CEEMDAN non riduce il numero di iterazioni necessarie per il calcolo di ciascuna IMF. Al contrario i due algoritmi sembrano eseguire grosso modo lo stesso numero di iterazioni, con l'eccezione di alcuni casi in cui questo numero diventa molto grande per la CEEMDAN.

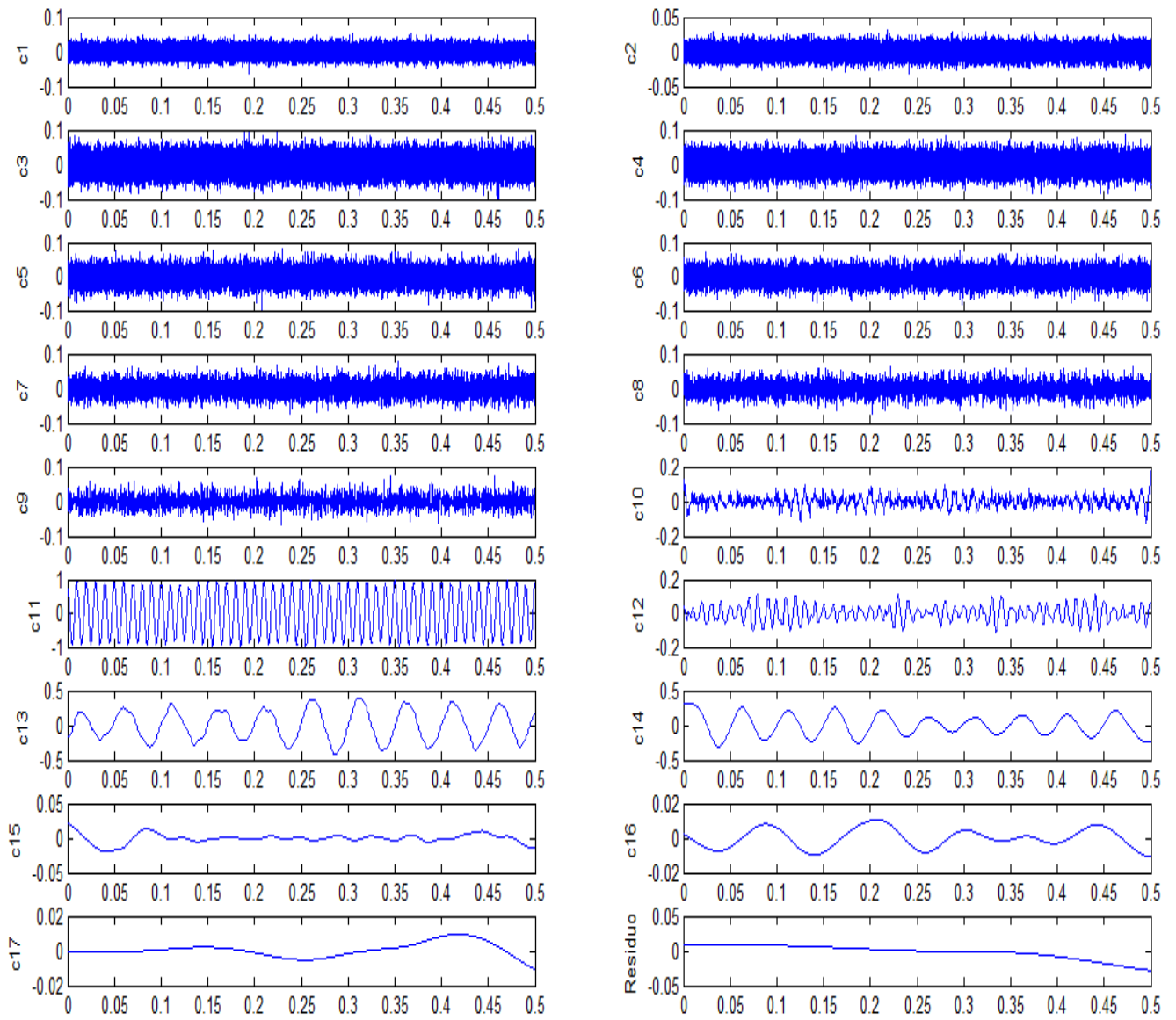
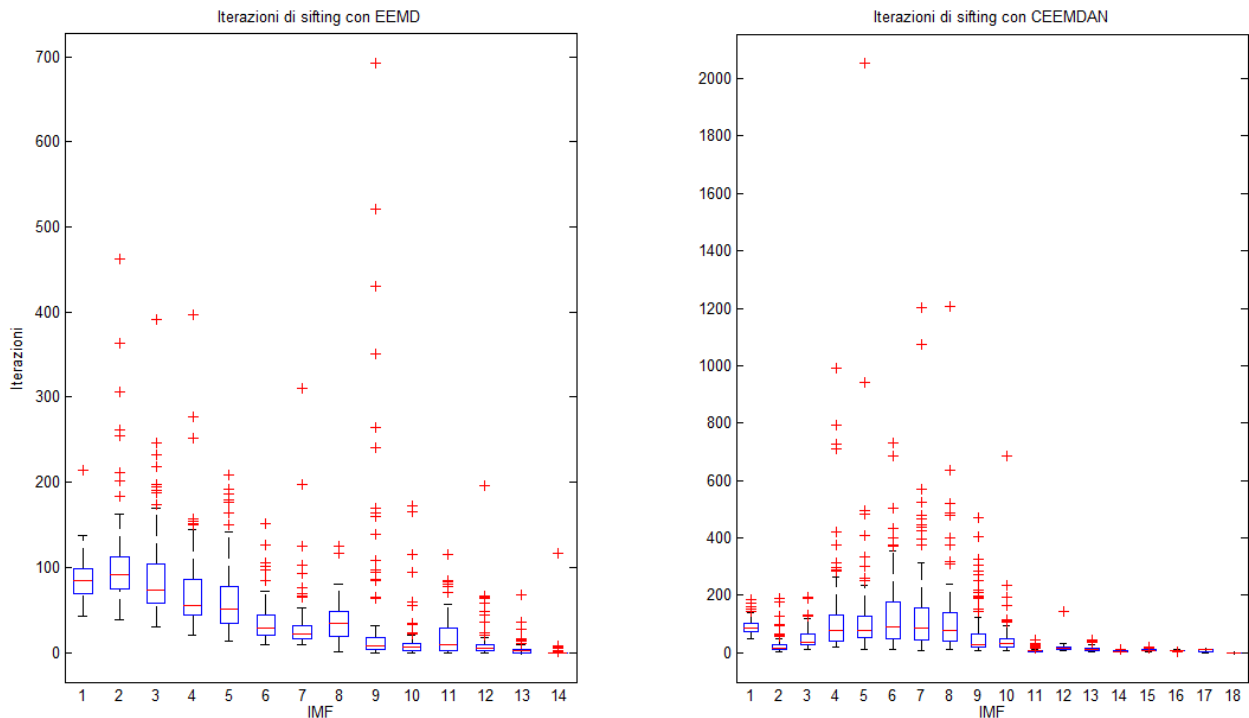


Figura 3.6 CEEMDAN del segnale in Equazione (3.13)



**Figura 3.7** Numero di iterazioni eseguite per il calcolo delle IMFs rispettivamente dalla EEMD (a sinistra) e CEEMDAN (a destra)

### 3.7 Il metodo delle Combined Mode Functions

Il mode mixing effect è il problema principale risultante dall'applicazione della Empirical Mode Decomposition per la analisi in tempo-frequenza di un segnale multiscala. Esso comporta la ripartizione di un livello tempo-frequenza in due o più Intrinsic Mode Functions, rendendole così prive di significato. Oltre alla strada della Ensemble Empirical Mode Decomposition e sue varianti, esiste un'altra via percorribile per superare questo ostacolo: il metodo delle Combined Mode Functions (CMFs) [17]. Questo metodo consiste nell'aggregare due o più IMFs contigue a formare una CMF:

$$c_s = c_i + c_{i+1} + \dots + c_{i+q} \quad (3.25)$$

Il problema principale nell'applicazione di questo metodo consiste nella scelta di un criterio in base al quale selezionare le IMFs che andrebbero combinate assieme, nonché nella scelta del numero K di CMFs da formare.

Grasso et al. [30] hanno dimostrato come la scelta del criterio in base al quale operare la suddivisione in gruppi può influenzare molto i risultati dell'analisi e l'efficienza della tecnica di monitoraggio. In questo studio infatti è stato mostrato

come il tasso di identificazione dei guasti venga influenzato dalla strategia con cui le IMFs vengono suddivise a formare le  $K$  CMFs desiderate.

Sebbene ad oggi in letteratura siano state proposte alcune regole in base alle quali effettuare questi raggruppamenti, nessuna di esse risulta essere automatica e applicabile in tutti i contesti.

Supponiamo ad esempio di prendere in considerazione ancora il segnale  $x(t)$  in Equazione (3.13). Come abbiamo visto, sebbene questo segnale sia costituito da 3 modi oscillatori, la Empirical Mode Decomposition genera 10 IMFs e un residuo. Questa sovra-decomposizione del segnale è chiaramente dovuta al fatto che la presenza del rumore comporta la nascita di nuovi punti di massimo e minimo e quindi all'insorgere del mode mixing.

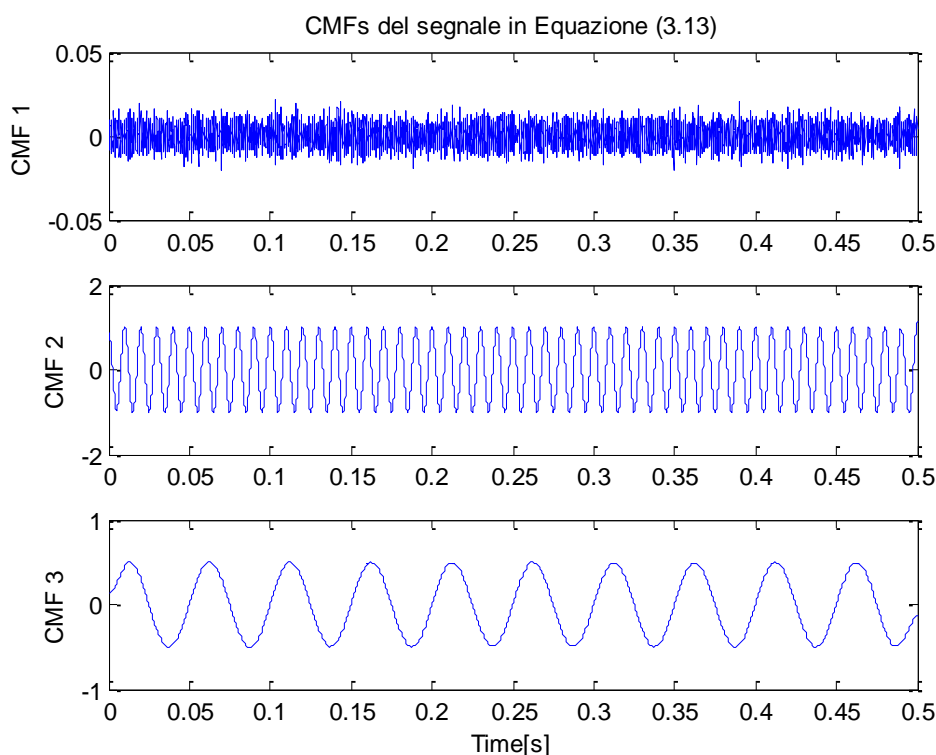
Guardando il risultato della decomposizione in Figura 3.3 è possibile notare che:

- le prime 5 IMFs hanno tutte una frequenza molto elevata e un'ampiezza limitata, il che suggerisce che potrebbero in realtà rappresentare solo la componente di rumore;
- le successive due IMFs  $c_6$  e  $c_7$  sembrano completarsi a vicenda, come se fossero in un certo senso "complementari". Se sommate potrebbero andare a rappresentare la componente sinusoidale ad alta frequenza  $x_1(t)$ .
- le ultime 3 IMFs e il residuo potrebbero essere parte del terzo ed ultimo modo oscillatorio che compone  $x(t)$ , ovvero la sinusoide a 20 Hz.

Questa analisi suggerirebbe di adottare il metodo delle Combined Mode Functions, ottenendo i risultati riportati in Figura 3.8.

Questo metodo rappresenta una soluzione più efficace ed efficiente al problema del mode mixing rispetto alla EEMD o alle sue varianti. Esso infatti non richiede di modificare il segnale in esame tramite aggiunta di rumore, richiede una sola decomposizione e inoltre permette di associare un significato fisico ad ognuno dei sotto-segnali risultanti. Nel nostro esempio, e.g., la prima CMF è molto simile a un rumore bianco, la seconda assomiglia molto alla componente  $x_1(t)$  a frequenza 100 Hz e infine la  $c_{53}$  ha un andamento che richiama molto quello della sinusoide  $x_2(t)$  a 20 Hz. Si può dunque affermare che questa tecnica è in grado di caratterizzare le diverse componenti in scala del segnale considerato.

La strategia che abbiamo adottato nell'applicare il metodo delle CMFs al nostro segnale di prova  $x(t)$  è però tutt'altro che automatica, dato che si è basata su una ispezione visiva delle IMFs ottenuta con la Empirical Mode Decomposition. In realtà sarebbe desiderabile avere a disposizione una procedura che in modo del tutto automatico generi le CMFs a partire dalle IMFs. Inoltre in questo caso il compito è stato facilitato dalla nostra conoscenza preliminare della struttura reale del segnale e quindi del numero  $K$  di modi oscillatori che lo compongono. Questo non sempre si verifica nelle applicazioni reali di analisi dei segnali.



**Figura 3.8 Combined Mode Functions per il segnale in Equazione (3.22)**

Anche i criteri suggeriti da Gao et al. [17] si basano esclusivamente su ispezioni di tipo visivo. In particolare questi ricercatori suggeriscono due criteri. Il primo consiste nell'ispezionare l'andamento delle IMFs ottenute e di accoppiare quei modi adiacenti che presentano una brusca e simultanea variazione di livello nello stesso intervallo temporale. Il secondo criterio invece è basato sul calcolo delle frequenze istantanee di ogni IMF e suggerisce di raggruppare quelle per le quali la frequenza istantanea varia improvvisamente nello stesso intervallo temporale. In [30] le IMFs derivanti dalla EMD sono state suddivise in  $K=3$  CMFs sulla base dell'analisi dell'andamento di due indicatori molto comuni, ovvero l'energia  $E_{i,j}$  associata all' $i$ -esimo modo  $c_{i,j}$  e il coefficiente di correlazione di Pearson  $\rho_{i,j}$  tra esso e il segnale  $x_j$  da cui deriva:

$$E_{i,j} = \sum_{t=1}^T c_{j,i,t}^2 \quad (3.26)$$

$$\rho_{i,j} = \frac{1}{T} \frac{\sum_{t=1}^T (x_{j,t} - \bar{x}_j)(c_{j,i,t} - \bar{c}_{j,i})}{\sqrt{\sum_{t=1}^T (x_{j,t} - \bar{x}_j)^2} \sqrt{\sum_{t=1}^T (c_{j,i,t} - \bar{c}_{j,i})^2}} \quad (3.27)$$

In molte applicazioni reali la scelta di come combinare insieme le IMFs di un segnale viene fatta con l'obiettivo di suddividerle in due gruppi e poi scartare uno dei due, in modo tale da isolare il segnale dalla sua componente di rumore o di trend.

Rato et al. [31] hanno indicato un valore soglia per il coefficiente di correlazione di Pearson al di sopra della quale la IMF viene considerata significativamente correlata col segnale  $x_j$ . Questa soglia viene calcolata come segue:

$$\lambda = \frac{\max(\rho_{i,j})}{10} \quad (3.28)$$

Liang *et al.* [25], e.g., hanno utilizzato una tecnica particolare per estrapolare la componente a bassa frequenza di un segnale e quindi eliminare la componente di rumore. Questa consiste nel decomporre il segnale in  $N$  IMFs tramite la Empirical Mode Decomposition. Successivamente per ogni valore di  $k=1,2,\dots,N$  il segnale viene parzialmente ricostruito sommando le prime  $K$  IMFs e quindi vengono calcolate la media e la deviazione standard campionaria del segnale ricostruito. A questo punto Liang et al. [25] suggeriscono di eseguire un t-Test con ipotesi nulla di media uguale a zero. Si identificherà un valore di  $k^*$  che è il minimo  $k$  per cui il t-Test conduce a rifiutare l'ipotesi nulla con livello di confidenza  $\alpha$ . La somma delle IMFs da  $K^*$  in poi costituisce la parte a bassa frequenza del segnale, e quindi può essere considerata come un'unica CMF. Le restanti IMFs invece costituiscono la parte a bassa frequenza del segnale, rappresentativa della componente di rumore, e vanno quindi scartate.

Boutana *et al.* [23] hanno adottato una tecnica differente per distinguere le IMFs significative da quelle ridondanti, basata sulla distanza di Minkowski tra una IMF e il segnale da cui è stata estratta. Gli autori affermano che il valore della distanza sarà grande per le IMFs "ridondanti", ovvero ad alta frequenza, mentre risulterà più basso per IMFs "appropriate", ovvero a bassa frequenza.

Un altro metodo per stimare la similarità tra le IMFs è quello proposto da Komaty *et al.* [45]. Questo si basa sulla stima della Probability Density Function (PDF) del segnale originario e delle  $n$  IMFs estratte con l'EMD. Questa stima viene fatta tramite l'utilizzo di uno stimatore di densità kernel.

Una volta stimate tutte le PDFs, queste vengono considerate come dei sottoinsiemi di uno stesso spazio Euclideo e il grado di similarità è misurato dalla distanza di Hausdorff (HD).

Secondo Komaty *et al.* [45] le IMFs che vanno considerate significative, ovvero rappresentative del segnale vero e proprio depurato dalla componente di rumore, sono quelle successive a quella per cui si verifica un picco nel grafico.

L'idea suggerita in [45] di caratterizzare la distribuzione dei dati in ciascuna IMF tramite la corrispondente PDF ha ispirato il metodo che viene proposto nel Capitolo 4 del presente lavoro, sebbene vi siano diversi aspetti per cui quest'ultimo si differenzia nettamente dal metodo suggerito in questo articolo. Si rimanda il lettore al Capitolo 4 per una discussione più dettagliata e approfondita.

Nel caso di studio trattato da Bu *et al.* [34] invece, l'obiettivo era quello di estrapolare i segnali corrispondenti al battito cardiaco e alla respirazione da un segnale di pressione registrato tramite un sensore piezoelettrico su cui dorme un paziente soggetto a disordini cardio-respiratori durante il sonno. Essendo noti in questo caso i range di frequenze in cui tipicamente si trovano il segnale respiratorio e quello cardiaco, Bu *et al.* [24] hanno raggruppato in una stessa CMF le IMFs che avevano un picco di frequenza appartenente a quei range.

Nella maggior parte dei casi applicativi tuttavia non sono noti i range di frequenza in cui tipicamente si trovano i modi oscillatori che compongono un segnale, e questo quindi comporta che sia difficile applicare la tecnica presentata in [24] come criterio per la formazione delle Combined Mode Functions.



## Capitolo 4

### Il metodo proposto

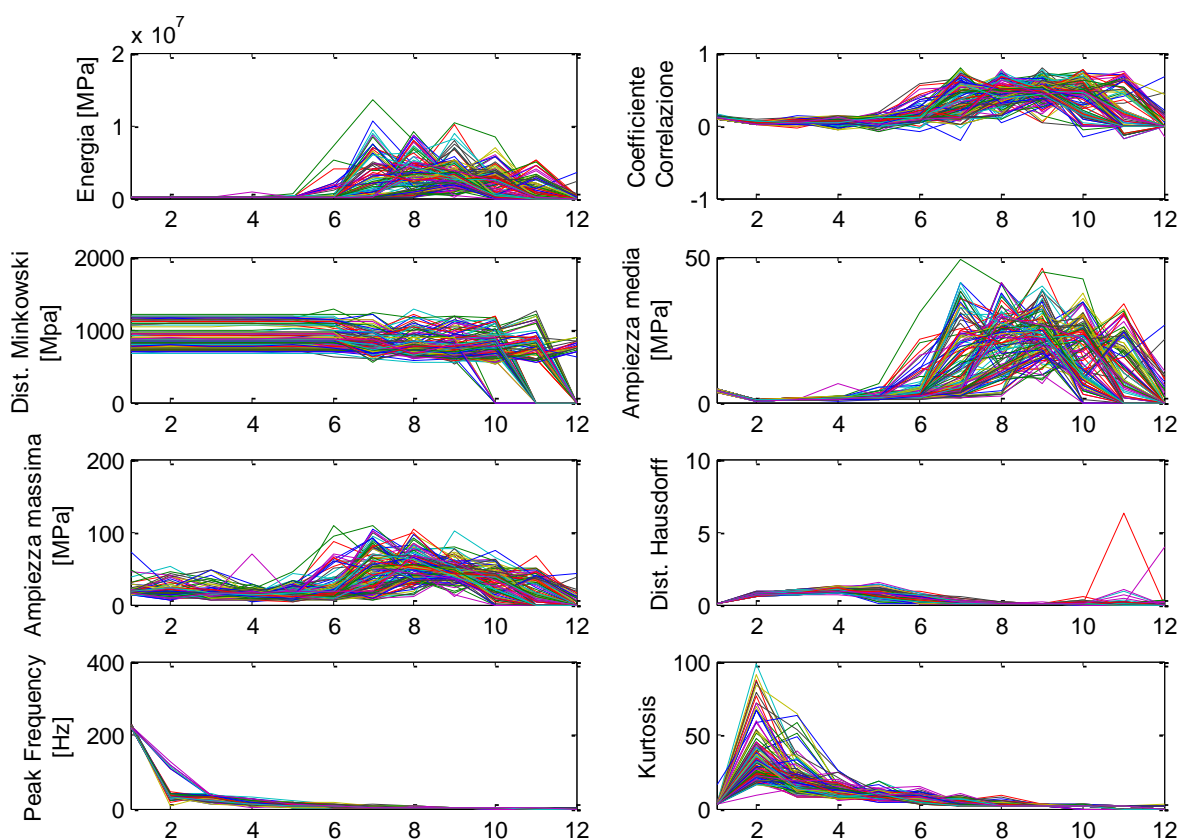
I principali problemi da affrontare quando si utilizza il metodo CMF sono essenzialmente di due tipi:

- a. Stabilire un criterio automatico sulla base del quale decidere come ripartire le IMFs in CMFs;
- b. Stabilire un criterio automatico per la scelta automatica del numero  $K$  di CMFs in cui ripartire le IMFs.

La prima problematica è stata affrontata in molte ricerche negli ultimi anni. In alcuni di questi lavori (e.g. in [17]), gli autori consigliano di realizzare la scelta relativa a come creare le CMFs raggruppando le IMFs semplicemente tramite una ispezione visiva della forma di quest'ultime: le IMFs che presentano uno stesso andamento o comunque un andamento molto simile devono essere sommate insieme per formare un'unica CMF. Questa metodologia è ovviamente molto efficace ma estremamente poco efficiente. Essa infatti presuppone la presenza di un operatore che indichi al sistema di monitoraggio come ripartire le IMFs derivanti dalla decomposizione di un segnale campionato. Questo non permetterebbe al sistema di monitoraggio di segnalare la presenza di profili anomali in maniera tempestiva. A questo proposito, tra l'altro, c'è da considerare che, come sottolineato da Grasso *et al.* [30], spesso i cicli di lavorazione che vengono realizzati da macchine come la WaterJet possono avere durate anche molto elevate e questo può comportare una elevata percentuale di processi non supervisionati.

In altri testi scientifici invece si consiglia di effettuare la suddivisione delle IMFs sulla base di indicatori sintetici, calcolati uno per ogni funzione. Gli indicatori più frequentemente utilizzati sono:

- Energia del segnale, come in Equazione (3.26);
- Coefficiente di correlazione lineare di Pearson tra la IMF e il segnale da cui deriva, come in Equazione (3.27);
- La distanza di Minkowski tra la IMF e il segnale da cui deriva, considerati come dei vettori  $P$ -dimensionali;
- L'ampiezza media o massima;
- L'indice di Kurtosis;
- La distanza di Hausdorff tra le Probability Distribution Functions stimate per la IMF e per il segnale da cui essa deriva;



**Figura 4.1** Andamento dei principali indicatori per le IMFs dei profilo WJ IC

- La peak frequency, ovvero la frequenza di picco nello spettro di frequenze ottenuto tramite Trasformata di Fourier.

In Figura 4.1 viene riportato l'andamento che questi indici assumono per i 131 profili WJ in controllo del dataset in esame.

Dall'analisi di questi grafici è possibile vedere come non tutti gli indicatori siano concordi nel suggerire la stessa ripartizione, così come non è semplice capire in maniera immediata e univoca quali IMFs andrebbero raggruppate insieme.

Ridurre una funzione come le IMFs al valore di un solo indicatore sintetico rappresenta probabilmente una strategia deficitaria, in quanto comporta una perdita di informazione notevole. Infatti potrebbe ad esempio capitare che due segnali abbiano forme e andamenti completamente diversi ma per i quali un certo indicatore (e.g. l'energia) assuma lo stesso valore e quindi li consideri del tutto uguali.

Al fine di superare questo ostacolo, in questo lavoro proponiamo un metodo per la separazione delle IMFs che si basa sulla stima della Probability Density Function (PDF) di ognuna di esse e sul confronto tra queste. Rispetto ai metodi basati sul calcolo di indicatori, questo metodo presenta il vantaggio di considerare la distribuzione di un segnale, e quindi la perdita di informazione è minore.

La prima problematica evidenziata in questo paragrafo, ovvero la necessità di poter determinare in maniera automatica il numero  $K$  di Combined Mode Functions, non ha ricevuto attenzione nell'ambito della ricerca scientifica. Tipicamente infatti nei casi in cui il metodo delle CMFs è applicato come soluzione all'effetto mode-mixing, gli autori ipotizzano sempre di conoscere a priori il valore di  $K$ . In altri casi invece l'obiettivo è semplicemente quello di separare la parte di segnale a bassa frequenza da quella di rumore ad alta frequenza, e quindi in questi casi  $K$  viene scelto automaticamente pari a 2. Nel seguito di questo studio proporremo alcuni metodi per determinare il numero ottimale di CMFs da estrarre a partire da un segnale in maniera del tutto automatica.

#### 4.1 Il calcolo delle CMFs

In questo lavoro il calcolo delle Combined Mode Functions viene proposto per affrontare il problema della sovra decomposizione dei segnali e catturare dunque le informazioni rilevanti contenute nel segnale tramite un numero limitato di modi oscillatori. A differenza di Grasso *et al.* [30] però, in questo studio viene proposto un metodo che suddivide le IMFs in  $K$  CMFs non sulla base di uno schema predeterminato uguale per ogni segnale, bensì sulla base delle caratteristiche delle IMFs di ogni singolo segnale.

Grasso *et al.* [30] infatti stabiliscono a priori che il numero di CMFs da estrarre dai segnali di pressione fosse  $K=3$ , e poi propongono uno schema di ripartizione delle IMFs così fatto:

$$c_{s1} = c_1 \quad (4.1)$$

$$c_{s2} = \sum_{i=2}^{k1} c_i \quad (4.2)$$

$$c_{s3} = \sum_{i=k1+1}^n c_i + r_n \quad (4.3)$$

In questo schema l'unico parametro per cui va stabilito un valore è  $k_1$ . Questo viene determinato guardando l'andamento dei grafici relativi all'energia, al coefficiente di correlazione e alle rispettive deviazioni standard per ciascuna IMF. Una volta stabilito il valore di  $k_1$  questo viene mantenuto uguale per ogni segnale del dataset, indipendentemente dal numero  $n$  di IMFs derivanti dalla Empirical Mode Decomposition.

Il metodo proposto in questa tesi invece prevede la possibilità di variare lo schema di ripartizione delle IMFs da segnale a segnale. In particolare esso si basa sulla stima della Probability Distribution Function per ciascuna IMF e prevede i seguenti step:

- 1) Per ognuna delle  $n$  IMFs si stima la Probability Distribution Function (PDF) tramite uno stimatore di densità Kernel, come se ognuno dei sottosegnali fosse in realtà una sequenza temporale di  $P$  realizzazioni di una variabile aleatoria, dove  $P$  è dato dal numero di punti campionati per ogni segnale. In particolare per ognuna delle IMFs va considerato uno stesso intervallo su cui valutare la PDF, e questo intervallo può avere per esempio come estremi il valore minimo e massimo del segnale da cui esse sono state estratte.
- 2) Per ogni coppia di IMFs adiacenti si calcola la distanza euclidea tra le PDFs.
- 3) Detto  $K$  il numero di CMFs che si intende creare, si considerano i  $K-1$  valori più grandi di distanza tra coppie di PDFs. Non si considera la distanza tra la PDF dell'ultima IMF e quella del residuo. Quest'ultimo infatti viene sempre aggregato all'ultima IMF.
- 4) Questi  $K-1$  punti rappresentano i punti in corrispondenza dei quali verrà eseguito un "taglio" tra le IMFs.
- 5) Tutte le IMFs che si trovano tra uno dei suddetti punti di taglio e il successivo vengono sommate per creare una CMF.

Questo metodo si ispira in parte a quello presentato da Komaty *et al.* [44], ma presenta diversi aspetti innovativi che lo rendono del tutto originale. Sebbene entrambi i metodi, infatti, si basino sulla stima della PDF di ogni IMF tramite stimatore kernel, il metodo qui proposto esegue dei confronti a coppie tra le IMFs stesse, e quindi non va a confrontare ogni IMF con in segnale originale. Inoltre la metrica utilizzata per il calcolo della distanza è quella Euclidea, e non di Hausdorff. Infine questo metodo prevede la possibilità di selezionare il

numero  $K$  di CMFs che si vogliono realizzare, e non si limita quindi a separare le IMFs “ridondanti” da quelle “significative” come in [45].

#### 4.2 Determinare il numero ottimale $K$ di CMFs

Nell'applicare del metodo delle Combined Mode Functions, oltre a definire una strategia per il raggruppamento delle IMFs in CMFs, è anche necessario poter determinare in maniera automatica il numero ottimale di CMFs da estrarre. Finora infatti abbiamo sempre supposto di conoscere il valore ottimale di  $K$ , considerandolo quindi come un dato di input dell'algoritmo. Nella realtà applicativa tuttavia l'utente di un sistema di controllo basato sulla Empirical Mode Decomposition non può sapere in generale qual è il numero esatto di modi oscillatori in cui il segnale dovrebbe essere scomposto. Si rende quindi necessario individuare un metodo per determinare in modo automatico il numero ottimale  $K$  di CMFs.

A tal fine possiamo prendere in prestito alcune nozioni dal mondo degli algoritmi di clustering. Il clustering è un processo il cui scopo è quello di scoprire una struttura ignota nascosta nelle osservazioni di un dataset. Esistono diversi algoritmi di clustering basati su logiche diverse, ma lo scopo è sempre quello di ripartire le  $n$  osservazioni  $p$ -dimensionali di un dataset in un certo numero  $K$  di gruppi (*clusters*), di modo tale che i punti appartenenti a uno stesso gruppo risultino essere più simili tra di loro che ai punti appartenenti a clusters diversi [51].

L'idea che sta alla base degli algoritmi di clustering dunque non è poi molto diversa da quella che sottosta il metodo delle CMFs: anche in questo caso infatti si ha la necessità di ripartire delle osservazioni (le IMFs estratte tramite EMD)  $P$ -dimensionali (dove  $P$  è la lunghezza del segnale oppure, nel caso del nostro algoritmo, il numero di intervalli della sua Probability Distribution Function) in un certo numero  $K$  di CMFs.

La differenza sostanziale sta nel fatto che nel nostro caso le osservazioni sono numerate da 1 a  $n$ , e un'osservazione  $i$ -esima non può appartenere a un cluster in cui non vi sia almeno una tra la IMF  $C_{i-1}$  e la IMF  $C_{i+1}$ , a meno che essa non costituisca da sola un cluster a sé stante. Fatta eccezione per questa differenza, anche nel caso del metodo delle CMFs lo scopo è quello di raggruppare (per poi sommare tra di loro) le IMFs che sono simili perché in realtà non sono altro che parti di uno stesso modo oscillatorio che è stato sovra-decomposto a causa del fenomeno del mode-mixing.

Molti algoritmi di clustering, come per esempio quello delle  $k$ -medie o dei  $k$ -medoidi hanno bisogno che gli venga fornito in input, oltre alla matrice  $n \times P$  delle osservazioni, anche il numero  $K$  di clusters in cui ripartirle, proprio come nel caso del nostro algoritmo di separazione delle IMFs. Il risultato del clustering può dipendere in modo marcato dalla scelta del valore di  $K$ .

In letteratura è possibile trovare una gran mole di lavori riguardanti metodi per la selezione del numero “corretto” di clusters in un dataset (si parla di *cluster validity methods* oppure di *cluster validation*). Milligan *et al.* [52] hanno effettuato una valutazione di ben 30 procedure diverse, utilizzando degli algoritmi di clustering di tipo gerarchico.

Una semplice regola che viene spesso utilizzata in letteratura è la cosiddetta *Rule of thumb* [53], ovvero una regola empirica che suggerisce di scegliere K come

$$K \approx \sqrt{n/2} \quad (4.4)$$

dove n rappresenta il numero di osservazioni che compongono il dataset. Nel nostro caso, considerando i 131 profili in controllo campionati, si va da un minimo di n=9 a un massimo di n=12 IMFs da clusterizzare. Applicando la rule of thumb quindi si otterrebbero valori di K oscillanti tra 2,12 e 2,45. In realtà sarebbe più opportuno applicare questa regola in casi di database più popolati, ovvero con un numero superiore di osservazioni (almeno qualche decina), ma comunque questi risultati possono essere presi come un’indicazione di massima su quale sia l’intorno di valori in cui cercare il K “ottimale”.

Più in generale però possiamo affermare che il processo di cluster validity si basa sul calcolo di un indice per diversi valori di K. In particolare, dati un dataset n×P, uno specifico algoritmo di clustering e un certo range di valori per il numero di clusters K, la procedura di cluster validity prevede i seguenti passi [54]:

- L’algoritmo di clustering viene ripetuto sul dataset per diversi valori di K in sequenza, da un certo Kmin a un certo Kmax fissati;
- Una volta ottenuta la partizione in clusters si calcola il valore di un certo indice per ogni valore di K;
- Si plotta poi l’andamento di questo indice contro il numero di clusters K e si seleziona il valore di K per cui la partizione appare essere ottimale a seconda di come l’indice va ottimizzato.

#### 4.2.1 Internal Validity Indexes

Esistono diversi indici che sono stati proposti in letteratura. Un primo gruppo di indici viene spesso indicato con il nome di *internal validity indexes*. Nel seguito ne riportiamo alcuni.

▪ *Indice di Dunn (DI)*

Questo indice fu introdotto per la prima volta da J. C. Dunn [55] nel 1973 e ha come scopo quello di individuare clusters che siano compatti e ben separati, ovvero con una piccola varianza tra le osservazioni appartenenti a un medesimo cluster e una grande distanza tra quelle appartenenti a clusters diversi. Questo si ha quando i centroidi dei diversi clusters, ottenuti facendo una media tra tutte le osservazioni ad essi appartenenti, sono relativamente molto distanti rispetto alla varianza intra-clusters. Per un certo numero  $K$  di clusters, l'indice di Dunn può essere definito come segue:

$$DI(K) = \min_{1 \leq i \leq K} \left\{ \min_{1 \leq j \leq K, j \neq i} \left\{ \frac{\delta(C_i, C_j)}{\max_{1 \leq k \leq K} \Delta_k} \right\} \right\} \quad (4.5)$$

In questa formula  $\Delta_k$  rappresenta il diametro del  $k$ -esimo clusters e può essere calcolato in diversi modi come per esempio la massima distanza tra due qualsiasi osservazioni appartenenti al cluster, oppure la distanza media tra tutte le coppie di osservazioni ad esso appartenenti oppure ancora la distanza di ciascuna osservazione dal centroide.

$\delta(C_i, C_j)$  rappresenta invece la distanza inter-cluster tra il cluster  $i$  e quello  $j$ . Anche in questo caso valgono le stesse considerazioni fatte per  $\Delta_k$  nel senso che questa distanza può essere calcolata come la distanza massima tra due osservazioni, ognuna appartenente a un cluster, oppure come la distanza minima oppure ancora come la distanza tra i centroidi dei due clusters.

Il  $K$  ottimale secondo questo indice sarebbe quello che massimizza il valore di  $DI(K)$ . Infatti tanto più ci si avvicina alla situazione di minima varianza intra-clusters e massima distanza inter-clusters, quanto più i  $\Delta_k$  saranno piccoli e le distanze  $\delta(C_i, C_j)$  saranno grandi.

▪ *Indice di Davies-Bouldin (DBI)*

Introdotto per la prima volta da D. L. Davies e D. W. Bouldin [56] nel 1979, anche questo indice può essere utilizzato al fine di valutare le partizioni in clusters. Esso si basa sulle grandezze  $S_i$  e  $M_{i,j}$  che sono date rispettivamente dalla media delle distanze euclidee delle osservazioni appartenenti al cluster  $i$ -esimo e il suo centroide, e dalla distanza euclidea tra i centroidi dei clusters  $i$  e  $j$ . A partire da queste grandezze è possibile calcolare i valori di  $R_{i,j}$  per ogni coppia di clusters e di  $D_i$  per ogni cluster  $i$ -esimo:

$$S_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \|x_j - A_i\| \quad (4.6)$$

$$M_{i,j} = \|A_i - A_j\| \quad (4.7)$$

$$R_{i,j} = \frac{S_i + S_j}{M_{i,j}} \quad (4.8)$$

$R_{i,j}$  assume valori molto piccoli nei migliori dei casi in cui i cluster  $i$ -esimo e  $j$ -esimo risultano essere ben separati. Per questo per ognuno dei  $K$  clusters viene considerato il caso peggiore:

$$D_i = \max_{j:i \neq j} R_{i,j} \quad (4.9)$$

Facendo una media di tutti i casi peggiori per ognuno dei  $K$  clusters si ottiene il Davies-Bouldin index per l'intera partizione:

$$DBI(K) = \frac{1}{K} \sum_{i=1}^K D_k \quad (4.10)$$

Ovviamente il  $K$  ottimale risulta essere quello in corrispondenza del quale DBI è minimo.

Nel nostro caso l'applicazione di questo indice per il processo di cluster validity non è una buona idea. Infatti quando il valore di  $K$  cresce e si avvicina al numero  $n$  di PDFs da clusterizzare molte delle osservazioni del dataset costituiscono dei cluster a sé stanti. In queste situazioni ovviamente le  $S_i$  per questi clusters risulteranno essere nulle e quindi necessariamente l'indice di Davies-Bouldin sarà più basso che per valori di  $K$  più piccoli.

- *Coefficiente di Silhouette*

Il coefficiente di Silhouette, calcolabile per ogni record di un dataset, è un'altra misura della bontà di una partizione in clusters. Esso rappresenta una misura di quanto l'osservazione sia simile alle osservazioni appartenenti al suo stesso cluster se messa a confronto con quelle degli altri  $K-1$  clusters [57]. Questo coefficiente viene calcolato come segue:

$$S_j = \frac{b_j - a_j}{\max(a_j, b_j)} \quad (4.11)$$

dove  $a_j$  è la distanza media di  $x_j$  dalle altre osservazioni che fanno parte del suo stesso cluster, mentre  $b_j$  è la minima distanza media di  $x_j$  dalle osservazioni



appartenenti a un cluster diverso dal suo, minimizzata sui restanti K-1 clusters. Facendo una media su tutte le n osservazioni del dataset, è possibile assegnare un coefficiente di Silhouette all'intera partizione.

Il coefficiente di Silhouette può assumere valori nel range [-1;1]. Un valore vicino a 1 indica che l'osservazione  $x_j$  è ben assegnata, ovvero risulta essere simile a quelle dello stesso cluster e dissimile da quelle di clusters diversi. Di conseguenza se tutte le n osservazioni del dataset presentano un elevato coefficiente di Silhouette allora la partizione è buona. Se al contrario molte osservazioni hanno un valore di coefficiente di Silhouette basso o addirittura negativo, questo significa che probabilmente abbiamo selezionato un valore di K troppo alto o troppo basso. In questo caso infatti il coefficiente di silhouette indica che molte osservazioni  $x_j$  sono male assegnate, in quanto risultano più vicine a un cluster vicino che non alle altre osservazioni dei loro cluster di appartenenza.

Uno dei difetti principali di questo indice è che difficilmente vengono penalizzate scelte di K troppo alto. In particolare, quando K si avvicina a n, molte osservazioni diventano clusters a sé stanti. In questa situazione ovviamente si annulla la distanza di  $x_j$  dalle altre osservazioni del cluster (dato che non esistono altre osservazioni per quel cluster) e il coefficiente di Silhouette assume valore 1 per ogni cluster "mono-osservazione". Il coefficiente di Silhouette medio per la partizione intera sarà tendenzialmente più alto che in casi di K più basso.

#### 4.2.2 Sum-of-squares-based indexes

Un secondo gruppo di indici proposti in letteratura per la cluster validation si basa sul calcolo di due grandezze che sono la Sum of Squares Within Clusters (SSW) e la Sum of Squares Between Clusters (SSB). Per questo motivo si parla di *sum-of-squares-based indexes*. La SSW rappresenta la varianza all'interno dei singoli gruppi, mentre la SSB rappresenta la varianza intra-gruppi.

$$SSW(K) = \sum_{i=1}^K \sum_{j \in C_i} \|x_j - A_i\| \quad (4.12)$$

$$SSB(K) = \sum_{i=1}^K n_i \|A_i - \bar{x}\| \quad (4.13)$$

Nell'Equazione (4.16)  $n_i$  rappresenta il numero di osservazioni appartenenti al cluster i-esimo,  $A_i$  il suo centroide e  $\bar{x}$  il valore medio di tutti i record del dataset. La somma di SSW e SSB eguaglia la Total Sum-of-Squares del dataset:

$$SST(K) = \sum_{j=1}^n \|x_j - \bar{x}\| \quad (4.14)$$

Di seguito riportiamo alcuni dei più importanti tra questi indici.

- *Indice di Ball e Hall*

Questo indice fu proposto per la prima volta nel 1965 [58]. Per un dato valore di  $K$ , si calcola come:

$$BH(K) = SSW / K \quad (4.15)$$

La  $SSW$  è una funzione monotona non-crescente del numero di clusters  $K$  per definizione.

- *Indice Calinski-Harabasz*

Calinski et al. [59] hanno proposto la seguente misura per la cluster validation:

$$CH(K) = \frac{SSB}{SSW} * \frac{n - K}{K - 1} \quad (4.16)$$

Esso è dato dal rapporto tra la  $SSB$  e la  $SSW$ , ciascuna divisa per i corrispettivi gradi di libertà. Il valore ottimale di  $K$  è quello che massimizza le differenze successive.

- *Indice di Hartigan*

L'indice di Hartigan [60], per un certo valore di  $K$ , è calcolato come:

$$H(K) = (n - K - 1) * \frac{SSW(K) - SSW(K + 1)}{SSW(K + 1)} \quad (4.17)$$

Dato che per costruzione la  $SSW$  è una funzione monotona non-crescente del numero di clusters  $K$ , questo rapporto misura il miglioramento relativo che si ha in termini di diminuzione della  $SSW$  quando si passa da una partizione con  $K$  a una con  $K+1$  clusters. Il coefficiente correttivo che lo pre-moltiplica ( $n-K-1$ ) può essere interpretato come un fattore di penalizzazione per alti valori di  $K$ . Il valore ottimale di  $K$  è quello successivo al punto di massimo della funzione  $H(K)$ .

Un ultimo metodo che si può menzionare per la determinazione del numero ottimale di clusters in un dataset è il cosiddetto *elbow method*. Questo metodo

consiste nel cercare il valore di  $K$  per cui si ha un “ginocchio” o “gomito” nel grafico che plotta la  $SSW$  contro il numero di clusters stesso. Per definizione di  $SSW$  in questo grafico si registrerà un calo ogni volta che si passerà da  $K$  a  $K+1$  clusters. Il metodo del ginocchio consiste nell’individuare quel valore di  $K$  tale per cui aggiungendo un altro cluster non si ottiene un netto miglioramento in termini di  $SSW$ , ovvero non si ottiene un grande vantaggio in termini di compattezza media dei clusters. Questo metodo può essere fatto risalire a Thorndiken [61]. E’ un metodo semplice ma allo stesso tempo intuitivo ed efficiente.

Alcuni studiosi, come ad esempio Tibshirani *et al.* [62], hanno proposto dei metodi per individuare in maniera automatica il gomito di questa curva. Il metodo che viene proposto in questo lavoro è il seguente: dato il grafico di  $SSW$  in funzione di valori di  $K$  che variano da un certo valore  $K_{min}$  (tipicamente pari a 1) a un certo valore  $K_{max}$ , il gomito della curva si trova in corrispondenza di quel valore  $K^*$  per cui è massima la distanza verticale tra la curva stessa e la retta che congiunge  $SSW(K_{min})$  a  $SSW(K_{max})$ .

## Capitolo 5

### Analisi dei risultati: simulation study

#### 5.1 Il calcolo delle CMFs

In questo capitolo vengono mostrati i risultati ottenuti applicando i metodi per il calcolo delle CMFs e per la selezione del valore ottimale di  $K$  descritti nel Capitolo 4 a segnali simulati.

In particolare si presentano i risultati ottenuti applicando i suddetti metodi ad un dataset di  $M=100$  profili multiscala simulati che presentano caratteristiche analoghe a quelle dei segnali di pressione WJ in controllo, eccezion fatta per una minore variabilità da profilo a profilo. Questi segnali di test risultano dalla somma di:

- una componente di errore bianco di ampiezza limitata  $err(t)$ ;
- una componente, che chiameremo  $x_1(t)$ , ad alta frequenza ma transitoria, i.e. localizzata solo in corrispondenza di determinati intervalli temporali;
- una componente a bassa frequenza  $x_2(t)$ .

Figura 5.1 e Figura 5.2 mostrano rispettivamente l'insieme delle  $M$  curve così generate, e il particolare di uno dei segnali di test  $x(t)$  con le sue componenti.

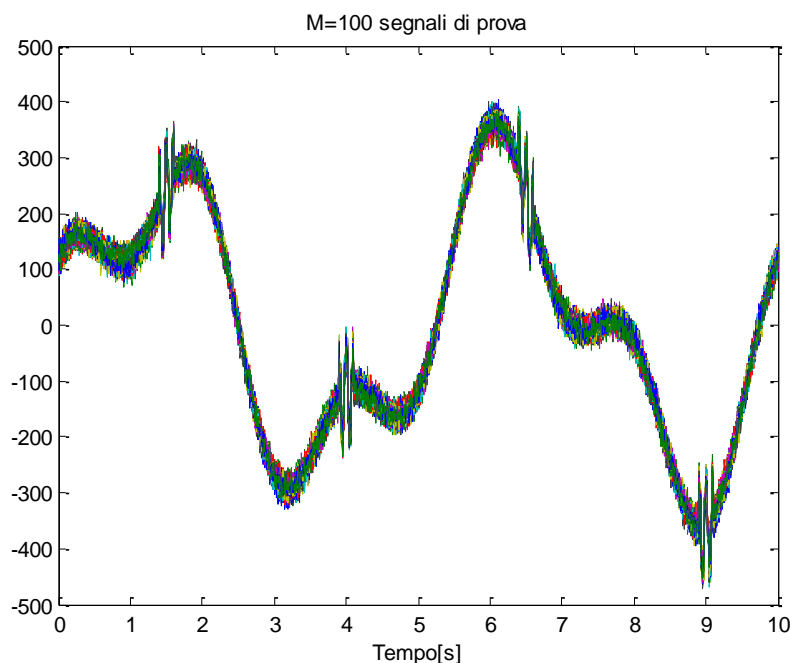
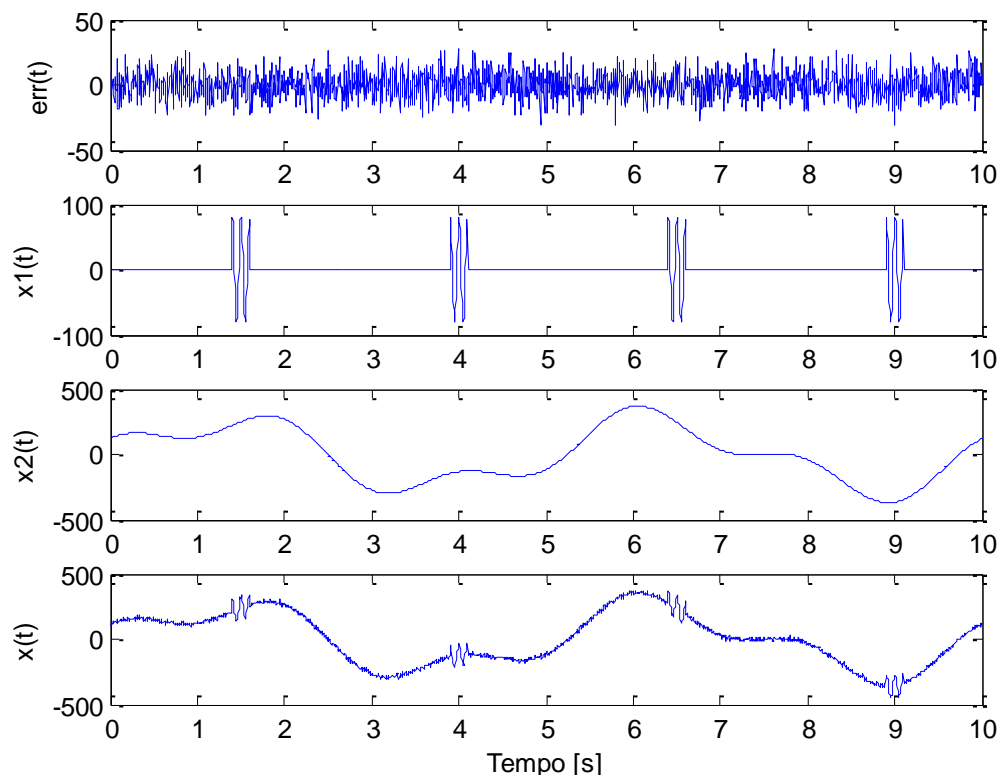


Figura 5.1 Insieme dei segnali del dataset simulato



**Figura 5.2** Uno dei segnali simulati  $x(t)$  e le sue componenti in frequenza

Le Figure da 5.3 a 5.7 mostrano gli step seguiti dall' algoritmo per creare le CMF nel caso in cui ipotizziamo di conoscere a priori che il numero  $K$  di livelli tempo-frequenza da individuare sia 3.

In particolar modo in Figura 5.3 vengono riportate le IMFs che la EMD estrae a partire dal profilo in esame. Una semplice analisi visiva porta ad affermare che la prima IMF, che rappresenta una componente ad altissima frequenza, è seguita da 4 segnali abbastanza simili tra di loro, pressoché nulli ad eccezione di alcuni intervalli temporali. Nel passare dalla quinta alla sesta IMF, infine, si nota una netta variazione delle caratteristiche dei segnali. I modi oscillatori si fanno più regolari, con ampiezze maggiori e frequenze minori.

A partire da queste IMFs l' algoritmo proposto stima la PDF di ognuna di esse. In figura 5.4 queste curve sono mostrate in rosso e sono sovrapposte a degli istogrammi (in blu) che rappresentano le frequenze relative per ognuno degli intervallini considerati. Essi sono stati ottenuti suddividendo in 200 parti uguali l' intervallo che va dal minimo al massimo del segnale in esame.

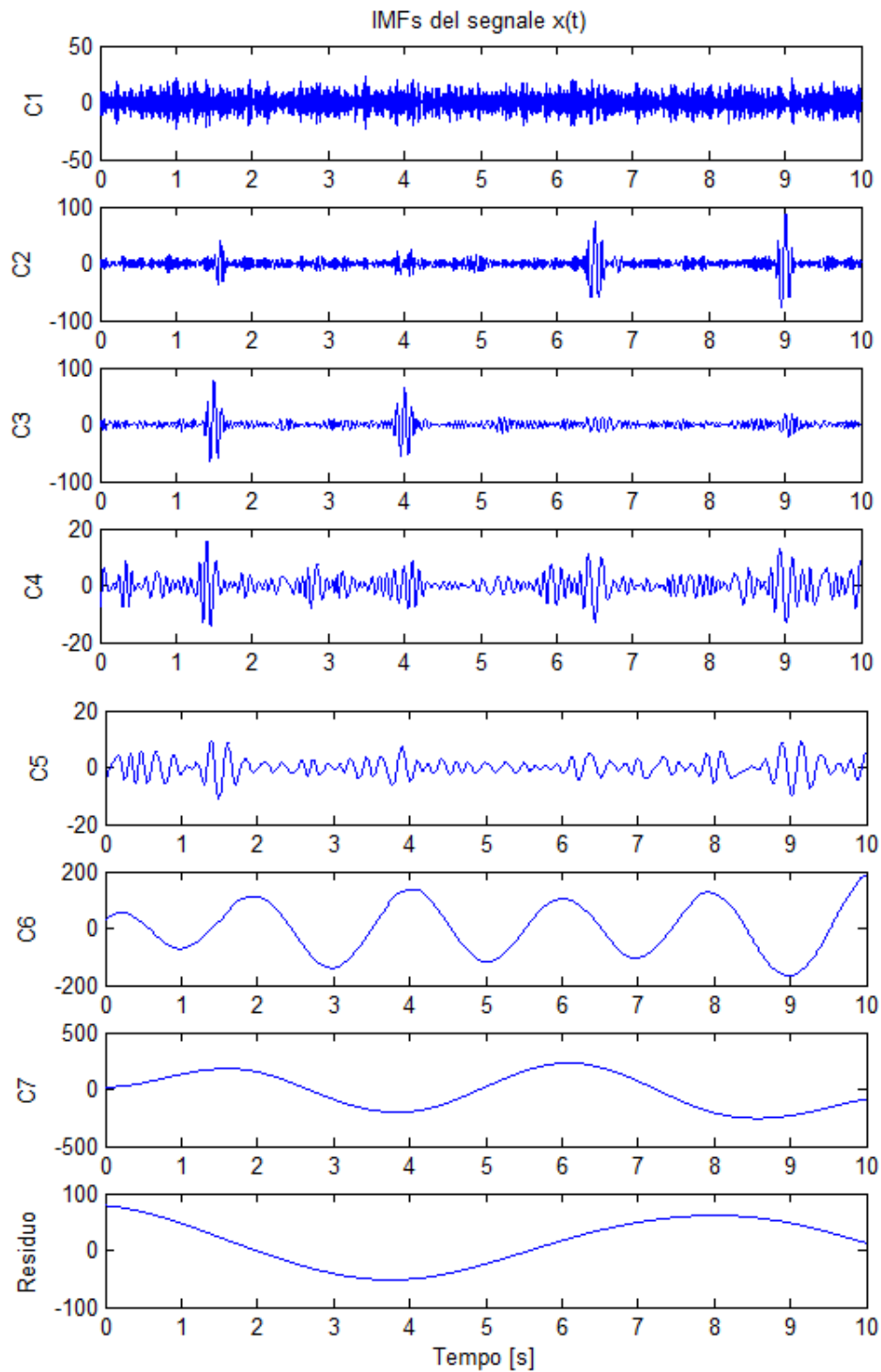


Figura 5.3 IMFs di un segnale simulato  $x(t)$

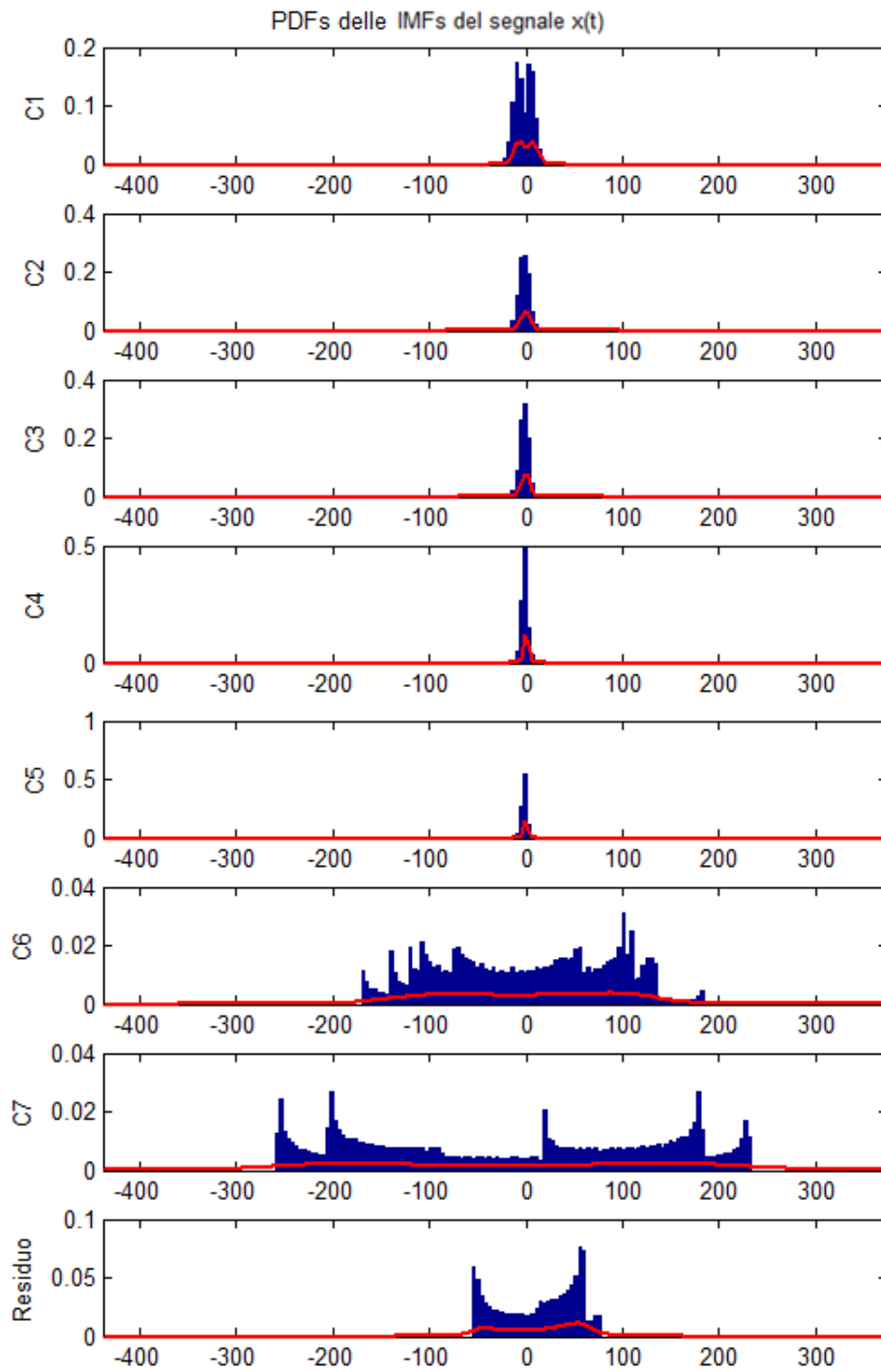
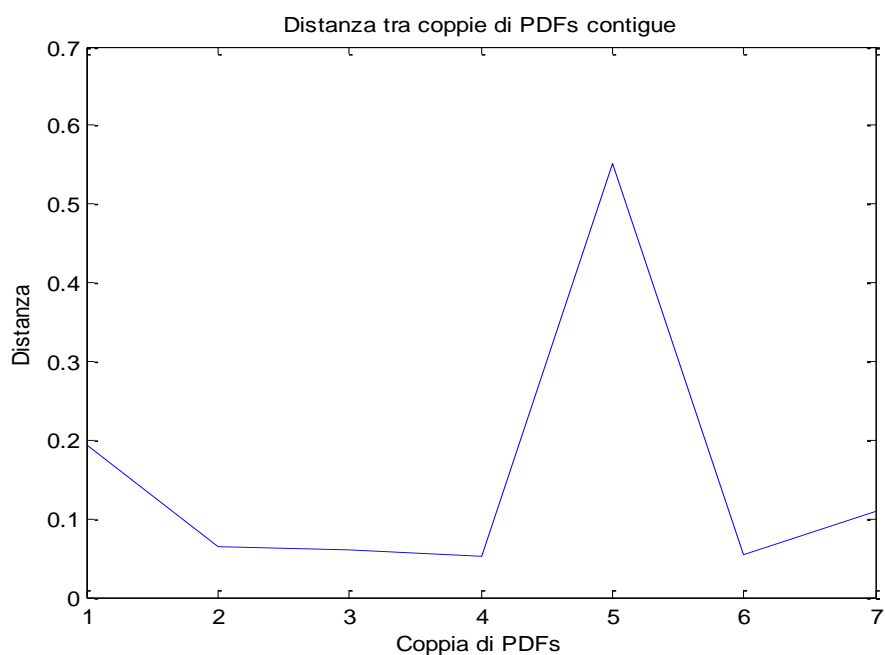


Figura 5.4 PDFs delle IMFs del segnale  $x(t)$

Una volta realizzata questa stima, l'algoritmo proposto passa a considerare ciascuna delle PDFs come un vettore di dimensione pari al numero di intervallini ricavati. Così facendo è possibile dare una misura della differenza tra le diverse PDFs semplicemente tramite una misura di distanza tra vettori. In questo lavoro è stata impiegata la distanza Euclidea. Figura 5.5 mostra l'andamento delle distanze tra coppie di PDFs contigue. In particolare modo, in corrispondenza del  $k$ -esimo valore in ascissa, viene riportata in ordinata la distanza Euclidea tra la PDF della IMF  $C_k$  e quella della IMF  $C_{k+1}$ . Dato che il metodo delle CMFs permette di accoppiare soltanto IMFs che siano contigue, il calcolo di queste distanze permette di capire quando si verifica un netto cambio di caratteristiche passando da una IMF a quella successiva. È in corrispondenza di questi "salti", infatti, che è necessario eseguire una separazione per distinguere livelli tempo-frequenza diversi. Nel caso specifico considerato, ad esempio, è netta la distanza tra la  $C_5$  e la  $C_6$ , così come tra la  $C_1$  e la  $C_2$ . Proprio queste quindi saranno le coppie di IMFs che verranno separate al fine di creare le 3 CMFs richieste. Tutte le altre coppie di IMFs, al contrario, sono pressoché simili tra di loro, come è confermato dal fatto che la distanza tra le loro PDFs è quasi nulla.



**Figura 5.5** Distanze tra PDFs contigue per il segnale  $x(t)$



In questo caso, dunque, lo schema di ripartizione ottimale è il seguente:

$$c_{s1} = c_1 \quad (5.1)$$

$$c_{s2} = \sum_{i=2}^5 c_i \quad (5.2)$$

$$c_{s3} = \sum_{i=6}^7 c_i + r_7 \quad (5.3)$$

Figura 5.6 mostra l'andamento dei livelli tempo-frequenza così ottenuti. Confrontando questi risultati con quelli di Figura 5.2 è possibile affermare che questo metodo permette di caratterizzare correttamente il contenuto informativo di questi segnali multiscala, in quanto separa efficacemente i diversi modi oscillatori che lo compongono.

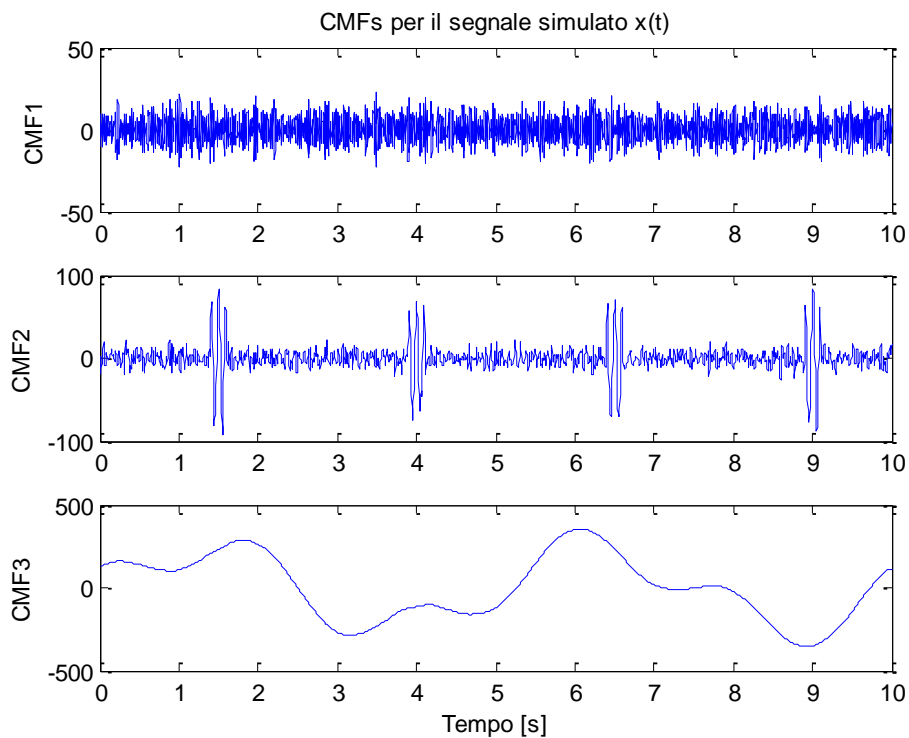


Figura 5.6 CMFs del segnale simulato  $x(t)$  con  $K=3$

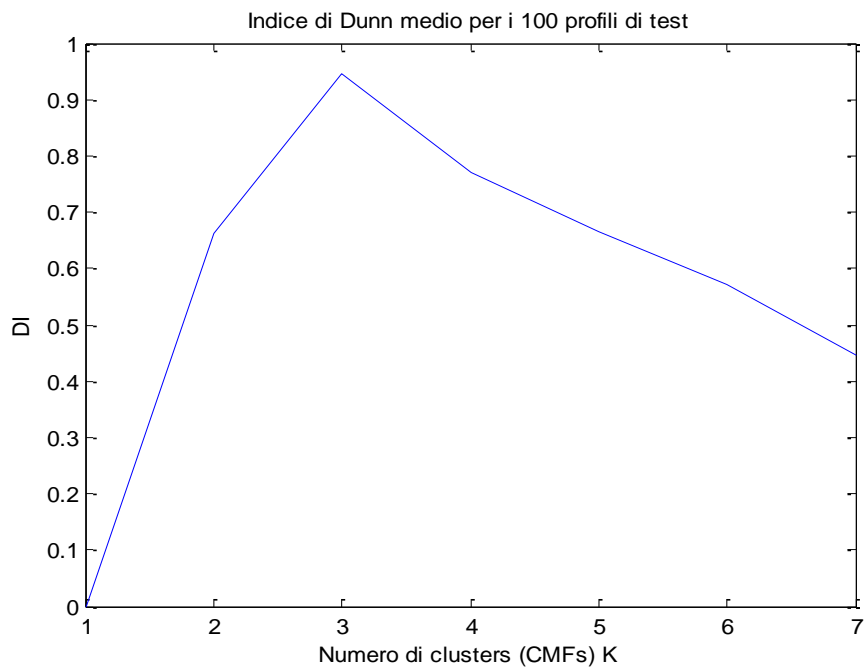
## 5.2 Scelta del K ottimale

Dato che l'utilizzo di indici diversi su uno stesso dataset comporta risultati discordanti, così come uno stesso indice può suggerire valori diversi di K per segnali diversi, è stata condotta un'analisi comparativa al fine di valutare l'affidabilità e la consistenza di questi indici. In particolare in questa sezione vengono riportati solo i risultati relativi a quelli indici che sono risultati maggiormente precisi (dove per precisione si intende una scarsa varianza dei risultati ottenuti in termini di K ottimale), ovvero:

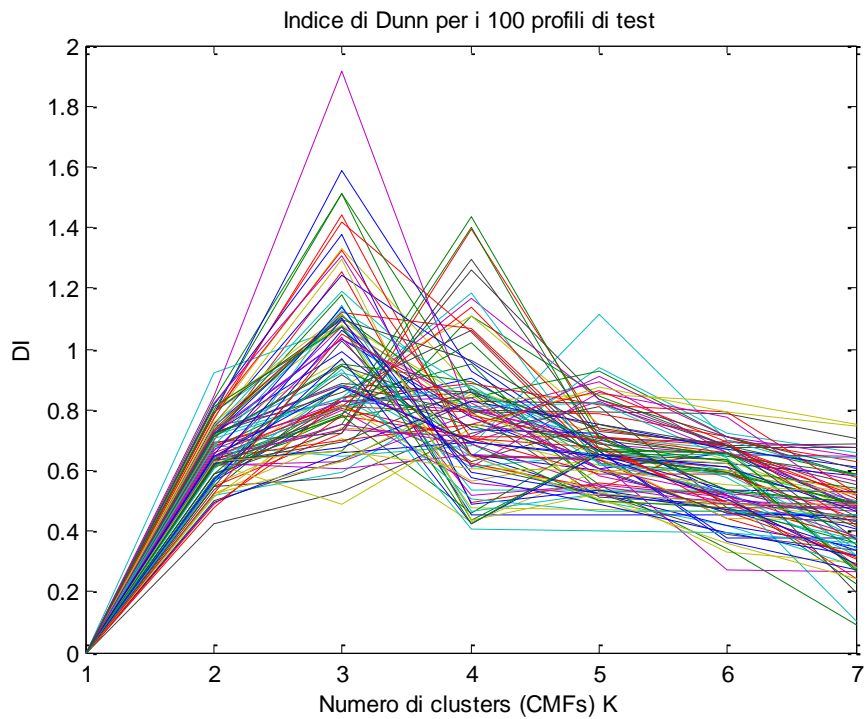
- Indice di Dunn;
- Elbow Method;
- Indice di Hartigan.

➤ *Indice di Dunn (DI)*

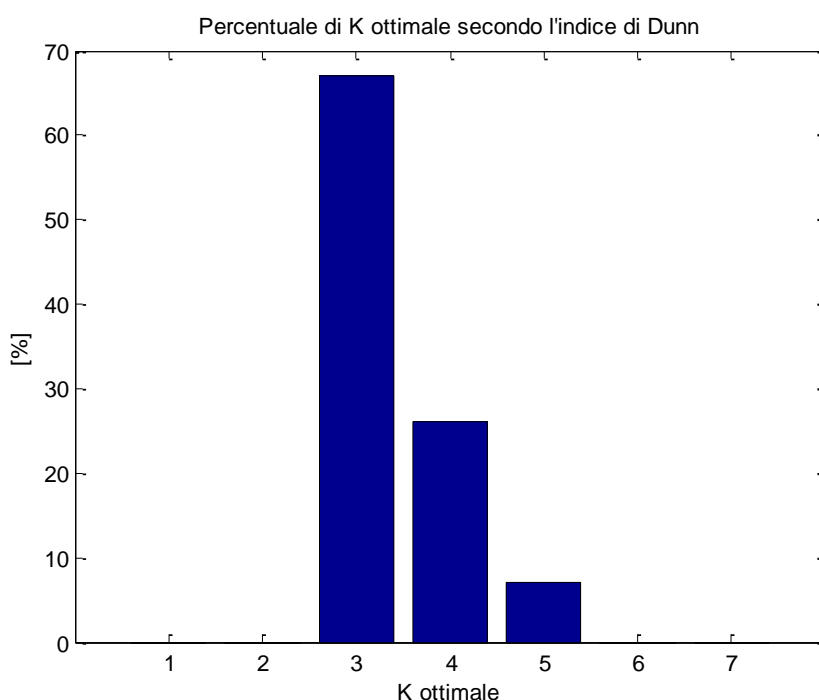
Se si considera la media dei valori che l'indice di Dunn assume per ogni segnale simulato in corrispondenza di ogni valore di K, l'andamento è quello riportato in Figura 5.7. Stando a questo grafico il valore ottimale di K sarebbe pari a 3. In realtà se si prendono in considerazione gli andamenti di questi grafici per ciascuno dei 100 segnali presi singolarmente (Figura 5.8), allora è possibile notare come i risultati non siano perfettamente consistenti, ma anzi vi sia una certa variabilità nel valore di K per il quale la curva ha un picco. Per alcuni profili il valore di K suggerito sarebbe 3, per altri 4 e per alcuni altri ancora 5. Figura 5.9 mostra, infine, la percentuale di profili che presentano un indice di Dunn massimo per ogni valore di K. Per la grande maggioranza dei profili (quasi il 70%), il numero di CMFs ottimale suggerito è 3.



**Figura 5.7** Indice di Dunn medio per il dataset simulato



**Figura 5.8** Indice di Dunn per ogni profilo del dataset simulato



**Figura 5.9** Distribuzione del K ottimale secondo l'indice di Dunn

➤ *The elbow method*

Questo metodo, che consiste nel trovare il valore di K per il quale si ha un ginocchio nel grafico della SSW, sembra essere caratterizzato da una grande stabilità. Come è possibile vedere in Figura 5.10 e Figura 5.11 infatti, l'andamento della SSW al variare di K è omogeneo per tutti i 100 segnali di test che costituiscono il dataset in esame. Queste curve sembrano appiattirsi tutte per  $K=3$ .

In effetti se si applica il metodo descritto precedentemente per il calcolo del ginocchio di una curva, si vede che nel 100% dei casi la massima distanza verticale tra la curva SSW e la retta sovrastante si ha esattamente per  $K=3$ .

➤ *Indice di Hartigan*

Se si prende in considerazione l'indice di Hartigan per la valutazione del numero K di CMFs più opportuno, questo suggerisce in modo piuttosto consistente di considerare soltanto 2 clusters, in quanto il passaggio da 1 a 2 clusters comporta il massimo miglioramento percentuale in termini di SSW della partizione.

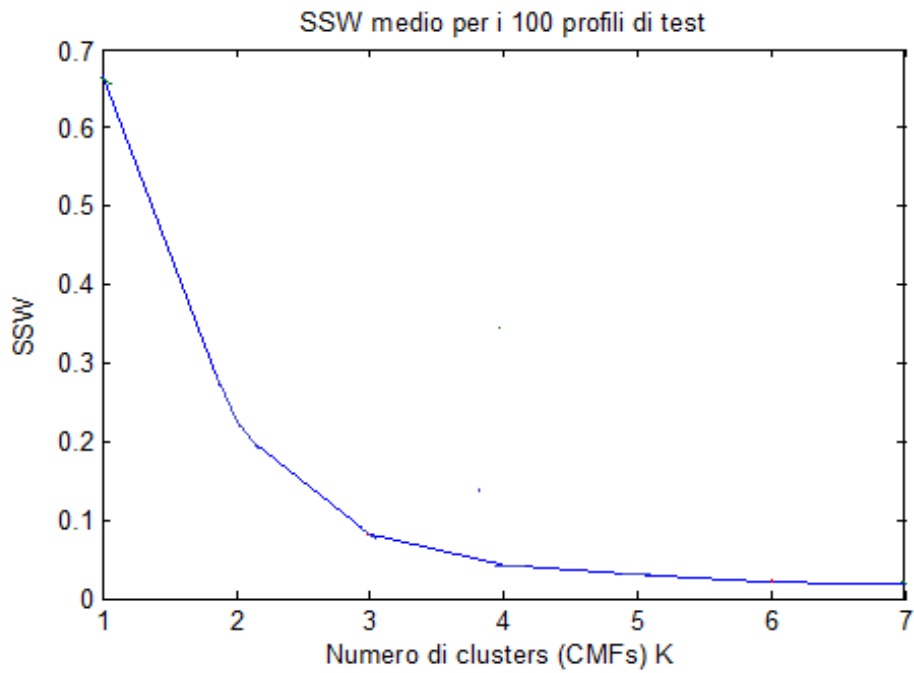


Figura 5.10 SSW medio per i segnali del dataset simulato

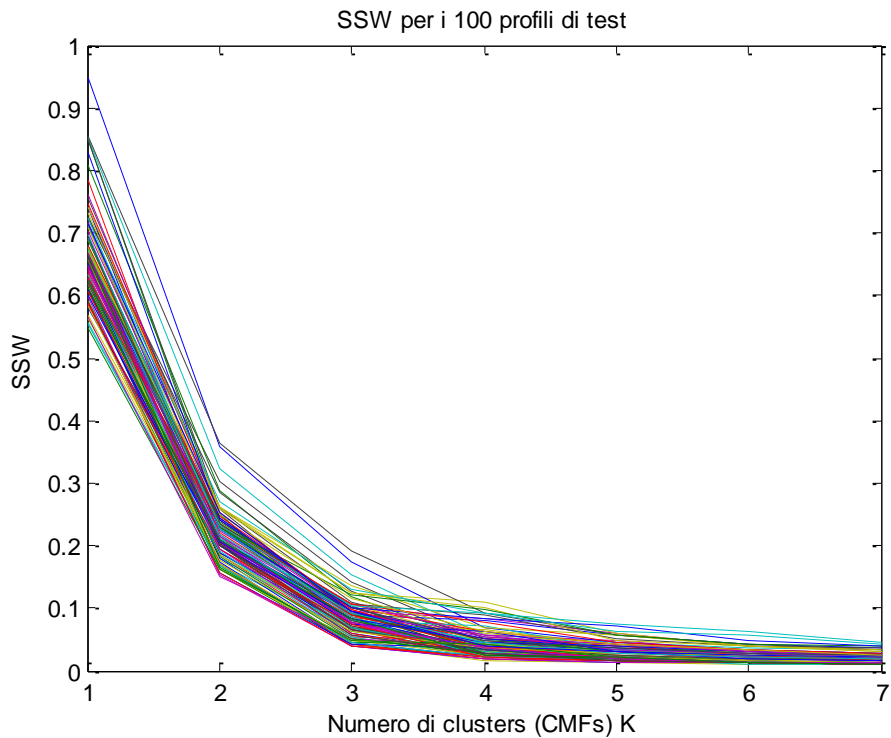
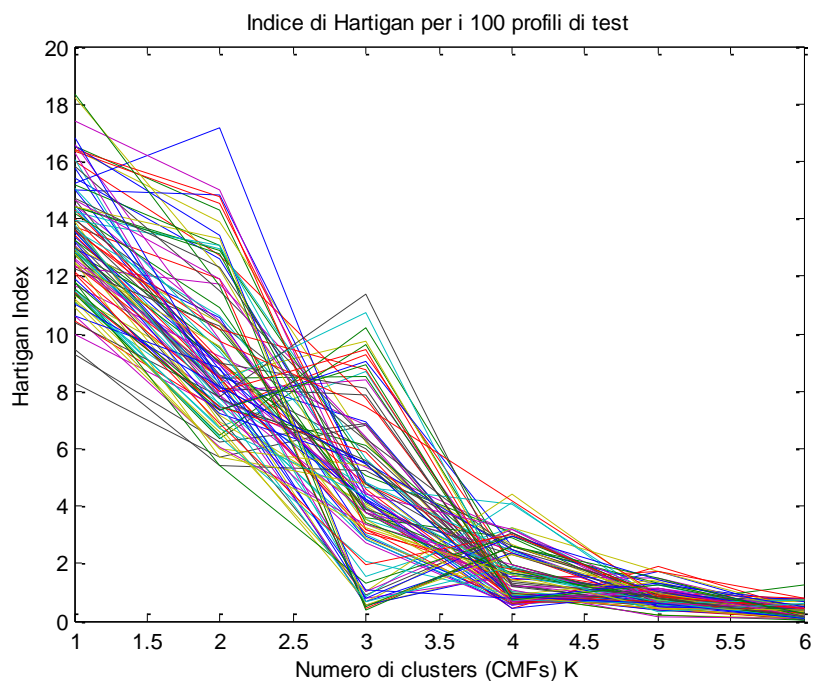
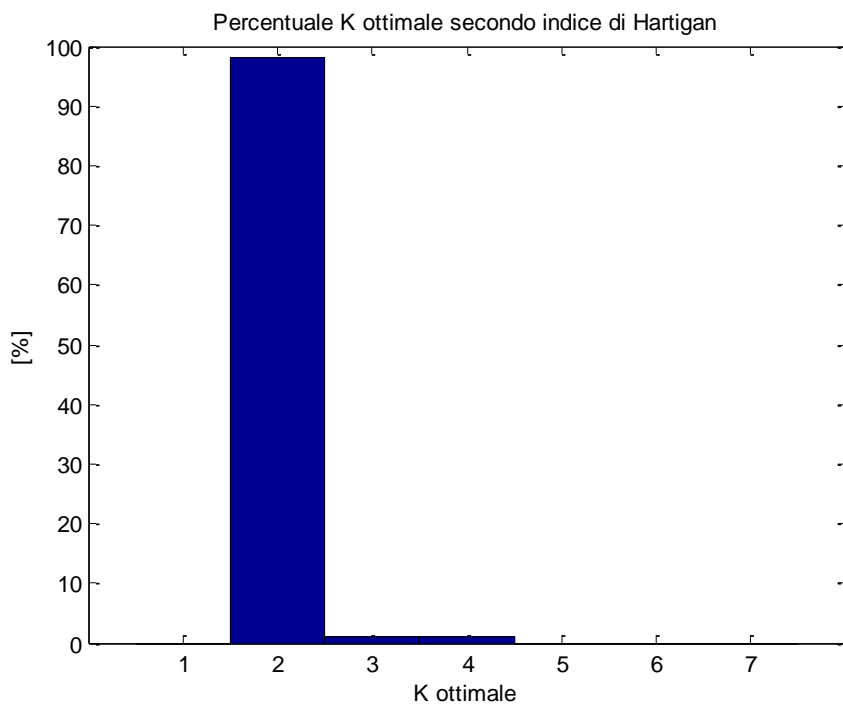


Figura 5.11 SSW per i segnali del dataset simulato



**Figura 5.12** Indice di Hartigan per i segnali del dataset simulato



**Figura 5.13** Distribuzione del K ottimale secondo l'indice di Hartigan

In Tabella 5.1 vengono riassunti i risultati ottenuti per ciascuno degli indici di cluster validation sul dataset di prova in termini di  $K$  ottimale. È possibile notare che il metodo del ginocchio basato sulla SSW risulta essere il più coerente di tutti.

A fronte di questa analisi è possibile affermare che l'elbow method basato sulla SSW è una buona tecnica per individuare il numero di CMFs in cui suddividere le IMFs.

Metodo	K=1	K=2	K=3	K=4	K=5	K=6	K=7	<b>K ottimale</b>
<b>Dunn</b>	-	-	68%	25%	7%	-	-	<b>3</b>
<b>Elbow Method</b>	-	-	100%	-	-	-	-	<b>3</b>
<b>Hartigan</b>	-	98%	1%	1%	-	-	-	<b>2</b>
<b>Davies-Bouldin</b>	-	12%	35%	41%	7%	5%	-	<b>4</b>
<b>Silhouette</b>	-	34%	25%	34%	7%	-	-	<b>?</b>
<b>Calinski-Harabasz</b>	-	81%	5%	14%	-	-	-	<b>2</b>

**Tabella 5.1 Distribuzione del  $K$  ottimale per il dataset di test secondo i diversi indici**

## Capitolo 6

### Analisi dei risultati: real case study

In questo capitolo si riportano i principali risultati ottenuti analizzando i dati reali acquisiti in lavorazioni di taglio waterjet. La prima parte mostra l'inefficacia delle tecniche EEMD e CEEMDAN. La seconda parte mostra i risultati ottenuti applicando il metodo proposto.

#### 6.1 Inefficacia della EEMD e della CEEMDAN

Applicando la EMD al segnale mostrato in Figura 2.3 si ottiene una scomposizione in  $n=10$  Intrinsic Mode Functions e un residuo, come in Figura 4.1. Dall'analisi visiva di questi segnali è possibile dedurre che anche in questo caso l'applicazione della EMD è stata soggetta all'effetto mode-mixing, che ha quindi causato una sovra decomposizione. Infatti alcune IMFs come per esempio  $C_2, C_3, C_4$  e  $C_5$  risultano essere molto simili tra di loro, ed è probabile che in realtà non siano altro che componenti diverse di uno stesso livello tempo-frequenza del segnale.

Come anticipato nel precedente capitolo, in letteratura scientifica sono state seguite principalmente due strade per risolvere il problema del mode-mixing e far sì che il segnale venga scomposto in un numero di sotto-segnali esattamente uguale al numero di livelli tempo-frequenza che lo costituiscono: la prima strada prevede di applicare delle variazioni all'algoritmo EMD, e ha portato alla nascita di algoritmi quali la EEMD e la CEEMDAN. La seconda strada invece è rappresentata dal metodo delle Combined Mode Functions.

In questo paragrafo si dimostra l'inefficacia di queste diverse soluzioni, andandole a testare sui profili di pressione in esame.

In letteratura viene spesso dichiarato che il numero di Intrinsic Mode Functions (IMFs) in cui i segnali vengono decomposti dagli algoritmi Ensemble Empirical Mode Decomposition (EEMD) e Complete Ensemble Empirical Mode Decomposition with Adaptive Noise (CEEMDAN) dipende essenzialmente da due parametri: uno è l'ampiezza dell'errore Gaussiano che viene aggiunto al segnale originale in ognuna delle prove dell'ensemble, e l'altro è il numero di prove di cui è fatto l'ensemble stesso.

In particolare, nell'eseguire queste prove, abbiamo agito sui due parametri in questo modo:

- L'ampiezza dell'errore Gaussiano da aggiungere al segnale in ogni prova dell'ensemble è definita dal valore di un parametro, chiamato  $N_{std}$ . Questo rappresenta il rapporto tra la deviazione standard del rumore



gaussiano aggiunto e la deviazione standard del segnale originale. In letteratura spesso viene consigliato di adottare un valore pari a 0.2. Nelle nostre prove abbiamo utilizzato quattro valori: 0.1, 0.2, 0.3, 0.4. Essi sono riportati sulle righe delle tabelle di seguito riportate.

- Il numero di prove che costituiscono ciascun ensemble è definito dal valore di un altro parametro, che si chiama  $N_{prove}$ , per il quale abbiamo assunto tre valori: 200, 400, 600. Essi sono riportati sulle colonne delle tabelle.

Come è possibile vedere dalle tabelle di seguito riportate relative al tempo computazionale, è chiaro che l'utilizzo di un numero superiore di prove per ogni ensemble diventa eccessivamente oneroso.

In questo lavoro vengono presentati i risultati ottenuti applicando questi due algoritmi ai segnali di pressione WaterJet IC. In particolare sono stati campionati casualmente 80 profili IC e si è cercato di valutare tre aspetti della decomposizione:

- La variabilità del numero di IMFs estratte dai diversi profili. Infatti, a causa soprattutto della presenza di una componente di errore cui spesso sono soggetti questo tipo di segnali, il numero di IMFs in cui la EMD li decompone non solo è superiore al numero di componenti vere dei segnali, ma non risulta neppure essere costante. Il numero  $n$  di IMFs solitamente varia di profilo in profilo.
- Il costo computazionale che queste tecniche comportano: tempi computazionali eccessivamente elevati renderebbero impossibile l'utilizzo di queste tecniche in ambito industriale, soprattutto se l'obiettivo è quello di avere una tecnica di monitoraggio delle condizioni dell'utensile che operi in tempo reale.
- La bontà della decomposizione, data dal grado di ortogonalità delle IMFs in cui un segnale viene decomposto. Una misura di questa bontà è data dal valore dell'Orthogonality Index (OI), che è stato introdotto da Huang nel 1998 [11]. Quanto più le IMFs sono ortogonali, tanto più è buona la decomposizione e tanto più l'OI è vicino al valore di zero.

$$OI = \sum_{p=1}^P \left\{ \sum_{\substack{k=1 \\ q \neq k}}^{n+1} \sum_{q=1}^{n+1} \frac{c_{k,p} c_{q,p}}{x_p^2} \right\} \quad (6.1)$$

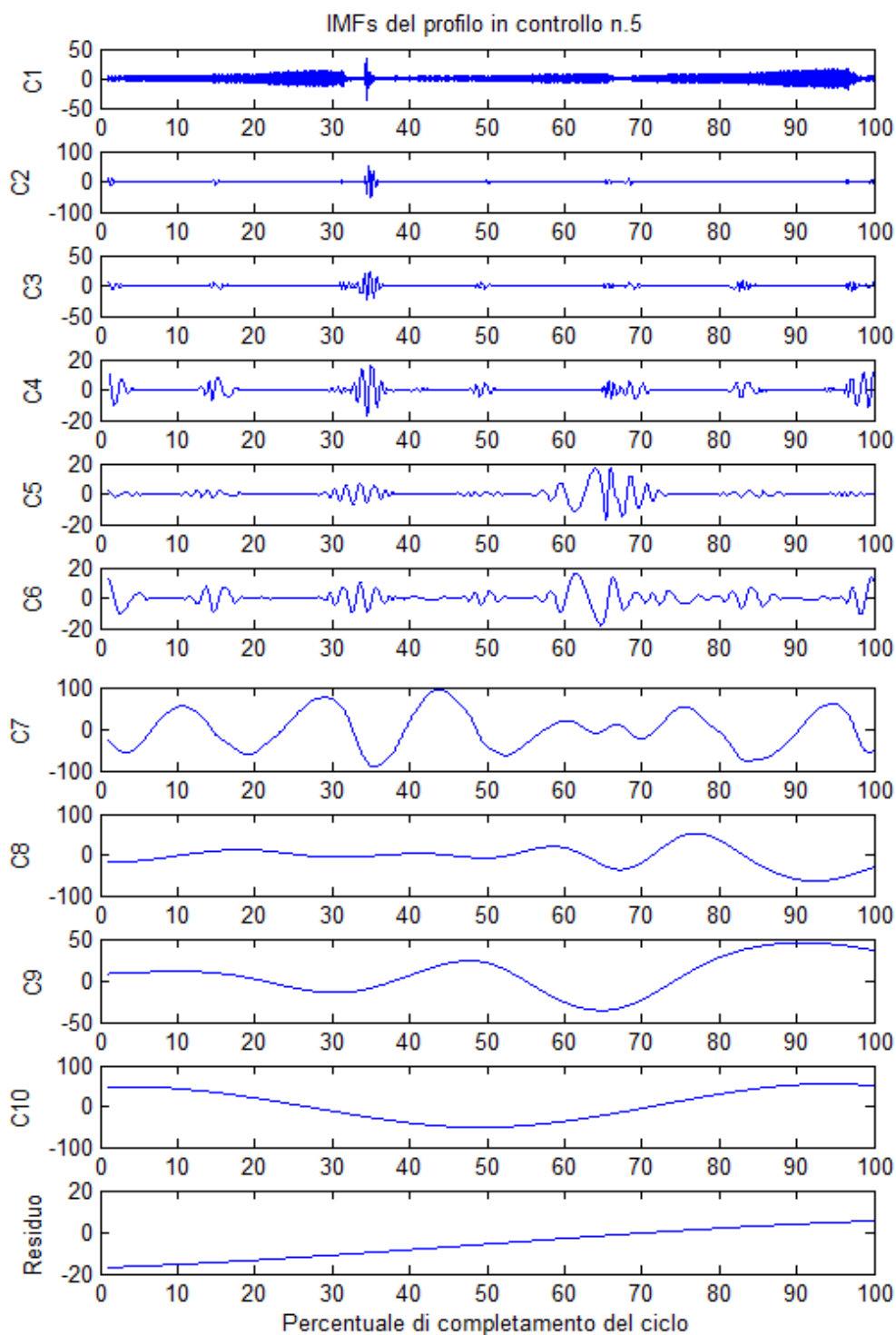


Figura 6.1 EMD del profilo WJ IC in Figura 2.3

➤ EMD

Applicando la EMD ai suddetti segnali di pressione si ottiene una grande variabilità del numero di IMFs estratte da ciascun segnale. Come è possibile vedere dalla Tabella 2.1, si oscilla tra un minimo di 9 e un massimo di 12 IMFs. Questo comporta delle difficoltà in ottica di monitoraggio statistico.

Numero di profili IC esaminati	80
Numero minimo di IMFs	9
Numero massimo di IMFs	12
Moda del numero di IMFs	11
Standard Deviation del numero di IMFs	0.7593
Tempo computazionale medio [s]	1.62
Indice di Ortogonalità medio	0.142

**Tabella 6.1 Risultati della EMD sui profili WJ IC**

➤ EEMD

Anche applicando la EEMD non si riesce a risolvere il problema del mode-mixing.

La situazione migliore in termini di variabilità del numero di IMFs è quella con  $N_{std}=0.3$  e  $N_{prove}=60$ .

Il problema relativo a questa combinazione di valori dei due parametri è rappresentato principalmente dal tempo computazionale. Com'è possibile vedere infatti i tempi crescono al crescere di  $N_{prove}$ , come era logico attendersi, e in particolare ci vogliono circa 3 minuti per eseguire la EEMD su un profilo se ciascun ensemble è costituito da 600 prove.

Numero medio di IMF con EEMD				Deviazione standard del numero di IMF con EEMD			
Nstd\Nprove	200	400	600	Nstd\Nprove	200	400	600
0,1	11,7875	11,9375	12,0375	0,1	0,566887	0,51173	0,514197
0,2	11,8125	11,9	11,975	0,2	0,452664	0,408765	0,449331
0,3	11,85	11,975	12	0,3	0,479979	0,317227	0,275589
0,4	11,8875	12,025	12,0625	0,4	0,389539	0,317227	0,367811

Orthogonality Index medio con EEMD				Deviazione standard dell'Orthogonality Index con EEMD			
Nstd\Nprove	200	400	600	Nstd\Nprove	200	400	600
0,1	0,373792	0,381627	0,382308	0,1	0,090719	0,084874	0,083811
0,2	0,402831	0,401265	0,401421	0,2	0,056839	0,0542	0,052744
0,3	0,429733	0,429103	0,431691	0,3	0,034401	0,034137	0,033721
0,4	0,43341	0,436508	0,436108	0,4	0,03332	0,033376	0,032475

Tempo computazionale medio [s] per l'estrazione delle IMFs con EEMD			
Nstd\Nprove	200	400	600
0,1	69,52154	140,0034	210,4016
0,2	59,70844	118,7092	178,5608
0,3	56,60508	113,0461	169,4535
0,4	55,26735	110,7862	165,839

Tabella 6.2 Risultati della EEMD sui profili WJ IC

➤ CEEMDAN

Applicando la CEEMDAN la situazione relativa alla variabilità del numero di IMFs migliora rispetto ai casi precedenti, sebbene il problema del Mode-Mixing non venga risolto.

A fronte di una minore variabilità e anche di una migliore ortogonalità delle IMFs, è necessario però sottolineare come i tempi di calcolo siano molto più alti nel caso della CEEMDAN rispetto alla EEMD. In particolare questi sono praticamente raddoppiati per ciascuna combinazione dei due parametri.

Numero medio di IMF con CEEMDAN				Deviazione standard del numero di IMF con CEEMDAN			
Nstd\Nprove	200	400	600	Nstd\Nprove	200	400	600
0,1	12,8125	12,8875	12,8375	0,1	0,479814	0,502997	0,403897
0,2	13,0875	13,0375	13,0625	0,2	0,284349	0,191182	0,243589
0,3	13,1125	13,0875	13,0375	0,3	0,317974	0,284349	0,191182
0,4	13,1375	13,075	13,05	0,4	0,346547	0,265053	0,21932

Orthogonality Index medio con CEEMDAN				Deviazione standard dell'Orthogonality Index con CEEMDAN			
Nstd\Nprove	200	400	600	Nstd\Nprove	200	400	600
0,1	0,199264	0,22693	0,24344	0,1	0,093289	0,090968	0,085843
0,2	0,247779	0,256109	0,25906	0,2	0,049998	0,045792	0,043473
0,3	0,277216	0,29253	0,298372	0,3	0,041837	0,041236	0,040678
0,4	0,290244	0,306184	0,31133	0,4	0,040511	0,035694	0,032498

Tempo computazionale medio [s] per l'estrazione delle IMFs con CEEMDAN			
Nstd\Nprove	200	400	600
0,1	120,6131	223,0996	325,7918
0,2	113,711	223,9112	325,0356
0,3	112,42	213,0807	313,4232
0,4	111,3797	214,0606	311,364

**Tabella 6.3 Risultati della CEEMDAN sui profili WJ IC**

➤ COMBINED MODE FUNCTIONS

Alla luce dei risultati derivanti dall'analisi effettuata, si evince chiaramente che, a differenza di quanto spesso viene sostenuto in letteratura, né la EEMD né tantomeno la CEEMDAN sono effettivamente in grado di porre rimedio al problema della sovra-decomposizione di un segnale. Anche applicando questi algoritmi infatti si continua a registrare un numero n di IMFs variabile da profilo a profilo e i modi oscillatori vengono comunque sovra-decomposti. Inoltre, i tempi computazionali richiesti da queste due varianti della Empirical Mode Decomposition non le rendono funzionali alla messa a punto di un sistema di monitoraggio on-line per profili multiscala.

**6.2 Il calcolo delle CMFs**

In questo paragrafo viene presentato l'algoritmo proposto per il calcolo delle CMFs applicato al dataset di profili WJ che costituiscono il caso di studio reale.

In particolare nel seguito si considera il segnale di pressione WaterJet in Figura 2.3 e le sue IMFs di Figura 6.1.

Figura 6.2 mostra gli istogrammi (in blu) delle frequenze relative calcolati per ognuna delle IMFs suddividendo l'intervallo che va dal minimo al massimo valore del segnale in 200 intervalli equamente ampi e le relative PDFs stimate tramite funzione Kernel (in rosso). In questa figura vanno notate le differenze di scala che vi sono tra gli assi dei diversi grafici. Le prime PDFs risultano essere molto schiacciate e alte, mentre le ultime assumono un andamento maggiormente appiattito e ampio. Questo è ovviamente dovuto alle differenze che intercorrono tra le diverse IMFs, dovute proprio al fatto che esse appartengono a livelli tempo-frequenza distinti.

Figura 6.3 invece mostra l'andamento della distanza euclidea tra le Probability Distribution Functions di coppie di IMFs adiacenti. In particolare in corrispondenza del k-esimo valore sull'asse delle ascisse viene riportato, in ordinata, il valore della distanza euclidea tra la PDF della IMF  $C_k$  e quella della IMF  $C_{k+1}$ .

Una volta calcolate queste distanze l'algorithmo seleziona in modo automatico le K-1 coppie di IMFs contigue per le quali risulta essere massima la distanza euclidea tra le PDFs.

Supponendo per il momento di sapere a priori che il numero di modi oscillatori che compongono il segnale di pressione è  $K=3$  (rimandando a dopo la discussione sui metodi per determinare il numero di CMFs ottimale), in questo caso verrebbero selezionati  $k=1$  e  $k=6$  quali punti di taglio, il che comporterebbe la seguente suddivisione delle IMFs in 3 CMFs:

$$c_{s1} = c_1 \quad (6.2)$$

$$c_{s2} = \sum_{i=2}^6 c_i \quad (6.3)$$

$$c_{s3} = \sum_{i=7}^{10} c_i + r_{10} \quad (6.4)$$

Con questa particolare scelta del valore di K, le CMFs assumono l'aspetto mostrato in Figura 6.3, in cui nella prima riga è stato riportato il segnale originale.

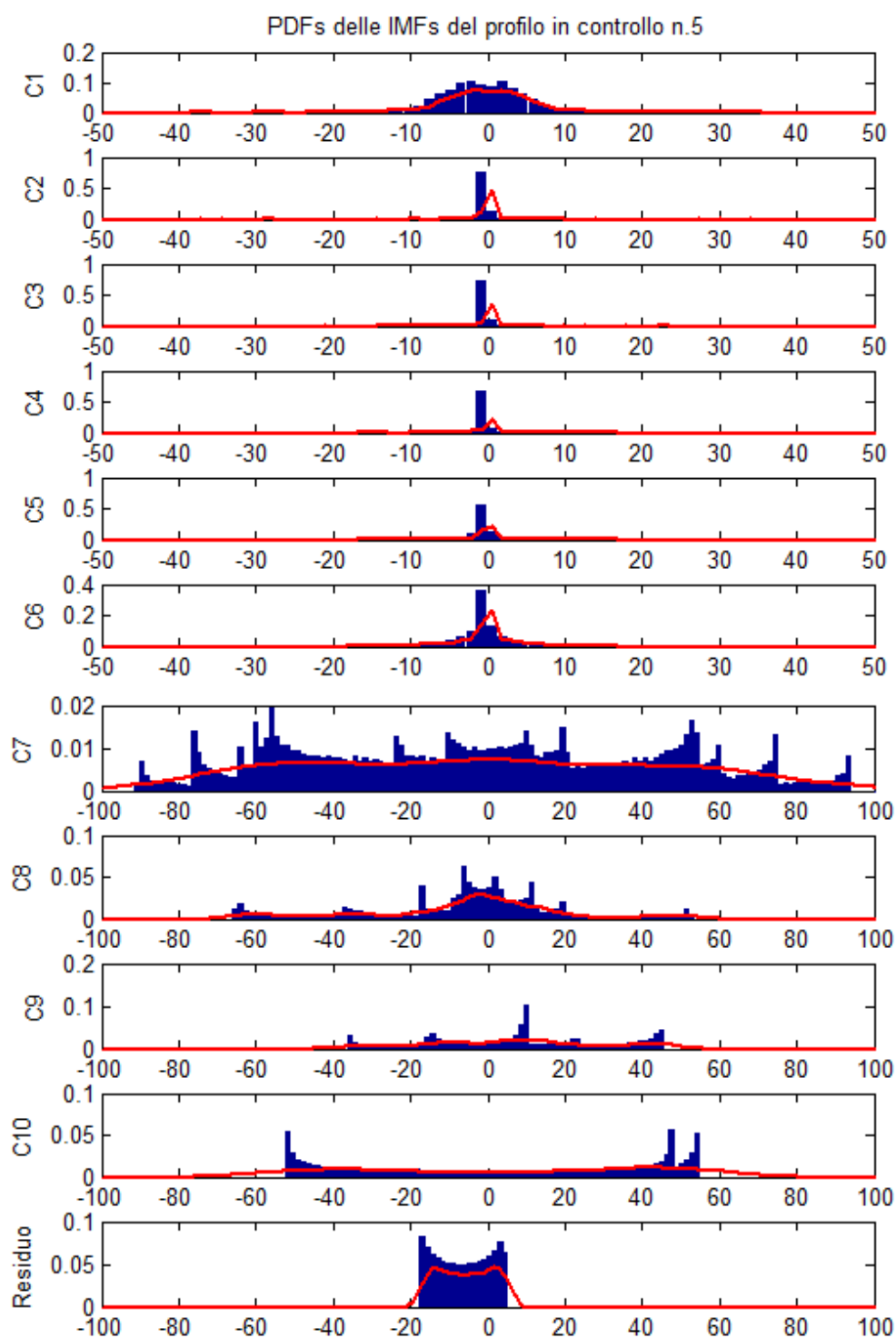
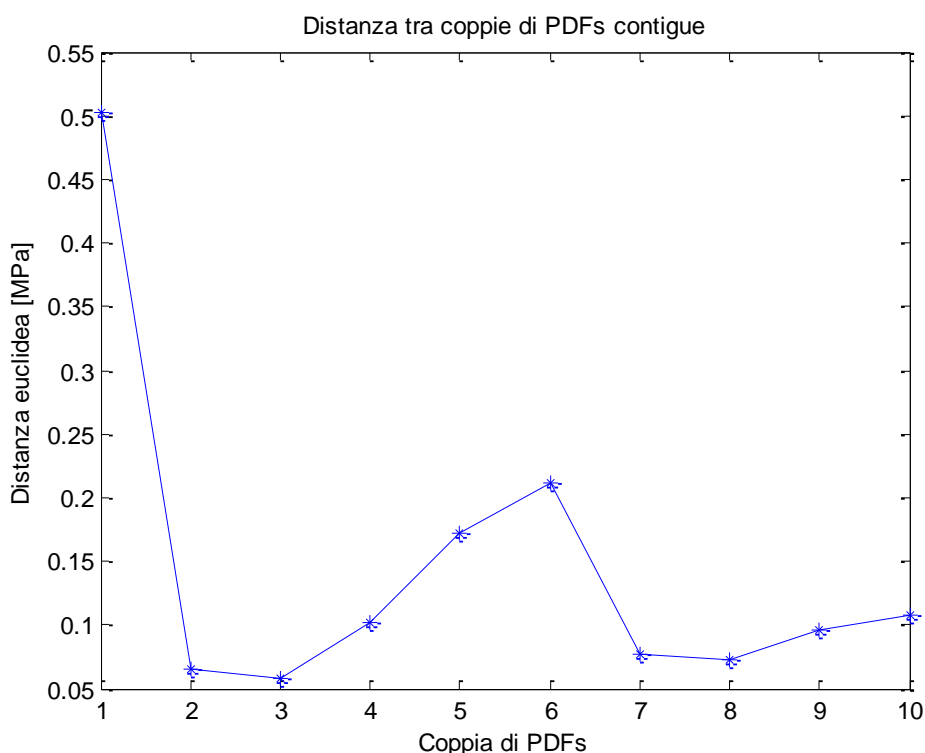


Figura 6.2 PDFs delle IMFs del profilo WJ IC di Figura 2.3



**Figura 6.3 Distanza tra coppie di PDFs contigue per il profilo WJ IC n.5**

Figura 6.5 mostra come si susseguono temporalmente le 3 fasi di cui si compone il ciclo di ogni pistone (*plunger*) della pompa intensificatrice. Questa figura ci permette di capire al meglio il significato delle 3 CMFs:

- La CMF1 rappresenta una componente di rumore ad altissima frequenza. In particolar modo questo rumore sembrerebbe avere una ampiezza che si amplifica man mano che il pistone prosegue la sua corsa verso la fine del cilindro e pressurizza l'acqua nella testa di taglio, salvo poi ridursi nel momento in cui si verifica un transitorio principale, ovvero la chiusura della valvola di ritegno e l'apertura di quella del plunger successivo;
- La CMF2 isola i transitori principali e quelli secondari, che sono perfettamente visibili rispettivamente in corrispondenza del passaggio dall'*active stroke* di un pistone a quello successivo e in corrispondenza degli istanti in cui terminano le fasi di aspirazione (*suction*) dei tre pistoni.
- La CMF3 infine modella l'andamento del segnale al netto della componente di errore e dei transitori. Essa rappresenta quindi le oscillazioni di bassa frequenza del segnale di pressione.



Questo dimostra come la Empirical Mode Decomposition sia effettivamente in grado di separare le diverse componenti in scala di un segnale come quello in esame.

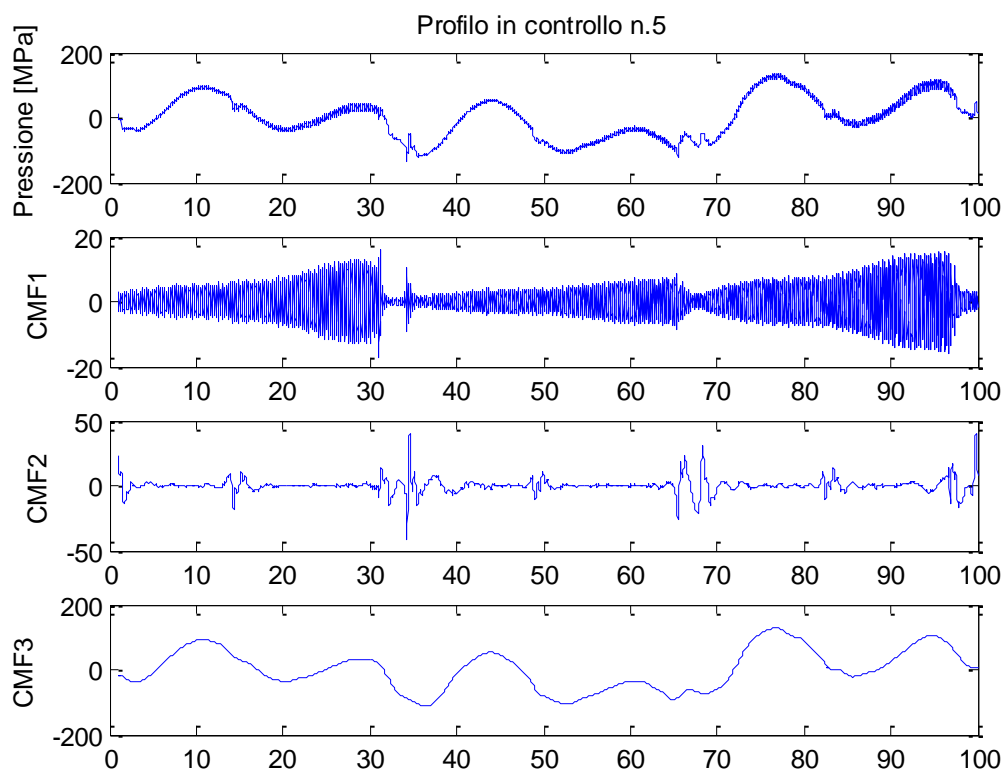


Figura 6.4 CMFs per il profilo WJ IC n.5 con K=3

<i>Plunger 1</i>	Active stroke	Suction		Pre-compression
<i>Plunger 2</i>	Pre-compression	Active stroke	Suction	
<i>Plunger 3</i>	Suction		Pre-compression	Active stroke

Figura 6.5 Fasi del ciclo dei tre gruppi cilindro-pistone della pompa intensificatrice

### 6.3 Scelta del K ottimale

Anche per i profili di pressione WJ sono stati valutati i diversi indici di cluster validation presentati nel Capitolo 4 al fine di determinare il valore ottimale del numero di CMFs.

Tabella 6.4 riporta i risultati ottenuti applicando questi indici ai 131 profili In Controllo. Anche in questo caso è possibile notare come l'utilizzo di indici diversi porti a conclusioni diverse circa il numero di livelli in scala più opportuno da estrarre dai segnali. Inoltre, alcuni di questi indici sembrano essere caratterizzati da una certa "variabilità" interna e quindi non sono del tutto affidabili. L'indice di Davies-Bouldin e il coefficiente di Silhouette, e.g., sono decisamente poco precisi nello stabilire quale debba essere il valore di K migliore, tanto che utilizzando questi indici risulterebbe difficile realizzare una scelta. D'altra parte l'elbow method sembra essere, anche in questo caso, il metodo maggiormente consistente. Anche alla luce dei risultati mostrati nel capitolo 5, l'idea di affidare la scelta del valore ottimale di K a questo metodo sembra essere quella migliore. Di conseguenza K=3 è il numero di CMFs che vanno estratte dai profili WJ al fine di caratterizzare al meglio il loro contenuto informativo su diversi livelli.

Metodo	K=1	K=2	K=3	K=4	K=5	K=6	K=7	K ott
<b>Dunn</b>	-	-	26,7%	42,8%	19,9%	6,9%	3,7%	4
<b>Elbow Method</b>	-	-	84,7%	15,3%	-	-	-	3
<b>Hartigan</b>	-	26,7%	42,8%	19,9%	6,9%	3,7%	-	3
<b>Davies-Bouldin</b>	-	-	19%	29,8%	26%	7,6%	17,6%	?
<b>Silhouette</b>	-	1,5%	15,3%	16,8%	13%	14,5%	38,9%	?
<b>Calinski-Harabasz</b>	-	85,5%	9,9%	4,6%	-	-	-	2

Tabella 6.4 Distribuzione del K ottimale per i profili WJ IC secondo i diversi indici

## Capitolo 7

# Combinazione di EMD e PCA per profile monitoring

L'approccio che viene proposto per il monitoraggio di profili risulta dalla combinazione della Empirical Mode Decomposition (EMD) e della Principal Component Analysis (PCA). In particolar modo la EMD serve a separare le diverse componenti in scala che costituiscono il segnale, mentre la PCA viene utilizzata al fine di ridurre la complessità del problema e sintetizzare le informazioni contenute nei segnali in un numero limitato di variabili. Il metodo che risulta dalla combinazione di queste due tecniche è lo stesso che è stato utilizzato da Grasso *et al.* in [30], e prende il nome di EMD-PCA.

In questo capitolo viene brevemente introdotta la PCA e si illustra nel dettaglio il principio di funzionamento dello schema di monitoraggio on-line di profili multiscala che si intende mettere a punto.

### 7.1 La Principal Component Analysis (PCA)

In una procedura di monitoraggio di profili si hanno due fasi distinte:

- La Fase I, anche detta Fase di addestramento (*training step*), in cui viene costruito il modello di riferimento per la successiva identificazione di guasti. Essa si basa su un dataset che è possibile rappresentare come una matrice  $\mathbf{X}$  di dimensioni  $M \times P$  come quella in Equazione 7.1, rappresentativa della condizione di buono stato di salute dell'utensile in lavorazione. In questa matrice  $M$  è il numero di profili campionati in condizioni di normale funzionamento della macchina utensile, mentre  $P$  è il numero di punti registrati per ogni profilo. Ogni riga di questa matrice dunque corrisponde a un profilo in controllo.

In questa fase è possibile distinguere una fase di *training* iniziale da una di *tuning*. La prima consiste nell'addestramento del modello, mentre nella seconda si realizza la stima dei limiti per le carte di controllo.

$$\mathbf{X} = \begin{bmatrix} x_{1,1} & \cdots & x_{1,P} \\ \vdots & \ddots & \vdots \\ x_{M,1} & \cdots & x_{M,P} \end{bmatrix} \quad (7.1)$$

- La Fase II, ovvero la fase di monitoraggio effettivo, in cui per ogni nuovo profilo campionato durante un ciclo di lavoro della macchina monitorata vengono calcolate una o più statistiche che vengono inserite all'interno delle carte di controllo allestite in Fase I.

Detta  $\mathbf{CS}_{j,k} = [CS_{j,1,k} \ CS_{j,2,k} \ \dots \ CS_{j,P,k}]^T$  la k-esima CMF ottenuta applicando il metodo descritto nel Capitolo 4 al profilo  $\mathbf{x}_j$  del dataset  $\mathbf{X}$ , è possibile costruire K matrici  $\mathbf{CS}_{1:M,k}$  di dimensioni M×P che raccolgano le k-esime CMFs calcolate per gli M profili di Fase I:

$$\mathbf{CS}_{1:M,k} = \begin{bmatrix} CS_{1,1,k} & CS_{1,2,k} & \dots & CS_{1,P,k} \\ \dots & \dots & \ddots & \vdots \\ CS_{M,1,k} & CS_{M,2,k} & \dots & CS_{M,P,k} \end{bmatrix} \quad (k = 1,2,\dots,K) \quad (7.2)$$

Su queste K matrici viene eseguita la Principal Component Analysis, al fine di valutare la stabilità nel tempo delle CMFs.

L'Analisi delle Componenti Principali (PCA) è una tecnica per la semplificazione dei dati utilizzata nell'ambito della statistica multivariata [46]. Lo scopo con cui viene applicata questa tecnica è la *feature reduction*, che consiste nel ridurre un dataset con un numero molto alto di variabili ad un numero limitato di nuove variabili (le Componenti Principali appunto) che risultano da una combinazione lineare delle variabili di partenza, il tutto preservando quanto più possibile la struttura di varianza-covarianza presente nei dati.

Nei casi come quello che stiamo trattando in cui la PCA viene applicata per monitorare dei segnali che si ripetono nel tempo la procedura di riferimento è quella delineata da Colosimo *et al.* [47].

Questa procedura prevede innanzitutto di eseguire una decomposizione spettrale della matrice di correlazione campionaria  $\mathbf{R}_{1:M,k}$  di  $\mathbf{CS}_{1:M,k}$ :

$$\mathbf{R}_{1:M,k} = \frac{1}{M-1} \sum_{j=1}^M (\mathbf{CS}_{j,k} - \overline{\mathbf{CS}}_k) * (\mathbf{CS}_{j,k} - \overline{\mathbf{CS}}_k)^T \quad (k = 1,2,\dots,K) \quad (7.3)$$

dove  $\overline{\mathbf{CS}}_k$  rappresenta la media delle k-esime CMFs calcolate per gli M profili di training:

$$\overline{\mathbf{CS}}_k = \frac{1}{M} \sum_{j=1}^M \mathbf{CS}_{j,k} \quad (k = 1,2,\dots,K) \quad (7.4)$$

Questa decomposizione consiste nel trovare le due matrici  $\mathbf{U}_k$  e  $\mathbf{L}_k$  che soddisfino la seguente uguaglianza:

$$\mathbf{U}_k^T * \mathbf{R}_{1:M,k} * \mathbf{U}_k = \mathbf{L}_k \quad (k = 1, 2, \dots, K) \quad (7.5)$$

$\mathbf{L}_k$  è una matrice diagonale di dimensioni  $P \times P$ , i cui elementi sulla diagonale principale rappresentano gli auto valori della matrice  $\mathbf{R}_{1:M,k}$ . Indicheremo questi autovalori come  $\lambda_{p,k}$  con  $p=1, 2, \dots, P$ .

$\mathbf{U}_k$  invece è una matrice ortonormale di dimensioni  $P \times P$  la cui  $p$ -esima colonna  $\mathbf{u}_{p,k}$  è il  $p$ -esimo autovettore di  $\mathbf{R}_{1:M,k}$ .

Si supponga di ordinare le colonne  $\mathbf{u}_{p,k}$  di  $\mathbf{U}_k$  in ordine decrescente rispetto ai rispettivi auto valori, i.e.  $\lambda_{1,k} > \lambda_{2,k} > \lambda_{3,k} > \dots > \lambda_{p,k}$ . Così facendo la prima PC  $\mathbf{u}_{1,k}$  rappresenterà la direzione di massima varianza delle  $k$ -esime CMFs di training, la seconda PC  $\mathbf{u}_{2,k}$  rappresenta la direzione, perpendicolare alla precedente, di massima varianza residua, e così via.

A questo punto è possibile calcolare, per ogni profilo  $\mathbf{x}_j$  di training un vettore  $\mathbf{z}_{j,k}$  il cui  $r$ -esimo elemento è il valore della  $r$ -esima PC calcolato per la  $k$ -esima CMF di quel profilo  $j$ -esimo. Esso si calcola come:

$$\mathbf{z}_{j,k} = \mathbf{U}_k^T * \mathbf{C}\mathbf{s}_{j,k} = [z_{j,1,k} \ z_{j,2,k} \ \dots \ z_{j,P,k}]^T \quad \begin{matrix} (k = 1, 2, \dots, K) \\ (j = 1, 2, \dots, M) \end{matrix} \quad (7.6)$$

Gli elementi  $z_{j,p,k}$  sono anche detti “punteggi” (*scores*).

Si può dimostrare che  $\mathbf{C}\mathbf{s}_{j,k} = \overline{\mathbf{C}\mathbf{s}}_k + \mathbf{U}_k \mathbf{z}_{j,k}$ , e cioè nota la matrice ortonormale  $\mathbf{U}_k$ , ogni CMF  $\mathbf{C}\mathbf{s}_{j,k}$  può essere ottenuto a partire dai suoi punteggi come

$$\mathbf{C}\mathbf{s}_{j,k} = \overline{\mathbf{C}\mathbf{s}}_k + \sum_{r=1}^P z_{j,r,k} \mathbf{u}_{r,k} \quad \begin{matrix} (k = 1, 2, \dots, K) \\ (j = 1, 2, \dots, M) \end{matrix} \quad (7.7)$$

Le Componenti Principali ottenute con l'Equazione 7.5 sono statisticamente indipendenti, ed ognuna di esse ha una varianza uguale al corrispondente auto valore. Quindi è possibile non considerare tutte le  $P$  Componenti Principali, bensì soltanto quelle più importanti (ovvero quelle associate a una varianza maggiore), scartando le restanti (ovvero quelle varianza più piccola).

Se si considerano tutte le  $P$  PCs allora è possibile ricostruire esattamente la CMF  $\mathbf{C}\mathbf{s}_{j,k}$  come mostrato in Equazione 7.7. Se invece si considerano solo le prime  $m_k$  PCs ( $m_k < P$ ), allora è possibile solamente dare una stima di  $\mathbf{C}\mathbf{s}_{j,k}$ :

$$\widehat{\mathbf{C}}\mathbf{s}_{j,k} = \overline{\mathbf{C}}\mathbf{s}_k + \sum_{l=1}^{m_k} z_{j,l,k} \mathbf{u}_{l,k} \quad \begin{array}{l} (k = 1, 2, \dots, K) \\ (j = 1, 2, \dots, M) \end{array} \quad (7.8)$$

Questo permette di ridurre il numero di variabili, preservando quanta più varianza possibile del dataset originale. In altre parole la PCA permette di individuare un insieme di curve ortonormali tali per cui l'espansione di ciascun profilo in termini di queste curve di base approssimi il profilo stesso nel miglior modo possibile. Ovviamente uno step importante è rappresentato dalla scelta del valore di  $m_k$  più opportuno per ognuna delle  $K$  CMFs. Esistono diversi approcci presentati in letteratura, alcuni dei quali verranno discussi in seguito. Una scelta errata del numero di PCs da considerare nel modello può causare ulteriori problemi allo schema di monitoraggio basato sui risultati della PCA.

## 7.2 Le carte di controllo

Una volta selezionato il numero di PCs  $m_k$  che si intende trattenere per ciascuna CMF  $k$ -esima, è possibile calcolare due statistiche per ognuno dei profili  $\mathbf{x}_j$ :

- La statistica  $T^2$  di Hotelling, che permette di identificare possibili deviazioni rispetto al modello di riferimento lungo le direzioni associate alle prime  $m_k$  Componenti Principali. Per la  $k$ -esima CMF derivante dal  $j$ -esimo profilo, questa si calcola come segue:

$$T_{j,k}^2(m_k) = \sum_{l=1}^{m_k} \frac{z_{j,l,k}^2}{\lambda_{l,k}} \quad (7.9)$$

- La statistica  $Q$ , che invece permette di individuare possibili deviazioni rispetto al modello di riferimento lungo le direzioni ortogonali alle prime  $m_k$  PCs. Per la  $k$ -esima CMF del profilo  $j$ -esimo si calcola come segue:

$$Q_{j,k}(m_k) = (\mathbf{C}\mathbf{s}_{j,k} - \widehat{\mathbf{C}}\mathbf{s}_{j,k})^T (\mathbf{C}\mathbf{s}_{j,k} - \widehat{\mathbf{C}}\mathbf{s}_{j,k}) \quad (7.10)$$

Quando si passa alla Fase II, per ogni nuovo profilo campionato  $\mathbf{x}_j$  si calcolano queste due statistiche utilizzando le matrici  $\mathbf{U}_k$  e  $\mathbf{L}_k$  e i valori di  $m_k$  stimati in Fase I. Ognuna di queste statistiche viene inserita all'interno di una carta di controllo.

I limiti di controllo vengono stimati nella fase di tuning della Fase I, calcolando il valore delle statistiche  $T^2$  e  $Q$  su un insieme di  $M$  profili in controllo diverso rispetto a quello con cui è stata effettuato l'addestramento del modello PCA e la scelta del numero di PCs.

Per la statistica  $T^2$  di Hotelling il limite di controllo è il seguente:

$$UCL_{T_k^2} = a_k F_\alpha(m_k, M - m_k) \quad (7.11)$$

In questa formula  $\alpha$  è l'errore di primo tipo,  $F_\alpha(m_k, M - m_k)$  è l' $(1 - \alpha)$ % percentile della distribuzione di Fisher con  $m_k$  e  $M - m_k$  gradi di libertà, e  $a_k$  si calcola come:

$$a_k = \frac{m_k(M + 1)(M - 1)}{M(M - m_k)} \quad (7.12)$$

Per la statistica Q invece si utilizza la formula per il limite di controllo suggerita da Nomikos e MacGregor [48]:

$$UCL_{Q_k} = g_k \chi_\alpha^2(h_k) \quad (7.13)$$

In questa formula  $\alpha$  è ancora l'errore di primo tipo e  $\chi_\alpha^2(h_k)$  è l' $(1 - \alpha)$ % percentile della distribuzione  $\chi^2$  con  $h_k$  gradi di libertà.  $g_k$  e  $h_k$  vengono stimati come segue:

$$h_k = \text{int} \left( \frac{2\bar{Q}_k^2}{\hat{\sigma}_{Q_k}^2} \right) \quad (7.14)$$

$$g_k = \frac{\hat{\sigma}_{Q_k}^2}{2\bar{Q}_k} \quad (7.15)$$

dove  $\bar{Q}_k$  e  $\hat{\sigma}_{Q_k}^2$  sono rispettivamente la media e la varianza campionaria per i valori della statistica Q calcolati per gli M profili in controllo del dataset di tuning per la k-esima CMF.

Nella presente tesi in realtà sono stati utilizzati dei limiti di controllo empirici, stimati a partire dalla distribuzione delle due statistiche per i profili di tuning. I limiti presentati in Equazione (7.11) e Equazione (7.13), infatti, sono validi solo qualora i dati soddisfino determinate ipotesi, che invece non vengono quasi mai rispettate dai dati reali come quelli che si sta analizzando. Il numero ristretto di profili in controllo a disposizione, inoltre, non consente di dare una stima ragionevole di questi limiti, e dunque l'utilizzo dei limiti di controllo empirici risulta essere una buona strategia per assicurarsi un ARL0 effettivo pari a quello target teoricamente fissato.

Per ognuna delle  $K$  CMFs quindi si avranno due carte di controllo: una carta  $T^2$  e una carta  $Q$ . Data che queste due carte si riferiscono a due insiemi disgiunti di PCs è necessario applicare la correzione di Dunn-Šidák [49] per eventi indipendenti. Se per ogni CMF si desidera un errore di primo tipo  $\alpha'$  allora l' $\alpha$  di ognuna delle due carte deve essere

$$\alpha = 1 - \sqrt{1 - \alpha'} \quad (7.16)$$

Inoltre, dato che il sistema di monitoraggio così creato segnalerà un allarme quando verrà violato il limite per almeno una delle  $2K$  carte di controllo ( $T^2$  e  $Q$  per ognuna delle  $K$  CMFs), allora è necessario adottare anche la correzione di Bonferroni [50] per controllare l'errore di primo tipo complessivo (*familywise first-type error rate*). Quindi, detto  $\alpha''$  il tasso di falso allarme globale desiderato per il sistema di monitoraggio, per ognuna delle  $K$  CMFs deve essere adottato un valore di  $\alpha'$  dato da:

$$\alpha' = \frac{\alpha''}{K} \quad (7.17)$$

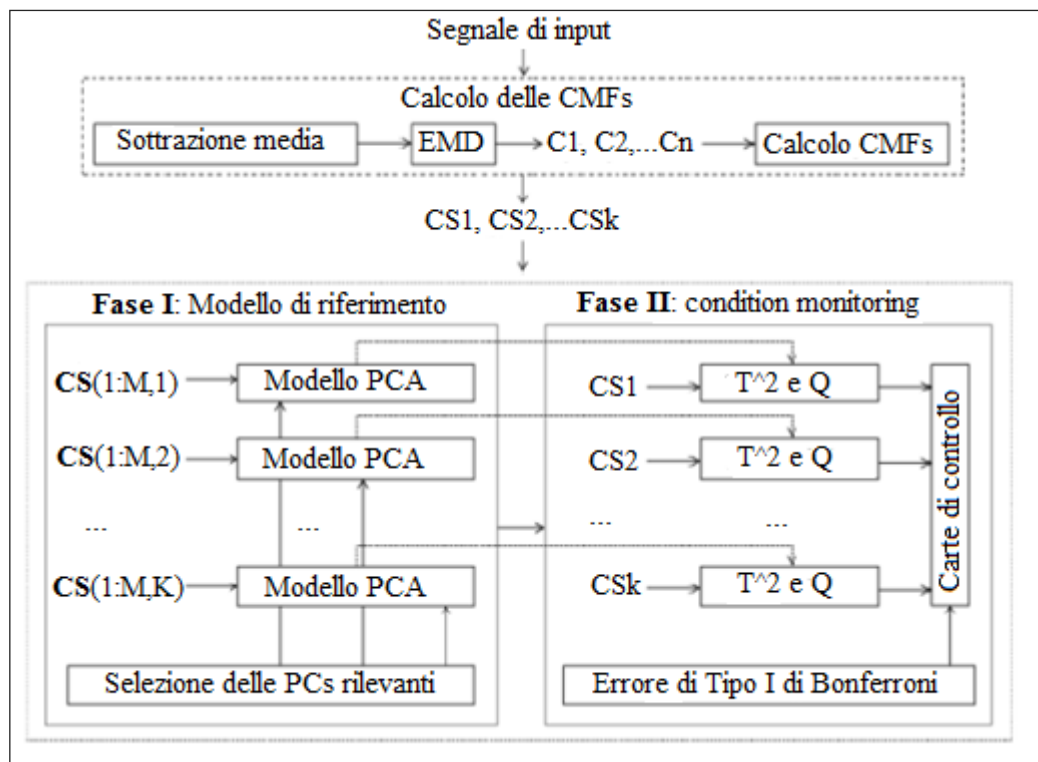


Figura 7.1 Schema di monitoraggio EMD-PCA proposto



### 7.3 Determinare il numero $m_k$ di Componenti Principali

In precedenza è stato sottolineato come uno step molto importante nella costruzione del sistema di monitoraggio di profili multiscala basato sul metodo EMD-PCA sia rappresentato dalla scelta del valore di  $m_k$  più opportuno per ognuna delle K CMFs. Esistono diversi approcci presentati in letteratura. Peres-Neto *et al.* [63], e.g., hanno presentato un'analisi comparativa di ben 20 diverse *stopping rules*. Nel seguito illustreremo alcune delle regole maggiormente utilizzate.

Un possibile approccio è stato proposto da Wold [64], che è basato su un algoritmo di cross-validazione.

Un altro approccio per la scelta del valore ottimale di  $m$  è quello proposto da Kaiser [65]: questa regola suggerisce semplicemente di ritenere nel modello solo le prime  $m$  PCs alle quali è associato un autovalore superiore alla media di tutti gli autovalori.

Lo *scree test* [66] è un'altra strategia per determinare  $m$  sulla base dell'analisi degli autovalori. Essa prevede di creare uno *scree plot* che riporti l'andamento degli autovalori ordinati in ordine decrescente e di individuare il valore di  $m$  per cui si verifica un livellamento/ginocchio in questa curva.

Un'altra strategia ancora si basa sulla percentuale di varianza spiegata dalle prime  $m$  PCs. Dato che l'autovalore di ogni PC ne costituisce anche la varianza spiegata, è possibile ritenere nel modello solo le prime  $m$  PCs che cumulano una varianza superiore a una certa percentuale (e.g. 80%) della varianza totale, data dalla somma di tutti gli autovalori. Dato che solitamente  $P > M$ , la matrice di correlazione campionaria  $R$  ha rango al più pari a  $M-1$ . Quindi possono essere ottenute al più  $M-1$  PCs. La varianza spiegata dalle prime  $m$  PCs può quindi essere calcolata come segue:

$$Var\_cum(m) = \frac{\sum_{l=1}^m \lambda_l}{\sum_{l=1}^{M-1} \lambda_l} \quad (7.18)$$

Quest'ultimo metodo è quello più ampiamente utilizzato e ha il vantaggio di garantire un confronto equo tra metodi diversi, in quanto garantisce che i confronti vengano eseguiti a parità di variabilità spiegata. Per questo motivo, questo metodo è quello applicato nel seguito.

### 7.4 Tasso di *fault detection* del sistema di monitoraggio

In questa sezione vengono illustrati i risultati ottenuti applicando il metodo EMD-PCA di cui sopra al dataset di profili di pressione WJ in esame. In

particolar modo i risultati si riferiscono alla scelta di  $K=3$  per il numero di CMFs da estrapolare da ciascun profilo di pressione.

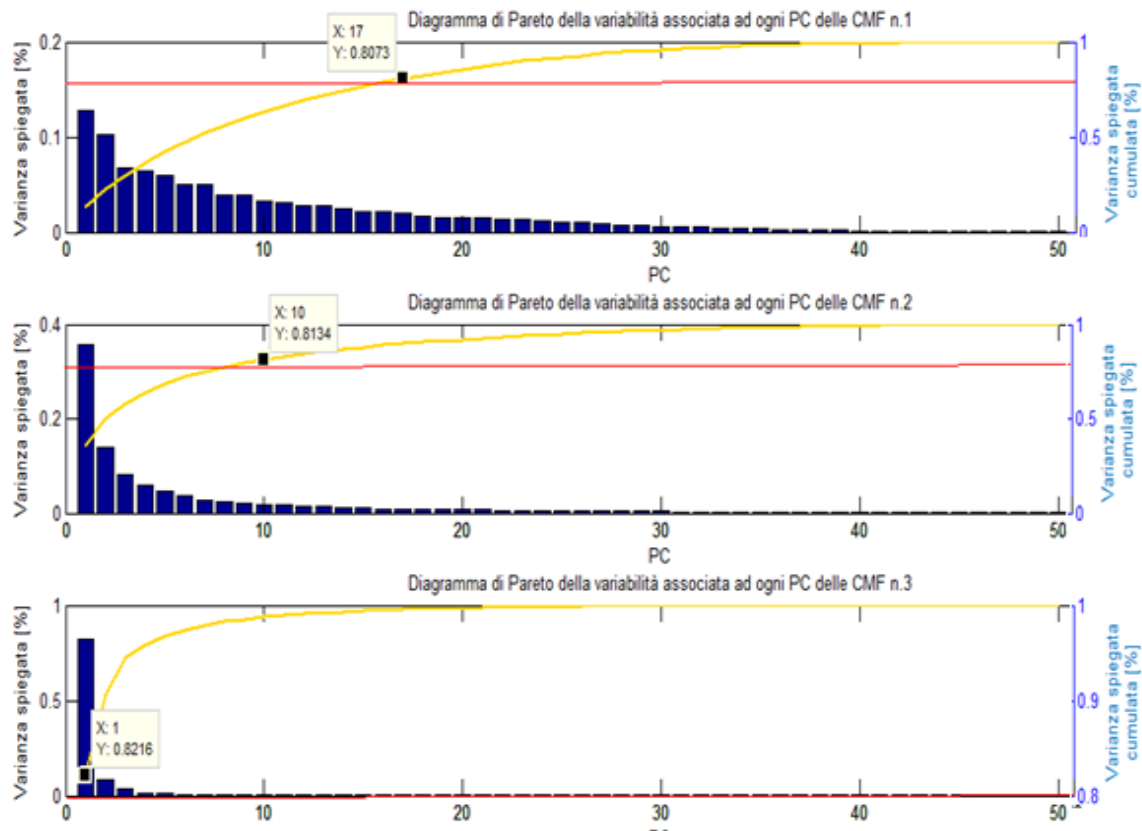
Nella fase di training è necessario, come precedentemente illustrato, addestrare il modello PCA per ognuna delle  $K$  CMFs. Questo addestramento viene eseguito sulla base di  $M_{train}$  profili in controllo e consiste fundamentalmente nello scegliere il numero di PCs  $m_k$  da considerare per la stima di ciascuna CMF  $k$ -esima.

Nel presente lavoro è stato implementato un metodo *naive*: si considerano nel modello le prime  $m_k$  PCs che congiuntamente spiegano almeno l'80% di variabilità del dataset di training. In Figura 7.2, e.g., vengono riportati tre grafici, uno per ogni CMF, che mostrano la variabilità spiegata da ciascuna PC (grafico a barre in blu riferito all'asse di sinistra), la variabilità spiegata cumulata (curva in arancio riferita all'asse delle ordinate di destra) e la soglia di variabilità spiegata cumulata imposta (linea rossa). Questi grafici si riferiscono ad un caso particolare in cui sono stati campionati casualmente  $M_{train} = 50$  profili di pressione in controllo dei 131 a disposizione.

Il primo valore in ascissa per cui la curva della variabilità cumulata spiegata supera il valore 80% viene considerato come il numero di PCs da considerare. E' possibile vedere come per la prima CMF in questo caso sono state selezionate le prime 17 PCs, per la seconda CMF le prime 10, mentre per la terza ed ultima CMF l'unica Componente Principale significativa è la prima, che da sola spiega più dell'82% di varianza delle 50  $CS_3$  di training. Dunque, se si tiene conto del fatto che inizialmente per ogni profilo registrato si avevano 4096 punti, è possibile affermare che la PCA permette di ridurre significativamente il numero di variabili da considerare e quindi il carico computazionale.

La fase di training viene poi seguita da una di tuning in cui utilizzando un altro gruppo di  $M_{tun}$  profili in controllo e il numero di PCs selezionate si realizza una stima dei limiti di controllo per le carte  $T^2$  e  $Q$ .

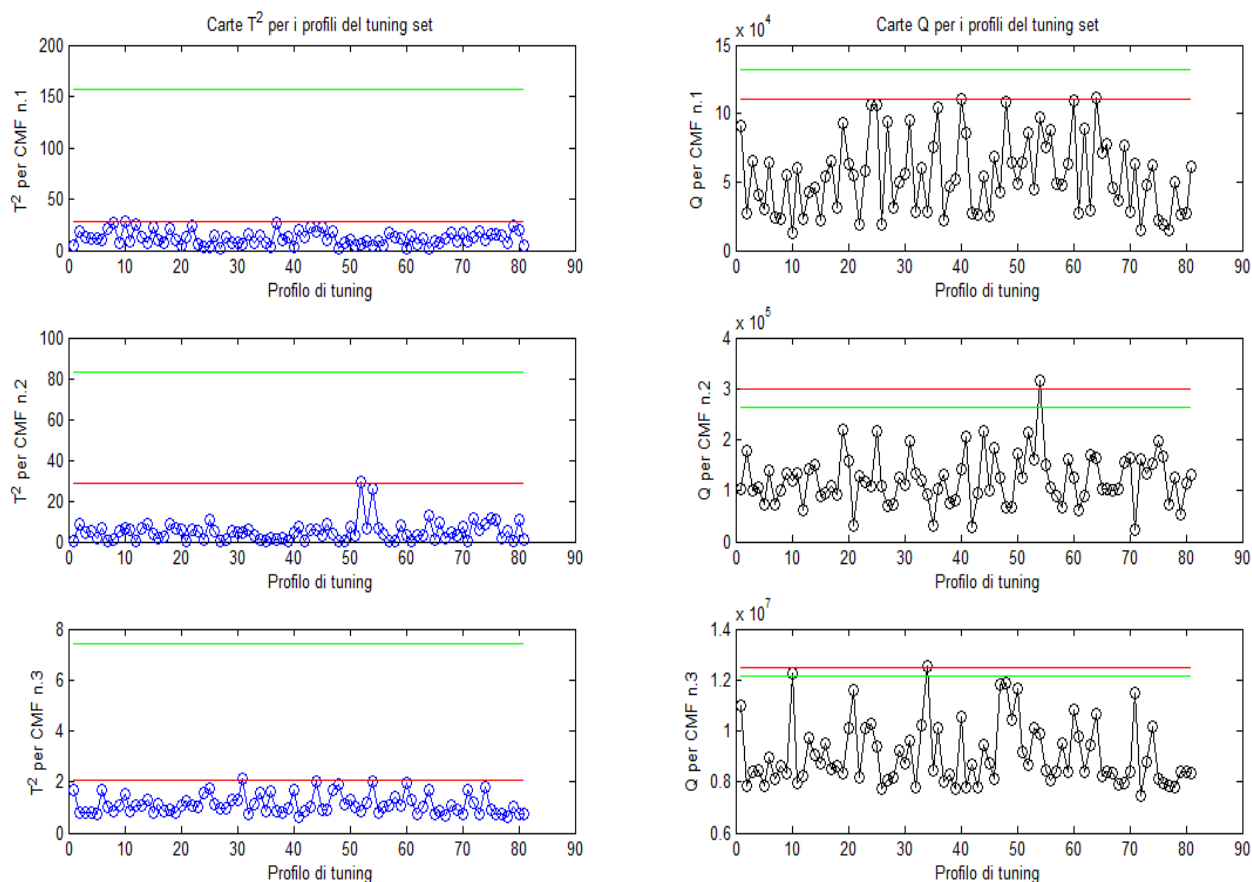
In Figura 7.3, e.g., vengono riportate le carte di controllo per le 3 CMFs estrapolate da ciascuno degli 81 profili in controllo che non sono stati utilizzati nella fase di training nell'esempio precedente. In questi grafici sono riportati sia i limiti teorici (in verde) che quelli empirici (in rosso) ottenuti con un errore di primo tipo complessivo target  $\alpha = 5\%$ . Come è possibile notare, soprattutto per le carte  $T^2$ , vi è una certa discrepanza tra i due tipi di limite. Questo è dovuto al fatto che i dati utilizzati non rispettano le ipotesi alla base della costruzione della carta  $T^2$ . L'utilizzo dei limiti empirici, quindi, costituisce un'ottima soluzione alternativa, dato che essi rispecchiano l'effettiva distribuzione della statistica in oggetto per i profili in controllo.



**Figura 7.2 Esempio di Fase di Training – diagrammi di Pareto**

Una volta progettato il sistema di monitoraggio si passa alla Fase II, che consiste nel calcolare le statistiche  $T^2$  e  $Q$  per ogni CMF di ogni nuovo profilo e confrontarle con i rispettivi limiti di controllo.

In Tabella 7.1 vengono riportati i tassi di fault detection media (su 100 ripetizioni, in cui ogni ripetizione consiste in una separazione random tra dati di training e dati di tuning all'intero del dataset di profili in-controllo) per ognuno dei guasti considerati in questo studio per la macchina WaterJet. Da questa tabella è possibile notare come, indipendentemente dal tipo di guasto considerato, la capacità del sistema di monitoraggio di individuare dei guasti cresce al crescere del numero di profili in controllo con cui il modello PCA viene addestrato. In particolare, utilizzando 100 profili per il training, sostanzialmente tutti i profili fuori controllo vengono segnalati dal sistema, fatta una piccola eccezione per il guasto C.



**Figura 7.3 Esempio di Fase di Tuning con limiti teorici (in verde) e limiti empirici (in rosso)**

Num profili training	Num profili tuning	Guasto A	Guasto B	Guasto C	Guasto D
25	106	92,72%	65,59%	85,68%	70,67%
50	81	98,74%	94,22%	94,24%	94,52%
75	56	99,81%	98,90%	96,67%	99,53%

**Tabella 7.1 Tasso di Fault Detection medio per il metodo EMD-PCA proposto**

Ovviamente il campionamento di un numero molto elevato di profili in controllo per l'addestramento del modello costituisce un costo. Sorge dunque un trade-off tra i costi di campionamento dei profili in controllo e i costi derivanti da una scarsa performance del sistema di monitoraggio dovuta all'inadeguatezza della fase di training.

### 7.5 Confronto con altri metodi

In questo paragrafo vengono mostrati i risultati ottenuti applicando la metodologia EMD-PCA per il monitoraggio dei profili di pressione WJ ma eseguendo la separazione delle IMFs in K CMFs con dei criteri diversi.

In particolar modo è stato calcolato il tasso di fault detection media che si ottiene separando le IMFs sulla base di tre indicatori, ovvero:

- Energia (Tabella 7.2);
- Coefficiente di Correlazione con il segnale originale (Tabella 7.3)
- Ampiezza (Tabella 7.4)

Dall'analisi di queste tabelle è possibile vedere come l'utilizzo del metodo proposto in questa tesi per la separazione delle IMFs porta a dei risultati migliori in termini di capacità di individuare e segnalare degli eventuali guasti. Il problema principale legato all'utilizzo di questi indicatori, infatti, risulta essere la loro scarsa capacità di distinguere la componente di rumore di fondo dai transitori, che quindi vengono fuse in un'unica CMFs. Di conseguenza questi indicatori separano la componente di bassa frequenza in due componenti distinte di cui è difficile dare anche una interpretazione fisica.

Num profili training	Num profili tuning	Guasto A	Guasto B	Guasto C	Guasto D
25	106	62,06%	6,37%	20,93%	8,87%
50	81	65,69%	11,73%	24,62%	11,90%
75	56	72,60%	21,06%	30,78%	18,89%

**Tabella 7.2 Tasso di Fault Detection medio con separazione delle IMFs basata sull'Energia**

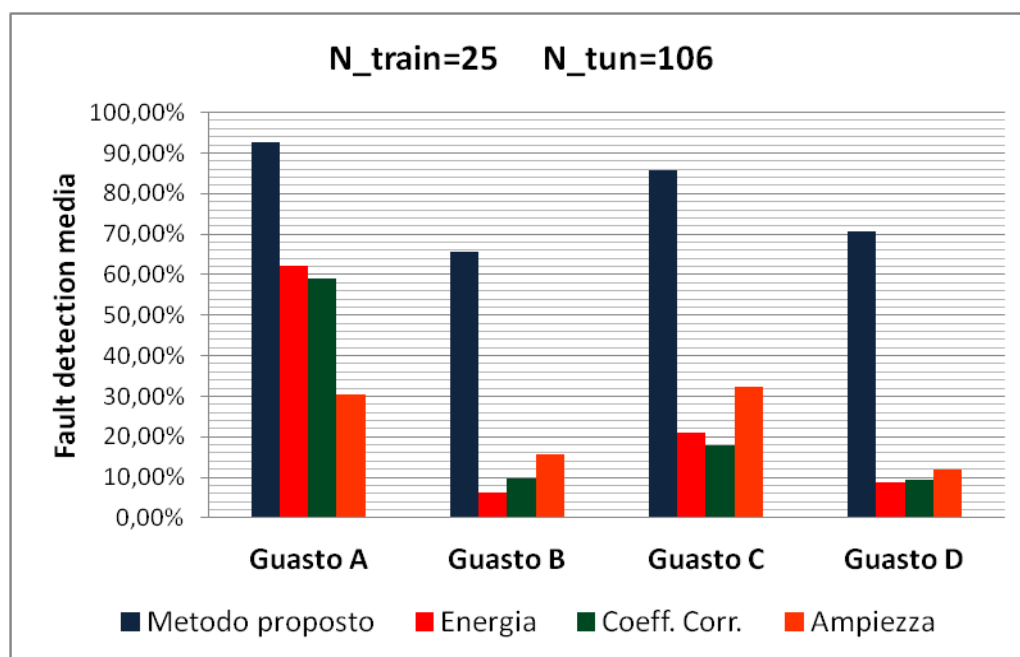
Num profili training	Num profili tuning	Guasto A	Guasto B	Guasto C	Guasto D
25	106	58,95%	9,73%	17,75%	9,38%
50	81	63,10%	12,71%	20,66%	13,24%
75	56	76,35%	28,53%	32,49%	27,33%

**Tabella 7.3 Tasso di Fault Detection medio con separazione delle IMFs basata sul Coefficiente di Correlazione**

Figura 7.4, Figura 7.5 e Figura 7.6 presentano un confronto grafico tra le prestazioni (in termini di fault detection media) dei quattro metodi presi in esame rispettivamente per ognuna delle 3 ripartizioni del dataset di profili in controllo eseguite.

Num profili training	Num profili tuning	Guasto A	Guasto B	Guasto C	Guasto D
25	106	30,47%	15,82%	32,38%	11,98%
50	81	32,62%	19,27%	34,67%	13,78%
75	56	36,06%	24,08%	38,22%	18,49%

**Tabella 7.4 Tasso di Fault Detection medio con separazione delle IMFs basata sull'Am piezza**



**Figura 7.4 Confronto delle performance con N\_train=25 e N\_tun=106**

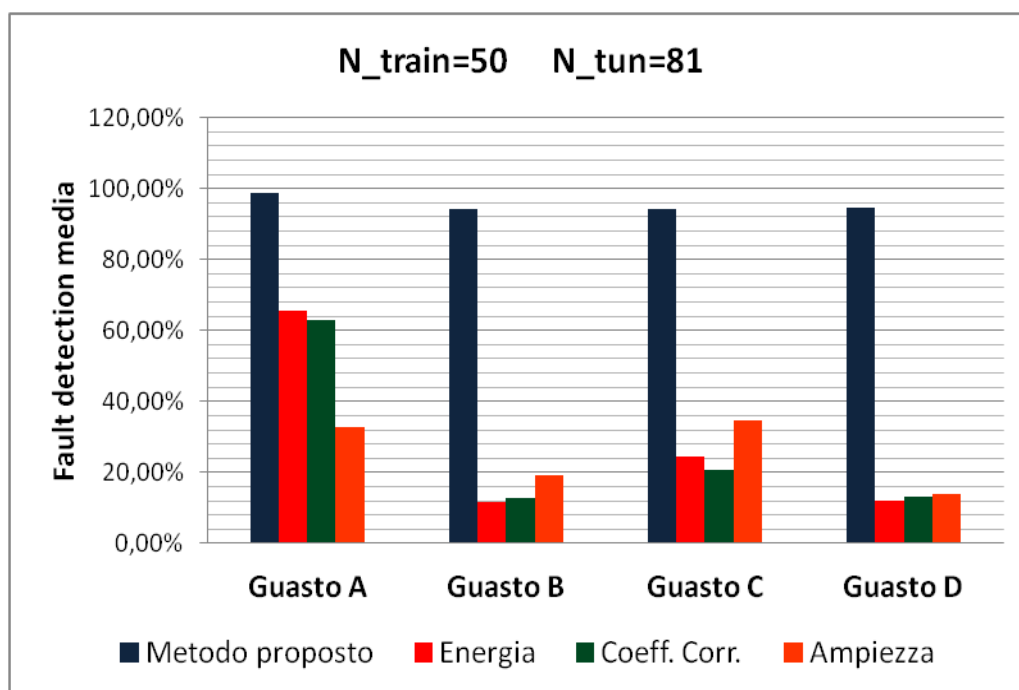


Figura 7.5 Confronto delle performance con N\_train=50 e N\_tun=81

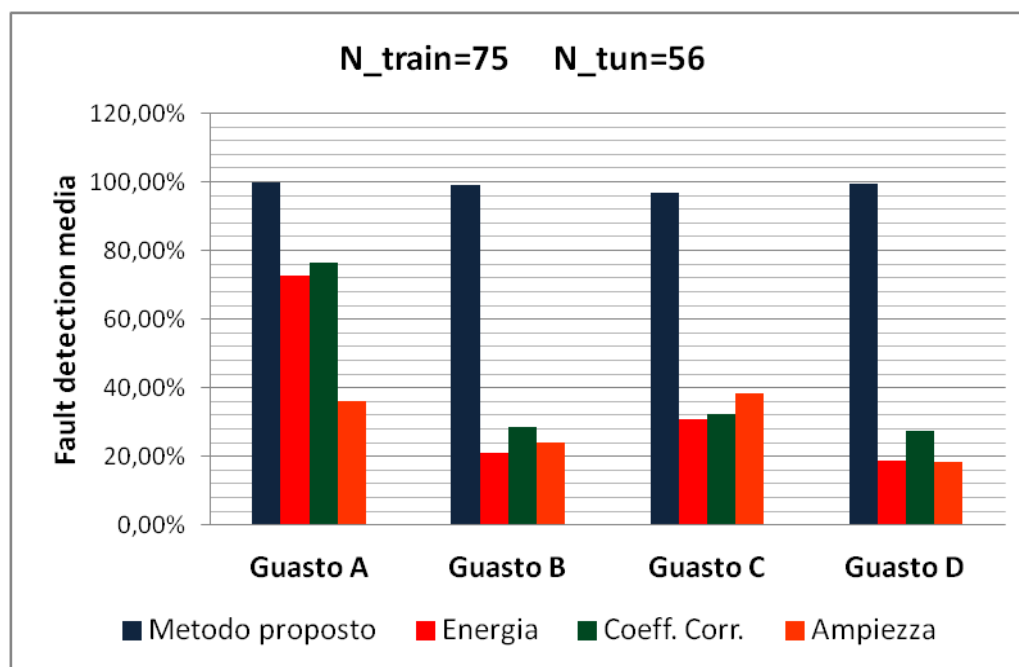


Figura 7.6 Confronto delle performance con N\_train=75 e N\_tun=56

## Capitolo 8

### Conclusioni

Questa tesi presenta uno schema di monitoraggio per profili multiscala basato su una tecnica di analisi nel dominio tempo-frequenza di un segnale nota come Empirical Mode Decomposition (EMD). Questo tipo di segnali, con cui spesso si ha a che fare nella pratica industriale, sono costituiti da diverse componenti in frequenza, e la separazione di queste componenti permette di migliorare le performance di un sistema di controllo. La EMD permette di ottenere questa separazione in maniera adattiva e non-parametrica, e rappresenta quindi una interessante alternativa rispetto a tecniche più tradizionali quali ad esempio la Wavelet Analysis o la STFT.

Questa tecnica viene testata su due dataset di profili multiscala: il primo è un insieme di profili simulati, con delle componenti transitorie ad alta frequenza che si sommano alle oscillazioni di bassa frequenza, mentre il secondo è costituito da un insieme di segnali reali di pressione di un getto idrico, registrati tramite un trasduttore durante diversi cicli di taglio di una macchina WaterJet.

La principale problematica relativa all'applicazione di questa tecnica è rappresentata dalla sovra decomposizione del segnale che può occorrere in presenza di una componente di rumore di fondo. In questa tesi viene dimostrata, sulla base dei segnali a disposizione, l'inefficacia e soprattutto l'inefficienza computazionale di tecniche tradizionalmente indicate come possibili soluzioni a questo problema, quali la EEMD e la CEEMDAN. Di conseguenza il metodo delle CMFs viene utilizzato per decomporre il contenuto informativo del segnale ad un numero limitato di livelli tempo-frequenza.

Un approccio del tutto innovativo viene proposto per la separazione delle IMFs in  $K$  CMFs. Questo metodo, basato sulla stima kernel delle Probability Distribution Functions di ogni IMF e sulle loro distanze, permette di identificare in maniera corretta e automatica quelle IMFs simili tra di loro che vanno sommate a formare un'unica CMF. Inoltre viene affrontato anche il problema della scelta del numero  $K$  di CMFs da estrapolare a partire da un segnale. Quest'ultima questione non era mai stata affrontata precedentemente nella letteratura scientifica, dato che questo valore è sempre stato assunto come un dato noto. Questo metodo quindi aumenta nettamente l'implementabilità a livello industriale di un sistema di monitoraggio di profili multiscala basato sulla EMD, in quanto non presuppone la presenza di un operatore che esegua un'analisi visiva delle IMFs per selezionare lo schema migliore di separazione, così come non richiede una conoscenza preliminare della struttura del segnale stesso.

Lo schema di monitoraggio vero e proprio combina la EMD con la Principal Component Analysis (PCA) al fine di individuare profili fuori controllo.



L'approccio risultante è semplicemente una estensione multiscala del metodo PCA precedentemente introdotto in letteratura per il monitoraggio di profili che si ripetono in maniera ciclica. Esso è noto in letteratura come EMD-PCA. La fase preliminare di selezione del numero  $K$  di CMFs da estrarre e di estrazione delle stesse da ogni segnale, infatti, viene seguita dalla messa a punto di due carte di controllo -  $T^2$  e  $Q$  - per ognuna delle CMFs. Un profilo viene considerato fuori controllo, e quindi derivante da una macchina soggetta a un guasto, qualora una delle due statistiche superi il limite di controllo per almeno una delle  $K$  CMFs.

I risultati dimostrano che i metodi proposti per la scelta automatica di  $K$  e per la separazione automatica delle IMFs risultano efficaci nel separare i diversi livelli tempo-frequenza che compongono ciascun segnale. Il tasso di fault detection medio, indice della capacità che il sistema di monitoraggio ha di identificare dei profili fuori controllo, risulta inoltre essere migliore rispetto ad altri metodi di separazione delle IMFs.

Il fatto che la EMD sia totalmente data-driven, i.e. la capacità di questa tecnica di analizzare un segnale qualsiasi, indipendentemente dalle sue features, rappresenta sicuramente un vantaggio rispetto alle tecniche tradizionali di analisi tempo-frequenza, ma può comportare anche delle difficoltà, in quanto i risultati ottenuti dipendono esclusivamente dalle caratteristiche del segnale. Un discorso analogo vale per i metodi introdotti nel presente lavoro per la scelta del numero di CMFs e per la separazione delle IMFs: trattandosi di un algoritmo di tipo numerico, non supportato da dimostrazioni di tipo teorico o matematico, i risultati ottenuti dipendono esclusivamente dall'andamento dei segnali in esame e dalle loro IMFs.

In futuro, dunque, la ricerca dovrebbe focalizzarsi sul cercare delle tecniche per l'estrazione delle CMFs di un segnale multiscala (e quindi per la caratterizzazione del suo contenuto informativo) che siano maggiormente supportate da un background teorico e che di conseguenza siano efficaci indipendentemente dal profilo da monitorare. L'algoritmo stesso di funzionamento della EMD, e.g., potrebbe essere studiato più approfonditamente e modificato in modo tale da ottenere una tecnica che veramente isoli le diverse componenti in scala di un profilo senza la necessità di ulteriori interventi.

D'altra parte sarebbe anche opportuno cercare di sviluppare nuove metriche per la scelta del numero ottimale di livelli tempo-frequenza da estrarre da un profilo. In altre parole, invece di prendere a prestito delle nozioni dal mondo del clustering, sarebbe più indicato sviluppare delle nuove misure *ad hoc* per il problema in esame, i.e. la scelta del numero di componenti in scala di un profilo.

## Bibliografia

- [1] Montgomery, Douglas C. "Introduction to statistical process control." *John Wiley & Sons, New York, NY* (2001).
- [2] Woodall, William H. "Current research on profile monitoring." *Produção* 17.3 (2007): 420-425.
- [3] Noorossana, Rassoul, Abbas Saghaei, and Amirhossein Amiri. *Statistical analysis of profile monitoring*. Vol. 865. John Wiley & Sons, 2012.
- [4] Bracewell, Ronald Newbold, and R. N. Bracewell. *The Fourier transform and its applications*. Vol. 31999. New York: McGraw-Hill, 1986.
- [5] Lighthill, Michael J. *An introduction to Fourier analysis and generalised functions*. Cambridge University Press, 1958.
- [6] Allen, J. "Short-term spectral analysis, and modification by discrete Fourier Transform." *IEEE Transactions on Acoustics Speech and Signal Processing* 25 (1977): 235-238.
- [7] Daubechies, Ingrid. "The wavelet transform, time-frequency localization and signal analysis." *Information Theory, IEEE Transactions on* 36.5 (1990): 961-1005.
- [8] Gao, Y., Q. Zhang, and X. Kong. "Wavelet-based pressure analysis for hydraulic pump health diagnosis." *Transactions of the ASAE* 46.4 (2003): 969-976.
- [9] Daubechies, Ingrid. "The wavelet transform, time-frequency localization and signal analysis." *Information Theory, IEEE Transactions on* 36.5 (1990): 961-1005.
- [10] Feng, Zhipeng, Ming Liang, and Fulei Chu. "Recent advances in time-frequency analysis methods for machinery fault diagnosis: A review with application examples." *Mechanical Systems and Signal Processing* 38.1 (2013): 165-205.
- [11] Huang, Norden E., et al. "The empirical mode decomposition and the Hilbert spectrum for nonlinear and non-stationary time series analysis." *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 454.1971 (1998): 903-995.
- [12] Peng, Z. K., Peter W. Tse, and F. L. Chu. "A comparison study of improved Hilbert-Huang transform and wavelet transform: application to

- fault diagnosis for rolling bearing." *Mechanical systems and signal processing* 19.5 (2005): 974-988.
- [13] Labus, Thomas J., and George A. Savanick, eds. *An Overview of Waterjet Fundamentals and Applications*. WaterJet Technology Association, 2001.
- [14] Pinheiro, Eduardo, Octavian Postolache, and Pedro Girão. "Empirical Mode Decomposition and Principal Component Analysis implementation in processing non-invasive cardiovascular signals." *Measurement* 45.2 (2012): 175-181.
- [15] Lei, Yaguo, et al. "A review on empirical mode decomposition in fault diagnosis of rotating machinery." *Mechanical Systems and Signal Processing* 35.1 (2013): 108-126.
- [16] Loutridis, S. J. "Damage detection in gear systems using empirical mode decomposition." *Engineering Structures* 26.12 (2004): 1833-1841.
- [17] Gao, Q., et al. "Rotating machine fault diagnosis using empirical mode decomposition." *Mechanical Systems and Signal Processing* 22.5 (2008): 1072-1081.
- [18] Yan, Ruqiang, and Robert X. Gao. "Rotary machine health diagnosis based on empirical mode decomposition." *Journal of Vibration and Acoustics* 130.2 (2008): 21007.
- [19] Junsheng, Cheng, Yu Dejie, and Yang Yu. "A fault diagnosis approach for roller bearings based on EMD method and AR model." *Mechanical Systems and Signal Processing* 20.2 (2006): 350-362.
- [20] Ricci, Roberto, and Paolo Pennacchi. "Diagnostics of gear faults based on EMD and automatic selection of intrinsic mode functions." *Mechanical Systems and Signal Processing* 25.3 (2011): 821-838.
- [21] Huang, Tian-li, Wei-xin Ren, and Meng-lin Lou. "The orthogonal hilbert huang transform and its application in earthquake motion recordings analysis." *14th World Conference on Earthquake Engineering, Beijing*. 2008.
- [22] Huang, Norden E., and Zhaohua Wu. "A review on Hilbert-Huang transform: Method and its applications to geophysical studies." *Reviews of Geophysics* 46.2 (2008).
- [23] Boutana, Daoud, Messaoud Benidir, and Braham Barkat. "On the selection of Intrinsic Mode Function in EMD method: Application on heart sound signal." *Applied Sciences in Biomedical and Communication Technologies (ISABEL), 2010 3rd International Symposium on*. IEEE, 2010.

- [24] Bu, Nan, Naohiro Ueno, and Osamu Fukuda. "Monitoring of respiration and heartbeat during sleep using a flexible piezoelectric film sensor and empirical mode decomposition." *Engineering in Medicine and Biology Society, 2007. EMBS 2007. 29th Annual International Conference of the IEEE*. IEEE, 2007.
- [25] Liang, Hualou, Qiu-Hua Lin, and J. D. Z. Chen. "Application of the empirical mode decomposition to the analysis of esophageal manometric data in gastroesophageal reflux disease." *Biomedical Engineering, IEEE Transactions on* 52.10 (2005): 1692-1701.
- [26] Huang, Wei, et al. "Engineering analysis of biological variables: an example of blood pressure over 1 day." *Proceedings of the National Academy of Sciences* 95.9 (1998): 4816-4821.
- [27] Weng, Binwei, Manuel Blanco-Velasco, and Kenneth E. Barner. "ECG denoising based on the empirical mode decomposition." *Engineering in Medicine and Biology Society, 2006. EMBS'06. 28th Annual International Conference of the IEEE*. IEEE, 2006.
- [28] Liang, Hualou, Zhiyue Lin, and R. W. McCallum. "Artifact reduction in electrogastrogram based on empirical mode decomposition method." *Medical and Biological Engineering and Computing* 38.1 (2000): 35-41.
- [29] Peng, Yonghong. "Empirical model decomposition based time-frequency analysis for the effective detection of tool breakage." *Journal of manufacturing science and engineering* 128.1 (2006): 154-166.
- [30] Grasso, M., P. Pennacchi, and B. M. Colosimo. "Empirical mode decomposition of pressure signal for health condition monitoring in waterjet cutting." *The International Journal of Advanced Manufacturing Technology* 72.1-4 (2014): 347-364.
- [31] Rato, R. T., M. D. Ortigueira, and A. G. Batista. "On the HHT, its problems, and some solutions." *Mechanical Systems and Signal Processing* 22.6 (2008): 1374-1394.
- [32] Wu, Zhaohua, and Norden E. Huang. "Ensemble empirical mode decomposition: a noise-assisted data analysis method." *Advances in adaptive data analysis* 1.01 (2009): 1-41.
- [33] Huang, Norden E., et al. "A confidence limit for the empirical mode decomposition and Hilbert spectral analysis." *Proceedings of the Royal Society of London. Series A: Mathematical, Physical and Engineering Sciences* 459.2037 (2003): 2317-2345.

- [34] Huang, Norden Eh, and Samuel S. Shen. *Hilbert-Huang transform and its applications*. Vol. 5. World Scientific, 2005.
- [35] Chang, Peter C., Alison Flatau, and S. C. Liu. "Review paper: health monitoring of civil infrastructure." *Structural health monitoring* 2.3 (2003): 257-267.
- [36] Pines, Darryll, and Liming Salvino. "Structural health monitoring using empirical mode decomposition and the Hilbert phase." *Journal of sound and vibration* 294.1 (2006): 97-124.
- [37] Li, Helong, Xiaoyan Deng, and Hongliang Dai. "Structural damage detection using the combination method of EMD and wavelet analysis." *Mechanical Systems and Signal Processing* 21.1 (2007): 298-306.
- [38] Roth, John T., et al. "Quality and Inspection of Machining Operations: Tool Condition Monitoring." *Journal of Manufacturing Science and Engineering* 132.4 (2010): 041015.
- [39] Elbestawi, Mo A., Mihaela Dumitrescu, and Eu-Gene Ng. "Tool condition monitoring in machining." *Condition Monitoring and Control for Intelligent Manufacturing*. Springer London, 2006. 55-82.
- [40] Zhang, Xun, et al. "Estimating the impact of extreme events on crude oil price: An EMD-based event analysis method." *Energy Economics* 31.5 (2009): 768-778.
- [41] Rilling, Gabriel, Patrick Flandrin, and Paulo Goncalves. "On empirical mode decomposition and its algorithms." *IEEE-EURASIP workshop on nonlinear signal and image processing*. Vol. 3. NSIP-03, Grado (I), 2003.
- [42] Torres, María Eugenia, et al. "A complete ensemble empirical mode decomposition with adaptive noise." *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on*. IEEE, 2011.
- [43] Zhang, Jian, et al. "Performance enhancement of ensemble empirical mode decomposition." *Mechanical Systems and Signal Processing* 24.7 (2010): 2104-2123.
- [44] Lei, Yaguo, et al. "Fault diagnosis of rotating machinery based on an adaptive ensemble empirical mode decomposition." *Sensors* 13.12 (2013): 16950-16964.
- [45] Komaty, Ali, Abdel Boudraa, and Delphine Dare. "Emd-based filtering using the Hausdorff distance." *Signal Processing and Information*

- Technology (ISSPIT), 2012 IEEE International Symposium on.* IEEE, 2012.
- [46] Jolliffe, Ian. *Principal component analysis*. John Wiley & Sons, Ltd, 2005.
- [47] Colosimo, Bianca Maria, and Massimo Pacella. "On the use of principal component analysis to identify systematic patterns in roundness profiles." *Quality and reliability engineering international* 23.6 (2007): 707-725.
- [48] Nomikos, Paul, and John F. MacGregor. "Multivariate SPC charts for monitoring batch processes." *Technometrics* 37.1 (1995): 41-59.
- [49] Šidák, Zbyněk. "Rectangular confidence regions for the means of multivariate normal distributions." *Journal of the American Statistical Association* 62.318 (1967): 626-633.
- [50] Dunn, Olive Jean. "Multiple comparisons among means." *Journal of the American Statistical Association* 56.293 (1961): 52-64.
- [51] Jain, Anil K., M. Narasimha Murty, and Patrick J. Flynn. "Data clustering: a review." *ACM computing surveys (CSUR)* 31.3 (1999): 264-323.
- [52] Milligan, Glenn W., and Martha C. Cooper. "An examination of procedures for determining the number of clusters in a data set." *Psychometrika* 50.2 (1985): 159-179.
- [53] Mardia, Kantilal Varichand, John T. Kent, and John M. Bibby. *Multivariate analysis*. Academic press, 1979.
- [54] Zhao, Qinpei, Mantao Xu, and Pasi Fränti. "Sum-of-squares based cluster validity index and significance analysis." *Adaptive and Natural Computing Algorithms*. Springer Berlin Heidelberg, 2009. 313-322.
- [55] Dunn, J. C. (1973). "A Fuzzy Relative of the ISODATA Process and Its Use in Detecting Compact Well-Separated Clusters". *Journal of Cybernetics* 3 (3): 32–57.
- [56] Davies, David L.; Bouldin, Donald W. (1979). "A Cluster Separation Measure". *IEEE Transactions on Pattern Analysis and Machine Intelligence*. PAMI-1 (2): 224–227.
- [57] Kaufman L., and P. J. Rousseeuw. *Finding Groups in Data: An Introduction to Cluster Analysis*. Hoboken, NJ: John Wiley & Sons, Inc., 1990.
- [58] Ball, Geoffrey H., and David J. Hall. *ISODATA, a novel method of data analysis and pattern classification*. STANFORD RESEARCH INST MENLO PARK CA, 1965.

- 
- [59] Caliński, Tadeusz, and Jerzy Harabasz. "A dendrite method for cluster analysis." *Communications in Statistics-theory and Methods* 3.1 (1974): 1-27.
- [60] Hartigan, John A. "Clustering algorithms." (1975).
- [61] Thorndike, Robert L. "Who belongs in the family?." *Psychometrika* 18.4 (1953): 267-276.
- [62] Tibshirani, Robert, Guenther Walther, and Trevor Hastie. "Estimating the number of clusters in a data set via the gap statistic." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 63.2 (2001): 411-423.
- [63] Peres-Neto, Pedro R., Donald A. Jackson, and Keith M. Somers. "How many principal components? Stopping rules for determining the number of non-trivial axes revisited." *Computational Statistics & Data Analysis* 49.4 (2005): 974-997.
- [64] Wold, Svante. "Cross-validatory estimation of the number of components in factor and principal components models." *Technometrics* 20.4 (1978): 397-405.
- [65] Kaiser, Henry F. "The application of electronic computers to factor analysis." *Educational and psychological measurement* (1960).
- [66] Cattell, Raymond B. "The scree test for the number of factors." *Multivariate behavioral research* 1.2 (1966): 245-276.