



SAPIENZA
UNIVERSITÀ DI ROMA

Università Sapienza di Roma - Dipartimento di Matematica

Master di II livello in "Calcolo Scientifico"

Tesina di tirocinio presso Thales Alenia Space s.p.A.

Metodi di interpolazione di serie temporali irregolari

Tutor interno:

Prof. Camillo Cammarota

Tutor esterno:

Ing. Carlo Ciancarelli

Tirocinante:

Dr. Daniele Manzo

2017/2018

Indice

Introduzione	1
1 Analisi di Dati Telemetrici da satellite	3
1.1 Il sistema satellitare	3
1.2 Analisi delle Telemetrie	7
1.3 Dataset	11
1.4 Contesto dell'analisi dei dati telemetrici	15
2 Metodi Numerici applicati	19
2.1 Interpolazione polinomiale	20
2.2 Modelli state-space: interpolazione con modello Kalman	28
2.3 Wavelet	41
2.3.1 Definizione e proprietà	41
2.3.2 Trasformate multiscale	52
2.3.3 Wavelet di Haar discrete	58
2.3.4 Wavelet Shrinkage	60
2.3.5 Dati con campionamento irregolare	64
3 Descrizione delle analisi	69
3.1 Analisi eseguite	69

4 Risultati delle analisi	75
Conclusioni	97
Bibliografia	99

Introduzione

Il presente lavoro ha avuto come obiettivo quello di analizzare serie temporali costituite da dati telemetrici satellitari. La caratteristica principale dei dati esaminati é costituita dal fatto che le serie temporali sono irregolari, a causa delle particolari procedure di acquisizione dei vari segnali a bordo del satellite; ciò determina inevitabilmente un campionamento non omogeneo delle misure delle grandezze fisiche osservate. Risulta pertanto necessario definire delle opportune tecniche di analisi per rendere questi dati omogenei, cioè per regolarizzare le serie temporali del dataset esaminato. Il lavoro é stato rivolto ad un numero finito di variabili osservabili, come temperature di apparati (elettronica), tensioni e correnti elettriche, parametri di avionica (posizioni, velocità angolari,...) che sono ritenuti essere significativamente rilevanti per il monitoraggio dello "stato di salute" e dell'efficienza operativa del satellite. L'intero processo di analisi é articolato in piú step successivi, che prevedono una fase preliminare in cui i campioni di dati ottenuti, che presentano un'irregolarità intrinseca dovuta al rate di campionamento del segnale acquisito, vengono regolarizzati, resi quindi omogenei dal punto di vista temporale, tramite opportune tecniche di interpolazione e analisi statistica. In tal modo, si ottengono set di dati regolari in cui i dati mancanti del set originario sono sostituiti con dati ricostruiti a partire da funzioni di interpolazione o tramite metodi statistici. Il lavoro svolto nel periodo di stage presso la Thales Alenia Space s.p.A., ha riguardato principalmente questo tipo di indagine, che risulta inserita all'interno di un ben piú ampio progetto che prevede nei suoi sviluppi futuri, una volta ottenuti dei dataset regolari, l'analisi dei trend e l'individuazione

di rumore ed eventuali stagionalità (periodicità) all'interno dei dati, dove queste ultime saranno sottratte dal dato originale così da evidenziare la presenza di possibili discontinuità o anomalie nel segnale rispetto al trend atteso, e la conseguente individuazione delle relative cause. In tal senso, è di particolare interesse determinare se esistono eventuali fenomeni precursori precedenti alla comparsa di eventi anomali all'interno della serie temporale acquisita. Una volta completate queste prime fasi di elaborazione si potrà passare ad effettuare un'analisi di correlazione tra le variabili per la ricerca di una possibile relazione di causa-effetto (correlazione statistica) tra queste, processo questo che potrà essere svolto, ad esempio, tramite metodi basati sulle Reti Bayesiane.

Capitolo 1

Analisi di Dati Telemetrici da satellite

1.1 Il sistema satellitare

Per introdurre il contesto nel quale abbiamo svolto l'analisi delle serie di dati, introduciamo alcuni concetti di base di un sistema satellitare. Un sistema satellitare é un insieme complesso di infrastrutture in parte in orbita e in parte a Terra, che operano in modo coordinato per lo svolgimento delle funzionalità e dei servizi richiesti. Comprende principalmente il segmento spaziale, composto da uno o piú satelliti eventualmente organizzati in costellazione, e il segmento terrestre, che comprende le infrastrutture a terra necessarie alla gestione e controllo dell'intero sistema, nonché a garantire le funzionalità e i servizi che devono essere erogati all'utente ([1],[2]). I sistemi satellitari si distinguono a seconda del tipo di missione richiesta, i principali sono i seguenti:

- sistema satellitare per le Telecomunicazioni, per i servizi di comunicazione di dati, fonía, immagini, ecc., tra utenti dislocati su ampie aree geografiche;
- sistema satellitare per l'Osservazione della Terra (o Telerilevamento), per i servizi di monitoraggio del territorio, delle condizioni atmosferiche, delle foreste, dei mari e degli oceani;
- sistema satellitare per la Meteorologia, per i servizi di previsione meteorologica;

- sistema satellitare di Navigazione, per i servizi di supporto alla navigazione aerea, di veicoli a terra (automobili, mezzi di trasporto, ecc.);
- sistema satellitare per Missioni Scientifiche, ad esempio per lo studio dei pianeti del sistema solare, delle comete e del sole.

La Figura 1.1 mostra schematicamente un sistema satellitare con due satelliti in orbita.

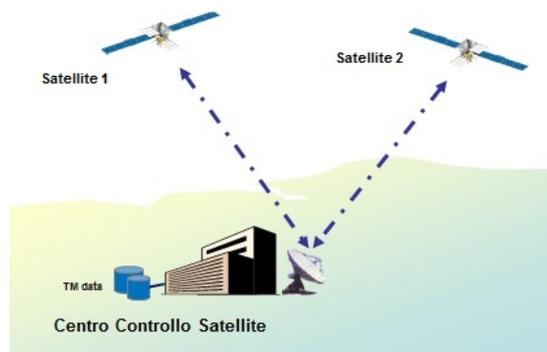


Figura 1.1: Schema di principio di un sistema satellitare.

In funzione della missione, la posizione orbitale può essere di tipo geostazionario (o GEO, dall'inglese Geostationary Earth Orbit), cioè il satellite si trova su un'orbita circolare ed equatoriale ad una altezza di circa 36.000 Km tale che il periodo di rivoluzione coincide con il periodo di rotazione della Terra; oppure può essere un'orbita terrestre bassa (o LEO, dall'inglese Low Earth Orbit), che è un'orbita attorno alla Terra di altitudine compresa tra 400 e 2000 km. Ulteriore classificazione orbitale riguarda i satelliti in orbita terrestre media (o MEO, dall'inglese Medium Earth Orbit) di altitudine oltre 2.000 km. A titolo di esempio, un satellite per l'osservazione della Terra è posizionato in un'orbita LEO ed ha un periodo di rivoluzione di circa 90 minuti, quindi viaggia a circa 27.400 km/h. Una delle funzioni fondamentali svolte dal segmento terrestre di un sistema satellitare riguarda il controllo e il monitoraggio del satellite in orbita, per garantirne il corretto funzionamento nel corso di tutta la sua vita operativa, tipicamente di durata

compresa tra 5 e 15 anni, in funzione della missione ([1],[2]). Infatti, occorre verificare costantemente che il satellite in orbita sia nella posizione orbitale richiesta (GEO, LEO, MEO) e che sia in condizioni operative nominali per lo svolgimento della missione per la quale é stato progettato e realizzato, ad esempio per l'acquisizione di immagini con sensori ottici o radar per l'osservazione di determinate aree di interesse, oppure per la ricezione e trasmissione dei segnali a banda larga in un sistema di telecomunicazione. L'infrastruttura del segmento terrestre che svolge il compito di controllo e monitoraggio del satellite é il Centro Controllo Satellite (o SCC, dall'inglese Satellite Control Center). Questa azione é duplice in quanto riguarda sia gli equipaggiamenti della piattaforma del satellite (avionica, dispositivi di alimentazione a pannelli solari, batterie, ecc.) e sia gli apparati del cosiddetto carico utile, cioè i dispositivi specifici per il tipo di missione richiesta. Ad esempio, per un satellite per l'osservazione della Terra tramite radar tali dispositivi comprendono principalmente l'antenna radar (detta anche SAR, dall'inglese Synthetic Aperture Radar, che identifica la tecnologia per l'acquisizione e la formazione di immagini tramite sensore radar), il processore SAR e il trasmettitore delle immagini verso la Terra ([3],[4]). La Figura 1.2 mostra un esempio di centro controllo satellite con le console degli operatori per la visualizzazione delle informazioni riguardanti la missione satellitare.



Figura 1.2: Esempio di Centro Controllo Satellite.

Il monitoraggio dello stato di funzionamento degli apparati a bordo del satellite é basa-

to sulla disponibilità di informazioni riguardanti i parametri caratteristici delle unità che si intendono monitorare. Le grandezze fisiche di interesse sono, ad esempio, la temperatura degli apparati, le tensioni e le correnti del sistema di alimentazione elettrica, le ruote di reazione per il controllo di assetto, i parametri della dinamica di volo del satellite, cioè la posizione, la velocità e l'accelerazione rispetto ad un sistema di riferimento cartesiano ortogonale tridimensionale solidale con il satellite in volo, normalmente indicato con la terna (roll, pitch, yaw). A titolo di esempio, la Figura 1.3 mostra uno schema di alto livello di alcune apparecchiature principali presenti in un satellite ([1],[2]).

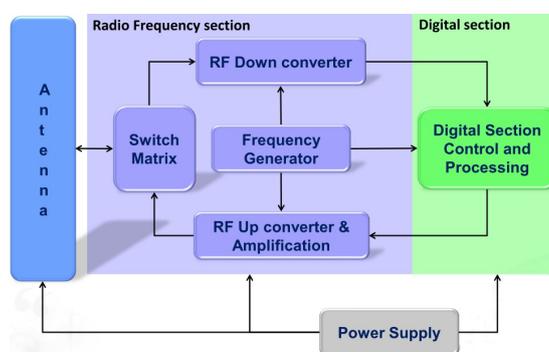


Figura 1.3: Esempio di schema di alto livello delle apparecchiature a bordo satellite.

Il numero complessivo di parametri normalmente misurati tramite sensori e dispositivi di misura posti a bordo del satellite è molto elevato e può superare le diecimila variabili. Tali parametri sono misurati sistematicamente tramite procedure automatizzate gestite dal calcolatore di bordo, i valori ottenuti sono codificati in forma digitale e memorizzati nella memoria di bordo. Nel momento in cui il satellite si trova nella posizione orbitale tale da essere in visibilità della stazione di terra, le misure raccolte e presenti nella memoria di bordo sono trasmesse al Centro di Controllo Satellite tramite il collegamento in radiofrequenza ([3]). Quindi l'insieme delle misure, indicate anche come dati telemetrici, costituisce un flusso di dati sistematicamente generato dagli apparati a bordo del satellite, trasferiti a Terra e memorizzati nel centro controllo satellite per le successive operazioni

di analisi e monitoraggio da parte del personale addetto alla supervisione e controllo del satellite stesso ([4]).

1.2 Analisi delle Telemetrie

I dati telemetrici generati dagli equipaggiamenti presenti a bordo di un satellite possono essere divisi in due classi principali:

- telemetrie cosiddette di housekeeping (Housekeeping Telemetry);
- telemetrie dei dati scientifici (Science Telemetry).

Le prime riguardano lo stato di funzionamento degli apparati tramite dispositivi di monitoraggio dedicati, mentre le seconde riguardano i dati effettivamente osservati/acquisiti in base alla specifica missione satellitare. L'analisi dei dati di housekeeping ha lo scopo di verificare il corretto funzionamento degli strumenti a bordo, in modo da poter modificare (anche in tempo reale) la loro configurazione per evitare errori o prevenire possibili malfunzionamenti; questo tipo di verifica richiede evidentemente una analisi in tempo reale delle telemetrie di housekeeping. L'analisi delle telemetrie dei dati scientifici ha invece lo scopo di verificare la qualità dei dati generati, ad esempio la qualità delle immagini acquisite da un sensore radar o ottico di un satellite di osservazione della Terra; il risultato di questa analisi è costituito normalmente da un insieme di parametri necessari per configurare la strumentazione a bordo affinché siano garantite le prestazioni attese. Quindi, considerato che queste informazioni non riguardano lo stato di funzionamento degli apparati, l'analisi delle telemetrie dei dati scientifici non richiede necessariamente di essere eseguita in tempo reale ([3]). Per quanto detto, l'analisi dei dati telemetrici riveste quindi una importanza vitale per la gestione operativa di ogni sistema satellitare al fine di garantire la continuità dei servizi relativi alla missione da realizzare (Telerilevamento,

Navigazione o Telecomunicazione). Questa disciplina tecnica si è sviluppata contestualmente alle fasi di progetto delle missioni spaziali ed è parte integrante delle attività di ingegneria per la progettazione, lo sviluppo e l'esercizio operativo dei sistemi satellitari ([4]). L'approccio classico nell'analisi delle telemetrie generate da satellite è quello di verificare che i valori delle variabili osservate siano entro soglie di guardia prestabilite. Il superamento di una soglia oltre un determinato intervallo di tempo determina l'attivazione una segnalazione di allarme. Questa metodologia quindi si basa sulla impostazione di allarmi a soglia fissa (*fixed thresholds*), come indicato in Figura 1.4, e sul controllo dei valori delle telemetrie rispetto a tali soglie. Il comportamento nel corso del tempo, la presenza di eventuali trend e la variabilità del dato nell'intervallo di valori consentiti di fatto non è esaminato e, conseguentemente, questo comporta un ritardo nell'identificazione di eventuali anomalie (*late detection*).

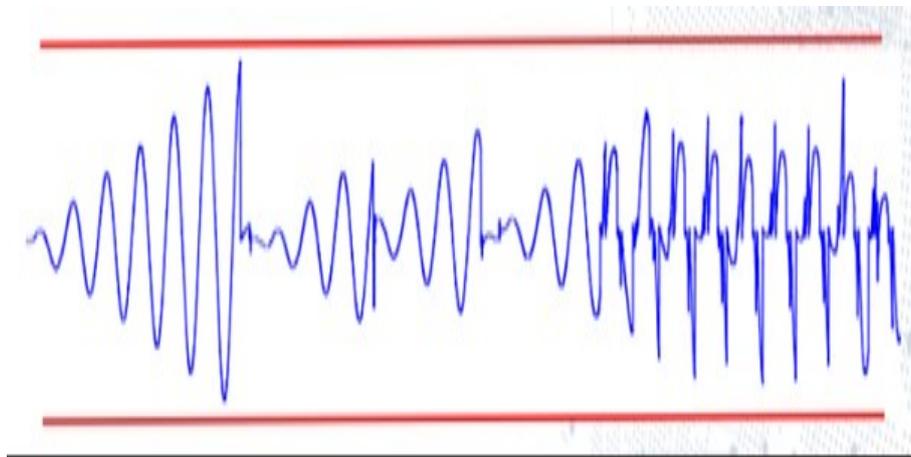


Figura 1.4: Esempio monitoraggio telemetrie a soglia fissa.

La Figura 1.5 mostra un semplice esempio di un segnale di telemetria che presenta un trend di tipo crescente fino a superare la soglia massima consentita.



Figura 1.5: Esempio telemetria che eccede la soglia fissa.

Nel corso delle attività operative, l'attivazione di un allarme causato dal superamento di una soglia comporta l'avvio di procedure di verifica da parte del personale del centro di controllo satellite per l'analisi e la risoluzione dell'anomalia dell'unità di bordo interessata, facendo anche ricorso alle funzioni di ridondanza presenti a bordo per le unità ritenute più critiche per la missione. Ovviamente, nelle situazioni più critiche, l'insorgenza di una anomalia a bordo del satellite può determinare una condizione di fuori servizio, con conseguenze che dipendono dal tipo di problema, fino all'interruzione parziale o totale dei servizi della missione ([2],[3],[4]). Per questa serie di motivazioni di natura tecnico-operativa con possibili risvolti anche dal punto di vista economico, la **tendenza delle attività di ricerca applicata in questo settore industriale è di affrontare il tema del monitoraggio dei dati telemetrici da satellite anche con strumenti analitici che consentano di incrementare la capacità di prevenire eventuali anomalie, piuttosto che attendere "passivamente" la loro insorgenza.** La complessità del problema richiede un lavoro congiunto tra più discipline tecniche, tra cui l'ingegneria, per quanto riguarda le conoscenze specifiche del progetto e delle funzioni del sistema satellitare, e la matematica applicata, per quanto riguarda l'individuazione dei metodi di analisi, la definizione di modelli e di algoritmi in grado di estrarre le informazioni desi-

derate. Rispetto all'approccio classico, si tratta quindi di esaminare sistematicamente le serie storiche nel tempo con l'uso di strumenti di analisi e algoritmi studiati per la tipologia di dati di interesse (analisi statistica, analisi dei trend, filtraggio, analisi predittiva, ecc.), con l'obiettivo di individuare eventuali "segni premonitori" o specifici indicatori in grado di segnalare un potenziale degrado delle prestazioni di una unità a bordo e quindi di permettere un intervento tempestivo da parte del centro di controllo satellite (*predictive maintenance*). Questo approccio relativamente recente nel panorama dell'analisi dei dati quindi si basa sulla capacità di opportuni algoritmi e tecniche numeriche di estrarre informazione (*knowledge retrieval*) dalle serie storiche delle variabili osservate. Un ulteriore passo nella direzione di estrazione di informazione dai dati è di utilizzare tecniche avanzate di data-analytics per la ricerca di correlazioni tra le variabili di telemetria osservate, cioè di individuare possibili relazioni di causa-effetto (correlazione statistica). In questo caso, l'obiettivo è di individuare le possibili dipendenze tra le variabili che descrivono il funzionamento del sistema satellite, in modo da risalire alle possibili cause che hanno determinato una anomalia. Ad esempio, una variazione di una tensione di alimentazione in ingresso ad un apparato, seppur graduale nel tempo e senza superamento delle soglie di guardia, può causare il degrado delle prestazioni dell'apparato stesso e di altri ad esso correlati dal punto di vista funzionale ([1],[2],[3]). La ricerca e sperimentazione di tecniche analitiche in grado di estrarre informazione dai dati richiede in primo luogo la conoscenza approfondita delle caratteristiche e della struttura dei dati a disposizione, come ad esempio la presenza di rumore o di irregolarità che dipendono dal sistema che li ha generati. Per questo motivo, nel paragrafo 1.3 descriveremo la composizione e le caratteristiche del dataset delle telemetrie satellitari che abbiamo esaminato in questo lavoro ([4]). Nel paragrafo 1.4 descriveremo il contesto del processo complessivo di elaborazione dei dati telemetrici satellitari necessario per raggiungere gli obiettivi suddetti, all'interno del quale si colloca la parte di elaborazione affrontata in questo lavoro.

1.3 Dataset

Il dataset analizzato nel presente lavoro comprende un totale di 26 variabili, come riportato nell'elenco in Tabella 1.8.

- Temperatura di apparati;
- Tensione di alimentazione e corrente elettrica;
- Errore di velocità sui tre assi (X,Y,Z);
- Errore angolare sui tre assi (X,Y,Z);
- Velocità ruote di reazione;
- Velocità angolari sui tre assi (X,Y,Z).

La costruzione del dataset è stata guidata dalla necessità di avere un numero di parametri limitato, ma al tempo stesso significativo per quanto riguarda le funzioni e variabili coinvolte nel contesto del controllo di assetto del satellite. La selezione di questo insieme di variabili di interesse (tra le migliaia di dati generati dal satellite) ha richiesto ovviamente il coinvolgimento di esperti di ingegneria di satellite. Esaminando nel dettaglio le serie storiche rese disponibili si osservano le seguenti principali caratteristiche:

1. il passo di campionamento non è uniforme: l'intervallo temporale tra due campioni successivi della serie non è costante, quindi siamo in presenza di serie irregolari; rispetto alla scala temporale di 1 s, si nota che, mediamente, è presente circa un campione ogni otto;
2. presenza di componenti stagionali, in buona parte riconducibili alle periodicità orbitali del sistema satellitare;

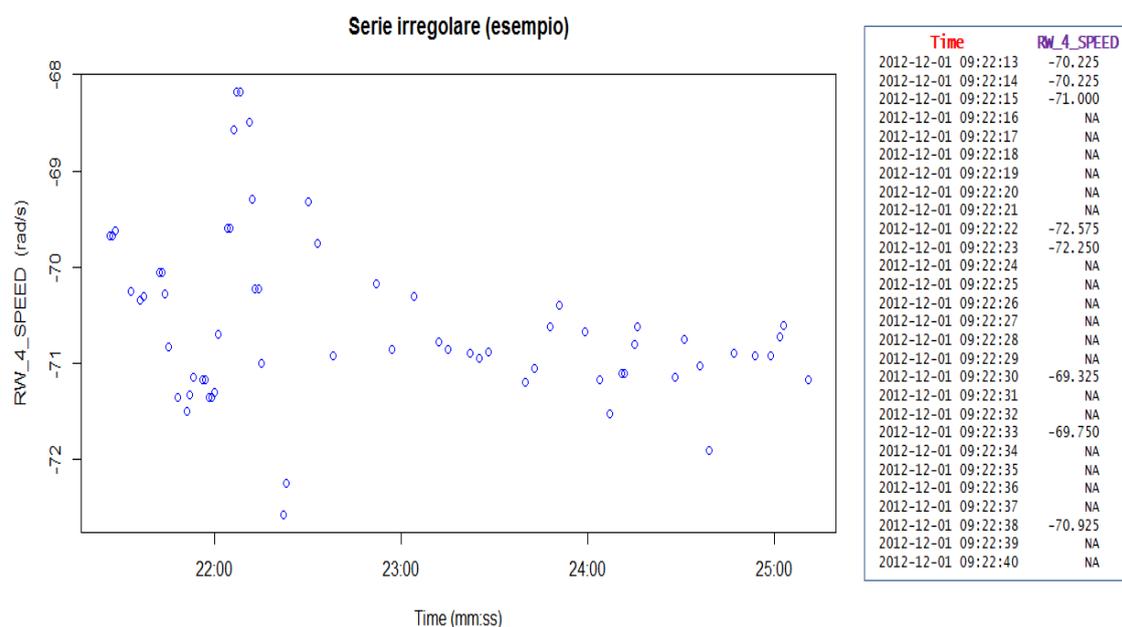


Figura 1.6: Esempio di serie temporale irregolare campionata al secondo. I dati mancanti (*missing-data*) sono indicati con NA.

3. presenza di discontinuit  evidenti rispetto all'andamento nominale della serie, probabilmente riconducibili alle manovre programmate per il controllo di assetto del satellite;

4. assenza di campioni in alcuni tratti delle serie, in certi casi di durata dell'ordine di alcune decine di minuti, dovuta all'interruzione del canale di telemetria durante il trasferimento dei dati a Terra.

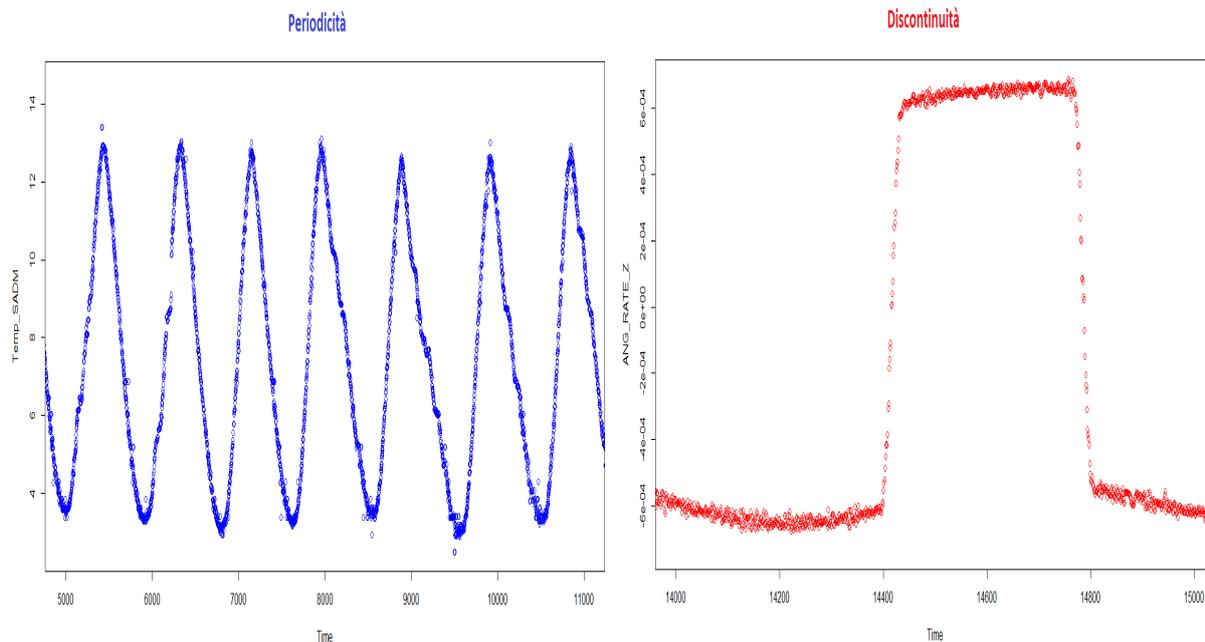


Figura 1.7: Esempi di serie temporali periodiche e discontinue.

Nelle Figure 1.6 e 1.7 sono riportati a titolo di esempio alcuni grafici delle serie storiche che mostrano alcuni tratti delle caratteristiche suddette. Tenendo conto delle caratteristiche delle serie di dati, in questo lavoro abbiamo concentrato l'attenzione principalmente su come rendere il passo di campionamento omogeneo tramite tecniche di interpolazione. Per quanto riguarda le variabili relative agli errori angolari, queste sono ottenute tramite un sistema di controllo in controreazione; l'uscita dal processo viene confrontata con il riferimento da un controllore, il risultato di questo confronto è l'errore del sistema di controllo. Questo parametro identifica quindi l'errore angolare misurato in RAD attorno all'asse Z ricavato dal controllore di assetto del sottosistema avionico. Identifica quindi quanto la posizione attuale si discosta da quella desiderata; ovviamente l'errore d'assetto deve rientrare nei range specificati dai requisiti relativi al puntamento del satellite.

No.	Acronimo	Nome del parametro
1	Temp_PCU	Temperatura Power Control Unit (Distribuzione di potenza)
2	Temp_BTA	Temperatura Batteria
3	Temp_SMU	Temperatura Spacecraft Management Unit (Computer di bordo)
4	Temp_Gyro	Temperatura Giroscopio
5	Temp_SADM	Temperatura Solar Array Drive Motor (Motore di rotazione dei Solar Array)
6	Temp_FSS	Temperatura Fine Sun Sensor (Sensore di sole)
7	Temp_MGM	Temperatura Magnetometer
8	BTA_Voltage	Tensione ai capi della Batteria
9	SAW_Current	Corrente fornita dai Solar Array
10	SADM_Voltage	Tensione di alimentazione Solar Array Drive Motor (Motore di rotazione dei Solar Array)
11	RATE_ERR_X	Errore di velocità asse X (su roll)
12	RATE_ERR_Y	Errore di velocità asse Y (su pitch)
13	RATE_ERR_Z	Errore di velocità asse Z (su yaw)
14	ANGLE_ERR_X	Errore angolare asse X (su roll)
15	ANGLE_ERR_Y	Errore angolare asse Y (su pitch)
16	ANGLE_ERR_Z	Errore angolare asse Z (su yaw)
17	RW_1_SPEED	Velocità ruota di reazione 1
18	RW_2_SPEED	Velocità ruota di reazione 2
19	RW_3_SPEED	Velocità ruota di reazione 3
20	RW_4_SPEED	Velocità ruota di reazione 4
21	GYRO_FILT_ANG_RATE_X	Velocità angolare filtrata asse X (su roll)
22	GYRO_FILT_ANG_RATE_Y	Velocità angolare filtrata asse Y (su pitch)
23	GYRO_FILT_ANG_RATE_Z	Velocità angolare filtrata asse Z (su yaw)
24	ANG_RATE_X	Velocità angolare asse X (su roll)
25	ANG_RATE_Y	Velocità angolare asse Y (su pitch)
26	ANG_RATE_Z	Velocità angolare asse Z (su yaw)

Figura 1.8: Tabella delle variabili del data-set di dati telemetrici.

1.4 Contesto dell'analisi dei dati telemetrici

L'analisi delle serie storiche del dataset condotta in questo lavoro si inquadra in un piú ampio progetto di ricerca e sviluppo per il trattamento delle telemetrie satellitari, che trae origine dalle considerazioni tecniche che abbiamo esposto nel paragrafo 2.2. Gli obiettivi generali di questa attività possono essere sintetizzati nei seguenti punti principali:

1. caratterizzare il comportamento delle variabili osservate;
2. monitorare ed analizzare i dati telemetrici con il supporto di strumenti analitici che consentano di estrarre informazione "nascosta" nei dati;
3. individuare, possibilmente in modo preventivo, caratteri irregolari nei dati che siano riconducibili a possibili anomalie degli apparati di bordo;
4. fornire supporto alle attività di ingegneria per le operazioni di satellite.

Come spesso avviene in molti procedimenti di elaborazione di dati, il processo di trattamento dei dati telemetrici é impostato logicamente come una sequenza di operazioni successive, progettate con specifici obiettivi, sulla base delle caratteristiche delle serie storiche in esame. Questo processo si compone di varie fasi successive di elaborazione, che é possibile schematizzare come segue:

1. **Regolarizzazione**

Essendo i dati telemetrici delle serie irregolari, come primo passo si rende necessario interpolare i dati disponibili per ottenere delle serie temporali con una griglia di campionamento a passo costante; infatti, molti metodi di analisi statistica si basano sull'ipotesi che i dati in input siano regolarmente spazati e, normalmente, non garantiscono i risultati in caso di serie irregolari. Il panorama delle tecniche disponibili in letteratura é molto ampio, e si tratta di identificare quali tra queste offrono le migliori prestazioni con il minor impatto nella ricostruzione dei dati mancanti.

2. Filtraggio e rimozione componenti stagionali

Questa fase di elaborazione ha l'obiettivo di ridurre il contenuto di rumore presente nelle serie temporali (adottando un opportuno modello di rumore, ad esempio di tipo additivo) e di sottrarre le componenti stagionali presenti nelle serie, dovute principalmente alla periodicità orbitale del sistema satellitare. Inoltre, ove possibile, si cerca di rimuovere alcuni specifici pattern presenti nelle serie dovuti a fenomeni che è possibile riconoscere come deterministici (quindi non casuali) inerenti l'operatività del sistema satellitare, come ad esempio le manovre di assetto imposte dal centro di controllo satellite. In questa fase il rischio che occorre valutare è duplice: da un lato, la possibile eliminazione di una parte dell'informazione utile presente nelle serie temporali e, dall'altro, l'introduzione di possibili distorsioni dovute alle specifiche tecniche di filtraggio.

3. Analisi predittiva e correlazione statistica

Questa fase di elaborazione è quella più complessa e si propone di analizzare le serie storiche (dopo essere state sottoposte ai trattamenti di regolarizzazione e filtraggio nelle fasi precedenti) per estrarre le informazioni "nascoste" riguardanti il loro andamento nel tempo o la mutua correlazione. Gli obiettivi sono di vario tipo, ad esempio individuare trend anomali, l'esistenza di correlazioni statistiche tra le serie temporali o identificare possibili relazioni di causa-effetto. Le tecniche note in letteratura sono di vario tipo e non sempre di immediata applicazione, in quanto richiedono una attenta verifica dell'applicabilità ai dati di interesse e della sensibilità dei risultati. A titolo di esempio, citiamo le reti Bayesiane, i metodi di Markov, le reti neurali convoluzionali (CNN).

Questa schematizzazione preliminare, sebbene non esaustiva e suscettibile di ulteriori fasi intermedie di elaborazione, descrive sinteticamente il contesto delle necessità di elabo-

razione per indirizzare gli obiettivi suddetti. é evidente che ciascuna fase richiede l'utilizzo di tecniche numeriche, algoritmi, modelli e procedimenti di analisi noti in letteratura, ma che devono essere attentamente selezionati ed adattati al problema di nostro interesse per poter fornire, auspicabilmente, i risultati attesi. La valutazione delle varie tecniche richiede anche un confronto tra i risultati ottenuti in termini di prestazione, distorsione, stima dell'errore, ecc., per la scelta della soluzione ottimale rispetto a criteri prefissati.

Il presente lavoro ha affrontato sostanzialmente il primo passo di questo processo, cioè l'applicazione e sperimentazione di alcuni metodi di interpolazione delle serie temporali irregolari, come descritto nei capitoli successivi.

Capitolo 2

Metodi Numerici applicati

I metodi numerici utilizzati sono alcuni ben noti metodi di interpolazione, distinti principalmente in metodi deterministici e metodi stocastici. La questione ha riguardato quindi la definizione di un modello plausibile per ricostruire quei data-points assenti nei segnali originari. I metodi deterministici (interpolazione matematica) sfruttano il legame esistente fra punti vicini, espresso da una legge esplicita i cui parametri hanno significato fisico, come per esempio l'interpolazione lineare, cubica e la serie di Fourier. I metodi stocastici (interpolazione statistica), invece, utilizzano il legame fra punti vicini espresso da una correlazione statistica (covarianza o funzione di correlazione) tra dati che può non avere significato fisico. Inoltre, l'interpolazione può avere un carattere globale (trend e serie di Fourier) o locale (spline), a seconda che siano utilizzati tutti i valori della serie oppure soltanto i valori dei punti vicini, per costruire il modello. I metodi di interpolazione applicati sono: *interpolazione polinomiale: lineare e spline* (cubica); *interpolazione Stineman* (variante dell'interpolazione lineare); *interpolazione tramite filtro Kalman*; *interpolazione tramite Wavelet*. Di seguito, forniremo una breve descrizione di ognuno di questi metodi.

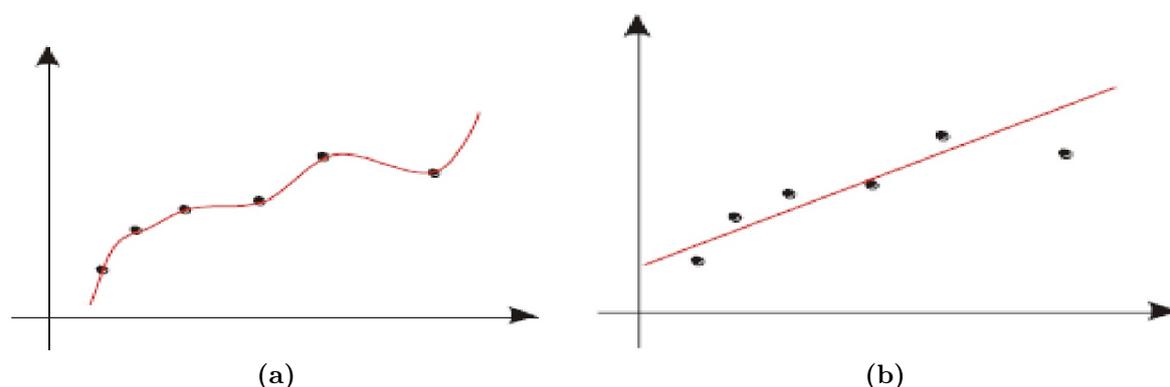


Figura 2.1: (a) Interpolazione matematica: calcola una funzione polinomiale con il vincolo che questa passi per tutti i punti noti del campione di dati; (b) Interpolazione statistica: calcola una funzione polinomiale che non necessariamente passa per i punti noti del campione di dati.

2.1 Interpolazione polinomiale

In molti problemi di natura matematica, nasce spesso l'esigenza di dover approssimare una funzione $f(x)$ nota analiticamente, oppure nota solo in un insieme discreto di punti, con una funzione $g(x)$ piú semplice, scelta in un insieme opportuno. La costruzione di tale funzione puó avvenire secondo due approcci: 1) dati affetti da errore: i dati vengono approssimati "nel loro insieme", processo denominato *migliore approssimazione*; 2) dati esatti: si cerca una funzione che passi per i valori assegnati, processo definito *interpolazione*. Le classi piú note di funzioni utilizzate per l'approssimazione e interpolazione sono: 1) funzioni polinomiali; 2) funzioni razionali; 3) funzioni trigonometriche (serie di Fourier); 4) funzioni polinomiali a tratti (*splines*). Tuttavia, la nostra attenzione si é concentrata solo su alcuni di questi, ritenuti essere piú adatti al tipo di analisi che si voleva eseguire. Il metodo di *interpolazione polinomiale* consiste nell'interpolare $n + 1$ punti noti di una serie, detti nodi (x_0, x_1, \dots, x_n) , con un polinomio di grado n , $P_n(x)$, calcolandone i coefficienti a partire dai valori assunti da dato polinomio nei punti noti, minimizzando un'opportuna funzione delle differenze (distanze) tra questi valori e i corrispondenti valori dei

punti noti; questo criterio di minimizzazione é solitamente quello dei minimi quadrati. In base al grado n del polinomio si possono quindi costruire le seguenti interpolazioni: 1) $n = 0 \rightarrow P_0(x) = a_0$ (funzione costante); 2) $n = 1 \rightarrow P_1(x) = a_0 + a_1x$ (retta); 3) $n = 2 \rightarrow P_2(x) = a_0 + a_1x + a_2x^2$ (parabola). In generale, per un grado n qualsiasi si ha il polinomio:

$$P_n(x) = \sum_{i=0}^n a_i x^i$$

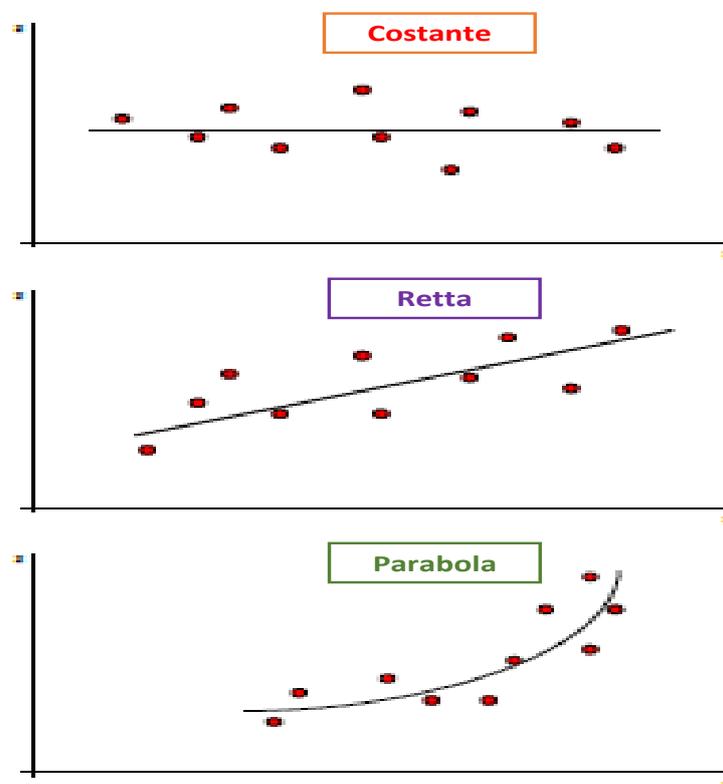


Figura 2.2: Rappresentazione schematica semplificata di un set di data-points interpolati (o fittati) con funzioni polinomiali: a) funzione costante (grado $n = 0$); b) funzione lineare o retta (grado $n = 1$); c) funzione quadratica o parabola (grado $n = 2$).

Applicando il criterio dei minimi quadrati per il calcolo dei coefficienti a_i , si minimizza

la funzione:

$$\sum_{i=1}^m (\hat{x}_i - x_i)^2$$

dove con \hat{x}_i si indicano i valori stimati dal polinomio, con x_i i valori noti e con m il numero di punti noti. Per le loro caratteristiche peculiari di solito questi metodi sono usati per rimuovere trend globali prima di applicare un interpolatore locale. In generale, la procedura di interpolazione polinomiale di un set discreto di dati, consiste nel determinare una funzione g , appartenente ad una classe prescelta di funzioni, che meglio approssima una funzione data f . Nel caso discreto, la funzione f é nota su un insieme di $n + 1$ punti (nodi) appartenenti ad un intervallo definito $[a, b]$. Questo é un caso caratteristico dell'analisi di dati sperimentali: talvolta il numero $n + 1$ di dati disponibili é piccolo, come nello studio di certi fenomeni fisici o biologici, talvolta é elevato, come nelle analisi statistiche e nello studio di serie temporali. Si puó anche supporre che i valori $f(x_i)$, in particolare quando sono rilevati in laboratorio, siano affetti da errori di misura. La funzione g che si costruisce serve per approssimare valori della f in punti diversi dai nodi o per dedurne proprietá di comportamento. Al caso discreto pu'ó essere ricondotto anche il caso continuo, in cui la funzione f é una funzione nota della matematica, in generale non razionale. Lo scopo 'e quello di produrre una funzione g piú semplice, cioé piú facilmente "trattabile" (ad esempio piú facilmente calcolabile, derivabile o integrabile) della f . In questo caso i nodi sono scelti in modo opportuno e i valori $f(x_i)$ sono calcolati con elevata accuratezza, sfruttando le proprietá note della funzione f . In generale, date le coppie di valori $(x_0, f_0), \dots, (x_n, f_n)$ con $x_i \in [a, b]$, e indicato con S lo spazio (o base ortonormale) delle funzioni definite su $[a, b]$; si cerca quindi una funzione $g \in S$ che soddisfi le condizioni di interpolazione come il passaggio per i nodi (valori noti) di $f(x)$, per cui:

$$g(x_i) = f_i, \quad i = 0, 1, \dots, n$$

; si vuole quindi valutare g nei punti $\bar{x} \neq x_i, i = 1, \dots, n$, con $\bar{x} \in [a, b]$. La scelta dello spa-

zio delle funzioni S ricade su spazi lineari a dimensione finita $n + 1$ (pari al numero di nodi presenti nell'intervallo considerato), per cui si può fissare una base $\phi_j(x)$, $j = 0, 1, \dots, n$ ed esprimere $g(x) \in S$ come:

$$g(x) = \sum_{j=0}^n a_j \phi_j(x)$$

dove trovare la funzione interpolante $g \in S$ equivale a trovare i coefficienti a_0, a_1, \dots, a_n tali che

$$g(x_i) = \sum_{j=0}^n a_j \phi_j(x_i) = f_i, \quad i = 0, 1, \dots, n$$

, che si riduce a risolvere un sistema di n equazioni lineari in n incognite, che sono i coefficienti a_i ($i = 0, 1, \dots, n$) della funzione interpolante g .

Nel caso dell'interpolazione polinomiale lo spazio $S = P_n$ delle funzioni di base $\phi(x)$ sono le funzioni polinomiali $1, x, x^2, x^3, \dots, x^n$, per cui la funzione interpolante g è data da:

$$g(x) = \sum_{j=0}^n a_j x^j$$

dove le funzioni di base sono

$$\phi_j(x) = x^j$$

, per cui la funzione interpolante è il polinomio

$$g(x) = p(x) = a_0 + a_1 x + a_2 x^2 + \dots + a_n x^n$$

. Trovare la funzione interpolante $g \in S$ equivale a trovare dei coefficienti a_0, a_1, \dots, a_n tali che

$$g(x_i) = \sum_{j=0}^n a_j \phi_j(x_i) = f_i \quad i = 0, 1, \dots, n$$

per cui essendo $\phi_j = x^j$ si ottiene:

$$g(x_i) = \sum_{j=0}^n a_j x_i^j = f_i \quad i = 0, 1, \dots, n$$

Il problema, quindi, consiste nel risolvere un sistema lineare nelle incognite a_0, a_1, \dots, a_n , dato da:

$$\begin{cases} a_0 + a_1x_0 + a_2x_0^2 + \dots + a_nx_0^n = f_0 \\ a_0 + a_1x_1 + a_2x_1^2 + \dots + a_nx_1^n = f_1 \\ \vdots \\ a_0 + a_1x_n + a_2x_n^2 + \dots + a_nx_n^n = f_n \end{cases}$$

Tuttavia, va sottolineato che il metodo di interpolazione polinomiale presenta delle notevoli limitazioni, poiché il comportamento del polinomio interpolante, in generale, peggiora al crescere del numero di nodi e quindi del grado del polinomio, poiché le derivate di f crescono rapidamente all'aumentare di n , fatto questo che determina un andamento oscillante del polinomio interpolatore che i dati o la funzione non presentano. Ciò implica che spesso è preferibile ricorrere a polinomi di grado basso anche quando sono noti molti dati. Se i nodi vengono scelti in modo opportuno il risultato dell'interpolazione polinomiale migliora nettamente. Nella pratica è abbastanza naturale considerare come nodi punti x_i di f equidistanti. Tuttavia, spesso si preferisce adottare una scelta più opportuna per la posizione dei nodi, tenendo conto della distribuzione dei dati noti o della funzione da interpolare, e quindi ridistribuendo i nodi in modo non omogeneo, come per esempio nel caso dei nodi di Chebyshev con cui è possibile addensare i nodi in alcune regioni dell'intervallo e renderli invece meno densi in altre. In generale, i metodi di interpolazione polinomiale sono in grado di modellizzare bene andamenti globali di una grande quantità di dati tramite polinomi di grado basso; inoltre, è possibile stimare con accuratezza il grado di approssimazione. Tuttavia, spesso risulta difficile dare un corretto significato fisico a polinomi di grado alto utilizzati per l'interpolazione; inoltre, l'utilizzo dei minimi quadrati come criterio di stima della migliore approssimazione dei dati, presuppone che i residui tra dati veri e dati costruiti dall'interpolazione abbiano una distribuzione nor-

male; di fatto, anche gli stessi minimi quadrati costituiscono un criterio di stima poco attendibile.

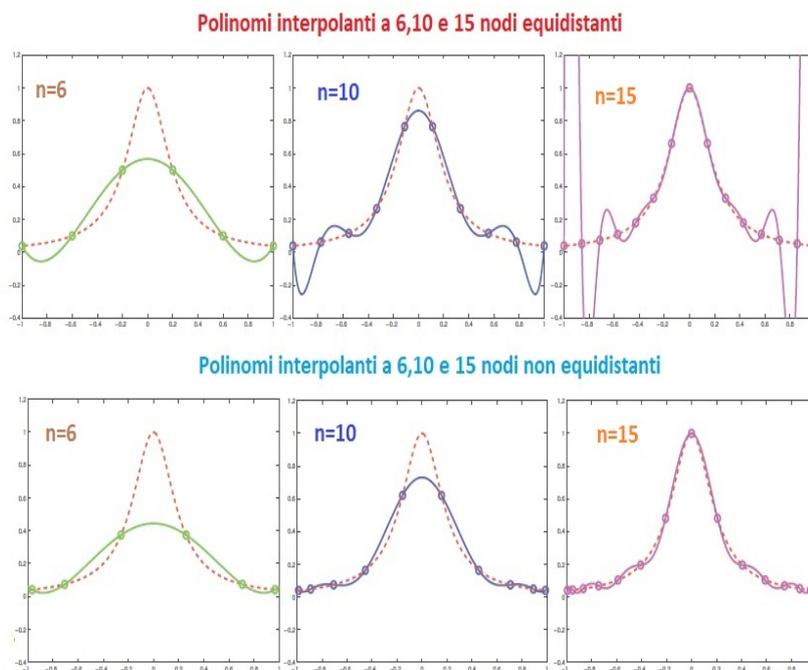


Figura 2.3: Esempio di interpolazione polinomiale di una funzione al variare del numero di nodi (6, 10, 15) nell'intervallo $[-1, 1]$, nel caso di nodi equidistanti (prima riga di grafici) e non equidistanti (seconda riga di grafici).

Come già accennato in precedenza, nell'interpolazione polinomiale spesso è poco conveniente aumentare il grado del polinomio interpolante, poiché così facendo si possono introdurre delle oscillazioni dei polinomi che poco si accostano ai dati reali. Una soluzione migliore è costituita dall'impiego di *funzioni polinomiali a tratti (splines)*. L'idea di base consiste nell'usare interpolanti polinomiali di grado basso, spezzando l'intervallo in considerazione in tanti sotto-intervalli ed interpolando in ciascun sotto-intervallo solo pochi punti. In linea di principio con l'utilizzo delle splines non è necessario che gli estremi dei sotto-intervalli coincidano con punti di interpolazione. In tal caso, si parla di

splines interpolanti nei nodi. Il caso piú semplice é costituito da funzioni lineari a tratti (polinomi di primo grado in ciascun sotto-intervallo). Assegnato, quindi, un intervallo $[a, b]$ ed gli $n + 1$ nodi $a = y_0, y_n = b$, una *funzione spline* di grado m e *nodi* y é una funzione polinomiale a tratti tale che:

$$\begin{cases} S_{m,y}(x) \in P_m, x \in [y_i, y_{i+1}] \\ S_{m,y}(x) \in C^{m-1}[a, b] \end{cases}$$

L'interpolazione o approssimazione é, quindi, data da un insieme di funzioni definite localmente e opportunamente raccordate. Su n intervalli ($n + 1$ nodi), si puó scrivere la funzione interpolante $g(x)$ come:

$$g(x) = \sum_{i=0}^{n-1} S_i(x)$$

dove le $S_i(x)$ sono le funzioni di base, diverse da zero solo nei rispettivi intervalli i -esimi $[x_i, x_{i+1}]$ e sono dei polinomi di grado basso. Il raccordo fra i tratti é realizzato imponendo vincoli sulle singole funzioni (polinomi). Si impone, infatti, di avere l'uguaglianza delle derivate fino all'ordine $r - 1$ sugli estremi di ciascun sotto-intervallo, cioé che le funzioni godano di una certa regolaritá:

$$\frac{\partial^j S_i(x_i)}{\partial x_i^j} = \frac{\partial^j S_{i+1}(x_i)}{\partial x_i^j}$$

dove $j = 0, \dots, r - 1$ e $i = 1, \dots, n - 1$, con $r = 0$ che indica nessun vincolo, $r = 1$ continuitá della funzione, $r = 1$ continuitá della funzione e della sua derivata. In totale, si ha quindi $(m + 1)n - m(n - 1) = n + m$ (dimensione dello spazio delle splines di grado m , con regolaritá $m - 1$ definite su $n + 1$ nodi).

Un ulteriore miglioramento é stato ottenuto con funzioni splines di base piú stabili, dette *B-splines*, le quali sono definite per ricorrenza a partire dalla funzione:

$$B_{i,0} = \begin{cases} 1 & \text{se } x \in [y_i, y_{i+1}) \\ 0 & \text{se } x < y_i, x \geq y_{i+1} \end{cases}$$

per cui si ottiene:

$$B_{i,m}(x) = \left(\frac{x - y_i}{y_{i+m} - y_i}\right)B_{i,m-1}(x) - \left(\frac{y_{i+m+1} - x}{y_{i+m+1} - y_{i+1}}\right)B_{i+1,m-1}(x)$$

Le B-splines hanno le proprietà di essere funzioni tutte positive ($B_{i,m}(x) > 0$, $x \in (y_i, y_{i+m+1})$), a supporto compatto ($\text{supp}B_{i,m}(x) = [y_i, y_{i+m}]$), sono opportunamente normalizzate ($\sum_{i=-\infty}^{+\infty} B_{i,m}(x) = 1$). In generale i nodi delle spline possono non coincidere con i punti di interpolazione; se invece coincidono si ottiene il sottoinsieme delle spline interpolanti nei nodi. In questo caso le $n + 1$ condizioni di interpolazione riducono il numero dei parametri liberi a $n + m - (n + 1) = m - 1$. Tra le spline interpolanti nei nodi le più usate sono quelle cubiche per le quali restano da determinare due condizioni ulteriori. Le scelte più comuni sono $S''(a) = S''(b) = 0$ e $S'(a) = S'(b)$, $S(a) = S(b)$. Le spline danno un grande vantaggio dal punto di vista computazionale, ma se usate per rendere più "smooth" i bordi di una data regione portano ad un calcolo errato dell'area e, inoltre, spline di ordine alto non possono modellare discontinuità, salti o punti angolosi. L'interpolazione di una funzione o di un profilo di data-points è data dalla funzione costituita dalla somma di tutte le spline:

$$g(x) = \sum_{i=1}^n f_i B(\text{frac}x - x_i \Delta x)$$

dove per ogni sotto-intervallo soltanto due spline sono non nulle, con f_i valore della funzione data nel nodo i -esimo e Δx il passo di ogni sotto-intervallo.

I principali vantaggi dell'utilizzo delle spline per l'interpolazione di dati è il ridotto tempo computazionale, la modellizzazione accurata di trend locali, e possono essere sfruttate efficacemente per la visualizzazione di dati. Tuttavia, gli svantaggi sono costituiti principalmente dal fatto che spline di ordine alto non sono in grado di modellare accuratamente discontinuità, salti o punti angolosi, introducono caratteristiche non presenti nel profilo dei dati, e infine non forniscono una stima degli errori di approssimazione.

Tra tutti i metodi di interpolazione polinomiale esistenti, abbiamo utilizzato soltanto i seguenti:

- interpolazione lineare;
- interpolazione spline (cubica);
- interpolazione Stineman.

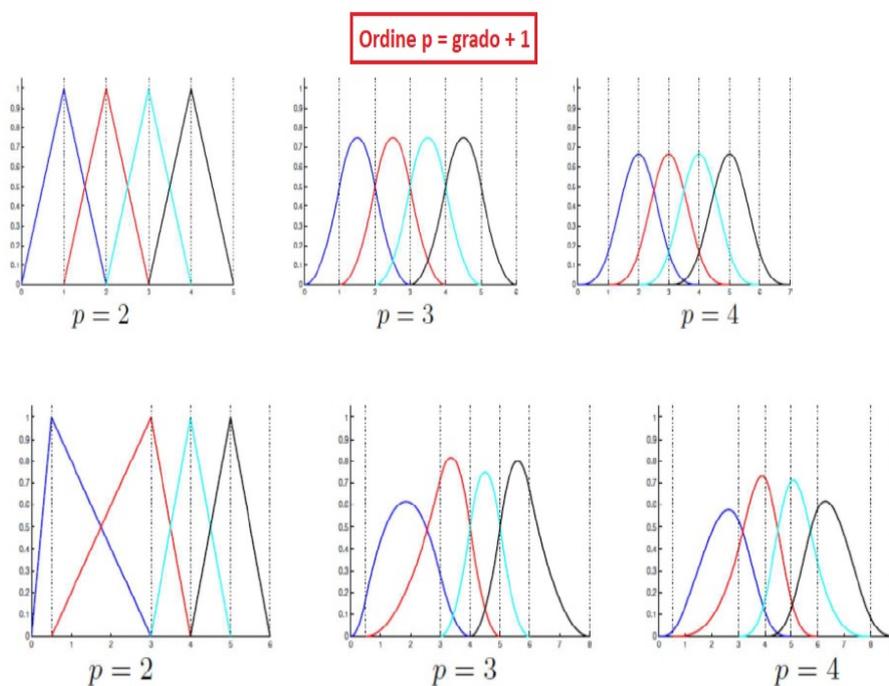


Figura 2.4: Funzioni *splines* di vari ordini p ($p = n + 1$, dove n è il grado dei polinomi utilizzati.)

2.2 Modelli state-space: interpolazione con modello Kalman

I modelli state-space rappresentano una classe ampia di algoritmi in grado di descrivere accuratamente molti casi reali di particolare interesse, alla pari con i modelli basati

sull'auto-regressione lineare (ARIMA), di cui costituiscono una generalizzazione. Questo tipo di modelli furono introdotti per la prima volta da Kalman nel 1960. Sebbene il primo modello fu originariamente introdotto per la ricerca nel campo delle missioni aerospaziali, fu successivamente impiegato per modellizzare dati economici, medici e nelle scienze ambientali ([6]).

Il modello state-space, nella sua forma elementare, sfrutta una relazione vettoriale auto-regressiva di primo grado come equazione di stato:

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t \quad (2.1)$$

dove l'equazione di stato determina la regola per la generazione del vettore di stato \mathbf{x}_t di dimensione $(p \times 1)$, a partire dal vettore di stato \mathbf{x}_{t-1} al tempo precedente di dimensione $(p \times 1)$, per gli istanti di tempo $t = 1, \dots, n$. Si assume che \mathbf{w}_t siano vettori rappresentanti rumore (o un qualsiasi disturbo o perturbazione) con distribuzione normale con media nulla, matrice di covarianza Q , indipendenti e identicamente distribuiti (iid). Nei modelli dinamici lineari, assumiamo che il processo parta con il vettore normale \mathbf{x}_0 che ha media μ_0 e matrice di covarianza Σ_0 di dimensione $(p \times p)$. Inoltre, i modelli state-space introducono un ulteriore componente additiva al modello, assumendo che non sia possibile osservare direttamente il vettore di stato \mathbf{x}_t , ma soltanto una sua trasformazione lineare con presenza di rumore additivo:

$$\mathbf{y}_t = A_t \mathbf{x}_t + \mathbf{v}_t \quad (2.2)$$

dove A_t è una matrice di osservazione (o misure) di dimensione $(q \times p)$. Il vettore dei dati osservati \mathbf{y}_t è di dimensione $(q \times 1)$, che può essere maggiore o minore della dimensione p della variabile di stato \mathbf{x}_t . Il rumore additivo osservato, \mathbf{v}_t , è assunto essere rumore bianco, con distribuzione Gaussiana, e con matrice di covarianza R di dimensione $(q \times q)$. Inoltre, si assume inizialmente, per semplicità, che \mathbf{x}_0 , \mathbf{w}_t e \mathbf{v}_t siano non-correlati;

questa ipotesi é in linea di principio non necessaria, ma rende piú facile la comprensione del modello state-space ([6]).

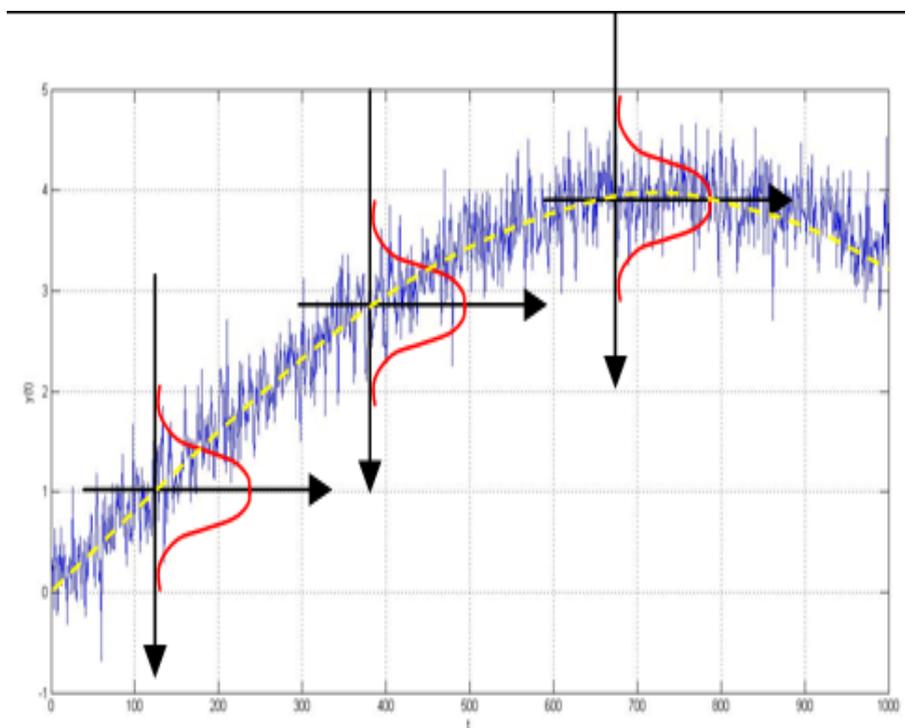


Figura 2.5: Rappresentazione grafica di un dataset con modello trend piú rumore additivo con distribuzione normale.

Come per i modelli ARIMA, variabili di input fissate, possono essere introdotte nelle variabili di stato \mathbf{x}_t o in quelle osservate \mathbf{y}_t . In tal caso, supponiamo di avere un vettore di input \mathbf{u}_t di diemnsione $(r \times 1)$ e possiamo quindi ridefinire il modello come segue:

$$\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \Omega \mathbf{u}_t + \mathbf{w}_t \quad (2.3)$$

$$\mathbf{y}_t = A_t \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{w}_t \quad (2.4)$$

dove Ω é una matrice di dimensione $p \times r$ e Γ una matrice $(q \times r)$. Sebbene esista un'equivalenza tra modelli ARIMA e modelli state-space stazionari, qualche volta risulta

piú semplice trattare dei dati con un tipo di modelli piuttosto che con gli altri. Nel caso di serie temporali con dati mancanti, sistemi multi-variati complessi, effetti misti e alcuni tipi di non-stazionarietà, sono piú facili da trattare con lo schema definito dai modelli state-space ([6]).

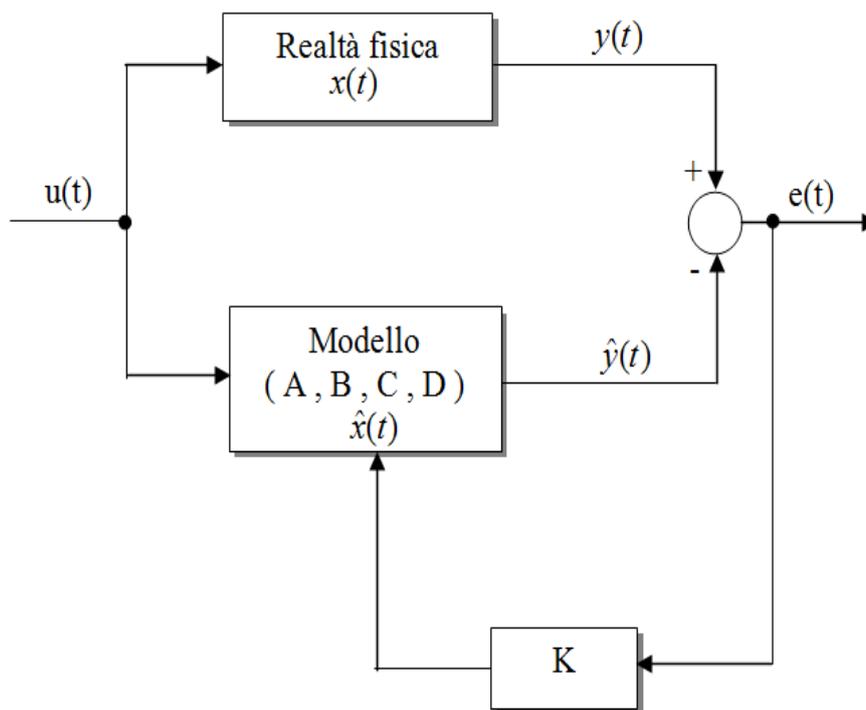


Figura 2.6: Schema del modello di Kalman.

Dal punto di vista prettamente pratico, lo scopo principale di qualsiasi analisi basata sui modelli state-space, definiti dal sistema di equazioni:

$$\begin{cases} \mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t \\ \mathbf{y}_t = A_t \mathbf{x}_t + \mathbf{v}_t \end{cases} \quad (2.5)$$

é di ottenere degli stimatori del segnale non-osservato sottostante \mathbf{x}_t , a partire dai dati $Y_s = \mathbf{y}_1, \mathbf{y}_2, \dots, \mathbf{y}_s$ fino al tempo s . Quando $s < t$, il problema é detto di previsione ("forecasting"), con $s = t$ di filtraggio ("filtering") e con $s > t$ di regolarizzazione ("smoo-

thing"). Oltre a queste stime, si vuole anche misurare il loro relativo grado di precisione nella stima ottenuta rispetto al dato originario \mathbf{x}_t . La risoluzione di questo tipo di problemi é rappresentato dall'applicazione di "filtri" e "smoother" di tipo Kalman. Per far ciò, introduciamo la seguente notazione:

$$\mathbf{x}_t^s = E(\mathbf{x}_t | Y_s) \quad (2.6)$$

e

$$P_{t_1, t_2}^s = E[(\mathbf{x}_{t_1} - \mathbf{x}_{t_1}^s)(\mathbf{x}_{t_2} - \mathbf{x}_{t_2}^s)'] \quad (2.7)$$

Quando $t_1 = t_2 (= t)$ possiamo scrivere P_{t_1, t_2}^s come P_t^s , per semplicitá. Per ottenere le equazioni di filtraggio e smoothing, ci si basa fortemente sull'ipotesi di Gaussianitá. Anche nel caso non-Gaussiano, gli stimatori che si ottengono sono gli stimatori del minimo errore quadratico medio, all'interno della classe di stimatori lineari. In tal senso, si puó pensare al valore di aspettazione E come un operatore di proiezione e Y_s come lo spazio delle combinazioni lineari di $\mathbf{y}_1, \dots, \mathbf{y}_s$; in questo caso, P_t^s é l'errore quadratico medio corrispondente ([6]). Se si assume che i processi siano di tipo Gaussiano, P puó essere pensato come una covarianza di errore condizionale, cioè si ha:

$$P_{t_1, t_2}^s = E[(\mathbf{x}_{t_1} - \mathbf{x}_{t_1}^s)(\mathbf{x}_{t_2} - \mathbf{x}_{t_2}^s)' | Y_s] \quad (2.8)$$

Ciò puó essere visto, per esempio, considerando che la matrice di covarianza tra $(\mathbf{x}_t - \mathbf{x}_t^s)$ e Y_s , per ogni t e s , é nullo; si puó affermare quindi che questi sono ortogonali. Inoltre, ciò implica che $(\mathbf{x}_t - \mathbf{x}_t^s)$ e Y_s sono indipendenti, per la normalitá, e quindi la distribuzione condizionale di $(\mathbf{x}_t - \mathbf{x}_t^s)$ dato Y_s é la distribuzione non-condizionale di $(\mathbf{x}_t - \mathbf{x}_t^s)$.

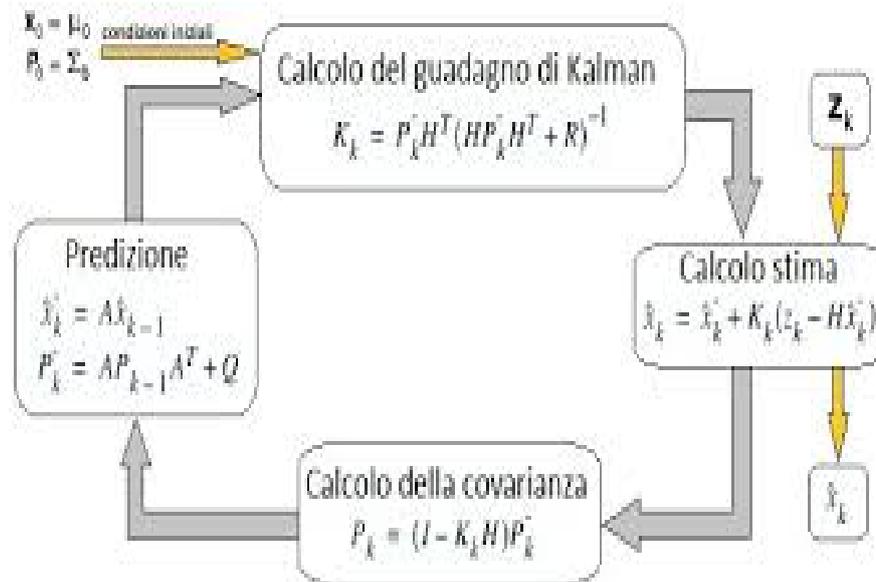


Figura 2.7: Fasi principali del modello di Kalman, con le relative operazioni eseguite (filtraggio, predizione, smoothing, correzione).

Su queste basi si possono derivare le equazioni per il filtraggio e lo smoothing dal punto di vista Bayesiano oppure tramite un approccio basato sul concetto di proiezione e la teoria della distribuzione normale multivariata. Come prima cosa, presentiamo il filtro di Kalman, che fornisce le equazioni di filtraggio e forecasting ([6]). Il nome "filtro" deriva dal fatto che \mathbf{x}_t^t è un filtro lineare delle variabili osservabili $\mathbf{y}_1, \dots, \mathbf{y}_t$, cioè:

$$\mathbf{x}_t^t = \sum_{s=1}^t B_s \mathbf{y}_s \quad (2.9)$$

per scelte opportune delle matrici B_s di dimensione $(p \times q)$. Il vantaggio rappresentato dall'utilizzo del filtro Kalman è che esso specifica come aggiornare il filtro da \mathbf{x}_{t-1}^{t-1} a \mathbf{x}_t^t una volta che una nuova osservazione \mathbf{y}_t è ottenuta, senza dover riprocessare l'intero data-set $\mathbf{y}_1, \dots, \mathbf{y}_t$.

Per il modello state-space rappresentato dalle equazioni 2.3 e 2.4, con condizioni iniziali $\mathbf{x}_0 = \mu_0$ e $P_0 = \Sigma_0$ per $t = 1, \dots, n$:

$$\mathbf{x}_t^{t-1} = \Phi \mathbf{x}_{t-1}^{t-1} + \Omega \mathbf{u}_t \quad (2.10)$$

$$P_t^{t-1} = \Phi P_{t-1}^{t-1} \Phi' + Q \quad (2.11)$$

con

$$\mathbf{x}_t^t = \mathbf{x}_t^{t-1} + K_t(\mathbf{y}_t - A_t \mathbf{x}_t^{t-1} - \Gamma \mathbf{u}_t) \quad (2.12)$$

$$P_t^t = [I - K_t A_t] P_t^{t-1} \quad (2.13)$$

dove

$$K_t = P_t^{t-1} A_t' [A_t P_t^{t-1} A_t' + R]^{-1} \quad (2.14)$$

é detto "guadagno di Kalman". La previsione per $t > n$ é ottenuta tramite la 2.10 e la 2.11 con condizioni iniziali x_n^n e P_n^n . Il filtro di Kalman restituisce in output gli "aggiornamenti" (detti anche "innovazioni") o "errori di previsione" ϵ_t :

$$\epsilon_t = \mathbf{y}_t - E(\mathbf{y}_t | Y_{t-1}) = \mathbf{y}_t - A_t \mathbf{x}_t^{t-1} - \Gamma \mathbf{u}_t \quad (2.15)$$

e le relative matrici di varianza-covarianza:

$$\Sigma_t = \text{var}(\epsilon_t) = \text{var}[A_t(\mathbf{x}_t - \mathbf{x}_t^{t-1}) + \mathbf{v}_t] = A_t P_t^{t-1} A_t' + R \quad (2.16)$$

per $t = 1, \dots, n$.

Le precedenti equazioni per il filtro Kalman continuano a valere anche nel caso in cui alcuni o tutti i parametri dell'equazione di stato e dell'equazione delle osservazioni siano dipendenti dal tempo ([6]).

Un tipico schema di aggiornamento é costituito dallo schema Bayesiano. Supponiamo di avere una serie uni-variata y_t di dati osservati che consiste di una componente di "trend", μ_t , e una componente di rumore, v_t , per cui si ha:

$$y_t = \mu_t + v_t \quad (2.17)$$

con $v_t \sim iid \mathcal{N}(0, \sigma_v^2)$. In particolare, assumiamo che il trend sia un random-walk dato da:

$$\mu_t = \mu_{t-1} + w_t \quad (2.18)$$

dove $w_t \sim iid \mathcal{N}(0, \sigma_w^2)$ é indipendente da v_t . Il modello, é ovviamente, di tipo state-space con equazione di stato data dalla 2.17. Per la fase di forecasting, calcoliamo la densitá a posteriori $p(\mu_t|Y_{t-1})$. Si conclude, che l'aggiornamento del forecasting é dato da:

$$\mu_t|Y_{t-1} \sim \mathcal{N}(\mu_t^{t-1}, P_t^{t-1}) \quad (2.19)$$

dove $\mu_t^{t-1} = \mu_{t-1}^{t-1}$ e $P_t^{t-1} = P_{t-1}^{t-1} + \sigma_w^2$.

Per l'aggiornamento del filtro si ottiene, invece, che:

$$\mu_t|Y_t \sim \mathcal{N}(\mu_t^t, P_t^t) \quad (2.20)$$

con

$$\mu_t^t = \mu_t^{t-1} + K_t(y_t - \mu_t^{t-1}) \quad (2.21)$$

dove abbiamo definito

$$K_t = \frac{P_t^{t-1}}{P_t^{t-1} + \sigma_v^2} \quad (2.22)$$

e

$$P_t^t = \left(\frac{1}{\sigma_v^2} + \frac{1}{P_t^{t-1}} \right)^{-1} = \frac{\sigma_v^2 P_t^{t-1}}{P_t^{t-1} + \sigma_v^2} = (1 - K_t) P_t^{t-1} \quad (2.23)$$

Successivamente, consideriamo il problema di ottenere stimatori per \mathbf{x}_t basati sull'intero campione di dati $\mathbf{y}_1, \dots, \mathbf{y}_n$, dove $t \leq n$, indicato con \mathbf{x}_t^n . Questi stimatori sono detti "smoother", poiché un plot della sequenza $\mathbf{x}_t^n; t = 1, \dots, n$ è tipicamente più regolare di quello delle previsioni $\mathbf{x}_t^{t-1}; t = 1, \dots, n$ o dei filtri $\mathbf{x}_t^t; t = 1, \dots, n$. Come è ovvio, il processo di smoothing implica che ogni valore stimato sia una funzione del presente, passato e futuro, mentre lo stimatore filtrato dipende solo dal presente e futuro, e la previsione, invece, solo dal passato, come è usuale attendersi ([6]).

Per i modelli state-space rappresentati dalle Eq.2.3 e 2.4 con condizioni iniziali \mathbf{x}_n^n e P_n^n , ottenute tramite il filtro Kalman per $t = n, n-1, \dots, 1$, le equazioni per le componenti "smoother" sono:

$$\mathbf{x}_{t-1}^n = \mathbf{x}_{t-1}^{t-1} + J_{t-1}(x_t^n - x_t^{t-1}) \quad (2.24)$$

$$P_{t-1}^n = P_{t-1}^{t-1} + J_{t-1}(P_t^n - P_t^{t-1})J_{t-1}' \quad (2.25)$$

dove

$$J_{t-1} = P_{t-1}^{t-1} \Phi' [P_t^{t-1}]^{-1} \quad (2.26)$$

Un caso particolarmente interessante di modelli state-space è costituito dai *modelli a trend local-level*. Questo tipo di modelli possono essere schematizzati nella loro forma più semplice tramite un random-walk, definito da:

$$\mu_t = \mu_{t-1} + w_t \quad (2.27)$$

con cui rappresentiamo una serie temporale di n osservazioni e dove $w_t \sim iid \mathcal{N}(0, 1)$ e $\mu_0 \sim \mathcal{N}(0, 1)$. In tal caso, immaginiamo di osservare una serie univariata y_t consistente di una componente di trend, μ_t , e una componente di rumore, $v_t \sim iid \mathcal{N}(0, 1)$, dove:

$$y_t = \mu_t + v_t \quad (2.28)$$

Le sequenze w_t e v_t e μ_0 sono tra loro indipendenti. Quindi, si applicano il filtro e lo smoother Kalman, usando i parametri reali. Nella Fig.2.9 sono riportati in alto i valori reali di μ_t come punti e le previsioni μ_t^{t-1} con $t = 1, \dots, n$ sovrapposte sullo stesso plot come linea continua. Inoltre, sono riportati i valori $\mu_t^{t-1} \pm 2\sqrt{P_t^{t-1}}$ come linee tratteggiate. Il grafico centrale mostra il filtro Kalman μ_t^t per $t = 1, \dots, n$ come linea continua e $\mu_t^t \pm 2\sqrt{P_t^t}$ come linee tratteggiate. Il terzo grafico illustra, invece, un plot analogo ai precedenti per lo smoother μ_t^n . Un importante caratteristica dello schema dei modelli state-space é la loro capacità di trattare accuratamente serie temporali irregolari nel tempo, cioè serie con dati mancanti, la cui analisi, come già discusso in precedenza, é lo scopo principale del presente lavoro. Di seguito, analizzeremo quali sono le modificazioni necessarie che vanno introdotte nei modelli state-space multi-variatati, per serie temporali con dati mancanti ([6]).

Definiamo, ad un dato istante di tempo t , la partizione del vettore di osservazioni $\mathbf{y}_t = (\mathbf{y}_t^{(1)'}, \mathbf{y}_t^{(2)'})'$ di dimensione $(q \times 1)$, dove la prima componente $(q_{1t} \times 1)$ é quella osservata e la componente $(q_{2t} \times 1)$ é quella non-osservata, per cui $q_{1t} + q_{2t} = q$. Quindi, possiamo riscrivere l'equazione delle osservazioni partizionata come segue:

$$\begin{pmatrix} \mathbf{y}_t^{(1)} \\ \mathbf{y}_t^{(2)} \end{pmatrix} = \begin{bmatrix} A_t^{(1)} \\ A_t^{(2)} \end{bmatrix} \mathbf{x}_t + \begin{pmatrix} \mathbf{v}_t^{(1)} \\ \mathbf{v}_t^{(2)} \end{pmatrix} \quad (2.29)$$

dove $A_t^{(1)}$ e $A_t^{(2)}$ sono, rispettivamente, le matrici delle osservazioni partizionate di dimensione, rispettivamente, $(q_{1t} \times p)$ e $(q_{2t} \times p)$, e

$$\text{cov} \begin{pmatrix} \mathbf{v}_t^{(1)} \\ \mathbf{v}_t^{(2)} \end{pmatrix} = \begin{bmatrix} R_{11t} & R_{12t} \\ R_{21t} & R_{22t} \end{bmatrix} \quad (2.30)$$

che denota la matrice di covarianza degli errori di misura tra le parti osservate e quelle non-osservate.

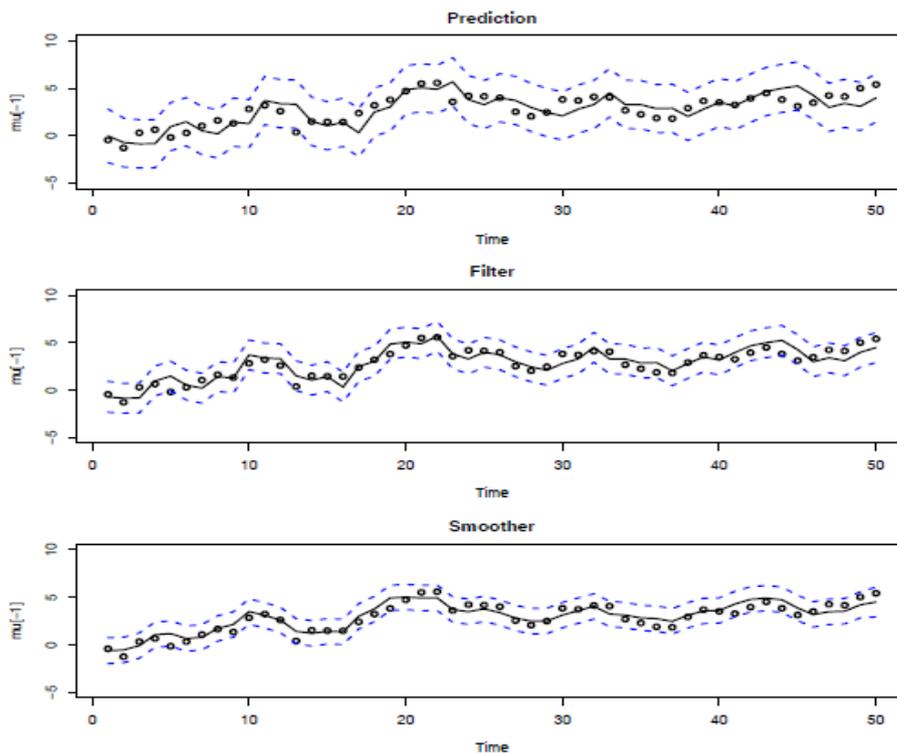


Figura 2.8: Esempio di filtraggio, previsione e regolarizzazione di un set di dati (cerchi), con curva di trend associata ad ogni singolo processo (linea continua) ed errori commessi nella stima rispetto ai dati veri (linee tratteggiate).

Nel caso di dati mancanti, dove $\mathbf{y}_t^{(2)}$ è la parte non-osservata, possiamo modificare l'equazione delle osservazioni, nelle Eq. 2.1 e 2.2, così che il modello state-space sia dato da:

$$\begin{cases} \mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \mathbf{w}_t \\ \mathbf{y}_t^{(1)} = A_t^{(1)} \mathbf{x}_t + \mathbf{v}_t^{(1)} \end{cases} \quad (2.31)$$

dove ora, l'equazione delle osservazioni é di dimensione q_{1t} al tempo t . In tal caso, si ha che le equazioni per il filtro di Kalman continuano a valere facendo opportune sostituzioni nella notazione originaria ([6]).

Se al tempo t non sono presenti delle osservazioni, quindi, imponiamo che la matrice di guadagno di Kalman, K_t , sia una matrice ($p \times q$) nulla, per cui si ha $\mathbf{x}_t^t = \mathbf{x}_t^{t-1}$ e $P_t^t = P_t^{t-1}$. Piuttosto che trattare con osservazioni di dimensione variabile é preferibile modificare il modello azzerando alcune componenti e conservando un'equazione delle osservazioni q -dimensionale. In particolare, il filtro Kalman continua a valere anche per il caso di serie temporali con dati mancanti se, al tempo t di aggiornamento, si sostituiscono:

$$\mathbf{y}_t = \begin{pmatrix} \mathbf{y}_t^{(1)} \\ \mathbf{0} \end{pmatrix}, \quad A_{(t)} = \begin{bmatrix} A_t^{(1)} \\ 0 \end{bmatrix}, \quad R_{(t)} = \begin{bmatrix} R_{11t} & 0 \\ 0 & I_{22t} \end{bmatrix} \quad (2.32)$$

per \mathbf{y}_t , A_t e R , rispettivamente, nelle Eq. 2.22, 2.23 e 2.24, dove I_{22t} é la matrice identitá ($q_{2t} \times q_{2t}$). Con le precedenti sostituzioni, i valori di aggiornamento nelle Eq. 2.25 e 2.26 saranno dati da:

$$\boldsymbol{\epsilon}_{(t)} = \begin{pmatrix} \boldsymbol{\epsilon}_t^{(1)} \\ \mathbf{0} \end{pmatrix}, \quad \Sigma_{(t)} = \begin{bmatrix} A_t^{(1)} P_t^{t-1} A_t^{(1)'} + R_{11t} & 0 \\ 0 & I_{22t} \end{bmatrix} \quad (2.33)$$

cosí da avere anche l'aggiornamento della probabilitá per il caso di serie temporali non-regolari. In questo modo, date le sostituzioni precedenti, la stima della probabilitá a partire dagli aggiornamenti della stessa puó essere effettuata, come nel caso di serie regolari ([6]).

Una volta ottenuti i valori filtrati dei dati mancanti, si stabilisce che i valori degli smoother possono essere processati, come nel caso di serie temporali complete, tramite i valori ottenuti dai dati mancanti filtrati. Nel caso di dati mancanti, gli stimatori delle variabili di stato sono indicati con:

$$\mathbf{x}_t^{(s)} = E \left(\mathbf{x}_t \mid \mathbf{y}_1^{(1)}, \dots, \mathbf{y}_s^{(1)} \right) \quad (2.34)$$

con matrice di varianza-covarianza degli errori, data da:

$$P_t^{(s)} = E \left\{ \left(\mathbf{x}_t - \mathbf{x}_t^{(s)} \right) \left(\mathbf{x}_t - \mathbf{x}_t^{(s)} \right)' \right\} \quad (2.35)$$

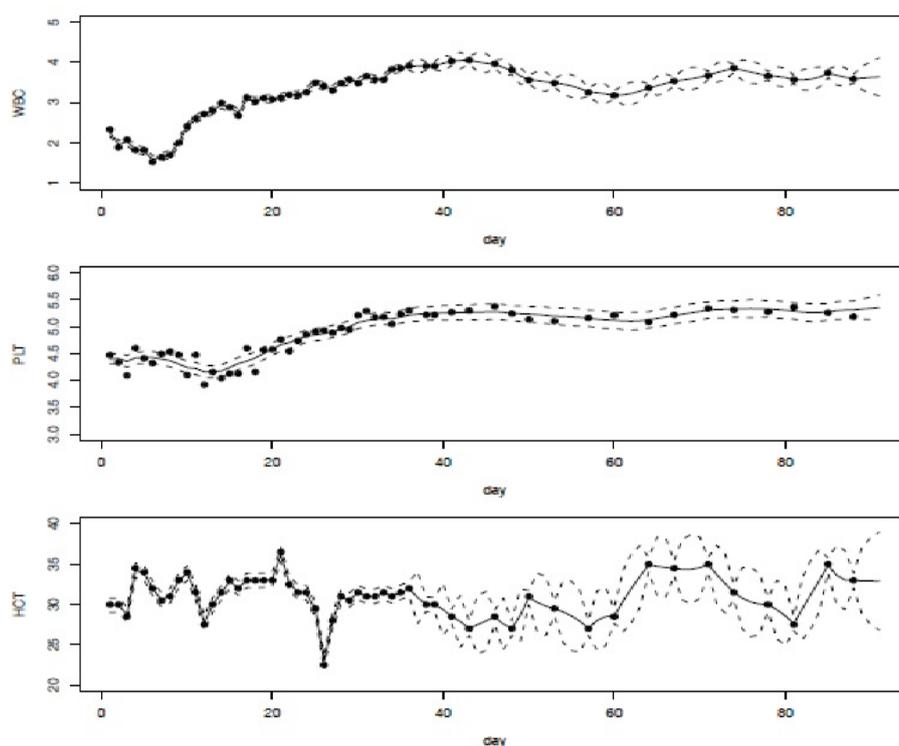


Figura 2.9: Esempio di valori regolarizzati per tre differenti grandezze: i dati (punti), valori regolarizzati (linee continue), e errori nella stima (linee tratteggiate).

Smoothed values for various components in the blood parameter tracking problem. The actual data are shown as points, the smoothed values are shown as solid lines, and 3 standard error bounds are shown as dashed lines.

2.3 Wavelet

2.3.1 Definizione e proprietà

Le **wavelet**, come suggerisce la parola stessa, sono tradotte letteralmente con il termine "ondine". Originariamente questo termine fu coniato nell'ambito degli studi geofisici per poi essere esteso allo studio di serie temporali e in molti altri ambiti dell'analisi statistica. Il loro sviluppo iniziale nel campo statistico avvenne nei primi anni '90 ([5],[7]). Prendiamo come esempio iniziale per la descrizione delle wavelet, quelle denominate "wavelet di Haar" (WH).

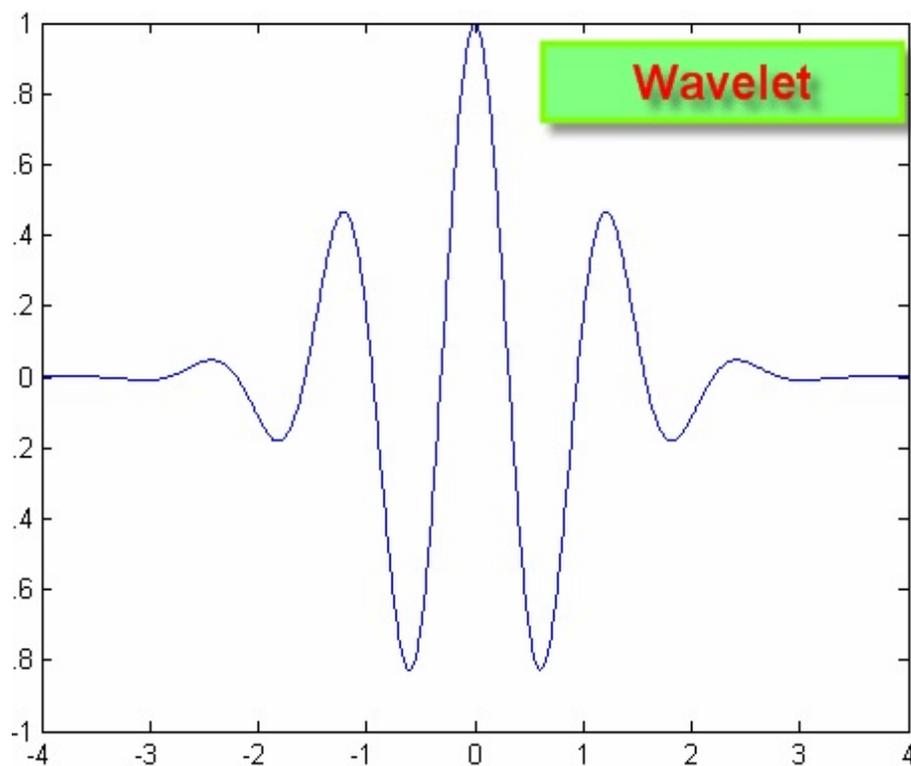


Figura 2.10: Rappresentazione semplificata di una *wavelet*

Si definisce wavelet "madre" di Haar una funzione matematica definita da:

$$\psi(x) = \begin{cases} 1 & x \in [0, \frac{1}{2}) \\ -1 & x \in [\frac{1}{2}, 1) \\ 0 & \text{altrimenti} \end{cases} \quad (2.36)$$

che costituisce la base per la nostra descrizione delle wavelet. Sebbene, la wavelet di Haar sia una forma estremamente semplificata delle wavelet, questa conserva tutte le caratteristiche essenziali delle wavelet. Due prime caratteristiche rilevanti sono l'oscillazione, infatti le WH oscillano, per cui si può affermare che:

$$\int_{-\infty}^{+\infty} \psi(x) dx = 0$$

come per tutte le altre forme di wavelet esistenti, e il supporto compatto (non tutti i tipi di wavelet sono funzioni a supporto compatto, tuttavia devono in ogni caso decadere rapidamente a zero). Quindi, le wavelet sono oggetti che oscillano ma decadono rapidamente.

A partire da una wavelet "madre", possiamo quindi generare wavelet tramite operazioni di dilatazione e traslazione. Dati gli interi j e k possiamo scrivere:

$$\psi_{j,k}(x) = 2^{j/2} \psi(2^j x - k) \quad (2.37)$$

Si può dimostrare che tali wavelet possono formare un set di funzioni di base (ortonormali), cioè:

$$\langle \psi_{j,k}, \psi_{j',k'} \rangle = \int_{-\infty}^{+\infty} \psi_{j,k}(x) \psi_{j',k'}(x) dx = \delta_{j,j'} \delta_{k,k'} \quad (2.38)$$

dove in generale $\delta_{m,n} = 1$ se $m = n$ e $\delta_{m,n} = 0$ se $m \neq n$, dove $\langle ; \rangle$ indica il prodotto interno ([5],[7]).

Inoltre, questo set di wavelet può formare delle basi per vari spazi di funzioni. Per

esempio, il set di funzioni wavelet $\{\psi_{j,k}(x)\}_{j,k \in \mathbb{Z}}$ può essere una base ortonormale completa per $L^2(\mathbb{R})$. In questo modo, data una funzione generica $f(x)$, possiamo decomporla nella seguente serie di Fourier generalizzata:

$$f(x) = \sum_{j=-\infty}^{+\infty} \sum_{k=-\infty}^{+\infty} d_{j,k} \psi_{j,k}(x) \quad (2.39)$$

dove, per l'ortogonalità delle wavelet, si ha:

$$d_{j,k} = \int_{-\infty}^{+\infty} f(x) \psi_{j,k}(x) dx = \langle f, \psi_{j,k} \rangle \quad (2.40)$$

con j e k interi. Il set di numeri $\{d_{j,k}(x)\}_{j,k \in \mathbb{Z}}$ é detto insieme dei *coefficienti wavelet* di $f(x)$.

Sebbene questa prima breve descrizione delle wavelet sia stata basata sulla forma piú semplice delle wavelet di Haar, per le altre forme esistenti di wavelet valgono le stesse considerazioni fatte. Molti tipi di wavelet sono piú appropriate per certi scopi pratici poiché sono piú regolari delle wavelet di Haar, per cui presentano anche delle proprietà di decadimento migliori nel dominio di Fourier cosí come nel dominio temporale ([5],[7]). A questo pnto rispondiamo ad una semplice domanda: *perché utilizziamo le wavelet?*. Per alcune tipologie di problemi, l'impiego delle wavelet é effettivamente vantaggioso, anche se in alcuni casi esistono altri metodi piú efficaci. Tuttavia, le wavelet si sono dimostrate essere uno strumento molto utile in quasi tutte le classi di problemi.

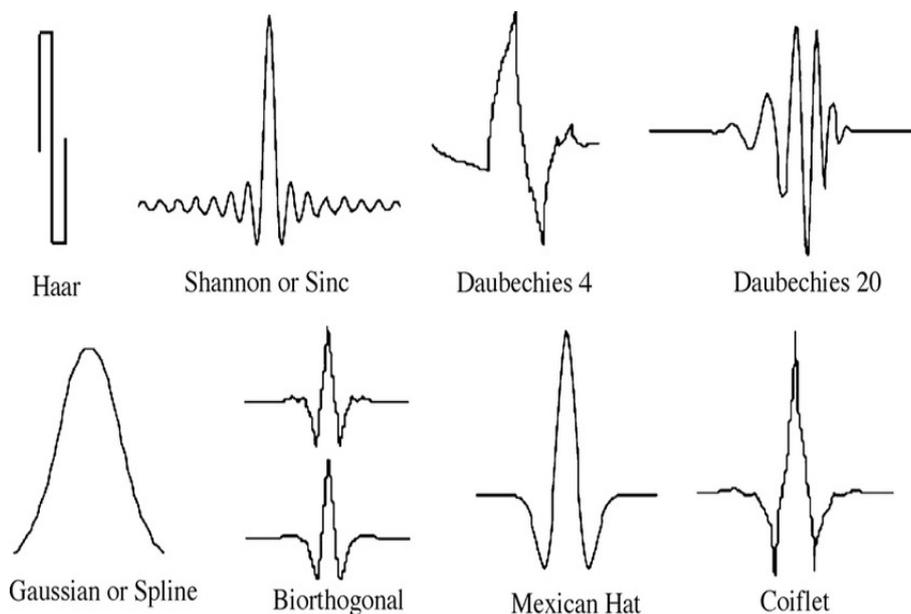


Figura 2.11: Varii tipi di *Wavelet*.

Analizziamo quindi nel dettaglio alcune delle piú importanti proprietá delle wavelet:

- Estrazione della struttura

L'equazione 2.40 mostra come calcolare i coefficienti wavelet $d_{j,k}$ di una funzione $f(x)$. Un altro modo di interpretare la Eq.2.40 é di usare il prodotto interno, e vedere che le $d_{j,k}$ rappresentano il "contenuto" di $\psi_{j,k}(x)$ che é presente all'interno di $f(x)$. Cosí, se il coefficiente $d_{j,k}$ é grande, ció significa che c' é qualche variazione ondulatoria (oscillazione) presente in $f(x)$ intorno a $2^{-j}k$ (assumendo che la wavelet sia localizzata intorno a '0'), con una lunghezza d'onda dell'oscillazione proporzionale a 2^{-j} .

- Localizzazione

Se la funzione $f(x)$ presenta una discontinuitá, questa influenzerá soltanto le wavelet $\psi_{j,k}(x)$ vicine ad essa. Soltanto quei coefficienti $d_{j,k}$, le cui wavelet associate $\psi_{j,k}(x)$ si sovrappongono alla regione della discontinuitá saranno influenzate da

questa ([5],[7]). Per esempio, per le WH, i soli coefficienti wavelet $d_{j,k}$ che possono essere influenzati da una discontinuità in x^* sono quelli per cui j, k soddisfano: $2^{-j}k \leq x^* \leq 2^{-j}(k+1)$.

Per le WH, le quali non si sovrappongono tra loro (supporti separati), soltanto una wavelet per scala si sovrappone con una discontinuità (o altre caratteristiche) della funzione $f(x)$. Questa proprietà è in contrapposizione con le basi di Fourier, costituite da funzioni seno e coseno a differenti frequenze: ogni funzione seno/coseno della base interagisce con una discontinuità, non importa dove questa sia localizzata, per cui questa influenza ogni coefficiente di Fourier della serie.

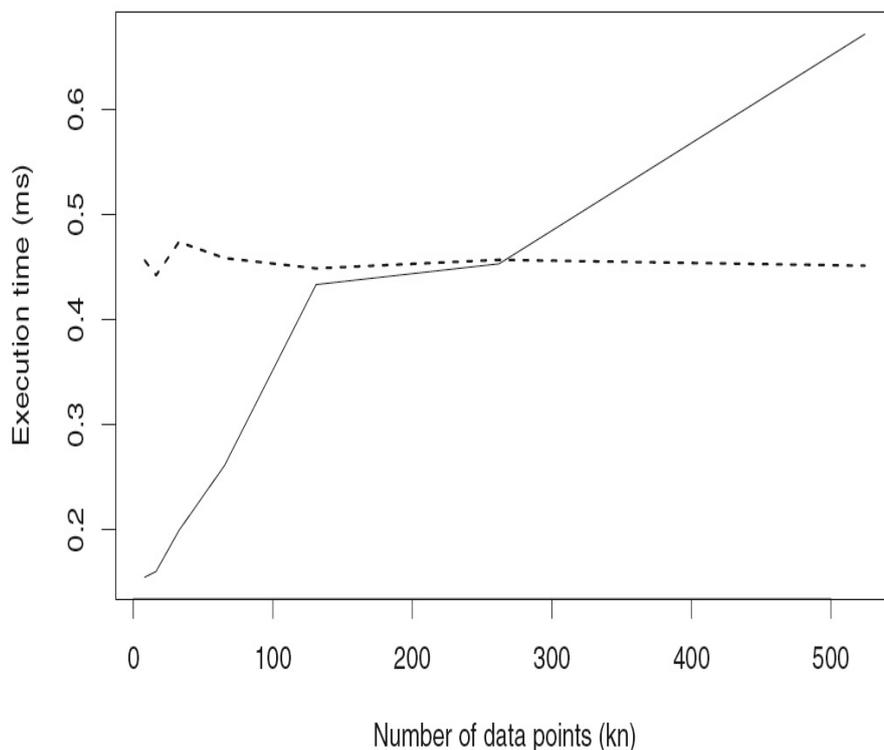


Figura 2.12: Tempi computazionali medi (divisi per n , numero di dati) per la trasformata di Fourier veloce (linea continua), e per la trasformata wavelet (linea tratteggiata)

Un campo in cui le wavelet hanno trovato grande applicazione è il trattamento delle

immagini, in cui le wavelet sono in grado di evidenziare le forti discontinuità presenti in un immagine dovute al bordo di uno o più oggetti, cioè regioni di transizione nette dove il livello di grigio dell'immagine varia rapidamente ([5],[7]). I bordi degli oggetti in un immagine sono molto evidenti nei coefficienti wavelet dell'immagine, in particolare alle scale piccole e medie. In tal caso, è possibile anche ottenere un approssimazione della funzione originaria (o immagine), usando tutte le wavelet (moltiplicate per i loro rispettivi coefficienti) fino a e inclusa una data scala specifica. Ciò è descritto da:

$$f_J(x) = \sum_{j=-\infty}^J \sum_{k=-\infty}^{+\infty} d_{j,k} \psi_{j,k}(x) \quad (2.41)$$

Così facendo si parla spesso di scale piccole, medie e grandi, o rispettivamente "grana fine", "grana media" e "grana grande", per cui alle scale più piccole corrispondono J più grandi. Ciò evidenzia come le wavelet siano in grado di estrarre informazioni rilevanti da un set di dati originario, a differenti scale e fornendo informazioni locali riguardo un dato sistema ([5],[7]).

- Efficienza

Nella Fig. 2.12 si possono confrontare direttamente i tempi di esecuzione delle trasformata wavelet (TW) e della trasformata di Fourier (TF) per uno stesso dataset di dimensione n (per differenti valori di n), ripetendo gli stessi calcoli più volte, e calcolando i tempi medi di esecuzione. Come si evince facilmente dal grafico, i tempi di esecuzione medi per le wavelet sembrano essere abbastanza costanti. Ciò suggerisce che il tempo computazionale per la TW deve essere proporzionale (cioè scalare) a n . Tuttavia, il tempo di esecuzione per la TF veloce, invece, aumenta all'aumentare di n . Si può dimostrare che lo "sforzo" computazionale della TW discreta è dell'ordine di n confrontato all'ordine $n \log(n)$ della TF veloce. Da queste

considerazioni, si può affermare che la TW è più veloce della TF veloce. Tuttavia, è necessario prestare particolare attenzione poiché:

1. le due trasformate eseguono differenti operazioni;
2. come risulta evidente dalla Fig. 2.12, la TF è più veloce della TW per $n \leq 125000$. Tuttavia, è chiaro come la TW sia comunque un algoritmo veloce.

I coefficienti crescono in numero con il livello di risoluzione con potenze 2^j al livello j . Questo tipo di problema è noto come "algoritmo a piramide" dovuto alla particolare distribuzione dei coefficienti wavelet. Nel caso della serie discreta di dati di 512 campioni, si hanno 511 coefficienti visibili e 1 non rappresentato nel plot. Ciò significa che abbiamo 512 coefficienti in totale pari al numero di campioni dalla funzione originaria. L'algoritmo a piramide richiede soltanto un numero fissato di calcoli per generare ogni coefficiente. Ciò conferma che la TW discreta può essere eseguita utilizzando N operazioni computazionali ([5],[7]).

- Sparsità

Un aspetto caratteristico delle wavelet è la natura "sparsa" delle trasformate wavelet di molte funzioni reali. In alto nella Fig. 2.13 è riportata una semplice funzione polinomiale a tratti $y(x)$ che presenta una discontinuità pronunciata da $y = 1$ a $y = 1/2$ a circa $x = 3/4$. In basso nella stessa figura sono invece illustrati i coefficienti wavelet associati alla funzione $y(x)$. Ogni singolo coefficiente è rappresentato da una piccola linea verticale. I coefficienti $d_{j,k}$ corrispondenti allo stesso livello di risoluzione j sono distribuiti lungo un'immaginaria linea orizzontale. Per esempio, i coefficienti a livello di risoluzione più alto (grana fine) corrispondenti a $j = 8$ sono posizionati lungo linea orizzontale più in basso. I coefficienti con 2^{-j} vicino 0 compaiono a sinistra del plot, mentre quelli vicini a 1 a destra. Si può vedere quindi

che i coefficienti sono vicini tra loro alle scale piú fini, poiché 2^{-j} diviene piú piccolo al crescere di j . Come si può facilmente notare dal plot dei coefficienti wavelet, ci sono pochi coefficienti non-nulli, all'incirca 10, e un coefficiente che risulta essere il piú significativo. Nel caso specifico della Fig. 2.13, si hanno 512 campioni non-nulli della funzione, di cui gran parte di questi sono maggiori di 0.2, che sono trasformati in circa 10 coefficienti wavelet non-nulli, di cui la maggior parte dei restanti sono nulli. Se si riportassero tutti i coefficienti wavelet su uno stesso plot, dove ognuno di questi é riscalato opportunamente, vedremo che i coefficienti piú significativi sono quelli in corrispondenza della discontinuitá della funzione. Ciò mette in evidenza che i coefficienti wavelet possono essere grandi quando le loro corrispondenti wavelet si sovrappongono alla caratteristica di interesse, quali per esempio le discontinuitá o salti, di una funzione o set discreto di dati; in altre parole, i coefficienti sono "clusterati" intorno alle posizioni che presentano forti discontinuitá, come per esempio i bordi di oggetti in un'immagine. La proprietá di sparsitá delle wavelet dipende dal fatto che le wavelet sono un set di basi non-condizionali per molti spazi di funzioni: "una base ortogonale che é anche una base non-condizionale per una classe di funzioni, é migliore di altre basi ortogonali nel rappresentare le funzioni di questa classe, poiché consente di comprimere l'energia in un numero minore di coefficienti". Le serie wavelet presentano convergenza non-condizionale, che significa che le somme parziali della serie convergono senza tener conto dell'ordine in cui i termini sono presi. Questa importante proprietá permette di eseguire delle procedure per ottenere stime ben-definite "immagazzinando" termini wavelet in ordine di dimensione (assoluta) dei coefficienti wavelet associati. Ciò non può essere sempre ottenuto, con altre basi, quali per esempio quelle di Fourier, per certi spazi di funzioni. Eseguendo la TW su una serie di dati non sempre si ottiene un insieme di coefficienti wavelet sparsi. Infatti, i coefficienti appaiono essere equidistribuiti e con una distribuzione

piú estesa. Infatti, se la distribuzione di dati iniziali é invariante alla trasformata wavelet, non si ha una "compressione". Caso tipico é costituito da una serie di variabili aleatorie con distribuzione normale standard indipendenti e identicamente distribuiti.

La TW conserva l'"energia" (norma quadra o somma dei quadrati dei valori) e, inoltre, puó comprimere un segnale in pochi coefficienti con valori grandi, mentre il rumore (aleatorio) rimane non-compresso. Dunque, fare la TW spesso migliora il rapporto segnale-rumore. La somma dei valori al quadrato della funzione ("energia") é la stessa dei coefficienti wavelet, ma molti di questi sono nulli o molto piccoli. Quindi, dato che l'energia é conservata, e molti coefficienti wavelet sono molto piccoli, alcune di questi devono essere necessariamente grandi.

In sintesi, le proprietá delle wavelet si possono riassumere cosí:

- sparsitá della rappresentazione dei coefficienti per un'ampia classe di funzioni incluse quelle con discontinuitá.
- La capacitá di "zoomare" per analizzare una funzione a differenti scale e inoltre la possibilitá di manipolare le informazioni contenute a tali scale.
- capacitá di rilevare o creare, e quindi, rappresentare caratteristiche locali di una funzione o di un set di dati.
- efficienza in termini di velocitá o tempi computazionali e di utilizzo della memoria per elaborare e immagazzinare dati.

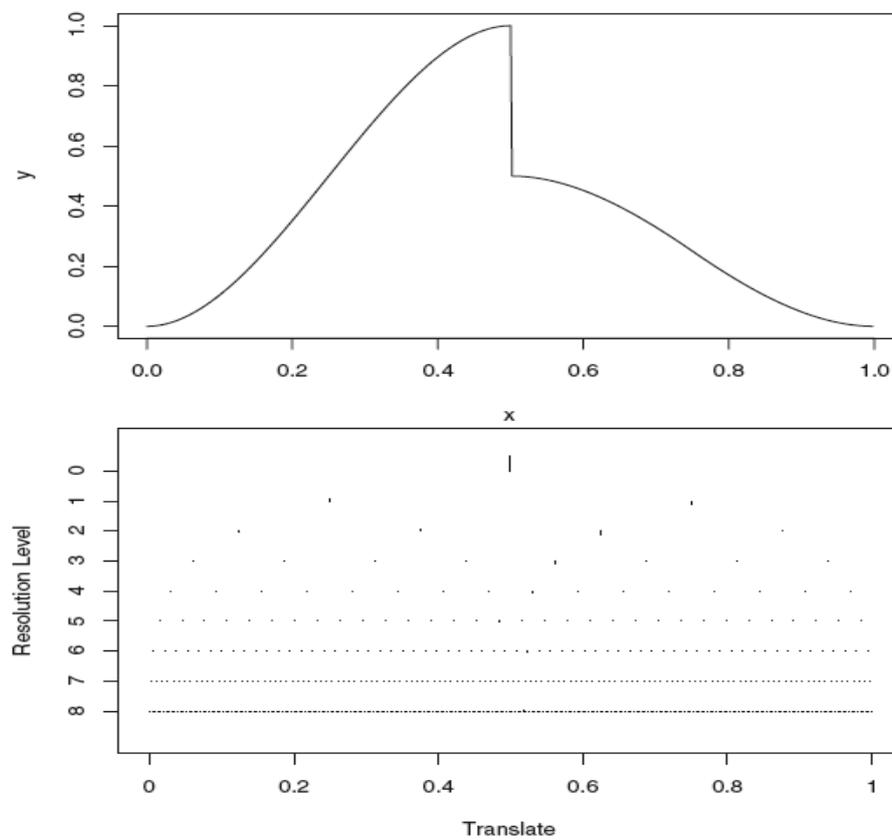


Figura 2.13: *In alto:* funzione polinomiale a tratti campionata in con 512 punti in $[0, 1]$. *In basso:* coefficienti wavelet della funzione polinomiale a tratti. Tutti i coefficienti sono rappresentati sulla stessa scala.

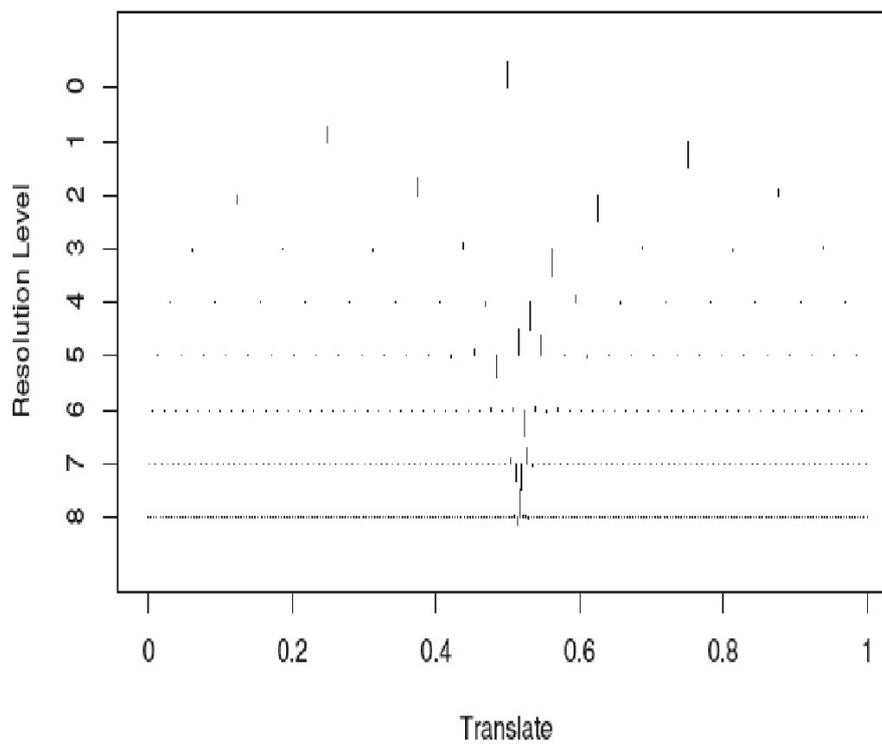


Figura 2.14: Coefficienti wavelet della funzione polinomiale a tratti. Ogni livello di scala dei coefficienti é stato ridimensionato separatamente per rendere i coefficienti piú grandi (in dimensione assoluta) di ogni riga con la stessa dimensione apparente nel plot.

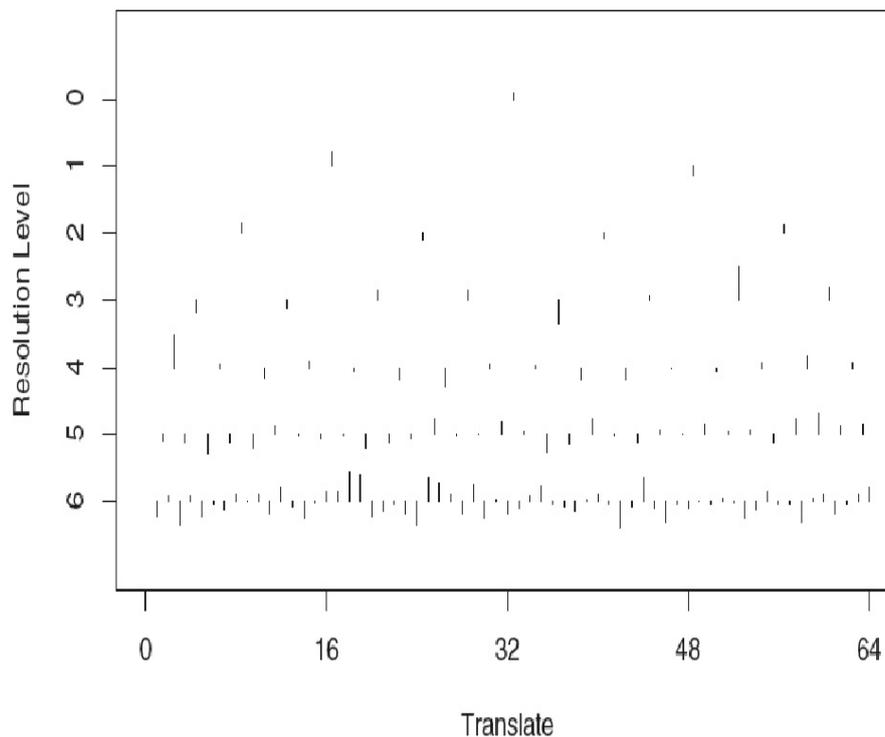


Figura 2.15: Coefficienti wavelet di una sequenza di 128 variabili normali standard indipendenti.

2.3.2 Trasformate multiscala

Prima di introdurre una trattazione formale piú rigorosa per le funzioni wavelet e per la TW, introdurremo i concetti fondamentali e il meccanismo di funzionamento dell'*analisi multiscala* ([5],[7]). Per i nostri scopi, dato che siamo interessati ad analizzare le proprietá statistiche di set di dati, tratteremo l'utilizzo delle TW per l'analisi di data-set.

Prendiamo una sequenza discreta di dati (vettori):

$$y = (y_1, y_2, \dots, y_n)$$

dove ogni y_i é un numero reale e i é un intero compreso tra 1 e n . Per semplicitá, assumiamo che il dataset sia una sequenza di n dati con $n = 2^J$ per $J \geq 0$. Chiamiamo

questo tipo di sequenza *diadica*.

L'informazione chiave che viene estratta da un'analisi multiscala é il dettaglio della sequenza a differenti scale e posizioni. Per "dettaglio" si intende il "grado di differenza" o "variazione" delle osservazioni del vettore a date scale e posizioni. Il primo step per ottenere il dettaglio é:

$$d_k = y_{2k} - y_{2k-1} \quad (2.42)$$

per $k = 1, \dots, n/2$. Se y_{2k} é molto vicino a y_{2k-1} , il coefficiente d_k sará molto piccolo. Se $y_{2k} = y_{2k-1}$, d_k é esattamente nullo. Se y_{2k} é molto diverso da y_{2k-1} , allora d_k sará molto grande. In d_k sono quindi riportate le differenze tra coppie successive di dati del vettore originario y . Tuttavia, $d_{k_{k=1}}^{n/2}$ non é il convenzionale vettore delle differenze. Infatti, in d_k non sono presenti le differenze come $y_3 - y_2$. La sequenza d_k rappresenta la differenza o dettaglio nelle posizioni $(2k - 1/2)$. Il set d_k fornisce informazioni a differenti posizioni. Tuttavia, ogni d_k ci dá soltanto informazioni riguardo un particolare y_{2k} e il suo vicino. Dato che nella sequenza non ci sono dati piú vicini di d_k , la sequenza d_k ci dá informazioni nei punti y_{2k} e intorno ad essi a scale di dettaglio piú piccole possibile ([5],[7]). A partire dalle scale piú piccole é possibile ottenere informazioni a scale piú grandi?

Il successivo step é costruire le somme:

$$c_k = y_{2k} + y_{2k-1} \quad (2.43)$$

per $k = 1, \dots, n/2$. In questo caso, la sequenza $c_{k_{k=1}}^{n/2}$ é un insieme di medie locali a differenti scale (scalate perché non divise per due, che si ha per una media), e in c_k sono contenute informazioni piú grossolane di quelle contenute nel vettore originario y . Dunque, l'operazione che trasforma y_k in c_k é analoga ad un'operazione di "regolarizzazione" ("smoothing") a media mobile, eccetto, come nel caso delle differenze d_k , per il fatto che mediamo su coppie consecutive che non si sovrappongono, al contrario delle medie mobili

convenzionali che mediano su coppie consecutive dei dati originari che si sovrappongono. Un aspetto importante é che ogni c_k contiene informazioni che si originano da y_{2k} e y_{2k-1} . In altre parole, include informazioni da due dati o osservazioni consecutive. Se si vuole ottenere un "dettaglio" piú grossolano di quello contenuto in d_k , dobbiamo confrontare due c_k vicini ([5],[7]).

Per far ciò dobbiamo introdurre un ulteriore notazione. Come prima cosa definiamo il dettaglio alla scala piú piccolo d_k . Possiamo quindi introdurre il dettaglio alla scala piú grande. Successivamente, definiremo il dettaglio per scale sempre piú grandi. Dunque, dobbiamo in qualche modo conservare traccia della scala del dettaglio a cui ci si trova. Ciò viene fatto introducendo un ulteriore pedice, j . La sequenza originale y consiste di $n = 2^J$ osservazioni. Il dettaglio al livello piú piccolo d_k , consiste di $n/2 = 2^{J-1}$ osservazioni, così il pedice aggiuntivo che scegliamo per il dettaglio del livello piú piccolo é $j = J - 1$, per cui indicheremo d_k con $d_{J-1,k}$ oppure utilizzeremo equivalentemente $d_{j,k}$; allo stesso modo, le medie a livello piú piccolo, dette "smooths" c_k sono rinominate con

$$c_{J-1,k}$$

.

Per ottenere il dettaglio alla scala piú grande successiva, si ripete l'operazione descritta in 2.42 applicata alle medie di livello piú piccolo, $c_{J-1,k}$:

$$d_{J-2,l} = c_{J-1,2l} - c_{J-1,2l-1} \quad (2.44)$$

con $l = 1, \dots, n/4$. Di nuovo, $d_{J-2,l}$ rappresenta la differenza o dettaglio della scala a cui ci si trova, tra $c_{J-1,2l}$ e $c_{J-1,2l-1}$ allo stesso modo del coefficiente di dettaglio piú piccolo definito nella 2.42.

A prima vista, nella 2.44 non risulta evidente che $d_{J-2,l}$ é a una differente scala da $d_{J-1,k}$. Tuttavia, scrivendo c_{J-1} in termini delle sue parti costituenti come definito in 2.43, si ha:

$$d_{J-2,l} = (y_{4l} + y_{4l-1}) - (y_{4l-2} + y_{4l-3}) \quad (2.45)$$

per lo stesso l come in 2.44. Ora risulta chiaro che $d_{J-2,l}$ é un insieme di differenze delle componenti (al variare di l) che sono delle medie di due dati della serie originaria. Quindi, questi $d_{J-2,k}$ possono essere pensati come "differenze alla scala 2" e $d_{J-1,k}$ come "differenze alla scala 1". In conclusione, nell'approccio a multiscala, abbiamo differenze tra valori che esistono a scale differenti.

Va sottolineato che spesso quando si parla di analisi wavelet e trasformate multiscala, si utilizzano indifferentemente i termini "scala", "livello" e "risoluzione". Più precisamente, si indica il "livello" con j e la scala con 2^j . Tuttavia, a volte si utilizza "scala" con il significato di "livello" ([5],[7]). Indicheremo con i j più grandi positivi le scale più piccole, e con i j piccoli le scale più grandi. Possiamo quindi iterare il processo di media nella 2.43 su $c_{J-1,k}$ per ottenere:

$$c_{J-2,l} = c_{J-1,2l} + c_{J-1,2l-1} \quad (2.46)$$

per $l = 1, \dots, n/4$. Come si evince facilmente sostituendo $l = 1$ nella 2.46, si ottiene una somma delle prime quattro osservazioni del data-set originario che costituisce una sorta di "media mobile locale" (ad esclusione di un fattore $1/4$). Ripetendo le procedure 2.36 e 2.37, possiamo produrre sia coefficienti di "dettaglio" che coefficienti "smooths" per scale progressivamente più grandi. Si noti che a partire da una data scala, questa aumenta di un fattore due ogni volta e il numero di coefficienti decresce ad ogni scala di un fattore due ([5],[7]). Il processo si arresta quando l'algoritmo produce un solo coefficiente "c". Ciò avviene quando si hanno $2^0 = 1$ coefficienti, e quindi questo coefficiente corrisponde al livello di indice $j = 0$ e che indicheremo con $c_{0,1}$. Nella Fig. 2.16 sono illustrate le operazioni in 2.36 e 2.37 secondo un diagramma a blocchi per un generico step dell'algoritmo multiscala.

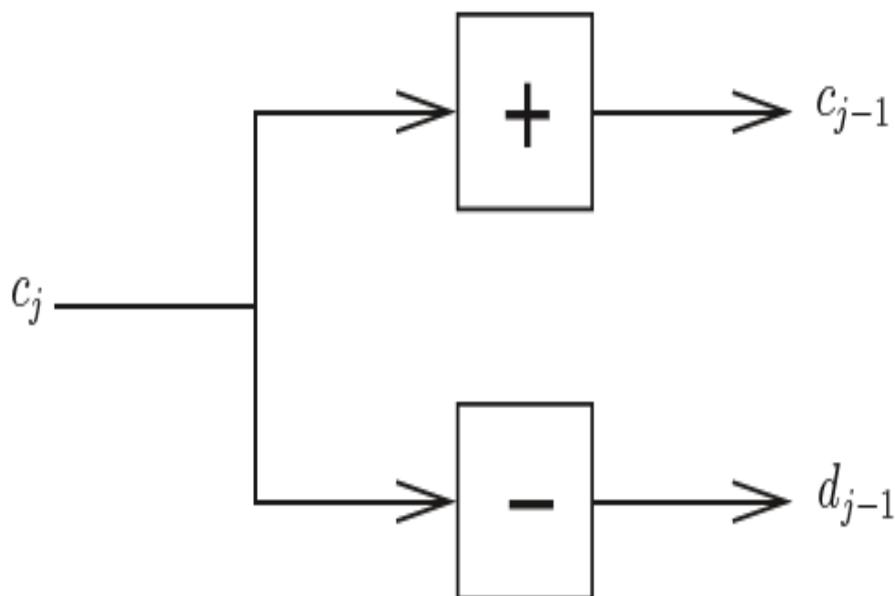


Figura 2.16: Step generico nella "trasformata multiscala". Il vettore di input, c_j , viene trasformato in due vettori di uscita, c_{j-1} e d_{j-1} , mediante le operazioni di addizione e sottrazione definite nelle Equazioni 2.36 e 2.37 per $j = J, \dots, 1$.

In sintesi, un vettore in input $c_j = (c_{j,1}, c_{j,2}, \dots, c_{j,m})$ è trasformato in due vettori di output c_{j-1} e d_{j-1} secondo le operazioni algebriche definite in precedenza. L'output c_{j-1} sarà poi dato come input ad un altro diagramma a blocchi per produrre i vettori c_{j-2} e d_{j-2} , e così via. Si noti che l'input iniziale y dell'intero algoritmo multiscala è comunque rappresentato nel diagramma a blocchi "one-step", dove $c^J = y$. Inoltre, va sottolineato che le lunghezze di c_{j-1} e d_{j-1} sono la metà della lunghezza di c_j , e così anche per tutti gli step successivi, per cui il numero di elementi in output di uno step è identico al numero di elementi in input ([5],[7]).

Nella Fig. 2.17 è raffigurata una tipica struttura piramidale invertita che rappresenta graficamente una trasformata multiscala.

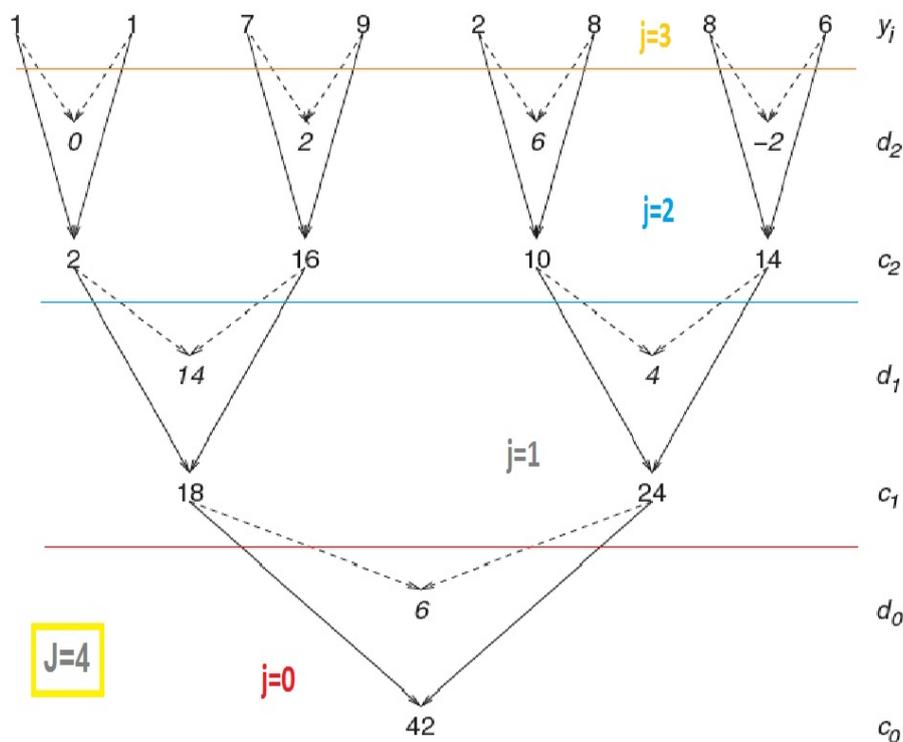


Figura 2.17: Rappresentazione grafica di una trasformazione multiscala. Le frecce tratteggiate raffigurano una sottrazione e numeri in corsivo i corrispondenti coefficienti di dettaglio $d_{j,k}$. Le frecce continue indicano l'addizione e i numeri corrispondenti indicano i coefficienti smooths $c_{j,k}$.

Si noti come a partire da 8 dati y_i in input si generino 4 coefficienti d_2 e 4 coefficienti c_2 , cioè si ottengono tanti coefficienti in output quanti sono i dati in input. I coefficienti derivati forniscono informazioni riguardo la sequenza originaria, come per esempio il coefficiente finale $c_0 = 42$ indica che la somma degli elementi dell'intera serie originaria è pari a 42. Tutto ciò può essere espresso in termini di wavelet, dicendo che i coefficienti di dettaglio $d_{j,k}$ sono i "coefficienti wavelet" e i coefficienti $c_{j,k}$ sono i cosiddetti "coefficienti di scaling" o "coefficienti wavelet padre". Finora abbiamo descritto un algoritmo definito come "trasformata wavelet discreta" (DWT) che può essere rappresentato da un algoritmo multiscala ([5],[7]). Le wavelet che sottostanno alla trasformata multiscala sono dette "wavelet di Haar". L'intera serie originaria può essere esattamente ricostruita utilizzando

soltanto i coefficienti wavelet $d_{j,k}$ e l'ultimo coefficiente $c_{0,0}$. In questo caso, avrebbe:

$$c_{j-1,2k} = (c_{j-2,k} + d_{j-2,k})/2 \quad (2.47)$$

$$c_{j-1,2k-1} = (c_{j-2,k} - d_{j-2,k})/2 \quad (2.48)$$

Tuttavia, se facciamo il calcolo dell'energia per i coefficienti in input e output otteniamo valori diversi. Ciò ci porta ad introdurre una trasformata wavelet che conservi l'energia tra dati prima e dopo l'applicazione della trasformata wavelet.

2.3.3 Wavelet di Haar discrete

Per far sì che l'energia si conservi si devono modificare le Eq. 2.36 e 2.37 in modo opportuno. Si osserva che affinché l'energia dei dati in input (y_{2k}, y_{2k-1}) data da $y_{2k}^2 + y_{2k-1}^2$ sia uguale a quella dei coefficienti in output (d_k, c_k) , data da $d_k^2 + c_k^2$ si deve introdurre la normalizzazione $1/\sqrt{2}$, per cui:

$$d_k = (y_{2k} - y_{2k-1})/\sqrt{2} \quad (2.49)$$

$$c_k = (y_{2k} + y_{2k-1})/\sqrt{2} \quad (2.50)$$

A volte è più conveniente utilizzare una differente normalizzazione per i coefficienti wavelet d_k data da:

$$d_k = \sum_{l=-\infty}^{\infty} g_l y_{2k-l} \quad (2.51)$$

dove

$$g_l = \begin{cases} 2^{-1/2} & l = 0 \\ -2^{-1/2} & l = 1 \\ 0 & \text{altrimenti} \end{cases}$$

L'Eq. 2.51 rappresenta quindi un'operazione di filtraggio con coefficienti di filtraggio dati da g_l per $l = -\infty$ a $+\infty$.

Per i coefficienti wavelet "padre" c_k si utilizza un'operazione di filtraggio simile alla precedente, data da:

$$c_k = \sum_{l=-\infty}^{\infty} h_l y_{2k-l} \quad (2.52)$$

dove

$$h_l = \begin{cases} 2^{-1/2} & l = 0 \\ 2^{-1/2} & l = 1 \\ 0 & \text{altrimenti} \end{cases}$$

Dato che l'output generato da una serie in input, si ottiene con semplici operazioni di addizione, sottrazione e riscalatura costante, non sorprende che si possa calcolare l'output dall'input tramite la moltiplicazione dei dati della serie originaria per una matrice W , per cui si ha:

$$d = Wy$$

con y vettore dei dati originari e d coefficienti della trasformata wavelet e con W matrice ortogonale, per cui questa conserva la norma dei due vettori d e y ; in altre parole, la lunghezza del vettore d è la stessa di quella del vettore y .

L'operazione precedente esegue una TW utilizzando una moltiplicazione matriciale invece dell'algoritmo multiscala. Se y è un vettore diadico, per cui la sua lunghezza è pari a

$n = 2^J$, la matrice W sarà di dimensione $n \times n$, di conseguenza il tempo computazionale necessario per eseguire l'operazione Wy è dell'ordine $O(n^2)$ ([5],[7]).

L'algoritmo multiscala, invece, è in grado di produrre gli stessi coefficienti wavelet della moltiplicazione matriciale, ma in $O(n)$ operazioni. Ciò fa della tecnica "piramidale" o multiscala un "algoritmo veloce", per cui si parla di "trasformata wavelet veloce" (FWT). Un altro esempio ben noto di algoritmo veloce è la trasformata di Fourier veloce (FFT), che calcola la trasformata di Fourier discreta (DFT) in $O(n \log n)$ operazioni. Le wavelet risultano essere più veloci della FFT, ma si deve osservare che le trasformate wavelet e di Fourier discrete calcolano trasformate abbastanza differenti tra loro. nei casi di interesse, $\log n$ è piccolo anche con n abbastanza grande ([5],[7]).

2.3.4 Wavelet Shrinkage

La TW nello schema degli algoritmi multiscala rientrano nella classe dei metodi statistici definiti "regressivi". In particolare, la TW è definita come una forma di regressione non-parametrica, e le tecniche sono denominate come *wavelet shrinkage*, stima della curva o regressione wavelet. In questo paragrafo, ci occupiamo del processo di wavelet shrinkage (WS) ([5]).

Ciò che si ha, in generale, è una funzione o serie di dati che sono "contaminati" da rumore additivo; applicando una TW, questa modifica, o più precisamente contrae i coefficienti wavelet della funzione o data-set rumoroso, e successivamente facendo la trasformata wavelet inversa (ITW) si ottiene una stima della funzione o data-set originario.

L'idea generale consiste nel considerare le osservazioni o dati $y = (y_1, \dots, y_n)$ che derivano dal modello:

$$y_i = g(x_i) + e_i \quad (2.53)$$

con $i = 1, \dots, n$ e dove $x_i = i/n$.

Lo scopo é di stimare la funzione ignota $g(x)$ per $x \in [0, 1]$, usando le osservazioni rumorose y_i . Nel modello base é di solito assunto che $e_i \sim \mathcal{N}(0, \sigma^2)$ siano indipendenti, detto "rumore bianco". L'idea generale alla base del metodo del WS discreto é che la trasformata wavelet discreta (DWT) viene applicata all'Eq. 2.53. Sia W la matrice della TW (i calcoli nella realtà sono eseguiti tramite l'algoritmo piramidale multiscala veloce di Mallat) e y , g e e rappresentano rispettivamente i vettori delle osservazioni, dei dati veri ignoti e del rumore ([5]).

Dato che la DTW é una trasformazione lineare, possiamo scrivere il modello wavelet trasformato come:

$$d^* = d + \epsilon \quad (2.54)$$

dove $d^* = Wy$, $d = Wg$ e $\epsilon = We$.

Affinché il processo di WS abbia successo, dobbiamo avere le seguenti caratteristiche:

1. Per funzioni o data-set che presentano discontinuitá a salti o altre irregolaritá, il vettore dei coefficienti wavelet d é un vettore "sparso".
2. L'*energia* della funzione o data-set vero $\sum_i g(x_i)^2$ é uguale alla somma dei quadrati dei coefficienti wavelet $\sum_{j,k} d_{j,k}^2$ secondo la relazione di Parseval. Tuttavia, per via della sparsitá dei coefficienti, l'energia del segnale originario g , é ora concentrata in pochi coefficienti senza alcuna perdita di informazione. Quindi, non solo i vettori d saranno sparsi ma i loro valori saranno grandi.
3. Dato che la W della DWT é una matrice ortogonale, ciò implica che la TW del rumore bianco, ϵ , é essa stessa rumore bianco. Ciò significa che il rumore non é concentrato come g , ma piuttosto é distribuito in modo uniforme su tutti i coefficienti wavelet.

Quindi, la tecnica dello "shrinkage" dei coefficienti wavelet fornisce una buona stima di $g(x)$. L'idea é che valori grandi dei coefficienti wavelet sperimentali d^* , siano costituiti da segnale "vero" e rumore, mentre coefficienti piccoli siano dovuti solo a rumore. Per stimare correttamente d , il processo di *thresholding*, ci porta ad ottenere una stima della funzione o data-set originario, \hat{g} . Per valutare se la stima ottenuta é valida, si calcola l'errore tra la stima $\hat{g}(x)$ e i valori veri $g(x)$ e si cerca di ottenere la stima \hat{g} che minimizza questo errore ([5]). La scelta piú comune per questo errore é l'errore quadrato integrato (ISE) dato da:

$$\hat{M} = \frac{1}{n} \sum_{i=1}^n [\hat{g}(x_i) - g(x_i)]^2 \quad (2.55)$$

L'errore di stima, cosí definito, dipende dalla stima \hat{g} che a sua volta dipende dalla sequenza degli errori e_i . Siccome, siamo interessati a conoscere cosa avviene con l'errore in media, definiamo il valro medio di ISE (MISE), o anche "rischio", come $M = E(\hat{M})$. Va sottolineato che M non é un semplice numero e puó dipendere dalla scelta del particolare stimatore utilizzato, dalla funzione vera, il numero di dati e le proprietá della serie del rumore e_i . Nel processo di WS é anche importante ricordare che, dato che l'errore M dipende dallo stimatore scelto, quindi questo non dipende solo dai parametri di "smoothing" scelti, ma anche dalla famiglia di wavelet utilizzate per eseguire il processo di regolarizzazione ([5]).

Thresholding

Introduciamo il cosiddetto processo di *thresholding* "universale" (ne esistono anche di altri tipi), in cui la "soglia" (*threshold*) é data da:

$$\lambda = \sigma \sqrt{2 \log n} \quad (2.56)$$

Nei casi reali il livello di rumore σ é stimato da $\hat{\sigma}$, cioè una stima della deviazione standard del rumore ϵ_i , e n é il numero di campioni del data-set ([5]).

Si può dimostrare che in un processo aleatorio (set di variabili random) con distribuzione Gaussiana (non necessariamente indipendenti) la piú grande delle variabili é della dimensione di circa $\sqrt{2\log n}$. Ciò implica che se i coefficienti wavelet sono coefficienti di rumore Gaussiano con varianza unitaria, se si sceglie una soglia pari a $\sqrt{2\log n}$, sará eliminato il rumore. Si può dimostrare che il rapporto segnale-rumore (SNR) é semplicemente il rapporto tra la deviazione standard campionaria del segnale e quella del rumore additivo. Un valore basso del SNR significa che la varianza del rumore risulta essere grande rispetto alla dimensione del segnale ([5]).

Nella Figura 2.18 sono plottati i coefficienti wavelet di un segnale impulsato con e senza rumore, e nell'ultimo plot l'effetto prodotto dal thresholding dei coefficienti. Come si può facilmente notare l'operazione di thresholding elimina i coefficienti piú piccoli presenti al livello di risoluzione maggiore, mentre i coefficienti caratteristici del segnale rimangono invariati o in alcuni casi divengono ancora piú evidenti, poiché i piccoli coefficienti presenti intorno a questi sono definitivamente eliminati. Eseguendo la trasformata wavelet inversa a partire dai coefficienti wavelet filtrati (regolarizzati) dal thresholding, si ricostruisce un segnale piú smooth rispetto a quello originario.

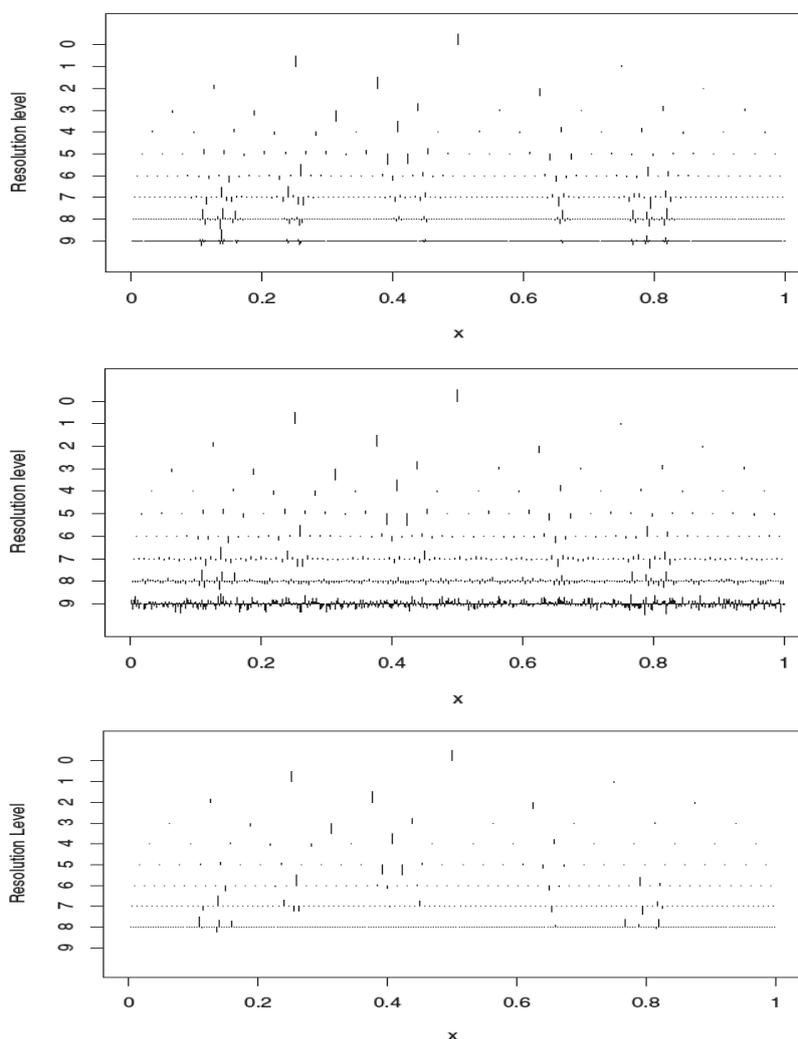


Figura 2.18: In alto: coefficienti ottenuti dalla trasformata wavelet discreta di un segnale impulsato. Al centro: coefficienti ottenuti dalla trasformata wavelet discreta di un segnale impulsato con rumore additivo. In basso: coefficienti wavelet dopo il *thresholding* del segnale impulsato rumoroso; la trasformata wavelet inversa a partire da questi coefficienti restituisce un segnale piú smooth rispetto a quello originario.

2.3.5 Dati con campionamento irregolare

Finora abbiamo discusso l'analisi con trasformata wavelet di serie di dati regolari. Nel presente lavoro siamo invece interessati a trattare l'analisi statistica di serie temporali con

dati campionati in modo irregolare nel tempo, cioè misure di una data variabile ottenute a intervalli di tempo non costanti ([5]).

Esistono tre metodi principali per trattare dati irregolari, che sono:

- trasformazioni;
- interpolazioni;
- metodi statistici con ipotesi di dati con distribuzione uniforme in x_i .

Per l'analisi statistica di dati irregolari, uno dei metodi più validi é rappresentato dal metodo di interpolazione di Kovac e Silverman ([5]). Questo metodo si basa sul prendere i dati irregolari, dove $x_i \in (0, 1)$, e di interpolare i valori y_i su una specifica griglia regolare. Quindi, il processo di wavelet shrinkage viene applicato ai valori interpolati sulla griglia regolare con thresholding dei coefficienti wavelet (questi sono i coefficienti dei valori correlati della funzione interpolata, e non i valori della funzione originaria assunti indipendenti).

Piú precisamente, questo metodo si basa sulla definizione di una griglia con spaziatura regolare t_0, \dots, t_{N-1} sull'intervallo $(0, 1)$, dove $N = 2^J$ con $J \in \mathbb{N}$. Si prendono quindi i valori discreti cosí definiti $t_k = (k + 1/2)/N$ per $k = 0, \dots, N - 1$ e si sceglie $N = 2^J$ tale che $J = \min \{j \in \mathbb{Z} : 2^j > n\}$.

A partire dai dati originari y_i su x_i , tramite interpolazione lineare si ottengono nuovi dati s_k , sulla griglia definita:

$$s_k = \begin{cases} y_1 & \text{se } t_k \leq x_1 \\ y_i + (t_k - x_i) \frac{y_{i+1} - y_i}{x_{i+1} - x_i} & \text{se } x_i \leq t_k \leq x_{i+1} \\ y_n & \text{se } t_k \geq x_n \end{cases} \quad (2.57)$$

Questa interpolazione puó essere espressa in forma matriciale come:

$$s = Ry \quad (2.58)$$

dove la matrice di interpolazione R dipende da t (sulla nuova griglia) e x (posizione dei dati originari). L'operazione di thresholding é applicata ai valori interpolati. Il primo passo del shrinkage é di fare la DWT, espressa in forma matriciale da:

$$d^* = Ws \quad (2.59)$$

dove la matrice di trasformata wavelet W é una matrice ortogonale di dimensione $N \times N$, associata alla *wavelet di Daubechies*. Se i dati originari y_i sono indipendenti e identicamente distribuiti con varianza σ^2 , si puó facilmente dimostrare che la matrice di varianza dei dati interpolati é data da:

$$\Sigma_s = \sigma^2 RR^T \quad (2.60)$$

Inoltre, le tecniche di thresholding possono essere modificate opportunamente per prendere in considerazione le differenti varianze dei coefficienti wavelet ([5]).

Questo tipo di analisi verrá descritta in modo piú approfondito nel successivo capitolo quando tratteremo l'analisi wavelet dei dati irregolari, in cui illustreremo nel dettaglio i passi seguiti per ricavare l'interpolazione wavelet di dataset campionati in modo irregolare nel tempo.

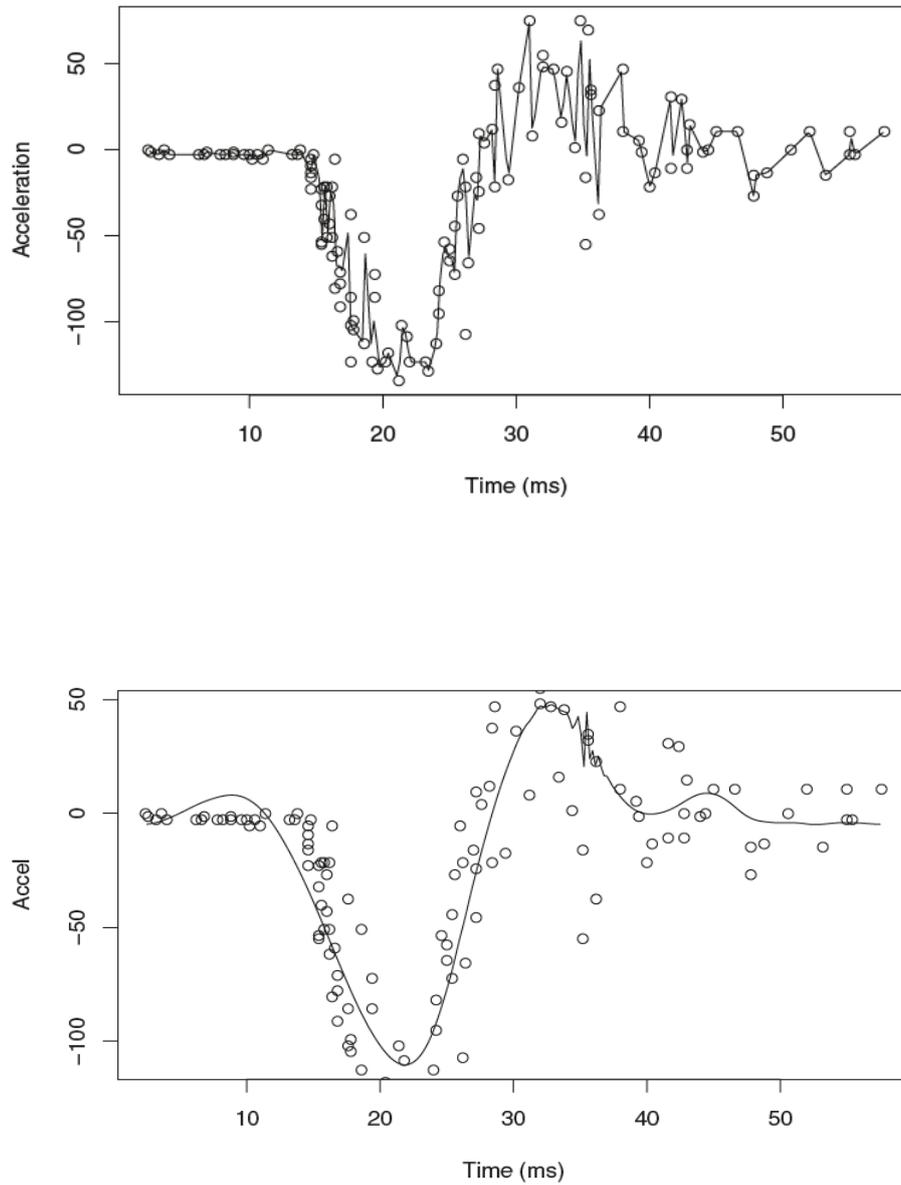


Figura 2.19: In alto: dati con campionamento irregolare (cerchi) e dati interpolati sulla griglia (linea). In basso: dati con campionamento irregolare (cerchi) e dati interpolati sulla griglia su cui è stato eseguito il *thresholding* (linea).

Capitolo 3

Descrizione delle analisi

3.1 Analisi eseguite

Nel presente lavoro sono state eseguite le seguenti analisi:

1. Interpolazione lineare;
2. Interpolazione spline;
3. Interpolazione Stineman;
4. Interpolazione Kalman;
5. Interpolazione Wavelet;

L'analisi é stata eseguita in R caricando le seguenti librerie:

- `library(astsa)`, `library(TSA)`, `library(tseries)` (librerie per la trattazione di serie temporali);
- `library(zoo)`, `library(splines)`, `library(stinepack)` (librerie per le interpolazioni lineare, spline, Stineman);
- `library(imputeTS)` (libreria per l'interpolazione Kalman);

- `library(wmts)`, `library(wavelets)`, `library(wavethresh)`, `library(waveband)`
(librerie per l'analisi wavelet).

Le serie temporali originarie delle variabili sono costituite da $N = 407337$ campioni, ma per poter effettuare un confronto tra le analisi eseguite con le interpolazioni lineare, spline, Stineman e Kalman con quella Wavelet, dobbiamo necessariamente selezionare un numero di campioni che sia pari a una potenza di due (2^n), in accordo al principio di funzionamento degli algoritmi multiscala come quello delle wavelet (vedi Capitolo 2).

Dato che le serie di dati, come già discusso in precedenza, sono state campionate con intervalli di tempo non-costanti (irregolari), per poter eseguire le tecniche di interpolazione polinomiale e quella di Kalman, abbiamo utilizzato alcune funzioni della libreria `zoo` in R, che ci hanno permesso di trasformare le serie originarie in serie campionate in modo omogeneo, introducendo dei tempi "fittizi" a intervalli regolari e attribuendo un NA (*missing-data*) in corrispondenza di questi tempi. Così facendo si sono ottenute delle serie con campionamento temporale regolare, su cui poi sono state eseguite le operazioni di interpolazione descritte precedentemente. Di seguito riportiamo le funzioni e operazioni eseguite in R per ottenere le serie regolarizzate in tempo e quelle usate per applicare i metodi di interpolazione a queste nuove serie ottenute:

```
times <- as.POSIXct(strptime(data$Time[n1:n2],
format="%Y-%m-%d_%H:%M:%S"))
data2 <- zoo(serie, times)
data3 <- merge(data2, zoo(, seq.POSIXt(min(times), max(times), 1)),
all=TRUE)
```

Nelle righe di codice precedenti sono presenti la funzione `as.POSIXct()` che consente di trasformare i valori dei tempi, espressi nel data-set originario nel formato ANNO-MESE-GIORNO ORA-MINUTI-SECONDI, in un oggetto della "classe" `POSIXct`, che le funzioni di `zoo` sono in grado di riconoscere. La successiva operazione consiste nel creare una nuova serie temporale `data2`, identica a quella originaria, ma nella "classe" `zoo`. Nella passo

successivo si uniscono (`merge`) la serie ottenuta per `data2` e una serie costituita da soli NA, che andranno a posizionarsi in corrispondenza dei tempi "fittizi" introdotti nella serie originaria. Quello che si ottiene quindi a conclusione di queste operazioni iniziali é una serie temporale `data3` regolare con campionamento temporale al secondo.

```
data_linearinterp <- na.approx(data3)
data_spline <- na.spline(data3)
data_stineman <- na.stinterp(data3)
data_Kalman <- na.kalman(data3)
```

Le funzioni di `zoo`, `na.approx()` e `na.spline()`, consentono di sostituire i dati mancanti con valori ottenuti da un'interpolazione lineare e da un'interpolazione spline cubica, rispettivamente. Queste funzioni restituiscono un oggetto con struttura analoga a quella della serie di input, dove gli NA sono stati sostituiti dai valori ottenuti dall'interpolazione.

La funzione `na.stinterp()`, utilizza il metodo dell'interpolazione Stineman, una variante dell'interpolazione cubica tale da preservare la monotonicit  del data-set che viene interpolato (normalmente la monotonicit  é garantita per l'interpolazione lineare ma non lo é per quella cubica).

La funzione `na.kalman()` esegue un processo di filtraggio (*filtering*) e regolarizzazione (*smoothing*) su dati di serie temporali strutturate. Inoltre, la funzione fornisce alcune opzioni per decidere con quale modello state-space eseguire l'operazione di ricostruzione dei dati mancanti, come per esempio i modelli ARIMA (modelli auto-regressivi) o i modelli basati sulla probabilit  massima (*maximum likelihood*). Inoltre, é possibile decidere di eseguire il metodo Kalman, impostando opportuni parametri per il *filtering* e lo *smoothing*.

Nelle successive righe di codice vengono illustrate le operazioni e le funzioni della libreria `wavethresh`, necessarie per eseguire il metodo della **trasformata Wavelet** secondo

lo schema di Kovac e Silverman (vedi Cap. 2):

```

time <- as.numeric(Time)
time01 <- (time-min(time))/(max(time)-min(time))
serieinterp <- makegrid(t=time01, y=data)
timegrid <- serie_interp$gridt*(max(time)-min(time)) + min(time)
serieirr <- irregwd(serieinterp)
serieirr2 <- serieirr
class(serieirr2) <- "wd"
seriethreshold <- threshold(serieirr2, policy="universal",
type="soft", dev=madmad)
seriewr <- wr(seriethreshold)

```

Inizialmente si convertono i valori del tempo `Time`, della serie analizzata, dal formato ANNO-MESE-GIORNO ORA-MINUTI-SECONDI a quello di vettore numerico `time`. Successivamente, si esegue la trasformazione del vettore dei tempi sulla griglia temporale nell'intervallo $[0, 1]$, ottenendo un nuovo vettore dei tempi `time01` con valori compresi tra 0 e 1. A partire dal vettore originario dei dati, `data`, i cui valori sono posizionati su una griglia bidimensionale regolare del tempo e dei dati (indicati con `gridt` e `gridy`, nell'oggetto `serieinterp`); si esegue, quindi, un'interpolazione dei dati sulla griglia, tramite la funzione `makegrid()`. Si ottiene una nuova serie temporale costituita da i valori originari posizionati sulla griglia con l'aggiunta dei valori ottenuti dall'interpolazione. A questo punto, si ricostruiscono i valori dei tempi, `timegrid` sulla scala temporale originaria a partire da quella sull'intervallo $[0, 1]$. Si esegue, quindi, la trasformata wavelet per le serie temporali irregolari, tramite la funzione `irregwd` applicata alla serie interpolata `serieinterp`. L'oggetto così ottenuto viene convertito in un oggetto della classe `wd` per eseguire le operazioni successive, come il plot dei *coefficienti wavelet* (indicati con `D` e `C` nell'oggetto della classe `wd`), il calcolo dell'energia (norma quadrata) e la stima delle bande di confidenza. Successivamente, si esegue l'operazione di *thresholding*, tramite la funzione `threshold`, sulla serie irregolare `serieirr2` ottenuta. Nella presente analisi si è scelto di applicare il processo di *thresholding* "universale", tramite l'opzione `policy`. Gli altri parametri, quali `type` permet-

tere di decidere quale tipo di soglia si vuole adottare (le scelte possibili sono **hard** e **soft**), mentre con **dev** si sceglie quale funzione usare per calcolare lo "spread" dei coefficienti in valore assoluto, e restituisce un valore in termini di varianza, che é legato al valore della soglia applicata. Infine, si esegue la trasformata wavelet inversa tramite la funzione **wr()**, che restituisce la serie temporale ricostruita **seriewr** a partire dalla serie originaria.

In conclusione, abbiamo calcolato e graficato le *bande di confidenza* (tramite la funzione **wave.band** della libreria **waveband**) per la serie ricostruita, tramite il metodo della trasformata wavelet, a partire dalla serie originaria.

```
confidence_band <- wave.band(data)
```


Capitolo 4

Risultati delle analisi

In questo capitolo descriveremo le analisi e i risultati ottenuti tramite le varie tecniche di interpolazione. Come prima cosa riportiamo in Fig. 4.1 i plot delle serie associate a quattro variabili del data-set:

- Temp_PCU (Temperatura Power Control Unit);
- BTA_Voltage (Tensione ai capi della Batteria);
- ANG_RATE_Z (Velocità angolare asse Z);
- ANG_ERR_Z (Errore angolare asse Z).

le quali sono state scelte per tipologia di grandezza misurata, che sono temperatura, tensione elettrica, velocità (parametro di avionica) ed errore angolare.

Per ogni serie è stato scelto un numero di campioni di $N = 16384$, pari a circa 1/20 dell'intero data-set. Ciò ci ha permesso di poter confrontare tra loro le serie interpolate con metodi differenti, poiché il metodo utilizzato per la trasformata wavelet, essendo basato su un algoritmo a multiscala, agisce solo su serie di dati con un numero di campioni pari ad una potenza di due ($N = 2^n$).

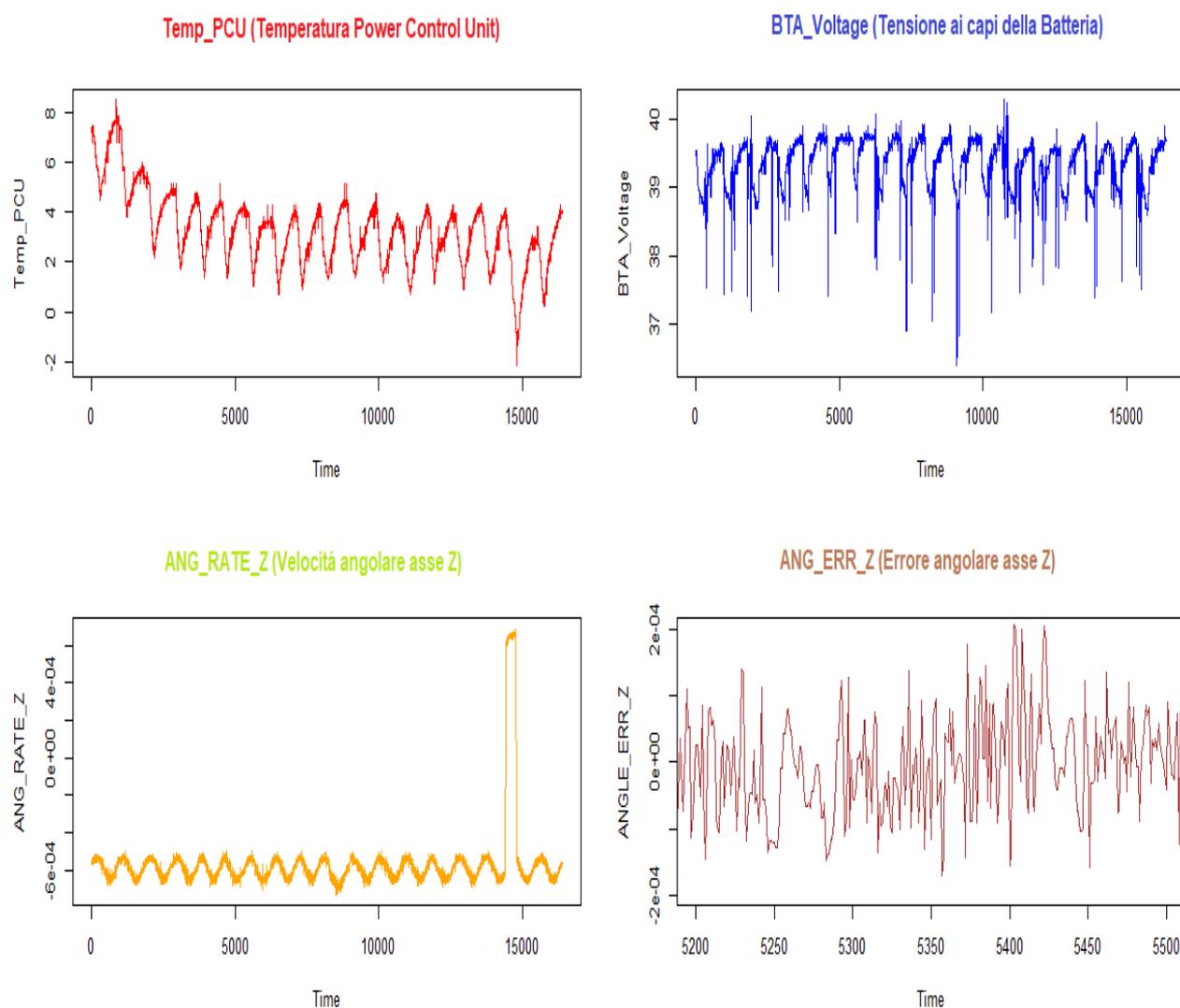


Figura 4.1: Serie temporali di $n = 16284$ campioni delle variabili: *Temperatura Power Control Unit* (Temp_PCU); *Tensione ai capi della Batteria* (BTA_Voltage); *Velocità angolare asse Z* (ANG_RATE_Z); *Errore angolare asse Z* (ANG_ERR_Z) in un range ristretto.

Nelle Fig. 4.2, 4.3, 4.4 e 4.5 sono riportati i grafici delle interpolazioni eseguite sulle quattro variabili scelte, secondo i metodi di analisi utilizzati. In tutte e quattro le comparazioni si evince come le curve stimate dall'interpolazione wavelet siano molto più significative, in quanto riproducono meglio il trend presente all'interno della sequenza dei

dati. Ciò é dovuto al fatto che i metodi di interpolazione polinomiale (lineare, spline cubica, Stineman) e quella basata sul filtering e smoothing di Kalman, cercano di "inseguire" la posizione dei dati, costruendo una semplice spezzata come nel caso dell'interpolazione lineare, o addirittura introducono sequenze di dati fittizi che non rispecchiano l'andamento medio dei dati reali.

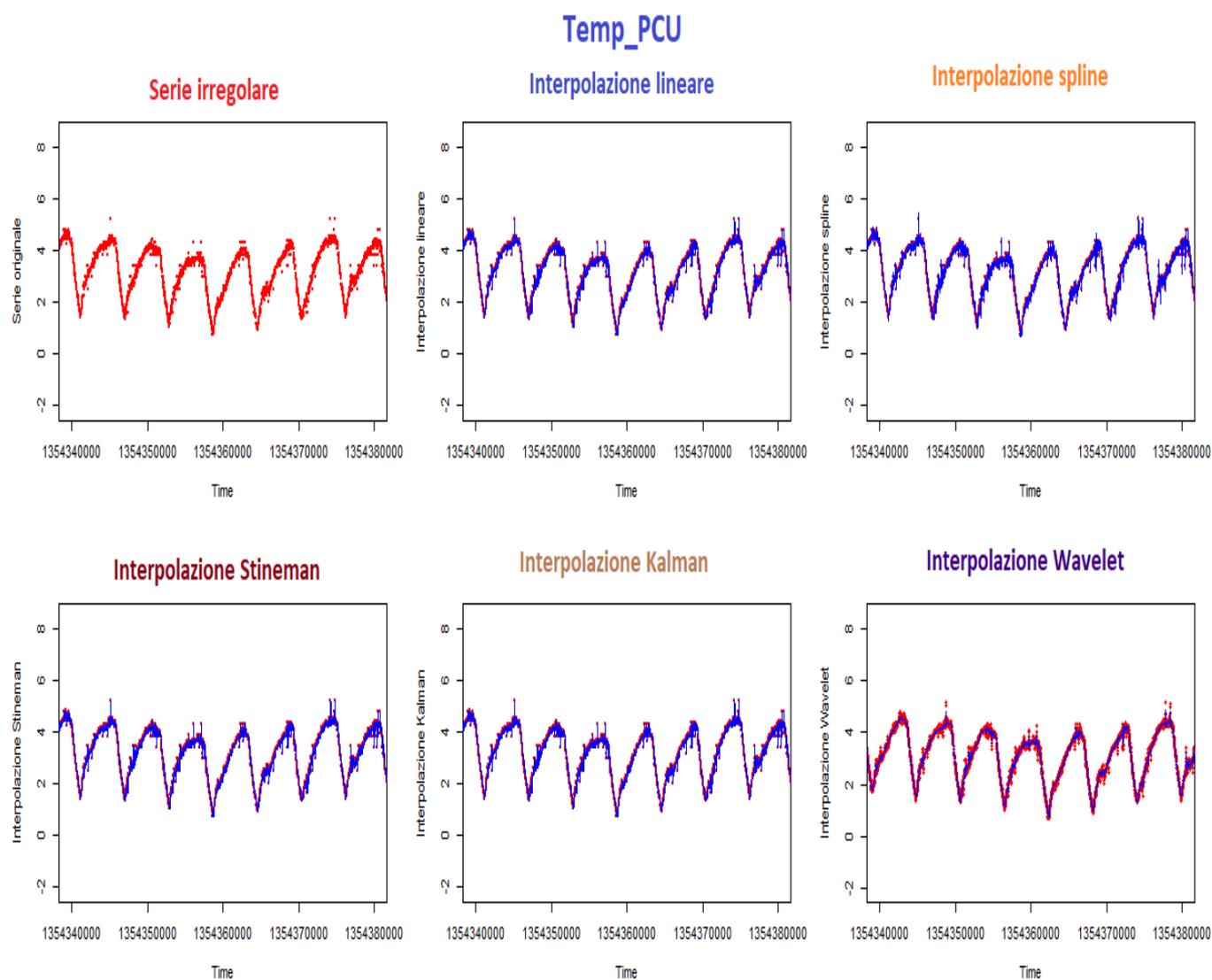


Figura 4.2: Confronto delle interpolazioni eseguite sulla serie temporale della variabile Temp_PCU: serie irregolare originaria; interpolazione lineare; interpolazione spline; interpolazione Stineman; interpolazione Kalman; interpolazione Wavelet.

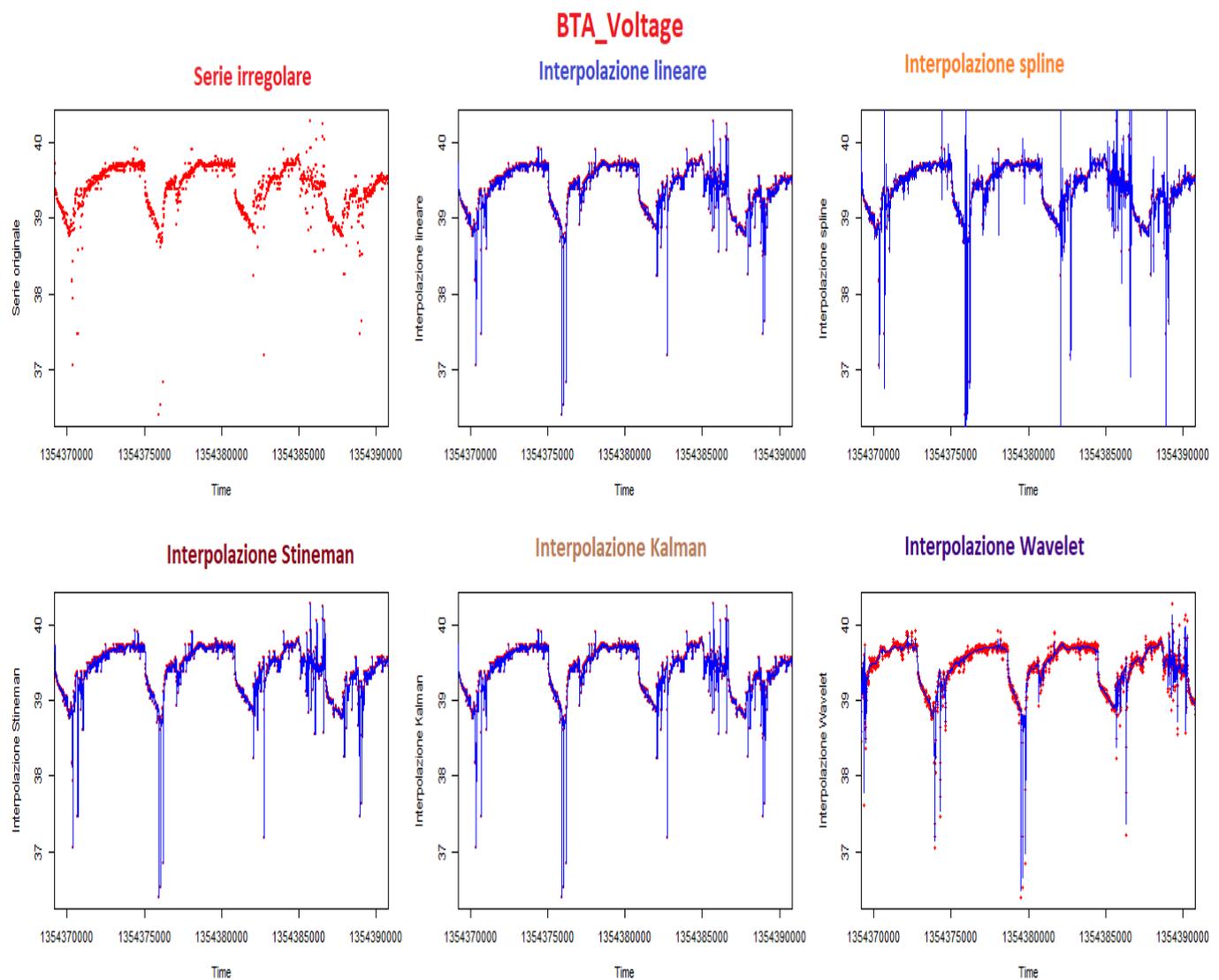


Figura 4.3: Confronto delle interpolazioni eseguite sulla serie temporale della variabile BTA_Voltage: serie irregolare originaria; interpolazione lineare; interpolazione spline; interpolazione Stineman; interpolazione Kalman; interpolazione Wavelet.

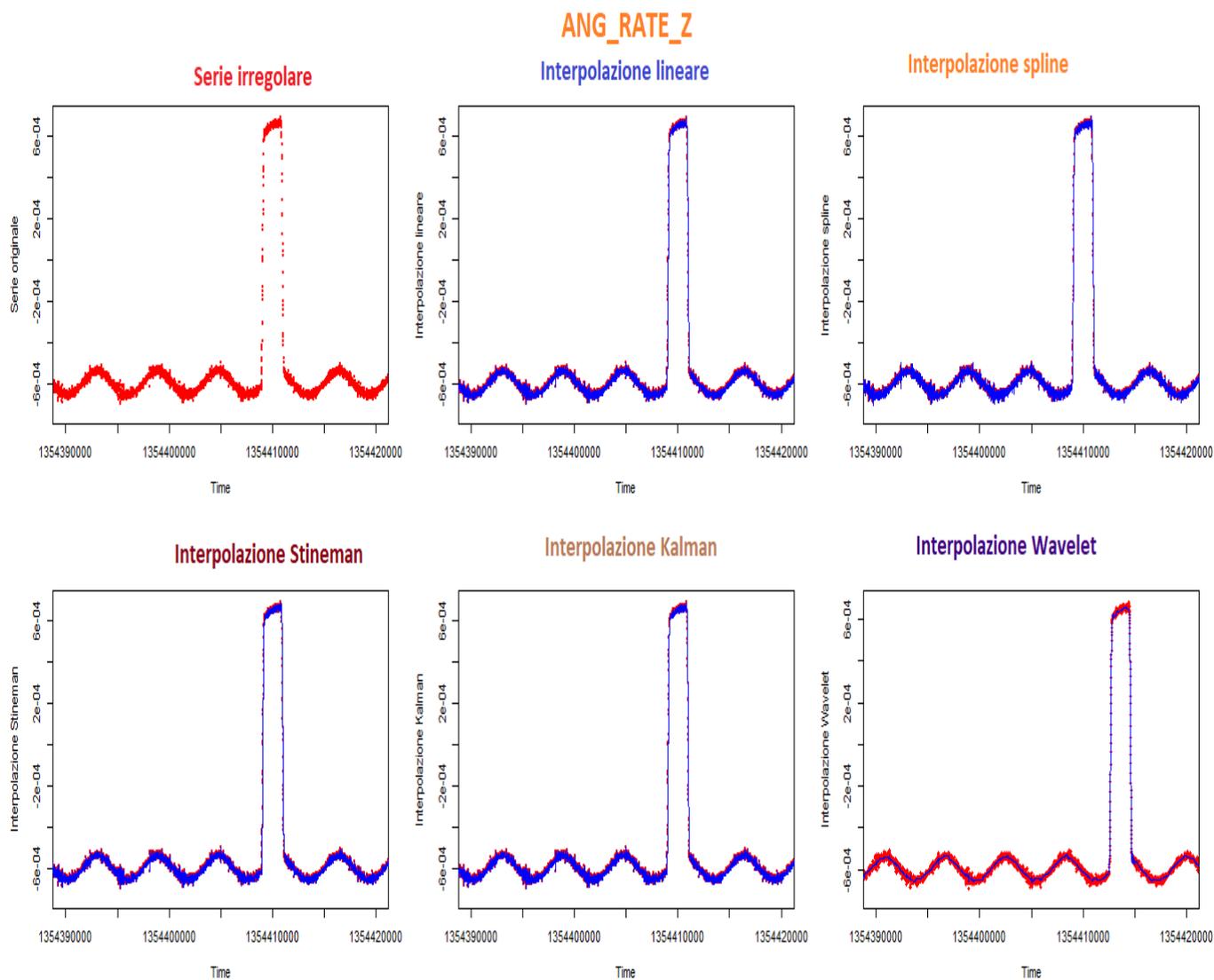


Figura 4.4: Confronto delle interpolazioni eseguite sulla serie temporale della variabile **ANG_RATE_Z**: serie irregolare originaria; interpolazione lineare; interpolazione spline; interpolazione Stineman; interpolazione Kalman; interpolazione Wavelet.

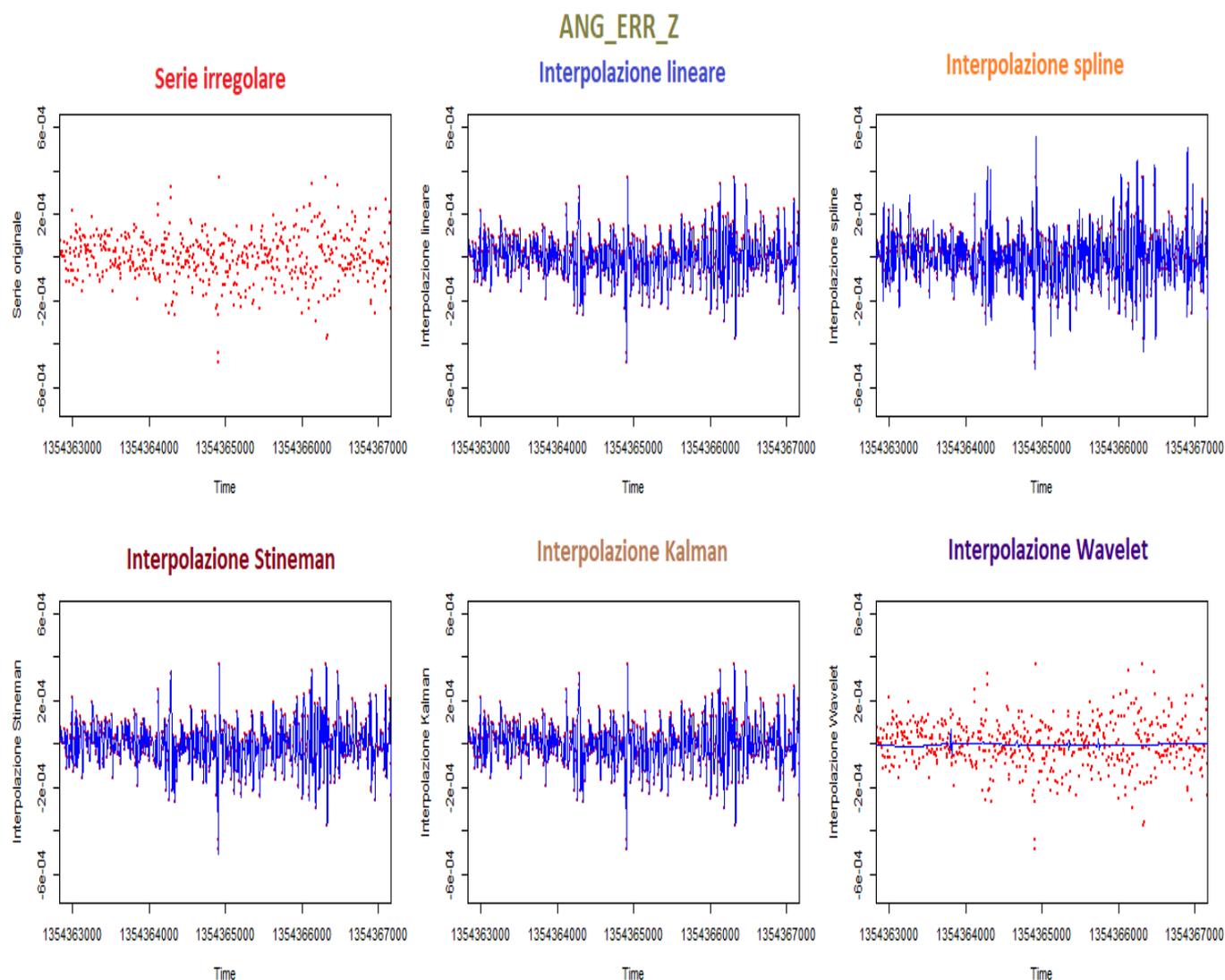


Figura 4.5: Confronto delle interpolazioni eseguite sulla serie temporale della variabile **ANG_ERR_Z**: serie irregolare originaria; interpolazione lineare; interpolazione spline; interpolazione Stineman; interpolazione Kalman; interpolazione Wavelet.

Questo aspetto é ancora piú evidente dal plot dell'interpolazione spline della variabile **BTA_Voltage**, dove le variazioni repentine non sono ben riprodotte dalla funzione polinomiale (cubica), che anzi introduce nuovi dati non reali. Ciò invece non avviene nel caso della trasformazione wavelet, che interpreta correttamente i dati anche in corrispondenza

dei punti che sono al di fuori del trend principale. Ciò é anche confermato dalle interpolazioni eseguite sulla variabile `ANG_RATE_Z`, dove la forte discontinuitá presente é ben fittata dalla curva stimata con la trasformata di wavelet, che conferma come la tecnica delle wavelet siamo in grado di rappresentare correttamente anche trend che presentano forti discontinuitá o irregolaritá nella serie dei dati. Nel caso delle interpolazioni eseguite sulla variabile `ANG_Z`, si nota come l'interpolazione costruita sulla trasformata wavelet produca una curva che rappresenta l'andamento medio dei dati (media vicino a zero), dovuto al fatto che la variabile analizzata non é una grandezza misurata direttamente come per esempio lo sono le variabili di temperatura, velocitá o tensioni elettriche, ma é una grandezza derivata (errore) costituita da una componente media piú rumore additivo; come vedremo piú avanti nell'analisi wavelet, il processo di thresholding elimina gran parte dei coefficienti (piccoli) che sono attribuiti al rumore. Come si puó facilmente osservare dai valori riportati nella tabella, il metodo della trasformata wavelet risulta essere un algoritmo molto efficiente in termini di tempi computazionali impiegati per svolgere le operazioni necessarie. Come giá discusso nella parte relativa alle analisi eseguite, la tecnica della trasformata wavelet per la ricostruzione di serie temporali irregolari, si basa su una trasformazione dei valori dei dati e dei relativi tempi di acquisizione, su una griglia regolare, su una successiva interpolazione, il processo del thresholding e, infine, la trasformata wavelet inversa per ottenere la curva interpolante dei dati. Ognuna di queste operazioni ha un costo in termini di efficienza computazionale, ma risulta tuttavia avere dei tempi inferiori o al massimo confrontabili in ordine di grandezze con quelli degli altri metodi, che invece eseguono una semplice operazione di costruzione di un polinomio interpolante passante per i dati (intepolazioni lineare, spline, Stineman), o la stima di una curva da analisi di tipo statistico. Inoltre, la tecnica delle wavelet offre una maggiore precisione nella rappresentazione del trend di una sequenza di dati. Dal punto di vista computazionale, il metodo Kalman risulta essere il piú lento, a causa dei processi di

filtraggio e regolarizzazione.

Tempi computazionali (sec)	Temp_PCU	BTA_Voltage	ANG_RATE_Z	ANG_ERR_Z
Interpolazione lineare	0,36	0,14	0,15	0,10
Interpolazione spline	0,33	0,14	0,13	0,12
Interpolazione Stineman	1,46	1,60	1,84	1,65
Interpolazione Kalman	5,07	4,14	4,16	4,47
Interpolazione Wavelet	0,16	0,14	0,16	0,16

Figura 4.6: Tabella dei tempi computazionali dei metodi di interpolazione utilizzati per l'analisi delle serie temporali associate alle quattro variabili del data-set.

Di seguito riportiamo nella Fig.4.7 la tabella con gli indici statistici (medie e varianze) delle serie delle variabili considerate, ricostruite tramite i differenti metodi di interpolazione. Osservando i valori ottenuti si nota in tutte e quattro le variabili come i vari metodi di interpolazione non varino significativamente la media della serie originaria. Ciò implica che i vari metodi di interpolazione non introducono ulteriori componenti che possano influenzare il comportamento medio o trend dei dati. Tuttavia, si può notare come la tecnica di interpolazione che sfrutta la trasformata wavelet determini una serie ricostruita con varianza minore rispetto a quella della serie originaria così come anche rispetto alla varianza delle serie ottenute dalle altre tecniche di interpolazione. Ciò é dovuto al fatto che il processo di thresholding introduce uno “smoothing” della serie originaria eliminando

così la componente rumorosa e quindi quei dati che si allontanano significativamente dal trend dei dati.

Temp_PCU	Serie Irregolare	Lineare	Spline	Stineman	Kalman	Wavelet
Media	3,306	3,324	3,324	3,324	3,324	3,324
Varianza	2,173	2,083	2,086	2,083	2,084	2,026

BTA_Voltage	Serie Irregolare	Lineare	Spline	Stineman	Kalman	Wavelet
Media	39,344	39,375	39,376	39,375	39,375	39,375
Varianza	0,133	0,111	0,126	0,112	0,108	0,097

ANG_RATE_Z	Serie Irregolare	Lineare	Spline	Stineman	Kalman	Wavelet
Media	-5,69E-04	-5,77E-04	-5,77E-04	-5,77E-04	-5,77E-04	-5,77E-04
Varianza	3,50E-08	2,87E-08	2,88E-08	2,87E-08	2,87E-08	2,80E-08

ANGLE_ERR_Z	Serie Irregolare	Lineare	Spline	Stineman	Kalman	Wavelet
Media	3,76E-07	3,73E-07	-5,17E-07	3,24E-07	3,73E-07	4,45E-07
Varianza	1,06E-08	6,88E-09	1,31E-08	7,88E-09	6,88E-09	5,20E-11

Figura 4.7: Tabelle degli indici statistici (media e varianza) delle serie associate alle variabili Temp_PCU, BTA_Voltage, ANG_RATE_Z, ANG_ERR_Z, e delle relative serie ottenute dalle interpolazioni eseguite.

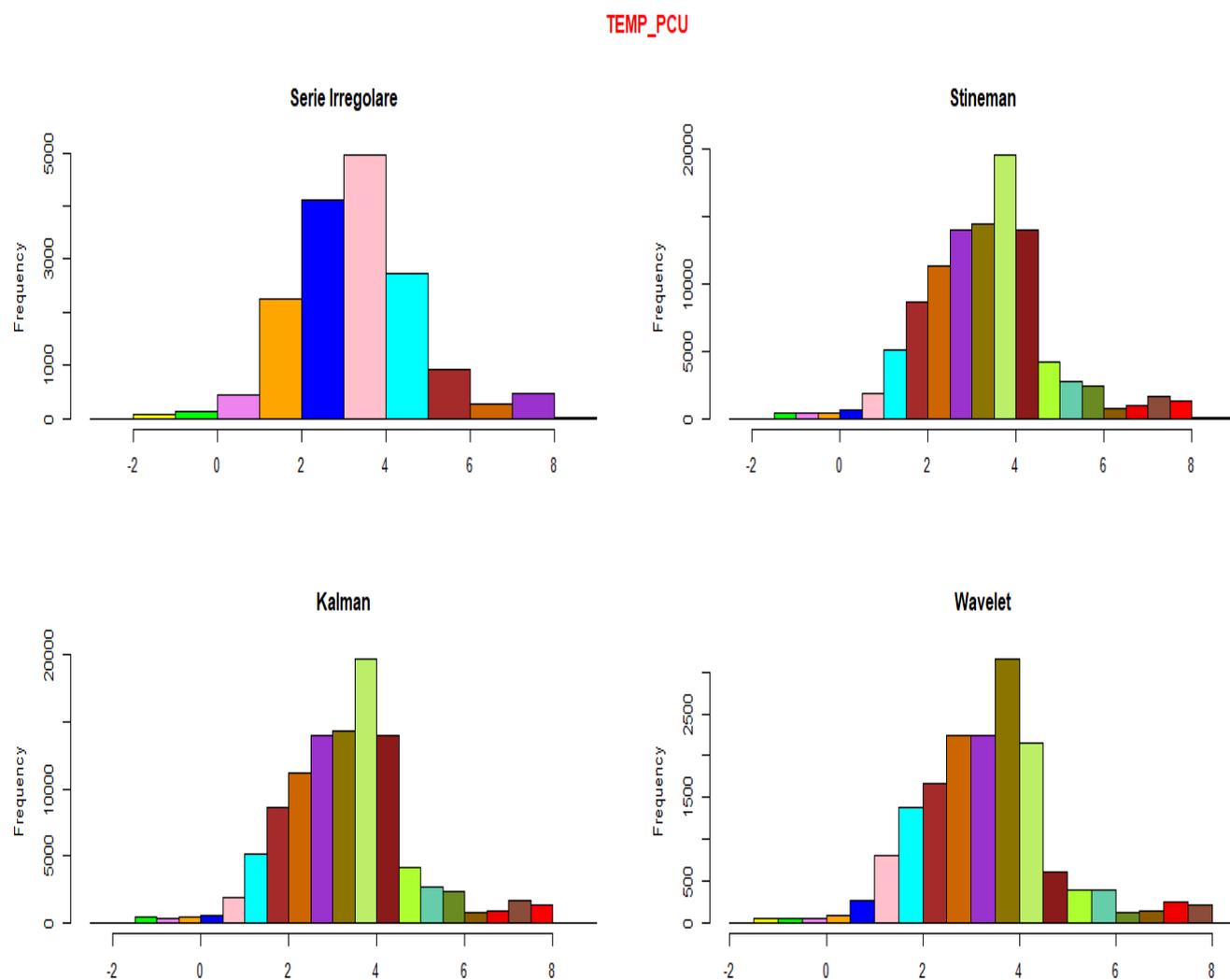


Figura 4.8: Istogrammi a confronto dei dati delle serie originaria, interpolata Stineman, interpolata Kalman, interpolata Wavelet per la variabile Temp_PCU.

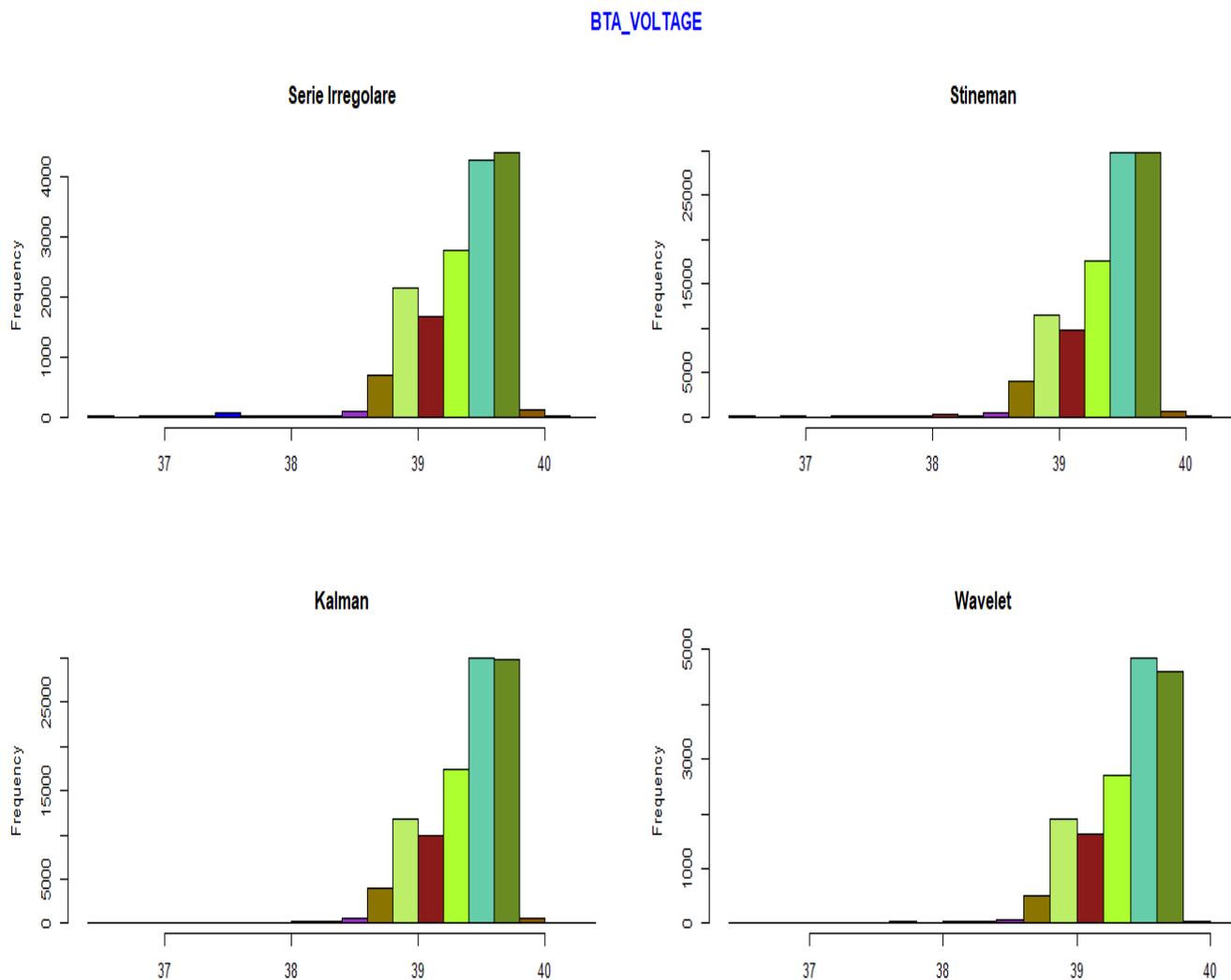


Figura 4.9: Istogrammi a confronto dei dati delle serie originaria, interpolata Stineman, interpolata Kalman, interpolata Wavelet per la variabile BTA_Voltage.

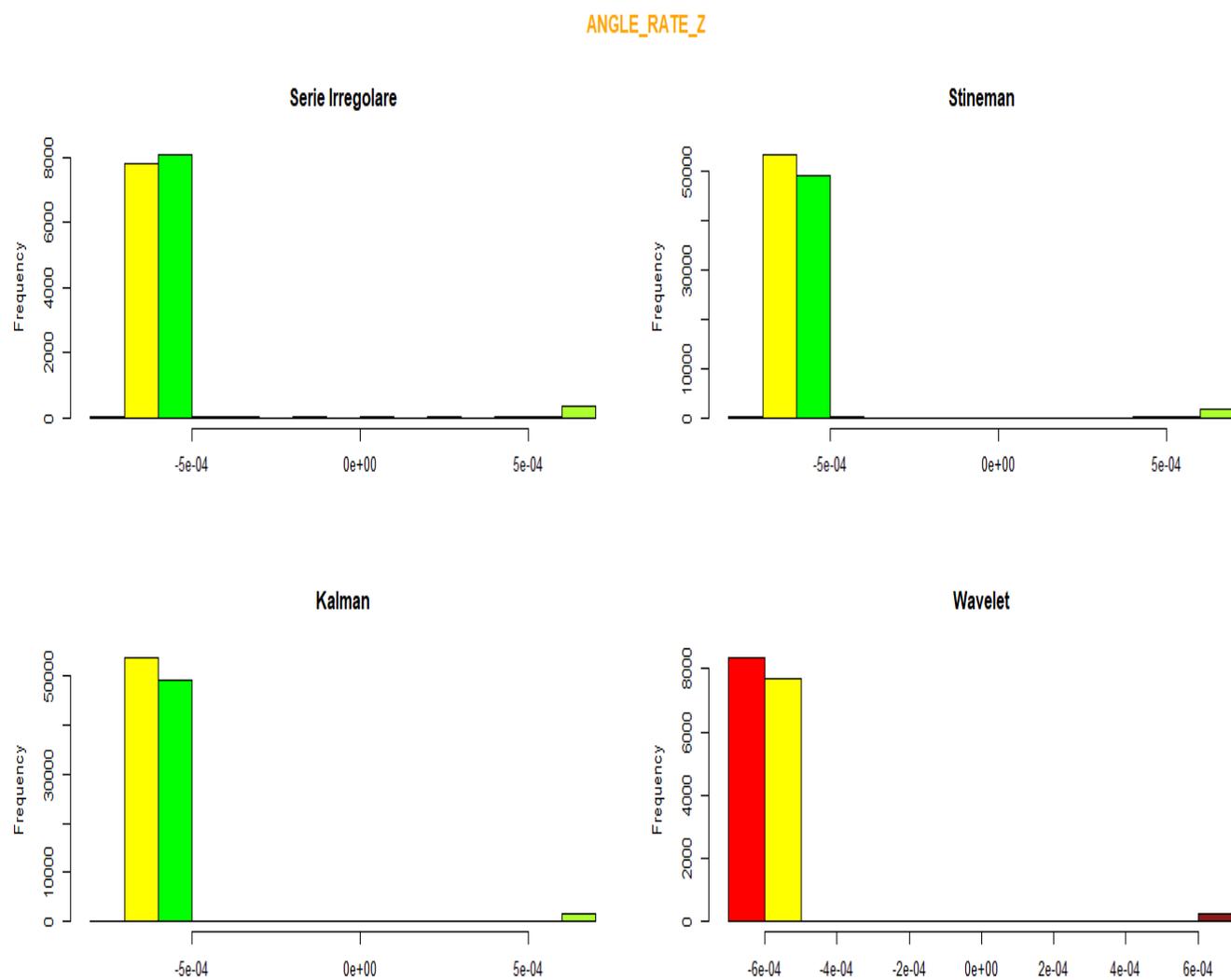


Figura 4.10: Istogrammi a confronto dei dati delle serie originaria, interpolata Stineman, interpolata Kalman, interpolata Wavelet per la variabile ANGLE_RATE_Z.

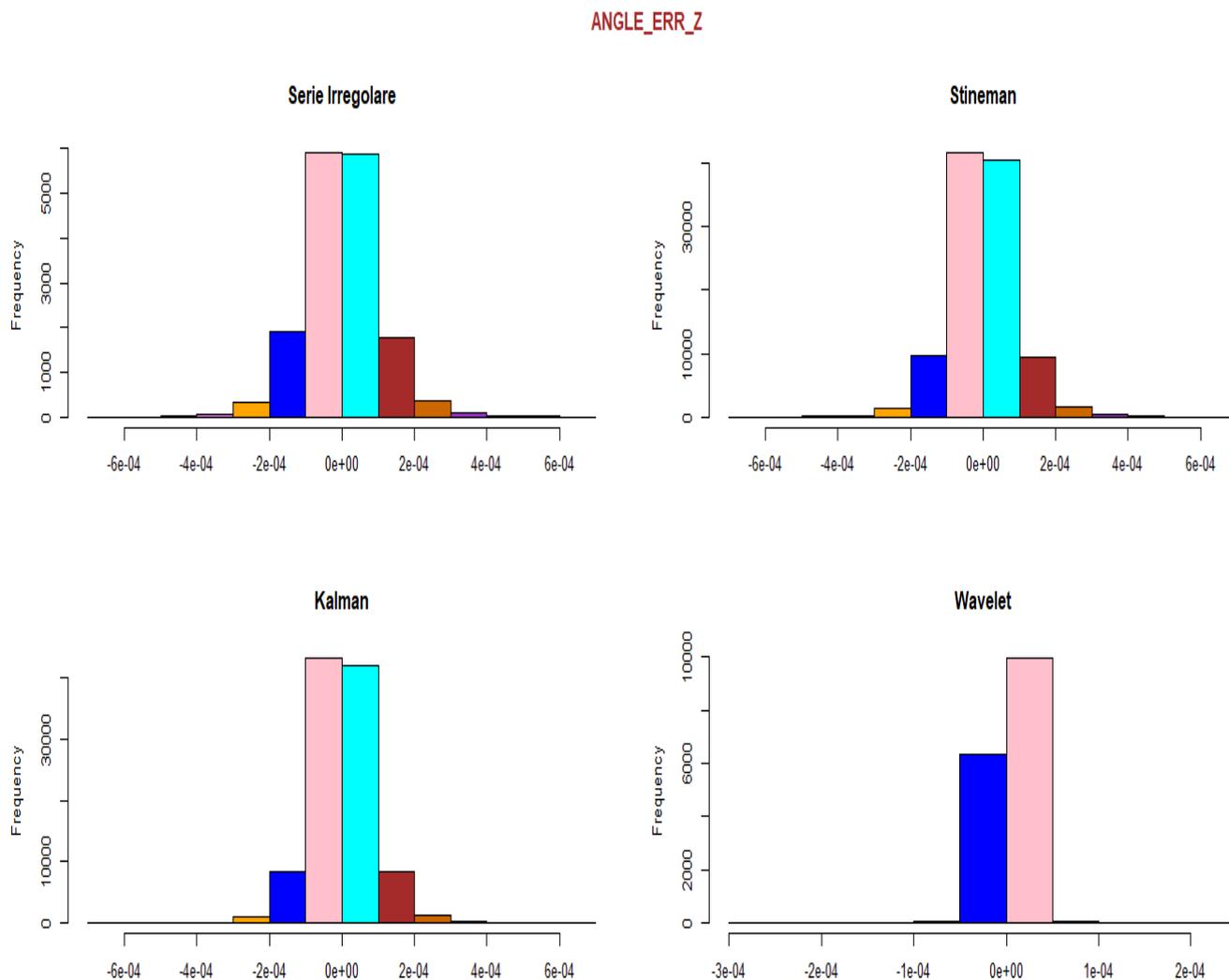


Figura 4.11: Istogrammi a confronto dei dati delle serie originaria, interpolata Stineman, interpolata Kalman, interpolata Wavelet per la variabile ANGLE_ERR_Z.

Quindi, il confronto degli indici statistici ci ha permesso di osservare che valori ottenuti per le medie calcolate sulla serie irregolare e le serie interpolate con i differenti metodi, mostrano che le interpolazioni non modificano il comportamento medio dei dati, per cui si conservano la struttura e caratteristica del dataset. I valori delle varianze, invece, per la serie irregolare e le serie interpolate con i differenti metodi, mostrano come l'interpolazione

con la trasformata Wavelet riduca la varianza della serie irregolare originaria; ciò è dovuto all'effetto di "smoothing" introdotto dal processo di *thresholding* dei dati. Tuttavia, il confronto tra gli istogrammi ottenuti dalla serie originaria e dalle serie interpolate non ci ha permesso di poter estrarre informazioni significative sulla ridistribuzione dei campioni ottenuti dai differenti metodi di interpolazione. Sarà quindi necessario considerare altri criteri di confronto statistico, per poter dedurre ulteriori informazioni riguardo i campioni ottenuti dalle serie interpolate.

Nelle Fig.4.12, 4.13, 4.15 e 4.14, sono riportati i plot dei coefficienti wavelet (d) prima e dopo l'applicazione del processo di thresholding alle serie. Come si può facilmente osservare, il thresholding elimina tutti quei coefficienti molto piccoli associati al rumore che vengono quindi filtrati, lasciando invariati solo quei coefficienti più grandi (linee verticali più lunghe), che contengono quindi le informazioni principali e più significative della sequenza di dati. Inoltre, si può notare come i coefficienti più piccoli, e quindi, eliminati sono quelli presenti alle scale di più alta risoluzione (in basso) ed eliminando quei coefficienti piccoli a confronto di quelli più grandi allo stesso livello di scala. La trasformata wavelet e il connesso meccanismo di thresholding, permettono quindi di estrarre le componenti più importanti contenute in una sequenza di dati, come per esempio il trend. Ciò è ancora più evidente nel caso dei coefficienti wavelet della variabile `ANG_ERR_Z`, dove, come già anticipato prima, è presente una componente di rumore dominante, per cui la trasformata wavelet e il thresholding restituiscono una serie a media quasi nulla, come è evidente dal fatto che molti dei coefficienti dei livelli di risoluzione maggiore sono eliminati tramite il processo thresholding, il quale produce quindi un filtraggio e smoothing dei dati.

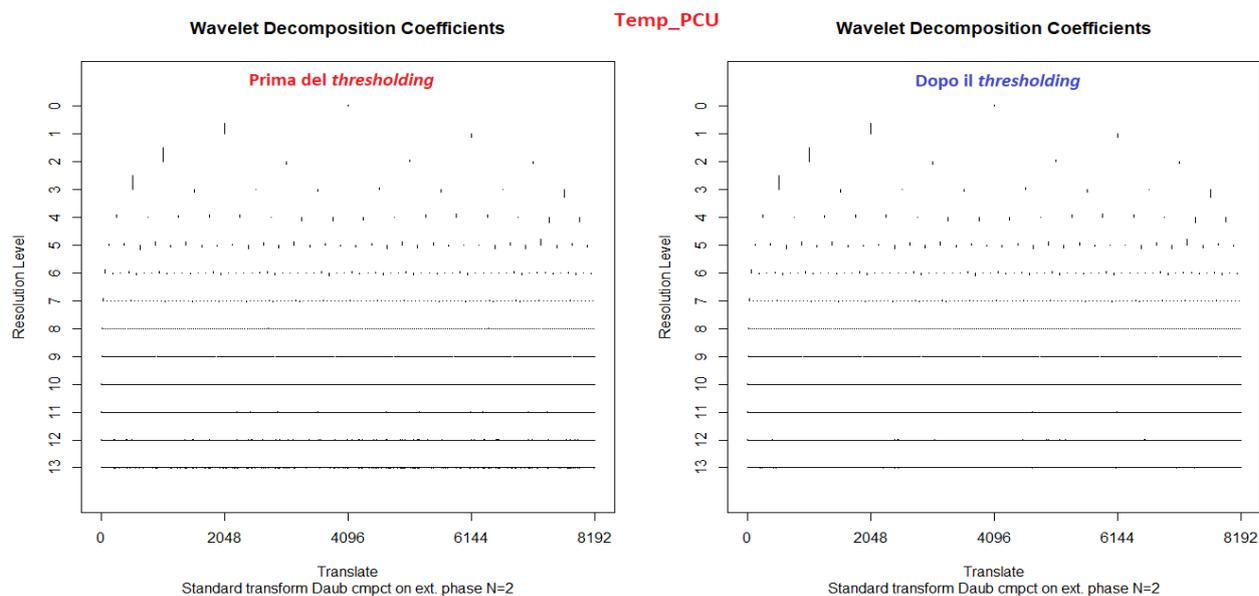


Figura 4.12: Plot dei coefficienti wavelet prima e dopo il processo di *thresholding* per la variabile Temp_PCU.

Nel caso del plot dei coefficienti wavelet della variabile ANG_RATE_Z si nota come i coefficienti dei livelli di piú alta risoluzione non compaiano nella finestra. Ció é dovuto al fatto che i coefficienti dei livelli 11, 12 e 13, sono in valore molto piccoli, inferiore al valore di precisione macchina, per cui non vengono visualizzati. In tal caso, il thresholding ha il solo effetto di rendere la serie piú regolare eliminando i piccoli coefficienti presenti alle scale di risoluzione maggiori. Altro aspetto significativo é che i plot dei coefficienti wavelet permettono di evidenziare le discontinuitá o forti irregolaritá presenti nelle sequenze di dati, come é visibile dal fatto che si hanno coefficienti significativamente grandi concentrati in corrispondenza della posizione del salto. Ció é evidente anche nel caso della variabile BTA_Voltage, dove i coefficienti in valore piú grande si concentrano in corrispondenza delle regioni di discontinuitá.

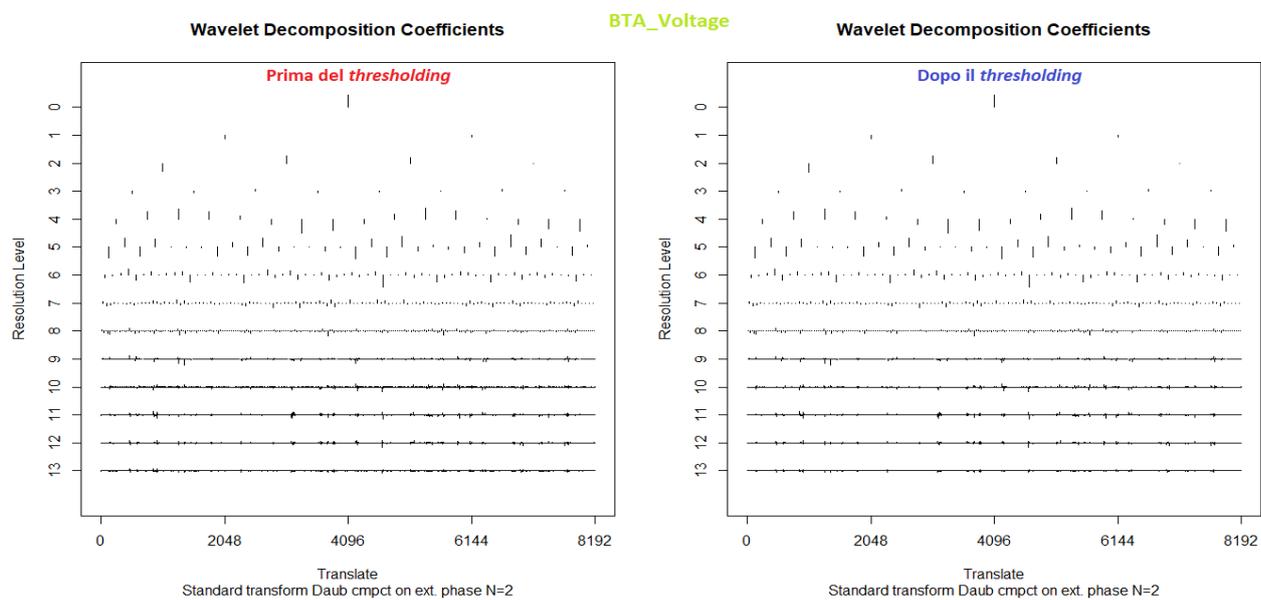


Figura 4.13: Plot dei coefficienti wavelet prima e dopo il processo di *thresholding* per la variabile BTA_Voltage.

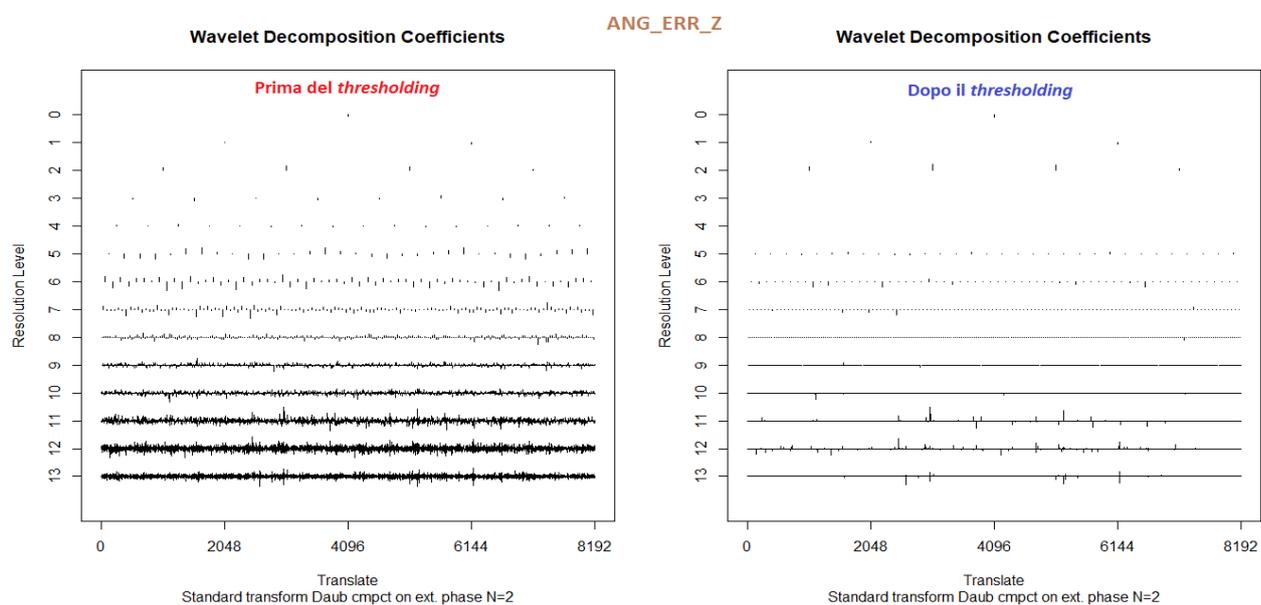


Figura 4.14: Plot dei coefficienti wavelet prima e dopo il processo di *thresholding* per la variabile ANG_ERR_Z.

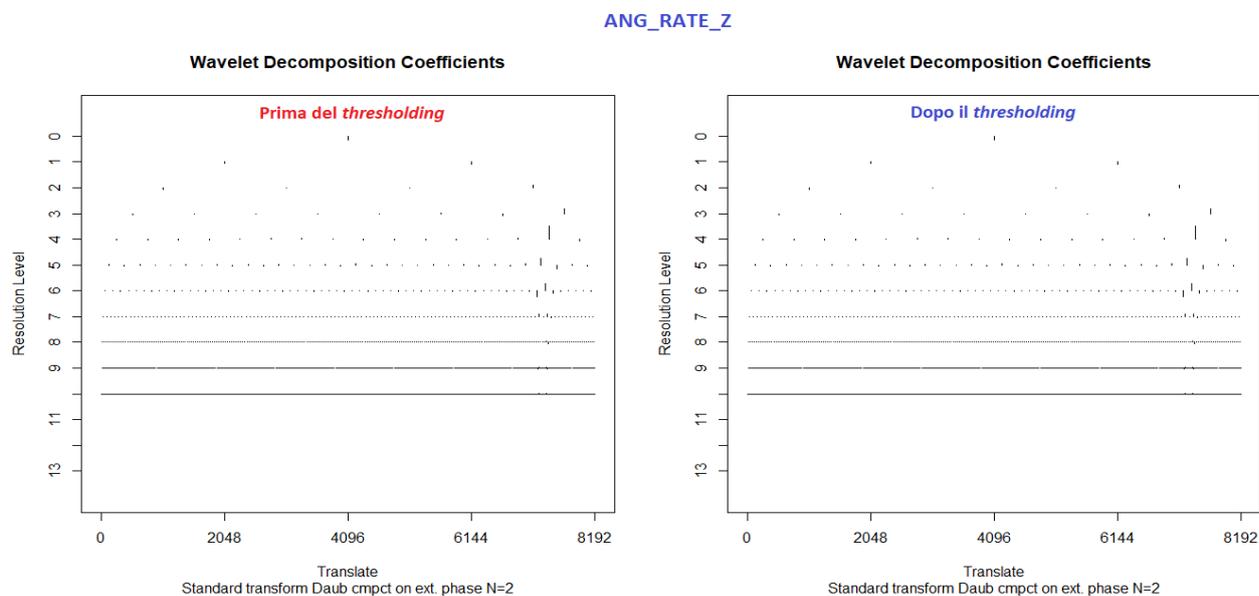


Figura 4.15: Plot dei coefficienti wavelet prima e dopo il processo di *thresholding* per la variabile ANG_RATE_Z .

Il metodo della trasformata wavelet, come visto dalle analisi precedenti, consente di stimare la curva che meglio si adatta ai dati, tramite il ‘denoising’ della serie, ottenuto dalla tecnica del *thresholding*. Nonostante gran parte dei coefficienti (piccoli in valore) siano eliminati, le informazioni principali del set di dati é preservato in pochi coefficienti di valore piú grande, fatto questo evidenziato dalla *sparsitá* stessa dei coefficienti. La curva stimata che meglio fitta i dati, sará quindi ricostruita a partire dai coefficienti ‘‘sopravvissuti’’ al processo di *thresholding*. Ció é confermato anche dalla conservazione dell’*energia* (somma dei quadrati) dei valori della serie originale e dei valori dei coefficienti ottenuti alle differenti scale. Osservando la tabella nella Fig.4.16 l’energia dei coefficienti al massimo livello di risoluzione (livello 13) é esattamente pari all’energia della serie originaria. Inoltre, come atteso il *thresholding* modifica il valore dell’energia di una percentuale irrisoria, a conferma che la serie ricostruita, dalla trasformata wavelet inversa, sui coefficienti piú grandi, é in grado di riprodurre correttamente le caratteristiche essenziali e peculiari dei dati analizzati

SERIE (N=16384 campioni)	Temp_PCU	BTA_Voltage	ANG_RATE_X	ANG_ERR_Z
Numero coeff. wavelet eliminati (thresholding)	8123	7851	8132	8175
Percentuale coeff. eliminati dal thresholding (%)	99,2	95,8	99,3	99,8
Energia serie	215131	25403464	0,0020714	0,0001124
Energia coeff. wavelet (livello 13)	215131	25403464	0,0020714	0,0001124
Energia coeff. wavelet + thresholding (livello 13)	215115	25403448	0,0020713	8,5E-05
Percentuale variazione energia (%)	0,0072	6,0E-05	0,0070	24,5

Figura 4.16: Tabella con il numero di coefficienti wavelet eliminati dal processo di thresholding, la percentuale di coefficienti eliminati, l'energia (somma dei quadrati) delle serie originarie e dei coefficienti prima e dopo il thresholding, e la corrispondente variazione percentuale.

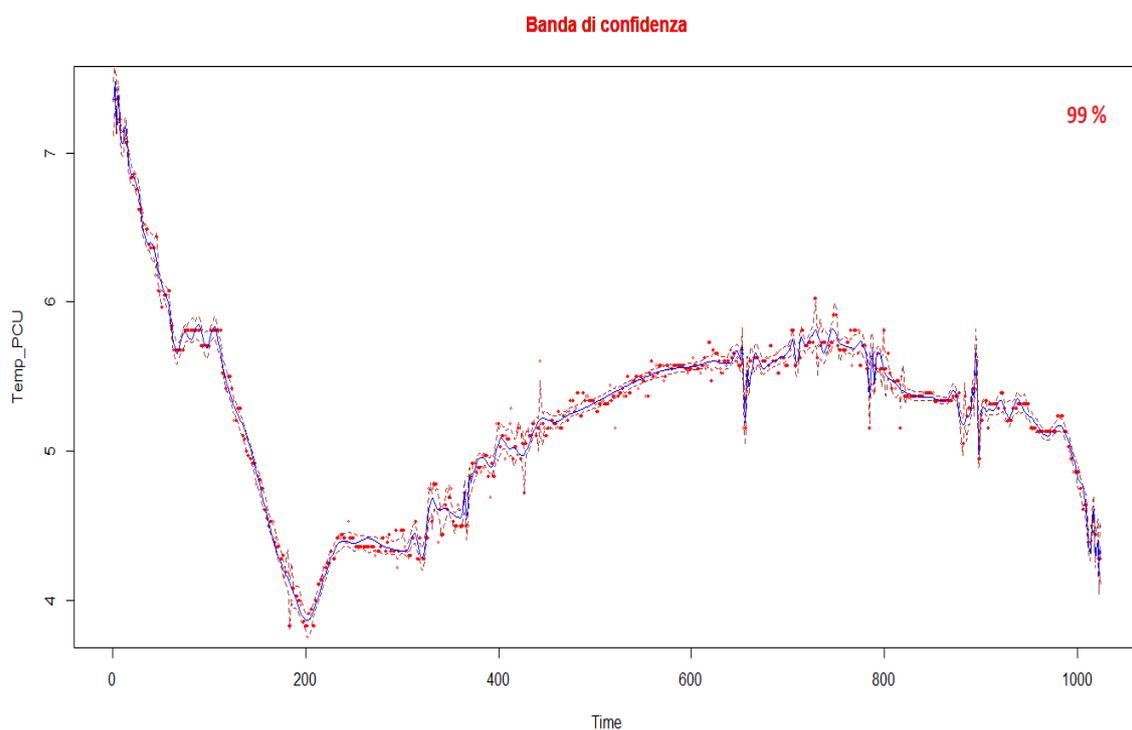


Figura 4.17: Banda di confidenza per l'interpolazione wavelet della variabile Temp_PCU: dati serie originaria (cerchi rossi), curva interpolante ottenuta da trasformata wavelet (linea blu continua), banda di confidenza al 99% (linee marroni tratteggiate).

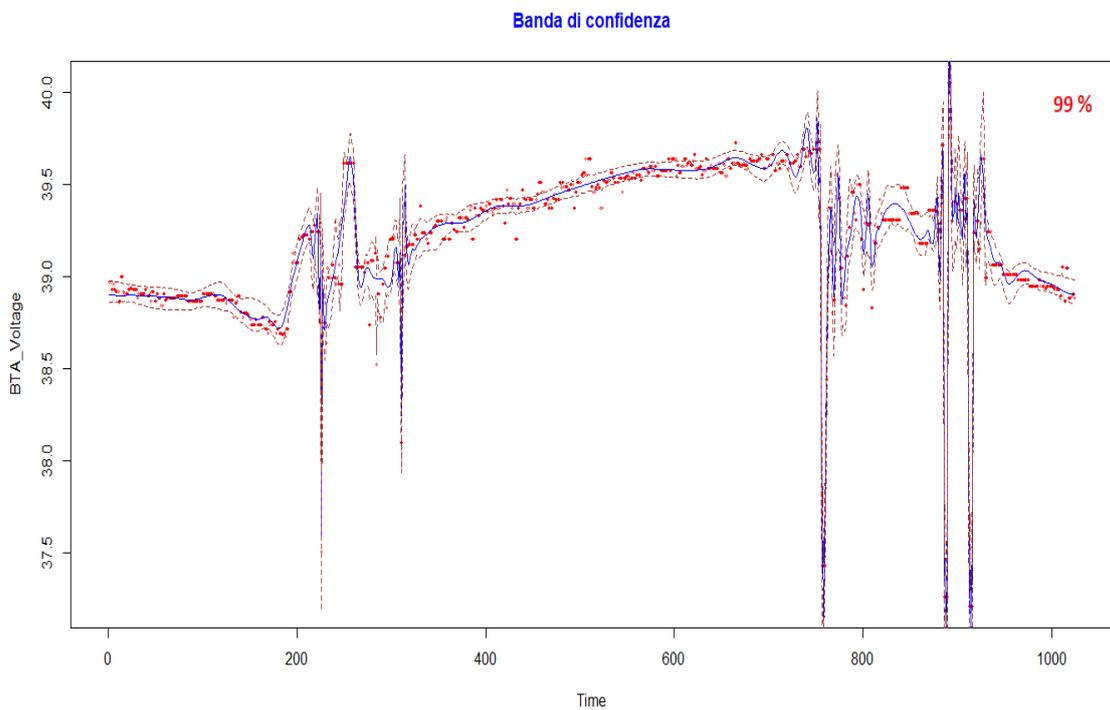


Figura 4.18: Banda di confidenza per l'interpolazione wavelet della variabile BTA_Voltage: dati serie originaria (cerchi rossi), curva interpolante ottenuta da trasformata wavelet (linea blu continua), banda di confidenza al 99% (linee marroni tratteggiate).

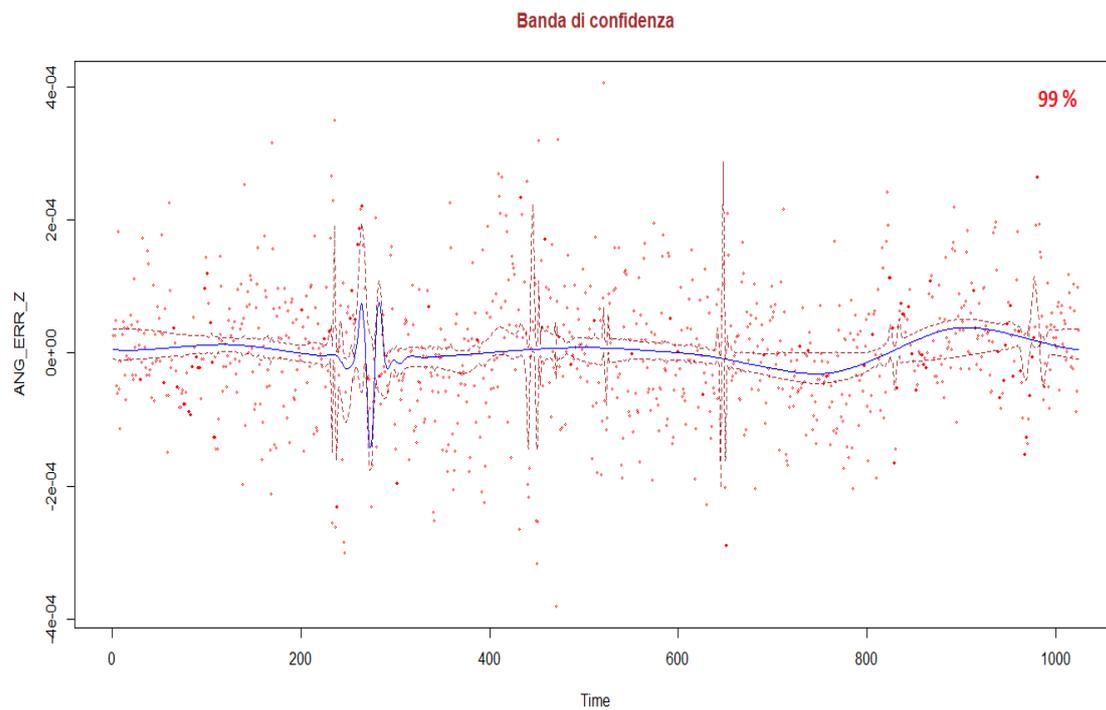


Figura 4.19: Banda di confidenza per l'interpolazione wavelet della variabile ANG_ERR_Z: dati serie originaria (cerchi rossi), curva interpolante ottenuta da trasformata wavelet (linea blu continua), banda di confidenza al 99% (linee marroni tratteggiate).

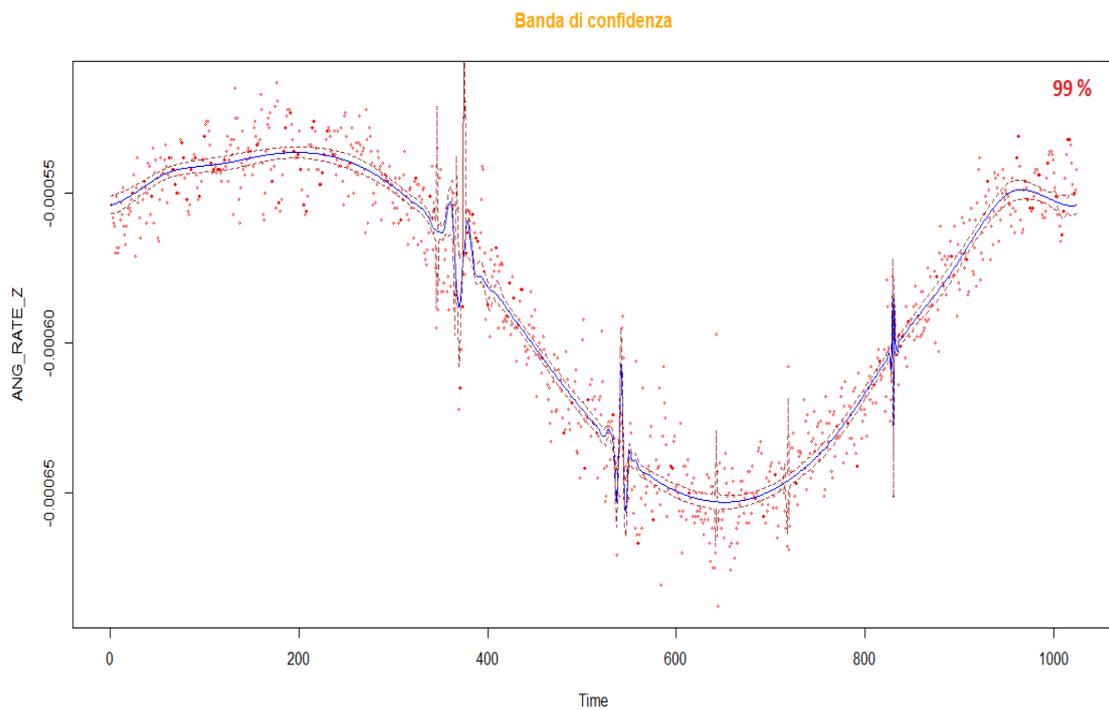


Figura 4.20: Banda di confidenza per l'interpolazione wavelet della variabile ANG_RATE_Z: dati serie originaria (cerchi rossi), curva interpolante ottenuta da trasformata wavelet (linea blu continua), banda di confidenza al 99% (linee marroni tratteggiate).

Conclusioni

In questo lavoro abbiamo affrontato i seguenti temi:

- Studio e caratterizzazione di serie di dati di telemetria da satellite: serie irregolari, presenza di stagionalità e discontinuità
- Panoramica di alcuni metodi matematici di interpolazione per serie irregolari: lineare, spline, stineman, kalman, wavelet
- Ricostruzione serie con passo regolare, confronto fra i risultati di interpolazione tramite indici statistici, efficienza computazionale
- Analisi tramite wavelet:
 - Stima curva di regressione non parametrica, taglio coefficienti (*thresholding*)
 - Verifica conservazione energia
 - Capacità di caratterizzazione locale delle serie
 - Bande di confidenza
- Wavelet, risultati ottenuti:
 - Consentono di rappresentare i caratteri delle serie esaminate
 - Consentono di filtrare la componente di rumore
 - Metodo candidato per trattare le serie per le fasi successive di elaborazione

- Sviluppi futuri:
 - Sperimentazione di altre tecniche wavelet
 - Sperimentazione di tecniche di filtraggio alternative, eliminazione stagionalità
 - Analisi di correlazione tra le serie

Bibliografia

- [1] ECSS-E-ST-70-31C (31 Luglio 2008) - *Space Engineering Ground Systems and Operations Monitoring and Control Data Definition*.
- [2] ECSS-E-ST-70C (31 Luglio 2008) - *Space Engineering Ground systems and operations*.
- [3] Wertz, James R., Larson, *Space Mission Analysis and Design*-Springer,Wiley.
- [4] C. Ciancarelli, A. Intelisano, S.G. Neglia, *Knowledge Retrieval Strategy for Satellites System Monitoring based on Data Analytics Techniques*, in Proc. of the 2017 conference on Big Data from Space (BiBS'17), Toulouse (France), pp.394-397.
- [5] G.P. Nason, *Wavelet Methods in Statistics with R*-Springer.
- [6] Shumway, Stoffer, *Time Series Analysis and Its Applications - With R Examples*-Springer-Verlag.
- [7] Donald B. Percival and Andrew T. Walden, *Wavelet Methods for Time Series Analysis*-Cambridge University Press, 2000.