

UNIVERSITA' POLITECNICA DELLE MARCHE

FACOLTÀ DI INGEGNERIA

Corso di Laurea triennale in Ingegneria Biomedica

Analisi della complessità delle serie temporali associate al cammino in pazienti con malattia di Parkinson

Analysis of the complexity of gait-related time series in patients with Parkinson's disease

Relatore:

Prof. Andrea Tigrini

Tesi di Laurea di:

Giacomo Cecconi

Correlatore:

Prof. Alessandro Mengarelli

A.A. 2025/2026

Indice

1 – Introduzione e stato dell'arte 1
1.1 La malattia di Parkinson 1
1.2 Alterazioni del cammino nella malattia di Parkinson 1
1.3 Variabilità del passo 2
1.4 Analisi frattale del cammino 3
1.5 *Detrended Fluctuation Analysis* (DFA) 5
1.6 *Extended Detrended Fluctuation Analysis* (EDFA) 6
1.7 *Fluctuation Analysis* (FA) 7
1.8 Correlazioni nel cammino 8
1.9 Obiettivo della tesi 9

2 – Materiali e metodi 10
2.1 Database 10
2.2 Estrazione della *stance phase* 12
2.3 *Pre-processing* delle serie temporali 15
2.3.1 Selezione del tratto valido 15
2.3.2 Trattamento dei valori mancanti e rimozione degli *outlier* 19
2.4 Esclusione delle registrazioni problematiche 20
2.5 Calcolo dei parametri 22
2.6 Analisi statistiche 23
2.6.1 Livello di significatività 23
2.6.2 Test di Mann–Whitney 23
2.6.3 Test ANOVA 24
2.6.4 Correlazione di Spearman 24

3 – Risultati 26
3.1 Risultati della varianza 26
3.2 Risultati della DFA 28
3.3 Risultati dell'EDFA 30
3.4 Risultati della FA 32

4 – Discussione 36
4.1 Discussione dei risultati 36
4.2 Limiti dello studio 40
4.3 Sviluppi futuri 44

5 – Conclusioni 45

6 – Bibliografia 47

Appendice – Codice MATLAB 49

1 – Introduzione e stato dell’arte

1.1 La malattia di Parkinson

La malattia di Parkinson è una patologia neurologica cronica e progressiva e rappresenta uno dei disturbi del movimento più comuni. È caratterizzata principalmente dalla progressiva degenerazione dei neuroni dopaminergici della *substantia nigra pars compacta* (SNpc), con conseguente riduzione della disponibilità di dopamina nei gangli della base. Tale alterazione è alla base dei principali disturbi motori associati alla malattia, tra cui tremore, rigidità, bradicinesia e instabilità posturale. La malattia di Parkinson presenta tuttavia una natura complessa e multisistemica, coinvolgendo anche regioni del sistema nervoso e sistemi di neurotrasmissione differenti da quelli dopaminergici. Alla componente motoria si affianca infatti un'ampia varietà di manifestazioni non motorie, alcune delle quali possono precedere di molti anni la comparsa dei disturbi motori [1].

Tra le manifestazioni motorie della malattia, le alterazioni dell'andatura rappresentano un problema particolarmente rilevante. Un'andatura patologica è infatti un sintomo comune e invalidante e, nei casi più gravi, può aumentare il rischio di cadute e compromettere l'autonomia funzionale del paziente. L'analisi del cammino assume pertanto un ruolo importante nella caratterizzazione delle alterazioni motorie associate alla malattia di Parkinson [2].

1.2 Alterazioni del cammino nella malattia di Parkinson

La malattia di Parkinson è associata a numerose alterazioni della locomozione, che possono compromettere significativamente l'efficienza e la sicurezza del cammino. Tra le principali alterazioni osservate nei pazienti rientrano la riduzione della velocità e della lunghezza del passo, modificazioni della cadenza e una maggiore asimmetria tra gli arti inferiori. La deambulazione è tipicamente rigida e acinetica, caratterizzata da una ridotta oscillazione delle braccia, che negli stadi più avanzati della malattia vengono mantenute in posizione addotta e flessa. Nei pazienti è inoltre frequente un'andatura caratterizzata da passi brevi e da una ridotta altezza di sollevamento dei piedi dal suolo [3].

Oltre alle modificazioni dei parametri spazio-temporali, il cammino parkinsoniano può essere caratterizzato da una maggiore irregolarità della sequenza dei passi. In particolare, le alterazioni del controllo motorio possono determinare una maggiore variabilità dei parametri temporali del cammino, rendendo la successione dei passi meno regolare rispetto a quella osservata nei soggetti sani [3].

La caratterizzazione del cammino nei pazienti con malattia di Parkinson non dovrebbe pertanto limitarsi alla valutazione dei valori medi dei parametri spazio-temporali, ma dovrebbe considerare anche la loro variabilità nel tempo. Quest'ultima può infatti fornire informazioni aggiuntive sulle modalità con cui il sistema motorio controlla e regola la locomozione, costituendo un possibile elemento utile per caratterizzare le alterazioni motorie associate alla malattia di Parkinson.

1.3 Variabilità del passo

Il ciclo del passo è costituito da due fasi principali: la *stance phase*, durante la quale il piede rimane a contatto con il terreno, e la *swing phase*, durante la quale il piede è sollevato e avanza verso il passo successivo. In condizioni di cammino fisiologico, la *stance phase* rappresenta circa il 60% del ciclo del passo, mentre la *swing phase* ne costituisce il restante 40%. Inoltre, all'inizio e alla fine della fase di appoggio sono presenti brevi periodi di doppio supporto, nei quali entrambi i piedi sono contemporaneamente a contatto con il terreno. Nella Figura 1.1 [3] è riassunta la sequenza delle fasi del ciclo del passo.

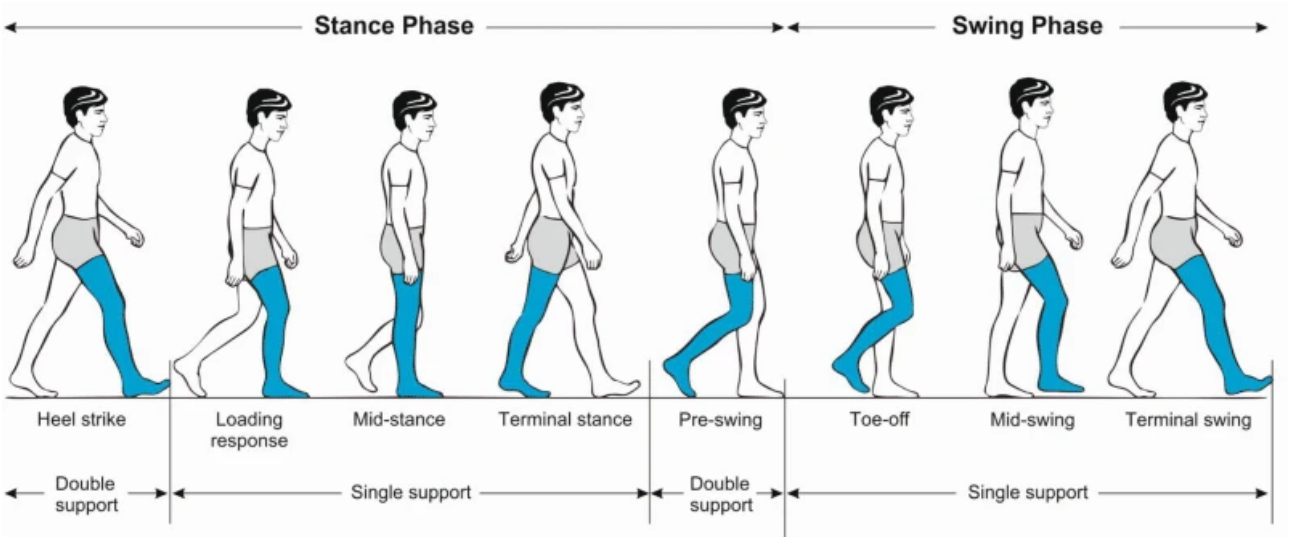


Figura 1.1 – Suddivisione del ciclo del passo nelle fasi di *stance* e *swing* [3].

Rispetto ai soggetti sani, nei pazienti affetti da malattia di Parkinson si osserva generalmente un aumento della variabilità dell'andatura, intesa come una maggiore dispersione dei valori di un determinato parametro da un passo al successivo. In condizioni fisiologiche, infatti, grandezze quali la lunghezza del passo o la durata della fase di appoggio (*stance phase*) tendono a mantenersi relativamente simili tra passi consecutivi. Nei pazienti con malattia di Parkinson, invece, possono verificarsi variazioni più ampie tra i singoli passi, indicando una maggiore irregolarità della sequenza temporale del cammino.

L'aumento della variabilità del cammino nei pazienti con malattia di Parkinson è stato documentato in diversi studi. Hausdorff et al. hanno analizzato le variazioni passo-passo della temporizzazione del ciclo del cammino, evidenziando nei pazienti con Parkinson livelli di variabilità significativamente superiori rispetto ai soggetti sani. In particolare, le misure di variabilità risultavano circa il doppio di quelle osservate nei controlli e mostravano inoltre una relazione con la gravità della malattia. È interessante osservare come tali differenze nella variabilità fossero presenti nonostante la durata media del ciclo del passo e di diverse sue sottofasi non differisse significativamente tra i gruppi, suggerendo che la variabilità possa fornire informazioni complementari rispetto ai soli valori medi dei parametri del cammino [4].

Tali evidenze sono state successivamente confermate da analisi più ampie della letteratura. Una metanalisi condotta da König et al. su 34 studi, comprendenti complessivamente 800 pazienti con malattia di Parkinson e 854 soggetti sani, ha evidenziato un aumento della variabilità in sei degli otto parametri del cammino considerati. Gli autori hanno inoltre osservato che l'entità dell'alterazione non è uniforme tra i diversi parametri, suggerendo che la variabilità debba essere valutata considerando specificamente la grandezza analizzata [5].

1.4 Analisi frattale del cammino

I metodi lineari di analisi dell'andatura, basati su modelli biomeccanici tradizionali, forniscono informazioni utili sui principali parametri del cammino, ma non consentono di caratterizzare completamente la struttura temporale della variabilità dell'andatura. Per questo motivo, negli ultimi decenni sono stati proposti metodi di analisi non lineare finalizzati a studiare l'organizzazione temporale della variabilità del cammino [6].

Tra questi, un ruolo rilevante è assunto dai metodi di analisi frattale, che consentono di caratterizzare le proprietà di correlazione e la struttura temporale delle serie costituite dai parametri del cammino. In particolare, l'analisi frattale può essere utilizzata per valutare la presenza di correlazioni a lungo termine nelle serie temporali degli intervalli di passo, fornendo informazioni complementari rispetto alle tradizionali misure della variabilità [6].
Molti segnali biomedici presentano una struttura invariante di scala, ovvero una struttura che si ripete a differenti scale di osservazione. Formalmente, un segnale $x(t)$ è invariante di scala se

$$x(ct) = c^H x(t)$$

dove H rappresenta l'esponente della legge di potenza associata alla struttura invariante di scala. Tale esponente può essere stimato mediante le analisi frattali e consente di caratterizzare le proprietà di scala del segnale biomedico considerato [7].

Le strutture invarianti di scala possono essere classificate in monofrattali e multifrattali. Una struttura monofrattale è caratterizzata da un singolo esponente della legge di potenza e presenta proprietà di invarianza di scala che non variano nel tempo e nello spazio. Al contrario, una struttura multifrattale è caratterizzata da variazioni della struttura invariante di scala nel tempo e nello spazio e, di conseguenza, non può essere descritta mediante un unico esponente. Essa viene pertanto rappresentata attraverso uno spettro multifrattale costituito da diversi esponenti della legge di potenza [7].

La Figura 1.2 mostra alcuni esempi di serie temporali caratterizzate da differenti proprietà di scala. In blu sono rappresentate, dall'alto verso il basso, una serie multifrattale, una serie monofrattale e un rumore bianco, mentre in rosso sono riportate le corrispondenti serie ottenute mediante integrazione. Queste ultime presentano un andamento assimilabile a un cammino casuale (*random walk*) e mostrano una struttura che si ripete a differenti scale di osservazione. In particolare, la serie multifrattale si distingue da quella monofrattale per la presenza di intervalli temporali caratterizzati da fluttuazioni di diversa ampiezza, riconducibili a variazioni locali delle proprietà di scala [7].

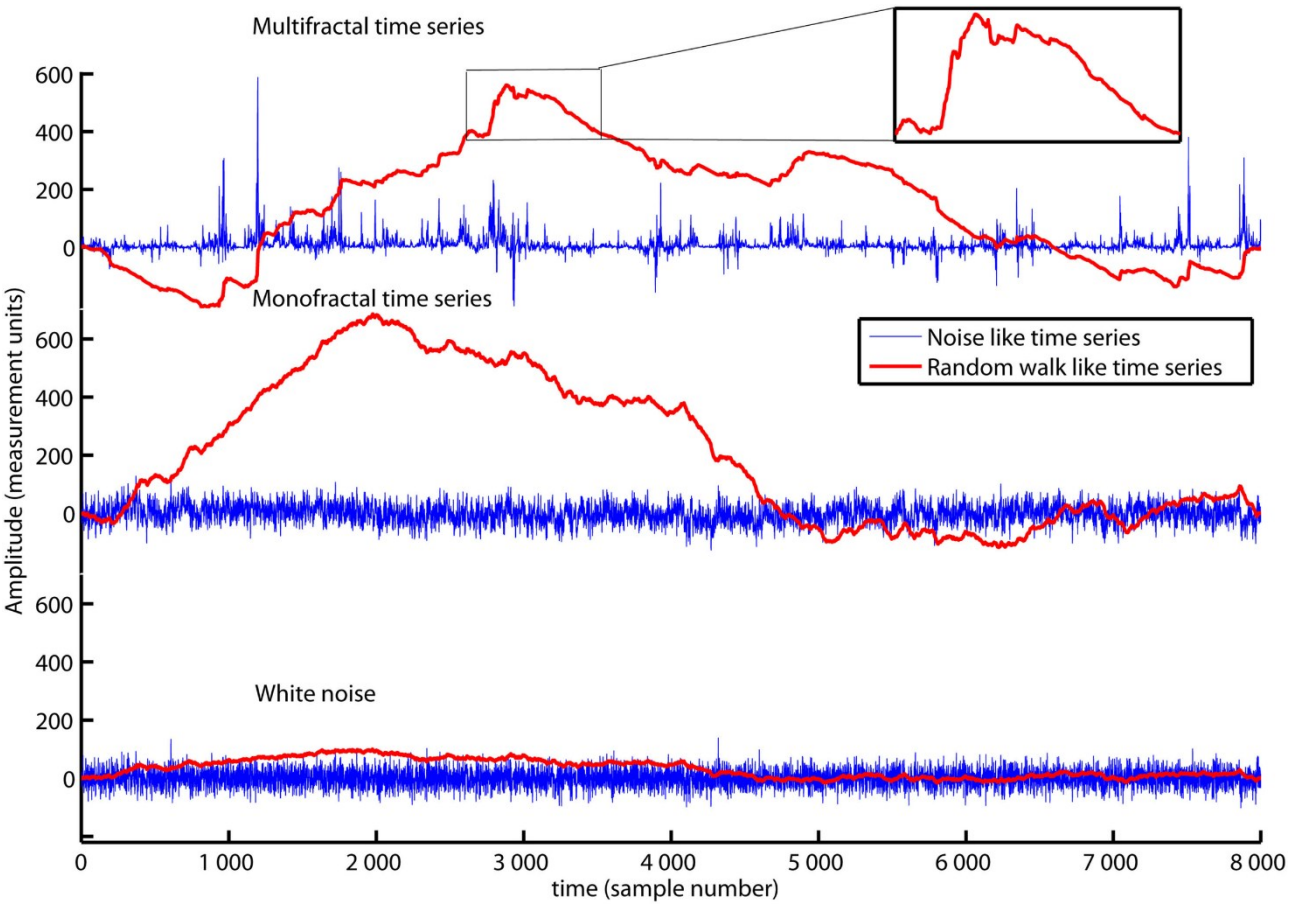


Figura 1.2 – Esempi di serie temporali: multifrattale (in alto), monofrattale (al centro) e rumore bianco (in basso). Le tracce blu rappresentano serie con andamento *noise-like*, mentre quelle rosse rappresentano le corrispondenti serie con andamento *random-walk-like* [7].

1.5 *Detrended Fluctuation Analysis* (DFA)

La *Detrended Fluctuation Analysis* (DFA) è un metodo di analisi frattale utilizzato per quantificare le proprietà di correlazione presenti in serie temporali, introducendo una procedura di *detrending* finalizzata a ridurre l'influenza di eventuali trend. In particolare, la DFA consente di valutare la presenza di correlazioni a lungo termine nelle fluttuazioni del segnale attraverso la stima di un esponente di scala, indicato con α. Tale parametro descrive il modo in cui l'ampiezza delle fluttuazioni varia al variare della scala temporale considerata [6].

Nell'analisi del cammino, la DFA viene applicata alle serie temporali dei parametri dell'andatura per caratterizzarne l'organizzazione temporale e valutare se le variazioni osservate da un passo al successivo presentino una struttura correlata oppure un comportamento maggiormente casuale. Essa fornisce pertanto informazioni complementari rispetto alle tradizionali misure lineari di variabilità, che descrivono principalmente l'entità delle fluttuazioni senza considerarne l'organizzazione nel tempo.

L'applicazione della DFA prevede una successione di passaggi finalizzati alla valutazione delle fluttuazioni della serie a differenti scale temporali [6].

1. Inizialmente, a partire dalla serie temporale $x(t)$, viene calcolato il profilo cumulativo mediante sottrazione del valore medio e successiva integrazione della serie. Si ottiene quindi la somma cumulativa $y(j)$:

$$y(j) = \sum_{t=1}^{j} [x(t) - \bar{x}], \qquad j = 1, \dots, N,$$

   dove $\bar{x}$ rappresenta il valore medio della serie temporale e N il numero totale di campioni che la compongono.
2. La serie temporale integrata $y(j)$ viene suddivisa in L finestre temporali non sovrapposte di lunghezza n.
3. Successivamente, in ciascuna delle L finestre viene applicato alla porzione corrispondente della serie $y(j)$ un adattamento polinomiale ai minimi quadrati di ordine k. La tendenza locale così stimata è indicata con $y_n(j)$ e rappresenta l'andamento della serie all'interno della relativa finestra.
4. La fluttuazione RMS (*Root Mean Square*) della serie integrata e detrendizzata viene quindi calcolata come segue:

$$F(n) = \sqrt{\frac{1}{N}\sum_{j=1}^{N}[y(j) - y_n(j)]^2}$$

5. I passaggi 2-4 vengono ripetuti per differenti valori del parametro n.

6. Se la serie presenta proprietà di invarianza di scala, la funzione di fluttuazione $F(n)$ segue una relazione di legge di potenza rispetto alla dimensione della finestra n:

$$F(n) \sim n^{\alpha}$$

dove α rappresenta l'esponente di scala. Applicando il logaritmo alla relazione, si ottiene una dipendenza lineare tra $\log F(n)$ e $\log n$. L'esponente α viene pertanto determinato come la pendenza della retta di regressione lineare ottenuta nel grafico log-log di $F(n)$ in funzione di n.

Il valore dell'esponente di scala α consente di caratterizzare la struttura delle correlazioni presenti nella serie temporale. Valori compresi tra 0 e 0,5 indicano un comportamento anti-persistente, mentre $\alpha = 0{,}5$ corrisponde all'assenza di correlazioni a lungo termine, caratteristica del rumore bianco. Valori compresi tra 0,5 e 1 indicano invece un comportamento persistente, per cui una variazione in una determinata direzione tende a essere seguita da variazioni nella stessa direzione. Per $\alpha \simeq 1$ si osserva un comportamento riconducibile al rumore $1/f$, mentre $\alpha \simeq 1{,}5$ è caratteristico del rumore browniano [6].

È inoltre importante considerare che la stima dell'esponente α può essere influenzata dalla lunghezza della serie temporale e dalla scelta dei parametri dell'algoritmo, quali l'intervallo delle dimensioni delle finestre e l'ordine del polinomio utilizzato per il *detrending* [6].

1.6 *Extended Detrended Fluctuation Analysis* (EDFA)

L'*Extended Detrended Fluctuation Analysis* (EDFA) rappresenta un'estensione della DFA convenzionale, proposta per migliorare la caratterizzazione di serie temporali non omogenee, nelle quali le proprietà di non stazionarietà possono variare tra differenti porzioni del segnale. Nella DFA convenzionale, la funzione di fluttuazione $F(n)$ fornisce una misura complessiva delle fluttuazioni della serie a una determinata scala temporale. L'EDFA introduce quindi un'ulteriore misura finalizzata a descrivere la variabilità delle fluttuazioni locali tra i diversi segmenti della serie [8].

Per ogni valore della dimensione della finestra n, vengono considerate le fluttuazioni RMS locali $F_{\mathrm{loc}}(n)$ calcolate nei diversi segmenti della serie. L'eterogeneità delle fluttuazioni locali viene quantificata attraverso la differenza tra il valore massimo e il valore minimo:

$$dF(n) = \max[F_{\mathrm{loc}}(n)] - \min[F_{\mathrm{loc}}(n)].$$

Se $dF(n)$ presenta una dipendenza di tipo legge di potenza dalla scala n, vale la relazione

$$dF(n) \sim n^{\beta}$$

dove β rappresenta un ulteriore esponente di scala introdotto dall'EDFA. Tale parametro caratterizza il modo in cui l'eterogeneità delle fluttuazioni locali varia al variare della scala temporale, fornendo informazioni sul grado di non omogeneità della serie.

L'EDFA permette quindi di ottenere due esponenti di scala complementari: α, già ricavato mediante la DFA convenzionale e associato alle proprietà delle correlazioni a lungo termine, e β, introdotto per caratterizzare l'eterogeneità delle fluttuazioni locali. L'utilizzo congiunto dei due parametri consente pertanto una caratterizzazione più completa delle serie temporali non omogenee rispetto a quella ottenibile mediante il solo esponente α [8].

1.7 *Fluctuation Analysis* (FA)

La *Fluctuation Analysis* (FA) è un metodo di analisi utilizzato per caratterizzare le proprietà di scala di una serie temporale attraverso lo studio della *diffusion of random walk process* ad essa associato. Analogamente alla DFA, la FA analizza il comportamento della serie a differenti scale; a differenza della DFA, tuttavia, non prevede l'adattamento locale mediante regressione polinomiale né la successiva rimozione del trend [9].

A partire dalla serie temporale viene costruito il corrispondente *random walk process*, le cui posizioni sono indicate con y_i. Considerando punti del cammino separati da un intervallo Δn, la dispersione viene valutata attraverso le differenze tra le rispettive posizioni. Essa può essere quantificata mediante la radice della funzione di struttura di secondo ordine S_2:

$$\sqrt{S_2(\Delta n)} = \sqrt{\langle (y_{i+\Delta n} - y_i)^2 \rangle}$$

dove le parentesi angolari indicano l'operazione di media, y_i rappresenta la posizione i-esima del *random walk process* e Δn la separazione, espressa in numero di campioni, tra i punti considerati. La grandezza così ottenuta descrive la dispersione del cammino al variare della scala di osservazione [9].

In presenza di un comportamento invariante di scala, la dispersione del cammino presenta una dipendenza di tipo legge di potenza dalla scala Δn:

$$\sqrt{S_2(\Delta n)} \sim (\Delta n)^H$$

dove H rappresenta l'esponente di Hurst. Di conseguenza, rappresentando $\sqrt{S_2(\Delta n)}$ in funzione di Δn su scala log-log, si ottiene un andamento lineare e H può essere determinato dalla pendenza della retta. L'esponente di Hurst consente di caratterizzare le proprietà di correlazione della serie: $H = 0{,}5$ corrisponde all'assenza di correlazione, valori di $H < 0{,}5$ indicano un comportamento anticorrelato, mentre valori di $H > 0{,}5$ indicano la presenza di correlazioni positive [9].

Poiché la FA non comprende una procedura di *detrending*, risulta adatta alla caratterizzazione di segnali stazionari e non consente di compensare gli effetti dovuti alla presenza di non stazionarietà. Rispetto alla DFA, inoltre, non richiede un adattamento polinomiale locale né la successiva rimozione del trend e considera solamente le differenze tra gli estremi degli intervalli analizzati, risultando pertanto computazionalmente più semplice [9].

1.8 Correlazioni nel cammino

Con il termine *stride* si indica un ciclo completo del cammino, compreso tra due contatti iniziali consecutivi dello stesso piede, e costituito quindi da due passi successivi. La variabilità del cammino non dipende solamente dall'entità delle fluttuazioni osservate tra passi consecutivi, ma anche dalla loro organizzazione temporale. In condizioni fisiologiche, le serie temporali degli intervalli di *stride* presentano generalmente correlazioni a lungo termine e un comportamento persistente. Nell'ambito della DFA, tale comportamento è caratterizzato da valori dell'esponente di scala α superiori a 0,5, mentre valori progressivamente più prossimi a 0,5 indicano un'organizzazione temporale maggiormente casuale [6].

L'importanza di considerare l'organizzazione temporale della variabilità, oltre alle tradizionali misure lineari, è stata evidenziata anche da Yano et al. In uno studio condotto su soggetti giovani sani, gli autori hanno analizzato gli intervalli di *stride* durante il cammino in tre differenti condizioni: controllo, svolgimento di un compito non cognitivo e svolgimento di un compito cognitivo mediante smartphone. La media e la deviazione standard degli intervalli di *stride* non hanno mostrato differenze statisticamente significative tra le condizioni, mentre l'esponente di scala α risultava significativamente ridotto durante il compito cognitivo. Tali risultati mostrano come l'analisi dell'organizzazione temporale delle fluttuazioni possa evidenziare modificazioni non riscontrabili attraverso le sole misure lineari della variabilità [10].

La Figura 1.3 riporta il confronto tra la media, la deviazione standard e l'esponente di scala α degli intervalli di *stride* nelle tre condizioni sperimentali considerate da Yano et al. [10].

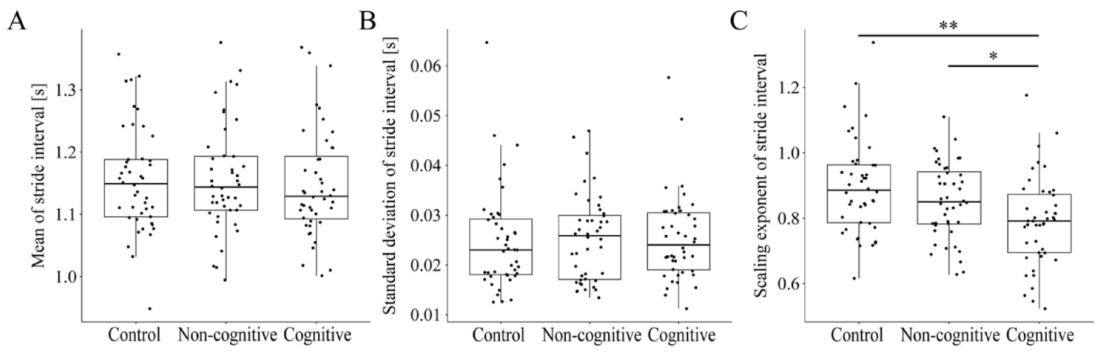


Figura 1.3 – Confronto tra media (A), deviazione standard (B) ed esponente di scala α ottenuto mediante DFA (C) degli intervalli di *stride* nelle condizioni di controllo, compito non cognitivo e compito cognitivo. Le differenze statisticamente significative tra le condizioni sono indicate mediante asterischi [10].

Hausdorff ha evidenziato come, nella malattia di Parkinson, le proprietà di scala frattale del cammino risultino alterate rispetto a quelle osservate nei soggetti sani e come tali proprietà descrivano un aspetto della dinamica dell'andatura distinto sia dalla lunghezza del passo sia dall'entità della variabilità. Nel cammino fisiologico, l'esponente di scala assume generalmente valori compresi tra circa 0,8 e 1, mentre nella malattia di Parkinson si osserva una tendenza verso valori più prossimi a 0,5, indicativa di una dinamica meno persistente e maggiormente casuale [11].

Warlop et al. hanno inoltre osservato che, nei pazienti con malattia di Parkinson, il deterioramento delle correlazioni a lungo termine della durata dello *stride* è associato a una maggiore compromissione motoria. Nei pazienti maggiormente compromessi, l'organizzazione temporale della variabilità del cammino tendeva infatti a presentare un comportamento più casuale [12].

Nel complesso, tali evidenze supportano l'impiego di metodi di analisi frattale per caratterizzare le alterazioni dell'organizzazione temporale del cammino associate alla malattia di Parkinson.

1.9 Obiettivo della tesi

Alla luce delle alterazioni della variabilità e dell'organizzazione temporale del cammino documentate nella malattia di Parkinson, l'analisi delle serie temporali relative alle durate della *stance phase* può fornire informazioni utili per caratterizzare tali modificazioni.

Nel presente lavoro sono stati presi in considerazione i dati relativi a soggetti sani e pazienti con malattia di Parkinson raccolti in un database. A partire dai dati grezzi è stato possibile procedere all'estrazione delle *stance* e, successivamente, al calcolo di diversi parametri basati sulle relative serie temporali. Le analisi statistiche hanno quindi permesso di valutare quali parametri fossero in grado di evidenziare differenze statisticamente significative non solo tra i soggetti patologici e il gruppo dei controlli, ma anche al variare del livello di gravità della malattia.

L'obiettivo della tesi è valutare se i parametri estratti dalle serie temporali della *stance phase* siano in grado di evidenziare differenze tra soggetti sani e pazienti affetti da malattia di Parkinson e di fornire informazioni associate alla gravità della patologia, valutata mediante la scala di Hoehn–Yahr, attraverso l'analisi delle differenze tra i diversi livelli di gravità e delle eventuali correlazioni con il punteggio della scala.

2 – Materiali e metodi

2.1 Database

Il database utilizzato in questo lavoro è il *Gait in Parkinson's Disease Database*, disponibile sulla piattaforma *PhysioNet* [2, 13, 14]. Esso raccoglie misure relative all'andatura di 93 soggetti affetti dalla malattia di Parkinson e 73 soggetti di controllo, per un totale di 306 registrazioni, ottenute nell'ambito di tre differenti studi clinici. Uno stesso soggetto può essere associato a più registrazioni, corrispondenti a differenti prove di cammino.

Nel database sono presenti file in formato testo, identificati da nomi quali *GaCo16_02.txt*, *JuPt06_07.txt* e *SiPt05_01.txt*. La nomenclatura dei file segue la seguente struttura:

1. Un prefisso che identifica lo studio di provenienza dei dati, ovvero “Ga”, associato allo studio di Galit Yogev et al. [15], “Ju”, associato allo studio di Hausdorff et al. [16], oppure “Si”, associato agli studi di Silvi Frenkel-Toledo et al. [17, 18];
2. Due lettere che indicano se il soggetto fa parte del gruppo dei controlli (Co) o se, invece, rientra tra quelli patologici (Pt);
3. Un numero a due cifre che identifica ogni soggetto all’interno dello specifico gruppo di appartenenza;
4. Un altro numero a due cifre, separato dal precedente mediante il carattere “_” (*underscore*), che identifica la prova di cammino eseguita dal soggetto. Nel caso dello studio "Ga", il valore "10" identifica una prova di *dual task*, nella quale il soggetto cammina mentre esegue un compito cognitivo di sottrazione seriale di 7.

La frequenza di campionamento utilizzata per la raccolta dei dati è pari a 100 Hz. Ciò significa che viene acquisito un campione ogni 10 ms.

Ogni file ha 19 colonne così suddivise:

- La prima colonna rappresenta il tempo (in secondi);
- Nelle colonne 2-9 sono riportate le forze di reazione verticale registrate tramite 8 sensori posizionati nel piede sinistro;
- Nelle colonne 10-17 sono riportate le forze di reazione verticale registrate tramite 8 sensori posizionati nel piede destro;
- Le colonne 18 e 19 riportano, rispettivamente, la forza di reazione verticale totale del piede sinistro e del piede destro.

Durante la fase di acquisizione dei dati, ai partecipanti è stato chiesto di camminare su un percorso pianeggiante, mantenendo il proprio passo abituale per circa due minuti consecutivi.

Le misure sono state acquisite mediante due solette sensorizzate, ciascuna dotata di otto sensori di forza distribuiti sulla superficie plantare. A partire dai segnali registrati dai singoli sensori viene ricavata, per ciascun istante temporale, la forza verticale complessiva esercitata dal piede sinistro e dal piede destro. Nel presente lavoro sono state utilizzate esclusivamente le serie temporali della forza totale dei due piedi, senza considerare la posizione dei singoli sensori.

Il database comprende inoltre un file denominato *demographics.xls*, contenente le principali informazioni anagrafiche e cliniche dei soggetti partecipanti allo studio. Il file è organizzato nelle seguenti colonne, da sinistra a destra:

- "ID", cioè il codice identificativo del soggetto, già descritto sopra;
- lo studio da cui i dati hanno origine ("Ga", "Ju" o "Si");
- il gruppo di appartenenza ("CO" o "PD");
- il numero del soggetto;
- il sesso;
- l'età;
- l'altezza (m);
- il peso (kg);
- Hoehn–Yahr, una scala per i pazienti parkinsoniani che classifica la progressione del morbo di Parkinson in funzione della compromissione motoria e della stabilità posturale;
- UPDRS, una scala che quantifica la gravità della malattia non solo sulla base dei sintomi motori, ma anche delle attività quotidiane, aspetti cognitivi e complicanze della terapia;
- UPDRSM, sottoscala motoria dell'UPDRS, che fornisce informazioni circa la compromissione motoria del paziente;
- TUAG (Timed Up and Go), che rappresenta il risultato del *Timed Up and Go*, un test clinico ampiamente utilizzato per valutare la mobilità funzionale e il rischio di caduta;
- "Speed_01 (m/sec)" rappresenta la velocità media del soggetto durante la camminata eseguita in condizioni normali (*single task*);
- "Speed_10" rappresenta la velocità media del soggetto durante la camminata in *dual task*.

Tra le variabili presenti nel file *demographics.xls*, particolare rilevanza assumono il gruppo di appartenenza del soggetto e la scala Hoehn–Yahr, utilizzati nel presente lavoro per distinguere i soggetti di controllo dai pazienti affetti dalla malattia di Parkinson e confrontare differenti livelli di gravità della patologia. Sebbene la scala Hoehn–Yahr preveda sette livelli di gravità (1, 1,5, 2, 2,5, 3, 4 e 5), nel database sono presenti esclusivamente soggetti appartenenti ai livelli 2, 2,5 e 3.

2.2 Estrazione della *stance phase*

Prima di procedere con l'applicazione delle tecniche di analisi della variabilità descritte nel capitolo precedente, è necessario ricavare, per ciascun soggetto, la serie temporale delle durate della fase di appoggio (*stance phase*) di entrambi gli arti inferiori. I segnali di forza registrati dal database rappresentano infatti misure grezze, dalle quali è necessario individuare gli istanti di inizio e di fine della fase di appoggio per ciascun passo. Tale procedura, nota come estrazione della *stance phase*, costituisce il punto di partenza per la costruzione delle serie temporali successivamente analizzate mediante varianza, DFA, EDFA e FA.

Nel presente lavoro l'obiettivo è determinare automaticamente l'inizio e la fine della *stance phase* per ciascun passo, così da ricavare la serie temporale delle durate della fase di appoggio. L'estrazione della fase di appoggio è stata ispirata al metodo descritto da Hausdorff et al. [19], fondato sull'utilizzo di sensori *footswitch* posizionati sotto il piede per individuare gli istanti di contatto con il suolo. Tale metodo si basa sull'analisi del segnale proveniente dal sistema *footswitch*. Come illustrato in Figura 2.1, l'individuazione della fase di appoggio avviene in tre passaggi: dapprima vengono identificati i fronti di salita e di discesa del segnale mediante una soglia prefissata; successivamente vengono individuati i minimi locali della derivata; infine, applicando un opportuno *offset* rispetto a tali minimi, vengono determinati l'istante di inizio (*Start Stance*) e di fine (*End Stance*) della fase di appoggio.

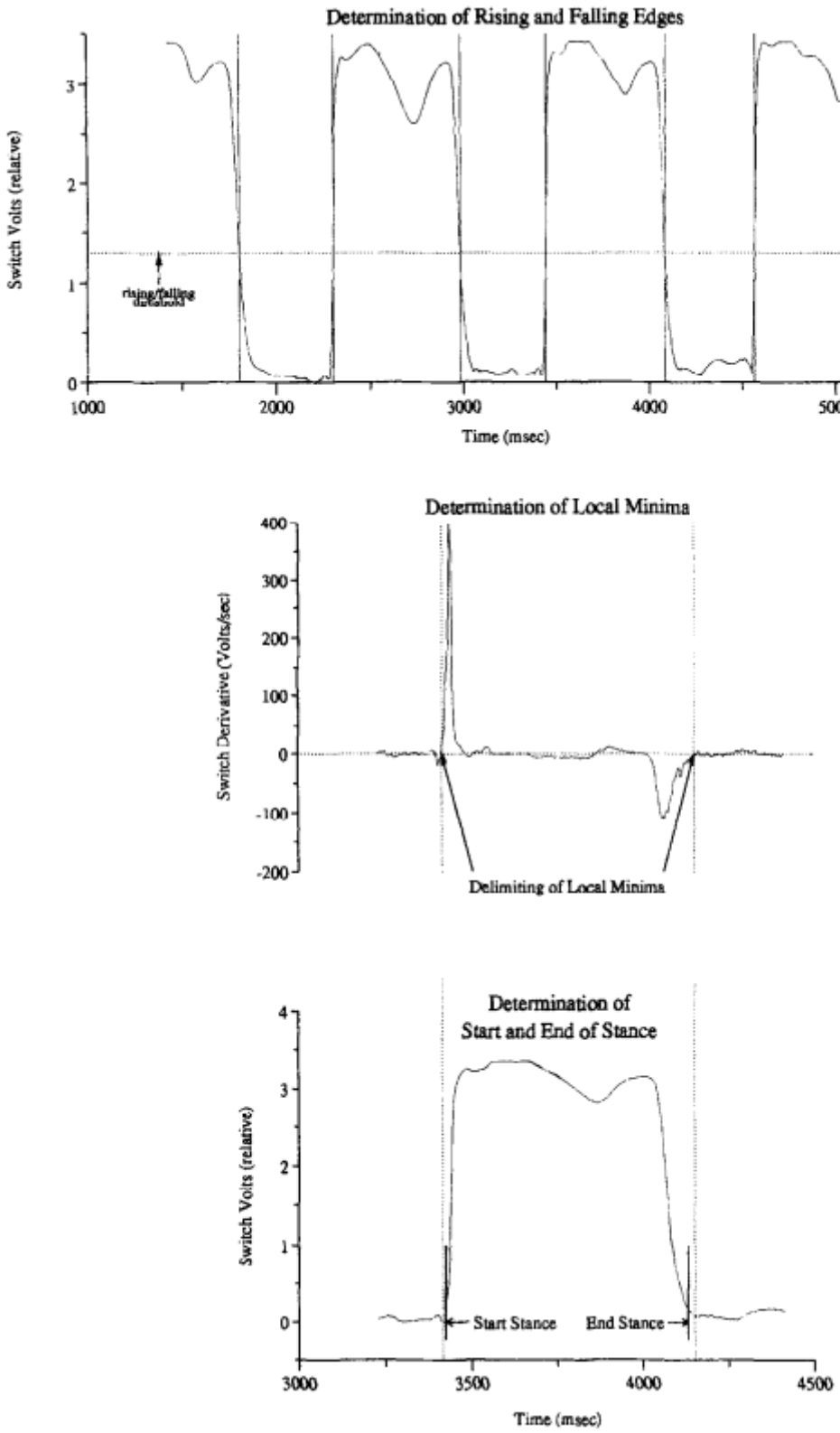


Figura 2.1 – Procedura per l'individuazione automatica dell'inizio e della fine della *stance phase*. Dall'alto verso il basso sono illustrati: (a) l'individuazione dei fronti di salita e discesa del segnale; (b) l'identificazione dei minimi locali della derivata; (c) la determinazione degli istanti di *Start Stance* ed *End Stance* [19].

Sebbene l'algoritmo originale sia stato sviluppato per segnali (espressi in volt) acquisiti mediante un sistema *footswitch*, in questo studio esso è stato adattato alle serie temporali della forza verticale

totale registrata dalle solette sensorizzate del database *PhysioNet*. In particolare, i criteri di individuazione dell'inizio e della fine della *stance phase* sono stati applicati ai segnali di forza dei due piedi, opportunamente elaborati mediante uno script sviluppato in ambiente MATLAB. Nei paragrafi successivi sono descritte le modifiche introdotte rispetto all'algoritmo originale, nonché i criteri adottati per l'individuazione automatica degli istanti di inizio (*Start Stance*) e di fine (*End Stance*) della fase di appoggio.

Poiché i segnali di forza sono inevitabilmente affetti da rumore e oscillazioni anomale dovute agli strumenti di misura e alle condizioni operative, la scelta delle soglie svolge un ruolo cruciale nell'individuazione delle fasi di appoggio. Le soglie devono infatti essere definite in modo da garantire un adeguato rapporto segnale/rumore, limitando l'effetto dei disturbi senza alterare significativamente l'andamento del segnale. Per ciascun segnale di forza è stata definita una soglia T pari al 5% del valore massimo registrato nella relativa acquisizione:

$$T = 0.05 \cdot \max(F)$$

dove F rappresenta la serie temporale della forza verticale totale del piede considerato. La scelta di una soglia relativa, anziché assoluta, consente di adattare automaticamente il criterio di rilevamento ai differenti livelli di forza presenti nei vari soggetti e nelle diverse prove di cammino.

Una volta definita la soglia, il segnale viene trasformato in una sequenza binaria:

- 1, quando la forza supera la soglia
- 0 negli altri casi.

Successivamente vengono individuate le transizioni della sequenza binaria:

- il passaggio da 0 a 1 identifica l'istante di inizio della *stance phase* (*Start Stance*);
- il passaggio da 1 a 0 identifica l'istante di fine della *stance phase* (*End Stance*).

Gli istanti così ottenuti vengono convertiti nei corrispondenti tempi mediante il vettore temporale del segnale.

Per garantire la correttezza dell'estrazione, gli istanti di inizio e fine vengono associati a coppie e vengono eliminate eventuali rilevazioni non valide, ossia quelle nelle quali l'istante finale precede quello iniziale.

La durata della *stance phase* viene quindi calcolata come

$$Stance = t_{End} - t_{Start}$$

Diversamente dall'algoritmo proposto da Hausdorff et al., nel presente lavoro non è stato utilizzato il criterio basato sui minimi locali della derivata e sull'*offset* temporale per correggere gli istanti di

contatto. Questa scelta è motivata dal fatto che i segnali analizzati non provengono da sensori *footswitch*, bensì dalle serie temporali della forza verticale totale fornite dal database *PhysioNet*. In tali segnali gli istanti di contatto risultano già ben identificabili mediante il semplice superamento della soglia, rendendo non necessaria la successiva correzione tramite *offset*. L'approccio adottato consente quindi di ottenere un'estrazione stabile delle *stance* mantenendo una procedura più semplice e direttamente applicabile ai dati disponibili.

2.3 *Pre-processing* delle serie temporali

Come già accennato nel paragrafo precedente, le serie temporali ottenute dall'estrazione delle *stance phase* possono essere affette da rumore e da altre anomalie che potrebbero influenzare le analisi successive. In questo paragrafo sono illustrate le procedure di *pre-processing*, implementate in ambiente MATLAB, adottate per migliorare la qualità dei dati e renderli idonei alle successive elaborazioni.

2.3.1 Selezione del tratto valido

Durante la fase di acquisizione possono verificarsi situazioni che introducono componenti indesiderate nei segnali registrati. Ad esempio, il soggetto può arrestarsi temporaneamente, cambiare direzione, iniziare a camminare prima dell'inizio della registrazione oppure continuare a muoversi dopo il termine della prova. A tali effetti si aggiungono il rumore strumentale e le interferenze introdotte dall'ambiente di misura, che possono alterare la qualità del segnale acquisito.

La Figura 2.2 mostra un esempio di segnale particolarmente rumoroso, relativo al file *GaCo13_10.txt*.

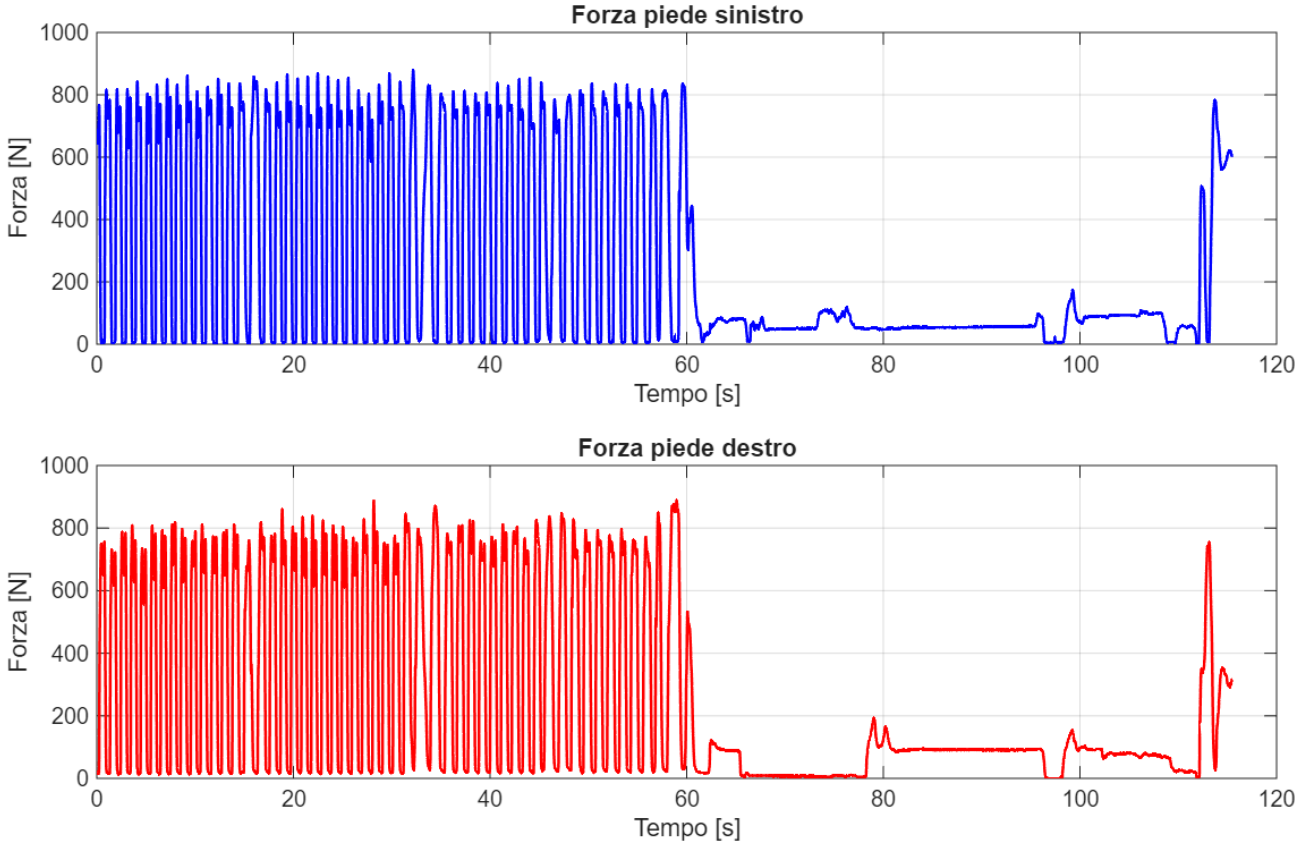


Figura 2.2 – Serie temporali della forza verticale totale del piede sinistro (in alto) e del piede destro (in basso) relative al file *GaCo13_10.txt*.

Oltre ai tratti rumorosi presenti all'inizio o alla fine della registrazione, possono verificarsi anche interruzioni del cammino o artefatti localizzati all'interno del segnale. Un esempio è riportato in Figura 2.3, relativa al file *SiPt05_01.txt*, nella quale sono presenti brevi segmenti non riconducibili alla normale deambulazione intervallati da periodi di cammino regolare. Tali situazioni richiedono l'impiego di procedure automatiche in grado di identificare e rimuovere le porzioni di segnale non valide senza compromettere le parti utili della registrazione.

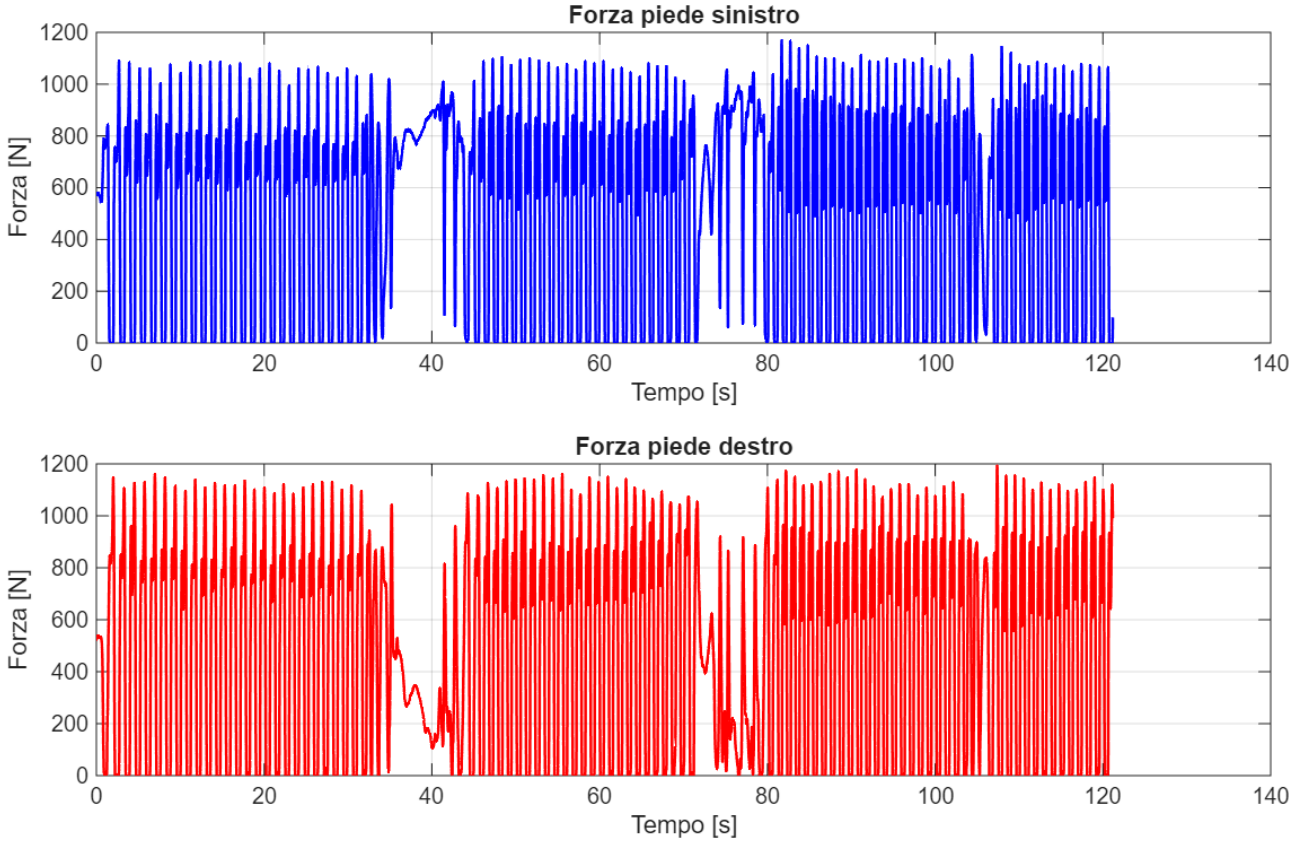


Figura 2.3 – Serie temporali della forza verticale totale del piede sinistro (in alto) e del piede destro (in basso) relative al file *SiPt05_01.txt*.

Per ovviare al problema degli artefatti da movimento, è stato sviluppato un algoritmo in grado di individuare e selezionare automaticamente esclusivamente i tratti "validi" del segnale, ossia quelli che soddisfano determinati requisiti. In particolare, assumendo come riferimento gli istanti di inizio della *stance phase* del piede sinistro (*Initial Contact,* IC*)*, l'algoritmo ricerca coppie di eventi IC consecutivi separati da un intervallo temporale superiore a una soglia prefissata, pari a 3 s. In tal caso, il cammino viene considerato interrotto. La registrazione viene quindi suddivisa in blocchi distinti e viene selezionato automaticamente il blocco continuo contenente il maggior numero di eventi IC, tagliando i tratti di segnale considerati non validi. La soglia di 3 s è stata scelta in quanto un intervallo di tempo superiore tra due passi consecutivi non è compatibile con una normale deambulazione e indica verosimilmente una pausa o un'interruzione della registrazione.

Le Figure 2.4 e 2.5 mostrano le serie temporali dei file *GaCo13_10.txt* e *SiPt05_01.txt* dopo l'applicazione dell'algoritmo di selezione del tratto valido.

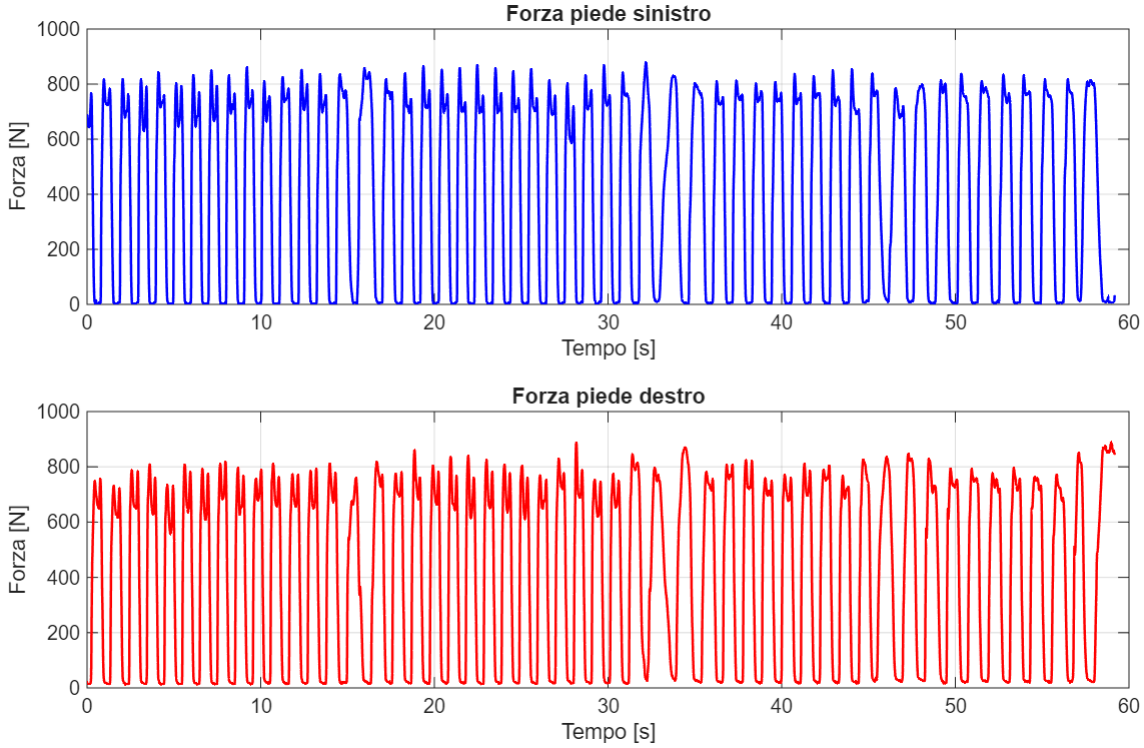


Figura 2.4 – Serie temporali della forza verticale totale del piede sinistro (in alto) e del piede destro (in basso) relative al file *GaCo13_10.txt*, dopo la selezione del tratto valido.

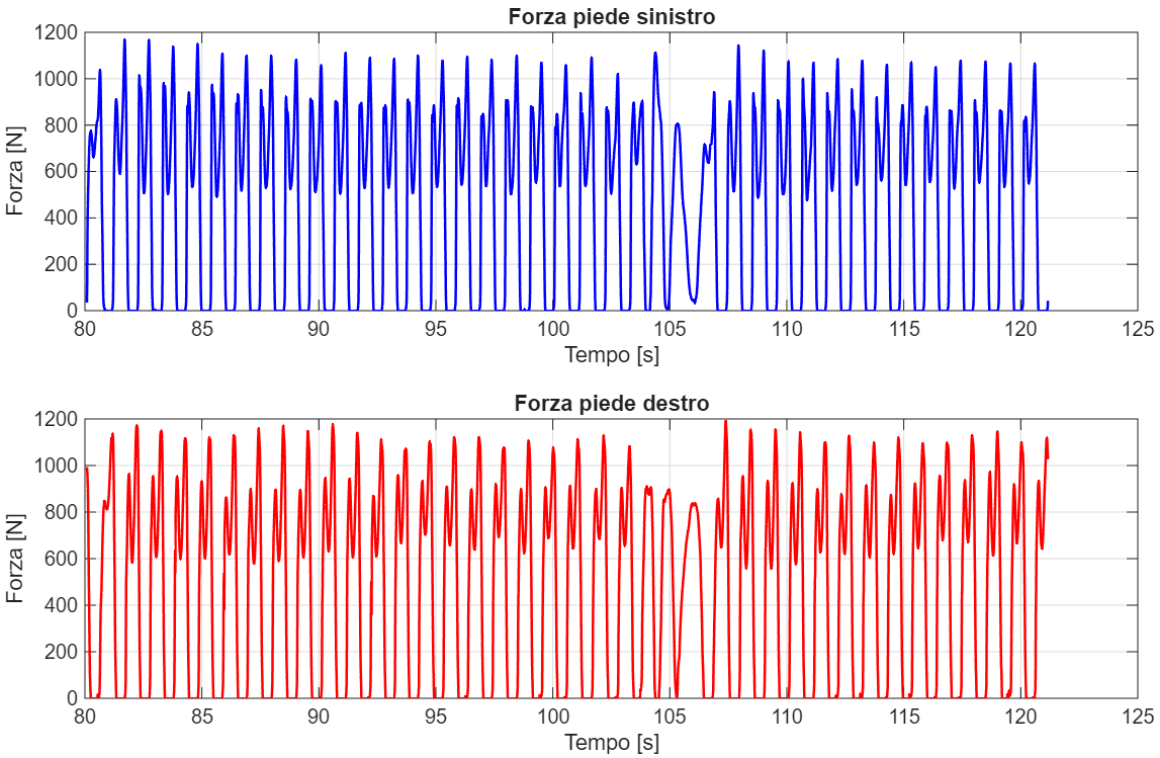


Figura 2.5 – Serie temporali della forza verticale totale del piede sinistro (in alto) e del piede destro (in basso) relative al file *SiPt05_01.txt*, dopo la selezione del tratto valido.

Come si può osservare, l'algoritmo elimina automaticamente le porzioni anomale presenti all'interno del segnale, mantenendo esclusivamente il blocco continuo contenente il maggior numero di eventi IC.

2.3.2 Trattamento dei valori mancanti e rimozione degli *outlier*

Prima dell'identificazione degli *outlier*, le serie della *stance phase* sono state analizzate per verificare l'eventuale presenza di valori mancanti (*NaN*). Tali valori possono derivare da anomalie durante la fase di acquisizione o da precedenti operazioni di elaborazione e, se non gestiti correttamente, possono compromettere le analisi successive.

L'individuazione dei valori mancanti è stata effettuata mediante la funzione MATLAB `isnan`, che consente di identificare automaticamente i valori *NaN* presenti nella serie temporale. Le osservazioni corrispondenti sono state eliminate prima delle successive elaborazioni.

Una volta rimossi gli eventuali valori mancanti, le serie sono state sottoposte a un'ulteriore procedura di controllo finalizzata all'identificazione di valori anomali (*outlier*), ossia osservazioni fisiologicamente improbabili che potrebbero influenzare le analisi statistiche.

Gli *outlier* sono stati identificati mediante il metodo basato sulla *Median Absolute Deviation* (MAD), implementato nella funzione MATLAB `isoutlier`. Tale metodo calcola la dispersione dei dati rispetto alla mediana della serie temporale, risultando particolarmente robusto nei confronti dei valori anomali. Un campione viene classificato come *outlier* quando la sua distanza dalla mediana supera una soglia definita in funzione della MAD.

Per evidenziare l'effetto delle procedure di eliminazione dei valori mancanti e degli *outlier*, la Figura 2.6 mostra la serie della *stance phase* relativa al file *SiPt39_01.txt* prima e dopo il *pre-processing*.

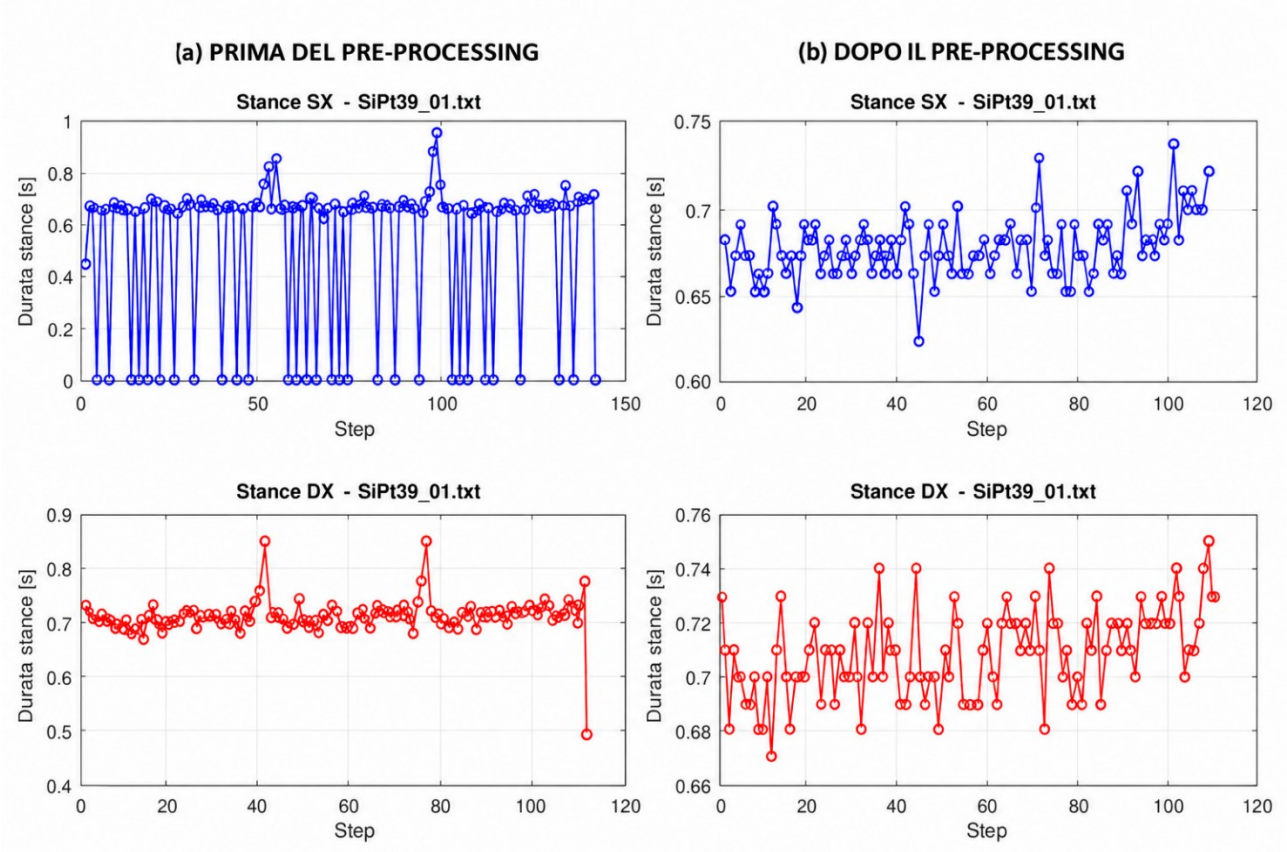


Figura 2.6 – Esempio dell'effetto del *pre-processing* sulla serie temporale della *stance phase* relativa al file *SiPt39_01.txt*. (a) Serie originale, caratterizzata dalla presenza di valori anomali e *stance* fisiologicamente non plausibili. (b) Serie ottenuta dopo la rimozione dei valori mancanti e degli *outlier*, utilizzata per le successive analisi della variabilità.

Dalla Figura 2.6 emerge come il *pre-processing* permetta di ottenere serie della *stance phase* prive di valori anomali e quindi idonee alle successive analisi della variabilità.

2.4 Esclusione delle registrazioni problematiche

La procedura di *pre-processing*, costituita dalle fasi di selezione del tratto valido, trattamento dei valori mancanti e rimozione degli *outlier*, consente nel complesso di attenuare gli effetti prodotti da artefatti indesiderati e valori anomali.

Tuttavia, nel caso del file *GaCo16_10.txt*, il segnale di forza ottenuto al termine del *pre-processing* risulta eccessivamente breve. La ridotta durata del segnale, unita alla presenza di artefatti che ne compromettono la qualità negli ultimi secondi della registrazione, si riflette inevitabilmente anche sulla serie della *stance phase*, che risulta costituita da un numero insufficiente di passi e presenta una qualità complessivamente inadeguata per le successive analisi. Per tali motivi, il file *GaCo16_10.txt* è stato escluso dal dataset utilizzato nel presente lavoro.

Le Figure 2.7 e 2.8 riportano, rispettivamente, le serie temporali della forza e della *stance phase* relative al file *GaCo16_10.txt*.

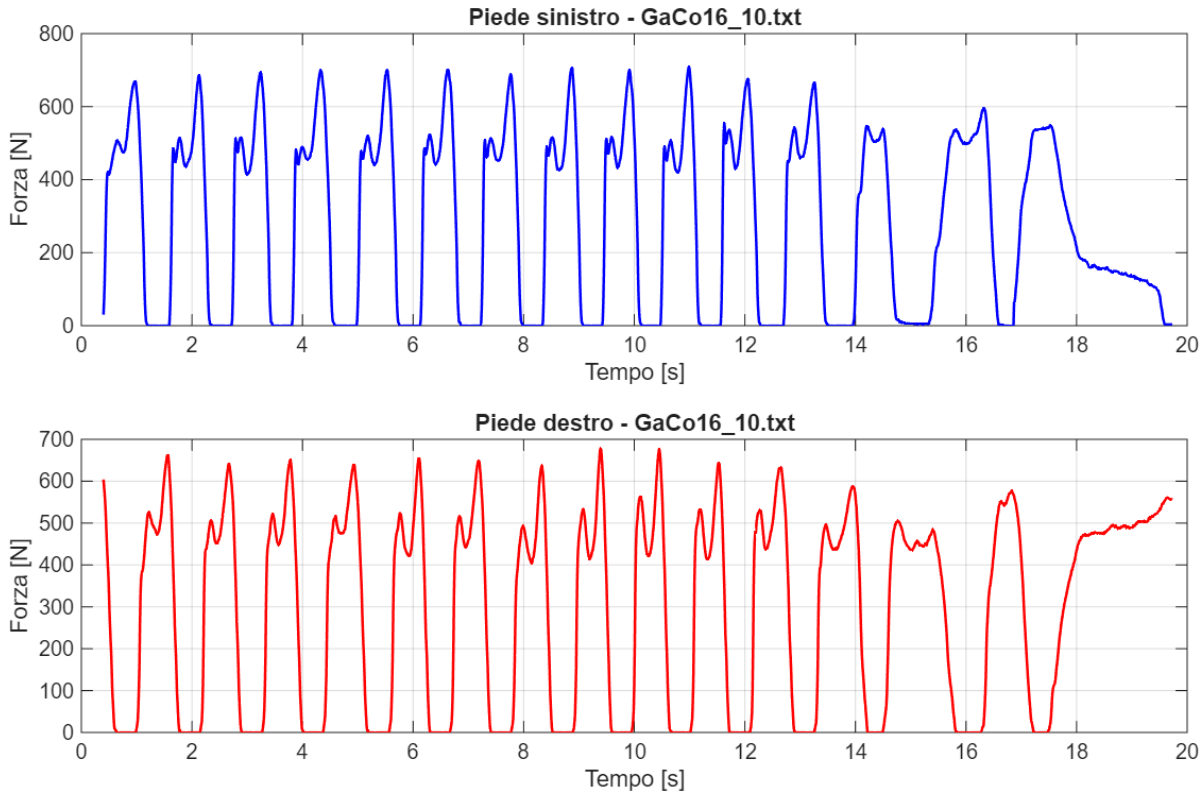


Figura 2.7 – Serie temporali della forza verticale totale del piede sinistro (sopra) e del piede destro (sotto) relative al file *GaCo16_10.txt*. La breve durata della registrazione e gli artefatti presenti nel tratto finale compromettono la qualità del segnale.

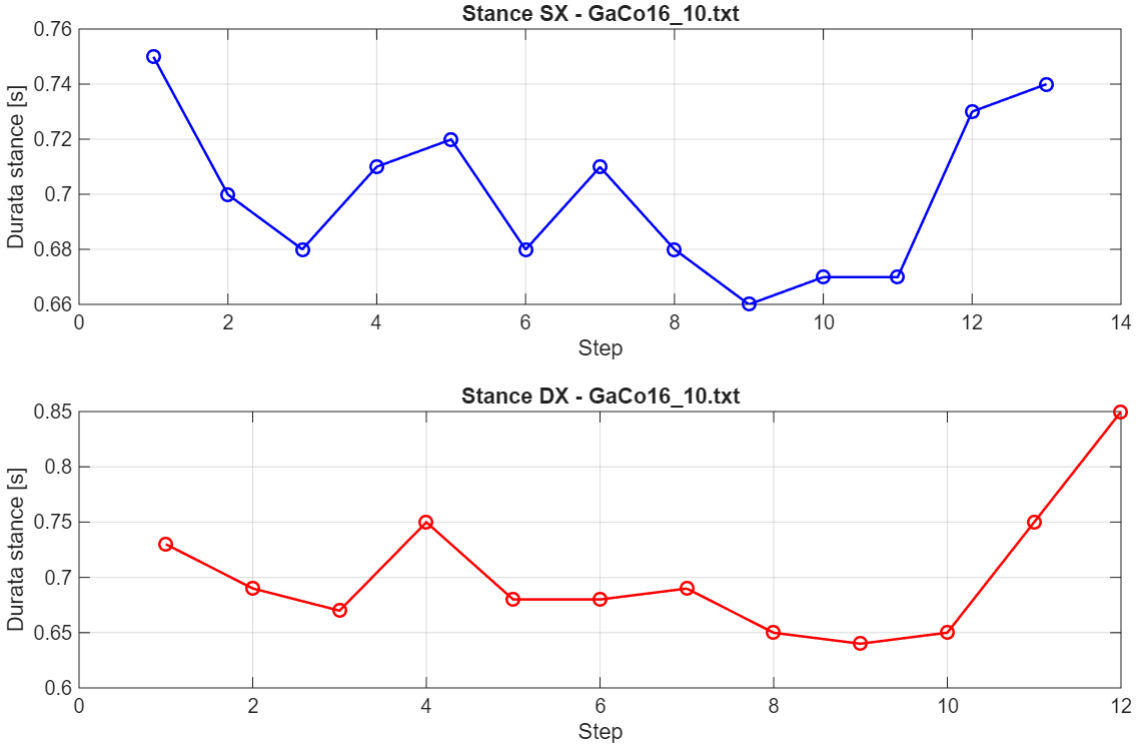


Figura 2.8 – Serie della durata della *stance phase* del piede sinistro (sopra) e del piede destro (sotto) relative al file *GaCo16_10.txt*. La ridotta numerosità delle *stance* (13 per il piede sinistro e 12 per il piede destro) rende la registrazione non idonea alle successive analisi della variabilità.

L'esclusione ha riguardato solamente questa registrazione, mentre tutte le altre sono risultate idonee alle successive analisi.

2.5 Calcolo dei parametri

I parametri di interesse sono stati calcolati mediante apposite funzioni definite in ambiente MATLAB e applicate alle serie temporali delle durate della *stance phase*, precedentemente ottenute e sottoposte a *pre-processing*. Per ciascuna registrazione, le analisi sono state eseguite separatamente per il piede sinistro e per il piede destro.

Per le analisi basate sulle proprietà di scala è stato definito un intervallo di valori della scala *s*. La scala minima è stata fissata pari a 10 elementi della serie, mentre la scala massima è stata determinata separatamente per i due piedi in funzione della lunghezza delle rispettive serie temporali, ponendola pari alla metà del numero di elementi della serie, arrotondata per difetto:

$$s_{\min} = 10$$

$$s_{\max,\mathrm{sx}} = \left\lfloor \frac{N_{\mathrm{sx}}}{2} \right\rfloor, \qquad s_{\max,\mathrm{dx}} = \left\lfloor \frac{N_{\mathrm{dx}}}{2} \right\rfloor$$

dove N_{sx} e N_{dx} rappresentano, rispettivamente, il numero di elementi delle serie della *stance phase* del piede sinistro e del piede destro. Tra il valore minimo e quello massimo sono stati generati 25 valori di scala distribuiti logaritmicamente; tali valori sono stati successivamente arrotondati all'intero più vicino ed eventuali duplicati sono stati rimossi. In questo modo, il limite superiore dell'intervallo di scala è stato adattato alla lunghezza di ciascuna serie temporale.

Per la DFA e l'EDFA è stato utilizzato un *detrending* polinomiale di secondo ordine. Gli esponenti α e β sono stati stimati mediante regressione lineare in scala logaritmica sulle rispettive funzioni di fluttuazione. Per la FA, l'esponente di Hurst *H* è stato analogamente ricavato dalla pendenza della regressione lineare in scala logaritmica.

Per ciascuna registrazione e separatamente per il piede sinistro e per il piede destro, sono stati quindi considerati la varianza delle durate della *stance phase*, l'esponente di scala α della DFA, il parametro β dell'EDFA e l'esponente di Hurst *H* della FA. I valori ottenuti sono stati successivamente utilizzati nelle analisi statistiche.

2.6 Analisi statistiche

Al fine di valutare le differenze tra soggetti sani e pazienti affetti dalla malattia di Parkinson e l'eventuale associazione dei parametri analizzati con la gravità della patologia, i valori ottenuti sono stati sottoposti ad analisi statistica. In particolare, il test di Mann–Whitney è stato utilizzato per confrontare il gruppo dei pazienti con quello dei controlli, il test ANOVA per confrontare le tre classi Hoehn–Yahr presenti nel database e il coefficiente di correlazione di Spearman per valutare l'eventuale associazione monotona tra ciascun parametro analizzato e il livello della scala Hoehn–Yahr. L'unità statistica considerata nelle analisi è stata la singola registrazione; pertanto, soggetti associati a più prove possono contribuire con più osservazioni.

2.6.1 Livello di significatività

Per tutti i test statistici eseguiti nell'analisi condotta è stato adottato un livello di significatività del 5%. Ciò significa che una differenza o una correlazione è stata considerata statisticamente significativa quando il corrispondente *p-value* è risultato inferiore a 0,05. In tali casi è stata rifiutata l'ipotesi nulla (*null hypothesis*, H_0), considerando il risultato statisticamente significativo al livello di significatività adottato. Il livello del 5% corrisponde a una probabilità del 5% di commettere un errore di I tipo, ossia di rifiutare l'ipotesi nulla quando essa è vera. Al contrario, valori superiori o uguali a 0,05 sono stati interpretati come indicativi dell'assenza di evidenze statisticamente sufficienti per rifiutare l'ipotesi nulla.

2.6.2 Test di Mann–Whitney

Il test di Mann–Whitney (o test *U*) è un test non parametrico utilizzato per confrontare due campioni indipendenti [20]. Esso si basa sul confronto dei ranghi delle osservazioni appartenenti ai due gruppi, senza richiedere l'assunzione di normalità delle distribuzioni.

La procedura prevede inizialmente di riunire le osservazioni dei due gruppi e ordinarle in senso crescente, assegnando a ciascuna di esse un rango. In presenza di valori coincidenti, viene assegnato il rango medio corrispondente. Successivamente, vengono calcolate le somme dei ranghi relative ai due gruppi e, a partire da queste, le rispettive statistiche *U*, nel modo seguente:

$$U_1 = n_1 n_2 + \frac{1}{2} n_1(n_1+1) - R_1$$

$$U_2 = n_1 n_2 + \frac{1}{2} n_2(n_2+1) - R_2$$

dove n_1 e n_2 rappresentano il numero di osservazioni appartenenti ai due gruppi, mentre R_1 e R_2 rappresentano la somma dei ranghi delle osservazioni dei due gruppi. La statistica *U* utilizzata nel test corrisponde al minore tra U_1 e U_2.

In questo studio, il test di Mann–Whitney è stato applicato separatamente ai parametri estratti dalle serie della *stance phase*, al fine di verificare l'eventuale presenza di differenze tra i soggetti di controllo e i pazienti affetti dalla malattia di Parkinson. L'ipotesi nulla H_0 assume l'assenza di differenze tra le distribuzioni dei due gruppi, mentre l'ipotesi alternativa H_1 assume la presenza di una differenza. Il risultato del test è stato valutato attraverso il *p-value*, considerando statisticamente significativi i confronti caratterizzati da $p < 0{,}05$.

Il test di Mann–Whitney è stato implementato in ambiente MATLAB mediante la funzione `ranksum`.

2.6.3 Test ANOVA

Il test ANOVA a una via (o analisi della varianza) è un test statistico utilizzato per verificare se le medie di tre o più gruppi distinti siano significativamente differenti tra loro. L'ipotesi nulla H_0 assume che le medie $\mu_1, \mu_2, \ldots, \mu_k$, relative ai k gruppi considerati, siano uguali tra loro:

H_0: $\mu_1 = \mu_2 = \ldots = \mu_k$

L'ipotesi alternativa H_1 stabilisce invece che non tutte le medie siano uguali tra loro, ovvero che almeno una di esse differisca dalle altre.

Come suggerisce il nome, il test ANOVA si basa sull'analisi della varianza. La statistica di test F è calcolata come rapporto tra la varianza intergruppo e la varianza intragruppo [21]. La prima misura la dispersione delle medie dei singoli gruppi rispetto alla media globale, mentre la seconda quantifica la variabilità delle osservazioni all'interno dei singoli gruppi. Un valore elevato di F indica una maggiore variabilità tra i gruppi rispetto a quella osservata al loro interno. A partire dal valore della statistica F viene quindi determinato il corrispondente *p-value*.

Nell'ambito dell'analisi statistica, il test ANOVA a una via è stato implementato in ambiente MATLAB mediante la funzione `anova1`, al fine di individuare eventuali differenze tra i tre livelli della scala Hoehn–Yahr presenti nel dataset.

2.6.4 Correlazione di Spearman

Il coefficiente di correlazione di rango di Spearman è una misura statistica non parametrica utilizzata per valutare il grado di associazione monotona tra due variabili, anche in presenza di una relazione non necessariamente lineare [22].

La correlazione di Spearman si basa sui ranghi delle osservazioni anziché sui valori originali delle variabili. I valori vengono quindi ordinati e sostituiti dai rispettivi ranghi; in presenza di valori coincidenti, viene assegnato il rango medio corrispondente. Il coefficiente ottenuto, indicato con ρ, esprime la direzione e l'intensità dell'associazione monotona tra le due variabili.

Il coefficiente ρ può assumere valori compresi nell'intervallo $[-1, 1]$. Valori pari a $+1$ e -1 indicano, rispettivamente, una perfetta associazione monotona positiva e negativa, mentre valori prossimi a 0 indicano un'associazione monotona debole o assente.

Nel presente lavoro, la correlazione di Spearman è stata utilizzata per valutare l'eventuale associazione monotona tra il livello della scala Hoehn–Yahr e i diversi parametri di interesse. Il risultato della correlazione è stato valutato considerando sia il coefficiente ρ, indicativo della direzione e dell'intensità dell'associazione, sia il relativo *p-value*, adottando il livello di significatività precedentemente definito.

Il coefficiente di correlazione di Spearman è stato calcolato in ambiente MATLAB mediante la funzione `corr`, specificando il metodo `Spearman`.

3 – Risultati

In questo capitolo sono riportati i risultati ottenuti dall'analisi delle serie temporali delle durate della *stance phase*, precedentemente estratte e sottoposte a *pre-processing*. Sono stati analizzati la varianza delle durate della *stance phase*, l'esponente di scala α della DFA, l'esponente β dell'EDFA e l'esponente di Hurst H della FA. Per ciascun parametro sono stati effettuati confronti tra soggetti sani e pazienti con malattia di Parkinson e tra i diversi livelli della scala Hoehn–Yahr, oltre ad analisi di correlazione con il livello della stessa scala.

3.1 Risultati della varianza

Il test di Mann–Whitney effettuato sulla varianza delle durate della *stance phase* ha evidenziato differenze statisticamente significative tra i soggetti sani e i pazienti con malattia di Parkinson sia per il piede sinistro ($p = 7{,}17 \times 10^{-4}$) sia per il piede destro ($p = 7{,}00 \times 10^{-6}$).

Le Figure 3.1 e 3.2 riportano, rispettivamente, i boxplot relativi alla varianza delle durate della *stance phase* del piede sinistro e del piede destro nei soggetti sani e nei pazienti con malattia di Parkinson.

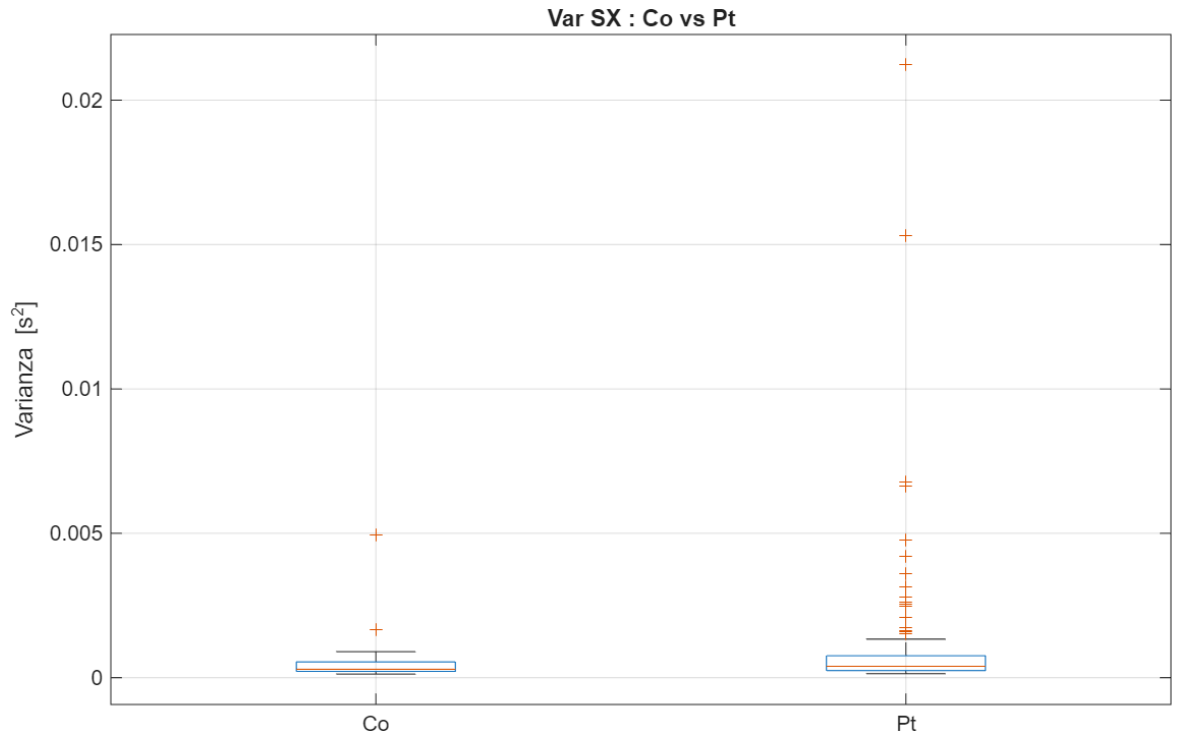


Figura 3.1 – Boxplot relativo alla varianza delle durate della *stance phase* del piede sinistro nei soggetti sani e nei pazienti con malattia di Parkinson.

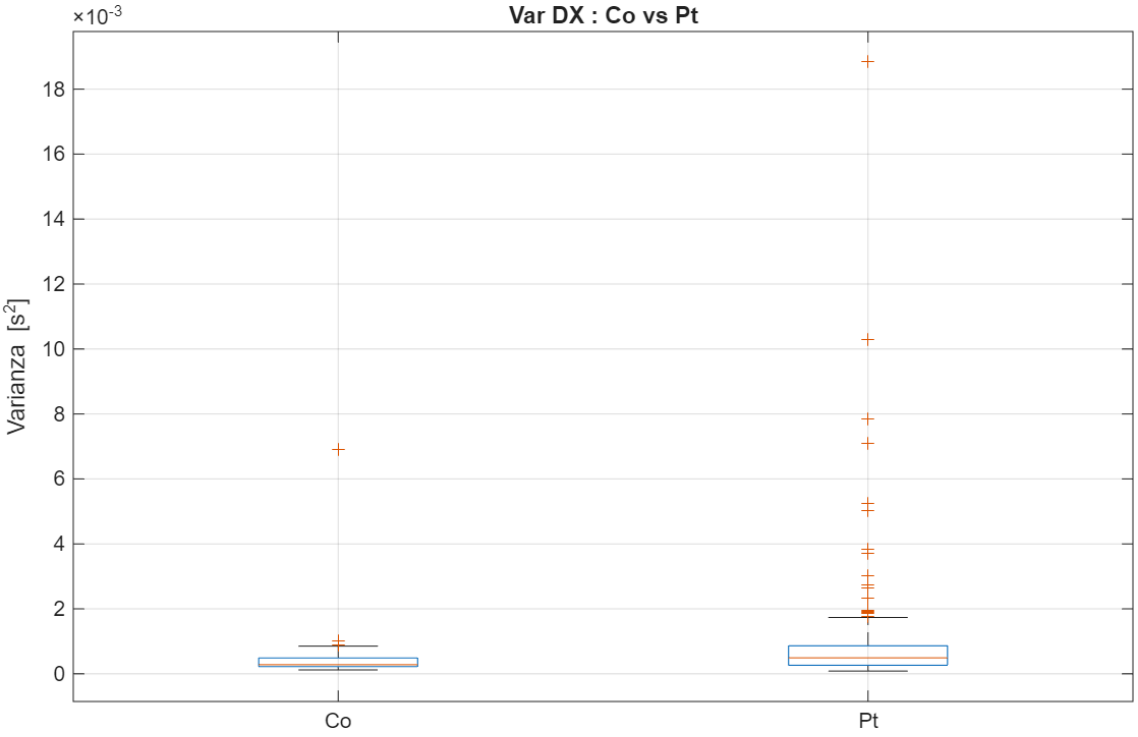


Figura 3.2 – Boxplot relativo alla varianza delle durate della *stance phase* del piede destro nei soggetti sani e nei pazienti con malattia di Parkinson.

Il confronto tra i tre livelli della scala Hoehn–Yahr mediante ANOVA ha evidenziato differenze statisticamente significative sia per il piede sinistro ($p = 8{,}13 \times 10^{-9}$) sia per il piede destro ($p = 7{,}39 \times 10^{-10}$).

L'analisi di correlazione di Spearman ha evidenziato un'associazione positiva, seppur debole, tra la varianza delle durate della *stance phase* e il livello della scala Hoehn–Yahr sia per il piede sinistro ($\rho = 0{,}158$; $p = 0{,}021$) sia per il piede destro ($\rho = 0{,}137$; $p = 0{,}046$).

I boxplot relativi alla varianza nei tre livelli della scala Hoehn–Yahr e i diagrammi relativi all'analisi di correlazione di Spearman sono riportati in Figura 3.3.

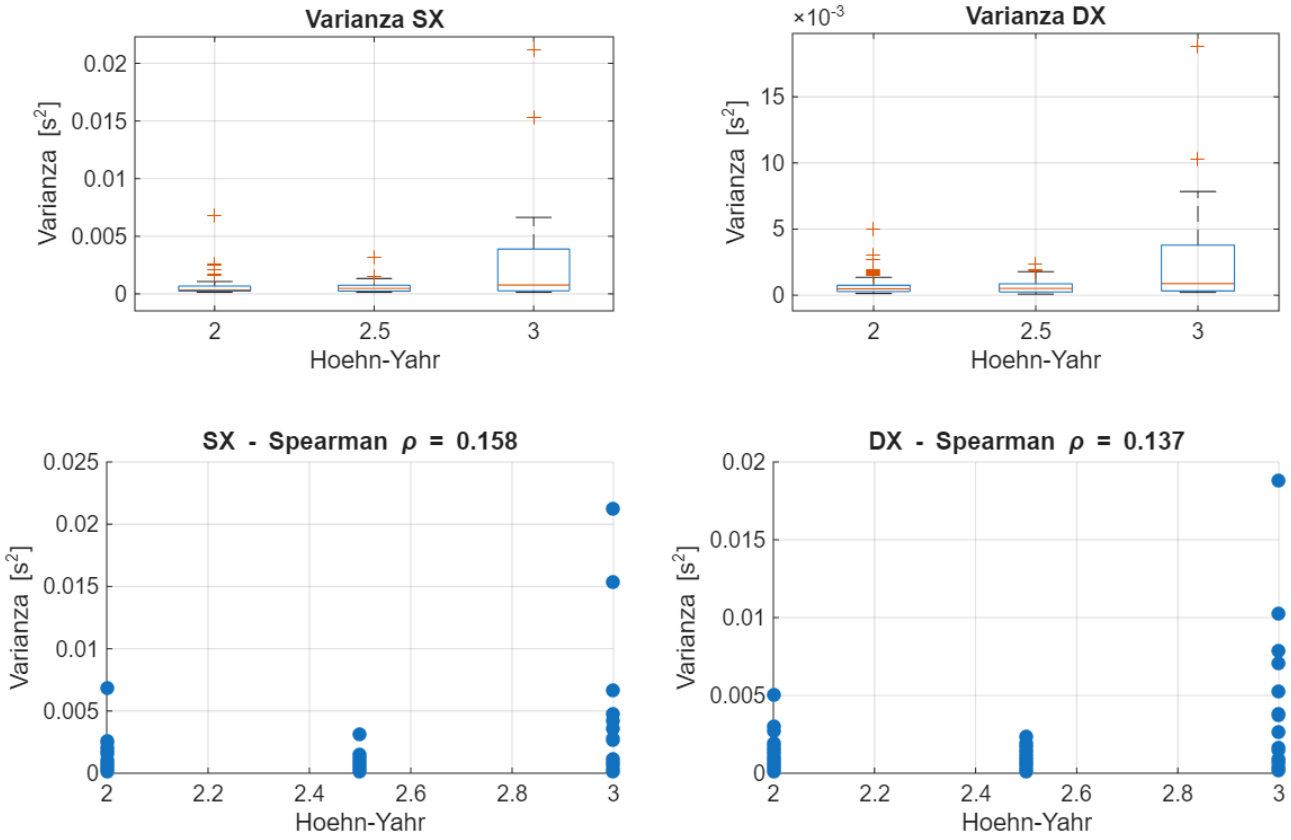


Figura 3.3 – Boxplot della varianza delle durate della *stance phase* e relativi diagrammi di correlazione di Spearman in funzione del livello della scala Hoehn–Yahr.

3.2 Risultati della DFA

Il test di Mann–Whitney applicato all'esponente di scala α della DFA non ha evidenziato differenze statisticamente significative tra i soggetti sani e i pazienti con malattia di Parkinson, né per il piede sinistro (p = 0,787) né per il piede destro (p = 0,437).

Le Figure 3.4 e 3.5 riportano, rispettivamente, i boxplot relativi all'esponente di scala α della DFA per il piede sinistro e per il piede destro nei soggetti sani e nei pazienti con malattia di Parkinson.

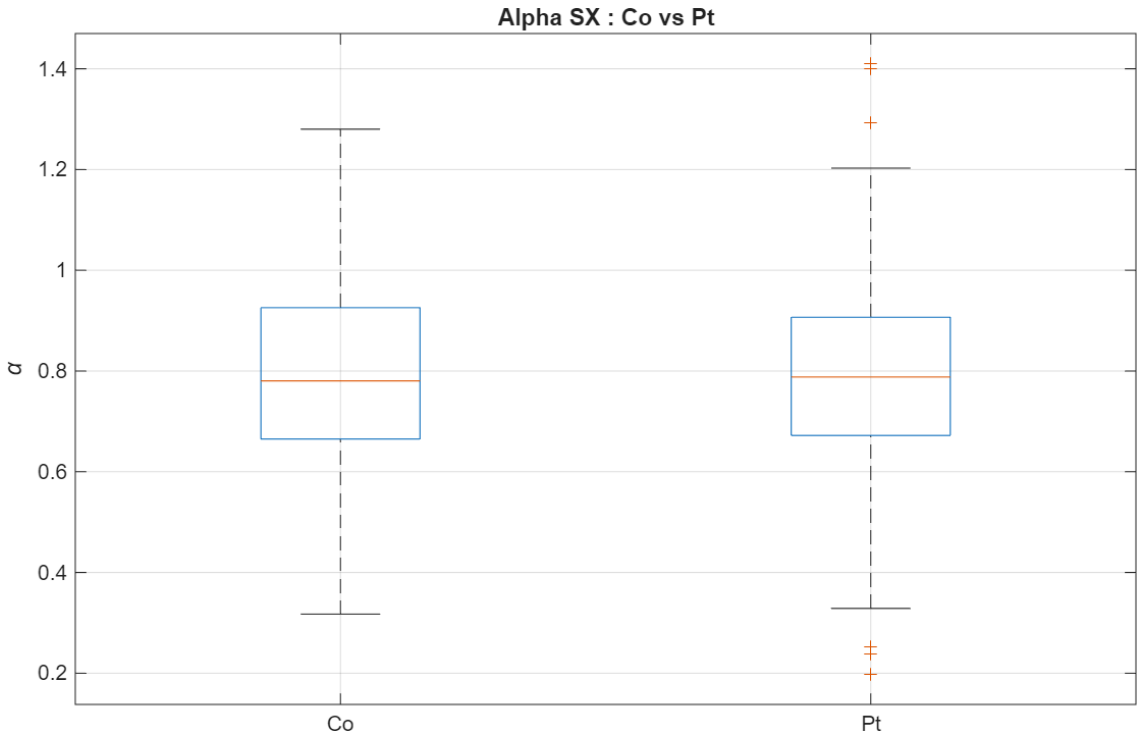


Figura 3.4 – Boxplot relativo all'esponente di scala α della DFA per il piede sinistro nei soggetti sani e nei pazienti con malattia di Parkinson.

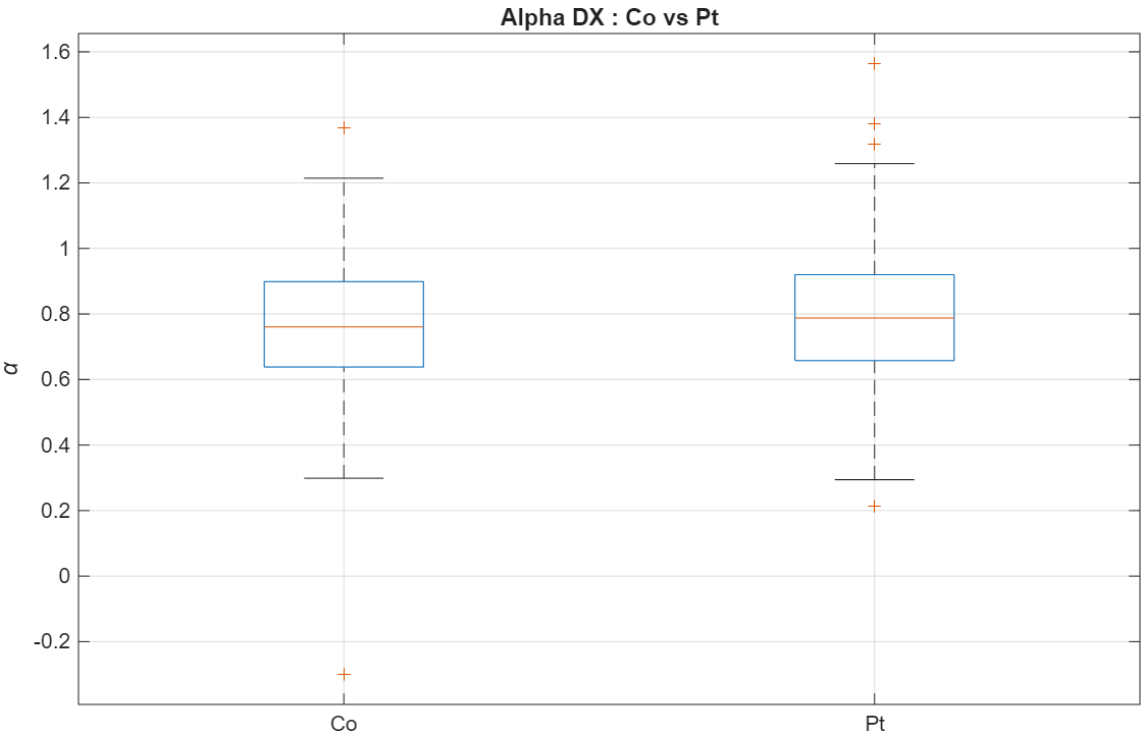


Figura 3.5 – Boxplot relativo all'esponente di scala α della DFA per il piede destro nei soggetti sani e nei pazienti con malattia di Parkinson.

Il confronto tra i tre livelli della scala Hoehn–Yahr mediante ANOVA non ha evidenziato differenze statisticamente significative né per il piede sinistro (p = 0,489) né per il piede destro (p = 0,624).

L'analisi mediante il coefficiente di correlazione di Spearman ha evidenziato, per il piede sinistro, un coefficiente prossimo a zero tra l'esponente α e il livello della scala Hoehn–Yahr ($\rho = 0{,}055$; $p = 0{,}423$), non statisticamente significativo. Anche per il piede destro il coefficiente è risultato prossimo a zero e non statisticamente significativo ($\rho = -0{,}037$; $p = 0{,}587$).

I boxplot relativi all'esponente α nei tre livelli della scala Hoehn–Yahr e i diagrammi relativi all'analisi di correlazione di Spearman sono riportati in Figura 3.6.

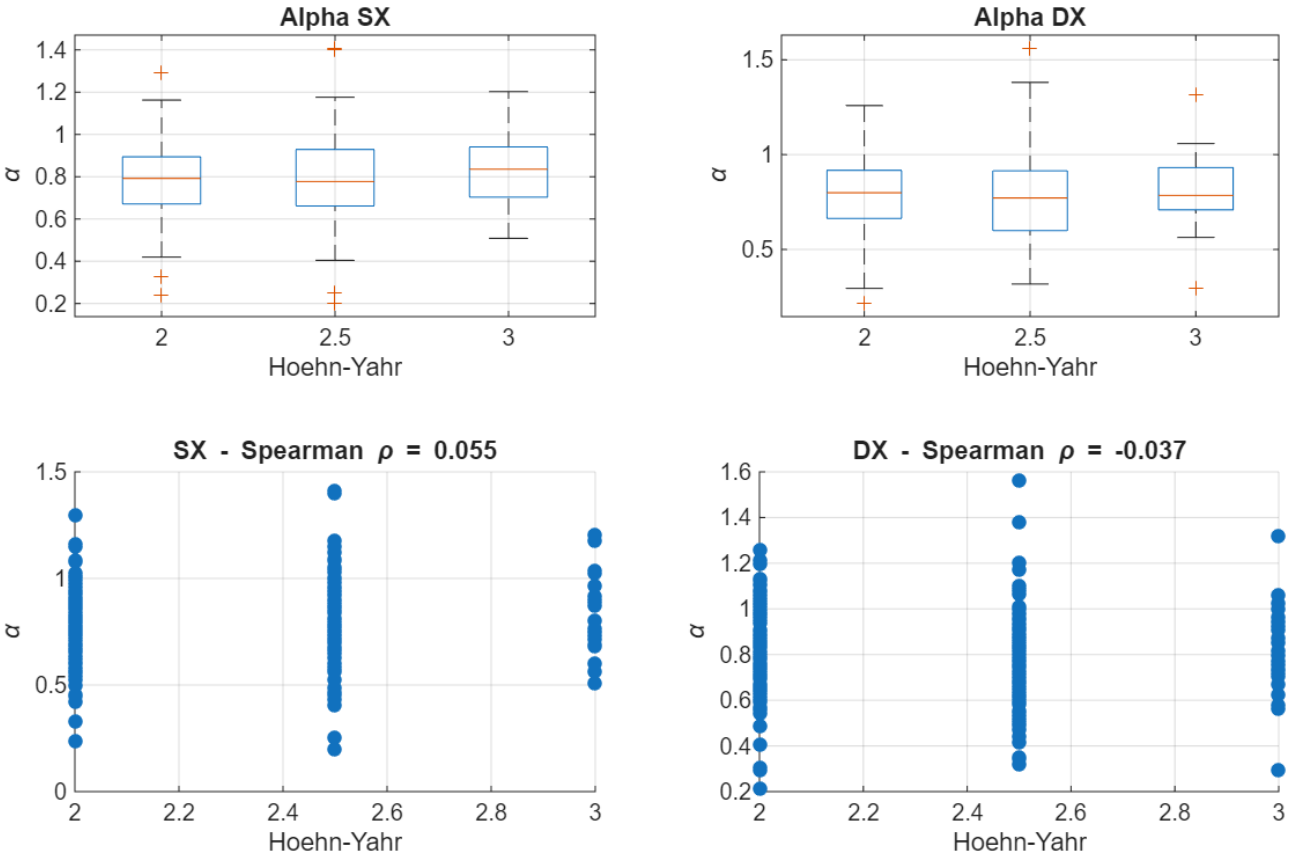


Figura 3.6 – Boxplot dell'esponente di scala α e relativi diagrammi di correlazione di Spearman in funzione del livello della scala Hoehn–Yahr.

3.3 Risultati dell'EDFA

Per quanto riguarda l'EDFA, l'analisi dei risultati è stata condotta sull'esponente β, poiché l'esponente α è già stato considerato nell'ambito della DFA.

Il test di Mann–Whitney applicato all'esponente β non ha evidenziato differenze statisticamente significative tra i soggetti sani e i pazienti con malattia di Parkinson, né per il piede sinistro ($p = 0{,}083$) né per il piede destro ($p = 0{,}274$).

Le Figure 3.7 e 3.8 riportano, rispettivamente, i boxplot relativi all'esponente β dell'EDFA per il piede sinistro e per il piede destro nei soggetti sani e nei pazienti con malattia di Parkinson.

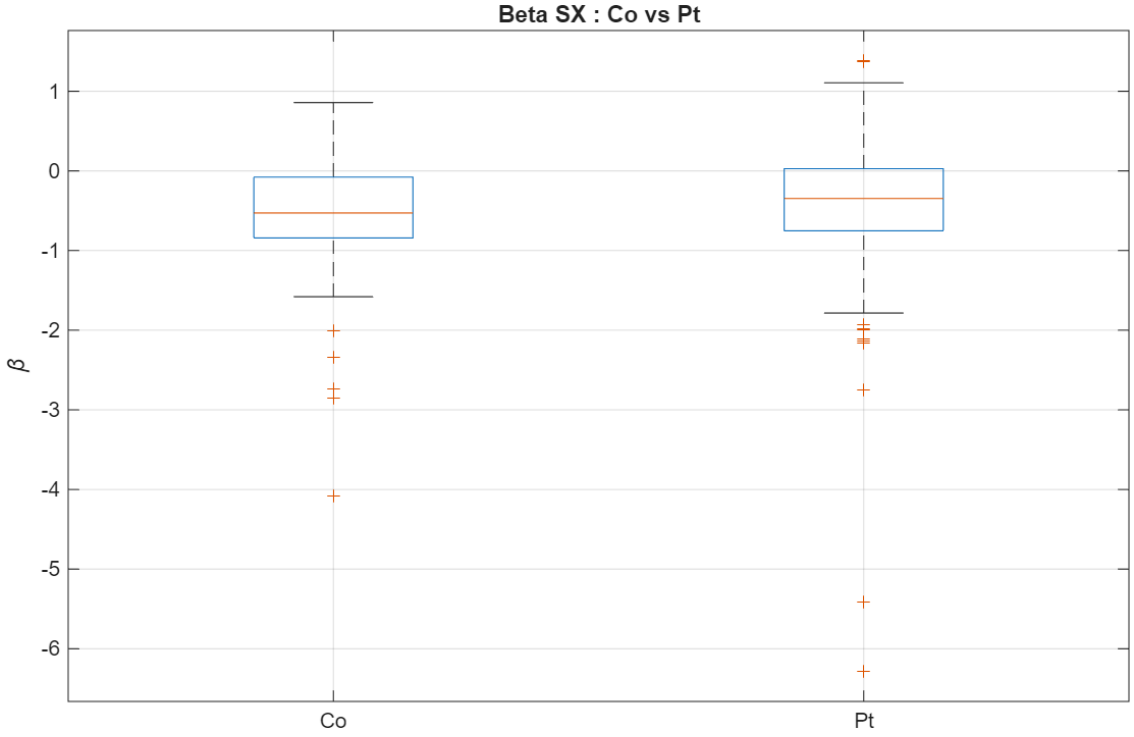


Figura 3.7 – Boxplot relativo all'esponente β dell'EDFA per il piede sinistro nei soggetti sani e nei pazienti con malattia di Parkinson.

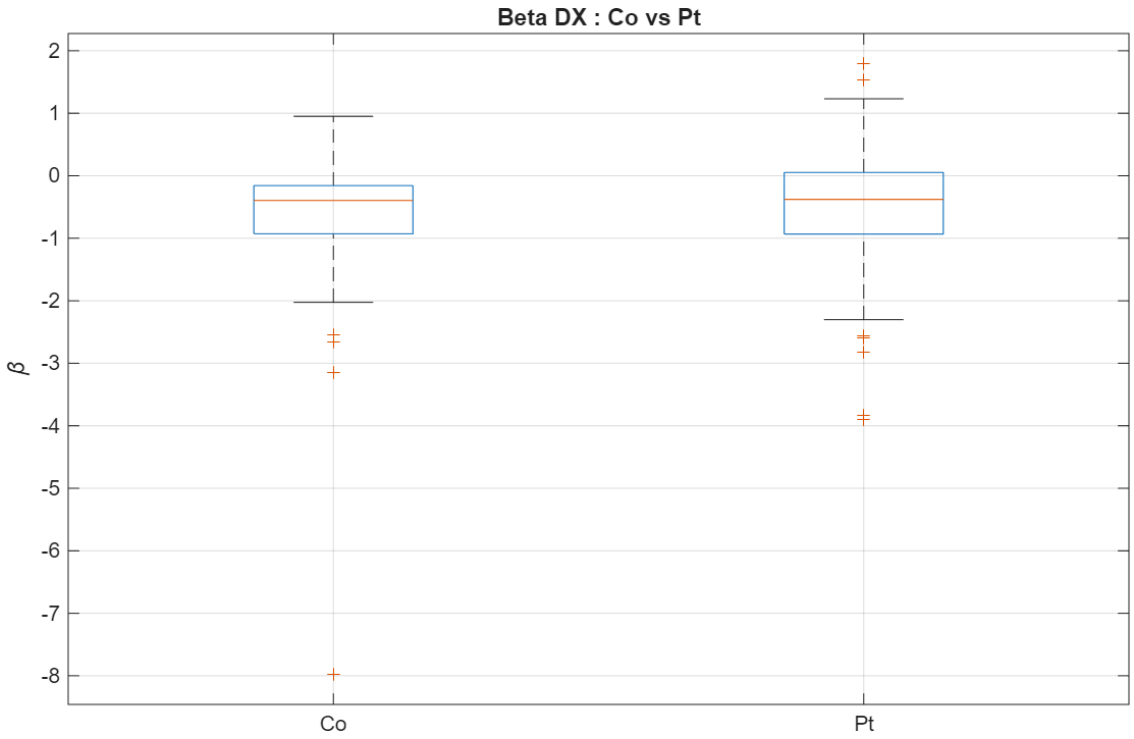


Figura 3.8 – Boxplot relativo all'esponente β dell'EDFA per il piede destro nei soggetti sani e nei pazienti con malattia di Parkinson.

Il confronto tra i tre livelli della scala Hoehn–Yahr mediante ANOVA non ha evidenziato differenze statisticamente significative né per il piede sinistro (p = 0,956) né per il piede destro (p = 0,422).

L’analisi mediante il coefficiente di correlazione di Spearman ha evidenziato coefficienti prossimi a zero e negativi sia per il piede sinistro ($\rho = -0{,}004$; $p = 0{,}959$) sia per il piede destro ($\rho = -0{,}013$; $p = 0{,}847$). In entrambi i casi, la correlazione non è risultata statisticamente significativa.

I boxplot relativi all’esponente β nei tre livelli della scala Hoehn–Yahr e i diagrammi relativi all’analisi di correlazione di Spearman sono riportati in Figura 3.9.

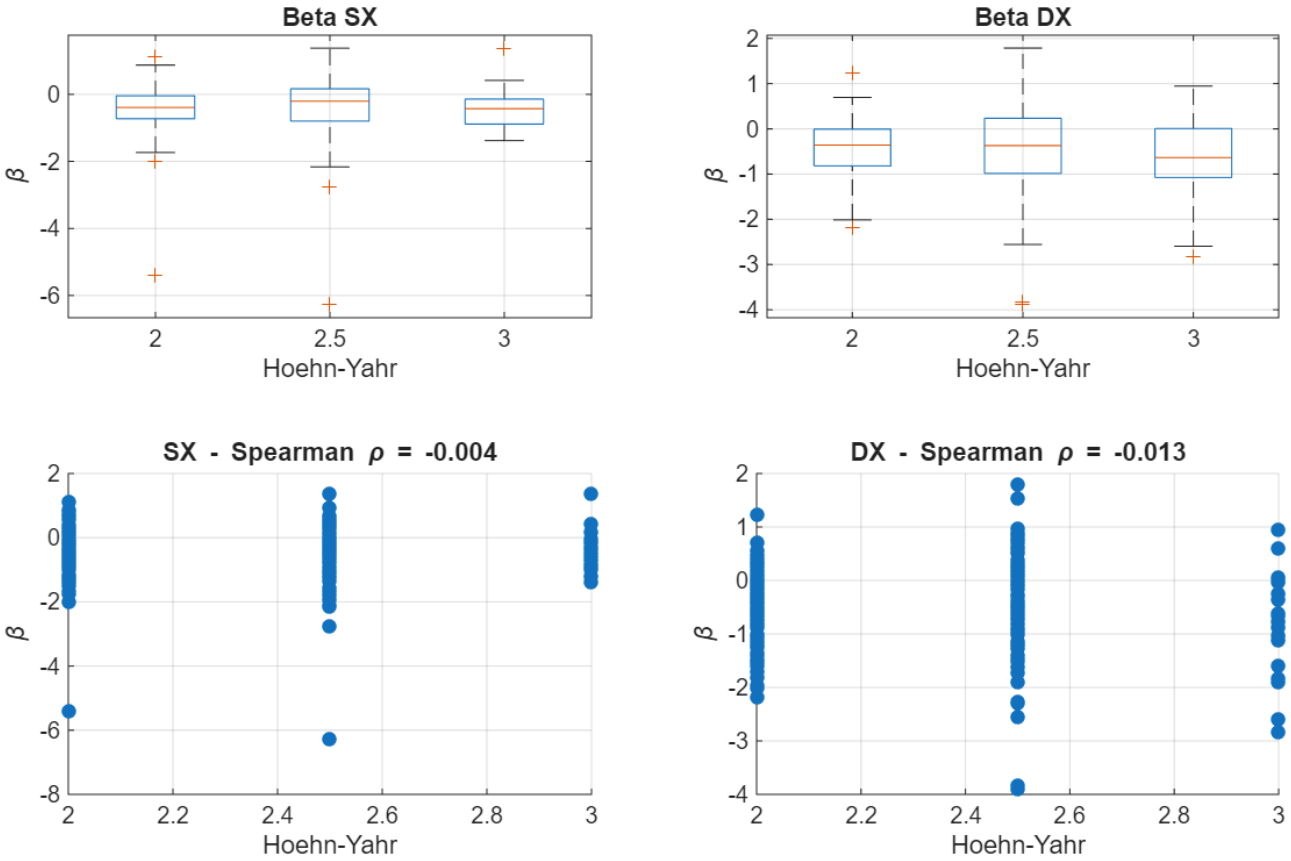


Figura 3.9 – Boxplot dell’esponente β dell’EDFA e relativi diagrammi di correlazione di Spearman in funzione del livello della scala Hoehn–Yahr.

3.4 Risultati della FA

Il test di Mann–Whitney applicato all’esponente di Hurst H della FA ha evidenziato una differenza statisticamente significativa tra i soggetti sani e i pazienti con malattia di Parkinson per il piede sinistro ($p = 0{,}009$), mentre per il piede destro la differenza non è risultata statisticamente significativa ($p = 0{,}204$).

I risultati sono riportati nei boxplot delle Figure 3.10 e 3.11.

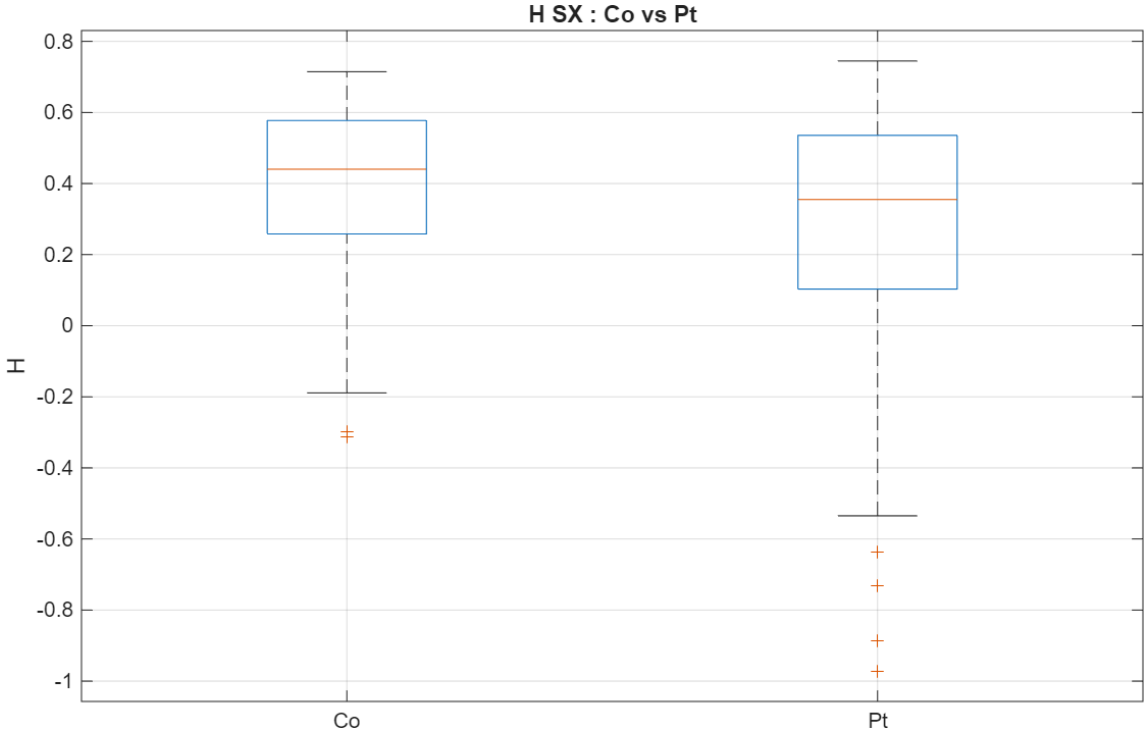


Figura 3.10 – Boxplot relativo all'esponente di Hurst *H* della FA per il piede sinistro nei soggetti sani e nei pazienti con malattia di Parkinson.

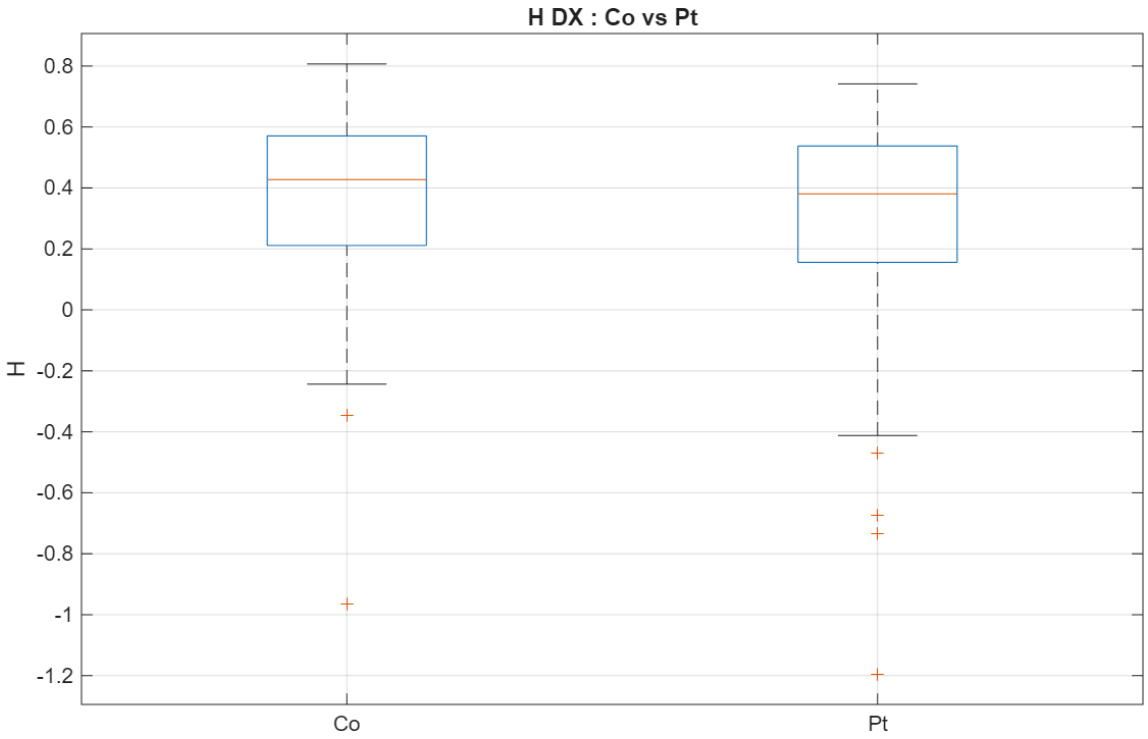


Figura 3.11 – Boxplot relativo all'esponente di Hurst *H* della FA per il piede destro nei soggetti sani e nei pazienti con malattia di Parkinson.

Il confronto tra i tre livelli della scala Hoehn–Yahr mediante ANOVA non ha evidenziato differenze statisticamente significative per il piede sinistro (p = 0,504), mentre è risultato statisticamente significativo per il piede destro (p = 0,043).

L’analisi mediante il coefficiente di correlazione di Spearman ha evidenziato coefficienti negativi per entrambi i piedi. Per il piede sinistro, la correlazione non è risultata statisticamente significativa ($\rho = -0{,}108$; $p = 0{,}116$), mentre per il piede destro è risultata statisticamente significativa ($\rho = -0{,}159$; $p = 0{,}020$).

La Figura 3.12 riporta i boxplot relativi all’esponente *H* nei tre livelli della scala Hoehn–Yahr e i diagrammi relativi all’analisi di correlazione di Spearman.

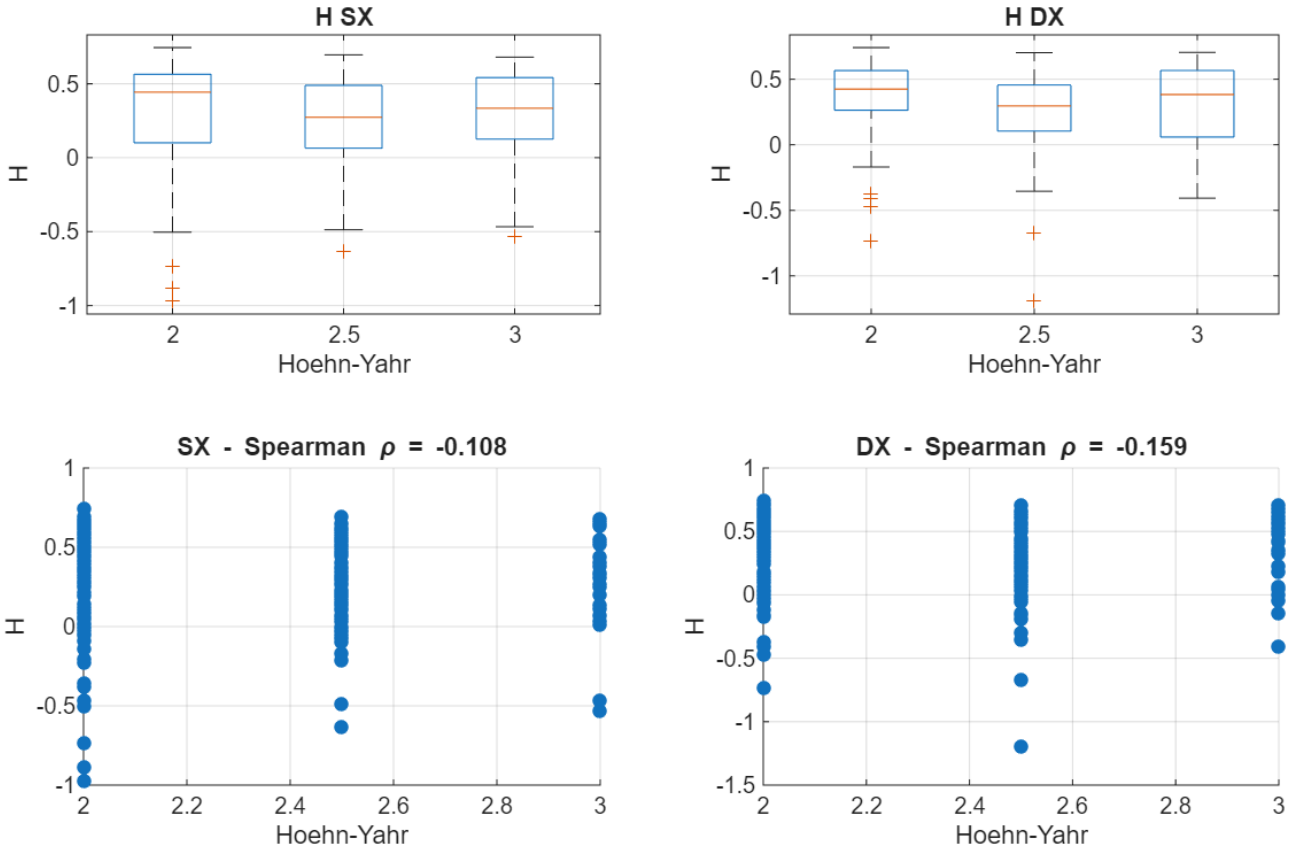


Figura 3.12 – Boxplot dell’esponente di Hurst *H* della FA e relativi diagrammi di correlazione di Spearman in funzione del livello della scala Hoehn–Yahr.

La Tabella 3.1 riassume i principali risultati delle analisi statistiche condotte sui parametri considerati, separatamente per il piede sinistro e per il piede destro.

Parametro	Piede	Mann–Whitney p	ANOVA p	Spearman ρ	Spearman p
Varianza	SX	**$7{,}17 \times 10^{-4}$**	**$8{,}13 \times 10^{-9}$**	0,158	**0,021**
	DX	**$7{,}00 \times 10^{-6}$**	**$7{,}39 \times 10^{-10}$**	0,137	**0,046**
DFA – α	SX	0,787	0,489	0,055	0,423
	DX	0,437	0,624	−0,037	0,587
EDFA – β	SX	0,083	0,956	−0,004	0,959
	DX	0,274	0,422	−0,013	0,847
FA – H	SX	**0,009**	0,504	−0,108	0,116
	DX	0,204	**0,043**	−0,159	**0,020**

Tabella 3.1 – Sintesi dei risultati delle analisi statistiche condotte sui parametri considerati.

Nota. I valori di p inferiori a 0,05, evidenziati in grassetto, indicano risultati statisticamente significativi. SX: piede sinistro; DX: piede destro.

4 – Discussione

4.1 Discussione dei risultati

I risultati ottenuti evidenziano un comportamento differente tra la varianza delle durate della *stance phase* e i parametri derivati dalle analisi frattali. In particolare, la varianza ha mostrato differenze statisticamente significative tra soggetti sani e pazienti con malattia di Parkinson per entrambi i piedi e differenze significative anche nel confronto tra i livelli della scala Hoehn–Yahr. Al contrario, gli esponenti α della DFA e β dell'EDFA non hanno evidenziato differenze statisticamente significative, mentre l'esponente di Hurst H della FA ha fornito risultati significativi soltanto in alcune delle analisi effettuate.

Per quanto riguarda la varianza, i risultati ottenuti indicano che la dispersione delle durate della *stance phase* differisce tra soggetti sani e pazienti con malattia di Parkinson. Tale risultato risulta coerente con quanto riportato in letteratura relativamente all'aumento della variabilità del cammino associato alla malattia. Hausdorff et al. hanno infatti evidenziato una maggiore variabilità passo-passo nei pazienti con malattia di Parkinson rispetto ai soggetti sani [4]. Analogamente, la metanalisi condotta da König et al. ha mostrato un aumento della variabilità in numerosi parametri dell'andatura, sottolineando tuttavia come l'entità delle alterazioni possa dipendere dalla specifica grandezza considerata [5].

La varianza rappresenta inoltre l'unico parametro analizzato per il quale il confronto tra i tre livelli della scala Hoehn–Yahr ha evidenziato differenze statisticamente significative per entrambi i piedi. Questo risultato indica che le medie della varianza differiscono tra almeno due dei tre livelli di gravità considerati. L'analisi di correlazione ha inoltre evidenziato un'associazione positiva statisticamente significativa tra la varianza e il livello Hoehn–Yahr per entrambi i piedi. Tuttavia, i valori dei coefficienti di Spearman ($\rho = 0{,}158$ per il piede sinistro e $\rho = 0{,}137$ per il piede destro) indicano che tali associazioni sono deboli. Pertanto, pur in presenza delle differenze evidenziate dall'ANOVA, la relazione monotona positiva tra gravità della malattia e varianza è di modesta entità.

Risultati differenti sono stati ottenuti mediante la DFA. L'esponente di scala α non ha evidenziato differenze statisticamente significative tra soggetti sani e pazienti con malattia di Parkinson per nessuno dei due piedi. Analogamente, non sono emerse differenze significative tra i livelli Hoehn–Yahr e i coefficienti di correlazione di Spearman sono risultati prossimi a zero. Nell'ambito delle serie temporali delle durate della *stance phase* analizzate, l'esponente α non ha quindi evidenziato una distinzione statisticamente significativa tra i due gruppi né una relazione sistematica con la gravità della malattia.

Tale risultato deve essere considerato alla luce del significato dell'esponente α. A differenza della varianza, che quantifica la dispersione complessiva dei valori, la DFA descrive l'organizzazione temporale delle fluttuazioni e le proprietà di correlazione presenti nella serie alle diverse scale [6]. Due serie temporali possono quindi presentare livelli differenti di variabilità pur mantenendo proprietà di correlazione simili. La presenza di differenze significative nella varianza, accompagnata dall'assenza di differenze significative nell'esponente α, non rappresenta pertanto un risultato contraddittorio, ma evidenzia come i due parametri descrivano aspetti differenti e complementari della dinamica del cammino.

A scopo esemplificativo, la Figura 4.1 mostra l'andamento della funzione di fluttuazione $F(s)$ al variare della scala s per una delle registrazioni analizzate, separatamente per il piede sinistro e per il piede destro. I corrispondenti valori dell'esponente di scala α, stimati mediante regressione lineare sui valori trasformati in scala logaritmica, sono riportati in legenda. La figura consente di osservare concretamente l'andamento della funzione di fluttuazione alle differenti scale considerate nell'analisi.

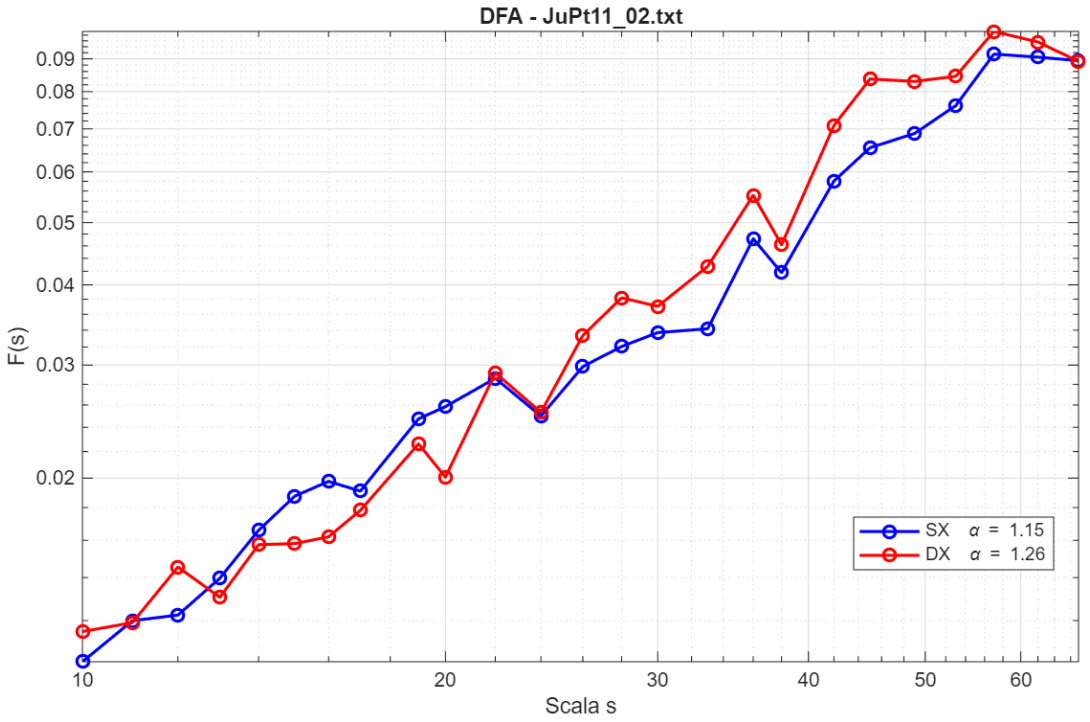


Figura 4.1 – Esempio dell'andamento della funzione di fluttuazione $F(s)$ in funzione della scala s ottenuto mediante DFA per il piede sinistro (SX) e per il piede destro (DX) nella registrazione *JuPt11_02.txt*. I corrispondenti valori dell'esponente di scala α sono riportati in legenda.

L'assenza di differenze significative osservata mediante DFA non coincide pienamente con alcune evidenze riportate in letteratura, nelle quali sono state descritte alterazioni delle correlazioni a lungo termine del cammino nei pazienti con malattia di Parkinson. Hausdorff ha evidenziato come le

proprietà di scala frattale del cammino risultino alterate nella malattia di Parkinson, con una tendenza dell'esponente di scala verso valori inferiori e più prossimi a 0,5, indicativi di un comportamento maggiormente casuale [11]. Warlop et al. hanno inoltre osservato un deterioramento delle correlazioni a lungo termine della durata dello *stride* associato a una maggiore compromissione motoria [12].

Nel confronto con tali studi è tuttavia importante considerare la grandezza analizzata. Le evidenze citate riguardano principalmente serie temporali degli intervalli di *stride*, mentre nel presente lavoro sono state analizzate le durate della *stance phase*. Lo *stride* rappresenta l'intero ciclo compreso tra due contatti iniziali consecutivi dello stesso piede, mentre la *stance phase* ne costituisce soltanto la fase di appoggio. Di conseguenza, le due serie temporali descrivono aspetti differenti del ciclo del passo e possono non presentare la stessa sensibilità alle alterazioni dell'organizzazione temporale del cammino.

Anche l'EDFA non ha evidenziato differenze statisticamente significative tra soggetti sani e pazienti con malattia di Parkinson né tra i diversi livelli della scala Hoehn–Yahr. Inoltre, i coefficienti di correlazione tra l'esponente β e il livello Hoehn–Yahr sono risultati prossimi a zero per entrambi i piedi. L'esponente β, introdotto nell'EDFA, descrive il modo in cui l'eterogeneità delle fluttuazioni locali varia al variare della scala e fornisce pertanto un'informazione complementare rispetto a quella ottenuta mediante l'esponente α della DFA [8]. Nei dati analizzati, tale caratteristica non sembra quindi modificarsi in maniera sistematica né in relazione alla presenza della malattia né al livello della scala Hoehn–Yahr.

L'assenza di risultati significativi sia per α sia per β rafforza la distinzione osservata rispetto alla varianza. Quest'ultima descrive l'entità della dispersione delle durate della *stance phase*, mentre DFA ed EDFA caratterizzano aspetti relativi all'organizzazione temporale e alle proprietà di scala delle fluttuazioni [6, 8]. È quindi possibile osservare differenze nell'ampiezza complessiva delle variazioni senza che queste siano necessariamente accompagnate da modificazioni altrettanto evidenti della struttura temporale della serie.

Un comportamento parzialmente diverso è stato osservato per la FA. L'esponente di Hurst H ha evidenziato una differenza statisticamente significativa tra soggetti sani e pazienti con malattia di Parkinson per il piede sinistro, mentre tale risultato non è stato riscontrato per il piede destro. Il confronto tra i livelli Hoehn–Yahr ha invece mostrato differenze statisticamente significative per il piede destro, ma non per il piede sinistro. L'analisi di Spearman ha infine evidenziato coefficienti negativi per entrambi gli arti. Per il piede destro la correlazione è risultata statisticamente significativa, ma il valore del coefficiente ($\rho = -0{,}159$) indica un'associazione debole; anche per il

piede sinistro il coefficiente è risultato negativo e di debole entità ($\rho = -0{,}108$), senza tuttavia raggiungere il livello di significatività adottato.

Il segno negativo dei coefficienti indica una tendenza di *H* a diminuire all'aumentare del livello Hoehn–Yahr. Nell'ambito della FA, una riduzione dell'esponente di Hurst corrisponde a uno spostamento verso un comportamento meno persistentemente correlato [9]. Tale andamento risulta qualitativamente coerente con le evidenze riportate da Hausdorff e Warlop et al., secondo cui nella malattia di Parkinson le correlazioni a lungo termine del cammino possono risultare ridotte e la dinamica temporale può assumere caratteristiche maggiormente casuali con l'aumentare della compromissione motoria [11, 12].

I risultati ottenuti mediante FA devono tuttavia essere considerati nel loro insieme. La differenza tra soggetti sani e pazienti è risultata significativa solamente per il piede sinistro, mentre sia il confronto tra i livelli Hoehn–Yahr sia l'associazione significativa con tale scala hanno riguardato solamente il piede destro. Non si osserva quindi un comportamento concordante tra i due arti tale da attribuire all'esponente *H* una capacità di distinguere i gruppi altrettanto consistente rispetto a quella mostrata dalla varianza. I risultati indicano piuttosto la presenza di alcune differenze nelle proprietà di correlazione temporale, che tuttavia non si manifestano in maniera uniforme nelle differenti analisi effettuate.

Nel complesso, il confronto tra i parametri considerati mette in evidenza due aspetti distinti della variabilità del cammino. Da un lato, la varianza, quale misura della dispersione delle durate della *stance phase*, ha fornito i risultati più consistenti nel confronto tra soggetti sani e pazienti con malattia di Parkinson. Dall'altro, i parametri derivati dalle analisi frattali hanno mostrato risultati meno uniformi: l'esponente α della DFA e l'esponente β dell'EDFA non hanno evidenziato differenze significative, mentre l'esponente *H* della FA ha mostrato alcune differenze e associazioni significative, ma non concordanti tra piede sinistro e piede destro.

Nel contesto delle serie temporali considerate, i risultati suggeriscono pertanto che le alterazioni associate alla malattia di Parkinson risultino maggiormente evidenti nell'entità della variabilità delle durate della fase di appoggio rispetto alla loro organizzazione temporale. Le misure lineari e frattali non devono tuttavia essere considerate alternative tra loro, poiché descrivono proprietà differenti e potenzialmente complementari del cammino [6]. La loro valutazione congiunta permette quindi di considerare sia la dispersione delle durate della *stance phase* sia le modalità con cui le relative fluttuazioni si organizzano nel tempo.

4.2 Limiti dello studio

I risultati ottenuti devono essere interpretati considerando alcuni limiti metodologici legati sia alle caratteristiche del database sia alle procedure di elaborazione e di analisi adottate. In particolare, assumono rilievo la diversa lunghezza delle serie temporali, la scelta dell'intervallo di scale utilizzato nelle analisi frattali, le operazioni di *pre-processing*, la presenza di più registrazioni riferite allo stesso soggetto e alcuni aspetti relativi alle analisi statistiche.

Un primo limite riguarda la lunghezza delle serie temporali delle durate della *stance phase*. Come descritto nel Capitolo 2, nelle analisi DFA, EDFA e FA la scala minima è stata fissata a 10 elementi, mentre la scala massima è stata definita separatamente per ciascun piede come metà della lunghezza della relativa serie, arrotondata per difetto. Tra questi estremi sono stati inizialmente generati 25 valori di scala distribuiti logaritmicamente, successivamente arrotondati e privati degli eventuali duplicati. Di conseguenza, la lunghezza della serie determina direttamente sia l'estensione dell'intervallo di scale disponibile sia il numero effettivo di punti utilizzati per stimare gli esponenti.

La Figura 4.2 permette di osservare concretamente tale aspetto confrontando due registrazioni caratterizzate da lunghezze molto differenti. Nella registrazione *JuPt05_01.txt* le serie sono costituite da 34 elementi sia per il piede sinistro sia per il piede destro; la scala massima risulta pertanto pari a 17 e, dopo l'arrotondamento e l'eliminazione dei duplicati, l'analisi viene effettuata sulle sole otto scale comprese tra 10 e 17. Nella registrazione *JuPt13_01.txt*, invece, le serie comprendono 145 elementi per il piede sinistro e 140 per il piede destro, con scale massime rispettivamente pari a 72 e 70 e un numero considerevolmente maggiore di scale effettivamente disponibili.

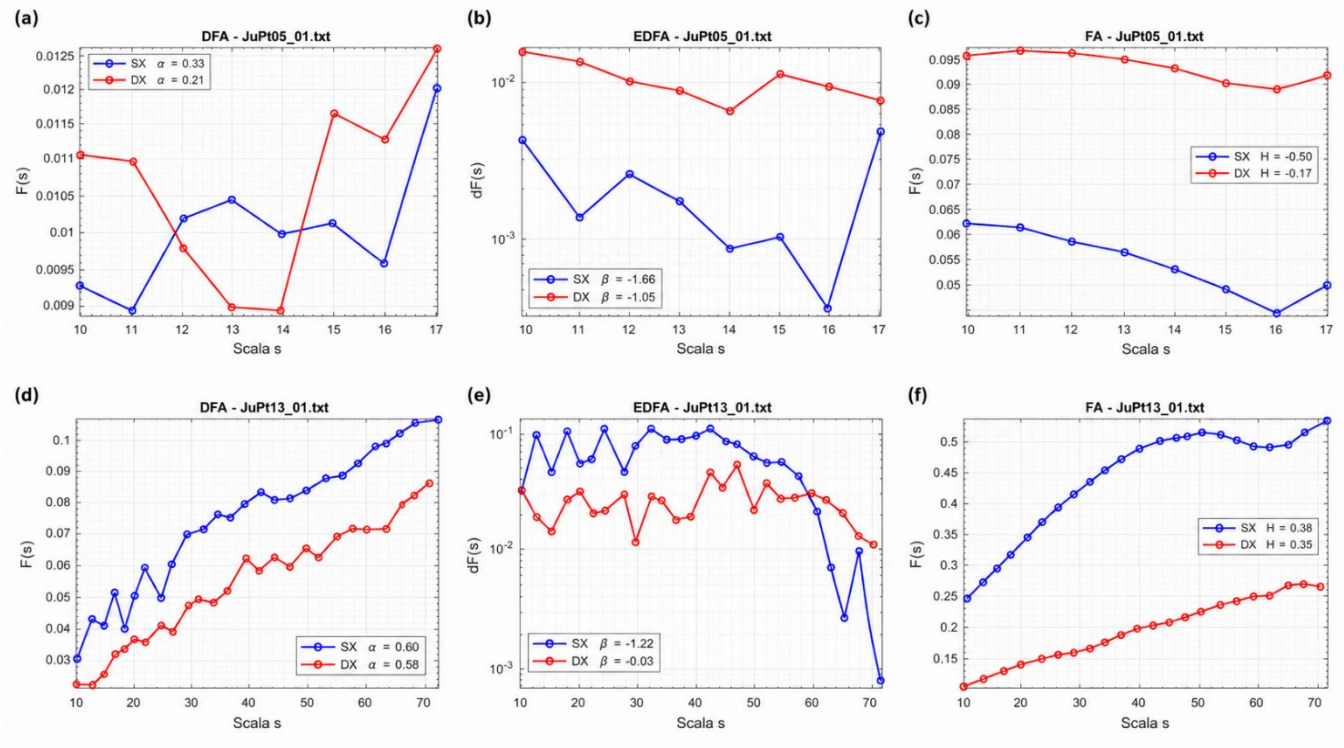


Figura 4.2 – Confronto tra i grafici ottenuti mediante DFA, EDFA e FA per due registrazioni caratterizzate da diversa lunghezza delle serie temporali della *stance phase*. I pannelli (a–c) si riferiscono alla registrazione *JuPt05_01.txt*, costituita da 34 elementi sia per il piede sinistro sia per il piede destro; i pannelli (d–f) si riferiscono alla registrazione *JuPt13_01.txt*, costituita da 145 elementi per il piede sinistro e 140 per il piede destro. Sono riportati, rispettivamente, i risultati della DFA (a, d), dell'EDFA (b, e) e della FA (c, f).

Il confronto mostra come una serie più lunga consenta di valutare le funzioni considerate su un intervallo di scale più esteso e mediante un numero maggiore di punti. Nelle serie più brevi, al contrario, la stima della pendenza nel grafico log-log viene effettuata a partire da un numero ridotto di valori. Questo aspetto può rendere la stima degli esponenti α, β e H maggiormente sensibile alle fluttuazioni dei singoli punti, senza che ciò permetta tuttavia di attribuire eventuali valori anomali esclusivamente alla lunghezza della serie.

Per l'EDFA questo aspetto può assumere un peso ancora maggiore, poiché $dF(s)$ è calcolata come differenza tra il valore massimo e il valore minimo delle fluttuazioni RMS locali. Quando il numero di porzioni della serie disponibili per il calcolo a una determinata scala è ridotto, tale grandezza può risultare maggiormente influenzata dalle singole fluttuazioni osservate.

La sensibilità delle analisi di scala alla numerosità dei dati è stata discussa anche da Bryce e Sprague [9]. Gli autori hanno evidenziato, attraverso l'analisi di serie simulate, la presenza nella DFA di effetti di dimensione finita capaci di introdurre curvature spurie nei *fluctuation plot* e conseguenti distorsioni nella stima dell'esponente. Gli stessi autori hanno inoltre messo in discussione l'idea che il *detrending* locale della DFA possa fornire una protezione generale nei confronti delle non stazionarietà, mostrando come trend non lineari possano influenzare il risultato

dell'analisi. In tale contesto, Bryce e Sprague propongono la FA come approccio più diretto per segnali stazionari, pur riconoscendo che anche l'analisi di serie brevi comporta difficoltà fondamentali nella stima delle proprietà statistiche del processo [9].

Tali considerazioni non implicano che i risultati ottenuti mediante DFA nel presente studio debbano essere considerati non validi, ma evidenziano la necessità di interpretarli con cautela, soprattutto quando le serie disponibili sono brevi. Inoltre, la FA non è immune dagli effetti dovuti alla limitata numerosità dei dati: Bryce e Sprague sottolineano infatti che, anche in questo caso, il ridotto numero di osservazioni disponibili può influenzare la stima della dispersione [9].

Un ulteriore limite riguarda il fatto che gli esponenti α, β e H sono stati stimati mediante un'unica regressione lineare in scala logaritmica, considerando tutte le scale disponibili. Questo approccio assume quindi che l'andamento delle funzioni analizzate possa essere descritto da una singola legge di potenza. Tuttavia, come visibile anche in alcuni grafici della Figura 4.2, l'andamento non risulta sempre perfettamente lineare e può presentare irregolarità. Poiché non è stata effettuata una valutazione specifica della bontà della regressione, gli esponenti ottenuti devono essere considerati come una sintesi complessiva del comportamento della serie nell'intervallo di scale analizzato.

Nel presente studio, in alcune registrazioni, sono state inoltre ottenute stime dell'esponente H esterne all'intervallo canonico compreso tra 0 e 1. Tali valori devono essere interpretati con particolare cautela, poiché possono indicare una limitata stabilità della stima e una non perfetta aderenza dell'andamento osservato a una singola legge di potenza nell'intervallo di scale considerato.

Devono inoltre essere considerate le possibili conseguenze delle procedure di *pre-processing*. La selezione automatica del blocco continuo contenente il maggior numero di eventi IC permette di eliminare pause, interruzioni e porzioni del segnale non riconducibili alla normale deambulazione. Successivamente, i valori mancanti vengono eliminati e gli *outlier* identificati mediante il criterio basato sulla mediana vengono rimossi. Tali operazioni sono necessarie per limitare l'influenza di artefatti e valori non plausibili; tuttavia, possono anche ridurre la lunghezza finale delle serie e, in alcuni casi, eliminare osservazioni estreme che potrebbero rappresentare una componente reale della variabilità patologica. In particolare, un criterio automatico potrebbe non consentire di distinguere sempre una pausa prolungata o un'alterazione marcata del passo da un'interruzione o da un artefatto della registrazione.

Un aspetto particolarmente rilevante riguarda inoltre la struttura del database. Il numero di registrazioni è superiore al numero dei soggetti, poiché uno stesso partecipante può essere associato a più prove di cammino. Nel presente lavoro ciascuna registrazione è stata elaborata separatamente e i relativi parametri sono stati inseriti come singole osservazioni nella tabella utilizzata per le

analisi statistiche. Di conseguenza, registrazioni appartenenti allo stesso soggetto possono contribuire più volte ai confronti statistici. In tali condizioni, l'assunzione di indipendenza tra tutte le osservazioni non è completamente garantita, poiché misure provenienti dallo stesso individuo possono essere tra loro correlate. Questo aspetto deve quindi essere considerato nell'interpretazione dei risultati statistici e, in particolare, dei livelli di significatività ottenuti.

A ciò si aggiunge l'eterogeneità delle condizioni sperimentali. Nel database sono infatti presenti differenti prove di cammino e, nello studio identificato dal prefisso "Ga", le registrazioni con suffisso "_10" corrispondono a prove in *dual task*, nelle quali il soggetto cammina svolgendo contemporaneamente un compito cognitivo. Poiché il tipo di compito può modificare l'andatura e la sua organizzazione temporale, l'analisi congiunta di prove svolte in condizioni differenti può introdurre una fonte aggiuntiva di variabilità non direttamente riconducibile alla presenza o alla gravità della malattia.

Anche la caratterizzazione della gravità clinica presenta alcune limitazioni. Sebbene la scala Hoehn–Yahr comprenda un intervallo più ampio di livelli, nel database analizzato sono rappresentati esclusivamente i livelli 2, 2,5 e 3. L'analisi è pertanto circoscritta a una porzione relativamente limitata della progressione clinica della malattia e non consente di stabilire se i parametri considerati presentino andamenti analoghi negli stadi iniziali o più avanzati.

Infine, devono essere considerati alcuni aspetti relativi all'analisi statistica. Sono stati effettuati numerosi confronti, considerando quattro parametri, entrambi i piedi e tre differenti approcci statistici. Non è stata applicata una correzione per confronti multipli; di conseguenza, all'aumentare del numero complessivo di test cresce anche la probabilità di ottenere almeno un risultato statisticamente significativo per effetto del caso. Questo elemento assume particolare rilievo per i risultati significativi caratterizzati da *p-value* relativamente prossimi alla soglia di 0,05.

Per il confronto tra i livelli Hoehn–Yahr è stata inoltre utilizzata un'ANOVA a una via. Il codice implementato non comprende verifiche esplicite delle assunzioni relative alla normalità dei residui e all'omogeneità delle varianze, né sono stati eseguiti confronti *post-hoc* nei casi in cui l'ANOVA è risultata significativa. Di conseguenza, per la varianza di entrambi i piedi e per l'esponente H del piede destro, i risultati consentono soltanto di affermare che almeno due dei tre livelli Hoehn–Yahr presentano medie differenti, senza stabilire quali coppie di livelli differiscano specificamente tra loro.

Nel complesso, tali limitazioni non annullano le evidenze emerse dall'analisi, ma definiscono il contesto entro cui devono essere interpretate. In particolare, la variabilità della lunghezza delle serie, la stabilità delle stime degli esponenti, la possibile dipendenza tra registrazioni dello stesso soggetto e l'eterogeneità delle condizioni sperimentali rappresentano elementi che possono aver

contribuito alla dispersione dei risultati e alla mancata significatività osservata per alcuni dei parametri frattali. Questi aspetti costituiscono inoltre il punto di partenza naturale per gli sviluppi futuri descritti nel paragrafo successivo.

4.3 Sviluppi futuri

Gli sviluppi futuri del presente lavoro potrebbero riguardare innanzitutto l'analisi di serie temporali più lunghe e caratterizzate da una maggiore uniformità nella durata delle registrazioni, così da ampliare l'intervallo di scale disponibile e ottenere stime più stabili degli esponenti α, β e H. Sarebbe inoltre utile approfondire la scelta dell'intervallo di scale e valutare quantitativamente la bontà delle regressioni utilizzate per la determinazione degli esponenti.

Un ulteriore sviluppo potrebbe riguardare la struttura statistica dei dati. Poiché alcuni soggetti sono associati a più registrazioni, studi successivi potrebbero adottare procedure in grado di tenere conto delle misure ripetute oppure ricavare, per ciascun soggetto, valori rappresentativi dei parametri analizzati.

Potrebbe inoltre risultare utile analizzare separatamente le diverse condizioni sperimentali presenti nel database, distinguendo le prove di cammino semplice da quelle svolte in condizioni di *dual task*, e ampliare il campione includendo pazienti appartenenti a un intervallo più esteso della scala Hoehn–Yahr.

Infine, studi futuri potrebbero affiancare agli indici già considerati ulteriori misure della complessità e dell'organizzazione temporale del cammino, al fine di verificare se approcci differenti siano in grado di evidenziare alterazioni non rilevate mediante DFA, EDFA e FA.

5 – Conclusioni

Il presente lavoro ha avuto l'obiettivo di analizzare la complessità delle serie temporali associate al cammino in pazienti con malattia di Parkinson, considerando in particolare le durate della *stance phase* ricavate dalle registrazioni della forza verticale totale. A partire dai segnali grezzi disponibili nel database *PhysioNet*, sono state estratte le fasi di appoggio dei due piedi e costruite le relative serie temporali, successivamente sottoposte a procedure di *pre-processing* e analizzate mediante misure lineari e frattali. Per ciascuna registrazione sono stati considerati la varianza delle durate della *stance phase*, l'esponente di scala α della DFA, l'esponente β dell'EDFA e l'esponente di Hurst H della FA.

Le analisi statistiche hanno evidenziato un comportamento differente tra la varianza e i parametri derivati dalle analisi frattali. La varianza ha mostrato differenze statisticamente significative tra soggetti sani e pazienti con malattia di Parkinson per entrambi i piedi e differenze significative anche nel confronto tra i livelli della scala Hoehn–Yahr. L'associazione con la gravità della malattia è risultata positiva e statisticamente significativa per entrambi i piedi, ma di debole entità.

Per quanto riguarda le analisi frattali, gli esponenti α della DFA e β dell'EDFA non hanno evidenziato differenze statisticamente significative né tra soggetti sani e pazienti con malattia di Parkinson né tra i differenti livelli Hoehn–Yahr. Anche le correlazioni con la gravità della malattia sono risultate prossime a zero. L'esponente di Hurst H ricavato mediante FA ha invece mostrato alcuni risultati significativi, ma non concordanti tra piede sinistro e piede destro: la differenza tra soggetti sani e pazienti è risultata significativa soltanto per il piede sinistro, mentre il confronto tra i livelli Hoehn–Yahr e la correlazione con tale scala sono risultati significativi soltanto per il piede destro.

Nel complesso, i risultati ottenuti suggeriscono che, nelle serie temporali analizzate, le alterazioni associate alla malattia di Parkinson siano risultate più evidenti nell'entità della variabilità delle durate della fase di appoggio rispetto alle proprietà di organizzazione temporale descritte dai parametri frattali. La varianza è infatti risultata il parametro che ha evidenziato in modo più coerente differenze tra i gruppi considerati, mentre DFA, EDFA e FA hanno fornito risultati meno uniformi.

Tali risultati non implicano tuttavia che le misure frattali siano prive di utilità. Le misure lineari e quelle basate sulle proprietà di scala descrivono infatti aspetti differenti della dinamica del cammino e possono quindi fornire informazioni complementari. L'analisi congiunta della dispersione delle durate della *stance phase* e della loro organizzazione temporale rappresenta pertanto un approccio

utile per caratterizzare in maniera più completa le modificazioni del cammino associate alla malattia di Parkinson.

L'interpretazione dei risultati deve infine tenere conto dei principali limiti metodologici individuati, tra cui la diversa lunghezza delle serie temporali, la presenza di più registrazioni riferite allo stesso soggetto, l'eterogeneità delle condizioni sperimentali e la limitata rappresentazione dei livelli Hoehn–Yahr. Studi futuri condotti su serie più lunghe e maggiormente omogenee, con una gestione specifica delle misure ripetute e delle diverse condizioni di cammino, potranno contribuire a chiarire ulteriormente il ruolo delle misure frattali nella caratterizzazione delle alterazioni del cammino associate alla malattia di Parkinson.

6 – Bibliografia

[1] L. V. Kalia e A. E. Lang, "Parkinson's disease," *The Lancet*, vol. 386, n. 9996, pp. 896–912, 2015, doi: 10.1016/S0140-6736(14)61393-3.

[2] J. M. Hausdorff, "Gait in Parkinson's Disease," *PhysioNet*, version 1.0.0, 2008, doi: 10.13026/C24H3N.

[3] W. Pirker e R. Katzenschlager, "Gait disorders in adults and the elderly: A clinical guide," *Wiener klinische Wochenschrift*, vol. 129, n. 3–4, pp. 81–95, 2017, doi: 10.1007/s00508-016-1096-4.

[4] J. M. Hausdorff, M. E. Cudkowicz, R. Firtion, J. Y. Wei e A. L. Goldberger, "Gait variability and basal ganglia disorders: Stride-to-stride variations of gait cycle timing in Parkinson's disease and Huntington's disease," *Movement Disorders*, vol. 13, n. 3, pp. 428–437, 1998, doi: 10.1002/mds.870130310.

[5] N. König, N. B. Singh, C. R. Baumann e W. R. Taylor, "Can gait signatures provide quantitative measures for aiding clinical decision-making? A systematic meta-analysis of gait variability behavior in patients with Parkinson's disease," *Frontiers in Human Neuroscience*, vol. 10, art. 319, 2016, doi: 10.3389/fnhum.2016.00319.

[6] A. Phinyomark, R. Larracy e E. Scheme, "Fractal analysis of human gait variability via stride interval time series," *Frontiers in Physiology*, vol. 11, art. 333, 2020, doi: 10.3389/fphys.2020.00333.

[7] E. A. F. Ihlen, "Introduction to multifractal detrended fluctuation analysis in Matlab," *Frontiers in Physiology*, vol. 3, art. 141, 2012, doi: 10.3389/fphys.2012.00141.

[8] A. N. Pavlov, O. N. Pavlova, O. V. Semyachkina-Glushkovskaya e J. Kurths, "Extended detrended fluctuation analysis: Effects of nonstationarity and application to sleep data," *The European Physical Journal Plus*, vol. 136, art. 10, 2021, doi: 10.1140/epjp/s13360-020-00980-x.

[9] R. M. Bryce e K. B. Sprague, "Revisiting detrended fluctuation analysis," *Scientific Reports*, vol. 2, art. 315, 2012, doi: 10.1038/srep00315.

[10] S. Yano, A. Nakamura, Y. Suzuki, C. E. Smith e T. Nomura, "Smartphone usage during walking decreases the positive persistency in gait cycle variability," *Scientific Reports*, vol. 14, art. 16410, 2024, doi: 10.1038/s41598-024-66727-1.

[11] J. M. Hausdorff, "Gait dynamics in Parkinson's disease: Common and distinct behavior among stride length, gait variability, and fractal-like scaling," *Chaos*, vol. 19, n. 2, art. 026113, 2009, doi: 10.1063/1.3147408.

[12] T. Warlop, C. Detrembleur, B. Bollens, G. Stoquart, F. Crevecoeur, A. Jeanjean e T. M. Lejeune, "Temporal organization of stride duration variability as a marker of gait instability in Parkinson's disease," *Journal of Rehabilitation Medicine*, vol. 48, n. 10, pp. 865–871, 2016, doi: 10.2340/16501977-2158.

[13] A. L. Goldberger, L. A. N. Amaral, L. Glass, J. M. Hausdorff, P. C. Ivanov, R. G. Mark, J. E. Mietus, G. B. Moody, C.-K. Peng e H. E. Stanley, "PhysioBank, PhysioToolkit, and PhysioNet: Components of a new research resource for complex physiologic signals," *Circulation*, vol. 101, n. 23, pp. e215–e220, 2000, doi: 10.1161/01.CIR.101.23.e215.

[14] T. Pollard, B. E. Moody, L.-W. H. Lehman, B. J. Gow, C. Fernandes, C. Xie, A. Johnson, R. G. Mark e T. Heldt, "PhysioNet as a global platform for biomedical research," *Nature Health*, vol. 1, n. 8, pp. 792–795, 2026, doi: 10.1038/s44360-026-00096-z.

[15] G. Yogev, N. Giladi, C. Peretz, S. Springer, E. S. Simon e J. M. Hausdorff, "Dual tasking, gait rhythmicity, and Parkinson's disease: Which aspects of gait are attention demanding?" *European Journal of Neuroscience*, vol. 22, n. 5, pp. 1248–1256, 2005, doi: 10.1111/j.1460-9568.2005.04298.x.

[16] J. M. Hausdorff, J. Lowenthal, T. Herman, L. Gruendlinger, C. Peretz e N. Giladi, "Rhythmic auditory stimulation modulates gait variability in Parkinson's disease," *European Journal of Neuroscience*, vol. 26, n. 8, pp. 2369–2375, 2007, doi: 10.1111/j.1460-9568.2007.05810.x.

[17] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger e J. M. Hausdorff, "Effect of gait speed on gait rhythmicity in Parkinson's disease: Variability of stride time and swing time respond differently," *Journal of NeuroEngineering and Rehabilitation*, vol. 2, art. 23, 2005, doi: 10.1186/1743-0003-2-23.

[18] S. Frenkel-Toledo, N. Giladi, C. Peretz, T. Herman, L. Gruendlinger e J. M. Hausdorff, "Treadmill walking as an external pacemaker to improve gait rhythm and stability in Parkinson's disease," *Movement Disorders*, vol. 20, n. 9, pp. 1109–1114, 2005, doi: 10.1002/mds.20507.

[19] J. M. Hausdorff, Z. Ladin e J. Y. Wei, "Footswitch system for measurement of the temporal parameters of gait," *Journal of Biomechanics*, vol. 28, n. 3, pp. 347–351, 1995, doi: 10.1016/0021-9290(94)00074-E.

[20] P. E. McKnight e J. Najab, "Mann-Whitney U test," in *The Corsini Encyclopedia of Psychology*, I. B. Weiner e W. E. Craighead, Eds., John Wiley & Sons, 2010, doi: 10.1002/9780470479216.corpsy0524.

[21] T. K. Kim, "Understanding one-way ANOVA using conceptual figures," *Korean Journal of Anesthesiology*, vol. 70, n. 1, pp. 22–26, 2017, doi: 10.4097/kjae.2017.70.1.22.

[22] H.-Y. Kim, "Statistical notes for clinical researchers: Covariance and correlation," *Restorative Dentistry & Endodontics*, vol. 43, n. 1, art. e4, 2018, doi: 10.5395/rde.2018.43.e4.

Appendice – Codice MATLAB

Nella presente appendice sono riportate le principali porzioni di codice MATLAB sviluppate per l'elaborazione dei segnali e l'analisi delle serie temporali considerate nel presente lavoro. Sono stati selezionati i frammenti ritenuti maggiormente significativi ai fini della descrizione e della riproducibilità della procedura, accompagnati, ove opportuno, da brevi commenti esplicativi. Il codice comprende le principali fasi di estrazione della *stance phase*, *pre-processing*, calcolo dei parametri e analisi statistica descritte nei capitoli precedenti.

```
clc;
clear;
close all;

%% ========================================================================
% CONFIGURAZIONE DEL DATABASE
% ========================================================================

% Percorso della cartella contenente il database
folder = '...\gait-in-parkinsons-disease-1.0.0';

% Studi considerati nell'analisi
groups = {'Ga', 'Ju', 'Si'};

% Struttura destinata alla raccolta dei risultati
all_results = {};


%% ========================================================================
% ANALISI DELLE REGISTRAZIONI
% ========================================================================

for g = 1:length(groups)

    groupName = groups{g};

    % Individuazione dei file appartenenti allo studio corrente
    files = dir(fullfile(folder, [groupName '\*.txt']));

    % Registrazioni escluse dall'analisi
    excluded_files = {};

    if strcmp(groupName, 'Ga')
        excluded_files = {'GaCo16_10.txt'};
    end

    for k = 1:length(files)

        filename = files(k).name;

        % Esclusione delle registrazioni considerate non idonee
        if ismember(filename, excluded_files)
            continue;
        end


        %% ------------------------------------------------------------------
        % CARICAMENTO DEI DATI
        % ------------------------------------------------------------------

        data = load(fullfile(folder, filename));

        % Colonna 1: tempo
```

```
% Colonna 18: forza verticale totale del piede sinistro
% Colonna 19: forza verticale totale del piede destro
t    = data(:,1);
F_sx = data(:,18);
F_dx = data(:,19);


%% ---------------------------------------------------------------
% INDIVIDUAZIONE PRELIMINARE DELLE STANCE PHASE
% ---------------------------------------------------------------

% La soglia viene posta al 5% del valore massimo della forza
% registrata separatamente per ciascun piede
threshold_sx = 0.05 * max(F_sx);
threshold_dx = 0.05 * max(F_dx);

% Estrazione preliminare degli Initial Contact del piede sinistro
% per l'individuazione delle eventuali interruzioni del cammino
[~, IC_sx_tmp, ~] = ...
    local_compute(F_sx, t, threshold_sx);


%% ---------------------------------------------------------------
% IDENTIFICAZIONE DEL TRATTO CONTINUO DI CAMMINO
% ---------------------------------------------------------------

% Un intervallo tra due Initial Contact superiore a 3 s viene
% considerato indicativo di un'interruzione della camminata
max_gap = 3;

step_intervals = diff(IC_sx_tmp);
break_idx = find(step_intervals > max_gap);

% Definizione degli estremi dei diversi blocchi continui
start_idx = [1; break_idx + 1];
end_idx   = [break_idx; length(IC_sx_tmp)];

% Numero di intervalli presenti in ciascun blocco
block_lengths = end_idx - start_idx;

% Selezione del blocco contenente il maggior numero di eventi IC
[~, best_block] = max(block_lengths);

idx_start = start_idx(best_block);
idx_end   = end_idx(best_block);

% Istanti iniziale e finale del tratto selezionato
t_start = IC_sx_tmp(idx_start);
t_end   = IC_sx_tmp(idx_end);


%% ---------------------------------------------------------------
% SELEZIONE AUTOMATICA DEL TRATTO VALIDO
% ---------------------------------------------------------------

valid_idx = t >= t_start & t <= t_end;

t    = t(valid_idx);
F_sx = F_sx(valid_idx);
F_dx = F_dx(valid_idx);


%% ---------------------------------------------------------------
% ESTRAZIONE DEFINITIVA DELLE STANCE PHASE
% ---------------------------------------------------------------

% Ricalcolo delle soglie sul solo tratto selezionato
threshold_sx = 0.05 * max(F_sx);
threshold_dx = 0.05 * max(F_dx);
```

```
% Estrazione degli eventi Initial Contact ed End Contact
% e calcolo della durata delle stance phase
[stance_sx, IC_sx, EC_sx] = ...
    local_compute(F_sx, t, threshold_sx);

[stance_dx, IC_dx, EC_dx] = ...
    local_compute(F_dx, t, threshold_dx);


%% ---------------------------------------------------------------
% PRE-PROCESSING DELLE SERIE TEMPORALI
% ---------------------------------------------------------------

% Rimozione degli eventuali valori mancanti
stance_sx = stance_sx(~isnan(stance_sx));
stance_dx = stance_dx(~isnan(stance_dx));

% Identificazione degli outlier mediante il metodo basato
% sulla mediana e sulla Median Absolute Deviation (MAD)
out_sx = isoutlier(stance_sx, 'median');
out_dx = isoutlier(stance_dx, 'median');

% Rimozione degli outlier
stance_sx = stance_sx(~out_sx);
stance_dx = stance_dx(~out_dx);


%% ---------------------------------------------------------------
% CALCOLO DELLA VARIANZA DELLA STANCE PHASE
% ---------------------------------------------------------------

var_sx = var(stance_sx);
var_dx = var(stance_dx);

% Statistiche descrittive
mean_sx = mean(stance_sx);
std_sx  = std(stance_sx);

mean_dx = mean(stance_dx);
std_dx  = std(stance_dx);


%% ---------------------------------------------------------------
% DEFINIZIONE DELLE SCALE DI ANALISI
% ---------------------------------------------------------------

% Scala minima
minScale = 10;

% Scala massima pari alla metà della lunghezza della serie
maxScale_sx = floor(length(stance_sx) / 2);
maxScale_dx = floor(length(stance_dx) / 2);

% Numero di scale inizialmente generate
numScales = 25;


%% ---------------------------------------------------------------
% CALCOLO DEI PARAMETRI DFA ED EDFA
% ---------------------------------------------------------------

% Piede sinistro
[alpha_sx, beta_sx, F_sx_dfa, dF_sx, s_sx] = ...
    edfa(stance_sx, ...
         minScale, ...
         maxScale_sx, ...
         numScales);
```

```
% Piede destro
[alpha_dx, beta_dx, F_dx_dfa, dF_dx, s_dx] = ...
    edfa(stance_dx, ...
         minScale, ...
         maxScale_dx, ...
         numScales);


%% ---------------------------------------------------------------
% CALCOLO DEL PARAMETRO FA
% ---------------------------------------------------------------

% Piede sinistro
[H_sx, F_FA_sx, s_FA_sx] = ...
    fa(stance_sx, ...
       minScale, ...
       maxScale_sx, ...
       numScales);

% Piede destro
[H_dx, F_FA_dx, s_FA_dx] = ...
    fa(stance_dx, ...
       minScale, ...
       maxScale_dx, ...
       numScales);


%% ---------------------------------------------------------------
% IDENTIFICAZIONE DEL GRUPPO CLINICO
% ---------------------------------------------------------------

if contains(filename, 'Co')

    clinicalGroup = 'Co';

elseif contains(filename, 'Pt')

    clinicalGroup = 'Pt';

else

    clinicalGroup = 'Unknown';

end


%% ---------------------------------------------------------------
% MEMORIZZAZIONE DEI RISULTATI
% ---------------------------------------------------------------

all_results = [all_results;
    {groupName, ...
     clinicalGroup, ...
     filename, ...
     mean_sx, ...
     std_sx, ...
     var_sx, ...
     alpha_sx, ...
     beta_sx, ...
     H_sx, ...
     length(stance_sx), ...
     mean_dx, ...
     std_dx, ...
     var_dx, ...
     alpha_dx, ...
     beta_dx, ...
     H_dx, ...
     length(stance_dx)}];
```

```
    end

end


%% ======================================================================
% CREAZIONE DELLA TABELLA COMPLESSIVA DEI RISULTATI
% ======================================================================

results_table = cell2table( ...
    all_results, ...
    'VariableNames', ...
    {'Study', ...
     'ClinicalGroup', ...
     'File', ...
     'MeanStance_sx', ...
     'StdStance_sx', ...
     'Variance_sx', ...
     'Alpha_sx', ...
     'Beta_sx', ...
     'H_sx', ...
     'Nsteps_sx', ...
     'MeanStance_dx', ...
     'StdStance_dx', ...
     'Variance_dx', ...
     'Alpha_dx', ...
     'Beta_dx', ...
     'H_dx', ...
     'Nsteps_dx'});


%% ======================================================================
% ANALISI STATISTICA: CONTROLLI VS PAZIENTI CON PARKINSON
% ======================================================================

% ----------------------------------------------------------------------
% Estrazione dei valori di alpha
% ----------------------------------------------------------------------

alphaCo_sx = results_table.Alpha_sx( ...
    strcmp(results_table.ClinicalGroup, 'Co'));

alphaPt_sx = results_table.Alpha_sx( ...
    strcmp(results_table.ClinicalGroup, 'Pt'));

alphaCo_dx = results_table.Alpha_dx( ...
    strcmp(results_table.ClinicalGroup, 'Co'));

alphaPt_dx = results_table.Alpha_dx( ...
    strcmp(results_table.ClinicalGroup, 'Pt'));


% ----------------------------------------------------------------------
% Estrazione dei valori di beta
% ----------------------------------------------------------------------

betaCo_sx = results_table.Beta_sx( ...
    strcmp(results_table.ClinicalGroup, 'Co'));

betaPt_sx = results_table.Beta_sx( ...
    strcmp(results_table.ClinicalGroup, 'Pt'));

betaCo_dx = results_table.Beta_dx( ...
    strcmp(results_table.ClinicalGroup, 'Co'));

betaPt_dx = results_table.Beta_dx( ...
    strcmp(results_table.ClinicalGroup, 'Pt'));
```

```
% ------------------------------------------------------------------------
% Estrazione dei valori di H
% ------------------------------------------------------------------------

HCo_sx = results_table.H_sx( ...
    strcmp(results_table.ClinicalGroup, 'Co'));

HPt_sx = results_table.H_sx( ...
    strcmp(results_table.ClinicalGroup, 'Pt'));

HCo_dx = results_table.H_dx( ...
    strcmp(results_table.ClinicalGroup, 'Co'));

HPt_dx = results_table.H_dx( ...
    strcmp(results_table.ClinicalGroup, 'Pt'));


% ------------------------------------------------------------------------
% Estrazione dei valori della varianza
% ------------------------------------------------------------------------

varCo_sx = results_table.Variance_sx( ...
    strcmp(results_table.ClinicalGroup, 'Co'));

varPt_sx = results_table.Variance_sx( ...
    strcmp(results_table.ClinicalGroup, 'Pt'));

varCo_dx = results_table.Variance_dx( ...
    strcmp(results_table.ClinicalGroup, 'Co'));

varPt_dx = results_table.Variance_dx( ...
    strcmp(results_table.ClinicalGroup, 'Pt'));


%% ------------------------------------------------------------------------
% TEST STATISTICO DI MANN-WHITNEY
% ------------------------------------------------------------------------

% DFA
[pa_sx, ha_sx] = ranksum(alphaCo_sx, alphaPt_sx);
[pa_dx, ha_dx] = ranksum(alphaCo_dx, alphaPt_dx);

% EDFA
[pb_sx, hb_sx] = ranksum(betaCo_sx, betaPt_sx);
[pb_dx, hb_dx] = ranksum(betaCo_dx, betaPt_dx);

% FA
[pH_sx, hH_sx] = ranksum(HCo_sx, HPt_sx);
[pH_dx, hH_dx] = ranksum(HCo_dx, HPt_dx);

% Varianza
[pv_sx, hv_sx] = ranksum(varCo_sx, varPt_sx);
[pv_dx, hv_dx] = ranksum(varCo_dx, varPt_dx);


%% ------------------------------------------------------------------------
% STAMPA DEI RISULTATI DEL TEST DI MANN-WHITNEY
% ------------------------------------------------------------------------

fprintf('\n');
fprintf('============================================\n');
fprintf('TEST DI MANN-WHITNEY: CONTROLLI VS PARKINSON\n');
fprintf('============================================\n');

fprintf('Alpha SX:    p = %.6f\n', pa_sx);
fprintf('Alpha DX:    p = %.6f\n', pa_dx);

fprintf('Beta SX:     p = %.6f\n', pb_sx);
fprintf('Beta DX:     p = %.6f\n', pb_dx);
```

```
fprintf('H SX:        p = %.6f\n', pH_sx);
fprintf('H DX:        p = %.6f\n', pH_dx);

fprintf('Varianza SX: p = %.6f\n', pv_sx);
fprintf('Varianza DX: p = %.6f\n', pv_dx);


%% ------------------------------------------------------------------------
% ESEMPIO DI RAPPRESENTAZIONE MEDIANTE BOXPLOT
% ------------------------------------------------------------------------

% Il seguente esempio mostra il confronto della varianza relativa
% al piede sinistro tra controlli e pazienti con Parkinson.
% La medesima procedura è stata applicata agli altri parametri
% e al piede destro.

figure;

boxplot( ...
    [varCo_sx; varPt_sx], ...
    [repmat({'Co'}, length(varCo_sx), 1); ...
     repmat({'Pt'}, length(varPt_sx), 1)]);

ylabel('Varianza [s^2]', 'Interpreter', 'tex');
title('Varianza SX: Co vs Pt');
grid on;


%% ========================================================================
% ASSOCIAZIONE DEI RISULTATI ALLE INFORMAZIONI CLINICHE
% ========================================================================

% Lettura del file contenente i dati clinici
demo = readtable( ...
    fullfile(folder, 'demographics.xls'));

% Copia della tabella dei risultati
results = results_table;


%% ------------------------------------------------------------------------
% IDENTIFICAZIONE DELLO STUDIO DI APPARTENENZA
% ------------------------------------------------------------------------

% Le prime due lettere del nome del file identificano lo studio
results.Study = extractBefore(results.File, 3);


%% ------------------------------------------------------------------------
% IDENTIFICAZIONE DI PAZIENTI E CONTROLLI
% ------------------------------------------------------------------------

results.Group = strings(height(results), 1);

for i = 1:height(results)

    if contains(results.File{i}, 'Pt')

        results.Group(i) = "PD";

    elseif contains(results.File{i}, 'Co')

        results.Group(i) = "CO";

    end

end
```

```
%% ------------------------------------------------------------------------
% ESTRAZIONE DEL NUMERO IDENTIFICATIVO DEL SOGGETTO
% ------------------------------------------------------------------------

results.Subjnum = zeros(height(results), 1);

for i = 1:height(results)

    fname = results.File{i};

    % Individuazione del numero successivo alle sigle Pt o Co
    token = regexp( ...
        fname, ...
        '(Pt|Co)(\d+)', ...
        'tokens');

    results.Subjnum(i) = ...
        str2double(token{1}{2});

end


%% ------------------------------------------------------------------------
% ASSOCIAZIONE TRA RISULTATI E DATI CLINICI
% ------------------------------------------------------------------------

% Le variabili usate come chiavi devono avere lo stesso tipo
results.Study = string(results.Study);
demo.Study    = string(demo.Study);
results.Group = string(results.Group);
demo.Group    = string(demo.Group);

% Associazione tra registrazioni e dati clinici
merged = innerjoin( ...
    results, ...
    demo, ...
    'Keys', {'Study', 'Group', 'Subjnum'});

% Verifica dell'abbinamento
assert(height(merged) == height(results), ...
    'Controllare gli abbinamenti con i dati clinici');

% Selezione dei pazienti con malattia di Parkinson
PD = merged(merged.Group == "PD", :);

assert(all(string(PD.ClinicalGroup) == "Pt"), ...
    'Nel gruppo PD è presente una registrazione di controllo');

%% ------------------------------------------------------------------------
% SUDDIVISIONE IN BASE ALLA SCALA HOEHN-YAHR
% ------------------------------------------------------------------------

HY2  = PD(PD.HoehnYahr == 2,   :);
HY25 = PD(PD.HoehnYahr == 2.5, :);
HY3  = PD(PD.HoehnYahr == 3,   :);


%% ========================================================================
% ANOVA A UNA VIA
% ========================================================================

% ------------------------------------------------------------------------
% DFA
% ------------------------------------------------------------------------

pa_sx = anova1( ...
    PD.Alpha_sx, ...
    PD.HoehnYahr, ...
```

```
    'off');

pa_dx = anova1( ...
    PD.Alpha_dx, ...
    PD.HoehnYahr, ...
    'off');


% ------------------------------------------------------------------------
% EDFA
% ------------------------------------------------------------------------

pb_sx = anova1( ...
    PD.Beta_sx, ...
    PD.HoehnYahr, ...
    'off');

pb_dx = anova1( ...
    PD.Beta_dx, ...
    PD.HoehnYahr, ...
    'off');


% ------------------------------------------------------------------------
% FA
% ------------------------------------------------------------------------

pH_sx = anova1( ...
    PD.H_sx, ...
    PD.HoehnYahr, ...
    'off');

pH_dx = anova1( ...
    PD.H_dx, ...
    PD.HoehnYahr, ...
    'off');


% ------------------------------------------------------------------------
% Varianza
% ------------------------------------------------------------------------

pv_sx = anova1( ...
    PD.Variance_sx, ...
    PD.HoehnYahr, ...
    'off');

pv_dx = anova1( ...
    PD.Variance_dx, ...
    PD.HoehnYahr, ...
    'off');


%% ------------------------------------------------------------------------
% STAMPA DEI RISULTATI DELL'ANOVA
% ------------------------------------------------------------------------

fprintf('\n');
fprintf('============================================\n');
fprintf('ANOVA A UNA VIA: CLASSI HOEHN-YAHR\n');
fprintf('============================================\n');

fprintf('Alpha SX:    p = %.9f\n', pa_sx);
fprintf('Alpha DX:    p = %.9f\n', pa_dx);

fprintf('Beta SX:     p = %.9f\n', pb_sx);
fprintf('Beta DX:     p = %.9f\n', pb_dx);

fprintf('H SX:        p = %.9f\n', pH_sx);
```

```
fprintf('H DX:        p = %.9f\n', pH_dx);

fprintf('Varianza SX: p = %.3e\n', pv_sx);
fprintf('Varianza DX: p = %.3e\n', pv_dx);


%% ========================================================================
% CORRELAZIONE DI SPEARMAN CON LA SCALA HOEHN-YAHR
% ========================================================================

% ------------------------------------------------------------------------
% DFA
% ------------------------------------------------------------------------

[rhoalpha_sx, pcorralpha_sx] = corr( ...
    PD.HoehnYahr, ...
    PD.Alpha_sx, ...
    'Type', 'Spearman');

[rhoalpha_dx, pcorralpha_dx] = corr( ...
    PD.HoehnYahr, ...
    PD.Alpha_dx, ...
    'Type', 'Spearman');


% ------------------------------------------------------------------------
% EDFA
% ------------------------------------------------------------------------

[rhobeta_sx, pcorrbeta_sx] = corr( ...
    PD.HoehnYahr, ...
    PD.Beta_sx, ...
    'Type', 'Spearman');

[rhobeta_dx, pcorrbeta_dx] = corr( ...
    PD.HoehnYahr, ...
    PD.Beta_dx, ...
    'Type', 'Spearman');


% ------------------------------------------------------------------------
% FA
% ------------------------------------------------------------------------

[rhoH_sx, pcorrH_sx] = corr( ...
    PD.HoehnYahr, ...
    PD.H_sx, ...
    'Type', 'Spearman');

[rhoH_dx, pcorrH_dx] = corr( ...
    PD.HoehnYahr, ...
    PD.H_dx, ...
    'Type', 'Spearman');


% ------------------------------------------------------------------------
% Varianza
% ------------------------------------------------------------------------

[rhovar_sx, pcorrvar_sx] = corr( ...
    PD.HoehnYahr, ...
    PD.Variance_sx, ...
    'Type', 'Spearman');

[rhovar_dx, pcorrvar_dx] = corr( ...
    PD.HoehnYahr, ...
    PD.Variance_dx, ...
    'Type', 'Spearman');
```

```
%% ------------------------------------------------------------------------
% STAMPA DEI RISULTATI DELLA CORRELAZIONE DI SPEARMAN
% ------------------------------------------------------------------------

fprintf('\n');
fprintf('=========================================\n');
fprintf('CORRELAZIONE DI SPEARMAN\n');
fprintf('=========================================\n');

fprintf( ...
    'Alpha SX:    rho = %.3f   p = %.6f\n', ...
    rhoalpha_sx, ...
    pcorralpha_sx);

fprintf( ...
    'Alpha DX:    rho = %.3f   p = %.6f\n', ...
    rhoalpha_dx, ...
    pcorralpha_dx);

fprintf( ...
    'Beta SX:     rho = %.3f   p = %.6f\n', ...
    rhobeta_sx, ...
    pcorrbeta_sx);

fprintf( ...
    'Beta DX:     rho = %.3f   p = %.6f\n', ...
    rhobeta_dx, ...
    pcorrbeta_dx);

fprintf( ...
    'H SX:        rho = %.3f   p = %.6f\n', ...
    rhoH_sx, ...
    pcorrH_sx);

fprintf( ...
    'H DX:        rho = %.3f   p = %.6f\n', ...
    rhoH_dx, ...
    pcorrH_dx);

fprintf( ...
    'Varianza SX: rho = %.3f   p = %.6f\n', ...
    rhovar_sx, ...
    pcorrvar_sx);

fprintf( ...
    'Varianza DX: rho = %.3f   p = %.6f\n', ...
    rhovar_dx, ...
    pcorrvar_dx);


%% ------------------------------------------------------------------------
% ESEMPIO DI FIGURA RIASSUNTIVA
% ------------------------------------------------------------------------

% La figura seguente rappresenta, a titolo di esempio, i valori
% dell'esponente alpha nelle diverse classi Hoehn-Yahr e la relativa
% associazione valutata mediante correlazione di Spearman.
% Figure analoghe sono state realizzate per beta, H e varianza.

figure;

subplot(2,2,1)

boxplot( ...
    PD.Alpha_sx, ...
    PD.HoehnYahr);

xlabel('Hoehn-Yahr');
```

```
ylabel('\alpha');
title('Alpha SX');
grid on;


subplot(2,2,2)

boxplot( ...
    PD.Alpha_dx, ...
    PD.HoehnYahr);

xlabel('Hoehn-Yahr');
ylabel('\alpha');
title('Alpha DX');
grid on;


subplot(2,2,3)

scatter( ...
    PD.HoehnYahr, ...
    PD.Alpha_sx, ...
    'filled');

xlabel('Hoehn-Yahr');
ylabel('\alpha');

title( ...
    sprintf( ...
    'SX - Spearman \\rho = %.3f', ...
    rhoalpha_sx));

grid on;


subplot(2,2,4)

scatter( ...
    PD.HoehnYahr, ...
    PD.Alpha_dx, ...
    'filled');

xlabel('Hoehn-Yahr');
ylabel('\alpha');

title( ...
    sprintf( ...
    'DX - Spearman \\rho = %.3f', ...
    rhoalpha_dx));

grid on;


%% ========================================================================
% FUNZIONE PER L'ESTRAZIONE DELLA STANCE PHASE
% ========================================================================

function [stance, IC_times, EC_times] = ...
    local_compute(signal, t, threshold)

    % Individuazione dei campioni per i quali la forza è superiore
    % alla soglia definita
    above_th = signal > threshold;

    % Calcolo delle transizioni del segnale binario
    edges = diff(double(above_th));

    % Transizione 0 -> 1: Initial Contact
    IC_idx = find(edges == 1);
```

```
    % Transizione 1 -> 0: End Contact
    EC_idx = find(edges == -1);

    % Gestione del caso in cui la registrazione inizi
    % durante una fase di appoggio
    if above_th(1) == 1
        IC_idx = [1; IC_idx];
    end

    % Gestione del caso in cui la registrazione termini
    % durante una fase di appoggio
    if above_th(end) == 1
        EC_idx = [EC_idx; length(signal)];
    end

    % Uniformazione del numero di eventi IC ed EC
    min_len = min(length(IC_idx), length(EC_idx));

    IC_idx = IC_idx(1:min_len);
    EC_idx = EC_idx(1:min_len);

    % Conversione degli indici in istanti temporali
    IC_times = t(IC_idx);
    EC_times = t(EC_idx);

    % Conservazione delle sole coppie di eventi temporalmente valide
    valid = EC_times > IC_times;

    IC_times = IC_times(valid);
    EC_times = EC_times(valid);

    % Calcolo della durata delle singole stance phase
    stance = EC_times - IC_times;

end


%% ========================================================================
% FUNZIONE PER DFA ED EXTENDED DFA
% ========================================================================

function [alpha, beta, F, dF, s] = ...
    edfa(signal, minScale, maxScale, numScales)

    % Rimozione del valore medio della serie
    signal = signal - mean(signal);

    % Costruzione del profilo cumulativo
    Y = cumsum(signal);

    % Generazione di scale distribuite logaritmicamente
    % tra il valore minimo e quello massimo
    s = round(logspace( ...
        log10(minScale), ...
        log10(maxScale), ...
        numScales));

    % Eliminazione delle eventuali scale duplicate introdotte
    % dall'arrotondamento
    s = unique(s);

    % Preallocazione dei vettori delle fluttuazioni
    F  = zeros(size(s));
    dF = zeros(size(s));


    % Analisi delle diverse scale
    for i = 1:length(s)
```

```
    boxSize = s(i);

    % Numero di segmenti completi disponibili
    % alla scala considerata
    N = floor(length(Y) / boxSize);

    % Fluttuazioni locali
    RMS = zeros(1, N);


    % Analisi dei singoli segmenti
    for j = 1:N

        idx = ...
            (j-1) * boxSize + 1 : ...
            j * boxSize;

        segment = Y(idx);

        x = (1:boxSize)';

        % Stima del trend locale mediante un polinomio
        % di secondo ordine
        p = polyfit( ...
            x, ...
            segment(:), ...
            2);

        trend = polyval(p, x);

        % Calcolo della fluttuazione RMS del segmento
        RMS(j) = sqrt(mean( ...
            (segment(:) - trend).^2));

    end


    % Fluttuazione complessiva alla scala corrente
    F(i) = sqrt(mean(RMS.^2));

    % Dispersione delle fluttuazioni locali impiegata
    % nell'Extended Detrended Fluctuation Analysis
    dF(i) = max(RMS) - min(RMS);

end


%% ---------------------------------------------------------------------
% STIMA DELL'ESPONENTE DFA ALPHA
% ---------------------------------------------------------------------

log_s = log(s);
log_F = log(F);

p_alpha = polyfit( ...
    log_s, ...
    log_F, ...
    1);

alpha = p_alpha(1);


%% ---------------------------------------------------------------------
% STIMA DELL'ESPONENTE EDFA BETA
% ---------------------------------------------------------------------

log_dF = log(dF);
```

```
    p_beta = polyfit( ...
        log_s, ...
        log_dF, ...
        1);

    beta = p_beta(1);

end


%% ========================================================================
% FUNZIONE PER LA FLUCTUATION ANALYSIS
% ========================================================================

function [H, F, s] = ...
    fa(signal, minScale, maxScale, numScales)

    % Rimozione del valore medio della serie
    signal = signal - mean(signal);

    % Costruzione del profilo cumulativo
    Y = cumsum(signal);

    % Generazione di scale distribuite logaritmicamente
    s = round(logspace( ...
        log10(minScale), ...
        log10(maxScale), ...
        numScales));

    % Eliminazione delle eventuali scale duplicate
    s = unique(s);

    % Preallocazione del vettore delle fluttuazioni
    F = zeros(size(s));


    % Analisi delle diverse scale
    for k = 1:length(s)

        scale = s(k);

        % Calcolo degli incrementi del profilo cumulativo
        % separati dalla scala considerata
        increments = ...
            Y(1+scale:end) - ...
            Y(1:end-scale);

        % Calcolo della fluttuazione RMS
        F(k) = sqrt(mean(increments.^2));

    end


    %% ------------------------------------------------------------------
    % STIMA DELL'ESPONENTE DI SCALA H
    % ------------------------------------------------------------------

    p = polyfit( ...
        log(s), ...
        log(F), ...
        1);

    H = p(1);

end
```