

ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA
SEDE DI BOLOGNA

Scuola di Ingegneria ed Architettura
Corso di Laurea Magistrale in Ingegneria Elettronica

TECNICHE DI MACHINE LEARNING PER IL
MONITORAGGIO DELLA LINEA DI
POTENZA FERROVIARIA

Elaborato in
Elaborazione dei segnali nei sistemi elettronici M

Laureando:
ANDREA ZANELLINI

Relatore:
Prof. MAURO MANGIA

Correlatori:
Prof. RICCARDO ROVATTI
Dott. ALEX MARCHIONI

ANNO ACCADEMICO 2020-2021

PAROLE CHIAVE

Catenaria

Analisi spettrale

Dataset sintetici

PCA

KPCA

RPCA

Distanza di Mahalanobis

Indice

Introduzione	7
1 Contesto applicativo	9
1.1 Il sistema della catenaria	9
1.2 Il pantografo	11
1.3 Il filo di contatto	12
1.4 Sistema di misura	16
2 Stima di spettro e analisi tempo-frequenza	21
2.1 Spettro di potenza	21
2.2 Stimatore a periodogramma	22
2.3 Stimatore a periodogramma modificato	25
2.4 Finestre di apodizzazione	26
2.5 Stimatore a minima varianza	29
2.6 Analisi tempo-frequenza	32
2.7 STFT	33
2.8 CWT	35
3 Rilevazione di anomalie	41
3.1 PCA	41
3.2 KPCA	45
3.3 RPCA	49
3.4 Score	51
3.5 ROC	55

4	Generazione di dataset sintetici	59
4.1	Costruzione dei dataset	60
4.2	Composizione del passaggio	63
3.2	Generazione di anomalie	66
5	Implementazione	73
5.1	Analisi spettrale	73
5.2	Generazione dataset di train	83
5.3	Generazione dataset di test	84
5.4	Detectors basati sulla PCA	86
5.5	Detectors basati sulla KPCA	96
5.6	Detectors basati sulla RPCA	100
5.7	Rilevazione di usura del filo di contatto	106
6	Conclusioni	111
	Ringraziamenti	115
	Elenco Figure	117
	Bibliografia	121

Introduzione

Con l'avvento dell'intelligenza artificiale, il monitoraggio di infrastrutture come ponti, grattacieli e linee ferroviarie, è diventato un tema di forte interesse ingegneristico. Posizionando dei sensori in punti strategici della struttura, è possibile acquisire periodicamente delle misure che permettono di valutarne lo stato di salute. L'elevato numero di sistemi di misura, congiuntamente ai lunghi tempi di monitoraggio, comporta la generazione di un'enorme mole di dati che necessita di particolari algoritmi per essere gestita e sfruttata proficuamente.

L'azienda RFI, gestore di gran parte dell'infrastruttura ferroviaria italiana, porta avanti da anni progetti di ricerca in collaborazione con le università, allo scopo di migliorare la qualità e la sicurezza del servizio offerto ai passeggeri ma anche di ridurre i costi di gestione. Nel 2020 ha avuto inizio il progetto con l'Università di Bologna di sensorizzare la linea e di strutturare delle tecniche di monitoraggio atte a rilevare la presenza di eventuali anomalie e a stimare il livello di usura del filo di contatto. Quest'ultimo obiettivo risulta essere particolarmente importante per l'azienda poiché permetterebbe di programmare con precisione la sostituzione del filo o intervenire tempestivamente in caso di un comportamento anomalo.

Le tecniche oggetto di questo elaborato afferiscono a quella branca del vasto del mondo del machine learning nota come "apprendimento non supervisionato", la cui caratteristica distintiva è quella di operare senza conoscere a priori le possibili anomalie. Questi algoritmi risultano particolarmente utili in una prima fase esplorativa, quando ancora non si conosce il sistema da monitorare e si ha la necessità di rilevare una serie di possibili eventi anomali. Ove fosse necessario, si potrebbe procedere caratterizzando quanto rilevato in fase esplorativa e allenando in

maniera supervisionata altre tipologie di algoritmi, in grado di cercare più efficacemente dei target noti.

La capacità di analisi dei risultati ottenuti e la corretta applicazione delle tecniche menzionate richiede una giusta conoscenza del sistema meccanico in esame, a partire dalla nomenclatura e dalla topologia fino ad arrivare al più rilevante comportamento dinamico. Infatti, il passaggio del treno da luogo ad un moto oscillatorio che, statisticamente, può essere modellato con un processo aleatorio non stazionario. In generale, il moto oscillatorio nel tempo è contraddistinto da un certo spettro nel dominio trasformato il cui studio risulta essere molto utile al fine di caratterizzare il processo. In questo elaborato verranno inizialmente affrontati gli aspetti teorici relativi al sistema meccanico e agli strumenti di machine learning, che verranno poi applicati nella seconda parte della trattazione.

Capitolo 1

Contesto applicativo

Lo scopo di questo capitolo è quello di descrivere la struttura meccanica della rete ferroviaria al fine di contestualizzare l'applicazione delle tecniche di elaborazione dell'informazione e evidenziare alcuni importanti aspetti che saranno particolarmente utili per l'analisi dei dati.

1.1 Il sistema della catenaria

Il sistema catenaria rappresenta l'insieme delle condutture che distribuiscono l'energia elettrica sulla linea e la trasmettono ai treni per mezzo di dispositivi di captazione di corrente, i cosiddetti pantografi [1]. Il nome deriva dalla forma geometrica assunta dalla corda portante, la quale sostiene il filo di contatto su cui scorre il contatto strisciante. Questo alimenta il motore elettrico posto a bordo del treno e consente lo spostamento dello stesso.

La struttura geometrica della catenaria consente di garantire una certa forza di contatto tra il pantografo e il cavo di alimentazione che è di fondamentale importanza al fine di trasmettere al treno la massima corrente possibile. La propagazione delle onde nei cavi rappresenta infatti un problema ben noto che costituisce il limite principale alla velocità dei treni [2]. Esistono diverse varianti della struttura, esse differiscono ad esempio per la composizione geometrica dei vari elementi o per il sistema sospensivo. In ogni caso, questa descrizione vuole essere un inquadramento generale dei vari componenti in gioco e delle loro funzioni.

Il sistema è formato dai seguenti componenti:

1. palo tubolare con attacchi snodati;
2. tirante;
3. morsetto di attacco per la fune;
4. isolatore;
5. staffa di sostegno per la fune;
6. mensola tubolare;
7. braccio di poligonazione;
8. tirantini di poligonazione;
9. morsetti di attacco;

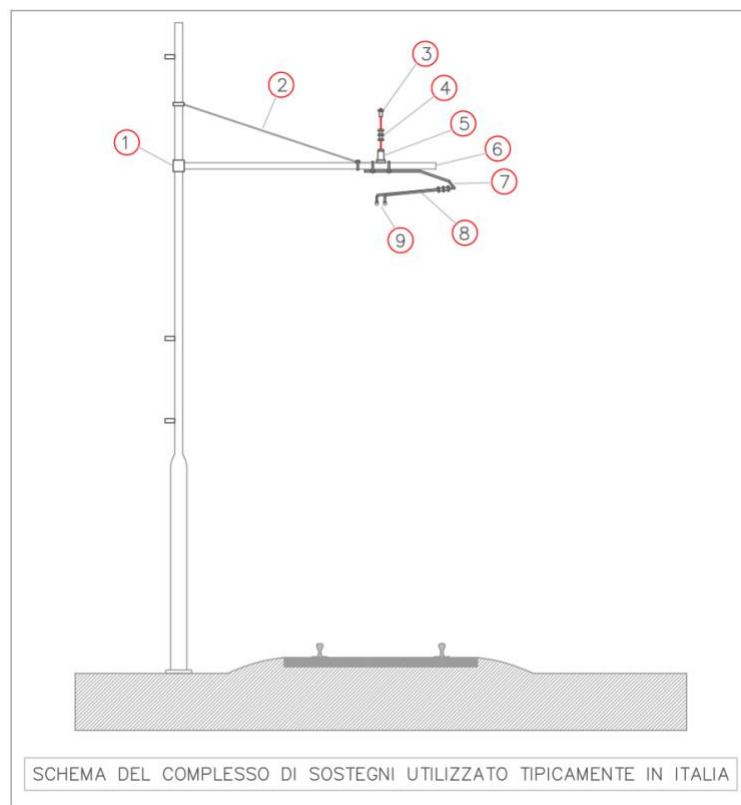


Figura 1 - Schematizzazione del complesso di sostegni utilizzato tipicamente in Italia.

10. filo di contatto;
11. pendini;

12. fune portante.

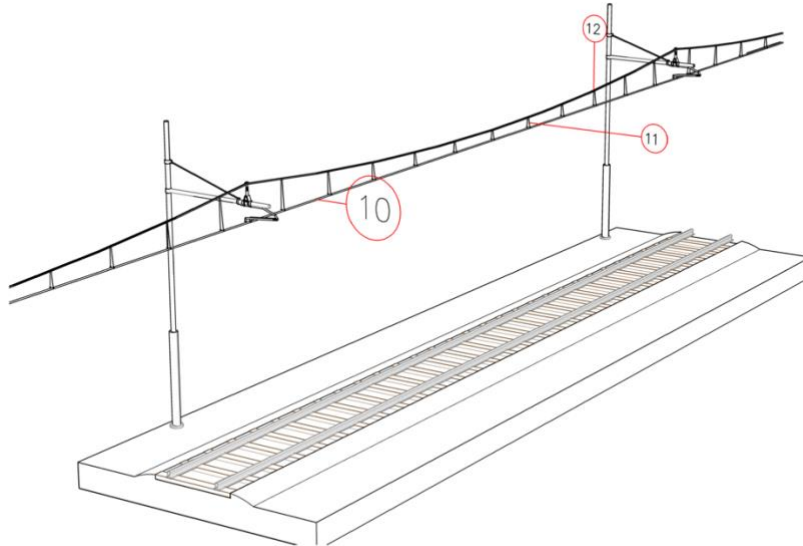


Figura 2 - Dettaglio in vista laterale del complesso di cavi.

Il palo sorregge, tramite la mensola tubolare, il complesso formato dal braccio di poligonazione e il tirantino. All'estremità del tirantino sono posti i morsetti di attacco che a loro volta sono ancorati al filo di contatto. Per limitare l'oscillazione del cavo al passaggio del treno, alle varie mensole è agganciata una fune di sostegno la quale vincola il cavo di contatto tramite i cosiddetti pendini.

1.2 Il pantografo

Con il termine pantografo si definisce, in Italia, un dispositivo di captazione dell'energia elettrica da una linea aerea di contatto, posto sull'imperiale di un veicolo ferroviario, tranviario o a guida vincolata, atto ad alimentarne le apparecchiature elettriche [5]. Esso è composto generalmente da una coppia di parallelogrammi a bracci snodabili alla cui estremità è posto un pattino che, strisciando sul filo di contatto, consente l'alimentazione del treno.



Figura 3 - Esempio di pantografo moderno.

La caratteristica principale che deve avere questo componente è quella di riuscire ad esercitare una forza verticale sufficiente ad assicurare sempre il contatto elettrico. Le oscillazioni del filo di alimentazione fanno sì che il pantografo debba spostarsi verticalmente per evitare l'interruzione del contatto elettrico. Con l'aumento della velocità dei treni crescono anche le sollecitazioni del sistema e conseguentemente risulta più complicato garantire questo aspetto. Le ultime versioni di pantografo si avvalgono sia di una spinta pneumatica ma anche di una spinta aerodinamica generata da apposite alette.

1.3 Il filo di contatto

Il filo di contatto è realizzato in rame ed è sagomato in modo tale da consentire l'attacco dei morsetti e allo stesso tempo fornire una buona superficie di contatto per la spazzola del pantografo.

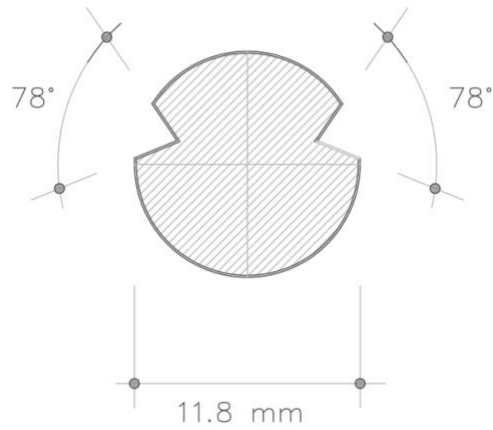


Figura 4 - Sezione tipica del filo di contatto.

Per limitare la curvatura del cavo sotto il suo stesso peso, viene posto in tensione tramite un sistema di masse e carrucole che esercitano una funzione di contrappeso. Al fine di contenere l'oscillazione è importante che il complesso di masse e carrucole mantenga il filo ad una tensione costante, cioè quella prevista in fase di progettazione.

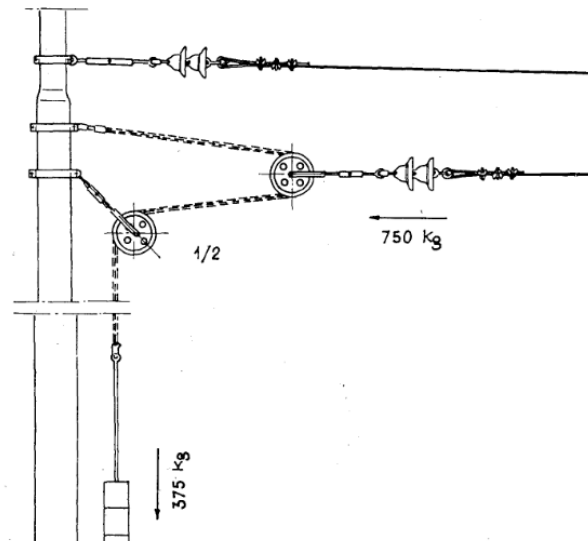


Figura 5 - Esempio di sistema sospensivo dei tiranti e del filo di contatto.

Il contatto strisciante, a causa dell'attrito tra la spazzola e il filo, tende a generare calore e a consumare le superfici di entrambe le parti. Per questo, il filo di contatto non è mai complanare con la rotaia, bensì è disposto a “zig-zag”; così facendo il punto di contatto sulla spazzola continuerà a variare, usurandola uniformemente. La superficie del filo di contatto si consuma di una certa percentuale ad ogni passaggio del pantografo. Una volta raggiunte specifiche soglie di spessore residuo si interviene sostituendo il cavo, prevenendo così possibili rotture che rappresentano un danno molto grave al servizio, sia per la sicurezza dei passeggeri, sia per i costi e i tempi di ripristino.

Diametro nominale [mm]	Diametro1	Diametro2	Diametro3
Tiro [kg]	< Tiro 1	Tiro 2	Tiro 3
Soglia di attenzione [mm]	$1A_i < s < 1A_s$	$2A_i < s < 2A_s$	$3A_i < s < 3A_s$
Soglia di Intervento a 30 giorni [mm]	$1I_i < s < 1I_s$	$2I_i < s < 2I_s$	$3I_i < s < 3I_s$
Soglia di difetto rilevante [mm]	$s < 2R_s$	$s < 2R_s$	$s < 3R_s$

Figura 6 - Tabelle di intervento e spessore residuo (i parametri inseriti sono stati indicati parametricamente a titolo d'esempio).

Chiaramente le soglie di intervento dipenderanno dal diametro del cavo, dal materiale, dal posizionamento e dal tensionamento. Per consentire il raggiungimento di velocità superiori è necessario ridurre l'ampiezza delle oscillazioni in modo tale da alzare la soglia a cui il pantografo si stacca dal contatto a causa della forza di inerzia verticale del cavo. Si procede aumentando la tensione del filo, che deve quindi essere composto da un materiale a maggior carico di rottura e/o di maggior diametro.

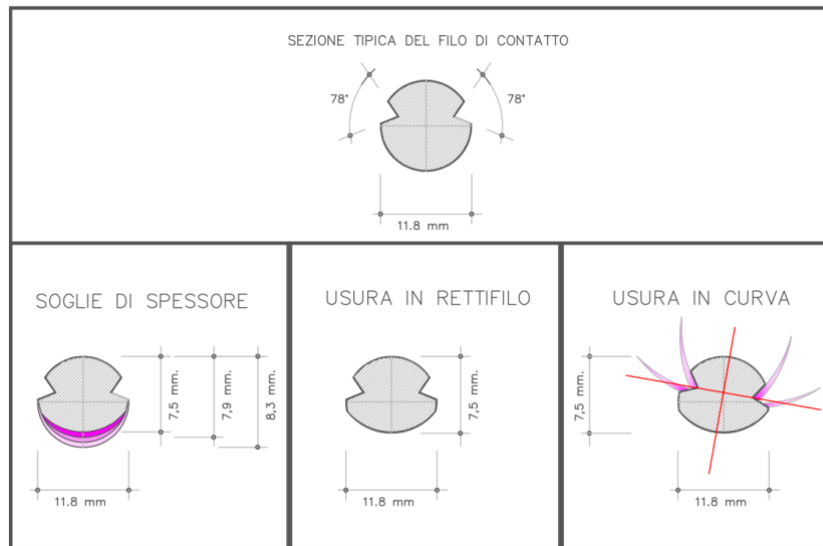


Figura 7 - Esempi di difetti dovuti all'usura.

Dal punto di vista fisico, si ha quindi una corda con estremi vincolati sulla quale, a seguito del passaggio del treno, si instaurano onde stazionarie [14]:

$$\xi_n(x, t) = A_n \sin k_n x \sin(\omega_n t + \varphi_n)$$

Dove con A_n si indica l'ampiezza dell'onda, k_n il numero d'onda, x la coordinata spaziale, t quella temporale, ω_n la pulsazione dell'onda e φ_n la sua fase.

Dall'imposizione dei due vincoli $\xi_n(0, t) = 0$ e $\xi_n(L, t) = 0$ si ottiene la seguente condizione sui numeri d'onda:

$$k_n = \frac{n\pi}{L} \text{ con } n = 1, 2, 3, ..$$

Con L lunghezza del cavo in metri.

Le autofrequenze in questo caso particolare sono quindi date da:

$$f_n = \frac{nv}{2L} \text{ con } n = 1, 2, 3, ..$$

La velocità di propagazione delle onde dipende dalla tensione del filo T [kg] e dalla densità lineare μ [$\frac{kg}{m}$] ed è data da:

$$v = \sqrt{\frac{T}{\mu}}$$

Combinando l'espressione della velocità con quella delle autofrequenze si ha:

$$f_n = \frac{n}{2L} \sqrt{\frac{T}{\mu}} \text{ con } n = 1, 2, 3, ..$$

Come anticipato, l'assottigliamento del cavo comporta una diminuzione della densità lineare e di conseguenza, a parità di tensione, un aumento di frequenza.

1.4 Il sistema di misura

L'obiettivo di monitorare un sistema meccanico come quello in questione può essere raggiunto in svariati modi. Il principio di base è quello di posizionare un certo numero di sensori che acquisiscono sequenze temporali con una certa frequenza di campionamento. La posizione, il numero e la tipologia dei sensori influenza sicuramente l'efficacia del monitoraggio ma risulta altrettanto importante la capacità di estrarre l'informazione acquisita sfruttando algoritmi di machine learning.

Per quanto riguarda la linea gestita da RFI, la prospettiva è quella di posizionare un certo numero di braccetti sensorizzati. Chiaramente questi braccetti avranno un costo superiore rispetto alla versione standard ma rappresentano una soluzione scalabile e quindi potenzialmente implementabile su vasta scala. Le acquisizioni dei braccetti sensorizzati potrebbero poi essere inviate ad un gateway che a sua volta compie delle elaborazioni e poi inoltra le informazioni utili verso un centro di calcolo e

stoccaggio. Il carico computazionale dovrà essere distribuito nella maniera più conveniente tra il gateway e il centro di calcolo ed eventualmente prevedere una certa elaborazione in prossimità dei sensori. Una volta definito il sistema dal punto di vista hardware sarà necessario individuare una strategia di acquisizione che permetta di catturare sequenze di una lunghezza tale da consentire il corretto monitoraggio del sistema, senza appesantire inutilmente l'elaborazione o richiedere unità di stoccaggio troppo capienti. In questo elaborato si è ipotizzato che il campionamento venisse attivato dall'arrivo del treno e che durasse per un certo tempo fissato, ad esempio 60s.

Ciò detto, per questa fase di pre-analisi non è stato possibile disporre del sistema di misura con braccetti strumentati e pertanto sono state acquisite alcune sequenze esemplificative tramite un potenziometro lineare ancorato da una parte ad una staffa di sostegno solidale al terreno e dall'altra al braccetto.

Di seguito una breve sequenza di immagini riassume il contesto in cui vengono acquisiti i passaggi:

- A riposo, la linea è pressoché ferma (o lievemente sollecitata dall'effetto del vento) e il potenziometro lineare riporta uno spostamento verticale nullo:

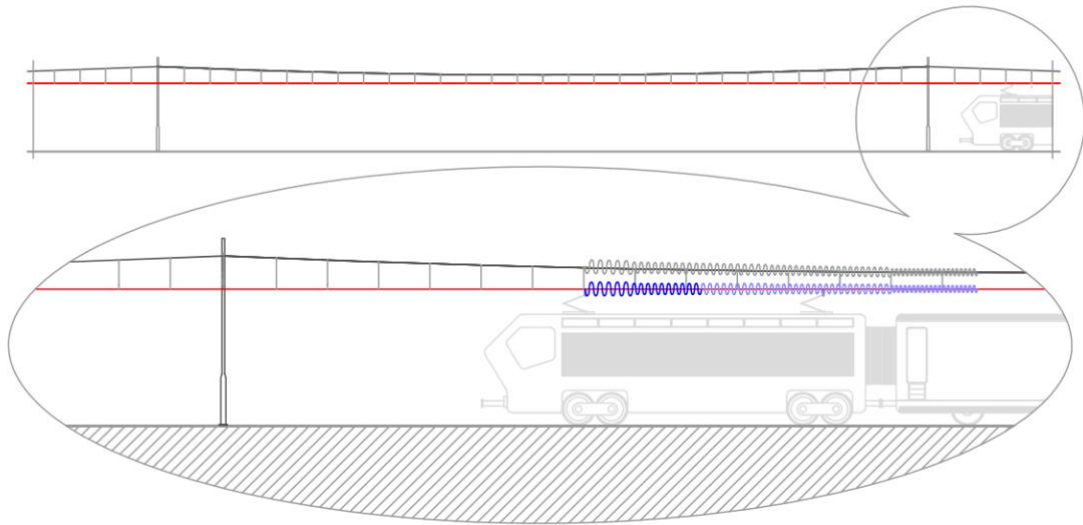


Figura 8 - Rappresentazione di quanto avviene al passaggio del pantografo. La frequenza delle oscillazioni è a puro scopo esemplificativo.

- All'arrivo del treno si origina un'oscillazione sempre più consistente che culmina con un picco in corrispondenza del passaggio del pantografo:

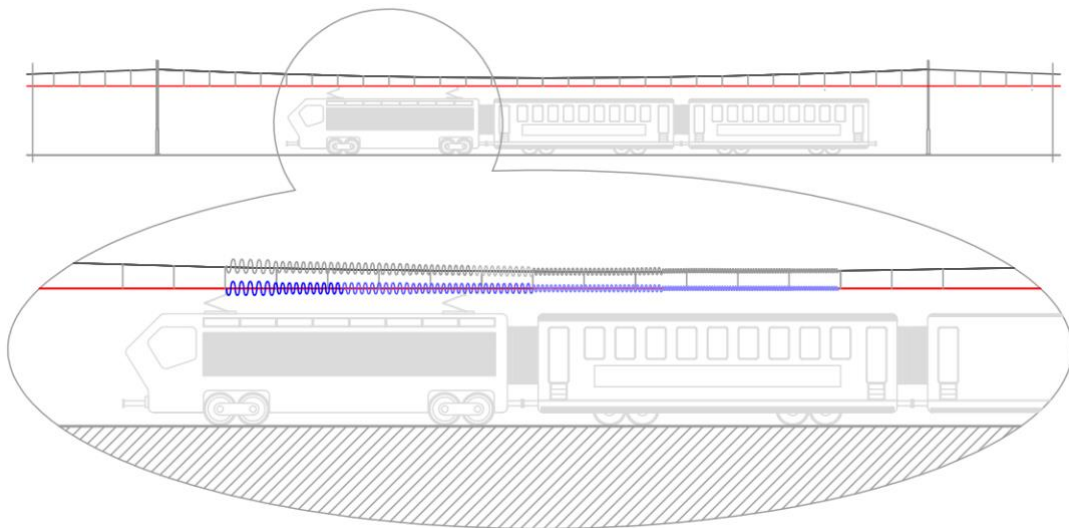


Figura 9 - Rappresentazione di quanto avviene al passaggio del pantografo. La frequenza delle oscillazioni è a puro scopo esemplificativo.

- A valle del transito, la linea oscilla per minuti con ampiezza sempre minore fino a ritornare alla condizione di riposo. In questa fase, il potenziometro lineare segue le oscillazioni della linea contraendosi ed estendendosi:

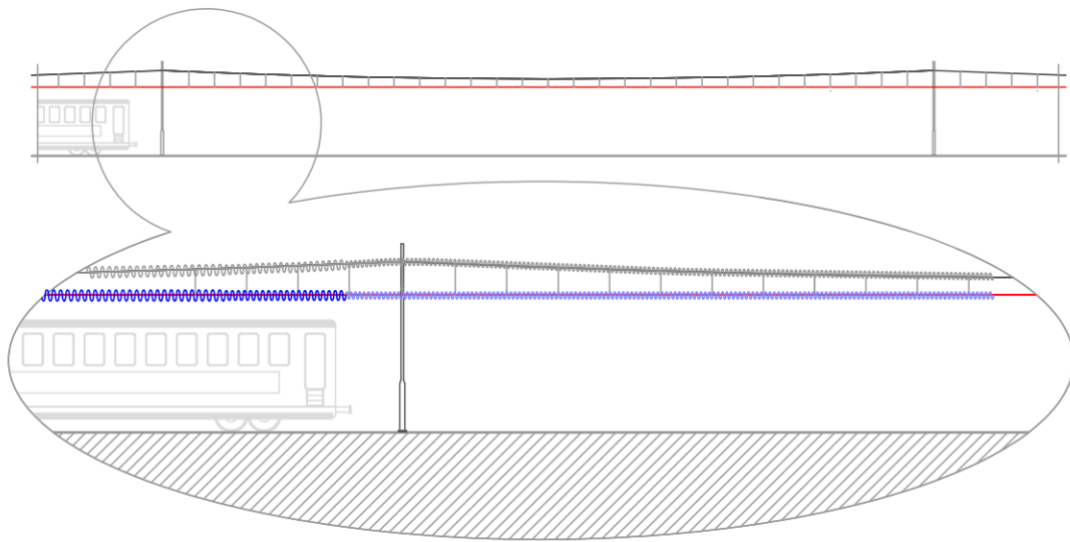


Figura 10 - Rappresentazione di quanto avviene al passaggio del pantografo. La frequenza delle oscillazioni è a puro scopo esemplificativo.

Capitolo 2

Stima di spettro e analisi tempo frequenza

Il sistema catenaria è di fatto un sistema meccanico elastico, in grado di oscillare ad una frequenza naturale che dipende dalle caratteristiche geometriche della struttura, dai materiali e da fattori ambientali come ad esempio la temperatura esterna. Risulta pertanto di particolare interesse analizzare nel dominio delle frequenze i segnali prodotti dal sistema di misura. In questo capitolo verranno inizialmente trattati gli aspetti teorici relativi allo spettro di potenza e ad alcuni metodi di stima, in secondo luogo verranno introdotti i concetti fondamentali dell'analisi tempo-frequenza ed i due principali strumenti volti a perseguire questa indagine.

2.1 Spettro di potenza

Dato un processo stocastico tempo-discreto x i cui campioni sono separati da un intervallo temporale P , tale che ogni sua realizzazione è a potenza finita, si definisce “Densità Spettrale di Potenza” o brevemente PDS la funzione [6]:

$$S_x^W(f) = P L_{\frac{1}{P}}(f) \lim_{N \rightarrow \infty} \frac{1}{2N + 1} \mathbb{E}[|\mathcal{F}[L_{2N}x](f)|^2]$$

Con $2N + 1$ campioni considerati. Nel caso tempo-discreto, la trasformata di Fourier da luogo ad una ripetizione periodica dello spettro ed è necessario dunque moltiplicare per la funzione $L_{\frac{1}{P}}(f)$ che ne limita il supporto all'intervallo $[-\frac{1}{2P}, \frac{1}{2P}]$.

La moltiplicazione per P dell'intera espressione riconduce il risultato ad una densità di potenza, la cui unità di misura è dunque i joule. In questa espressione sono anche presenti un limite per N che tende ad infinito e l'operatore aspettazione. Dal punto di vista pratico queste due operazioni non sono realizzabili poiché sarebbe necessario conoscere rispettivamente infiniti campioni e l'intera caratterizzazione statistica del processo. Perciò si parla di *stima* di spettro e non di *misura*.

2.2 Periodogramma

L'approccio di questo stimatore è quello di rimuovere dall'espressione della densità spettrale di potenza i due passaggi che la rendono inapplicabile nella pratica.

Si ottiene quindi la seguente espressione:

$$\hat{S}_x^W(f) = P L_{\frac{1}{P}}(f) \frac{1}{2N+1} |\mathcal{F}[L_{2N}x](f)|^2$$

che può essere espansa ottenendo:

$$\hat{S}_x^W(f) = P L_{\frac{1}{P}}(f) \frac{1}{2N+1} \sum_{j=-N}^N \sum_{k=-N}^N x_j^* x_k e^{2\pi i j P f} e^{-2\pi i k P f}$$

Naturalmente verrà introdotto un certo errore di stima che può essere quantificato tramite i concetti di polarizzazione e non consistenza.

2.2.1 Polarizzazione

$$\mathbb{E}[\hat{S}_x^W(f)] = P L_{\frac{1}{P}}(f) \frac{1}{2N+1} \sum_{j=-N}^N \sum_{k=-N}^N \mathbb{E}[x_j^* x_k] e^{2\pi i j P f} e^{-2\pi i k P f} =$$

$$= L_{\frac{1}{P}}(f) \frac{P}{2N+1} \sum_{j=-N}^N \sum_{k=-N}^N \mathbb{E}[x_j^* x_k] e^{-2\pi i(k-j)Pf}$$

limitandosi per semplicità al caso stazionario si ha che $\mathbb{E}[x_j^* x_k] = C_x(k-j)$:

$$\mathbb{E}[\hat{S}_x^W(f)] = L_{\frac{1}{P}}(f) \frac{P}{2N+1} \sum_{j=-N}^N \sum_{k=-N}^N C_x(k-j) e^{-2\pi i(k-j)Pf}$$

ponendo ora $d = k - j$ si possono riscrivere le sommatorie:

$$\begin{aligned} \mathbb{E}[\hat{S}_x^W(f)] &= L_{\frac{1}{P}}(f) \frac{P}{2N+1} \sum_{d=-2N}^{2N} (2N+1-|d|) C_x(d) e^{-2\pi i d P f} = \\ &= L_{\frac{1}{P}}(f) P \sum_{d=-2N}^{2N} \left(1 - \frac{|d|}{2N+1}\right) C_x(d) e^{-2\pi i d P f} = \\ &= L_{\frac{1}{P}}(f) P \sum_{d=-\infty}^{\infty} B_N(d) C_x(d) e^{-2\pi i d P f} \end{aligned}$$

idealmente, se la funzione $B_N(d) = \left(1 - \frac{|d|}{2N+1}\right)$ fosse uguale ad 1 si avrebbe che, per il teorema di Wiener-Khinchine, l'aspettazione della stima coincide con lo spettro vero. Si noti però che tale risultato è ottenuto per $N \rightarrow \infty$:

$$\lim_{N \rightarrow \infty} B_N(d) = \lim_{N \rightarrow \infty} \left(1 - \frac{|d|}{2N+1}\right) = 1$$

Si deduce quindi che lo stimatore a periodogramma è sì polarizzato, ma asintoticamente tende a coincidere con lo spettro vero. Questa proprietà è da tenere in considerazione quando si applica ad esempio ad una sequenza temporale, maggiore sarà il numero dei campioni della serie minore sarà la polarizzazione dello

stimatore. Al contrario, l'applicazione a sequenze troppo corte comporterebbe un errore di stima elevato.

2.2.2 Non consistenza:

Per lo studio del grado di consistenza dello stimatore è conveniente svolgere la trattazione a partire da un processo gaussiano bianco. In primo luogo, questa scelta rende lo stimatore non polarizzato e permette di focalizzare i calcoli solo sulla varianza. La trattazione che porta al risultato finale è piuttosto laboriosa e pertanto si è preferito sintetizzare alcuni passaggi chiave.

Lo spettro reale di un processo gaussiano bianco è costante all'interno della banda e si indica tale valore con P . Si consideri ora la covarianza tra la stima di due frequenze differenti, rispettivamente f' e f'' :

$$\begin{aligned} \gamma_{\hat{S}_x^W(f')\hat{S}_x^W(f'')} &= \mathbb{E}[\hat{S}_x^W(f')\hat{S}_x^W(f'')] - \mathbb{E}[\hat{S}_x^W(f')]\mathbb{E}[\hat{S}_x^W(f'')] = \\ &= \mathbb{E}[\hat{S}_x^W(f')\hat{S}_x^W(f'')] - P^2 \end{aligned}$$

Sostituendo l'espressione di $\hat{S}_x^W(f)$, sviluppando le varie sommatorie e propagando opportunamente l'aspettazione si ottiene il seguente risultato:

$$\gamma_{\hat{S}_x^W(f')\hat{S}_x^W(f'')} = \frac{P^2}{(2N+1)^2} \frac{\sin^2((2N+1)\pi P(f' - f''))}{\sin^2(\pi P(f' - f''))}$$

Come si può notare, la covarianza risulta nulla per tutti gli $f' = k_1/(2N+1)P$ e $f'' = k_2/(2N+1)P$ con $k_1, k_2 \in \mathbb{Z}$ tali che $k_1 - k_2$ non è multiplo di $2N+1$. Si ha quindi che, contrariamente a quanto ci si aspetta per un processo gaussiano bianco, le stime delle due frequenze sono incorrelate ed è impossibile predirne linearmente una conoscendo l'altra.

Per $f', f'' \rightarrow f$ si ottiene la varianza associata allo stimatore:

$$\sigma_{\hat{s}_x^W(f)}^2 = \lim_{f', f'' \rightarrow f} \gamma \hat{s}_x^W(f') \hat{s}_x^W(f'') = P^2 \frac{(2N+1)^2}{(2N+1)^2} = P^2$$

A differenza della polarizzazione, la varianza purtroppo non decresce con N e lo stimatore è pertanto non consistente. La non consistenza dello stimatore comporta un effetto pratico che contraddistingue le stime di spettro effettuate con periodogramma, cioè la presenza di una seghettatura che, in alcuni casi, rende lo spettro di difficile interpretazione.

2.3 Periodogramma modificato

Il problema principale dello stimatore a periodogramma è la non consistenza, tuttavia è possibile attuare una modifica che permette di ridurla significativamente nel caso di processi stazionari. L'idea è quella di suddividere la sequenza di campioni in M finestre di lunghezza tale da poter assumere che il contenuto di ciascuna di esse sia indipendente dal contenuto delle altre finestre. Applicando il periodogramma in ogni finestra si ottengono M differenti risultati s_j che possono poi essere mediati tra loro:

$$s = \frac{1}{M} \sum_{j=0}^{M-1} s_j$$

dal punto di vista della varianza si ottiene:

$$\sigma_s^2 = \frac{1}{M} \sum_{j=0}^{M-1} \sigma_{s_j}^2$$

si nota facilmente che all'aumentare del numero di finestre, la varianza viene ridotta.

Purtroppo, questo risultato non è gratuito, se da una parte si è ridotta la varianza, dall'altra si è suddiviso il segnale in finestre, riducendo di un fattore M il numero di campioni N che vengono elaborati dallo stimatore e peggiorando di conseguenza la polarizzazione. Al fine di ottenere una buona stima tramite periodogramma modificato, disporre di un numero elevato di campioni rappresenta una condizione necessaria.

Un ulteriore accorgimento da prendere in considerazione è l'introduzione di una certa sovrapposizione tra le finestre. Così facendo si hanno chiaramente delle sotto-sequenze correlate tra loro ma viene virtualmente aumentato il numero di campioni della sequenza, comportando un miglioramento in termini di polarizzazione.

2.4 Finestre di apodizzazione

L'operazione di finestrazione, introdotta con il periodogramma modificato, aggiunge un interessante grado di libertà inerente alla "forma" della funzione finestra. È possibile dimostrare che alcune tipologie di finestre riducono la polarizzazione dello stimatore migliorandone la localizzazione in frequenza mentre altre consentono invece di ottenere una stima più precisa della PSD nell'intorno di una certa frequenza. Poiché la funzione finestra è a supporto compatto, è sempre possibile esprimerla rispetto alla finestra rettangolare, il cui spettro di ampiezza è dato dal cosiddetto "kernel di Dirichlet":

$$X[k] = A_0 \frac{\sin(\pi k)}{\sin\left(\frac{\pi k}{N}\right)}$$

Qualsiasi finestra a supporto compatto sarà quindi contraddistinta da un lobo principale e da tanti lobi secondari.

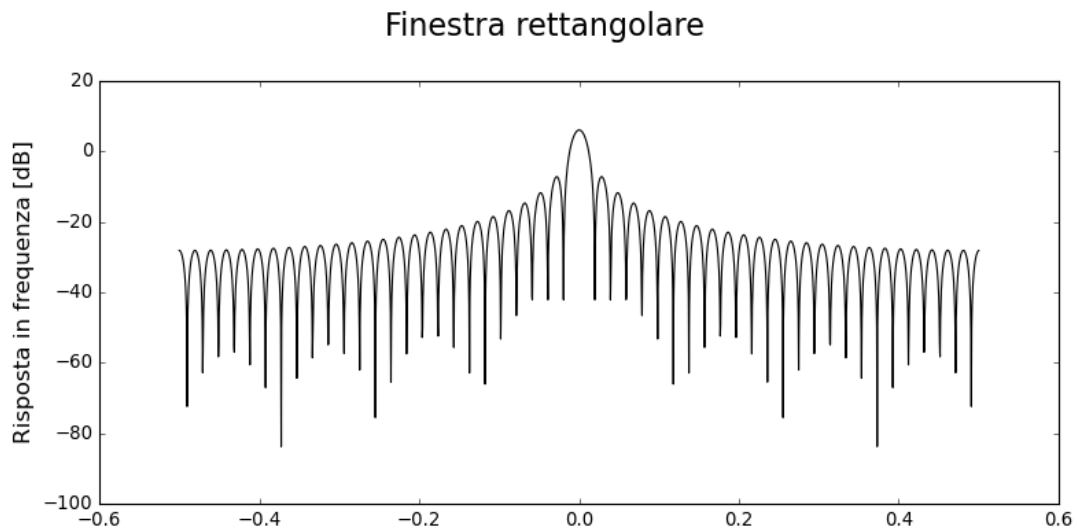


Figura 11 - Spettro d'ampiezza della finestra rettangolare.

Osservando la funzione si possono determinare tre parametri chiave che variano da finestra a finestra:

- Larghezza del lobo principale
- Ampiezza lobi secondari
- Roll-off rate

La larghezza del lobo principale determina la capacità della finestra di localizzare in frequenza, ad esempio nel caso in cui si voglia stimare con precisione la posizione di un picco. In merito a questo aspetto, la finestra rettangolare è quella che presenta la minima larghezza del lobo principale. Il roll-off rate, che rappresenta la velocità di decadimento della ampiezza dei lobi, unitamente alla loro ampiezza, determina la precisione di stima della PDS ad una data frequenza. Si noti inoltre che la lunghezza della finestra vincola la minima frequenza che lo stimatore è in grado di distinguere o per meglio dire la risoluzione in frequenza. Di seguito vengono proposti due esempi:

Finestra di Blackman:

$$x[n] = a_0 - a_1 \cos\left(\frac{2\pi n}{N-1}\right) + a_2 \cos\left(\frac{4\pi n}{N-1}\right)$$

Con $a_0 = \frac{1-\alpha}{2}$, $a_1 = 0.5$, $a_2 = \frac{\alpha}{2}$ e $\alpha = 0.16$.

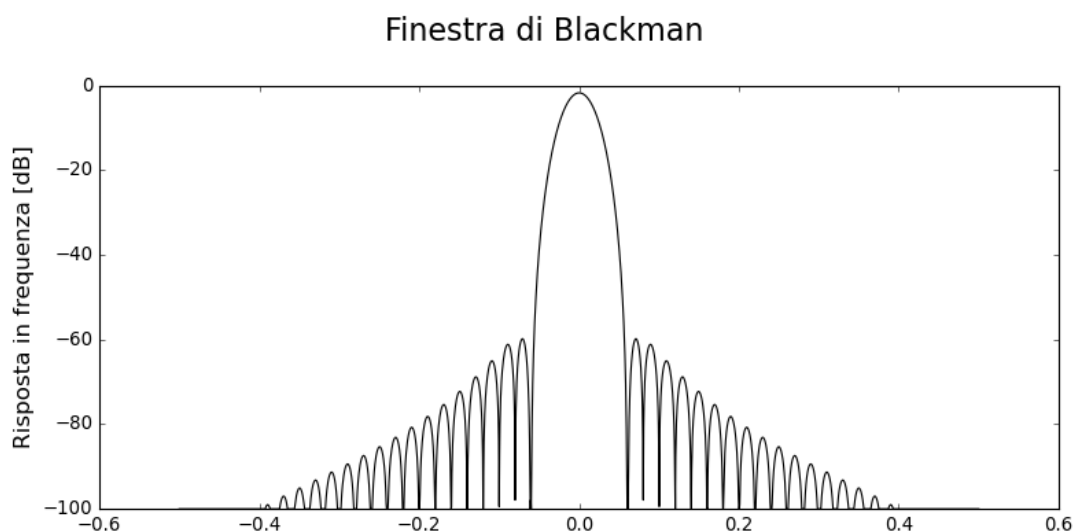


Figura 12 - Spettro d'ampiezza della finestra di Blackman.

Al contrario di quella rettangolare, la finestra di Blackman denota un lobo principale molto più largo, a fronte di un roll-off maggiore e in generale un'ampiezza dei lobi laterali inferiore. Questa finestra può essere applicata efficacemente nel caso si volesse stimare con precisione la PSD di uno spettro dolce (privo di picchi).

Finestra di Hamming:

$$x[n] = 0.54 - 0.46 \cos\left(\frac{2\pi n}{N-1}\right)$$

Questa variante può essere vista come un valido compromesso tra i due aspetti citati. Presenta un lobo principale sicuramente più stretto rispetto alla finestra di Blackman e un'ampiezza dei lobi secondari inferiore rispetto alla finestra rettangolare.

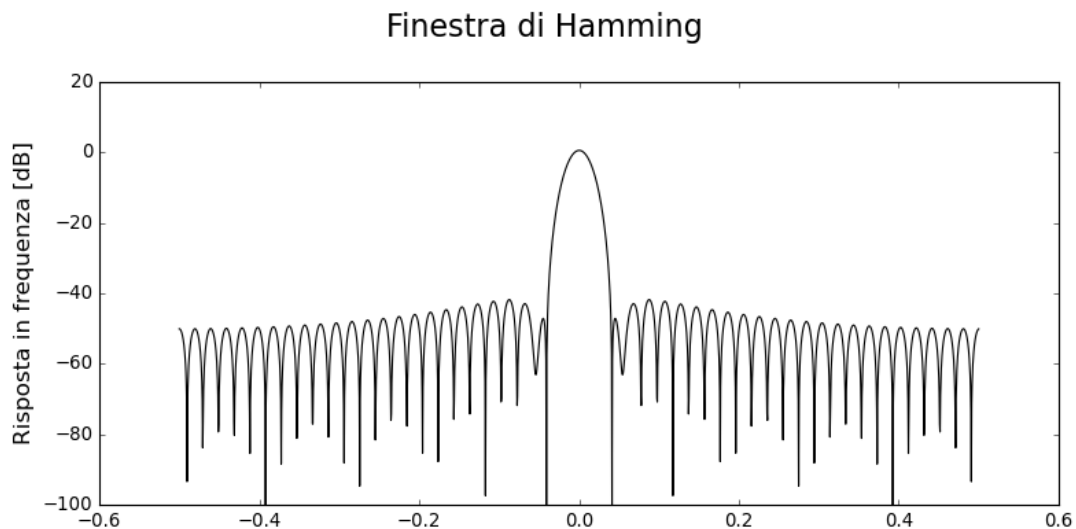


Figura 13 - Spettro d'ampiezza della finestra di Hamming.

2.5 Stimatore a minima varianza

L'approccio di questo stimatore è basato sulla suddivisione dello spettro del processo in intervalli di frequenze sui quali effettuare la stima della PSD. La selezione in frequenza viene realizzata tramite un filtro passa-banda. Di seguito viene esplicitata la relazione ingresso-uscita del filtro nel dominio del tempo, in cui si è indicato con $h = (h_0^*, h_1^*, \dots, h_{M-1}^*)^t$ il vettore dei coefficienti della risposta impulsiva e con $x_k = (x_k, x_{k-1}, \dots, x_{k-M+1})^t$ il vettore dei campioni del processo che, per semplicità, si assume essere stazionario:

$$w_k = \sum_{j=0}^{N-1} h_j x_{k-j} = h^\dagger x_k$$

Idealmente, per selezionare ogni singola frequenza, sarebbe necessario un filtro con funzione di trasferimento assimilabile ad una delta di Dirac. Nell'applicazione reale si ha a disposizione un numero finito di coefficienti del filtro che rende impossibile ottenere il risultato sopracitato. È possibile però fare in modo che la funzione di trasferimento del filtro sia il più possibile stretta nell'intorno della frequenza di interesse minimizzando la potenza in uscita e imponendo che la risposta in ampiezza del filtro sia unitaria. Così facendo, essendo l'ampiezza fissata, per minimizzare la potenza in uscita la funzione di trasferimento dovrà necessariamente stringersi in frequenza.

È opportuno quindi esprimere la potenza in uscita dal filtro in relazione ai coefficienti della risposta impulsiva:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \mathbb{E} \left[\sum_{k=-\frac{T}{2}}^{\frac{T}{2}} |w_k|^2 \right] = \mathbb{E}[|w_0|^2] = \mathbb{E}[h^\dagger x_0 x_0^\dagger h] = h^\dagger \mathbb{E}[x_0 x_0^\dagger] h = h^\dagger R_x h$$

Si noti la comparsa della matrice di correlazione del processo poiché questa contraddistinguerà alcuni importanti aspetti che verranno discussi più avanti.

In merito al vincolo sulla risposta in ampiezza, vengono raggruppati gli esponenziali complessi relativi alla trasformata di Fourier nel vettore

$$s = (1, e^{-2\pi i P f}, \dots, e^{-2\pi i (M-1) P f})^t:$$

$$\mathcal{F}[h](f) = \sum_{k=0}^{M-1} h_k e^{-2\pi i k P f} = s^t h^* = 1$$

Si ha dunque il seguente problema di minimizzazione vincolata:

$$\begin{cases} \min_h h^\dagger R_x h \\ s^\dagger h = 1 \end{cases}$$

La soluzione di questo problema può essere ottenuta sfruttando il teorema dei moltiplicatori di Lagrange:

$$\begin{cases} L(h, \lambda) = h^\dagger R_x h - \lambda^* (s^\dagger h - 1) \\ \nabla_{h^*} L(h, \lambda) = R_x h + \lambda s = 0 \end{cases}$$

tramite il quale si ha:

$$h = -\lambda R_x^{-1} s.$$

Sostituendo questo risultato nell'espressione del vincolo si ottiene:

$$s^\dagger h = s^\dagger (-\lambda R_x^{-1} s) = 1$$

da cui si può ricavare λ :

$$\lambda = \frac{1}{s^\dagger R_x^{-1} s}$$

La risposta impulsiva del filtro che meglio seleziona una determinata frequenza è data da:

$$h = \frac{R_x^{-1} s}{s^\dagger R_x^{-1} s}$$

Le ragioni elencate ad inizio del capitolo sull'impossibilità di misurare lo spettro di potenza si applicano anche in questo caso, rendendo impossibile una misura della matrice di correlazione. La stima della matrice inversa della matrice di correlazione è di norma un'operazione costosa in termini computazionali e per applicare questa tecnica è necessaria una certa potenza di calcolo.

Sfruttando il fatto che R_x è Toeplitziana e simmetrica, si può limitare la stima alla prima riga della matrice. Un possibile stimatore consistente privo di polarizzazione è il seguente:

$$\hat{R}_x(k) = \frac{1}{N-k} \sum_{j=0}^{N-k-1} x_j^* x_{j+k}$$

Tuttavia, questo stimatore potrebbe produrre matrici che non sono definite positive, pertanto è opportuno modificarlo in modo tale che non si verifichi tale problema. La seguente variante introduce però una certa polarizzazione:

$$\hat{R}_x(k) = \frac{1}{N} \sum_{j=0}^{N-k-1} x_j^* x_{j+k}$$

Riassumendo, si ha dunque che la PSD stimata tramite stimatore a minima varianza può essere espressa come:

$$\hat{S}_x^W(f) = \frac{1}{S^\dagger \hat{R}_x^{-1} S}$$

2.6 Analisi tempo-frequenza

Gli strumenti descritti fino a questo punto mappano il dominio temporale nel dominio delle frequenze. Tale trasformazione ha senso se il processo analizzato

soddisfa l'ipotesi di stazionarietà, poiché la stima di spettro codifica l'informazione di *quando* si presentano certe armoniche nella fase, la quale è ignorata nella rappresentazione della PSD [8]. Esistono però dei metodi alternativi che consentono di analizzare anche i processi non-stazionari e non-lineari, preservando il tempo come parametro. In generale, la trasformazione in questione è ridondante perché mappa il dominio del tempo nel dominio tempo-frequenza:

$$t \rightarrow (\Omega, \tau)$$

In pratica, ciò avviene discretizzando il piano tempo-frequenza in cosiddette “mattonelle” le cui forme e dimensioni definiscono la localizzazione in tempo e frequenza. Il formato delle mattonelle viene determinato, come verrà discusso in seguito, in base al tipo di strumento utilizzato. Esiste però un concetto fondamentale che vale universalmente, noto come principio di indeterminazione di Heisenberg, che asserisce l'impossibilità di ottenere contemporaneamente illimitata localizzazione temporale e frequenziale. In termini matematici, il prodotto tra la durata temporale σ_t e la banda σ_Ω (in questo caso determinate dalle dimensioni della mattonella) è limitato inferiormente:

$$\sigma_t \sigma_\Omega \geq \frac{1}{2}$$

La condizione per cui si ottiene l'uguaglianza si verifica per la finestra gaussiana.

2.7 STFT

La STFT o “Short Time Fourier Transform” affronta il problema della rappresentazione tempo-frequenza applicando la trasformata di Fourier su una finestra scorrevole nel tempo. Assumendo che tale finestra abbia una lunghezza tale da suddividere il segnale in sotto-sequenze stazionarie, si otterrà uno spettro per ogni

intervallo temporale. Nel caso tempo-continuo, data una funzione finestra $w(t)$ reale e simmetrica per cui $\|w(t)\| = 1$, si definisce STFT:

$$\text{STFT}_s(\Omega, \tau) = \int_{-\infty}^{+\infty} s(t)w^*(t - \tau)e^{-i\Omega t} dt = \langle s(t), w^*(t - \tau)e^{-i\Omega t} \rangle$$

Per analizzare l'andamento dell'energia in tempo e frequenza si utilizza lo spettrogramma, cioè:

$$S(\Omega, \tau) = |\text{STFT}_s(\Omega, \tau)|^2$$

Nel caso della STFT la discretizzazione del piano tempo-frequenza avviene secondo una griglia $(k\Omega_0, nT_0)$ $k, n \in \mathbb{Z}$ per cui le mattonelle hanno forma rettangolare e area costante su tutto il piano. Sostituendo nell'espressione della STFT si ottiene:

$$\text{STFT}_s(k\Omega_0, nT_0) = \int_{-\infty}^{+\infty} s(t)w^*(t - nT_0)e^{-ik\Omega_0 t} dt = \text{STFT}_s[k, n]$$

L'area della mattonella definisce la risoluzione della STFT, anche in questo caso essa è legata alle proprietà della funzione finestra $w(t)$. Assumendo che $w(t)$ sia passabasso con frequenza di taglio Ω_b si deve innanzitutto garantire l'assenza di aliasing scegliendo $T_0 < \frac{\pi}{\Omega_b}$ e $\Omega_0 < 2\Omega_b$. In merito alla risoluzione valgono poi le considerazioni fatte nel paragrafo 2.4, seppur in un contesto diverso. Riassumendo, per avere una buona risoluzione tempo-frequenza si desidera avere una funzione finestra il cui lobo principale sia stretto e il decadimento dei lobi laterali sia veloce.

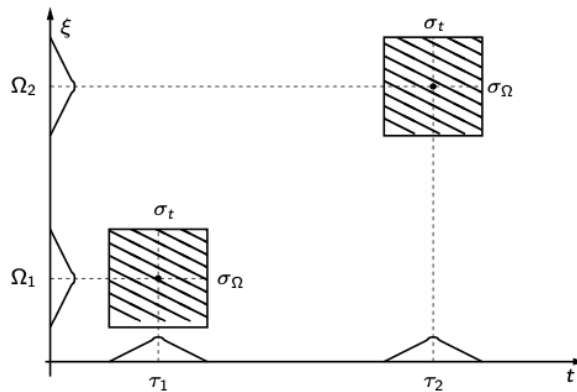


Figura 14 - Discretizzazione del piano tempo-frequenza della STFT.

Per ricondursi al caso pratico implementabile al calcolatore si introduce la versione discreta della STFT. A partire dalla definizione di STFT, si considerano un numero discreto e finito di N frequenze $\omega_k = \frac{2\pi}{N}k$ con $k = 0, 1, \dots, N - 1$. Così facendo, i filtri determinati dalla finestra risultano essere passa-banda centrati nell'intorno della frequenza selezionata.

$$\text{STFT}[k, n] = \text{STFT}_s(e^{i\omega}, n) \Big|_{\omega = \frac{2\pi}{N}k} = \sum_m w[m]s[n + m]e^{-i\frac{2\pi}{N}km}$$

Il calcolo racchiuso nell'ultimo passaggio può essere svolto in maniera efficiente tramite la DTF o "Discrete Fourier Transform" e ammette strategie ricorsive particolarmente efficaci dal punto di vista della complessità computazionale.

2.8 CWT

Il limite principale della STFT è che la sua risoluzione è fissata ed è costante su tutto il piano tempo-frequenza. Sono possibili, infatti, delle situazioni in cui il segnale in esame è composto da componenti ad alta frequenza di breve durata e componenti a bassa frequenza di lunga durata. Un'alternativa per rimuovere questo limite è quella

di fare in modo che la risoluzione dipenda dalle coordinate del piano e quindi che i rapporti $\frac{\Delta t}{t}$ e $\frac{\Delta f}{f}$ siano costanti. Così facendo, è possibile ottenere una risoluzione nel tempo arbitrariamente elevata ad alte frequenze e una risoluzione arbitrariamente elevata alle basse frequenze, pur rispettando il principio di Heisenberg. Questo concetto è la base su cui si sviluppa la teoria della trasformata Wavelet.

Per ottenere quanto appena descritto, la CWT o "Continuous Wavelet Transform" utilizza versioni scalate della stessa funzione $\psi(t) \in L^2(\mathbb{R})$ nota come base wavelet:

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|a|}} \psi\left(\frac{t-b}{a}\right)$$

Per poter essere considerata una base wavelet, la $\psi(t)$ deve soddisfare la condizione di ammissibilità:

$$C_\psi = \int_{-\infty}^{+\infty} \frac{|\Psi(\omega)|^2}{|\omega|} d\omega < \infty$$

Con $\Psi(\omega) = \mathcal{F}\{\psi(t)\}$ e $\lim_{\omega \rightarrow \pm\infty} \Psi(\omega) \rightarrow 0$ che si dimostra essere equivalente a $\Psi(\omega) = 0$. La $\Psi(\omega)$ ha dunque un comportamento passa-banda e un andamento oscillatorio nel tempo da cui deriva il nome "wavelet".

La formula di analisi della CWT è pur sempre una proiezione su di un insieme di base che però, a differenza della STFT, non è composto da esponenziali complessi bensì da funzioni wavelet opportunamente scalate e traslate:

$$\text{CWT}_s(a, b) = \int_{-\infty}^{+\infty} s(t) \psi_{a,b}^*(t) dt = \langle s(t), \psi_{a,b}^*(t) \rangle$$

Nel caso della CWT si parla di piano tempo-scala anziché tempo-frequenza. Esiste tuttavia un legame tra scala e frequenza:

$$a(\omega) = \frac{\omega_\psi}{\omega}$$

$$\omega_\psi = \operatorname{argmax} |\Psi(\omega)|$$

Dove con ω_ψ si è indicata la posizione in frequenza del picco della trasformata di Fourier della funzione wavelet.

Nel caso di volesse osservare l'andamento dell'energia sul piano tempo-scala sarebbe necessario fare uso del cosiddetto scalogramma:

$$C_s(a, b) = |CWT_s(a, b)|^2$$

Come spiegato per la STFT, si necessita di discretizzare il piano tempo-scala in modo da rendere applicabile la trasformata nel caso pratico. In questo caso si sceglie:

$$a = a_0^{-k}$$

$$b = nb_0 a_0^{-k}$$

Con $a_0 > 1$ e $b_0 > 0$ e $k, n \in \mathbb{Z}$. Sostituendo nell'espressione della wavelet traslata e scalata si ha:

$$\psi_{a,b}(t) \Rightarrow \psi_{kn}(t) = \frac{1}{\sqrt{|a_0|^{-k}}} \psi\left(\frac{t - nb_0 a_0^{-k}}{a_0^{-k}}\right) =$$

$$= |a_0|^{k/2} \psi(a_0^k t - nb_0)$$

Ponendo $a_0 = 2$ e $b_0 = 1$, nota in letteratura come “scelta diadica”, si ottiene una base ortonormale:

$$\psi_{kn}(t) = 2^{k/2} \psi(2^k t - n)$$

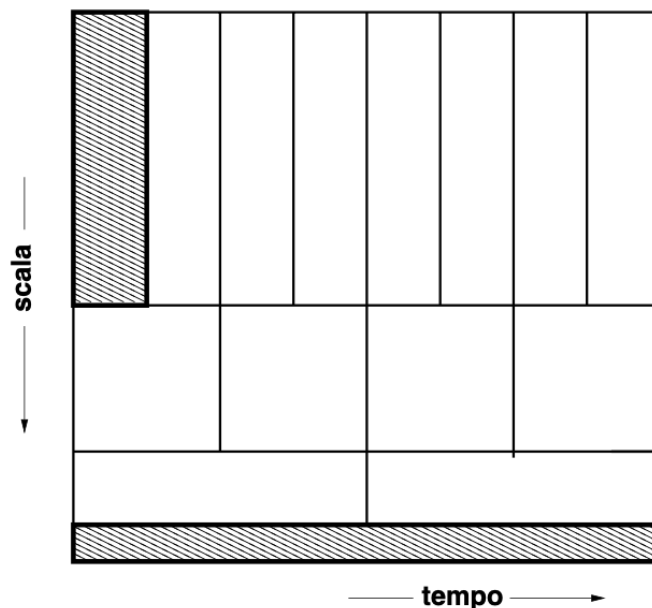


Figura 15 - Discretizzazione del piano tempo-scala della CWT.

Le uniche condizioni che deve soddisfare la funzione wavelet sono l'appartenenza a $L^2(\mathbb{R})$ e la condizione di ammissibilità. Esistono quindi infinite possibili funzioni di base che possono essere adoperate per l'analisi wavelet. Grazie al contributo di autorevoli autori come Ingrid Daubechies, Jean Morlet, Yves Meyer e altri, sono state studiate delle famiglie di wavelet che denotano determinate caratteristiche. Innanzitutto, vengono distinte tra wavelet continue e discrete ma nel caso di questo elaborato, trattandosi di CWT, ci si dedicherà solamente a quelle continue. Di seguito vengono presentate due famiglie che torneranno utili più avanti.

Wavelet di Morlet

La scelta di presentare questa famiglia di funzioni wavelet deriva dalla sua analogia con la finestra gaussiana che, come anticipato, consente di avere la massima risoluzione tempo-frequenza:

$$\psi(t) = \frac{B_0}{\sqrt{2\pi}} e^{i\omega_0 t} e^{-\frac{t^2}{2}}$$
$$\Psi(\omega) = B_0 e^{-\frac{(\omega - \omega_0)^2}{2}}$$

Tuttavia, per raggiungere il corretto parallelismo con la gaussiana è necessario applicare il logaritmo all'argomento. Si ottiene quindi:

$$\Psi(\omega) \sim e^{-\frac{(\omega_0 \log \omega)^2}{2}}$$

Naturalmente $\omega > 0$.

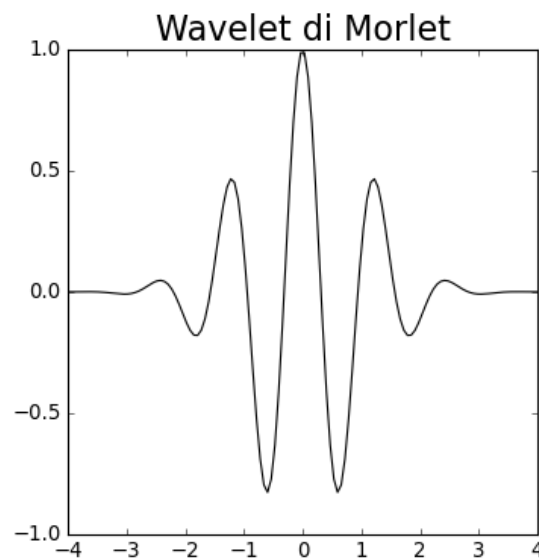


Figura 16 - Esempio di Wavelet di Morlet.

Wavelet di Shannon

Una possibile alternativa è quella della famiglia di Shannon. Questo insieme di wavelet è definito dalla seguente espressione:

$$\psi(t) = \text{sinc}\left(\frac{t}{2}\right) \cos\left(\frac{3\pi t}{2}\right)$$

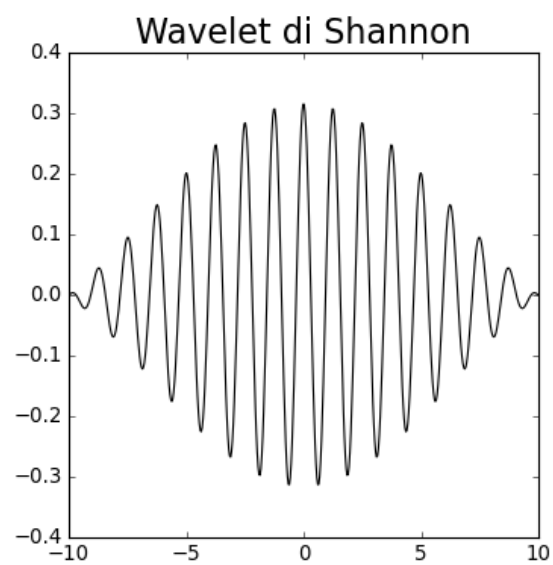


Figura 17 - Esempio di Wavelet di Shannon.

Capitolo 3

Rilevazione di Anomalie

Il compito di monitorare un sistema complesso consiste nel rilevare la presenza di eventi anomali, distinguendo quelle generate da veri e propri malfunzionamenti da quelli che invece sono attribuibili al rumore. L'approccio utilizzato in questo elaborato fa riferimento a tecniche non supervisionate, caratterizzate dall'attribuzione di un punteggio (detto "score") a ciascun punto del dataset che determina quanto il punto in questione sia "lontano" dal comportamento nominale del processo.

3.1 PCA

La PCA o "Analisi delle Componenti Principali" è una tecnica che permette di ridurre la dimensionalità di un set di dati mantenendone le cosiddette componenti principali, cioè gli elementi che ne esprimono il maggior contenuto informativo. L' algoritmo determina la trasformazione lineare ottima che permette di proiettare lo spazio dei dati $X_{n \times m}$ in un nuovo sistema cartesiano tale per cui la variabile con maggiore varianza viene proiettata sul primo asse, la seconda sul secondo asse e così via.

$$Y_{n \times m} = A_{n \times m} X_{n \times m}$$

A tal proposito è necessario che venga adattata una distribuzione gaussiana allo spazio dei dati.

Nel caso della PCA, la ricerca del sottospazio principale $A_{n \times m}$ avviene risolvendo il seguente problema di minimizzazione vincolata:

$$\begin{cases} \min \|X - A\| \\ \text{rango}(A) \leq k \end{cases}$$

Per cui si evidenzia l'utilizzo della norma L_2 .

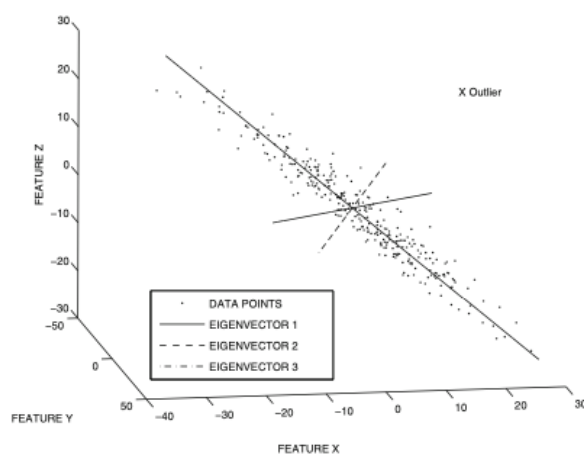


Figura 18 - Rappresentazione nello spazio tridimensionale delle componenti principali di una distribuzione.

Per poter applicare con efficacia questo metodo, X deve:

- Rappresentare le realizzazioni di un processo stazionario
- Avere variabili gaussiane
- Essere a media nulla
- Essere a varianza unitaria

Questa tecnica prevede la stima della matrice di covarianza associata a X tramite il seguente stimatore:

$$\Phi = \frac{1}{n-1} X^T X$$

e successivamente di estrarne gli autovalori e gli autovettori:

$$\Phi = \frac{1}{n-1} X^T X = P \Lambda P^T$$

in cui Λ è una matrice diagonale i cui elementi sono gli autovalori di Φ disposti in ordine decrescente. Le colonne della matrice P contengono invece gli autovettori di Φ . Riassumendo, la matrice X è scomponibile nel prodotto di due matrici, la prima detta "score" contenente le componenti principali e la seconda detta "loading":

$$X = T P^T$$

Un'implementazione molto efficace dal punto di vista computazionale consistente nell'applicare la "Decomposizione ai Valori Singolari" o SVD a X :

$$X = U \Sigma V^T$$

Dove le colonne di U sono gli autovettori "sinistri" di X , le colonne di V gli autovettori "destri" e gli elementi di Σ (matrice diagonale) sono gli autovalori di X .

L'aspetto chiave della PCA consiste nel considerare solamente $l < m$ componenti principali pur mantenendo la desiderata percentuale del contenuto informativo del segnale. Così facendo si delineano due nuovi sottospazi ortogonali tra loro, detti PCS "Spazio delle componenti principali" e RS "Spazio residuale":

$$\begin{aligned} \text{PCS} &= \text{Span}\{\hat{P}_l\} \\ \text{RS} &= \text{Span}\{\tilde{P}_{m-l}\} \end{aligned}$$

3.1.1 Scelta del numero di componenti

Tra i possibili metodi esistenti in letteratura per la determinazione del numero di componenti, ne sono stati studiati alcuni empirici ed altri frutto di un processo di ottimizzazione. Procedendo in maniera empirica potremmo attuare il cosiddetto criterio della "Percentuale cumulativa di varianza" o CPV, richiedendo all' algoritmo di ricavare quella trasformazione lineare ottima che permette di esprimere una data percentuale della varianza del dataset. In questo caso, il numero di componenti viene dedotto direttamente dalla percentuale di varianza richiesta secondo la seguente formula:

$$CPV(l) = 100 \frac{\sum_{j=1}^l \lambda_j}{\sum_{j=1}^m \lambda_j} \%$$

Un altro possibile metodo potrebbe essere quello di considerare il gomito formato dal profilo della varianza espressa da ciascuna componente (in scala logaritmica). Così facendo si mantiene gran parte della varianza scartando le componenti attribuibili in gran parte al rumore.

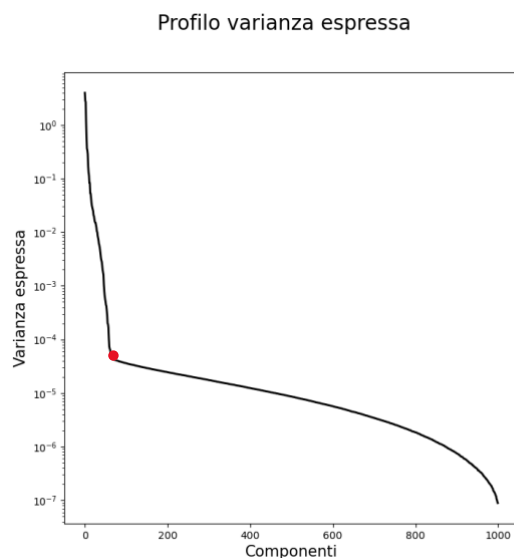


Figura 19 - Andamento della varianza espressa da ciascuna componente. Il "gomito" che divide i due andamenti è stato evidenziato con un punto rosso.

3.2 KPCA

Esistono tuttavia dei casi in cui lo spazio campionario non è linearmente separabile, non esiste cioè un possibile iperpiano su cui proiettare i punti del dataset in grado di preservarne il contenuto informativo. Si consideri ad esempio il seguente caso:

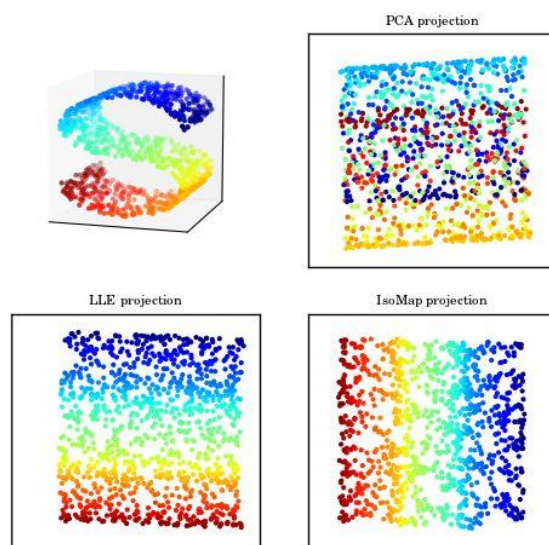


Figura 20 - Esempio di varietà tri-dimensionale e proiezione su piani.

La superficie bi-dimensionale in alto a sinistra rappresentata nello spazio tri-dimensionale è non lineare e pertanto la sua proiezione su di un qualsiasi piano (es. PCA) comporta il mescolamento dei punti, impedendo di distinguerne la provenienza e perdendo conseguentemente gran parte del contenuto informativo.

L'intuizione su cui si basa la Kernel PCA o KPCA è quindi quella di mappare il dataset di partenza in uno spazio di dimensionalità superiore tramite una trasformazione non-lineare chiamata "Kernel". In questo modo il nuovo spazio può essere processato linearmente come avviene per la PCA.

Indicando con Ψ il kernel, la KPCA opera cercando la soluzione del seguente problema di ottimizzazione:

$$\min \sum_i^t |\Psi(x_i) - UU^T\Psi(x_i)|$$

dove con U è indicata la matrice sinistra della scomposizione di Ψ :

$$\Psi(X) = U\Sigma V^T$$

Per stimare la matrice di covarianza si opera il cosiddetto "Trucco del Kernel" che permette di ricavare la matrice del Kernel semplicemente tramite i prodotti scalari tra i vettori di X mappati nel nuovo spazio:

$$K = \Psi(X)\Psi(X)^T$$

Nel caso della KPCA non è necessario avere X a media nulla ma questa ipotesi deve essere trasferita a valle della trasformazione non lineare, quindi Ψ deve essere a media nulla. Quanto appena descritto è riassumibile nei seguenti passaggi da seguire nell'implementazione della KPCA:

1. Scelta del Kernel $k(x_i, x_j)$
2. Calcolo di matrice K
3. Rendere K a media nulla
4. Estrarre gli autovalori e autovettori di K
5. Proiettare i dati nel nuovo spazio

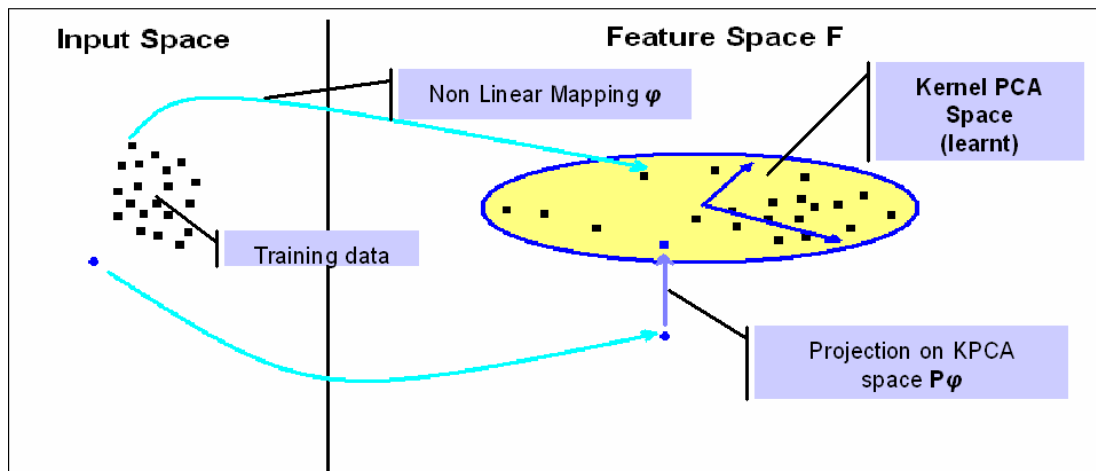


Figura 21 - Schematizzazione grafica dell' algoritmo della KPCA.

3.2.1 Scelta del kernel

La scelta del kernel da applicare con la KPCA influenza fortemente la bontà dei risultati ottenuti. Molti dei kernel sono inoltre contraddistinti da una serie di parametri, al variare dei quali vengono modificate le proprietà intrinseche della funzione. Si ha quindi un numero rilevante di gradi di libertà da fissare, pratica per la quale non esistono particolari algoritmi. Una possibilità (applicabile anche in altri contesti) consiste nel variare ciascun parametro tenendo fissi i rimanenti in modo da ricavare conseguentemente i parametri migliori al fine della riduzione di dimensionalità. I kernel più utilizzati in letteratura sono ad esempio:

RBF (Radial Basis Function)

È definito come:

$$k(x, y) = \exp(-\gamma \|x - y\|^2)$$

Ed è noto come kernel gaussiano nel caso in cui $\gamma = \sigma^{-2}$.

Cosine similarity

Rappresenta il prodotto scalare tra i due vettori normalizzato secondo la norma L_2 :

$$k(x, y) = \frac{xy^T}{\|x\| \|y\|}$$

La norma L_2 proietta i due vettori sulla sfera unitaria mentre il loro prodotto scalare esprime il coseno dell'angolo compreso tra essi.

3.3 RPCA

Nel caso della PCA, l'ottimizzazione del sottospazio che meglio esprime le direzioni di massima varianza del dataset avviene sfruttando la norma L_2 . Nel paragrafo 3.1 è stato spiegato che la PCA tenta di adattare una distribuzione gaussiana alla distribuzione dello spazio dei dati. Nel senso di L_2 , tale approccio è tuttavia molto sensibile alla presenza di campioni che si discostano notevolmente dalla distribuzione. Per meglio comprendere questo concetto si osservi il seguente esempio, in cui viene rappresentata la regressione di un insieme di punti:

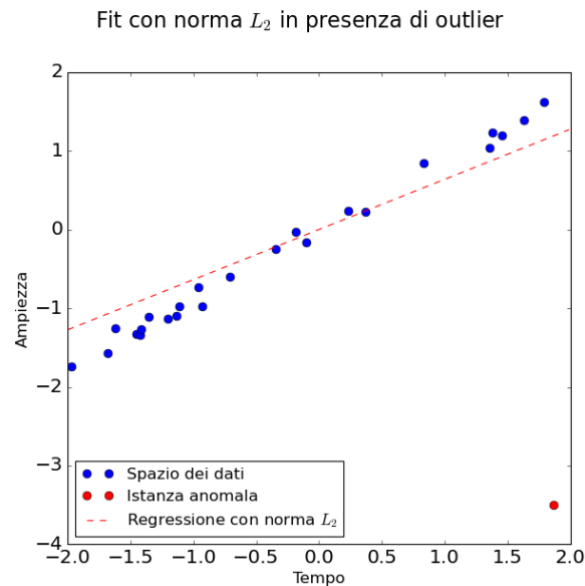


Figura 22 - Esempio di regressione effettuata tramite norma L_2 in presenza di outlier.

Si nota infatti che la retta di regressione tende ad essere “sbilanciata” dalla presenza del campione anomalo.

Di recente, Candès, Emmanuel J., e altri hanno presentato una variante della PCA che basa il suo funzionamento sulla applicazione della norma L_1 [12]. Questa nuova tecnica è nota come “Robust Principal Component Analysis” proprio perché, tramite questo accorgimento, è possibile limitare l’influenza di istanze anomale sulla determinazione della distribuzione. Tornando all’esempio precedente si evince come questa idea sia valida dal punto di vista appena descritto:

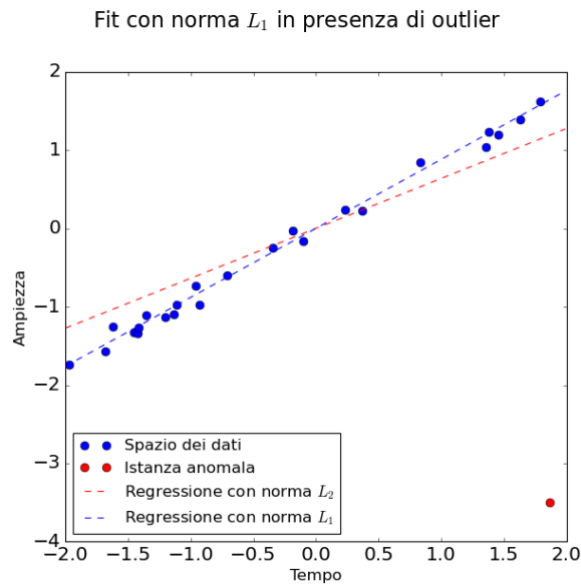


Figura 23 - Esempio di regressione tramite norma L_1 in presenza di outlier.

Si ipotizzi di avere una certa matrice X contenente i campioni del processo e di volerla scomporre nella somma di due componenti, la prima nota come componente a “basso rango” L (dall’inglese “low-rank”) e la seconda nota con il nome di componente “sparsa” S :

$$X = L + S$$

L contiene l’informazione che caratterizza il segnale nelle sue componenti principali mentre S contiene rumore e istanze anomale. Si noti che il problema in questione è mal posto poiché esistono infinite possibili soluzioni alla decomposizione di X nella somma di due matrici [13]. Fortunatamente, le moderne tecniche di ottimizzazione sono in grado di regolarizzare il problema consentendo di ottenere un’unica soluzione. Nel caso della RPCA il problema di ottimizzazione è il seguente:

$$\min_{L,S} \text{rank}(L) + \|S\|_0 \text{ s.t. } L + S = X$$

In sostanza si desidera minimizzare il rango di L (limitando il contenuto informativo alle componenti principali) con S matrice sparsa, in modo tale che la loro somma restituisca X . Purtroppo, questa formulazione non è risolvibile direttamente: l'operatore *rank* e la norma L_0 non sono convessi, per cui la sua risoluzione del problema di minimizzazione (a sua volta non convesso) non è garantita in un tempo limitato. Si attua dunque un'operazione nota con il nome di "rilassamento della convessità" tramite l'introduzione di operatori sostitutivi per il *rank* e per la norma L_0 , rispettivamente la norma nucleare L_* (cioè la somma degli autovalori) e la norma L_1 . Si ottiene il seguente problema di ottimizzazione convesso:

$$\min_{L,S} \|L\|_* + \lambda_0 \|S\|_1 \quad s. t. \quad L + S = X$$

Il metodo appena proposto consente quindi di estrarre le componenti principali di X con particolare robustezza nei confronti della presenza di istanze anomale o in generale di campioni che si discostano significativamente dalla distribuzione principale.

3.4 Score

Per distinguere un'istanza anomala all'interno del dataset è necessario introdurre un punteggio o "score" che rappresenta quanto "dista" ciascun punto del dataset di test dal comportamento nominale. Chiaramente, campioni in linea con quelli tipici denoteranno uno score basso mentre comportamenti anomali uno score più alto. Esistono numerose statistiche di variabilità in letteratura, ciascuna con le proprie specifiche caratteristiche.

3.4.1 Score SPE

Lo score SPE misura la proiezione dei vettori di X sullo spazio residuale P_{res} (indicato precedentemente con la dicitura "RS"):

$$SPE = x^T P_{res} P_{res}^T x$$

A seconda della tipologia di anomalia, questa potrebbe presentare caratteristiche tali da non alterare le misure nelle loro componenti principali ma bensì in quelle residuali. Di seguito una breve derivazione permette di ottenere un'espressione semplice da calcolare:

$$\begin{aligned} SPE &= x^T P_{res} P_{res}^T x = \\ &= \|P_{res}^T x\|^2 = \\ &= \|x\|^2 - \|P_{pc}^T x\|^2 = \\ &= \|x - P_{pc}^T x\|^2 = \\ &= \|x - \hat{x}\|^2 = \\ &= \|\tilde{x}\|^2 \end{aligned}$$

Questo però non può essere applicato direttamente nel caso della KPCA poiché lo spazio dei dati viene mappato in un nuovo spazio tramite una trasformazione non lineare. È necessario quindi adattare la definizione a questa casistica:

$$\begin{aligned} SPE &= \Psi(x)^T P_{res} P_{res}^T \Psi(x) = \\ &= \Psi(x)^T (I - P_{pc} P_{pc}^T) \Psi(x) = \\ &= \Psi(x)^T \Psi(x) - \Psi(x)^T P_{pc} P_{pc}^T \Psi(x) = \\ &= k(x, x) - k^T(x) C k(x) \end{aligned}$$

Dove $C = P_{pc} P_{pc}^T$.

3.4.2 Score T^2

In maniera del tutto ortogonale allo score SPE, il T^2 quantifica la proiezione di X nello spazio principale PCS . Per definizione:

$$T^2 = x^T P_{pc} \Lambda_{pc}^{-1} P_{pc}^T x$$

Anche in questo caso è possibile derivare dalla definizione una semplice espressione:

$$\begin{aligned} T^2 &= x^T P_{pc} \Lambda_{pc}^{-1} P_{pc}^T x = \\ &= \left\| \Lambda_{pc}^{-\frac{1}{2}} P_{pc}^T x \right\|^2 = \\ &= \sum_{j=0}^{n-1} \left(\frac{1}{\sqrt{\Lambda_j}} e_j^T x \right)^2 \end{aligned}$$

Nel caso particolare della KPCA lo score T^2 assume una forma che deve tenere conto dell'applicazione del kernel:

$$T^2 = k(x)^T P_{pc} \Lambda_{pc}^{-1} P_{pc}^T k(x)$$

3.4.3 Distanza di Mahalanobis

L'indicatore in questione appartiene all'insieme delle tecniche che si basano sulla distribuzione probabilistica dei vettori dello spazio dei dati [9]. Il dataset viene assunto come normalmente distribuito attorno alla propria media $\bar{\mu}$ secondo una distribuzione gaussiana multivariata. Nota la matrice di covarianza Σ associata al set di dati si può dunque esprimere la distribuzione con la seguente espressione:

$$f(\bar{X}) = \frac{1}{\sqrt{|\Sigma|}(2\pi)^{d/2}} e^{\left[-\frac{1}{2}(\bar{X}-\bar{\mu})\Sigma^{-1}(\bar{X}-\bar{\mu})^T\right]}$$

Il termine all'esponente rappresenta metà del quadrato della cosiddetta distanza di Mahalanobis tra i vettori dei dati X e il centroide del dataset $\bar{\mu}$:

$$\text{Mahalanobis}(\bar{X}, \bar{\mu}, \Sigma) = \sqrt{(\bar{X} - \bar{\mu})\Sigma^{-1}(\bar{X} - \bar{\mu})^T}$$

La distanza di Mahalanobis differisce da quella euclidea per il fatto che in quella di Mahalanobis la distanza tra X e $\bar{\mu}$ viene pesata dalla correlazione tra i vari attributi.

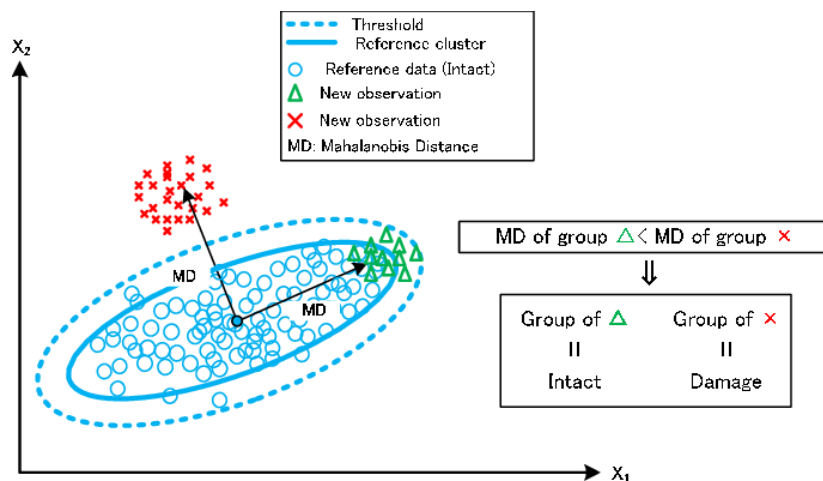


Figura 24 - Schematizzazione grafica della distanza di Mahalanobis.

Il concetto chiave è infatti che questo approccio riconosce che le diverse direzioni di correlazione hanno differenti varianze e di conseguenza, un punto molto distante dal centro dello spazio nella direzione a maggior varianza è meno probabilmente un punto anomalo rispetto ad un altro che è meno distante dal centro ma in una direzione in cui la varianza è minima.

3.5 ROC

Nel caso della rilevazione di anomalie esistono quattro possibili scenari e la probabilità che si verifichino determina la bontà del detector. I quattro casi sono:

- Vero positivo: il detector rileva correttamente la presenza di un'anomalia
- Vero negativo: il detector rileva correttamente l'assenza di un'anomalia
- Falso positivo: il detector rileva erroneamente la presenza di un'anomalia
- Falso negativo: il detector non rileva la presenza di un'anomalia

Di seguito una tabella riassuntiva delle quattro casistiche:

		Valore vero		Totale
		P	N	
Predizione risultato	P'	Vero positivo	Falso positivo	P'
	N'	Falso negativo	Vero negativo	N'
Totale		P	N	

Figura 25 - Tabella delle possibili casistiche nell'ambito della rilevazione di anomalie.

Rappresentando la distribuzione degli score in caso positivo e negativo si ottengono due campane, nell'immagine rispettivamente azzurra e rossa. Il numero dei falsi positivi e falsi negativi è determinato dalla sovrapposizione delle due campane. È facilmente intuibile che il caso ideale sarebbe l'assenza di sovrapposizione tra le due distribuzioni:

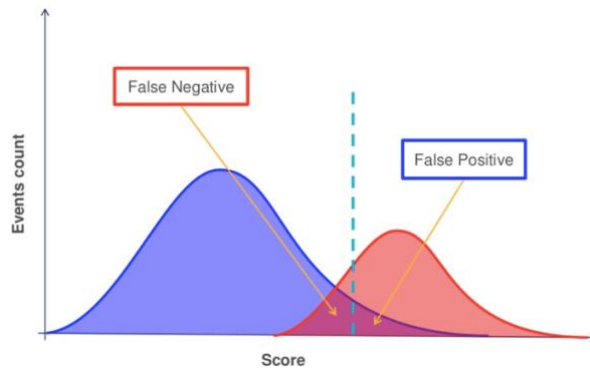


Figura 26 - Rappresentazione delle distribuzioni dei risultati positivi e negativi.

La linea verticale tratteggiata rappresenta la soglia che si pone per distinguere tra positivi e negativi e determina inevitabilmente anche il rapporto tra falsi positivi e falsi negativi.

Si definiscono quindi le seguenti quantità:

- La frazione di veri positivi rispetto al totale dei positivi, $TPR = \frac{TP}{P} = \frac{TP}{TP+FN}$
- La frazione di veri negativi rispetto al totale dei negativi, $FPR = \frac{FP}{N} = \frac{FP}{FP+TN}$
- L'accuratezza del detector, $ACC = \frac{TP+TN}{P+N}$

Con TPR si indica il "True positive rate" o anche detto "Sensibilità" mentre con il FPR si indica il "False positive rate" o "1-Specificità". Al fine di valutare la bontà dei vari detector vengono graficati questi due indicatori al variare della soglia di distinzione tra positivi e negativi, ottenendo così la cosiddetta curva del ROC:

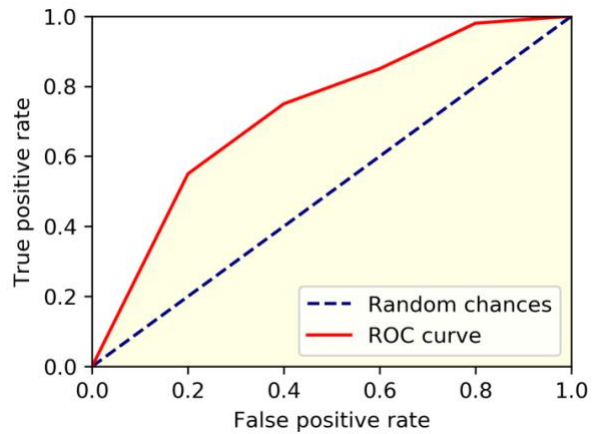


Figura 27 - Esempio di una possibile curva del ROC e della sua area sottesa (AUC).

La diagonale principale indica la performance del classificatore randomico. Più la curva associata al detector in esame si avvicina al punto (0,1) migliore sarà la performance di questo detector. Calcolandone l'area sottesa o AUC ("Area Under Curve") si ottiene la probabilità che il risultato del classificatore applicato ad un'occorrenza anomala sia superiore a quello del classificatore applicato ad un'occorrenza non anomala.

Capitolo 4

Generazione di dataset sintetici

Il set di dati fornito per la fase di pre-analisi risulta insufficiente al fine di applicare le tecniche di machine learning oggetto di questa tesi. È risultato pertanto conveniente realizzare un generatore di dati sintetici, in grado di rispecchiare il più possibile quelli reali, che consentisse di costruire dataset arbitrariamente grandi. L'idea alla base del generatore è quella di comporre il segnale con due contributi principali, il primo riferito al picco causato dal passaggio del pantografo e il secondo attribuito alle oscillazioni del sistema di funi e tiranti.

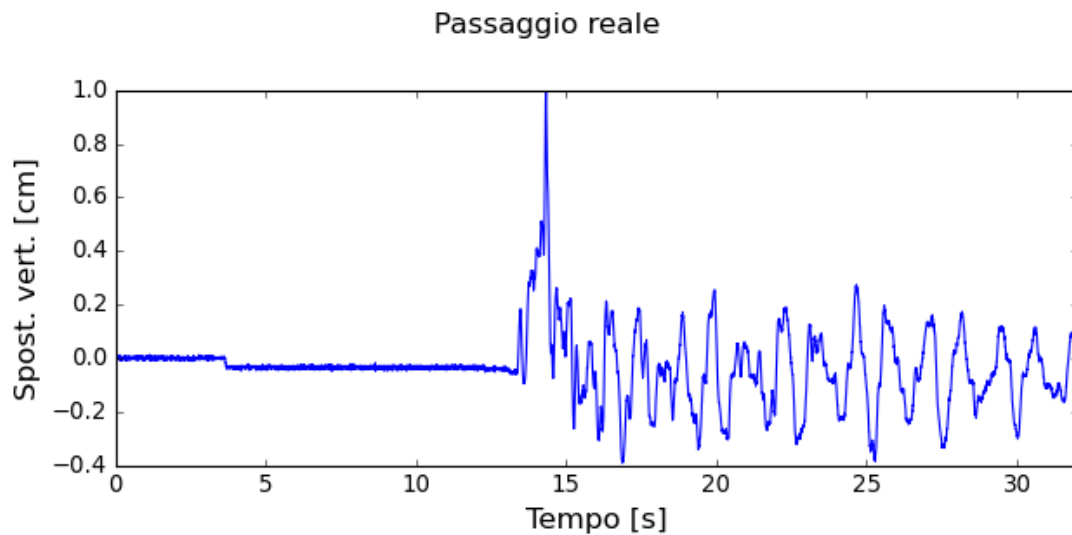


Figura 28 - Andamento nel tempo di un passaggio avvenuto nel 2017

L'intervallo temporale antecedente al picco non presenta particolare interesse per questa fase di studio preliminare, pertanto si è deciso di costruire sequenze che cominciano in corrispondenza del passaggio del pantografo e hanno una durata che può essere fissata in fase di generazione. Dal punto di vista pratico si è immaginato un sistema di misura che viene attivato dalla rilevazione del picco e acquisisce per un tempo prestabilito.

L'approccio adottato è quello di sfruttare le informazioni relative al contenuto spettrale dei dati reali per comporre dei segnali tramite sovrapposizione di sinusoidi. La variabilità della frequenza e della ampiezza dei vari picchi è stata modellata tramite distribuzione gaussiana.

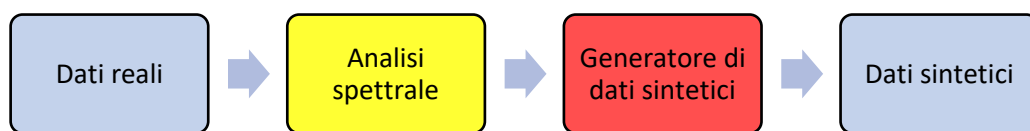


Figura 29 – Schema a blocchi del procedimento seguito per la generazione dei dataset sintetici.

4.1 Costruzione dei dataset

Il generatore in questione consiste nella classe "CatenaryDataGen" i cui parametri di input sono i vettori delle medie e delle deviazioni standard sia per quando riguarda le frequenze delle armoniche principali sia per le relative ampiezze. È necessario fornire anche le informazioni relative alla durata e all'ampiezza del picco (sempre tramite valore medio e deviazione standard). In fine è richiesto di specificare la deviazione standard del rumore gaussiano bianco (a media nulla). Di seguito una tabella riassuntiva degli input del generatore.

Parametro	Significato
d	Durata della sequenza [s]
fs	Frequenza di campionamento [Hz]
m	Numero di passaggi che costituiscono il dataset
a_mean	Vettore contenente i valori medi delle ampiezze delle sinusoidi che compongono l'oscillazione [cm]
a_std	Vettore contenente le deviazioni standard delle ampiezze delle sinusoidi che compongono l'oscillazione [cm]
f_mean	Vettore contenente i valori medi delle frequenze delle sinusoidi che compongono l'oscillazione [Hz]
f_std	Vettore contenente le deviazioni standard delle frequenze delle sinusoidi che compongono l'oscillazione [cm]
p_h_mean	Valore medio dell'altezza del picco [cm]
p_h_std	Deviazione standard dell'altezza del picco [cm]
p_d_mean	Valore medio della durata del picco [s]
p_d_std	Deviazione standard della durata del picco [s]
noise_std	Deviazione standard del rumore gaussiano bianco
date	Data di inizio del dataset

Figura 30 - Tabella riassuntiva dei parametri di input del generatore

L'output restituito dal metodo di costruzione della classe è una serie temporale il cui indice è in formato "Timestamp" in modo tale da essere facilmente integrabile con la classe che genera le anomalie che verrà descritta in seguito. La colonna riferita alle misure racchiude la concatenazione di m passaggi. Ogni dataset avrà dunque una colonna relativa alle misure e $m \times d \times f_s$ righe. Nel seguente esempio sono stati generati $m = 100$ passaggi della durata di $d = 60$ s e campionati a $f_s = 100$ Hz.

	val
2021-02-01 00:00:00.000	-0.008799
2021-02-01 00:00:00.010	0.040667
2021-02-01 00:00:00.020	0.037516
2021-02-01 00:00:00.030	0.089139
2021-02-01 00:00:00.040	-0.099517
...	...
2021-02-01 01:39:59.950	-0.172990
2021-02-01 01:39:59.960	-0.161928
2021-02-01 01:39:59.970	-0.246716
2021-02-01 01:39:59.980	-0.274773
2021-02-01 01:39:59.990	-0.359135

600000 rows × 1 columns

Figura 31 - Esempio di serie sintetica.

Dall'osservazione delle misure reali si è riscontrato che la velocità a cui passa il treno determina, come è naturale che sia, l'ampiezza del picco e delle oscillazioni. Ipotizzando dunque che sulla stessa linea possano passare diversi treni a varie velocità, si è deciso di preservare questo aspetto anche per la generazione sintetica, in modo tale da rendere il tutto più verosimile. Nello specifico, ciò che determina la velocità dell'occorrenza in questione è un'estrazione casuale.

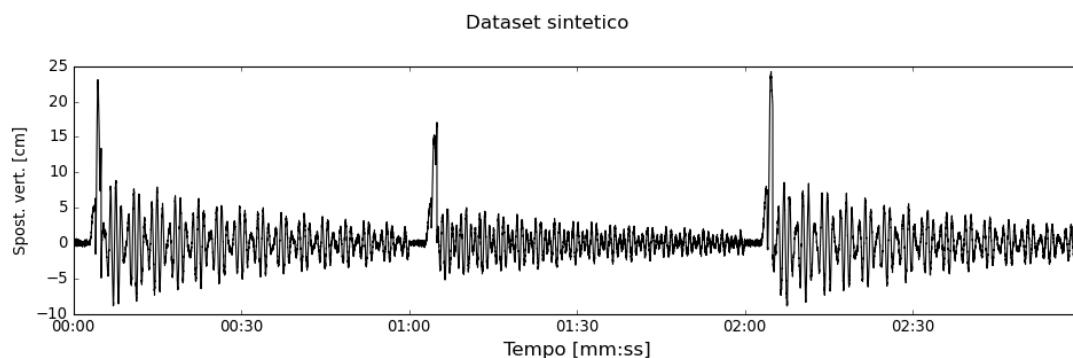


Figura 32 - Esempio di dataset sintetico composto da 3 passaggi, due alla velocità di 160 km/h e uno a 180 km/h.

4.2 Composizione del passaggio

Come anticipato, la generazione della singola occorrenza, determinata dal transito del pantografo che causa un'oscillazione del sistema meccanico, può essere scomposta in due eventi principali:

- Il picco, cioè il massimo innalzamento del cavo di contatto;
- L'oscillazione smorzata, seguente all'attraversamento del pantografo.

Una volta determinata l'ampiezza del picco p_h , viene creato un impulso rettangolare lungo p_d e di ampiezza p_h . Successivamente viene moltiplicato per una salita esponenziale che nel tempo p_d cresce da 0 a 1. In questo modo viene modellato l'innalzamento progressivo causato dal pantografo.

$$x_1(t) = \begin{cases} e^{\frac{t}{\tau_1}} - 1 & t < p_d \\ 0 & \text{altrove} \end{cases}$$

con $\tau_1 = \frac{p_d}{\log(2)}$.

Già dai primi istanti in cui la linea viene sollecitata, si originano delle oscillazioni (alle frequenze tipiche del sistema) che in questo caso vengono sommate al picco.

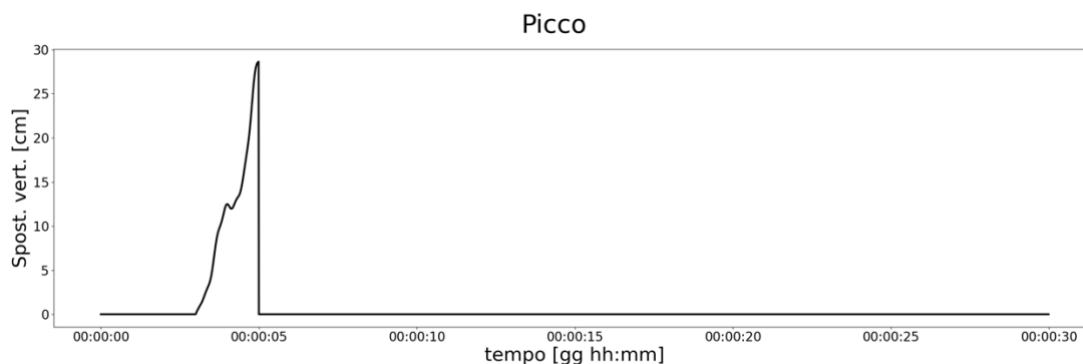


Figura 33 – Esempio di picco sintetico.

Successivamente viene aggiunto il post-passaggio formato dalla sovrapposizione di sinusoidi. Per determinare i vettori a_{mean} , a_{std} , f_{mean} , f_{std} sono stati stimati gli spettri dei passaggi a disposizione. Ogni passaggio è caratterizzato principalmente da un ristretto numero di componenti predicibili che hanno però una certa variabilità in frequenza e ampiezza. Dati n passaggi si avranno dunque n valori in ampiezza e frequenza relativi a ciascuna armonica.

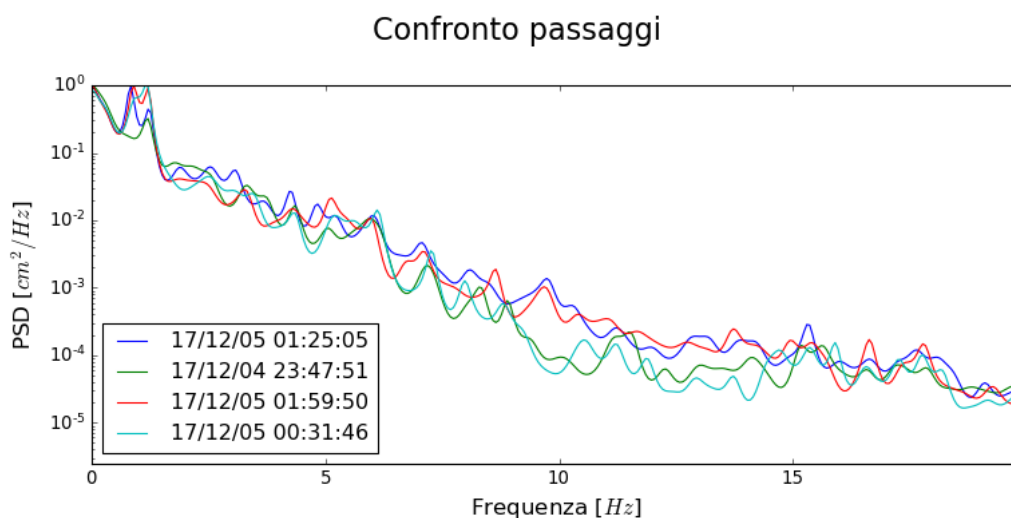


Figura 34 – Stima di spettro di potenza dei quattro passaggi del 2017

Come si può osservare, salvo i due picchi principali, la maggior parte delle armoniche ha ampiezza molto contenuta. Si noti anche che in alcuni casi la variabilità in ampiezza è piuttosto marcata, alcuni picchi sono quasi assenti in alcuni spettri e dominanti in altri.

A titolo esemplificativo, sono state estrapolate le informazioni relative a 9 componenti. Questo metodo consente di arricchire lo spettro anche in un secondo momento, ove fosse necessario.

Armoniche	1°	2°	3°	4°	5°	6°	7°	8°	9°
f_{mean}	0.0903	0.7995	0.8837	1.1657	1.2288	1.9299	2.4958	3.1093	3.3658
f_{std}	0.0024	0.0202	0.0138	0.028	0.0100	0.0105	0.1033	0.0523	0.0458
a_{mean}	1.1011	4.1820	2.6047	1.5423	0.5387	0.5387	0.5387	0.5154	0.6451
a_{std}	0.1577	1.0626	0.3695	0.4488	0.1367	0.1367	0.1367	0.0850	0.2413

Figura 35 - Esempio dei parametri estratti dai dati reali utili a modellare la variabilità in frequenza ed ampiezza delle varie armoniche.

Vengono quindi estratti in maniera casuale da una distribuzione gaussiana l'ampiezza e la frequenza di ciascuna armonica e con questi vengono generate le sinusoidi da sommare al segnale contenente il picco.

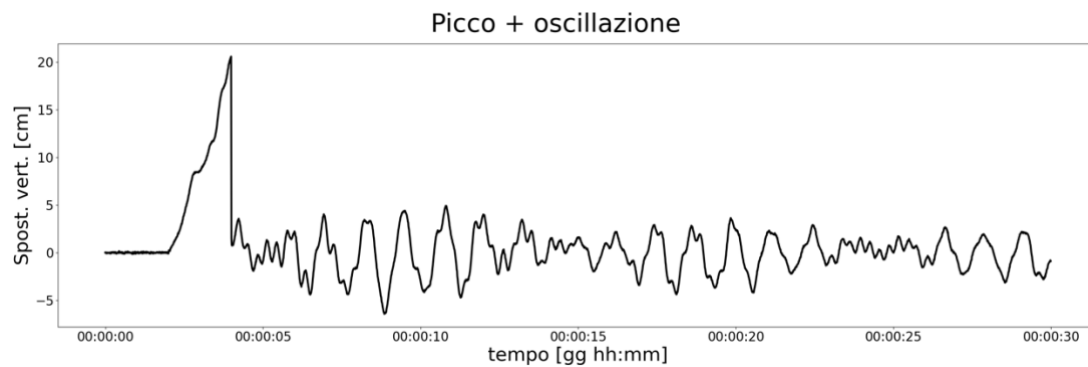


Figura 36 - Passaggio sintetico completo formato da picco e oscillazione.

Al fine di modellare lo smorzamento del sistema meccanico, l'oscillazione è stata moltiplicata per un decadimento esponenziale il cui coefficiente di smorzamento è stato stimato osservando l'involuppo dei passaggi reali.

$$x_2(t) = \begin{cases} \left[\sum_{i=1}^n a_i * \sin(2\pi f_i t) \right] * e^{-\frac{t}{\tau_2}} & \text{per } t > p_d \\ 0 & \text{altrove} \end{cases}$$

Infine, è stato aggiunto del rumore gaussiano bianco. Così facendo si introduce un'ulteriore fonte di casualità che da una parte rende i segnali sintetici più verosimili e dall'altra complica il lavoro dei detector, che dovranno essere di conseguenza più robusti. In questo caso il parametro che permette di dosarne l'energia è la deviazione standard σ . Si richiama per completezza la distribuzione gaussiana:

$$p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{x^2}{2\sigma^2}}$$

Sommando il contributo del picco, dell'oscillazione post-passaggio e del rumore bianco (indicato con $x_3(t)$), si ottiene il segnale sintetico:

$$y(t) = x_1(t) + x_2(t) + x_3(t)$$

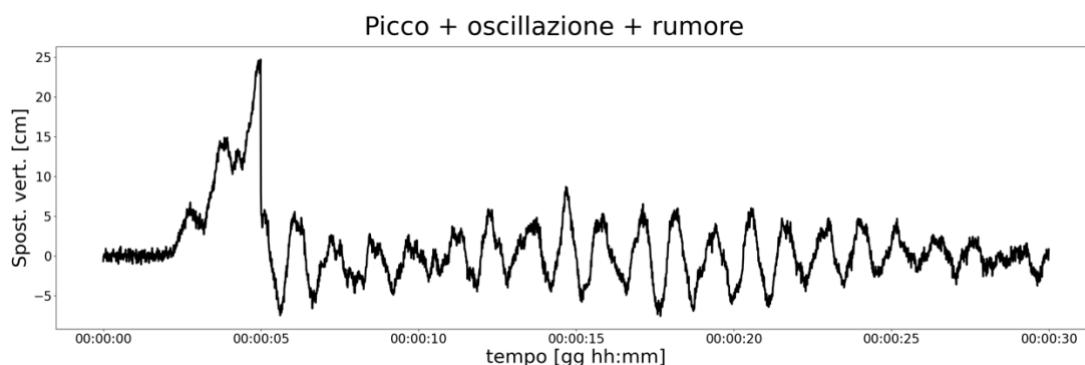


Figura 37 - Esempio di passaggio sintetico.

4.3 Generatore di anomalie

In questa fase preliminare in cui non si conoscono con precisione i possibili malfunzionamenti e soprattutto il loro effetto sui segnali acquisiti, risulta innanzitutto opportuno focalizzare l'attenzione sulla rilevazione di possibili anomalie attinente al

sistema di misura. Di seguito viene elencato un sottoinsieme degli effetti di anomalie sul sensore:

Impulso a gradino

Al segnale originale si somma un impulso rettangolare di ampiezza A_1 e durata T_1 :

$$a_1(t) = A_1 \text{rect}(tT_1)$$

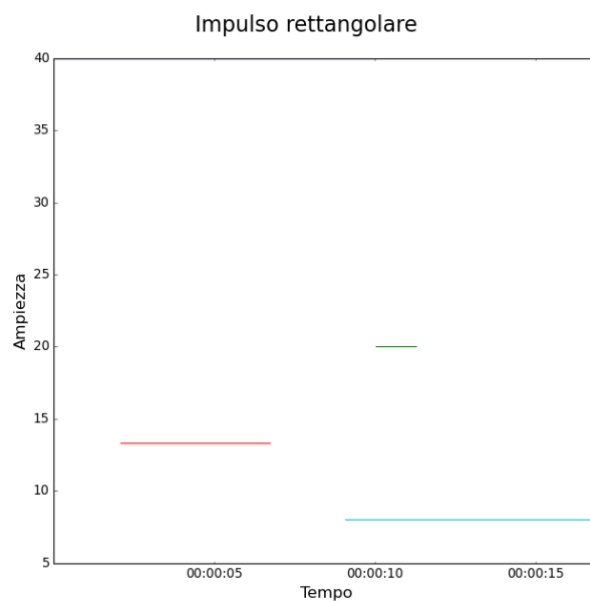


Figura 38 - Impulsi rettangolari di varia ampiezza e durata.

Impulso triangolare

Analogo al precedente ma ora si considera una funzione triangolare così definita:

$$a_2(t) = A_2 \begin{cases} t & \text{per } t < \frac{T_2}{2} \\ -t + \frac{T_2}{2} & \text{per } t \geq \frac{T_2}{2} \end{cases}$$

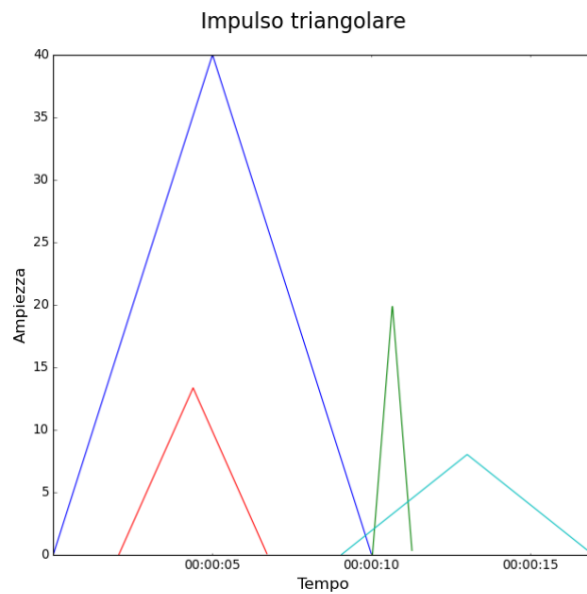


Figura 39 - Impulsi triangolari di varia ampiezza e durata.

Jammer

Per “jammer”, si intende l’interferenza di una sorgente esterna che disturba la misura aggiungendo un segnale con una certa banda. Il disturbo in questione tipicamente si posiziona in una banda “libera” dello spettro del segnale originale.

$$a_3(t) = A \sum_{j=0}^{N-1} \cos(2\pi f_j t)$$

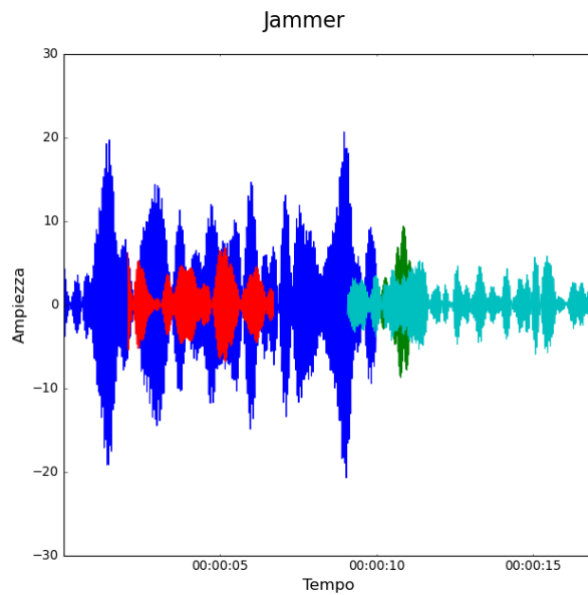


Figura 40 - Jammer di varia ampiezza, durata e banda.

Anomalia periodica

Data una determinata frequenza f , l'anomalia periodica consiste in una sinusoide che si aggiunge al segnale originale. La comparsa e la scomparsa di questa anomalia avviene più o meno "dolcemente" seguendo il profilo di una funzione coseno rialzato:

$$a_4(t) = A \cos(2\pi ft)$$

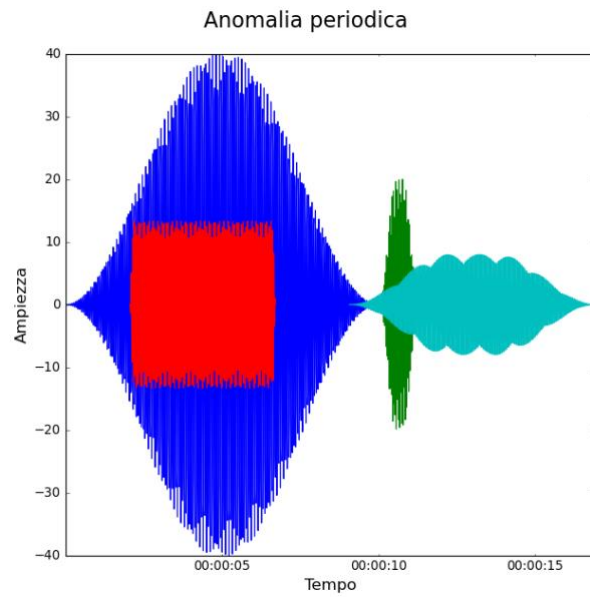


Figura 41 - Anomalie periodiche di varia ampiezza, durata e frequenza.

AWGN

Questa anomalia aggiunge ulteriore rumore bianco gaussiano con una certa deviazione standard per un certo intervallo temporale.

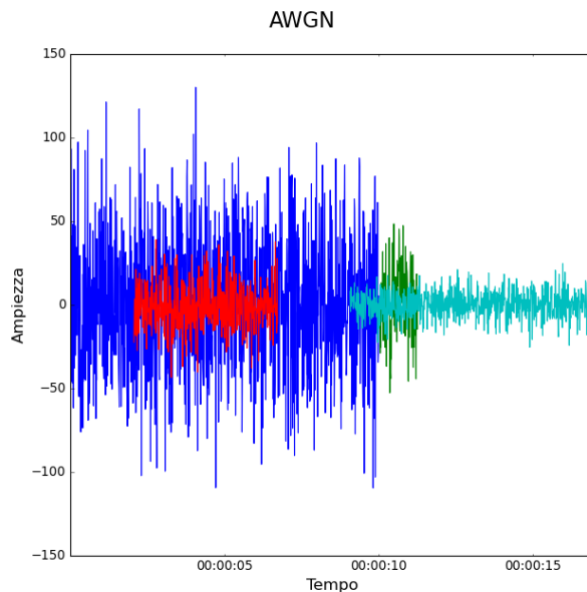


Figura 42 - AWGN di varia ampiezza e durata.

Le anomalie appena descritte sono state generate tramite la classe “Anomalies_timeseries” realizzata da Andriy Entsell. La logica del generatore prevede di inserire come parametri di input l’istante in cui comincia l’anomalia (in formato “timestamp”), la sua ampiezza, la sua durata e alcuni parametri specifici, ad esempio la banda (nel caso del jammer).

4.4 Simulazione di consumo del filo di contatto

Uno dei principali motivi per cui sono necessari interventi di manutenzione della linea è rappresentato dall’assottigliamento del filo di contatto causato dal passaggio dei treni. Calando la densità lineare, a parità di tensione, le frequenze naturali tendono a crescere. È stato dunque previsto un metodo che genera sequenze di passaggi contraddistinti da un contenuto spettrale le cui armoniche tendono a crescere nel tempo di una certa quantità prestabilita.

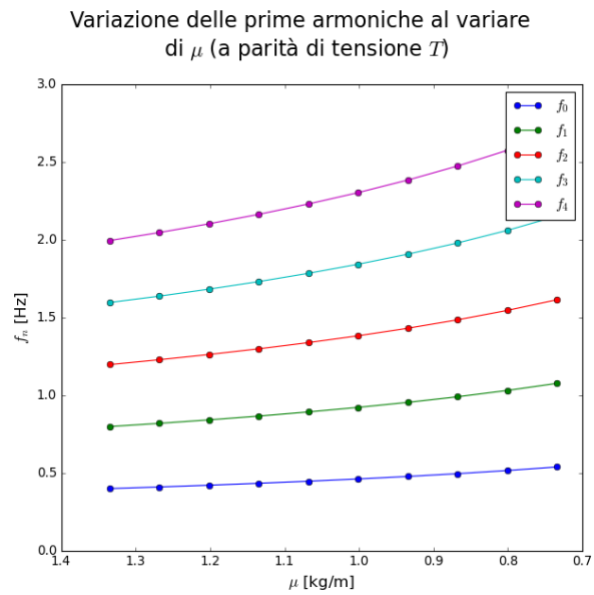


Figura 43 - Incremento in frequenza dei modi fondamentali a seguito del consumo del cavo.

In particolare, seguendo le leggi introdotte nel capitolo 1, si fornisce in ingresso al generatore il calo percentuale della densità lineare ad ogni passaggio e si ottiene un dataset che tiene conto dell'assottigliamento del cavo. Chiaramente, questo approccio simula in maniera del tutto ideale il fenomeno di consumo e rottura del cavo del quale non si hanno misure.

Capitolo 5

Implementazione

Alla luce degli aspetti più puramente teorici affrontati nei capitoli precedenti, si è poi operato al fine di rilevare la presenza di anomalie all'interno di un dataset e di stimare il consumo del filo al crescere del numero di passaggi. Il flusso di lavoro seguito per raggiungere i due obiettivi viene dunque raccontato in questo capitolo.

5.1 Analisi spettrale

Per la fase di pre-analisi, RFI ha fornito due set da quattro passaggi ciascuno, i primi quattro campionati a 200 Hz sono stati misurati nel 2017. I restanti quattro sono del 2019 e sono stati campionati alla massima frequenza possibile per lo strumento di misura, cioè 85470.08547 Hz. Partendo dai parametri tipici della catenaria italiana, è possibile calcolare approssimativamente la frequenza fondamentale e le sue armoniche principali:

$$L = 60 \text{ m}$$

$$T = 3059 \text{ kg}$$

$$\mu = 1,335 \text{ kg/m}$$

Si ha quindi:

$$f_n = \frac{n}{2L} \sqrt{\frac{T}{\mu}} \text{ con } n = 1,2,3,\dots$$

$$f_1 = \frac{1}{2 * 60 \text{ m}} \sqrt{\frac{3059 \text{ kg}}{1,335 \frac{\text{kg}}{\text{m}}}} =$$

$$= 0,39 \text{ Hz}$$

Trattandosi di un sistema meccanico caratterizzato da oscillazioni a bassa frequenza, la frequenza di campionamento utilizzata per le sequenze del 2017 è sicuramente ampiamente sufficiente.

5.1.1 Analisi tempo-frequenza

Data la natura non-stazionaria dei passaggi, si è operata in prima battuta un'analisi tempo-frequenza tramite gli strumenti descritti nel capitolo 2.

STFT

I principali aspetti di cui tenere conto nell'applicazione della STFT riguardano la scelta della funzione finestra, la sua lunghezza e la percentuale di sovrapposizione tra finestre adiacenti. Tenendo conto di quanto appena detto sulle frequenze in gioco, per avere una buona risoluzione in frequenza sarebbe opportuno che la finestra contenesse una decina di periodi T dell'oscillazione a minor frequenza:

$$W_L = 10 * T =$$

$$= 10 * f_s * \frac{1}{f_1} =$$

$$= 10 * 200 \text{ Hz} * \frac{1}{0,39 \text{ Hz}} =$$

$$= 5128 \text{ campioni}$$

Chiaramente, il numero di periodi contenuti nella finestra può essere diminuito o aumentato nel caso si volesse prediligere rispettivamente la risoluzione temporale o quella frequenziale. La funzione finestra scelta è quella gaussiana poiché garantisce la maggior risoluzione tempo-frequenza.

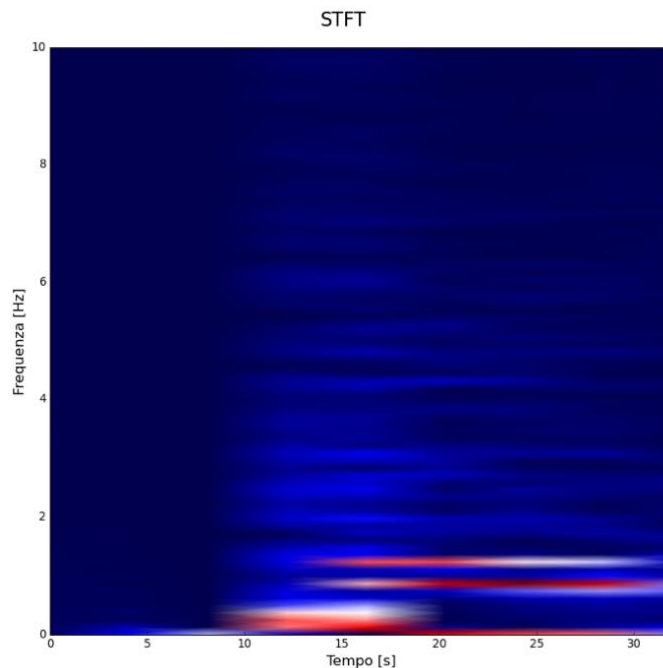


Figura 44 - STFT di un passaggio del 2017.

Nella fase antecedente al picco si intravedono le stesse armoniche del post-passaggio ma queste denotano chiaramente un'ampiezza non significativa. Si ha poi un contributo nell'intervallo $[0, 0.5]$ Hz che corrisponde temporalmente al passaggio del pantografo. Da questo momento in poi ha origine la fase di post-passaggio, nella quale sono visibili due armoniche principali attorno a 0.87 e 1.12 Hz e una serie di armoniche superiori di ampiezza nettamente inferiore. Le oscillazioni in generale tendono a mantenere la stessa frequenza lungo tutta la sequenza ma si può notare che alcune hanno origine più tardi rispetto al picco. Trattandosi di un filo con estremi vincolati si avranno riflessioni che dopo un certo tempo raggiungono il punto di misura.

CWT

Noti i vantaggi della CWT rispetto alla STFT, è stata effettuata un'analisi wavelet con la Morlet. Il numero di scale adottate è 256.

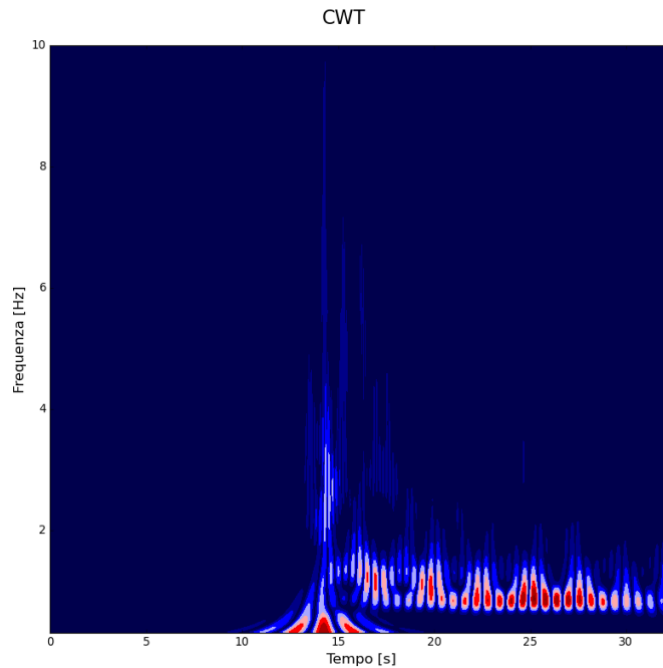


Figura 45 - Applicazione della CWT ad un passaggio del 2017.

Ora appare ancora più evidente quanto detto per la STFT in merito a riflessioni e battimenti. Si nota una sorta di intermittenza temporale dell'armonica a 1.1 Hz.

5.1.2 Stima di spettro

Considerando quanto appena descritto con l'analisi tempo-frequenza, risulta di particolare interesse concentrarsi sul post-passaggio e stimare con precisione le armoniche che lo compongono sia dal punto di vista della frequenza che della relativa ampiezza. L'obiettivo di questo passo è quello di estrarre i valori che verranno poi dati in pasto al generatore. Noto il comportamento fortemente passa-

basso evidenziato nell'analisi tempo-frequenza, l'approccio adottato consiste nell'individuare quelle armoniche principali nella banda [0,10] Hz e modellarne la variabilità sia in frequenza che in ampiezza. I metodi descritti nel capitolo 2 sono stati implementati in modo tale da sfruttare i pregi e limitarne i difetti.

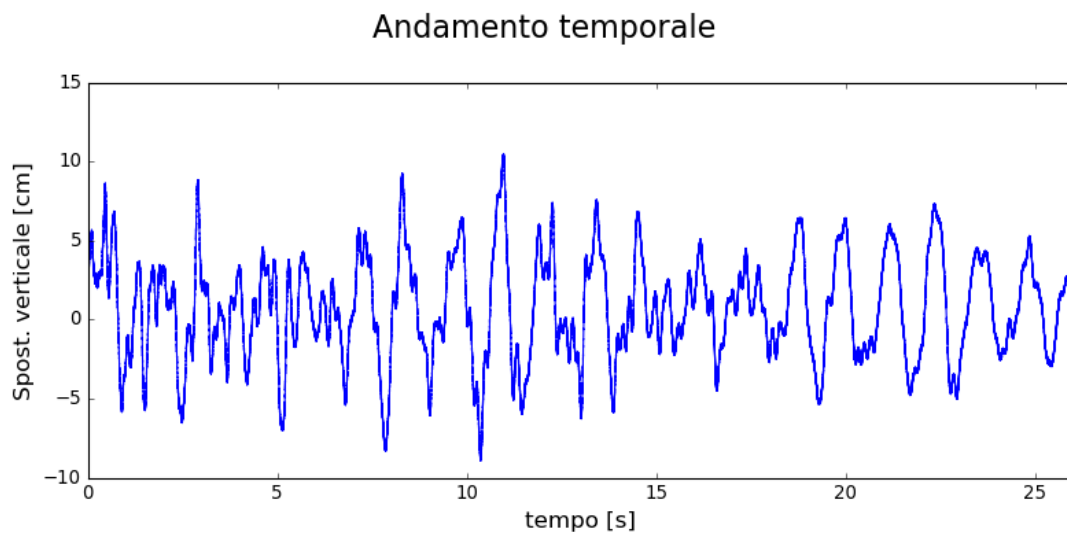


Figura 46 - Selezione del post-passaggio.

Naturalmente, vista la natura “stazionaria” dell’analisi di Fourier, si è scartato l’intervallo di tempo contenente il picco, applicando i metodi di stima solo al post-passaggio.

Periodogramma

Il pregio di questo stimatore è di essere asintoticamente non polarizzato. Disporre di un elevato numero di campioni, come nel caso del set del 2019, risulta quindi molto conveniente e permette di abbattere la polarizzazione. La sequenza in questione è di lunghezza 2234694 campioni.

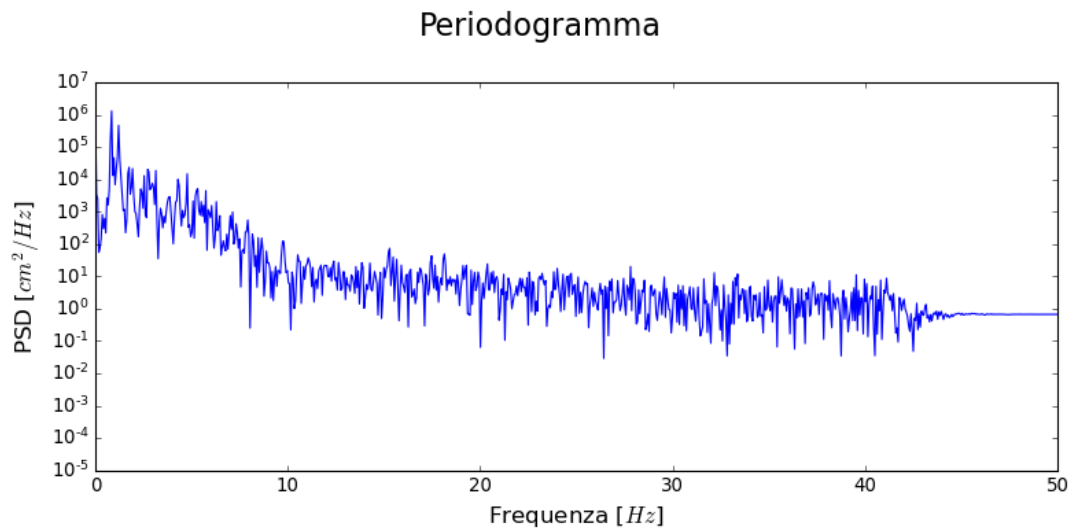


Figura 47 – Stima di spettro di potenza di un passaggio del 2017 tramite periodogramma.

Il risultato ottenuto conferma che la densità spettrale di potenza è concentrata principalmente entro i 10 Hz. Si possono distinguere facilmente due picchi piuttosto marcati al di sotto dei 2 Hz e altri meno pronunciati a frequenze superiori. Tuttavia, appare complicato distinguere le vere e proprie armoniche dal disturbo introdotto dallo stimatore.

Periodogramma modificato

Al fine di ridurre la varianza del periodogramma sacrificandone in parte la polarizzazione, è stata introdotta una finestrazione che permette di mediare più stime dello stesso spettro. L'introduzione di una funzione finestra riduce implicitamente il numero di campioni con cui si stima lo spettro e pertanto peggiora la risoluzione in frequenza. Noto tale compromesso, si è scelta una funzione finestra rettangolare di lunghezza pari a metà della lunghezza dell'intera sequenza, cioè 1117347 campioni.

Periodogramma modificato

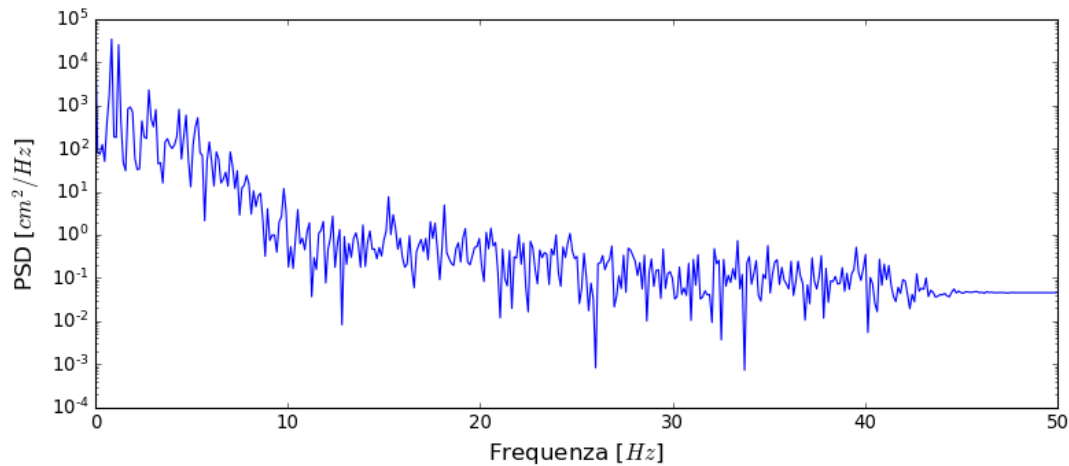


Figura 48 – Stima di spettro di potenza di un passaggio del 2017 tramite periodogramma modificato.

Così facendo si è ridotto il disturbo introdotto dallo stimatore pur mantenendo ben distinti i picchi principali.

Stimatore a minima varianza

A valle di una prima analisi effettuata con i due stimatori precedenti, si è deciso di applicare lo stimatore a minima varianza, concentrando l'indagine nell'intervallo 0-10 Hz, in modo tale da stimare con precisione le armoniche con ampiezza maggiore. Al contrario dello stimatore a periodogramma, quello a minima varianza risulta di difficile implementazione nel caso di sequenze molto lunghe, poiché richiede la stima dell'inversa di una matrice di correlazione. Il costo computazionale è elevato ed aumenta in maniera considerevole all'aumentare della lunghezza della sequenza e dell'ordine del filtro che si vuole implementare. Per questo motivo, si è deciso di utilizzare i dati del 2017 opportunamente decimati di un fattore 2 per ricondurli ad una frequenza di campionamento pari a 100 Hz. La scelta dell'ordine del filtro è stata fatta cercando di massimizzare la risoluzione in frequenza. Si è riscontrato che già con un ordine pari a 511 l'algoritmo non converge e riporta risultati non

interpretabili, pertanto si è scelto un ordine pari a 255. Per questa fase sono stati normalizzati gli andamenti temporali.

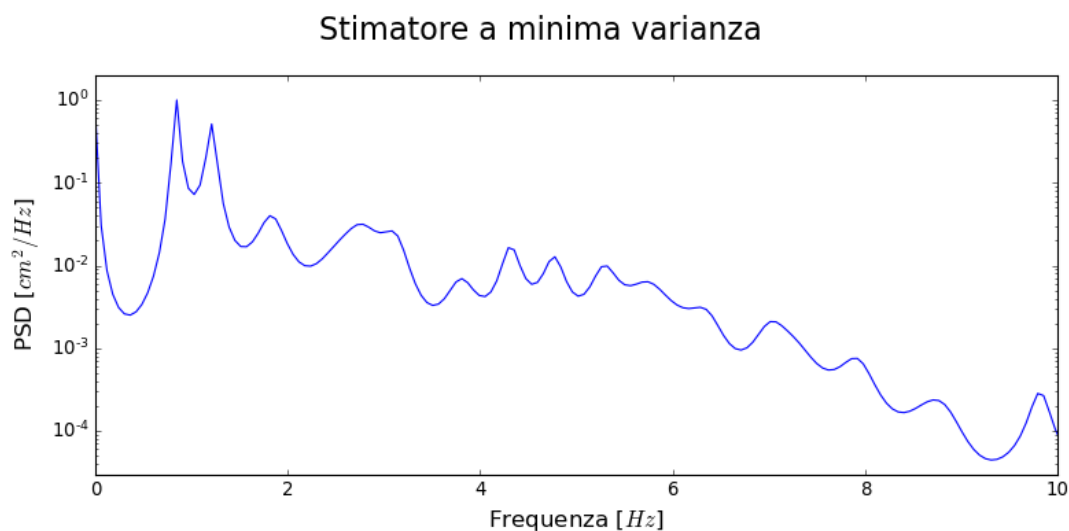


Figura 49 - Spettro di potenza di un passaggio del 2017 stimato con stimatore a minima varianza in scala logaritmica.

Il profilo ottenuto è molto più intellegibile, l'andamento è meno seghettato e consente l'individuazione delle armoniche. Per migliorare ulteriormente la prominenza, utile ad agevolare il confronto tra gli spettri dei vari passaggi, si è scelta una scala lineare:

Confronto passaggi

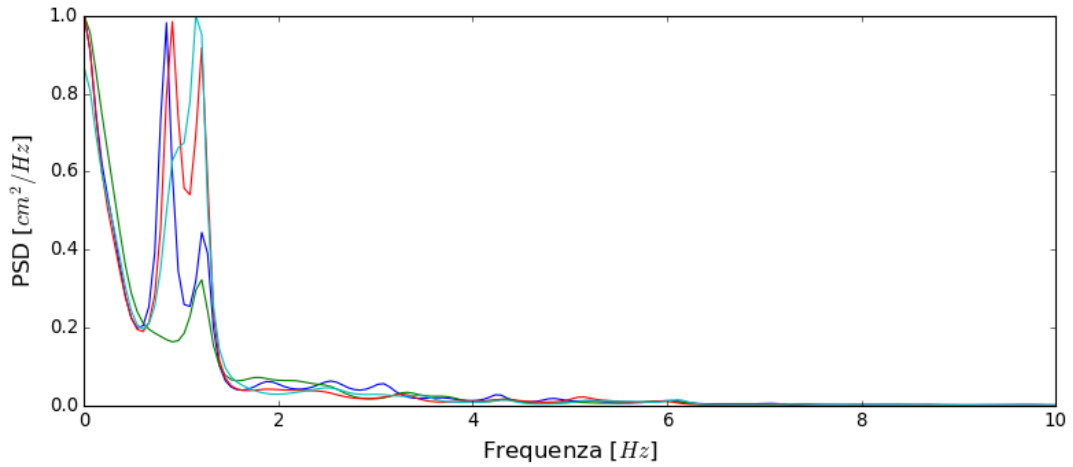


Figura 50 - Confronto tra i quattro passaggi del 2017. Dettaglio dell'intervallo di frequenze [0, 10] Hz in scala lineare.

Le due armoniche principali, che essendo molto vicine danno luogo al fenomeno dei battimenti, hanno un'ampiezza piuttosto variabile. Oltre alle prime due armoniche si evincono una serie di ulteriori picchi di ampiezza molto inferiore. Di seguito un focus sulla banda [1.5, 10] Hz:

Confronto passaggi

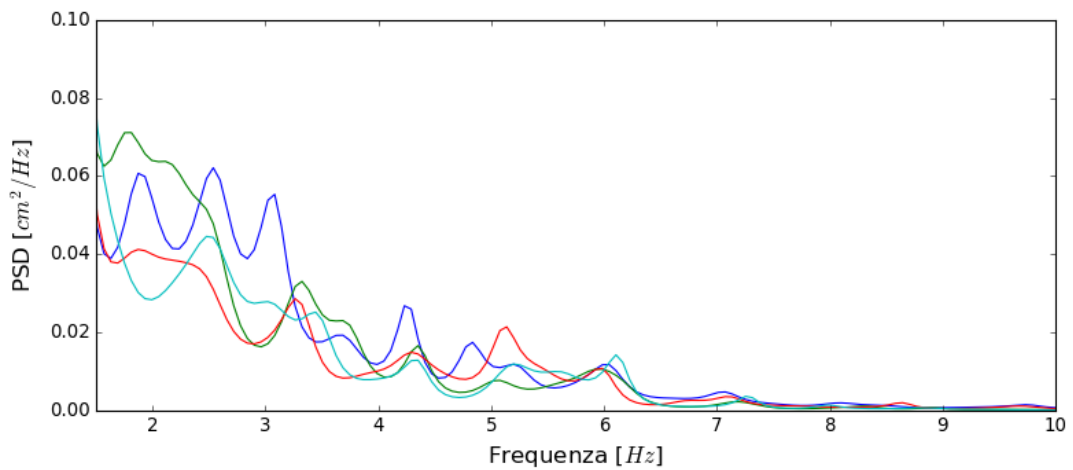


Figura 51 - Focus sull'intervallo [1.5, 10] Hz.

In questo intervallo la situazione è più complessa, alcuni picchi sono presenti in certi passaggi e completamente assenti in altri. Sono stati fatti diversi tentativi volti all'estrarre sistematicamente e in maniera automatica i parametri di ciascuna armonica ma è risultato piuttosto difficoltoso ottenere dei risultati soddisfacenti a causa della consistente variabilità in ampiezza e frequenza dei vari picchi. Si è proceduto quindi in maniera grafica e sono stati estratti i seguenti parametri:

```

f[0] = [0.858, 0.916, 0.915, 0.817] #0.86
a[0] = [0.925, 0.956, 0.647, 0.176]

f[1] = [1.205, 1.204, 1.147, 1.203] #1.21
a[1] = [0.437, 0.911, 1.000, 0.324]

f[2] = [1.892, 1.887, 1.720, 1.739] #1.8
a[2] = [0.065, 0.041, 0.032, 0.042]

f[3] = [2.526, 2.444, 2.522, 2.463] #2.5
a[3] = [0.062, 0.035, 0.045, 0.052]

f[4] = [3.040, 3.010, 3.010, 3.010] #3.0
a[4] = [0.056, 0.002, 0.019, 0.017]

f[5] = [3.200, 3.270, 3.325, 3.441] #3.2
a[5] = [0.042, 0.029, 0.033, 0.025]

f[6] = [3.692, 3.715, 3.725, 3.675] #3.7
a[6] = [0.019, 0.023, 0.010, 0.008]

f[7] = [4.244, 4.287, 4.357, 4.328] #4.3
a[7] = [0.027, 0.015, 0.016, 0.013]

f[8] = [4.817, 4.894, 4.794, 4.898] #4.8
a[8] = [0.017, 0.004, 0.003, 0.002]

f[9] = [5.107, 5.195, 5.204, 5.056] #5.1
a[9] = [0.021, 0.012, 0.017, 0.007]

f[10] = [6.020, 6.133, 5.967, 5.963] #6.0
a[10] = [0.011, 0.014, 0.001, 0.001]

```

Figura 52 - Vettori contenenti le frequenze e le ampiezze delle armoniche estratte dai dati reali.

Con questi vettori sono stati calcolati i vettori delle medie e delle varianze delle ampiezze e delle frequenze di ciascun' armonica:

	1	2	3	4	5	6	7	8	9	10	11
<i>f_mean</i>	0.876	1.12	1.81	2.49	3.02	3.31	3.701	4.304	4.851	5.141	6.021
<i>f_std</i>	0.042	0.025	0.08	0.036	0.013	0.088	0.019	0.043	0.046	0.062	0.068
<i>a_mean</i>	0.676	0.668	0.045	0.048	0.024	0.032	0.015	0.018	0.006	0.014	0.007
<i>a_std</i>	0.313	0.292	0.012	0.01	0.02	0.006	0.006	0.005	0.006	0.005	0.006

Figura 53 - Parametri delle gaussiane che modellano la variabilità in frequenza ed ampiezza delle varie armoniche.

5.2 Generazione del dataset di train

I vettori appena estratti vengono forniti in ingresso alla classe che genera i dataset sintetici. Per i passi successivi risulta rilevante la scelta di alcuni parametri, in particolare la lunghezza di ogni passaggio e il numero di passaggi. Si è deciso di generare passaggi lunghi 60 s e di generarne un numero sufficiente a garantire che in metodi come la PCA, la stima della matrice di covarianza sia accurata. Il limite superiore in merito al numero delle occorrenze è determinato dal carico computazionale che incrementa sensibilmente all'aumentare della lunghezza del dataset, aumentando considerevolmente i tempi di elaborazione. Si è riscontrato che un numero di passaggi pari a 10000 è più che sufficiente. Per mantenere una certa fedeltà rispetto al set reale, è stata inoltre fissata la frequenza di campionamento a 100 Hz. Complessivamente, il dataset generato ha le seguenti dimensioni $60 \times 100 \times 10000 = 6 \cdot 10^7$ campioni. Il dataset di train prevede l'alternanza di passaggi a 160 km/h e 180 km/h e quindi denoterà un susseguirsi di istanze di cui alcune saranno in media meno ampie (sia nel picco che nelle oscillazioni) e altre più marcate. La percentuale di rumore è stata determinata anch'essa osservando i dataset reali.

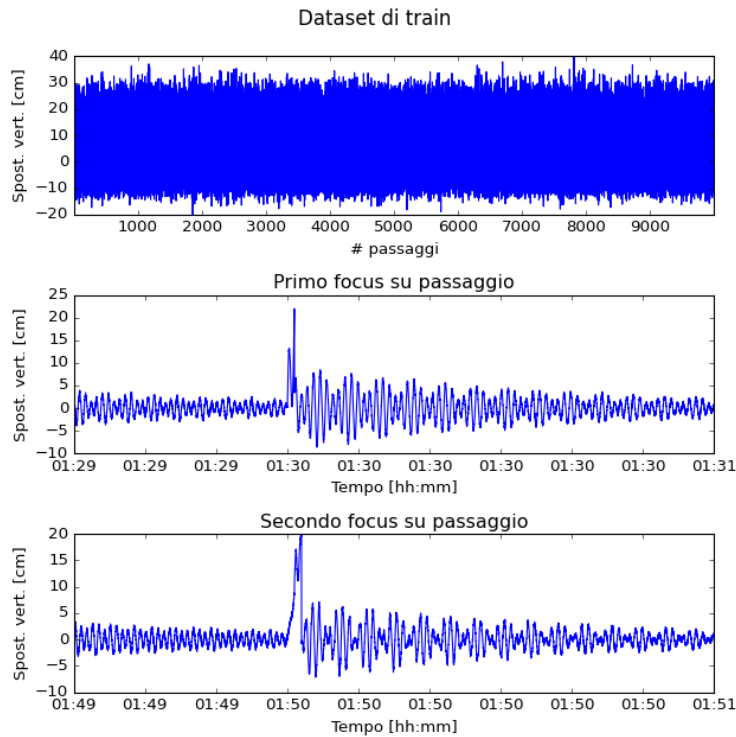


Figura 54 - Esempio di Dataset di train con focus su un paio di passaggi.

5.3 Generazione del dataset di test

Il dataset di test viene generato con gli stessi crismi di quello di train, fatta eccezione per l'aggiunta delle anomalie. Per una prima fase esplorativa si è scelto di aggiungere un'occorrenza per ciascuna tipologia di anomalia. Tutte le anomalie hanno la stessa durata e la stessa ampiezza, quest'ultima viene rapportata alla deviazione standard della sequenza del dataset di train. La loro manifestazione temporale è stata cadenzata con regolarità soprattutto per facilitarne la visione in questa fase di esplorazione preliminare. Nel seguente grafico sono indicati con un "1" gli istanti temporali in cui è presente un'anomalia mentre con "0" i campioni non anomali. Per le analisi seguenti l'ordine temporale delle anomalie è: impulso rettangolare, jammer, AWGN, anomalia periodica, impulso triangolare. In particolare, il jammer copre la banda [10, 16] Hz mentre l'anomalia periodica ha una frequenza pari a 0.4

Hz. Le cinque anomalie sono state posizionate rispettivamente al minuto 00:30, 01:30, 02:30, 03:30 e 04:30. Per semplicità verranno indicati sull'asse dei tempi i numeri corrispondenti ai passaggi anomali. Ad esempio, il 150-esimo passaggio è affetto dall'anomalia AWGN. Sempre a scopo esplorativo, si è scelto di generare inizialmente anomalie piuttosto consistenti di durata 60s e ampiezza pari a metà della deviazione standard del dataset di train.

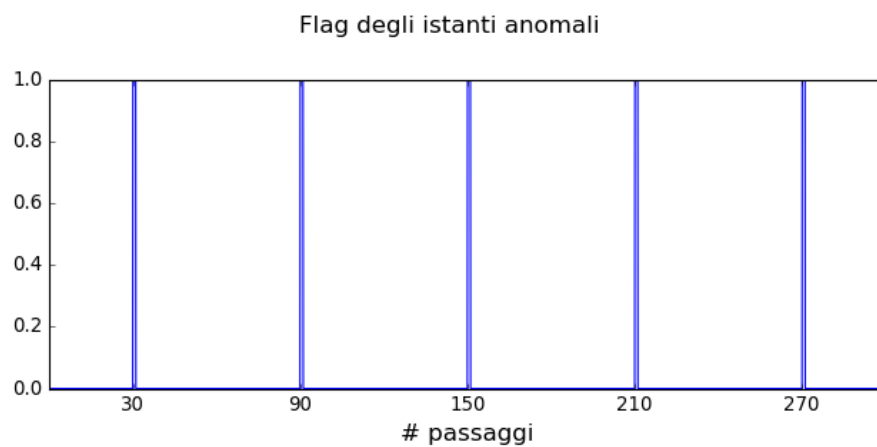


Figura 55 - Identificazione degli istanti temporali in cui sono presenti le anomalie utile a calcolare il ROC.

In questo caso non è necessario stimare una matrice di covarianza e quindi è possibile considerare anche 1 solo passaggio (che potenzialmente è quello che avverrebbe nel caso pratico). Tuttavia, per migliorare la visualizzazione si è scelto di considerare qualche centinaio di passaggi, nello specifico 300.

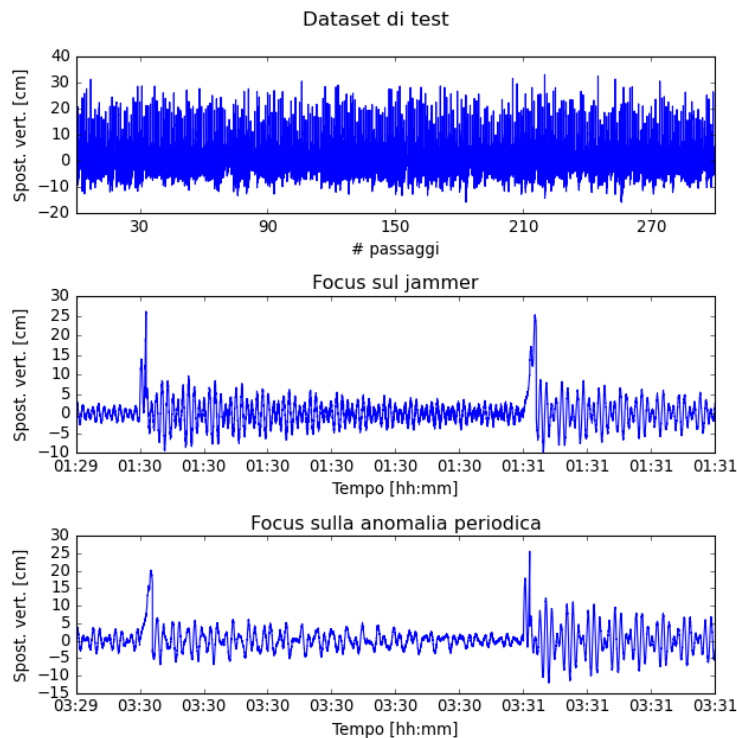


Figura 56 - Esempio di Dataset di test con focus su un paio di passaggi anomali.

5.4 Detectors basati sulla PCA

Per poter applicare correttamente la PCA è necessario riorganizzare il dataset, scomponendolo in finestre di una certa lunghezza e ponendone una sotto l'altra, questo produrrà una matrice che ha un numero di colonne pari al numero di campioni della finestra. Poiché i segnali in questione non sono stazionari, l'utilizzo della finestra risulta particolarmente utile per fare in modo che il segnale sia il più possibile stazionario all'interno della finestra stessa. Tuttavia, la lunghezza della finestra determina anche la minima frequenza che è possibile indagare, esiste quindi un trade-off tra questi due aspetti. È inoltre possibile prevedere un certo overlap. Seguendo il discorso introduttivo fatto sul principio di funzionamento della PCA, gli step eseguiti sono:

1. Suddividere in finestre i dataset di train e di test
2. Rendere il dataset a media nulla
3. Applicare la PCA al dataset di train in modo tale che l' algoritmo ottimizzi la trasformazione lineare che esprime la varianza desiderata.
4. Applicare il modello allenato al dataset di test
5. Calcolare i possibili score

La lunghezza della finestra è stata fissata a valle di alcuni tentativi esplorativi, volti a trovare il miglior compromesso per il caso in esame. Noto il limite inferiore dovuto alla minima frequenza analizzabile, si è sperimentato che con finestre più lunghe di 10 s non si ottenevano buoni risultati. Sono state provate dunque finestre di 2 s, 5 s e 10 s e, alla fine, si è optato per fissare 10 s (nonché 1000 campioni) per tutti i detectors che verranno descritti da qui in poi. Si è scelto inoltre di impostare l'overlap al 50%.

In secondo luogo, è stato necessario visualizzare il profilo degli autovalori e della varianza espressa, in scala logaritmica, al fine di individuare il numero di componenti principali.

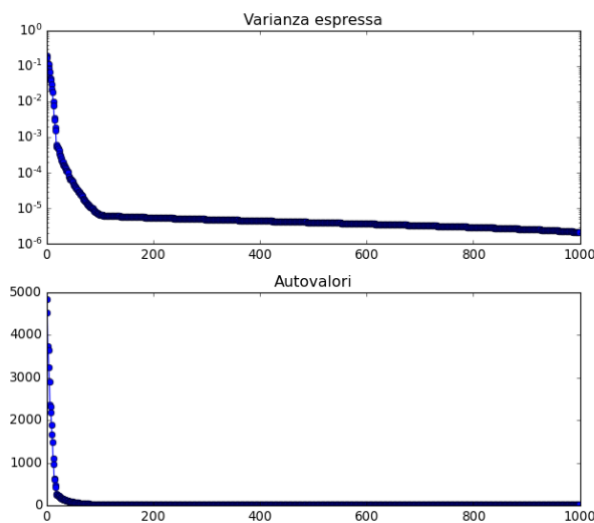


Figura 57 - Profilo degli autovalori e della varianza espressa da ciascuno di essi in scala logaritmica.

Come si evince, dopo un decadimento esponenziale si ha un cambio di pendenza, in questo caso ciò avviene circa alla 100-esima componente. Il modello viene di conseguenza ottimizzato imponendo all'algoritmo di mantenere le prime 100 componenti. I tempi di elaborazione sono veramente contenuti nonostante il numero elevato di passaggi. Dal punto di vista computazionale l'SVD è un metodo piuttosto consolidato e rende la PCA molto performante sotto questo punto di vista. Successivamente si applica il modello ottenuto al dataset di test, ottenendone una proiezione che verrà utilizzata per calcolare i vari score.

5.4.1 PCA + T^2

Viene ripresa per comodità di lettura la formulazione ottenuta nel capitolo 3 per questo score:

$$T^2 = \sum_{j=0}^{n-1} \left(\frac{1}{\sqrt{\Lambda_j}} e_j^T x \right)^2$$

Nell'implementazione pratica, gli autovalori Λ_j e gli autovettori e_j vengono estratti dal modello allenato, mentre con x si indica il dataset di test. Di seguito viene presentato l'andamento temporale dello score, il quale rappresenta quanto è anomala ciascuna finestra del dataset.

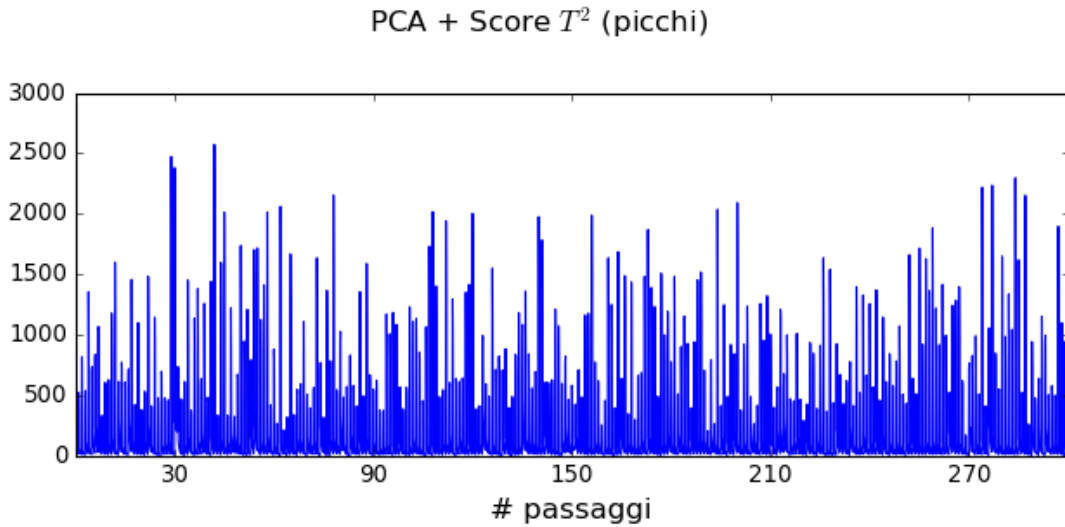


Figura 58 - Score T^2 del dataset di test d'esempio.

Si nota chiaramente un elevato disturbo, attribuibile alla non-stazionarietà del segnale, specialmente per quanto concerne il picco. In questo caso d'esempio si è scelto di aggiungere anomalie consistenti (con deviazione standard d'ampiezza pari alla metà di quella del dataset) per rendere comunque visibile la presenza di tratti in cui lo score raggiunge valori elevati per un tempo prolungato (corrispondente con la durata dell'anomalia). Nonostante ciò, il disturbo sovrasta notevolmente le anomalie come si può notare anche dal seguente dettaglio:

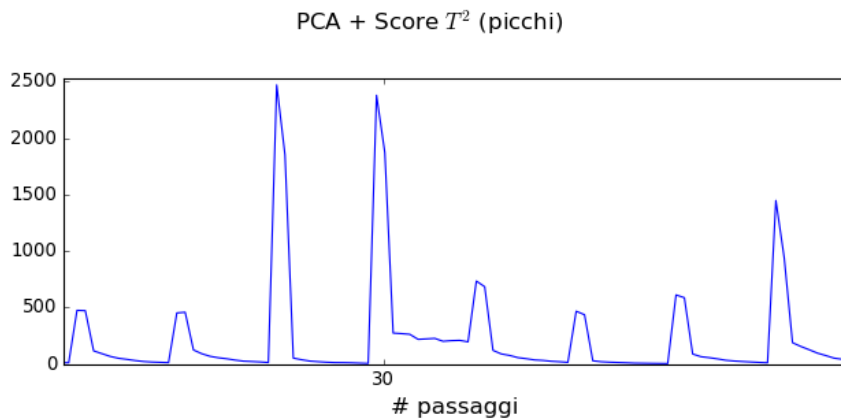


Figura 59 - Dettaglio sull'anomalia presente al trentesimo passaggio, cioè l'impulso rettangolare.

5.4.2 PCA + SPE

Lo score SPE è stato calcolato seguendo questa espressione:

$$\text{SPE} = \|x - \hat{x}\|^2$$

Dove con x si indica il dataset di train e con \hat{x} la proiezione del dataset di train sul modello allenato. Il risultato ottenuto in questo esempio è anch'esso fortemente compromesso dalla non-stazionarietà generale del segnale che culmina nei picchi.

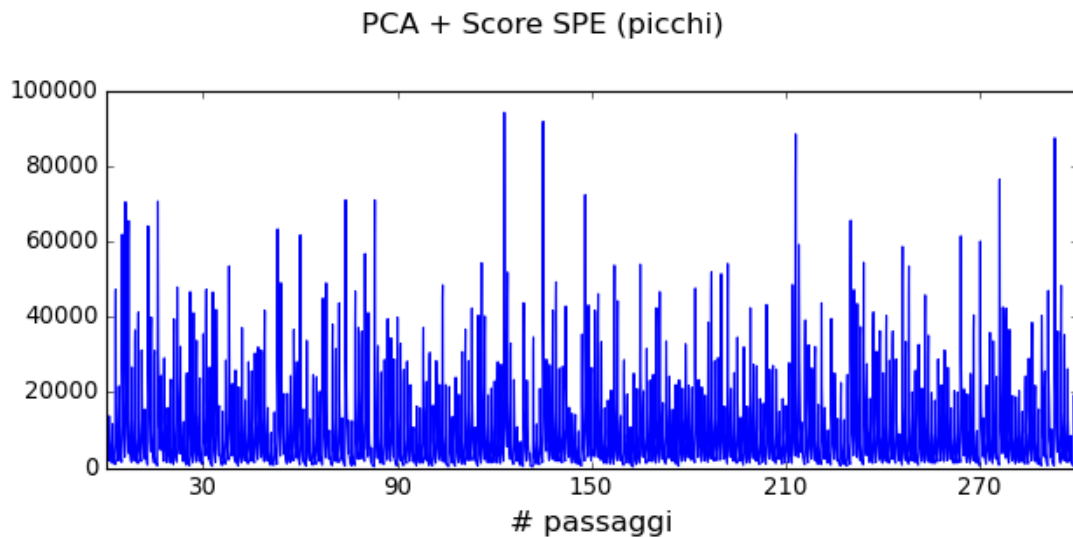


Figura 60 - Score SPE del dataset di test d'esempio.

A differenza dello score T^2 , l'SPE sembra avere difficoltà anche a rilevare la presenza delle anomalie. In questo caso si hanno praticamente solo falsi positivi causati dai picchi e dal decadimento esponenziale.

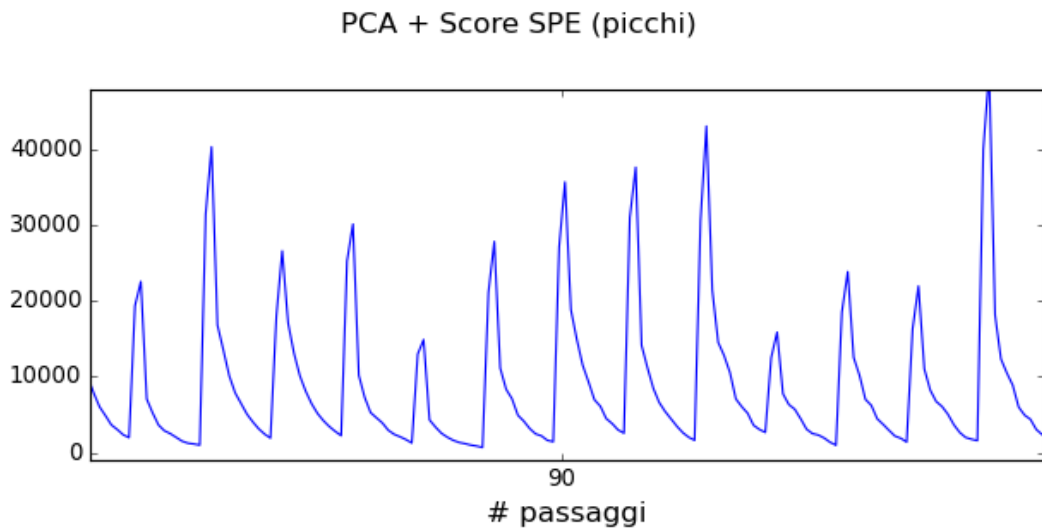


Figura 61 - Dettaglio del passaggio numero 90 in cui è presente il jammer.

5.4.3 PCA + Mahalanobis

Lo score di Mahalanobis si presenta come meno suscettibile al disturbo introdotto dalla non-stazionarietà ma anch'esso non è in grado di produrre risultati accettabili. Similmente al T^2 si rilevano più facilmente le anomalie che abitano lo spazio principale.

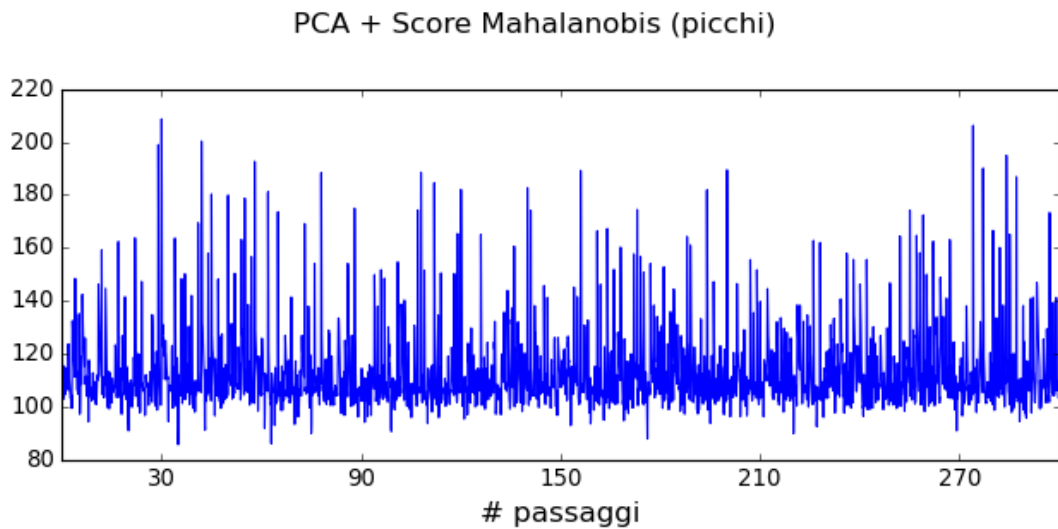


Figura 62 - Distanza di Mahalanobis della proiezione del dataset di test d'esempio dalla proiezione del dataset di train.

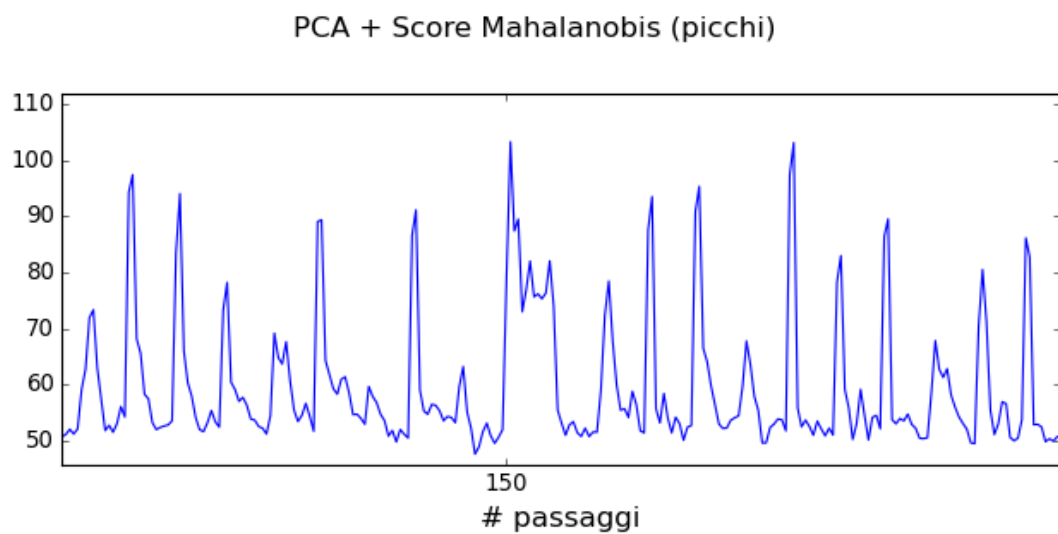


Figura 63 - Dettaglio del passaggio numero 150 in cui è presente l'AWGN.

5.4.4 ROC - PCA

Già in questo caso di esempio, in cui l'ampiezza delle anomalie è volutamente considerevole, si ottengono dei risultati non soddisfacenti. La non-stazionarietà causa un'elevata percentuale di falsi positivi che compromette l'esito della rilevazione.

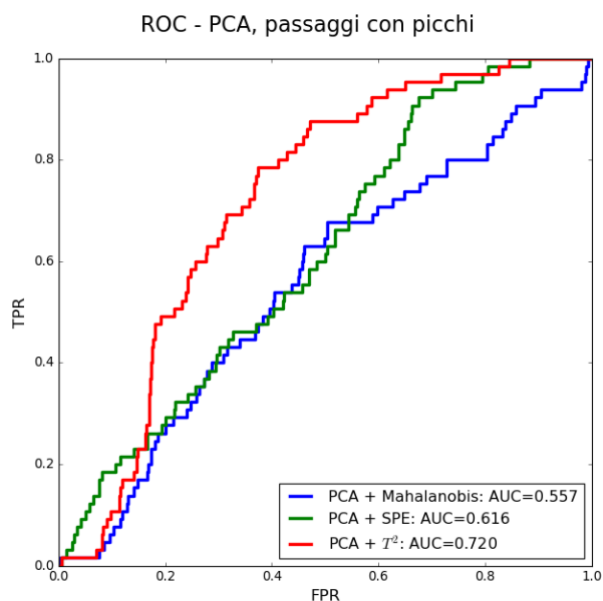


Figura 64 - Curva del ROC dei tre detectors nel caso d'esempio.

In ogni caso, stando all'AUC ottenuta, i tre detectors restituiscono risultati analoghi. È necessario chiarire che il risultato ottenuto dipende fortemente dalla tipologia di anomalie presenti nel dataset, dalla loro ampiezza e durata e dalla durata del dataset stesso. Se ad esempio si avessero solo anomalie AWGN si otterrebbero risultati migliori. Al contrario, il jammer appare già come piuttosto difficile da rilevare e un dataset affetto solo da anomalie di tipo jammer comporterebbe un AUC molto peggiore.

Inoltre, è bene far notare che la scelta della soglia che permette di distinguere tra anomalia e comportamento nominale deve essere scelta accuratamente. Osservando le curve del ROC dei tre detectors si nota che impostando una soglia in grado di rilevare tutte le anomalie si rileveranno inevitabilmente anche una quantità inaccettabile di falsi positivi. Al fine di ottenere uno strumento efficace è necessario che vi sia un certo margine che permetta di scegliere una soglia tale da rilevare anche le più piccole anomalie ma limitando il più possibile la segnalazione di falsi positivi.

Visti i limiti dell'implementazione appena descritta, si è pensato di provare a migliorare i risultati concentrando l'analisi unicamente sul post-passaggio. A tal proposito sono stati generati dataset di train e test composti da passaggi senza il picco. Successivamente sono stati seguiti gli stessi passaggi appena descritti e si è ottenuto il seguente risultato:

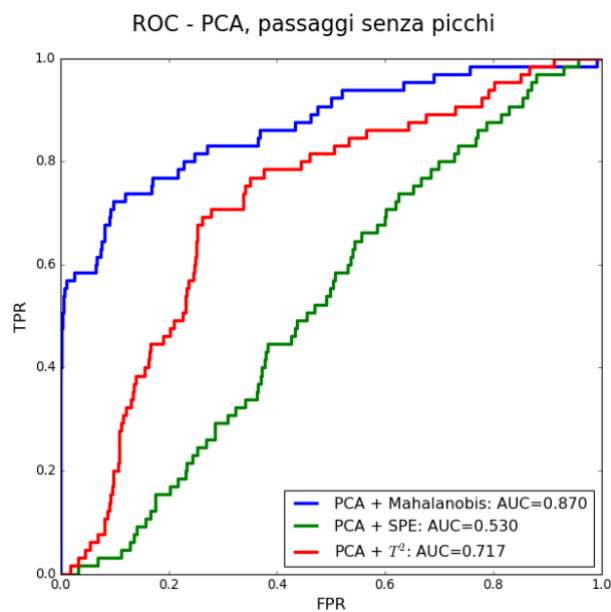


Figura 65 - Curva del ROC per i tre detectors nel caso d'esempio con passaggi senza picchi.

Come si può notare si ottiene un notevole miglioramento degli score T^2 e di Mahalanobis. Nonostante ciò, va ribadito che in questi esempi le anomalie sono piuttosto consistenti e quindi "facili" da rilevare.

Nel tempo si nota che gli score T^2 e di Mahalanobis sono più "puliti" e le anomalie sono più evidenti:

PCA + Score T^2 (no picchi)

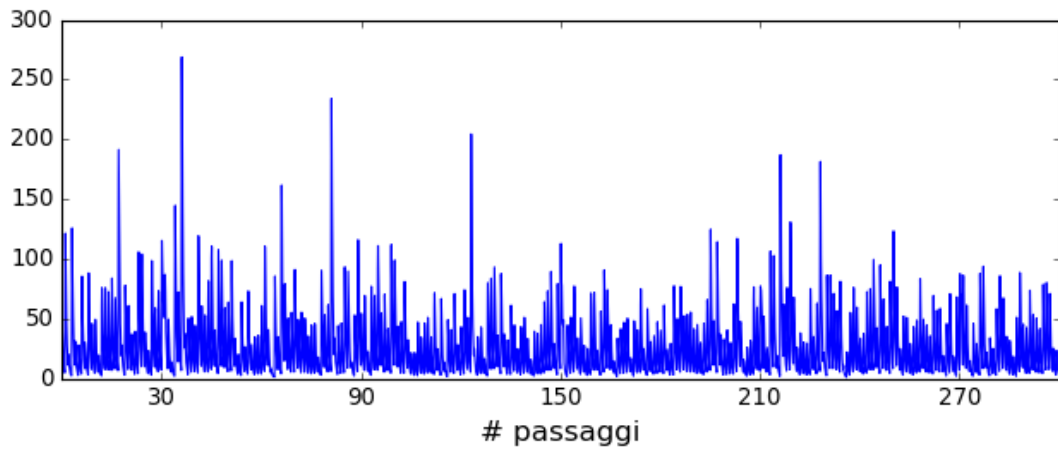


Figura 66 - Score T^2 nel caso di dataset senza picchi.

PCA + Score Mahalanobis (no picchi)

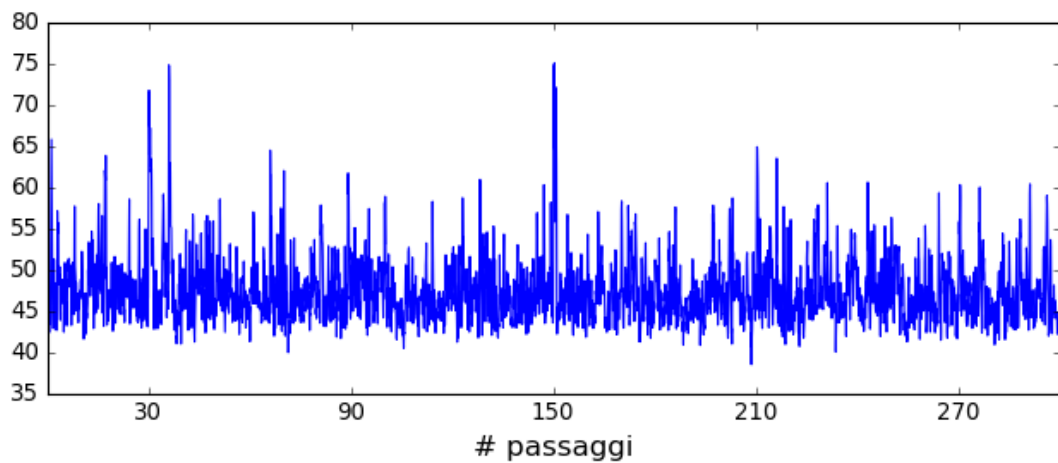


Figura 67 - Score di Mahalanobis nel caso di dataset senza picchi.

La presenza del disturbo generato dal decadimento esponenziale unitamente alla scarsa capacità di rilevare le anomalie, non consente di utilizzare la PCA nella sua versione classica.

5.5 Detectors basati sulla KPCA

Considerando la presenza del picco e del decadimento esponenziale è plausibile che una trasformazione lineare non sia in grado di estrarre le componenti principali dai dataset. Si è tentato dunque di applicare la KPCA sperando di ottenere dei miglioramenti rispetto a quanto ottenuto con la PCA. La chiave per raggiungere i migliori risultati con questo strumento consiste nella scelta del kernel e dei suoi parametri. Tecnicamente, sarebbe stato opportuno implementare delle strategie di ottimizzazione che, in maniera rigorosa ed automatica, fossero in grado di ricavare il kernel ottimo per questo tipo di segnali. Per ragioni temporali si è preferito dedicare questo tempo allo studio di uno strumento aggiuntivo (cioè la RPCA), limitando la scelta del kernel ad una serie di tentativi empirici. In seguito alla prova sommaria dei kernel disponibili si è deciso di puntare sul kernel “cosine” poiché, in prima istanza, ha dato luogo i migliori risultati. Solo per quanto riguarda la KPCA, per ragioni di spazio di memoria insufficiente, non è stato possibile allenarla con 10000 passaggi ma si è dovuto considerarne la metà, cioè 5000. Per poter applicare la funzione kernel, l’algoritmo necessita di allocare più di 50 GB di memoria, il che richiede sicuramente l’utilizzo di un centro di calcolo adeguato. Come fatto nel caso della PCA, si è valutato il profilo degli autovalori al fine di selezionare il numero di componenti principali. Anche per la KPCA si è scelto di tenere le prime 100 componenti.

5.5.1 KPCA + T^2

Lo score T^2 non presenta risultati soddisfacenti. Similmente alla PCA, il detector soffre ancora della presenza dei picchi:

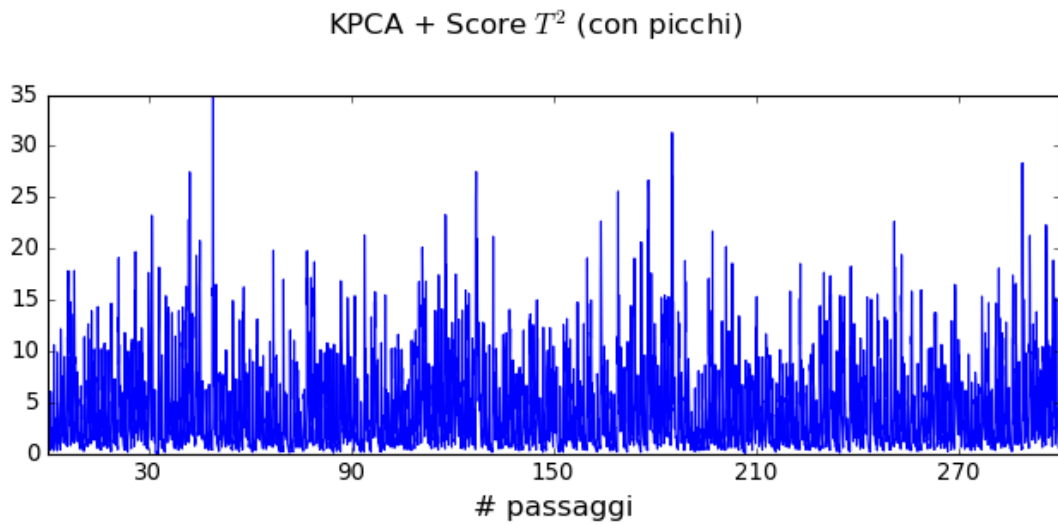


Figura 68 - Score T^2 del dataset di test d'esempio.

Il dataset senza picchi consente di ottenere risultati sicuramente migliori ma comunque non accettabili.

5.5.2 KPCA + SPE

Per quanto riguarda lo score SPE, i risultati ottenuti con la KPCA sono leggermente migliori rispetto alla PCA:

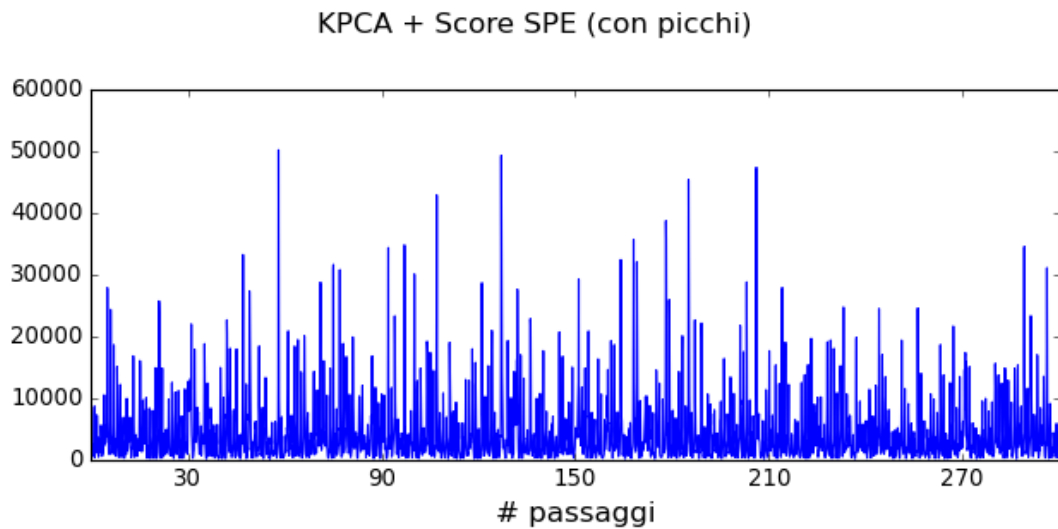


Figura 69 - Score SPE del dataset di test d'esempio.

Anche per l'SPE si ha un leggero miglioramento se vengono rimossi i picchi ma quanto ottenuto con la KPCA è in questo caso molto peggiorativo rispetto alla PCA.

5.5.3 KPCA + Mahalanobis

Nonostante sia stato più che dimezzato il disturbo rispetto alla PCA, la presenza dei picchi limita ancora l'efficacia di questo detector. In ogni caso, per quanto riguarda la distanza di Mahalanobis, si ha un certo miglioramento rispetto alla PCA:

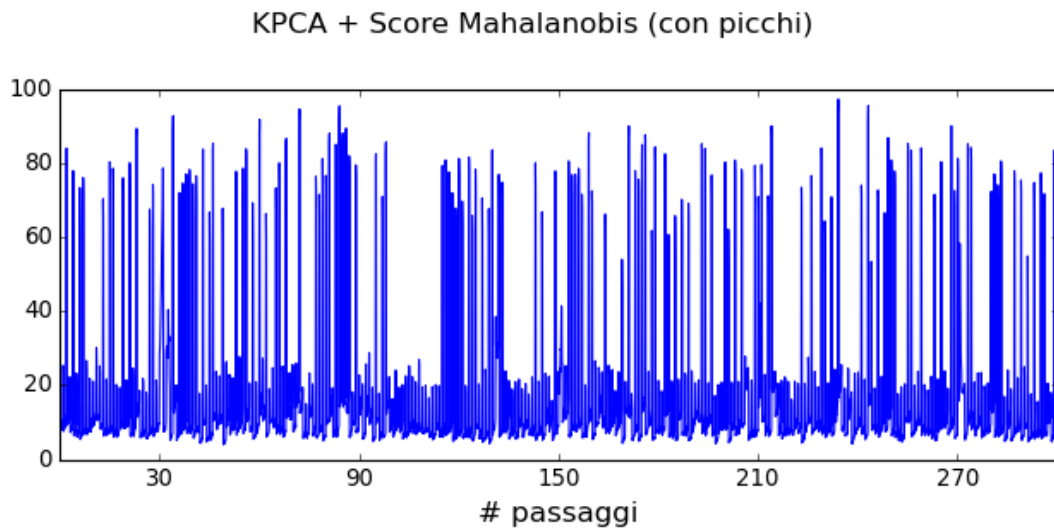


Figura 70 - Score di Mahalanobis del dataset di test d'esempio.

L'applicazione al post-passaggio comporta anche in questo caso la riduzione del disturbo, consentendo di ottenere risultati rilevanti ma comunque non sufficienti ad eleggere questo detector come un buon candidato.

5.5.4 ROC - KPCA

Purtroppo, complice in parte il numero limitato di kernel e di possibili parametri che sono stati provati, la KPCA non ha risolto il problema che affligge i detectors basati su PCA. In generale si è ottenuto però un miglioramento in merito allo score di Mahalanobis che ora assume performance non trascurabili ma comunque non soddisfacenti. Anche in questo caso vengono rappresentate le curve del ROC in questa casistica “agevolata” in cui, si ribadisce, le anomalie sono volutamente consistenti.

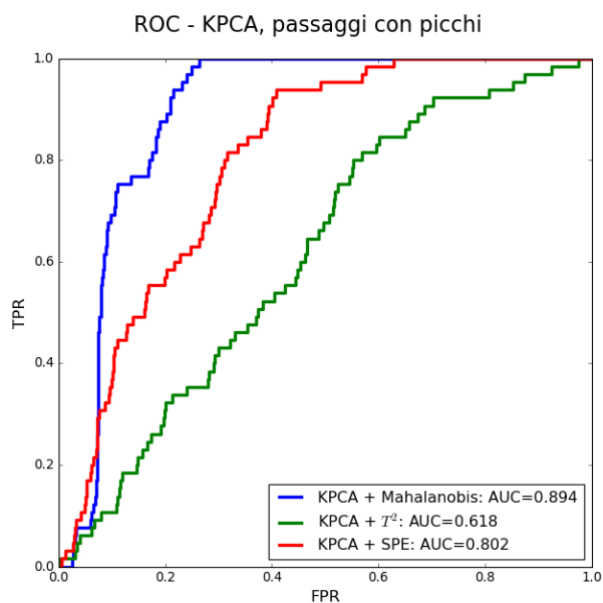


Figura 71 - Curva del ROC associata ai tre detectors basati su KPCA.

5.6 Detectors basati sulla RPCA

La causa delle scarse performance dei detectors precedenti, specialmente nel caso in cui non si volesse scartare i picchi, è attribuibile alla presenza di false anomalie, le quali compromettono la capacità dell'algoritmo di identificare le componenti principali dello spazio dei dati. Il caso in esame risulta piuttosto adatto ad essere affrontato tramite strumenti che denotino una certa robustezza statistica. Si è pensato quindi di tentare l'applicazione della RPCA al fine di limitare il contributo del disturbo che rende di difficile distinzione la presenza di anomalie. I passi seguiti sono analoghi a quelli visti nel caso della PCA salvo per l'applicazione del modello che in questo caso non è più incentrato sulla scelta del numero di componenti principali ma sulla determinazione di alcuni parametri che giocano un ruolo chiave nella soluzione del problema di ottimizzazione descritto nel capitolo 3.

In particolare, i parametri su cui si è operato sono λ , cioè il peso attribuito alla componente sparsa e "max_iter", cioè il numero massimo di iterazioni.

Secondo quanto descritto nel lavoro di Candès, Emmanuel J., e altri [12], il valore ottimale di λ è $1/\sqrt{n}$ dove con n si indica il massimo tra il numero di finestre e il numero di campioni per finestra. Questo algoritmo è molto lento a convergere e richiede un numero elevato di istanze per separare la componente sparsa da quella principale. Sicuramente, rispetto alla PCA e alla KPCA, la RPCA denota un costo computazionale molto superiore. È necessario però tenere conto del fatto che l'ottimizzazione fatta sul dataset di train viene svolta tipicamente un numero limitato di volte (in alcuni casi addirittura solo una) e dopo aver ottenuto il modello allenato si proietta il dataset di test su di esso, operazione molto leggera dal punto di vista computazionale. Visti i tempi necessari alla convergenza si è deciso di limitare il numero massimo di iterazioni a 1000. Dopo alcuni tentativi si è riscontrato inoltre che la regola $\lambda = 1/\sqrt{n}$ suggerita dagli autori consente di ottenere risultati soddisfacenti. L'implementazione dell'algoritmo prevede una soglia che distingue automaticamente le componenti principali da quelle residuali, nel caso del dataset privo di picchi il processo di ottimizzazione ha valutato come principali le prime 349 componenti.

5.6.1 RPCA + Mahalanobis

Combinando la RPCA con la distanza di Mahalanobis si ottengono degli ottimi risultati anche in presenza di anomalie di durata e ampiezza minime (anche pari a 1/10 della deviazione standard del dataset). Per potersi rendere conto della maggior efficacia di questo nuovo detector si è scelto, a titolo di esempio, di mantenere i parametri relativi alle anomalie usati per la PCA.

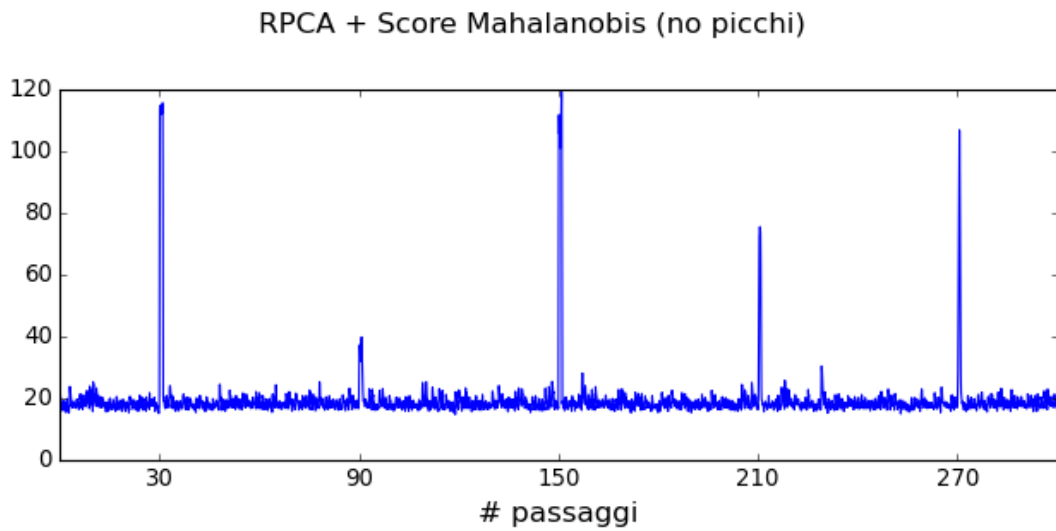


Figura 72 - Distanza di Mahalanobis della proiezione del dataset di test d'esempio dalla proiezione del dataset di train.

Si nota facilmente che il disturbo è fortemente abbattuto rispetto alla PCA e ciò consente di distinguere nettamente la presenza delle cinque anomalie.

5.6.2 RPCA + SPE

Si è implementato in maniera del tutto analoga a quanto fatto in precedenza anche lo score SPE nel caso della RPCA. Il risultato ottenuto è anch'esso molto valido:

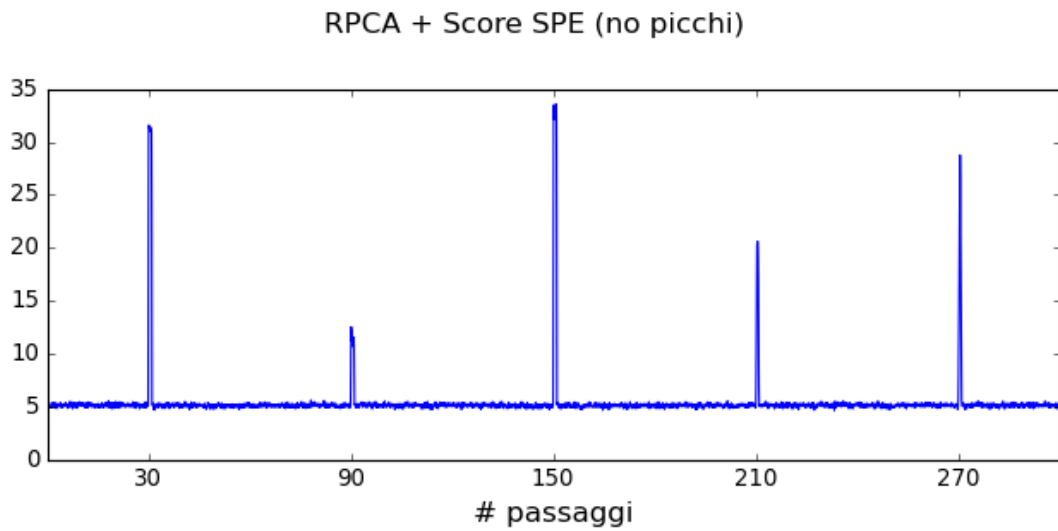


Figura 73 - Score SPE del dataset di test d'esempio.

5.6.2 RPCA + T^2

Infine, si può notare che anche lo score T^2 migliora notevolmente presentando dei risultati confrontabili con quelli degli altri due scores:

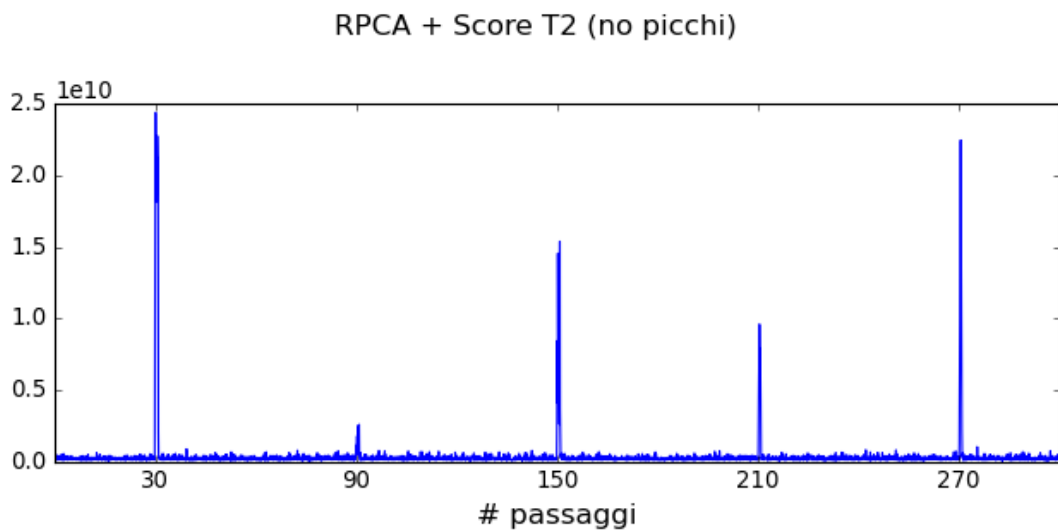


Figura 74 - Score T^2 del dataset di test d'esempio.

5.6.3 ROC

Ad ulteriore conferma di quanto riscontrato con i due esempi precedenti, è stata disegnata la curva del ROC. Grazie alla quasi totale assenza di falsi positivi e all'ottima capacità di rilevare le vere anomalie, tutti e tre i detectors mostrano un AUC molto alto.

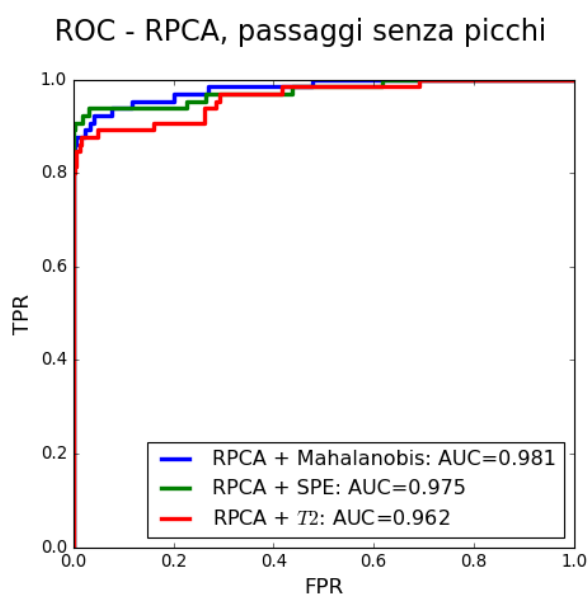


Figura 75 - Curva del ROC per i tre detectors nel caso d'esempio con passaggi senza picchi.

Per generalizzare il risultato ottenuto nel facile caso d'esempio, sono stati generati diversi dataset di test con anomalie sempre meno consistenti. Si è indicato con γ la deviazione standard del dataset di test (3.14 cm) e con ϕ la durata di un passaggio (60 s). Le anomalie iniettate vengono dunque parametrizzate in percentuale rispetto a γ e ϕ .

In questo primo grafico viene mostrato il valore di AUC al calare dell'ampiezza delle anomalie avendo fissato la loro durata a 10% ϕ (6 s):

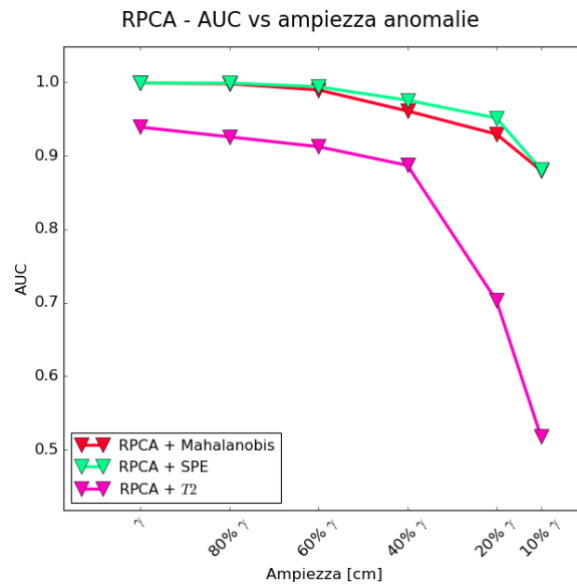


Figura 76 - Andamento dell'AUC al variare dell'ampiezza delle anomalie avendo fissato la loro durata a 60s.

In questo secondo grafico si mostra come varia l'AUC al calare della durata delle anomalie. L'ampiezza degli eventi anomali è pari a 10% γ :

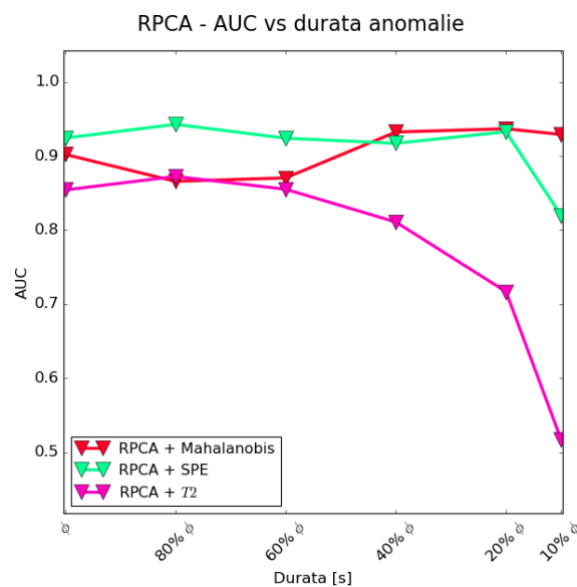


Figura 77 - Andamento dell'AUC al variare della durata delle anomalie avendo fissato la loro ampiezza ad un quarto della deviazione standard del dataset di train.

In generale, gli score SPE e di Mahalanobis sembrano presentare risultati migliori e di preservare la performance anche alla luce di anomalie minime in termini sia di durata che di ampiezza.

Purtroppo, nonostante la RPCA riesca a produrre ottimi risultati in assenza di picchi, provando ad applicarla al dataset completo si ottengono risultati migliori rispetto alla PCA ma comunque non soddisfacenti:

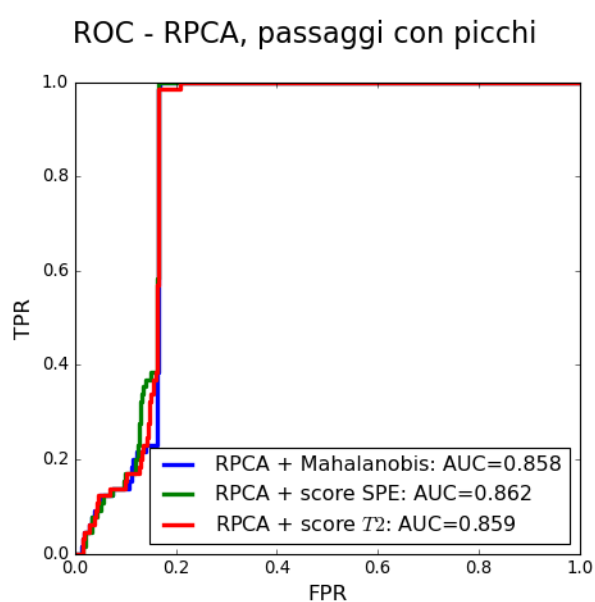


Figura 78 - Curva del ROC riferita ai detectors basati su RPCA nel caso di dataset completi di picco.

5.7 Rilevazione di usura del filo di contatto

Visti gli scarsi risultati ottenuti con i primi detectors è parso inutile descriverne i risultati nel caso della stima di usura del filo. Di seguito vengono quindi presentati i risultati ottenuti direttamente applicando i detectors più efficaci, cioè quelli basati sulla RPCA. Si ipotizzi che ciascun passaggio riduca la massa del cavo di contatto di una certa frazione e che ciò comporti una variazione della frequenza delle armoniche principali che compongono il segnale. Nel tempo ci si aspetta quindi un progressivo

deterioramento dello score che, superata una certa soglia, determina l'imminente manutenzione o sostituzione del filo. Nel caso in esame sono stati generati 2500 passaggi.

Partendo dal modello definito nel capitolo 1 che per comodità viene qui ripreso:

$$f_n = \frac{n}{2L} \sqrt{\frac{T}{\mu}} \text{ con } n = 1, 2, 3, \dots$$

Si è espressa la frequenza in relazione alla variazione normalizzata della densità lineare μ :

$$\begin{aligned} f_n \left(\frac{\Delta\mu}{\mu} \right) &= \frac{n\sqrt{T}}{2L} \sqrt{\frac{1}{\mu - \Delta\mu}} = \\ &= \frac{c}{\sqrt{\mu} \sqrt{1 - \frac{\Delta\mu}{\mu}}} = \\ &= \frac{\frac{c}{\sqrt{\mu}}}{\sqrt{1 - \frac{\Delta\mu}{\mu}}} \end{aligned}$$

Dove $c = \frac{n\sqrt{T}}{2L}$. Così facendo, è possibile associare la variazione di frequenza delle armoniche alla diminuzione percentuale di μ .

Come si può notare, di passaggio in passaggio, tutti e tre gli score tendono a peggiorare. Questo si può meglio osservare applicando un filtro a media mobile con una finestra piuttosto lunga, ad esempio 720 punti dello score, corrispondenti ad 1 ora. Così facendo si valuta complessivamente lo score mediandolo su 60 passaggi, il che è ragionevole essendo l'usura un processo lento e progressivo.

RPCA - Score SPE (no picchi)

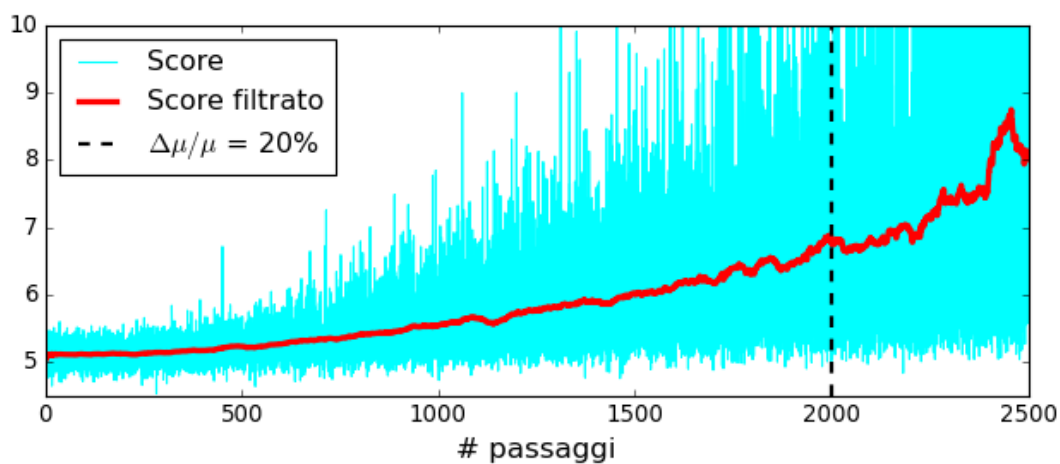


Figura 79 - Deterioramento progressivo dello score SPE a causa della complessiva variazione della frequenza delle armoniche dovuta all'assottigliamento del filo.

Lo score di Mahalanobis rileva in maniera più marcata questo trend progressivo:

RPCA - Score Mahalanobis (no picchi)

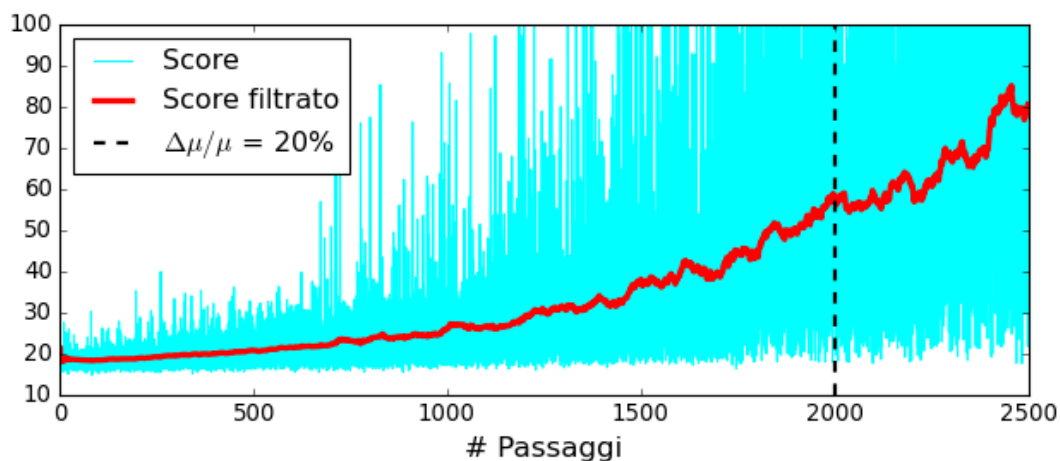


Figura 80 - Deterioramento progressivo dello score di Mahalanobis a causa della complessiva variazione della frequenza delle armoniche dovuta all'assottigliamento del filo.

Anche tramite lo score T^2 si evidenzia la deriva causata dalla modifica della frequenza delle armoniche:

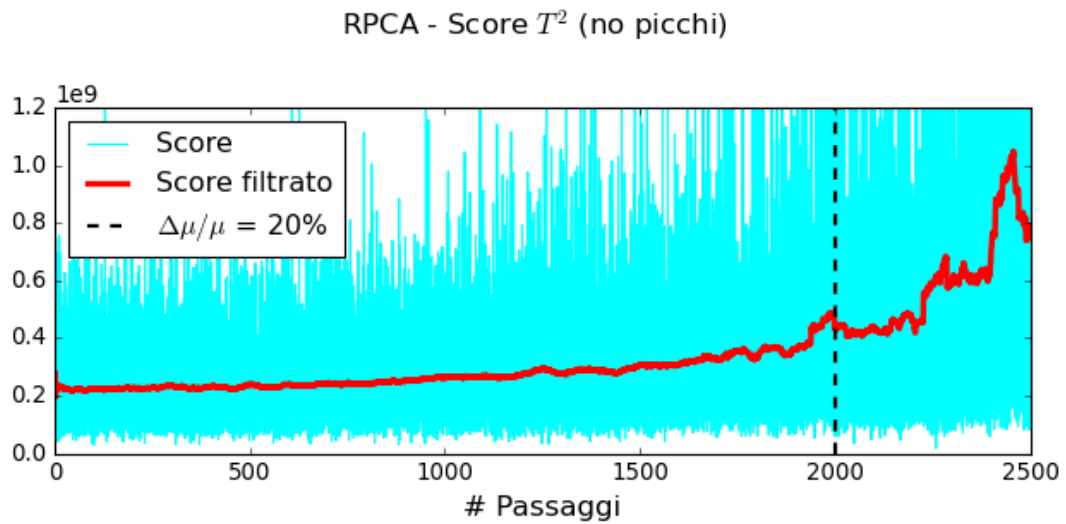


Figura 81 - Deterioramento progressivo dello score T^2 a causa della complessiva variazione della frequenza delle armoniche dovuta all'assottigliamento del filo.

Capitolo 6

Conclusione

A seguito dei risultati ottenuti in questa fase esplorativa è stato possibile trarre delle conclusioni che possono sicuramente essere sfruttate nelle fasi successive dei lavori con RFI. Riassumendo, il flusso operativo di questo elaborato è stato il seguente:

1. Analisi del sistema catenaria e del sistema di misura;
2. Analisi spettrale e T-F dei dataset forniti da RFI al fine di estrarne le caratteristiche utili a generare dei dataset sintetici analoghi a quelli reali;
3. Sintesi di dataset sintetici necessari all'applicazione degli algoritmi di machine learning che tipicamente richiedono una grande quantità di dati;
4. Applicazione di alcuni di detector al fine di rilevare la presenza di anomalie.

A tal proposito sono stati valutati:

- a. detectors basati su PCA;
 - b. detectors basati sulla KPCA;
 - c. detectors basati sulla RPCA;
5. Applicazione della RPCA al fine di rilevare l'effetto del consumo del filo di contatto.

Come è naturale che sia, si è riscontrato che alcuni strumenti non si prestano ad essere applicati con successo nel caso in esame, mentre altri sono in grado di produrre buoni risultati e sono candidati ad essere testati sul campo.

È dunque il momento di condensare quanto visto nei capitoli precedenti, discutendo separatamente la rilevazione di eventi anomali e la rilevazione dell'effetto del consumo della fune di contatto:

Rilevazione di eventi anomali

Alla luce di alcuni tentativi fatti applicando anomalie consistenti, si è subito capito che gli strumenti basati sulla PCA non permettono di raggiungere i risultati sperati. La causa di questo fatto è attribuibile alla non-stazionarietà dei segnali dovuta alla presenza del picco e al decadimento esponenziale che avviene nel post-passaggio. Si è tentato poi di applicare la KPCA sperando che tramite l'applicazione del kernel si potesse in qualche modo limitare gli effetti delle non-linearità. Tuttavia, non si è ottenuto alcun significativo miglioramento e pertanto si è deciso di studiare uno strumento più recente che dal punto di vista teorico prometteva di essere adatto a superare le problematiche viste. È stata dunque applicata la RPCA al fine di limitare il disturbo introdotto dalla presenza del picco e del decadimento esponenziale. In generale, tutti i detectors presentano migliori risultati se ci si limita ad analizzare la fase di post-passaggio, ma nel caso della RPCA questo segna lo spartiacque tra un'efficacia mediocre e un risultato molto promettente. Si è visto infatti che gli score di Mahalanobis, SPE e T^2 se applicati a valle della RPCA, sono in grado di rilevare efficacemente la presenza di anomalie limitando quasi totalmente la segnalazione di falsi positivi.

Consumo del filo di contatto

Dal punto di vista teorico, a parità di tensione del filo, il consumo del cavo produce armoniche a frequenza sempre maggiore. Si è quindi ipotizzato che lo score del detector utilizzato assumesse un trend crescente, sintomo di un comportamento gradualmente sempre più anomalo. Visti i risultati deludenti dei detectors basati su PCA e KPCA, si è deciso di focalizzare questa analisi direttamente sulla RPCA.

Effettivamente, i risultati ottenuti rispecchiano quanto sperato e suggeriscono che, nelle fasi successive a questo lavoro, si possano effettuare misurazioni su un intervallo di tempo prolungato, al fine di stabilire un'associazione tra lo score e l'effettivo deterioramento.

In definitiva, si suggerisce di implementare un algoritmo in grado di rilevare la fine del picco, di selezionare solo la fase di post-passaggio e di applicare la RPCA abbinata ai tre possibili score SPE, T^2 e distanza di Mahalanobis.

Tuttavia, per futuri approfondimenti, si suggerisce inoltre di:

- Consolidare l'efficacia statistica dei detectors, applicandoli a dataset di test in cui l'iniezione delle anomalie avvenga in maniera casuale. Sarebbe utile generare anomalie in quantità, ampiezza, durata e tipologia sempre diverse al fine di poter stabilire una performance media dei detectors.
- Approfondire la determinazione degli iper-parametri utilizzati per la KPCA in modo tale esplorarla in maniera esaustiva ed eventualmente scartarla ove si confermasse avere scarse performances.

Ringraziamenti

Il raggiungimento di un traguardo importante non è mai merito unicamente di un singolo ma è sempre frutto del contributo di più persone. Ho riflettuto a lungo sul percorso inusuale che ho fatto in questi due anni e sono arrivato alla conclusione che nulla è stato fatto in maniera totalmente indipendente e autosufficiente, molti dei contributi rappresentano veri e propri sacrifici, alcuni il limpido rispetto per il mio percorso formativo e altri possono sembrare anche solo piccoli favori ma se ho deciso di citarli è perché sono certo che abbiano facilitato il raggiungimento di questo obiettivo.

Nonostante io abbia frequentato in presenza solo una minima parte delle lezioni, ho avuto modo di conoscere alcuni dei miei compagni di corso. Non potendo presenziare alle lezioni ho capito quanto sia importante, soprattutto nei primi tempi, il contatto con l'avanzamento dei corsi e anche solo poter scambiare dei pareri con i propri compagni. Va a loro la mia gratitudine, in particolare ad Andriy, Francesco e Igor, i quali, penso per amicizia e semplice spirito di collaborazione, mi hanno sempre aggiornato con foto di interi quaderni di appunti e decine di GB di registrazioni. Un aiuto del genere non è scontato ed è stato fondamentale. Questo elaborato è frutto di mesi di lavoro in cui ho imparato tanto e ho capito cosa mi piace fare. Ringrazio Alex, Mauro e Riccardo per aver ritenuto che fossi in grado di portare avanti questo progetto e per avermi trasmesso la loro passione.

La qualità dell'esperienza lavorativa fatta in questi due anni ha ampiamente giustificato la fatica di portare avanti entrambi i percorsi. Fin dai primi tempi è stato chiaro, mi sono ritrovato in un team di ottime persone con le quali ho stretto rapporti di amicizia. Dal mio punto di vista è stato proprio questo l'aspetto che mi ha

permesso di avere ogni sera le energie per studiare. In questi due anni ho anche imparato molto, spesso perché proprio i miei colleghi hanno avuto la pazienza e l'entusiasmo di insegnarmi qualcosa. Sento anche la necessità di aggiungere un ringraziamento in particolare per Mario, Edda e Riccardo per aver sempre rispettato le mie esigenze e avermi permesso di organizzare i miei studi nel migliore dei modi.

Il tempo dedicato al lavoro e allo studio ha senza dubbio occupato gran parte, ed in alcuni momenti totalmente, le mie giornate. Questa situazione mi ha insegnato a valorizzare il poco tempo rimasto cercando di dedicarlo alle persone a me care. Caterina ha condiviso con me questo percorso, da sempre con fiducia e con la convinzione che sarei stato in grado di gestirne le difficoltà. Sicuramente non è stato facile, spesso abbiamo dovuto posticipare le esperienze che desideriamo fare ma i momenti passati insieme sono stati bellissimi e hanno alimentato la consapevolezza dell'importanza del nostro rapporto. Va a lei un sincero ringraziamento per la pazienza e la fiducia nei miei confronti.

Quando ho fatto la scelta di iniziare a lavorare ho anche scelto di non rinunciare ai miei amici nonostante fosse spesso complicato esserci. Abbiamo consolidato delle relazioni che ritengo essere imprescindibili.

Penso che i sacrifici più grandi li abbiano fatti i miei genitori. Innanzitutto, mi hanno guidato dandomi dei consigli che si sono sempre rivelati giusti, aiutandomi a fare quello che volevo nel migliore dei modi. Si sono dati da fare per garantirmi di poter dedicare tutto il tempo a mia disposizione allo studio, a volte anche rinunciando alle proprie comodità e ai propri spazi. I miei familiari e le persone a loro vicine hanno contribuito assicurando un sostegno non indifferente. Vi faccio i miei complimenti, siete il mio riferimento principale.

Elenco delle figure

Figura 1 - Schematizzazione del complesso di sostegni utilizzato tipicamente in Italia.....	10
Figura 2 - Dettaglio in vista laterale del complesso di cavi.	11
Figura 3 - Esempio di pantografo moderno.	12
Figura 4 - Sezione tipica del filo di contatto.	13
Figura 5 - Esempio di sistema sospensivo dei tiranti e del filo di contatto.....	13
Figura 6 - Tabelle di intervento e spessore residuo (i parametri inseriti sono stati indicati parametricamente a titolo d'esempio).....	14
Figura 7 - Esempi di difetti dovuti all'usura.	15
Figura 8 - Spettro d'ampiezza della finestra rettangolare.....	27
Figura 9 - Spettro d'ampiezza della finestra di Blackman.....	28
Figura 10 - Spettro d'ampiezza della finestra di Hamming.....	29
Figura 11 - Discretizzazione del piano tempo-frequenza della STFT.	35
Figura 12 - Discretizzazione del piano tempo-scala della CWT.....	38
Figura 13 - Esempio di Wavelet di Morlet.....	39
Figura 14 - Esempio di Wavelet di Shannon.	40
Figura 15 - Raffigurazione nello spazio tridimensionale delle componenti principali di una distribuzione.	42
Figura 16 - Andamento della varianza espressa da ciascuna componente. Il "gomito" che divide i due andamenti è stato evidenziato con un punto rosso.....	44
Figura 17 - Esempio di varietà tri-dimensionale e proiezione su piani.....	45
Figura 18 - Schematizzazione grafica dell'algoritmo della KPCA.	47
Figura 19 - Esempio di regressione effettuata tramite norma L_2 in presenza di outlier.	49
Figura 20 - Esempio di regressione tramite norma L_1 in presenza di outlier.	50
Figura 21 - Schematizzazione grafica della distanza di Mahalanobis.	54
Figura 22 - Tabella delle possibili casistiche nell'ambito della rilevazione di anomalie.	55
Figura 23 - Rappresentazione delle distribuzioni dei risultati positivi e negativi.	56
Figura 24 - Esempio di una possibile curva del ROC e della sua area sottesa (AUC).	57
Figura 25 - Andamento nel tempo di un passaggio avvenuto nel 2017.....	59
Figura 26 – Schema a blocchi del procedimento seguito per la generazione dei dataset sintetici.	60
Figura 27 - Tabella riassuntiva dei parametri di input del generatore.....	61
Figura 28 - Esempio di serie sintetica.	62
Figura 29 - Esempio di dataset sintetico composto da 3 passaggi, due alla velocità di 160 km/h e uno a 180 km/h.....	62

Figura 30 – Esempio di picco sintetico.	63
Figura 31 – Stima di spettro di potenza dei quattro passaggi del 2017.....	64
Figura 32 - Esempio dei parametri estratti dai dati reali utili a modellare la variabilità in frequenza ed ampiezza delle varie armoniche.	65
Figura 33 - Passaggio sintetico completo formato da picco e oscillazione.....	65
Figura 34 - Esempio di passaggio sintetico.....	66
Figura 35 - Impulsi rettangolari di varia ampiezza e durata.	67
Figura 36 - Impulsi triangolari di varia ampiezza e durata.	68
Figura 37 - Jammer di varia ampiezza, durata e banda.	69
Figura 38 - Anomalie periodiche di varia ampiezza, durata e frequenza.....	70
Figura 39 - AWGN di varia ampiezza e durata.....	71
Figura 40 - Incremento in frequenza dei modi fondamentali a seguito del consumo del cavo.	72
Figura 41 - STFT di un passaggio del 2017.....	75
Figura 42 - Applicazione della CWT ad un passaggio del 2017.....	76
Figura 43 - Selezione del post-passaggio.....	77
Figura 44 – Stima di spettro di potenza di un passaggio del 2017 tramite periodogramma.	78
Figura 45 – Stima di spettro di potenza di un passaggio del 2017 tramite periodogramma modificato.	79
Figura 46 - Spettro di potenza di un passaggio del 2017 stimato con stimatore a minima varianza in scala logaritmica.....	80
Figura 47 - Confronto tra i quattro passaggi del 2017. Dettaglio dell'intervallo di frequenze [0, 10] Hz in scala lineare.....	81
Figura 48 - Focus sull'intervallo [1.5, 10] Hz.	81
Figura 49 - Vettori contenenti le frequenze e le ampiezze delle armoniche estratte dai dati reali.....	82
Figura 50 - Parametri delle gaussiane che modellano la variabilità in frequenza ed ampiezza delle varie armoniche.	83
Figura 51 - Esempio di Dataset di train con focus su un paio di passaggi.....	84
Figura 52 - Identificazione degli istanti temporali in cui sono presenti le anomalie utile a calcolare il ROC.	85
Figura 53 - Esempio di Dataset di test con focus su un paio di passaggi anomali....	86
Figura 54 - Profilo degli autovalori e della varianza espressa da ciascuno di essi in scala logaritmica.....	87
Figura 55 - Score T_2 del dataset di test d'esempio.....	89
Figura 56 - Dettaglio sull'anomalia presente al trentesimo passaggio, cioè l'impulso rettangolare.....	89
Figura 57 - Score SPE del dataset di test d'esempio.	90
Figura 58 - Dettaglio del passaggio numero 90 in cui è presente il jammer.....	91
Figura 59 - Distanza di Mahalanobis della proiezione del dataset di test d'esempio dalla proiezione del dataset di train.....	92
Figura 60 - Dettaglio del passaggio numero 150 in cui è presente l'AWGN.....	92
Figura 61 - Curva del ROC dei tre detectors nel caso d'esempio.....	93

Figura 62 - Curva del ROC per i tre detectors nel caso d'esempio con passaggi senza picchi.	94
Figura 63 - Score T_2 nel caso di dataset senza picchi.	95
Figura 64 - Score di Mahalanobis nel caso di dataset senza picchi.	95
Figura 65 - Score T_2 del dataset di test d'esempio.	97
Figura 66 - Score SPE del dataset di test d'esempio.	98
Figura 67 - Score di Mahalanobis del dataset di test d'esempio.	99
Figura 68 - Curva del ROC associata ai tre detectors basati su KPCA.	100
Figura 69 - Distanza di Mahalanobis della proiezione del dataset di test d'esempio dalla proiezione del dataset di train.	102
Figura 70 - Score SPE del dataset di test d'esempio.	103
Figura 71 - Score T_2 del dataset di test d'esempio.	103
Figura 72 - Curva del ROC per i tre detectors nel caso d'esempio con passaggi senza picchi.	104
Figura 73 - Andamento dell'AUC al variare dell'ampiezza delle anomalie avendo fissato la loro durata a 60s.	105
Figura 74 - Andamento dell'AUC al variare della durata delle anomalie avendo fissato la loro ampiezza ad un quarto della deviazione standard del dataset di train.	105
Figura 75 - Curva del ROC riferita ai detectors basati su RPCA nel caso di dataset completi di picco.	106
Figura 76 - Deterioramento progressivo dello score SPE a causa della complessiva variazione della frequenza delle armoniche dovuta all'assottigliamento del filo.	108
Figura 77 - Deterioramento progressivo dello score di Mahalanobis a causa della complessiva variazione della frequenza delle armoniche dovuta all'assottigliamento del filo.	108
Figura 78 - Deterioramento progressivo dello score T_2 a causa della complessiva variazione della frequenza delle armoniche dovuta all'assottigliamento del filo.	109

Bibliografia

- [1] <https://www.rfi.it/it/Sicurezza-e-tecnologie/tecnologie/energia/energia-distribuzione.html>
- [2] Nåvik, Petter, Anders Rønquist, and Sebastian Stichel. "A wireless railway catenary structural monitoring system: Full-scale case study." *Case Studies in Structural Engineering* 6 (2016): 22-30.
- [3] Van, Olivier Vo, Jean-Pierre Massat, and Etienne Balmes. "Waves, modes and properties with a major impact on dynamic pantograph-catenary interaction." *Journal of Sound and Vibration* 402 (2017): 51-69.
- [4] https://it.wikipedia.org/wiki/Linea_aerea_di_contatto
- [5] Policicchio, Franco. *Lineamenti di infrastrutture ferroviarie*. Vol. 2. Firenze University Press, 2007.
- [6] *Elementi di Teoria Statistica dei Segnali, modelli e metodi per l'Ingegneria dell'Informazione*, Riccardo Rovatti May 30, 2005
- [7] Rioul, Olivier, and Martin Vetterli. "Wavelets and signal processing." *IEEE signal processing magazine* 8.4 (1991): 14-38.
- [8] Corso di Analisi Spettrale tenuto dal Prof. Niccolò Attilio Speciale.
- [9] Aggarwal, Charu C. "Outlier analysis." *Data mining*. Springer, Cham, 2015.

[10] Mnassri, Baligh, Bouchra Ananou, and Mustapha Ouladsine. "Fault detection and diagnosis based on PCA and a new contribution plot." IFAC Proceedings Volumes 42.8 (2009): 834-839.

[11] Alam, Md Ashad, and Kenji Fukumizu. "Hyperparameter selection in kernel principal component analysis." (2014).

[12] Candès, Emmanuel J., et al. "Robust principal component analysis?" Journal of the ACM (JACM) 58.3 (2011): 1-37.

[13] <https://youtu.be/yDpz0PqULXQ>

[14] Focardi, Sergio, et al. Fisica generale. Ambrosiana, 2007.