

Università degli Studi di Firenze



FACOLTÀ DI INGEGNERIA

*CORSO DI LAUREA IN INGEGNERIA PER LA TUTELA
DELL'AMBIENTE E DEL TERRITORIO*

Anno accademico 2005/2006

Corso di Modellistica dei Sistemi Ambientali

Analisi Serie Temporale Ossigeno Disciolto Laguna Orbetello

Confronto Wavelet-Spline

Prof. Marsili- Libelli Stefano

Studenti : Burberi Lorenzo

Ciappi Giulia

Indice

1. Introduzione.....	3
2. Analisi serie temporale di partenza.....	5
3. Detrending della serie temporale.....	9
4. Denoising con splines e controllo con analisi in frequenza.....	12
5. Analisi dei residui.....	15
6. Wavelets e controllo con analisi spettrale	19
7. Confronto Splines - Wavelet e conclusioni.....	29
Appendice A – Storia delle Wavelet.....	35
Appendice B – Trasformata di Fourier, STFT, funzioni Wavelts.....	37
Appendice C – Principali Funzioni Matlab impiegate.....	39
Bibliografia.....	42

1. Introduzione

E' stata condotta un'analisi di una serie temporale relativa alla concentrazione di ossigeno disciolto nella Laguna di Orbetello, lato Levante, nel periodo Luglio-Settembre 2001.

I dati analizzati sono complessivamente 1800, campionati con tempo T_s pari a 1 h.

Nello specifico i dati relativi ai mesi in esame sono:

- 336 dati nel mese di Luglio;
- 744 dati nel mese di Agosto;
- 720 dati nel mese di Settembre.

L'obiettivo iniziale era quello di analizzare l'intera stagione estiva 2001 (Giugno-Settembre), ma l'analisi in seguito è stata ristretta ai mesi sopra citati per l'assenza di dati nel mese di Giugno e nei primi quindici giorni di Luglio.

Ai 1800 dati di partenza ne sono stati sottratti 37, relativi agli ultimi giorni di Settembre, in quanto il loro andamento si discostava eccessivamente dai precedenti valori di concentrazione, denotando possibili errori di rilevazione.

I passaggi che sono stati seguiti nel trattare la serie di partenza sono i seguenti:

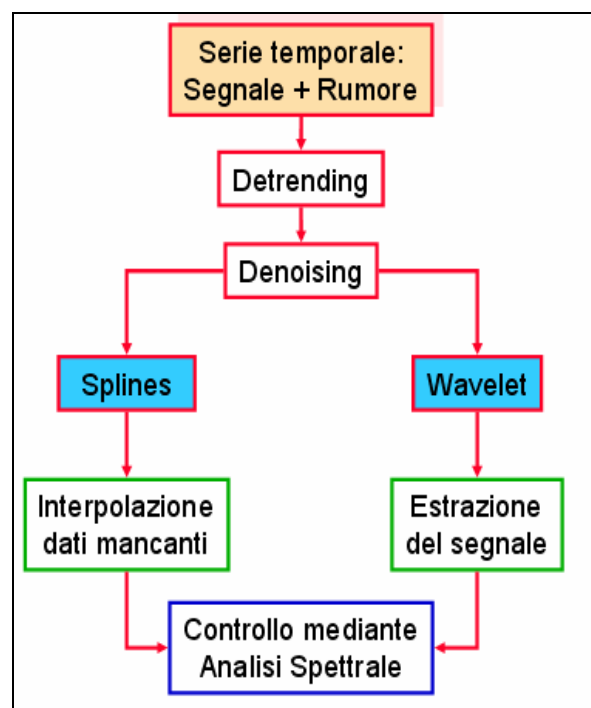


Figura 1.1 – Fasi analisi serie temporale.

I dati iniziali contenevano sia segnale che rumore; l'obiettivo è stato quello di studiare il segnale eliminando il più possibile il rumore.

Tale procedura è stata eseguita con due diverse tecniche:

- **funzioni Splines**, che sono utili per lo smoothing e l'interpolazione dei dati mancanti;
- **funzioni Wavelet**, che permettono di estrarre l'informazione deterministica con ottime proprietà di denoising, consentendo anche la derivazione dei segnali filtrati.

In entrambi i casi l'analisi in frequenza è stata un utile supporto per valutare la bontà dello studio condotto.

2. Analisi serie temporale di partenza

La prima operazione eseguita è stata l'analisi della serie di partenza, con l'individuazione di eventuali ciclicità e andamenti anomali:

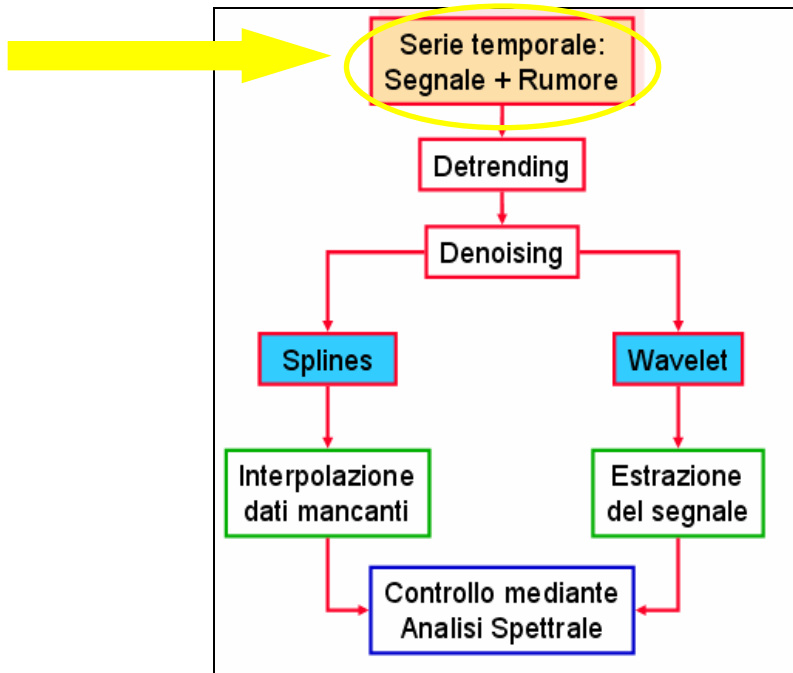


Figura 2.1 – Analisi serie temporale.

L'andamento delle concentrazioni di DO in funzione del tempo di campionamento T_s è il seguente:

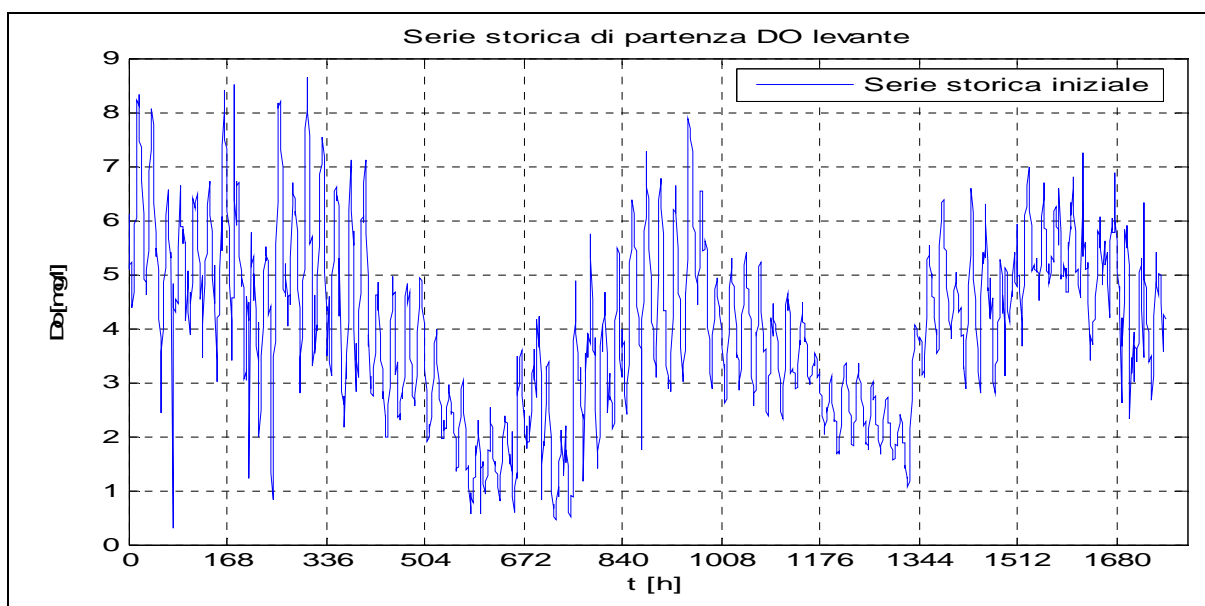


Grafico 2.1 – Andamento Do in funzione del tempo di campionamento.

Come si evidenzia dal grafico, i valori di DO sono variabili entro un range di 0.5 mg/l - 8.5 mg/l. Si nota inoltre la presenza di cicli di circa 6 settimane, relativi alle ultime due di Luglio - intero mese di Agosto e all'intero mese di Settembre, come messo in evidenza nel grafico seguente:

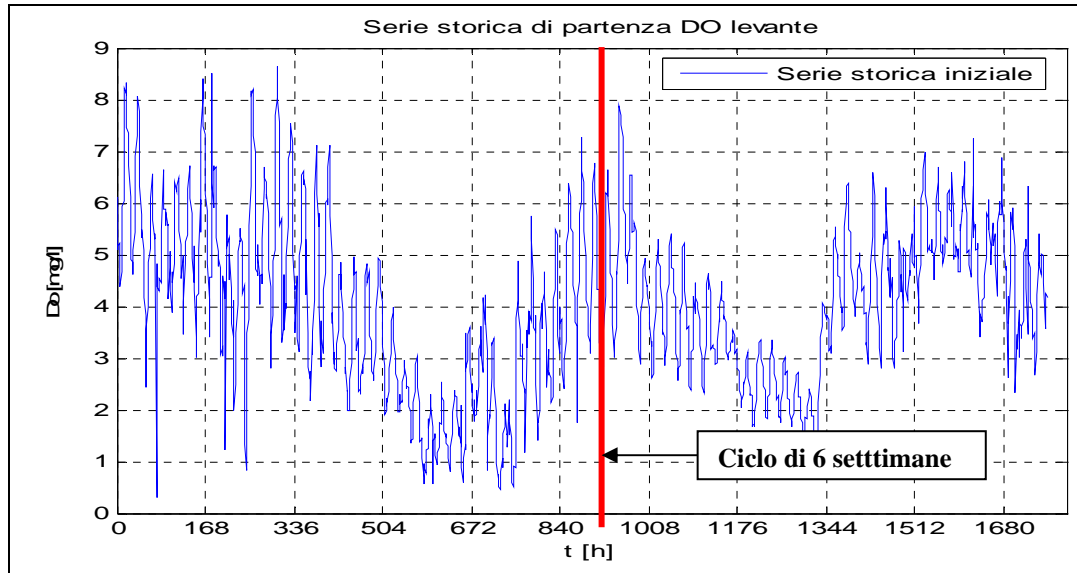


Grafico 2.2 – Cicli presenti nella serie temporale.

Si notano, inoltre:

- crisi anossiche in Luglio nelle prime ore della mattina;
- repentino incremento di DO nell'arco di un'ora agli inizi di Settembre.

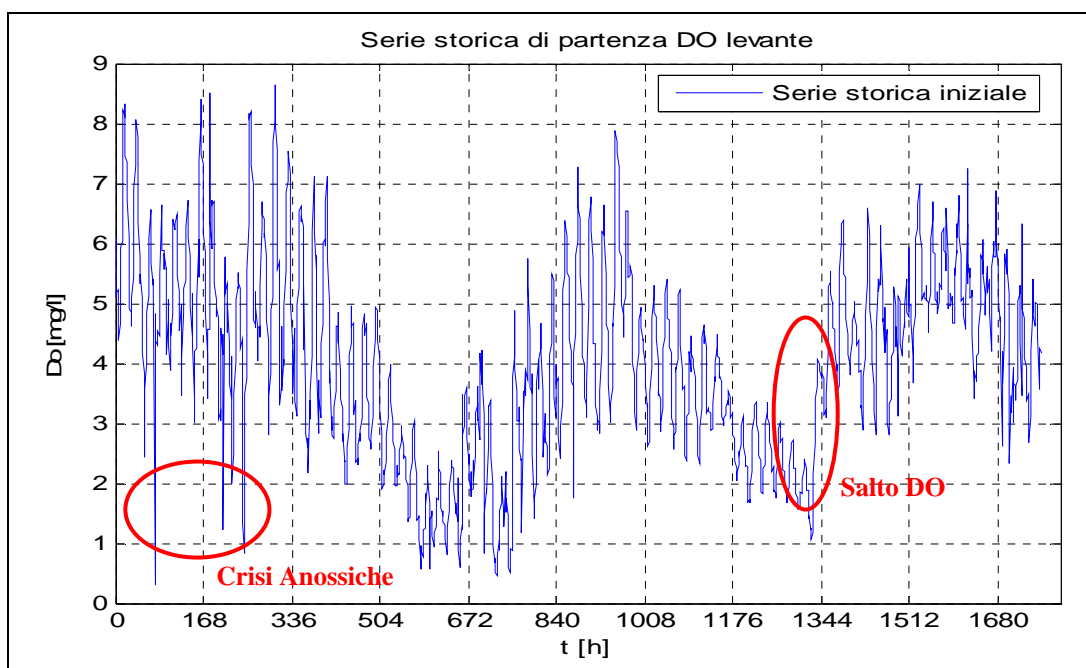


Grafico 2.3 –Evidenza valori anomali nella serie.

Tali valori anomali saranno considerati come outliers, trattati successivamente, sostituendoli con i valori corrispondenti della serie a cui è stato applicato lo smoothing con valore di p prefissato.

Come introdotto, inoltre, i 1800 dati di partenza sono stati depurati degli ultimi 37 campioni misurati nel mese di Settembre, a causa del loro andamento anomalo rispetto ai valori di concentrazione precedenti.

I valori eliminati infatti si aggiravano attorno a 16-18 mg/l, a causa di probabili errori di misura.

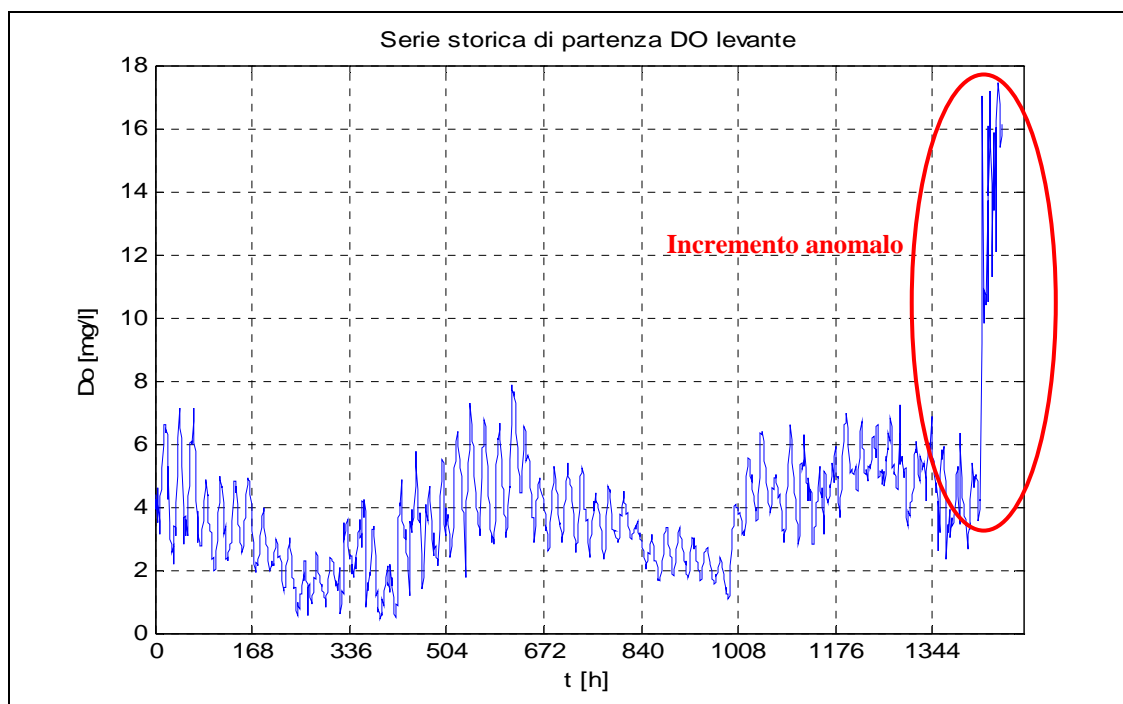


Grafico 2.4 –Evidenza valori anomali nella serie nel periodo di fine Settembre.

La serie storica è stata sottoposta ad analisi in frequenza per evidenziare la presenza di rumore e artefatti, oltre che per individuare eventuali ciclicità del segnale.

In particolare sono stati evidenziati:

- ciclo di 6 settimane;
- ciclo di 24 ore;
- ciclo di 12 ore.

Tali cicli sono stati messi in risalto nei grafici seguenti:

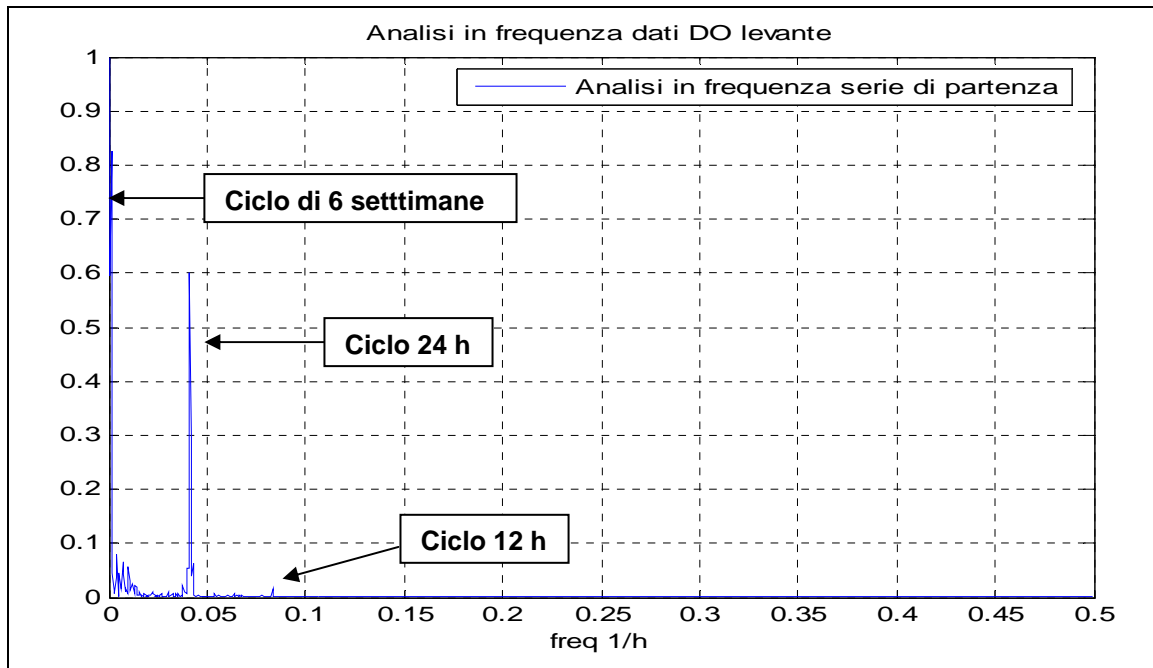


Grafico 2.5 –Analisi in frequenza serie iniziale in scala lineare.

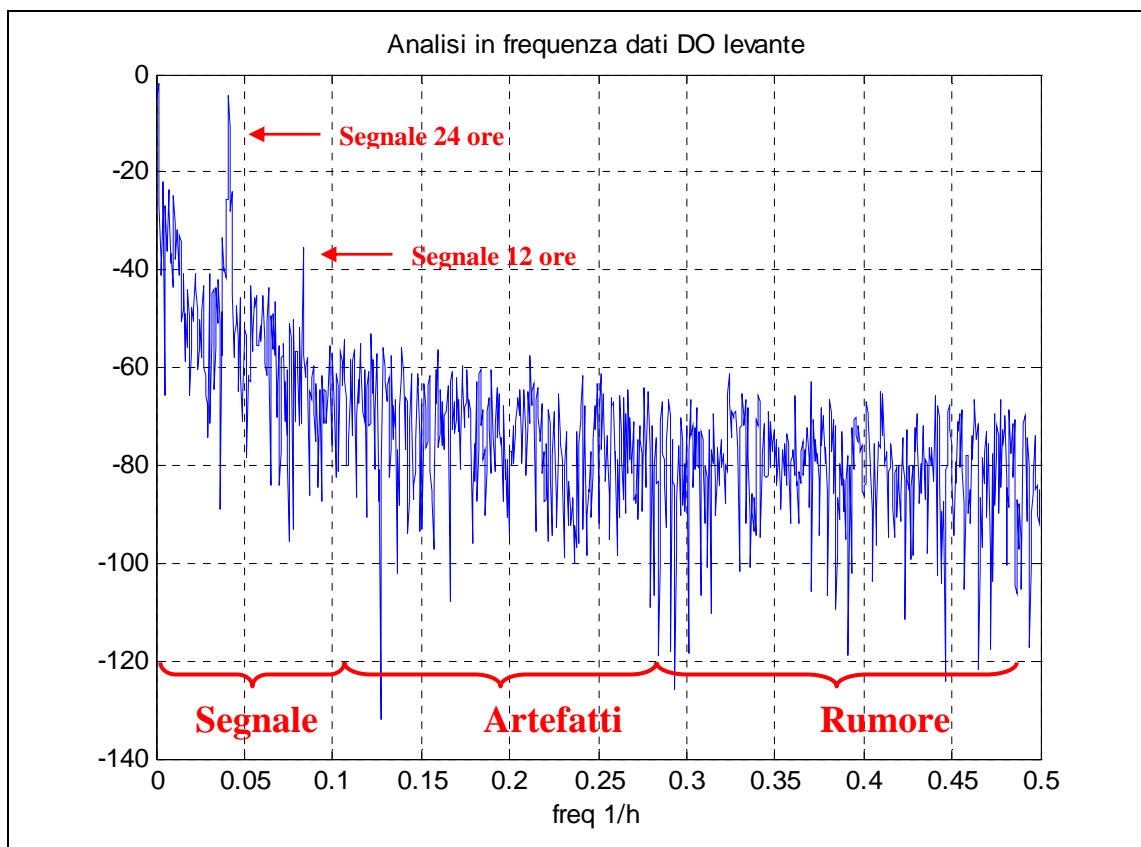


Grafico 2.6 –Analisi in frequenza serie iniziale in scala bilogarithmica.

3. Detrending della serie temporale

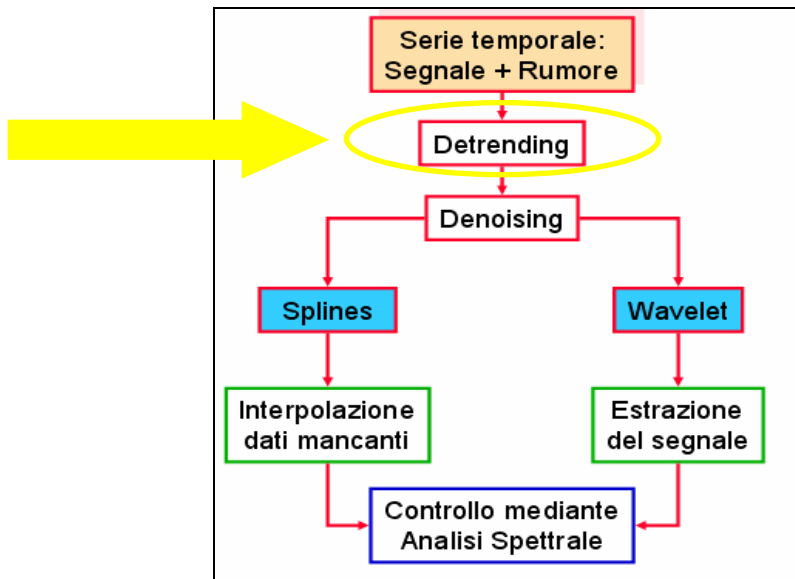


Figura 3.1 – Detrending serie temporale.

La serie di partenza è stata sottoposta ad un’ulteriore analisi al fine di individuare un eventuale presenza di trend, che implica un suo comportamento non stazionario.

Questa analisi è stata condotta tracciando la retta di regressione; un’eventuale presenza di trend può essere evidenziata dal suo coefficiente angolare $b(1)$ diverso da 0.

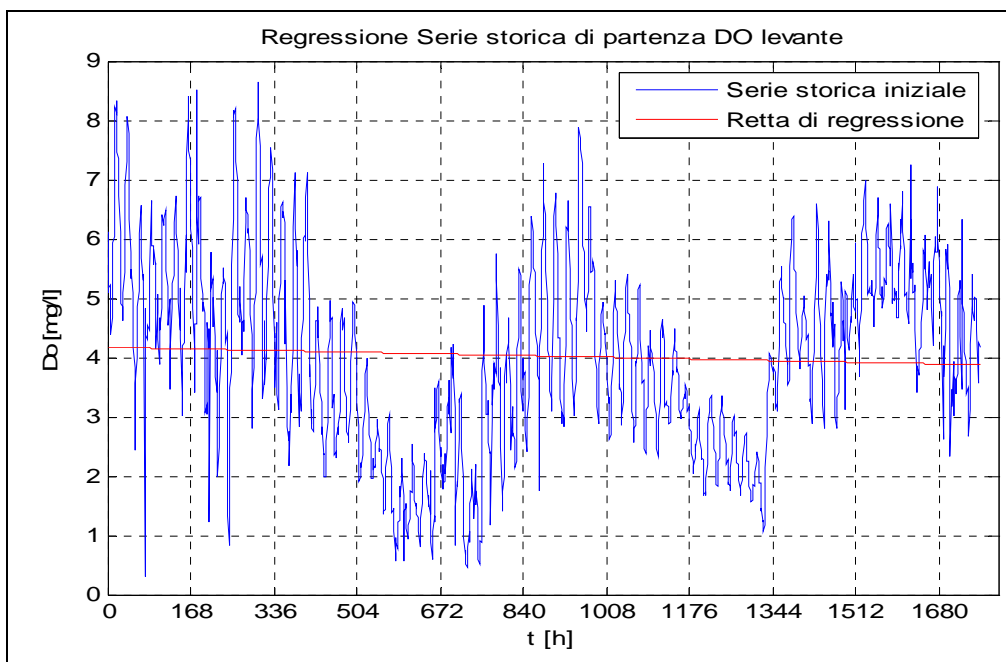


Grafico 3.1 – Regressione lineare serie iniziale.

La trend line è definita come: $y = b(2) + b(1) x$.

Nel caso in esame:

➤ $b(1) = - 1.68 \cdot 10^{-4}$;

➤ $b(2) = 4.17$.

Il valore di $b(1)$ conduce ad affermare che il trend, seppur presente, è così limitato da potere essere trascurato.

Per completezza di analisi, è stata comunque eseguita la detrendizzazione della serie iniziale, come riportato nel grafico seguente:

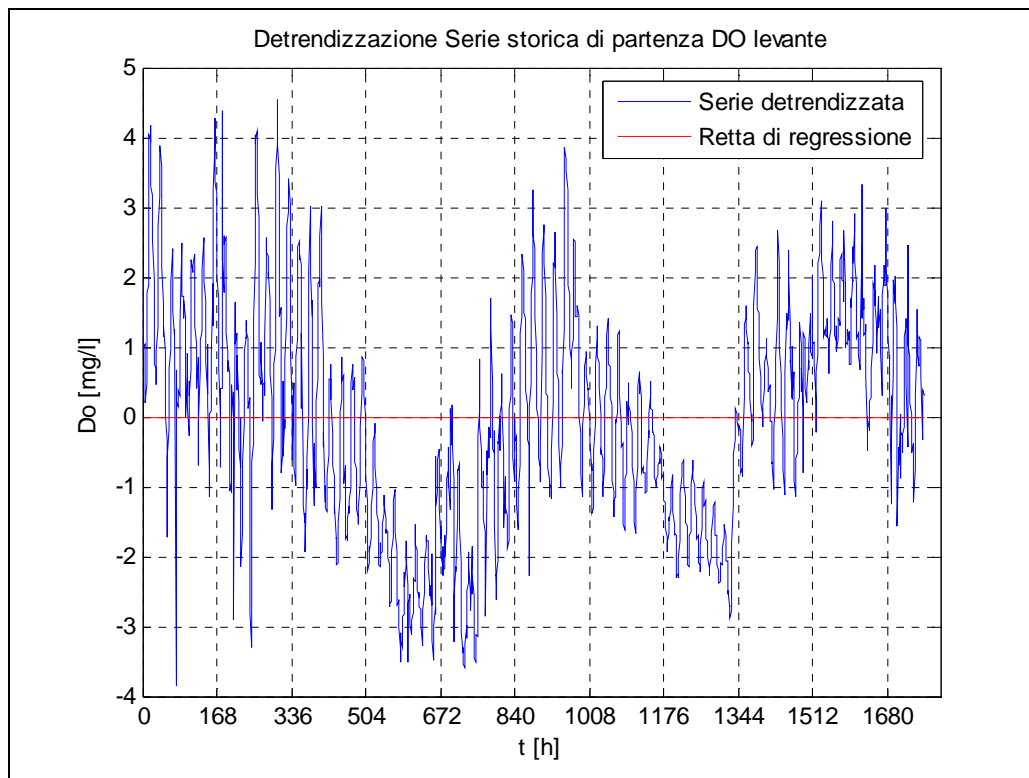


Grafico 3.2 – Detrendizzazione serie iniziale.

Per mettere in risalto la scarsa presenza di trend nella serie di partenza, quest'ultima è stata confrontata con quella sottoposta a detrendizzazione, traslata di una quantità pari a $b(2)$, in modo da fare coincidere l'intercetta relativa all'asse y.

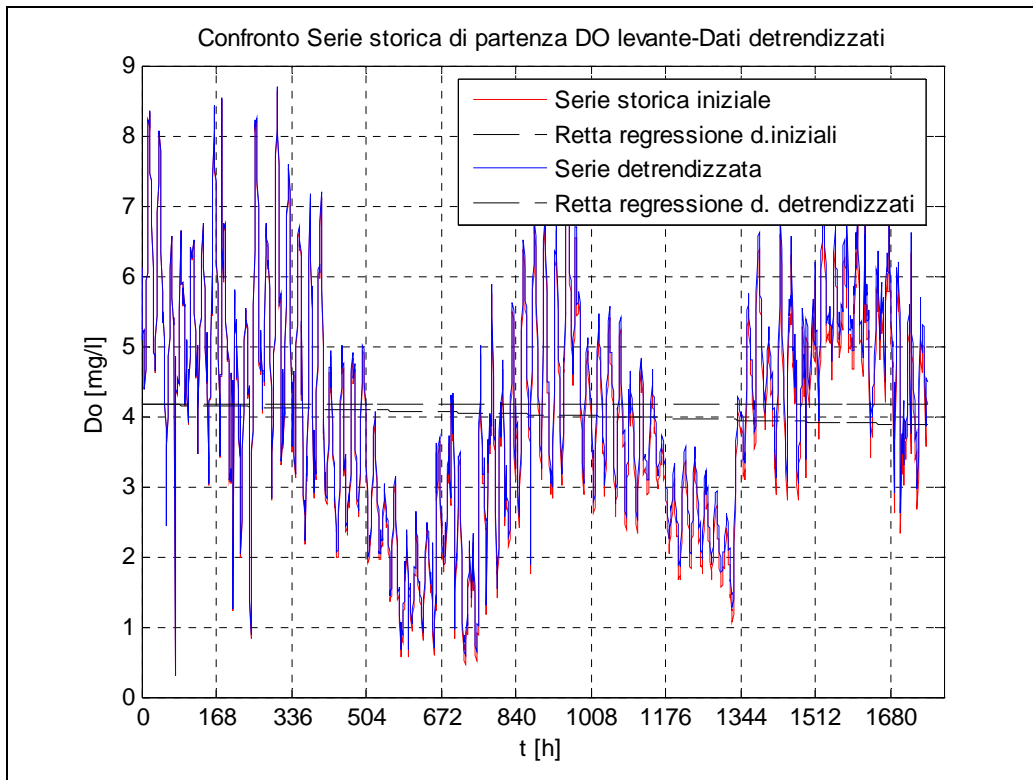


Grafico 3.3 – Confronto serie iniziale – serie detrendizzata.

Viste le considerazioni riportate sopra, per le successive analisi è stata considerata la serie iniziale.

4. Denoising con splines e controllo con analisi in frequenza

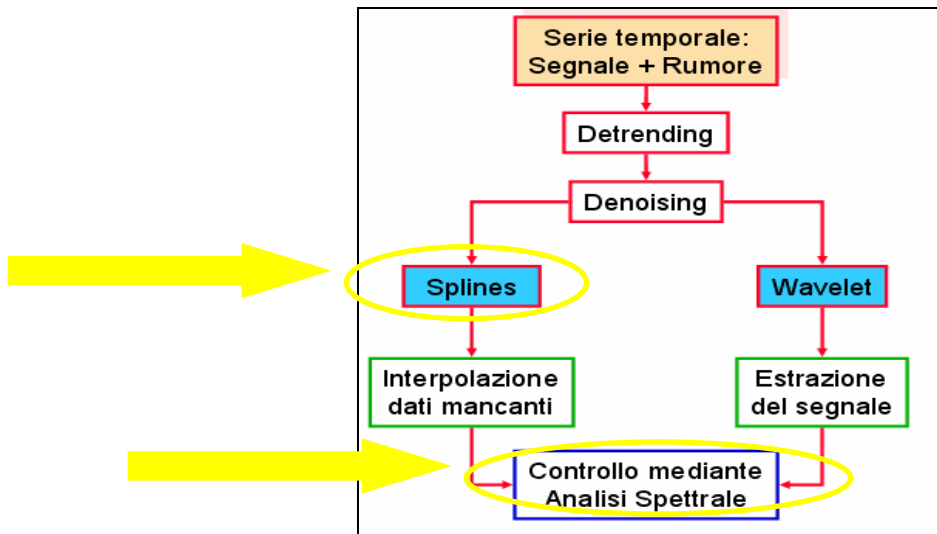


Figura 4.1 – Splines e analisi spettrale serie temporale.

Le funzioni splines sono un semplice strumento per il denoising e per l’interpolazione dei dati mancanti, con il pregio di conservare la continuità delle derivate; nel caso studiato, sono state impiegate per eliminare il rumore dalla serie di partenza, che può mascherare importanti informazioni contenute nei dati e rendere difficoltosa la derivazione del segnale stesso.

Le funzioni splines si basano sulla seguente relazione:

$$p \sum_{i=1}^n (y(i) - y_s(i))^2 + (1 - p)R$$

Fedeltà ai dati
Smoothing

dove:

- R = scabrezza dei dati;
- p = parametro di smoothing → determina l’importanza relativa fra le due esigenze contrastanti di smoothing e fedeltà ai dati.

Si possono avere due condizioni estreme:

- p = 0 → retta dei minimi quadrati;
- p=1 → interpolazione naturale dei dati;
- 0 < p < 1 → soluzione approssimante.

La scelta di p viene effettuata sulla base del tipo di dati in esame con l'ausilio dell'analisi in frequenza, che consente di individuare la frequenza del segnale, degli artefatti e del rumore; valutando i risultati dell'analisi in frequenza in corrispondenza dei diversi valori di p , è possibile scegliere il suo valore più idoneo al fine di conservare il segnale, abbattendo il più possibile rumore e artefatti.

Nel caso in esame, la scelta del parametro di smoothing è stata effettuata valutando vari casi:

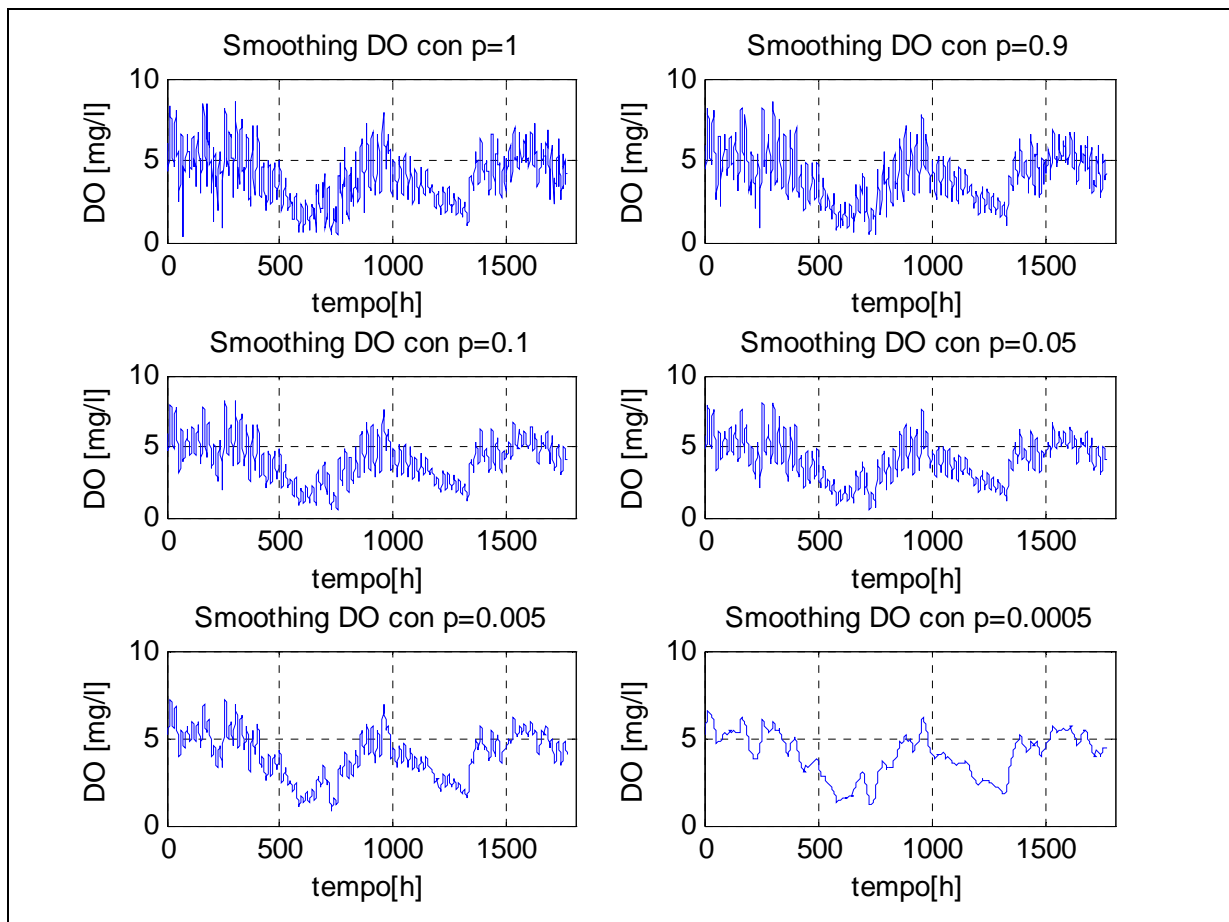


Grafico 4.1 – Smoothing con p diversi.

Dal grafico 4.1 si evince che, decrementando il valore di p , la serie in esame viene notevolmente lisciata, abbattendo i picchi del segnale e tendendo alla retta dei minimi quadrati.

Per la stessa serie è stata condotta l'analisi in frequenza in scala bilogarithmica con i medesimi valori di p .

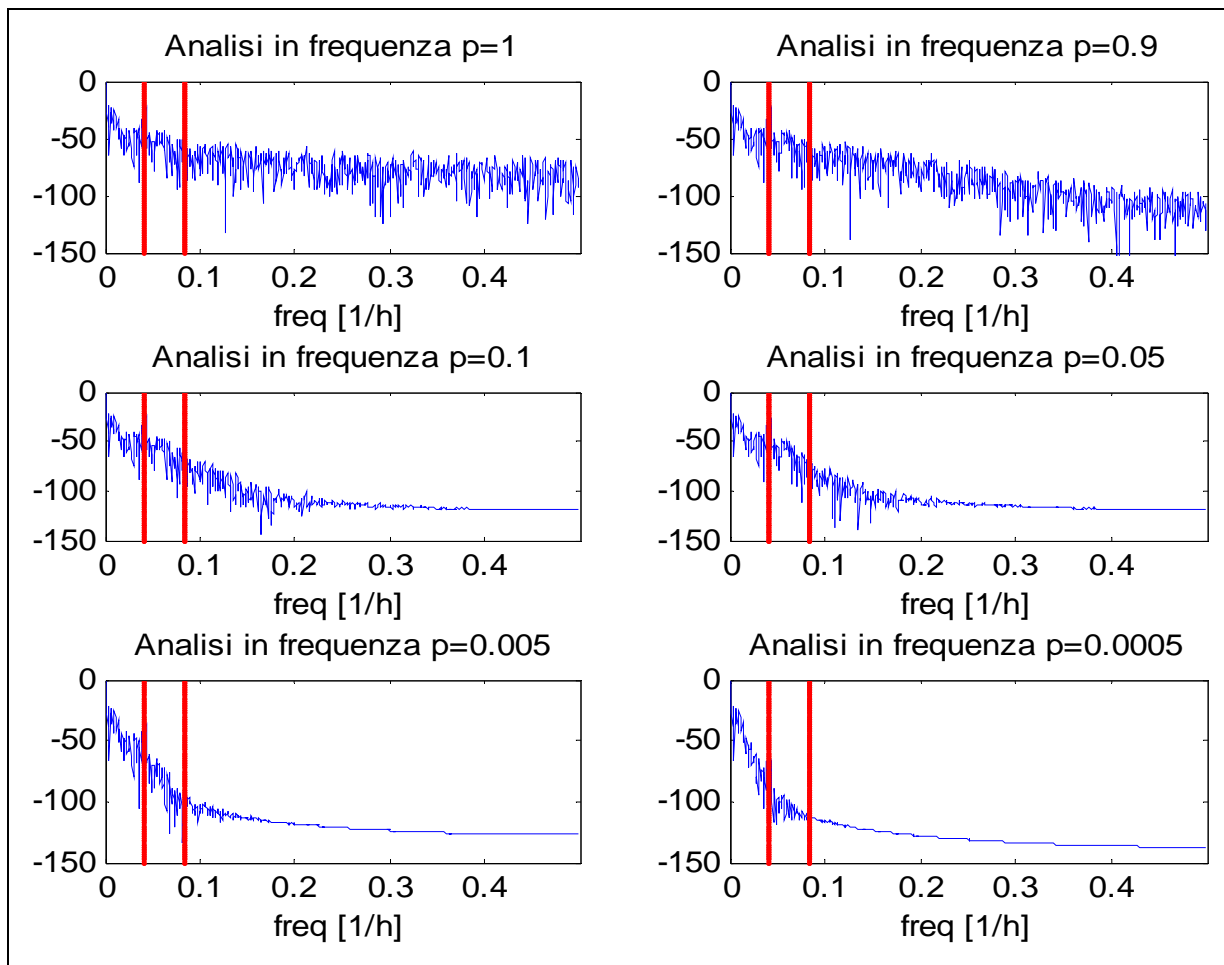


Grafico 4.2 – Analisi in frequenza con p diversi.

Da quest'ultima analisi in frequenza emerge che:

- valori elevati di p permettono di conservare il segnale, ma non eliminano il rumore;
- valori bassi di p mettono in evidenza componenti stagionali ed eliminano in modo marcato artefatti e rumore, a discapito del segnale iniziale (cicli 24 e 12 ore, evidenziati nel grafico 4.2 da linee rosse).

Tenendo conto delle esigenze dello studio condotto, il valore di p pari a 0.1 consente di effettuare uno smoothing che abbatte i picchi più elevati della serie iniziale, non discostandosi eccessivamente da essa; allo stesso tempo, dall'analisi in frequenza, emerge la conservazione dei cicli delle 12 e 24 ore, a fronte di un buon abbattimento del rumore.

5. Analisi dei residui

La valutazione dei residui viene effettuata paragonando il segnale originario con quello sottoposto a smoothing, mediante la loro differenza; nel caso in esame viene utilizzata per vari scopi:

- individuazione dei residui superiori ad una data soglia: outliers;
- considerazioni sulla base della loro distribuzione statistica;
 - analisi della distribuzione approssimata dall'istogramma;
 - paragone di tale distribuzione con la distribuzione gaussiana.

Come detto, per testare l'efficacia dell'analisi condotta, si valuta l'istogramma dei residui e si paragona tale distribuzione con una distribuzione gaussiana avente la stessa media e varianza.

L'eventuale non gaussianità della distribuzione dei residui è riconducibile ad un errore del modello.

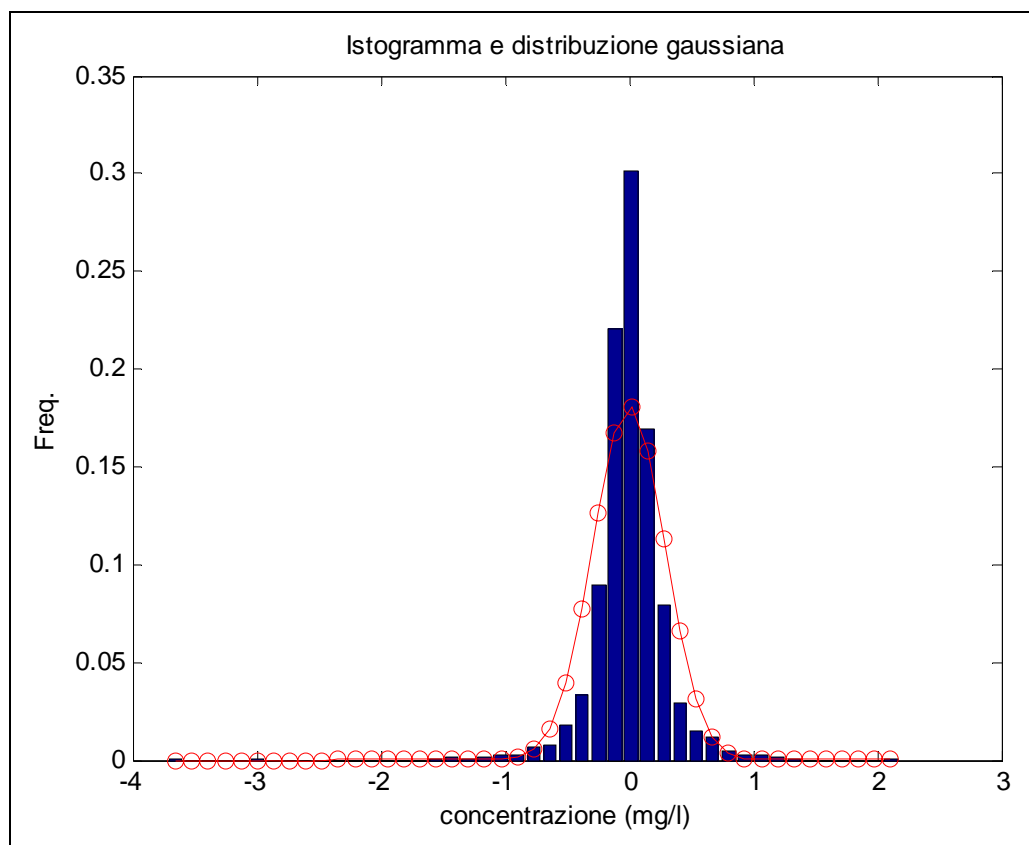


Grafico 5.1 – Istogramma e gaussiana dei residui.

Nel caso in esame, con un p pari a 0.1, la gaussiana si adatta in modo abbastanza corretto all'istogramma a conferma della bontà del modello adottato; tuttavia, non è detto che il miglior smoothing produca residui gaussiani ed incorrelati.

Si riporta inoltre la ricerca degli outliers, osservazioni anomale localizzate molto lontano dalla media del campione, comportando una non omogeneità dello stesso. Gli outliers possono provocare danni notevoli in quanto invalidano l'ipotesi di normalità degli errori. La loro presenza si traduce in residui anomali, ovvero in situazioni in cui il modello sottostima o sovrastima notevolmente il fenomeno oggetto di studio. Se i residui anomali hanno tutti lo stesso segno, la distribuzione dei residui sarà asimmetrica (positivamente o negativamente); se invece i residui anomali si distribuiscono più o meno equamente attorno allo zero allora la distribuzione tenderà ad essere platicurtica (cioè con code più pesanti, in senso probabilistico, del normale).

Solitamente conviene individuare ed eliminare gli effetti di disturbo degli outliers, mediante test di normalità dei residui o analisi visiva del grafico degli stessi.

E' stata adottata quest' ultima metodologia, con il seguente procedimento:

- filtraggio dei dati mediante smoothing effettuato con $p = 0.1$;
- calcolo dei residui;
- individuazione degli outliers intesi come dati i cui residui superano una soglia prefissata pari a $\pm 3\sigma$.

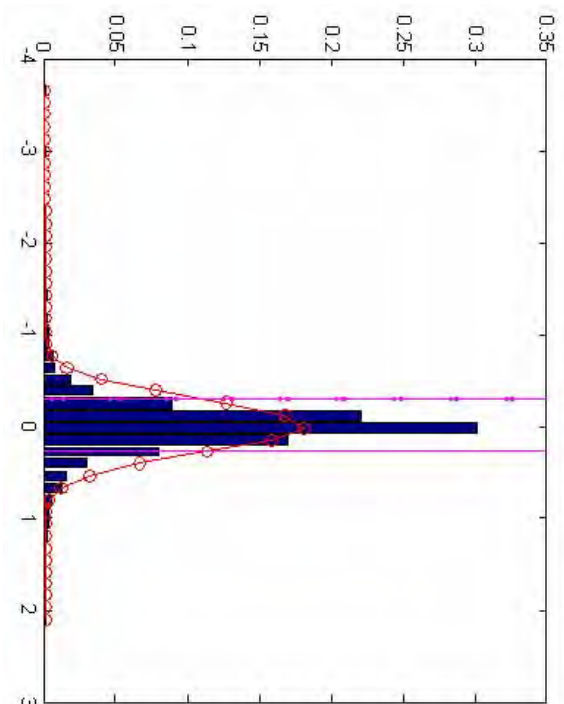
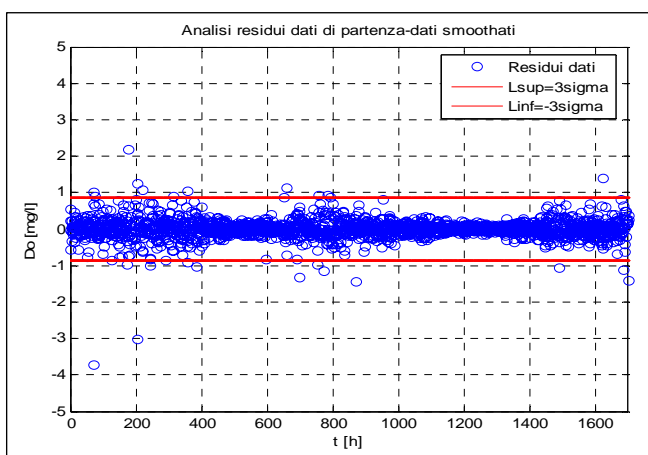


Grafico 5.2 – Individuazione degli outliers.

Il numero complessivo degli outliers è pari a 29.

Aggirandosi il loro valore nell'intorno dello 0, questo implica una distribuzione platicurtica con appesantimento di code come dimostra il grafico 5.1 e 5.2.

Gli outliers individuati sono stati sostituiti con i corrispondenti valori della serie sottoposta a smoothing, ottenendo una nuova serie temporale che viene confrontata con quella di partenza.

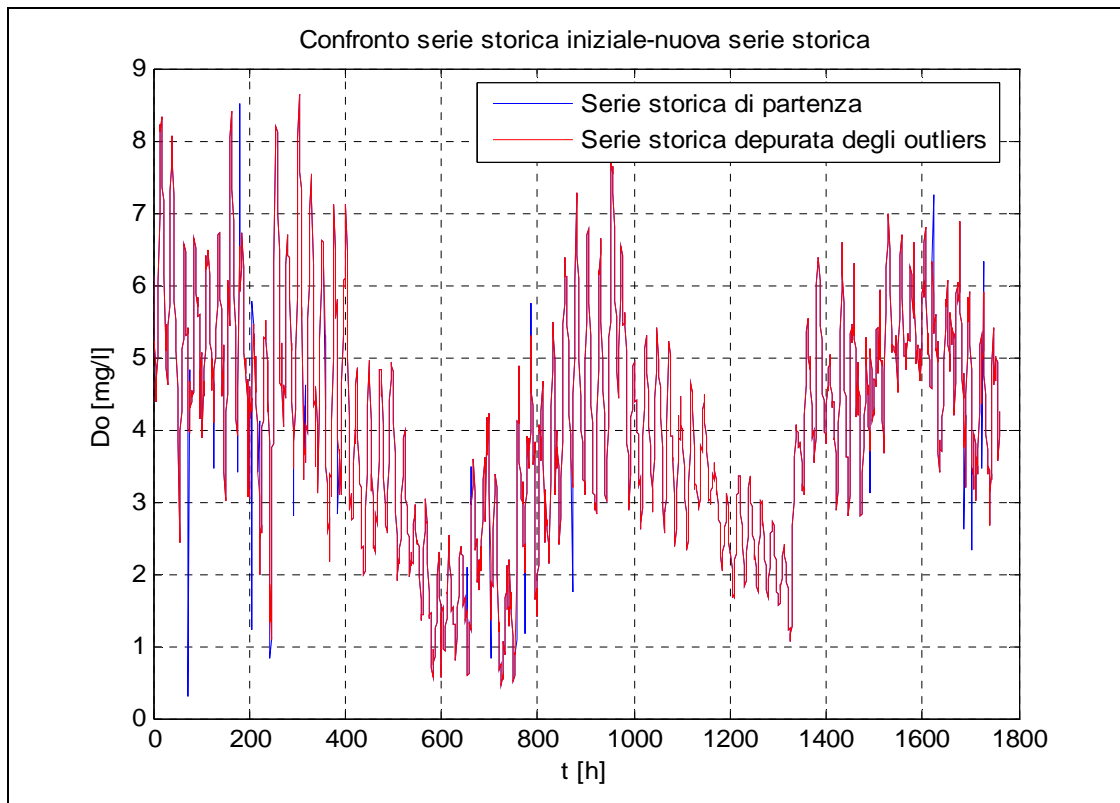


Grafico 5.3 – Confronto serie di partenza – serie trattata.

La nuova serie è stata sottoposta ad un'analisi in frequenza per verificare il mantenimento dei cicli di 12 e 24 ore e confrontata con l'analisi in frequenza della serie iniziale.

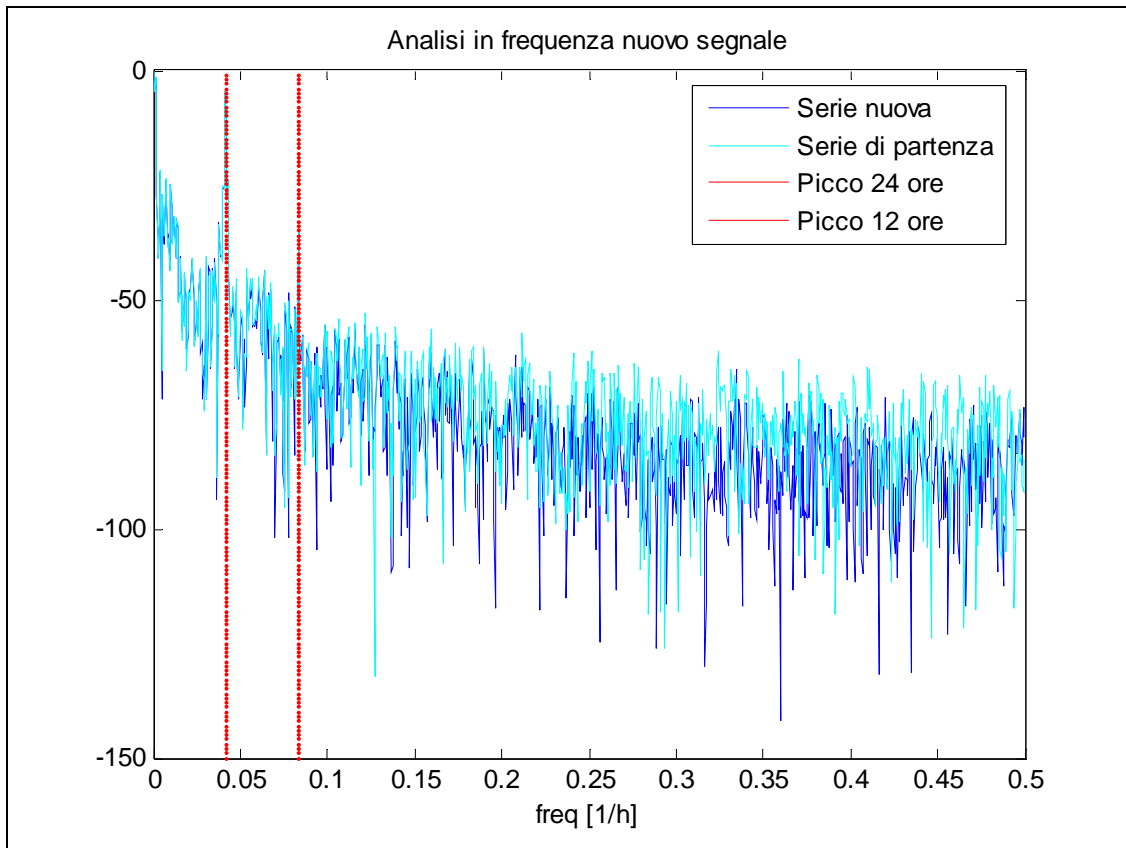


Grafico 5.4 – Analisi in frequenza serie trattata.

Come si nota dal grafico, i picchi delle 12 e 24 h vengono mantenuti anche nella nuova serie e inoltre viene leggermente abbattuto il rumore, attorno alle frequenze superiori a 0.2 /h.

6. Wavelets e controllo con analisi spettrale

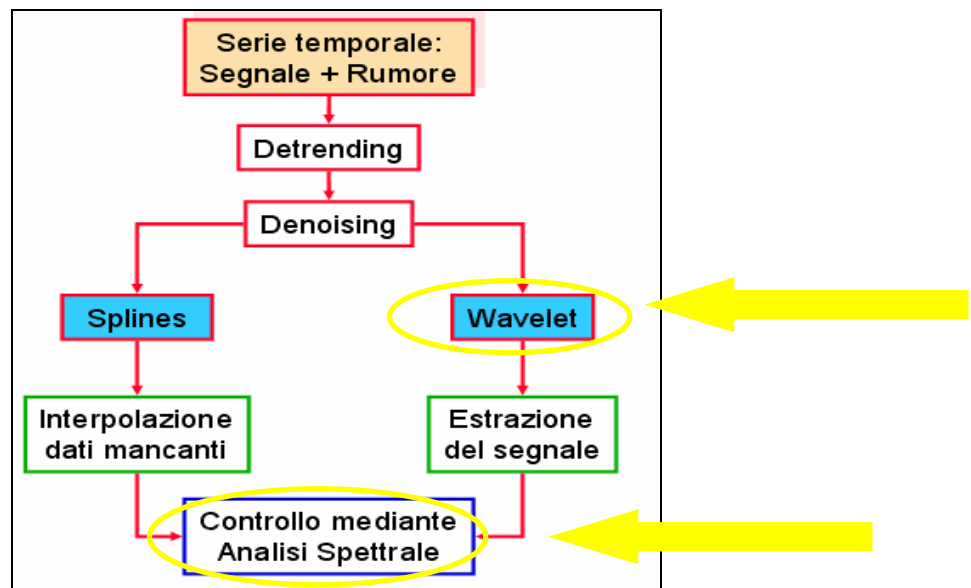


Figura 6.1 – Wavelets e analisi spettrale serie temporale.

Le Wavelets sono una famiglia di funzioni a base radiale (RBF), utili nell'approssimazione di funzioni. Rappresentano un'estensione dell'analisi di Fourier, che tuttavia presuppone segnali di durata infinita (sinusoidi) e quindi, nel trasformare il segnale dal dominio del tempo a quello della frequenza, comporta perdita di informazioni relative alla scala temporale.

Le Wavelets, invece, essendo funzioni a durata finita nel tempo che utilizzando la trasformazione tempo \rightarrow scala, consentono di combinare l'analisi nel tempo e nella frequenza mediante una tecnica di finestre adattative che garantisce l'analisi locale di eventi di breve durata senza perdere la potenza dell'analisi in frequenza.

La Trasformata Wavelet può essere eseguita come:

- Continuous Wavelet Transform (CWT);
- Discrete Wavelet Transform (DWT).

Per l'analisi condotta è stata utilizzata la Trasformata Wavelet Discreta (DWT), che consente di effettuare una decomposizione gerarchica del segnale in termini di approssimazioni e dettagli.

Le approssimazioni contengono la parte a bassa frequenza del segnale, mentre i dettagli contengono la parte ad alta frequenza.

Procedendo con una decomposizione multilivello per dimezzamenti successivi della scala, si applica lo schema riportato di seguito, caratterizzato dall'implementazione iterativa di filtri passa – alto ($g(n)$) e passa – basso ($h(n)$):

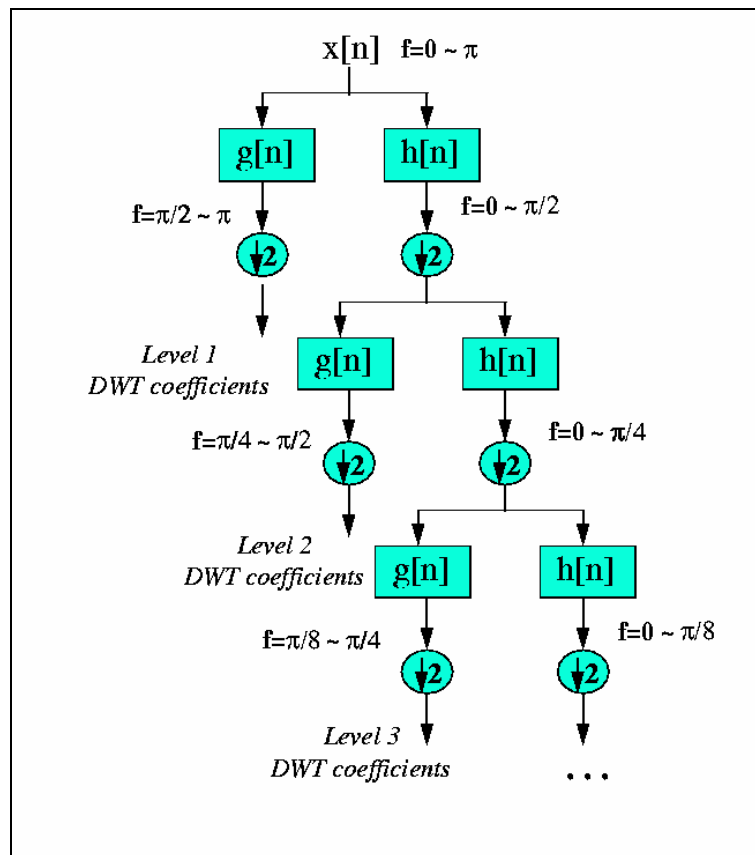


Figura 6.1 – Wavelets e analisi spettrale serie temporale.

Il numero di campioni delle sequenze di dettaglio e di approssimazione si riduce di un fattore 2 ad ogni passo, ma rispetto alla Trasformata di Fourier non viene persa la localizzazione temporale.

La DWT del segnale originale è poi ottenuta concatenando i coefficienti individuati dalla decomposizione, partendo da quello dell'ultimo livello e risalendo fino al coefficiente di primo livello.

Nei confronti della serie di partenza è stata condotta un' analisi con le Wavelets arrivando ad una decomposizione di livello 8.

Si riporta di seguito il confronto tra il segnale originale e le approssimazioni ai vari livelli.

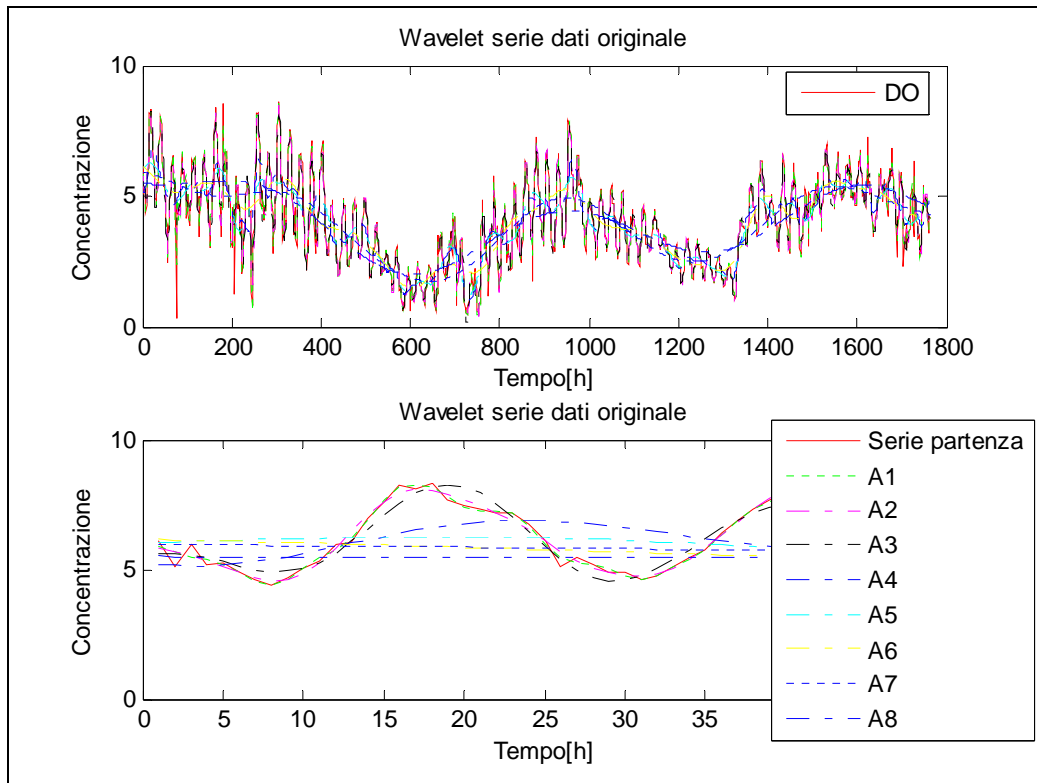


Grafico 6.1 – Approssimazioni con le Wavelet e segnale originale.

Segue inoltre l'andamento dei dettagli ai vari livelli.

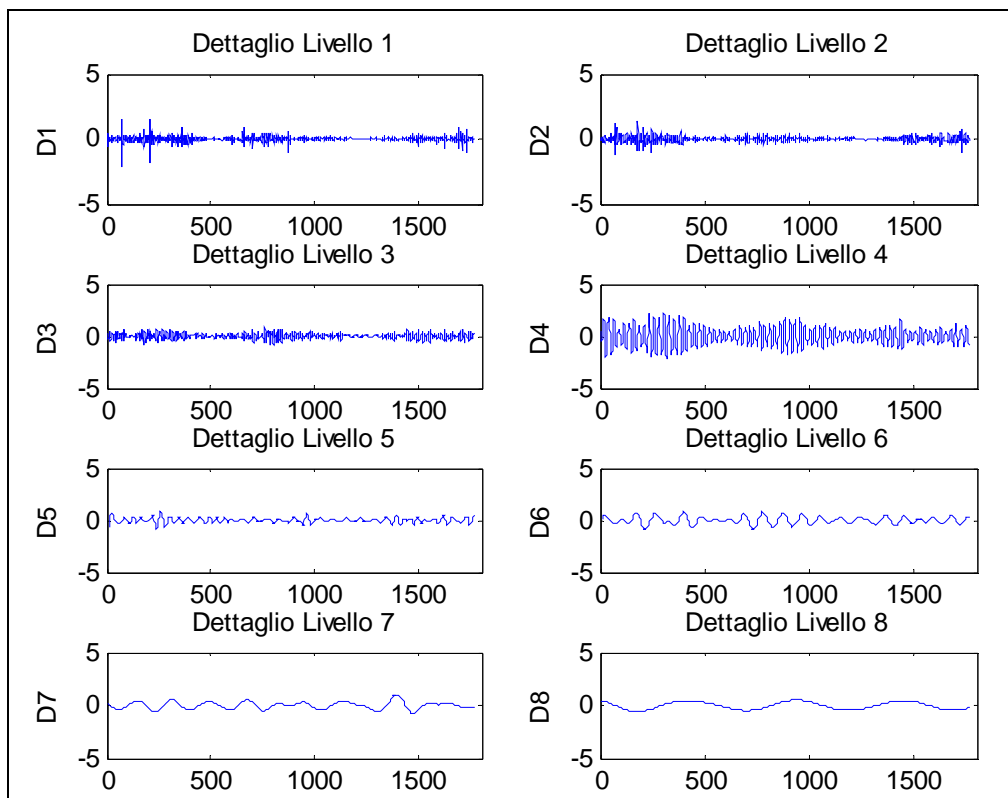


Grafico 6.2 – Dettagli delle Wavelet e segnale originale.

I grafici mettono in evidenza come all'aumentare del livello di analisi il segnale venga sempre più liscio nelle sue decomposizioni di approssimazione e dettaglio, con abbattimento di rumore, ma contemporanea perdita di informazioni. In particolare le approssimazioni effettuano uno smoothing del segnale proporzionale al grado dell'approssimazione stesso; come evidenziato dal grafico 6.1, in corrispondenza di alti livelli di approssimazioni, il segnale appare completamente liscio.

L'analisi tramite le Wavelet è preferibile all'analisi con le splines in quanto dà l'idea delle informazioni contenute nelle decomposizioni.

Per meglio comprendere come viene mantenuto il segnale con i suoi cicli ed abbattuto il rumore ad ogni livello di analisi, è stata effettuata una analisi in frequenza delle approssimazioni e dei dettagli.

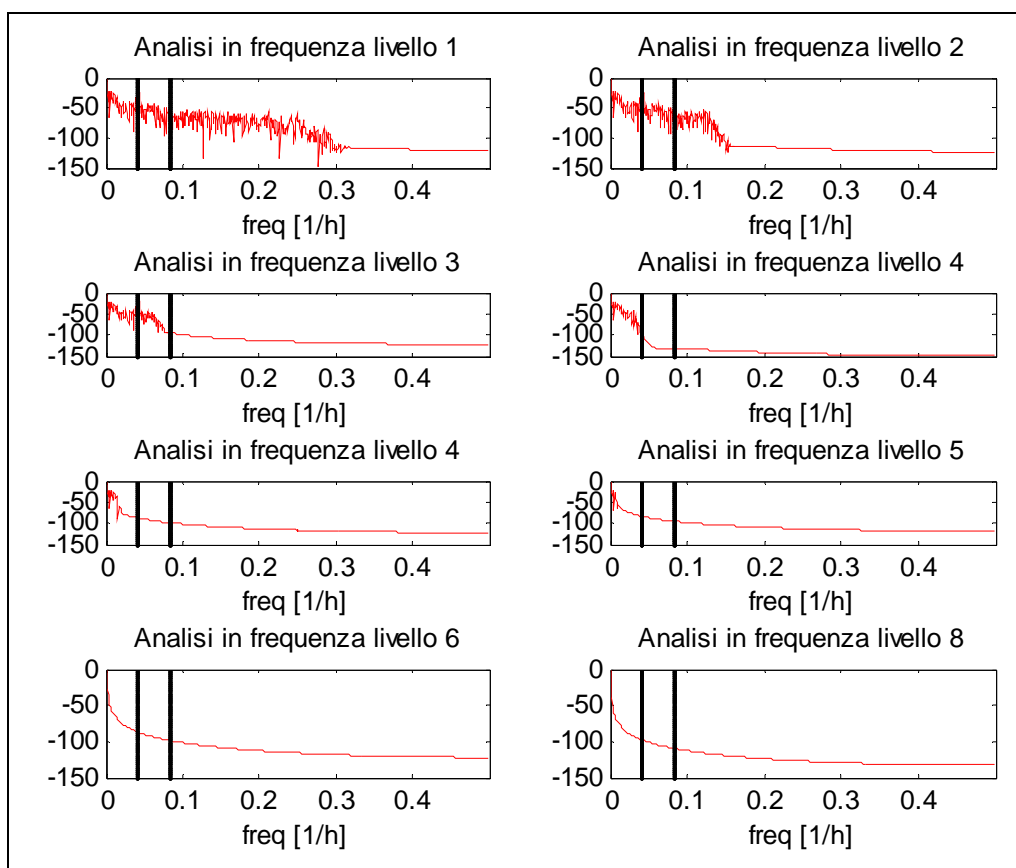


Grafico 6.3 – Analisi in frequenza delle approssimazioni.

All'aumentare del livello di analisi (per i livelli superiori al terzo) viene meno l'identificabilità dei cicli di 12 e 24 ore con un contemporaneo abbattimento del rumore.

Analizzando contemporaneamente l'analisi spettrale dei dettagli, si nota come, incrementando il livello, i dettagli si spostino verso le basse frequenze. Emerge inoltre che, per azione del filtraggio, l'analisi in frequenza dei dettagli si inializzi poco prima dell'interruzione del segnale in frequenza delle approssimazioni.

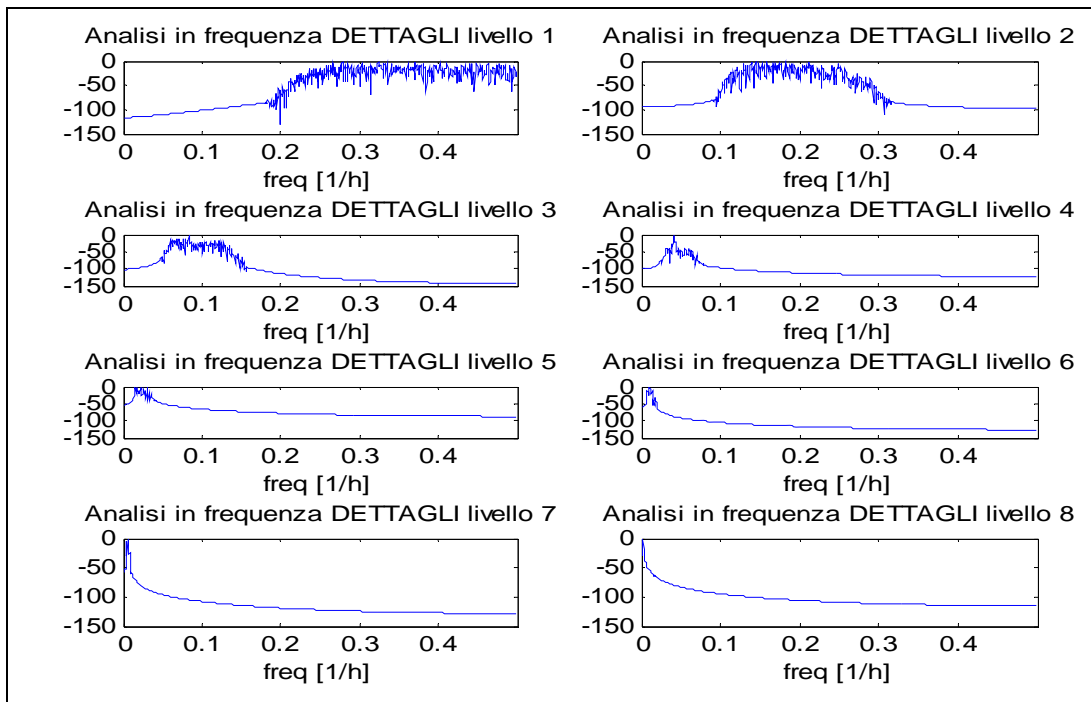


Grafico 6.4 – Analisi in frequenza dei dettagli.

Una discriminante per individuare il miglior livello di analisi ed interrompere correttamente la decomposizione del segnale, è data dall' analisi della sovrapposizione dell'andamento in frequenza delle approssimazioni e dei dettagli. Il livello in corrispondenza del quale i due andamenti in frequenza presentano spettri distinguibili si presenta come la migliore scelta possibile, a fronte di una adeguata conservazione del contenuto informativo del segnale.

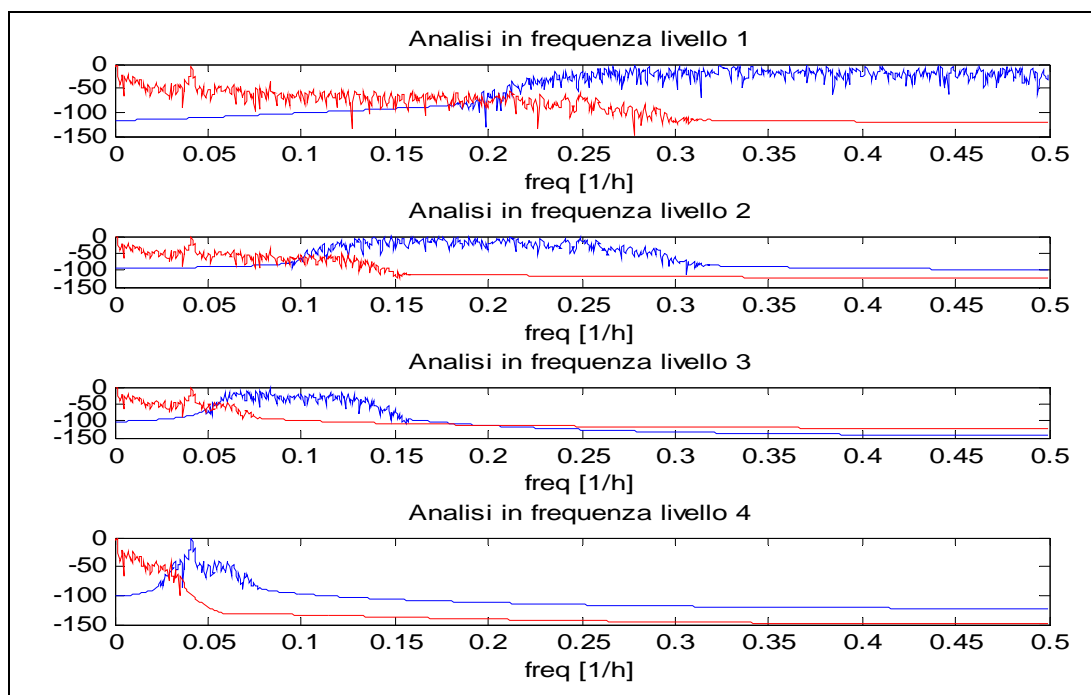


Grafico 6.5 – Sovrapposizioni analisi in frequenza dettagli-approssimazioni.

Come evidenziato dai grafici sopra riportati, al livello 1 l'andamento in frequenza delle approssimazioni e dei dettagli è caratterizzato da una eccessiva sovrapposizione e viene inoltre mantenuta la presenza del rumore nelle approssimazioni.

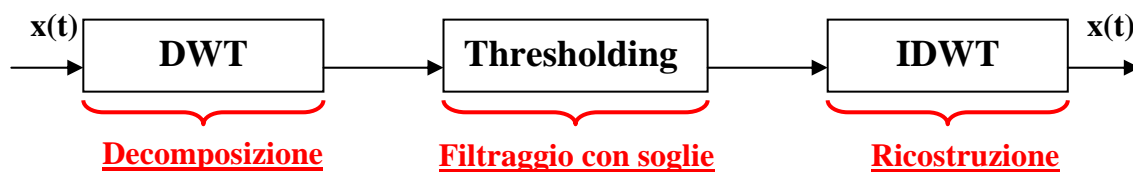
Aumentando il livello di analisi, diminuisce la zona di sovrapposizione e viene abbattuto più efficacemente il rumore, a fronte di una perdita di informazioni nel segnale.

Tenendo conto di tali considerazioni, il livello 2 appare il più idoneo nel caso in esame; viene infatti abbattuto sufficientemente il rumore, mantenuti i cicli a 12 e 24 ore ed inoltre si ha una ridotta sovrapposizione dei due spettri in frequenza (0.05 /h).

Ai livelli 3 e 4, la sovrapposizione viene ulteriormente ridotta, ma si perde l'evidenza della ciclicità a 12 ore.

La DWT è stata utilizzata anche per eseguire il denoising, grazie alle sue eccellenti proprietà di localizzazione.

La procedura standard di denoising si articola in tre passi:



- 1) Decomposizione: si sceglie la Wavelet ed il livello di decomposizione, calcolando i coefficienti di approssimazione e dettaglio;
- 2) Thresholding: per ogni livello si esegue una procedura di thresholding, che può essere hard o soft per i coefficienti dei dettagli. Il metodo soft-thresholding sceglie un valore di threshold e imposta a 0 gli elementi il cui valore assoluto è più piccolo di quello di threshold. Il metodo hard-thresholding si limita ad annullare gli elementi il cui modulo è minore del valore di threshold. Quest'ultimo è un metodo più semplice, ma crea delle discontinuità. In generale è preferibile il metodo soft.

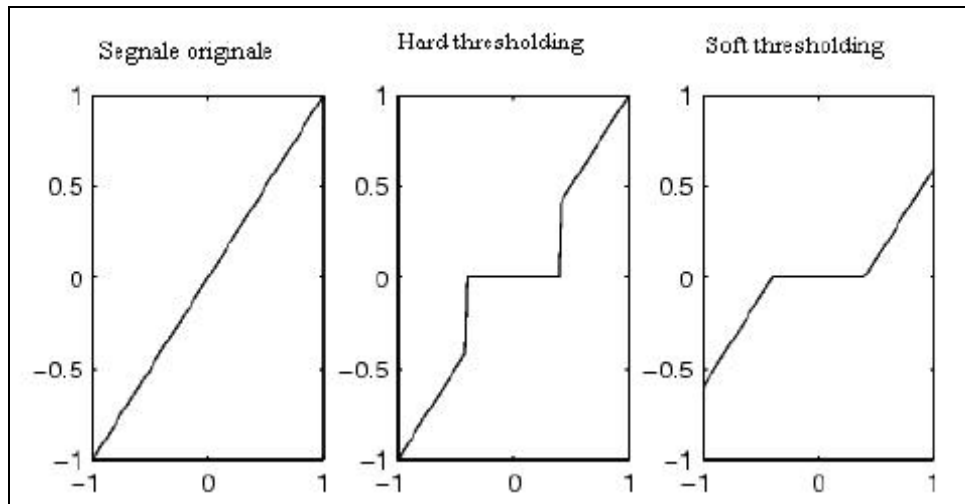


Figura 6.2 – Tecnica di thresholding.

- 3) Ricostruzione : si ricostruisce il segnale usando i coefficienti di approssimazione originali e quelli di dettaglio modificati.

Nel caso trattato è stato condotto un denoising della serie di partenza fino a livello 8, applicando una tecnica di soft-thresholding. L'individuazione della soglia di filtraggio, 'threshold', può avvenire con quattro diverse metodologie:

Option	Threshold Selection Rule
'rigrsure'	Selection using principle of Stein's Unbiased Risk Estimate (SURE)
'sqtwoolog'	Fixed form threshold equal to $\sqrt{2 \cdot \log(\text{length}(s))}$
'heursure'	Selection using a mixture of the first two options
'minimaxi'	Selection using minimax principle

Tabella 6.1 – Regole di selezione del threshold.

E' stata utilizzata la seconda metodologia che prevede di utilizzare un valore di soglia fisso pari a:

$$\sqrt{2 \log(k)}$$

dove k è la lunghezza del segnale.

Si riporta di seguito il confronto tra la serie iniziale e quella sottoposta a denoising ai vari livelli di analisi, secondo il metodo illustrato.

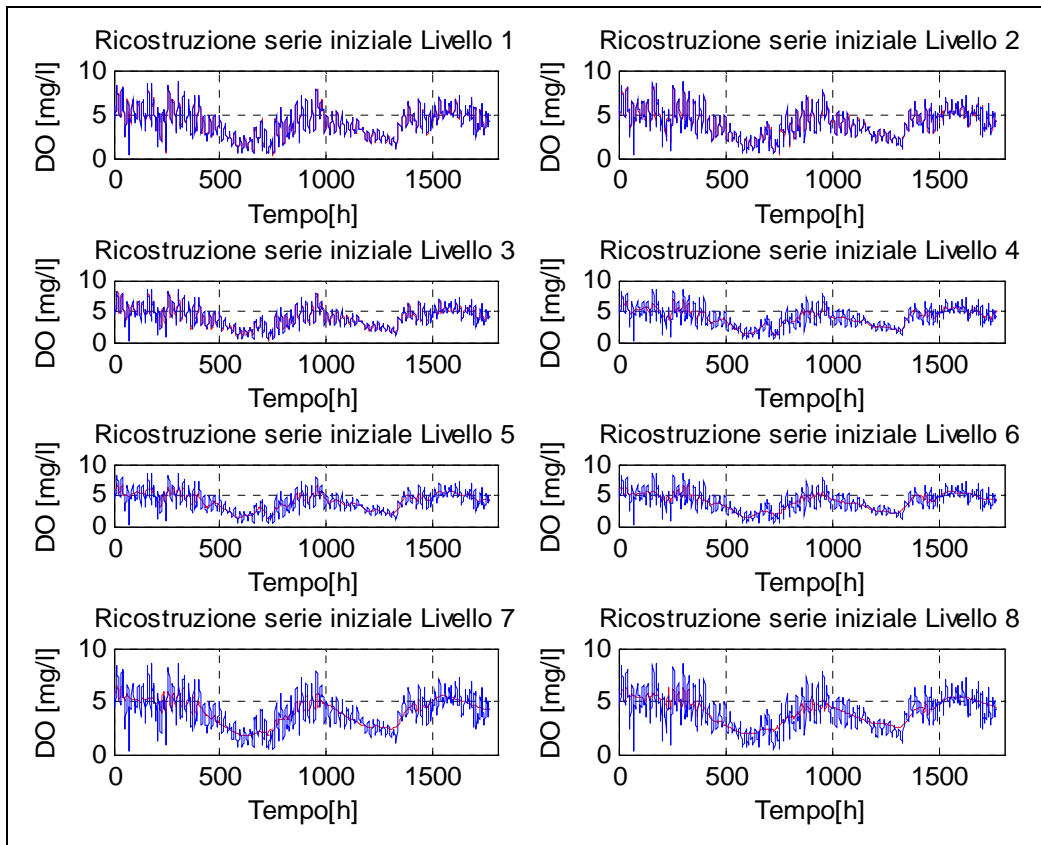


Figura 6.6 – Confronto serie iniziale – serie sottoposta a denoising .

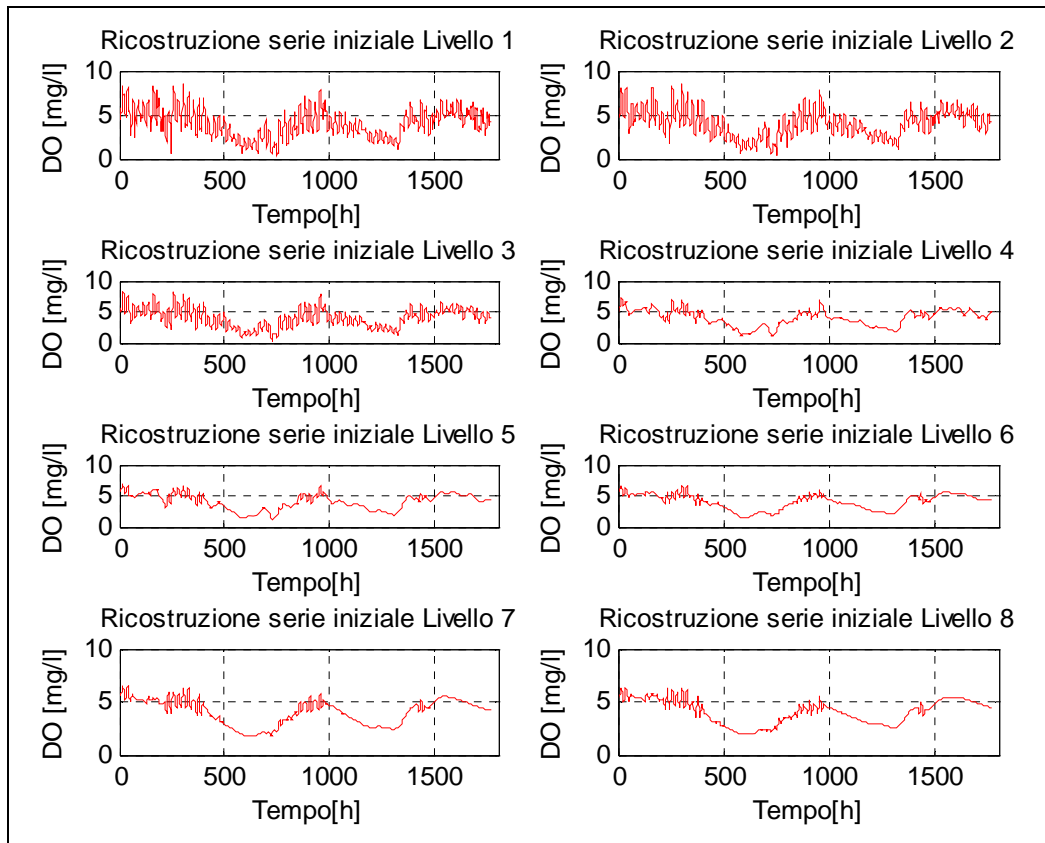
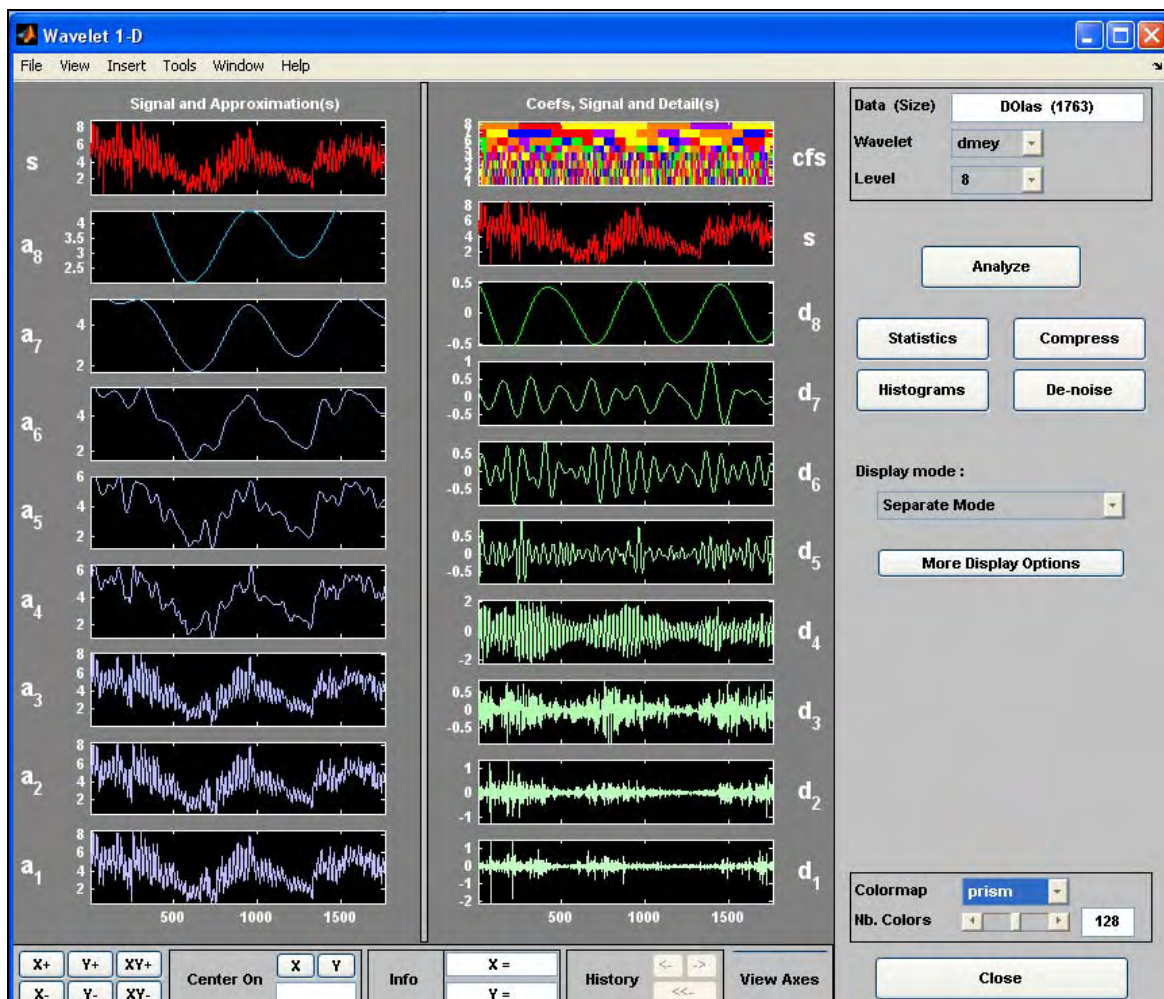
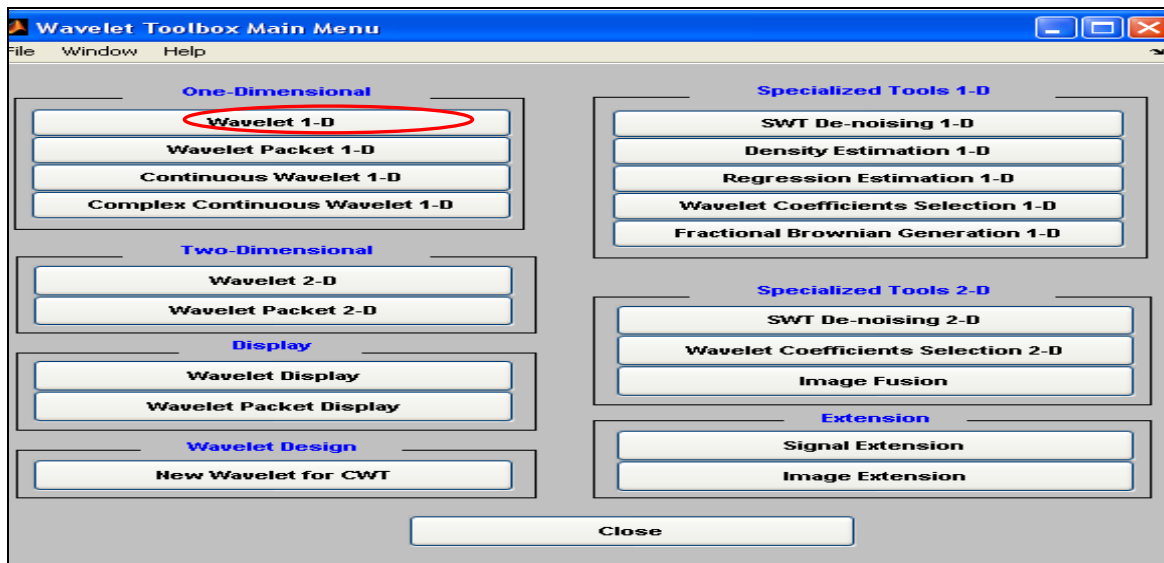


Figura 6.7 – Serie sottoposta a denoising.

Per conferma dell'analisi condotta, si riporta il risultato ottenuto con la medesima procedura eseguita mediante la *toolbox Wavelet* di Matlab.



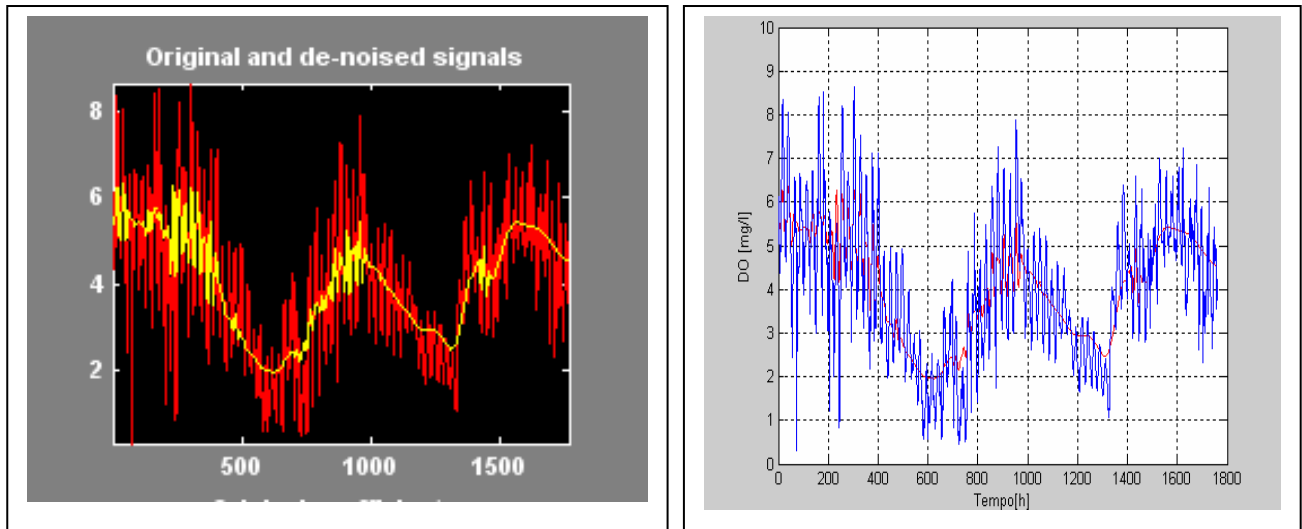


Grafico 6.8 – Confronto denoising Livello 8.

Il grafico 6.8, relativo al denoising a livello 8, dimostra come il risultato sia equivalente.

Date le considerazioni relative alle sovrapposizioni degli spettri in frequenza, che hanno portato all'individuazione del secondo livello come il più idoneo per le esigenze di studio, si riporta anche il confronto tra le due analisi a tale livello.

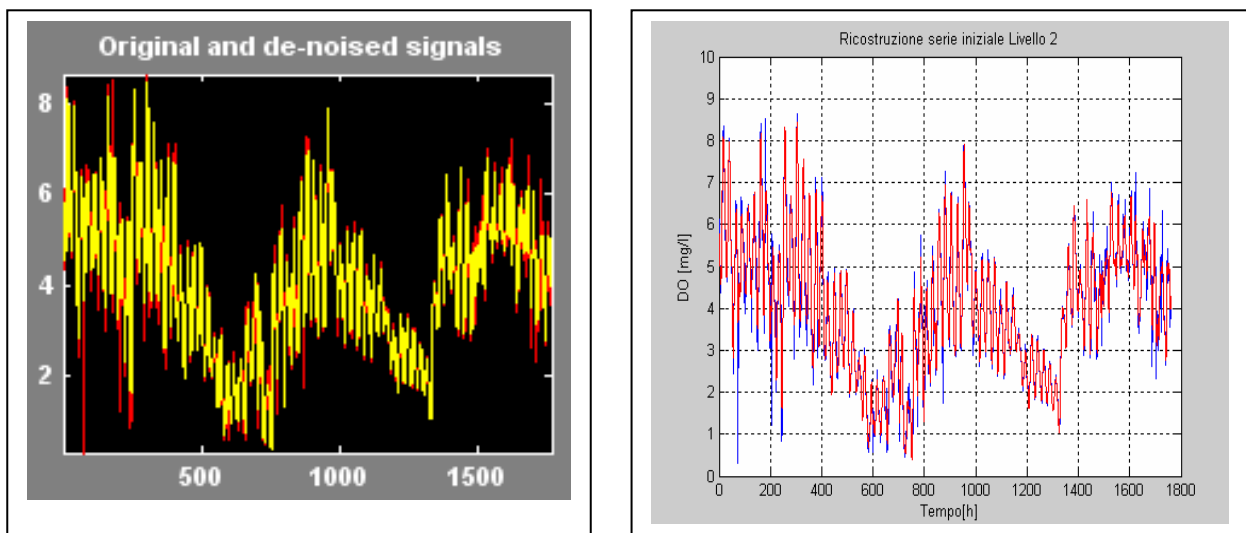


Grafico 6.9 – Confronto denoising Livello 2.

I due grafici riportati sono coincidenti; a livello 2 si ottiene un segnale in cui è stato sufficientemente eliminato il rumore pur conservando una aderenza al segnale originale.

7. Confronto Splines – Wavelet e conclusioni

Le funzioni splines e le funzioni Wavelets rappresentano due metodi per eliminare il rumore da una serie temporale; si riporta di seguito un confronto tra i due metodi applicati ai dati relativi all’ossigeno disciolto nella Laguna di Orbetello, lato Levante, periodo Luglio – Settembre 2001, mettendo in evidenza analogie e differenze.

Le funzioni splines conducono a risultati diversi agendo sul parametro p di smoothing; in particolare, nel caso in esame, sono stati valutati i risultati ottenuti con sei diversi valori di p applicati in ordine decrescente.

I grafici seguenti testimoniano come per p elevati si mantiene la fedeltà alla serie iniziale, mentre, decrementando il valore p si agisce direttamente lisciando la serie ed effettuando uno smoothing più efficace. La scelta del p e la quantità di rumore da eliminare dipende dai dati iniziali e dall’obiettivo posto.

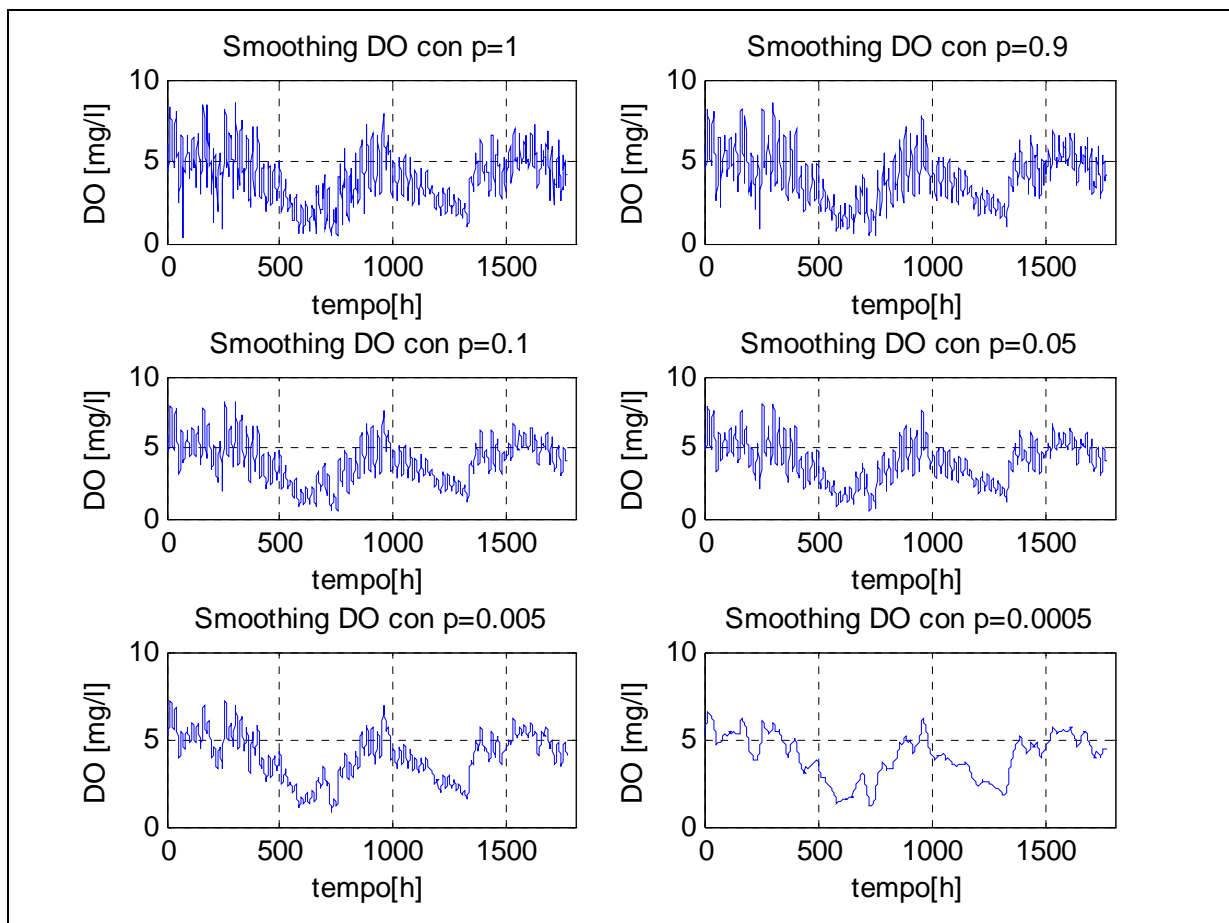


Grafico 7.1 – Splines con p diversi.

La stessa operazione di smoothing viene ottenuta agendo sui diversi livelli delle approssimazioni derivanti dall'utilizzo delle Wavelets; anche in questo caso, all'aumentare del livello di approssimazione, la serie iniziale risulta sempre più liscia, in modo simile al diminuire del p nelle funzioni splines.

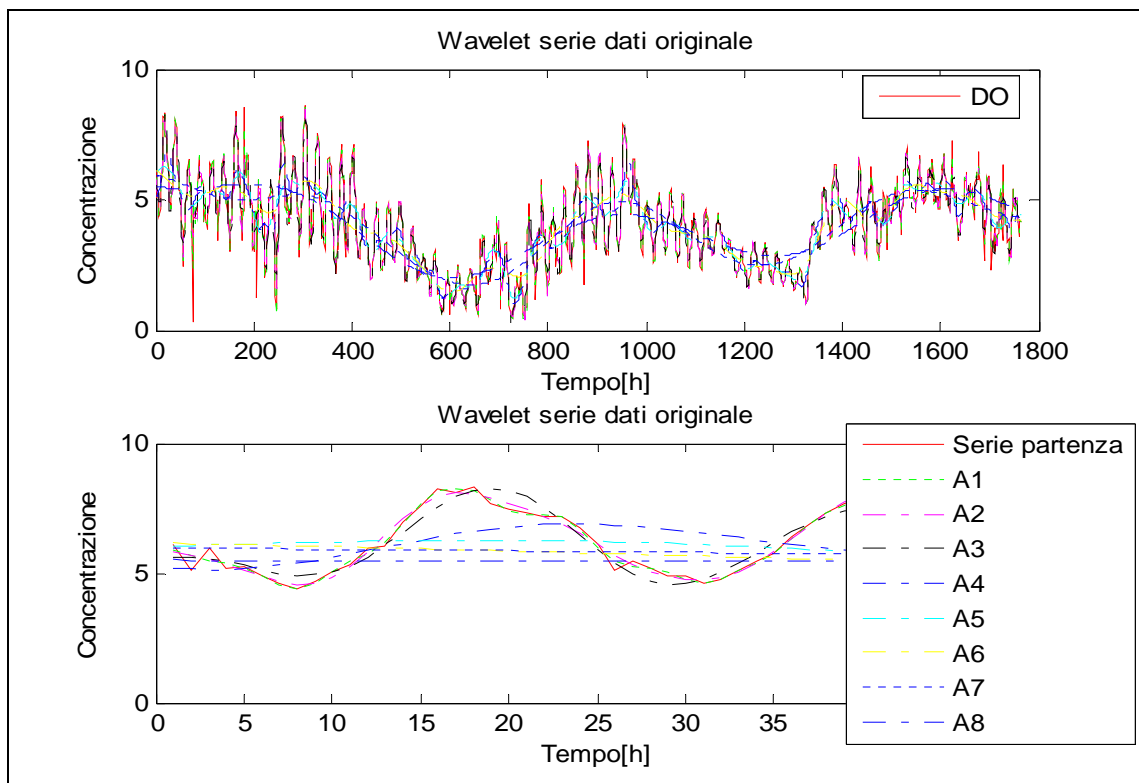


Grafico 7.2 – Wavelet con approssimazioni diverse.

La dimostrazione che una diminuzione del p e un aumento del livello di analisi porta a conseguenze simili in termini di abbattimento del rumore e conservazione del segnale del rumore è data dall'analisi in frequenza.

I grafici seguenti mostrano che:

- una diminuzione del parametro di smoothing p porta ad un elevato abbattimento del rumore, ma anche ad una contemporanea perdita del contenuto informativo e dell'evidenza delle ciclicità orarie;
- un aumento del livello di analisi conduce ad i medesimi risultati, con la possibilità di distinguere meglio le informazioni dal rumore (contenuto nell'analisi dei dettagli) ad ogni livello.

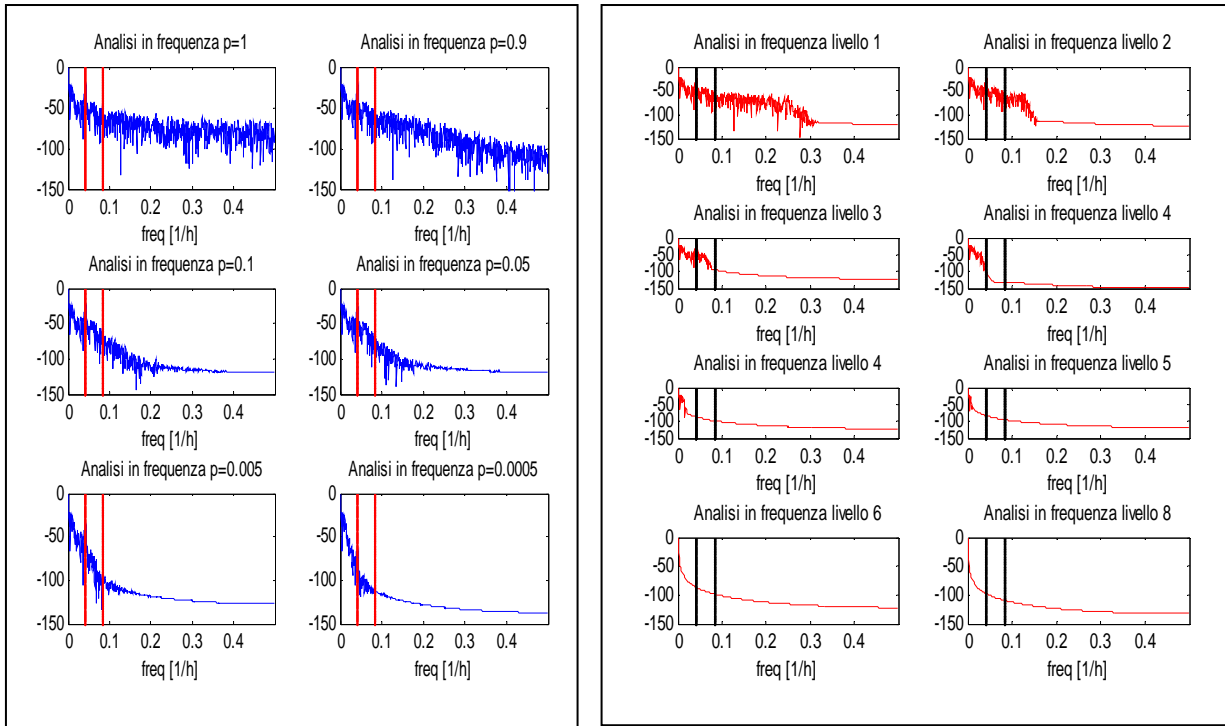


Grafico 7.3 – Confronto analisi in frequenza Splines – Wavelet.

Come dimostra il grafico 7.3, l'utilizzo delle Wavelet consente una eliminazione più marcata del rumore, già ai primi livelli delle analisi, in quanto la parte di dettaglio associata alle elevate frequenze viene filtrata e non passa alle successive analisi.

I dettagli possono essere ugualmente sottoposti ad analisi in frequenza, dimostrando come il loro contenuto sia associato ad elevate frequenze e venga ridotto a livelli successivi di analisi.

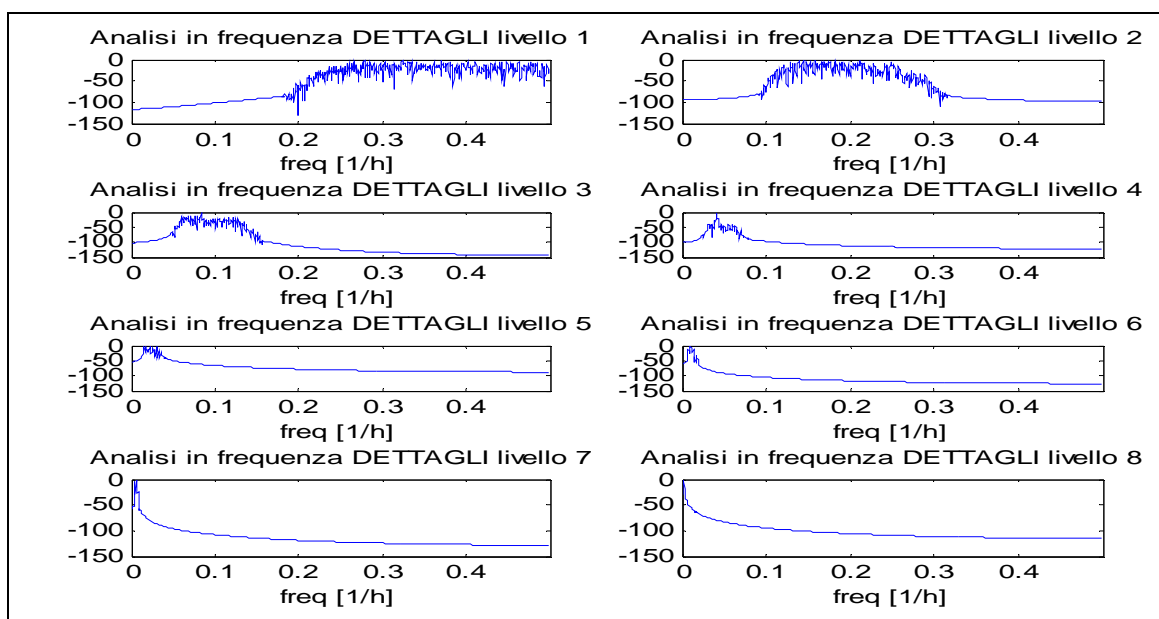


Grafico 7.3 – Dettagli Wavelet.

Sia l'analisi mediante splines che quella mediante Wavelets consente di ottenere una ricostruzione di una nuova serie iniziale.

Le due procedure applicate sono:

- affiancando l'analisi delle splines, con p adeguato, (nel nostro caso p uguale a 0.1) a quella dei residui, si individuano gli outliers presenti nella serie e si provvede a sostituirli con i corrispondenti valori della serie sottoposta a smoothing. Questa procedura consente di ottenere una nuova serie in cui sono stati eliminati i valori anomali e abbattuto leggermente il rumore.
- nell'analisi condotta mediante Wavelet viene direttamente eliminato il rumore dai dettagli con procedura di filtraggio mediante thresholding, utilizzando il livello di decomposizione più idoneo (nel caso in esame il secondo). Si ottiene, quindi, una nuova serie con i coefficienti di approssimazione inalterati e quelli di dettaglio modificati.

Le due procedure sono molto simili, in quanto, in entrambi i casi, viene eseguito uno smoothing ed individuati valori anomali sulla base di soglie.

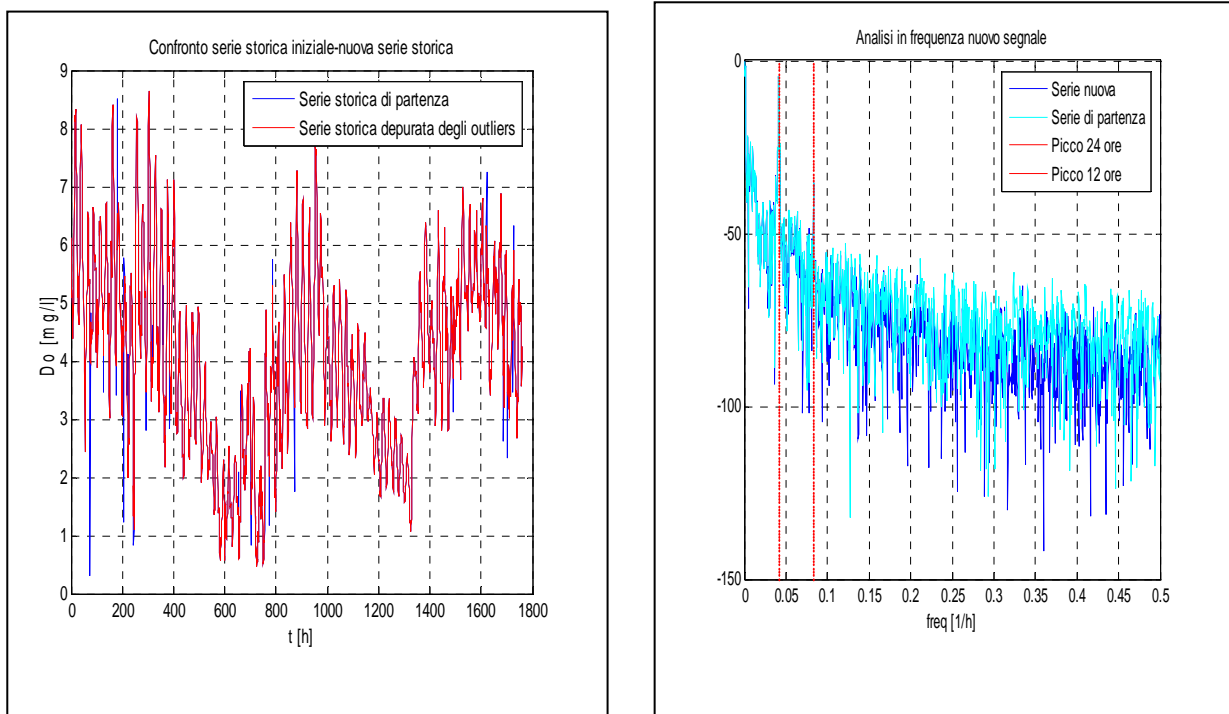


Grafico 7.4 – Nuova serie con splines.

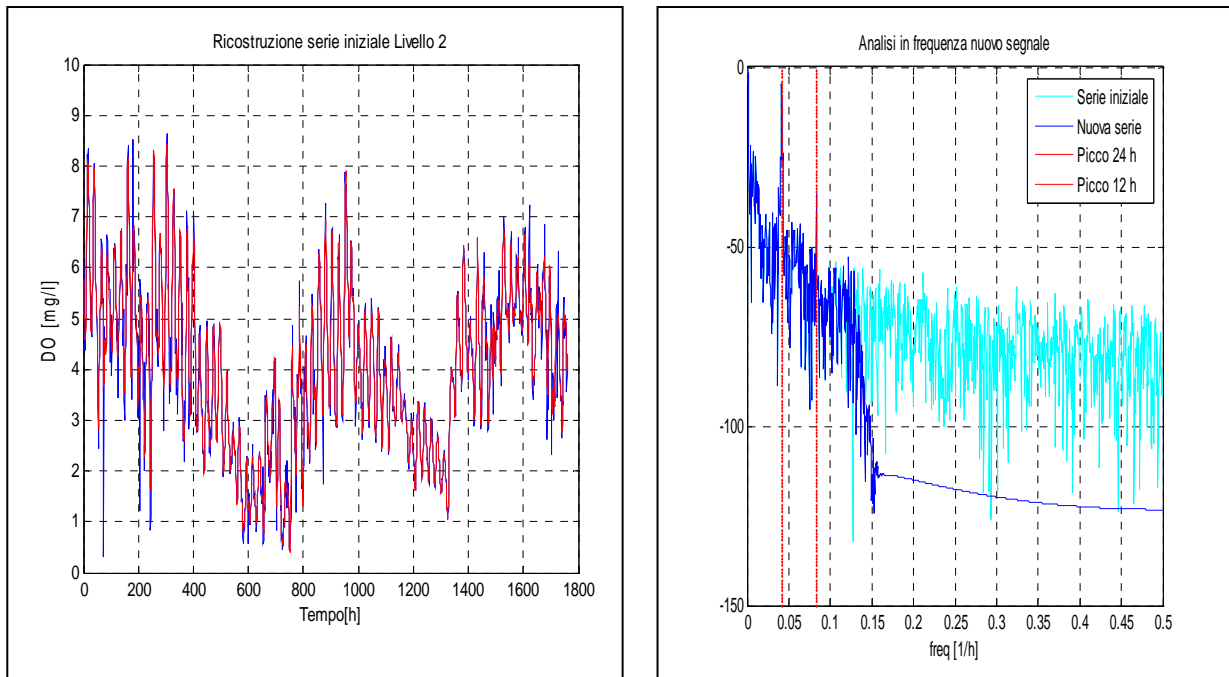


Grafico 7.5 – Nuova serie con Wavelet.

Come si può notare dai grafici 7.4 – 7.5 relativi all’analisi in frequenza, la ricostruzione con Wavelet, basandosi direttamente su una tecnica di denoising, consente un maggiore abbattimento del rumore nella serie ricostruita.

In conclusione, dall’analisi relativa alla concentrazione di ossigeno disciolto nella Laguna di Orbetello, lato Levante, nel periodo Luglio-Settembre 2001, è possibile concludere che:

- L’analisi della serie di partenza ha evidenziato che tale serie è stata ottenuta con un buon campionamento che non ha portato ad avere un eccessivo numero di dati anomali, ad eccezione degli ultimi 37 dati di Settembre, che non sono stati presi in esame nelle successive analisi;
- I dati analizzati sono caratterizzati da trend contenuto e questo implica una loro stazionarietà;
- Il processo di smoothing, eseguito mediante funzioni splines con diversi valori di p , ha portato ad individuare nel valore di $p = 0.1$ il più idoneo tra quelli analizzati; tale valore infatti consente un buon abbattimento del rumore e degli artefatti, con conservazione del segnale, come emerge dall’analisi in frequenza;
- Le funzioni Wavelet consentono un notevole abbattimento del rumore nella serie di partenza; il livello più idoneo per la decomposizione del segnale è il secondo, che, oltre ad

abbattere il rumore e mantenere il segnale, consente una ridotta sovrapposizione degli spettri in frequenza delle approssimazioni e dei dettagli;

- Il confronto tra Wavelets e splines mette in evidenza come entrambe le funzioni consentano di ottenere una nuova serie di partenza in cui il rumore viene ridotto; gli effetti più marcati in termini di riduzione del rumore si hanno con l'uso delle Wavelets, che consentono inoltre un controllo delle informazioni contenute nelle approssimazioni, ad ogni livello di decomposizione.

Appendice A – Storia delle Wavelets

La teoria e le applicazioni delle Wavelets hanno dominato indubbiamente le riviste matematiche ed ingegneristiche dell'ultimo decennio; pochi altri sviluppi matematici, infatti, hanno goduto di questa grande attenzione e sono stati applicati in così tanti rami.

Le Wavelets sono state utilizzate principalmente per elaborare correttamente i segnali; infatti, nell'analisi dei segnali vi erano talvolta alcune lacune che la teoria delle Wavelets ha rimosso.

Si analizza di seguito, da un punto di vista storico, i contributi che hanno condotto alla formazione delle Wavelets.

Le Wavelets rappresentano una estensione della trasformata di **Fourier**, nel 1807, infatti, il matematico francese elaborò l'idea di approssimare una funzione complessa come somma di semplici funzioni, che sono a loro volta ottenute da una "funzione prototipo".

Tale funzione, detta anche "funzione di base", poteva essere costruita come funzione a blocchi in modo che, di conseguenza, la funzione iniziale, rappresentante il segnale originale, potesse essere scomposta in più funzioni a blocchi. Il vantaggio di questa scomposizione è quello di poter passare da una funzione complessa di partenza ad una serie di funzioni semplici.

Come funzioni a blocchi di base, Fourier utilizzò sinusoidi di varia frequenza; il loro andamento provvedeva a rappresentare il contenuto in frequenza della funzione o del segnale iniziale.

Il limite nella scelta delle sinusoidi come funzioni base è che hanno un andamento compatto nel dominio della frequenza, ma non nel dominio del tempo; nel dominio del tempo, infatti, tendono all'infinito e non sono idonee, quindi, nel rappresentare un segnale non stazionario.

La trasformata di Fourier (TFR) consentiva dunque di ricavare informazioni analizzando un segnale nel dominio della frequenza, ma non dava alcuna informazione di carattere temporale.

La prima modifica alla trasformata di Fourier, che consentisse una analisi del segnale anche su scala temporale, portò alla trasformata breve di Fourier (STFT).

L'idea alla base delle STFT era quella di segmentare il segnale, usando una finestra temporale, e analizzarlo contemporaneamente in frequenza; l'uso della finestra temporale avrebbe consentito una rappresentazione tempo-frequenza.

Il primo ad elaborare quest'idea fu **Denis Gabor** nel 1946 per rappresentare un segnale di comunicazione, impiegando funzioni a base oscillatoria.

Subito dopo, nel 1947, **Jean Ville** ideò una trasformata simile a quella di Fourier per rappresentare l'energia di un segnale in un piano tempo-frequenza (la Wigner- Ville Transform).

Dal 1947 al 1970 sono state poi elaborate molte altre teorie, che partivano dalla STFT, modificando la modalità di costruzione della finestra temporale. Il limite di queste teorie era l'uso di un'unica

finestra temporale. Questo limite emerse nel 1970, quando **J. Morlet**, un ingegnere geofisico, si stava confrontando con problemi di analisi di segnali che avevano molte componenti in alta frequenza, con brevi intervalli temporali, e poche componenti a bassa frequenza, con lunghi intervalli temporali.

Le tecniche STFT erano in grado sia di analizzare le alte frequenze, usando delle finestre temporali strette, sia di analizzare le componenti a bassa frequenza, con finestre più ampie; tale analisi, però, doveva essere eseguita separatamente. J. Morlet ebbe l'idea di analizzare il segnale con finestre con banda di frequenza diversa; queste finestre erano tutte generate da dilatazioni e compressioni di una curva di riferimento gaussiana e garantivano una compattezza sia nel dominio del tempo che in quello della frequenza. Morlet nominò tali funzioni di base come 'Wavelet of constant shape'. Tale intuizione venne seguita da ulteriori studi; nel 1980 Morlet con un altro studioso, **A. Grossman**, elaborò una trasformata, partendo dalla Wavelet definita, e la sua inversa.

Nel 1984 lo studio di Morlet venne ripreso da un matematico francese **Yves Meyer**, il quale notò che c'era una grande ridondanza nella scelta delle funzioni di base, dette Wavelets; riprendendo tali studi, Meyer costruì delle Wavelets ortogonali con una buona localizzazione temporale e buona frequenza, facendo riferimento alle Wavelets utilizzate nel 1909 da un matematico tedesco, **Alfred Haar**. Questo matematico aveva costruito delle Wavelets ortonormali estremamente semplici, che non vennero mai utilizzate per la loro scarsa localizzazione temporale; Meyer riprese e migliorò tali Wavelets.

Attorno agli anni '90, uno studente laureato da Grossman, **Ingrid Daubechies**, venne ad introdurre una importante innovazione. Fu lui, infatti, l'inventore della Trasformata Wavelet Discreta. Egli, infatti, individuò delle funzioni Wavelets che consentirono una buona discretizzazione temporale, con corretti parametri di scala e con maggiore libertà nella scelta delle funzioni di base. Inoltre, assieme ad un altro studioso, **Stephane Mallat**, si impegnò nel passare da un segnale continuo ad uno discreto. Nel 1986, infatti, i due svilupparono una analisi multirisoluzione per la Trasformata Wavelet Discreta; il segnale venne suddiviso in bande di frequenza diadica con una serie di filtri passa - alto e passa - basso.

Le ultime innovazioni nella teoria delle Wavelets furono quelle introdotte nel 1988 con lo sviluppo delle basi ortonormali di supporto, introdotte da **Albert Cohen** e dalla introduzione di Wavelet biortogonali.

Appendice B – Trasformata di Fourier, STFT, funzioni Wavelts

➤ Trasformata di Fourier

E' la più antica nella rappresentazione delle funzioni usando basi ortonormali:

$$x(t) = \sum_{k=-\infty}^{\infty} c_k e^{jk(\frac{2\pi}{T})t} \quad \text{con} \quad c_k = \frac{1}{T} \int_T x(t) e^{-jk(\frac{2\pi}{T})t} dt$$

dove x(t) è il segnale da analizzare, T è il periodo del segnale e c_k sono i coefficienti di Fourier che rappresentano le componenti spettrali di x(t).

Il limite della trasformata di Fourier è che la funzione complessa alla frequenza $\frac{2\pi jk}{T}$ non è analizzabile nel tempo dal momento che si estende all'infinito; in altri termini, in una analisi di segnali non stazionari, le esponenziali complesse analizzano il segnale globalmente nel tempo, senza localizzare temporalmente in esso il contenuto spettrale. Questo non rappresenta un problema nell'analisi di un segnale stazionario, poiché tutte le componenti spettrali sono presenti in qualsiasi istante di tempo; per segnali non stazionari, al contrario, il contenuto spettrale cambia nel tempo e questo rende non idonea la trasformata di Fourier.

➤ Trasformata breve di Fourier

Tale trasformata deriva da una modifica della trasformata di Fourier; la STFT consente l'analisi di segnali non stazionari, trasformandoli in segmenti stazionari e analizzando, successivamente, ogni segmento in frequenza con Fourier:

$$S(T, f) = \int x(t) w^*(t-T) e^{2\pi j f t / T} dt \quad \text{con} \quad x(t) = \iint S(T, f) w^*(t-T) e^{2\pi j f t / T} dt df$$

dove w(t) è la finestra di trasferimento dall'analisi temporale a quella in frequenza, f e T sono la frequenza e il parametro di traslazione, * è l'operatore coniugato complesso e S(T,f) è la STFT di x(t) alla frequenza f secondo la traslazione T nel tempo.

Il limite di questo approccio è che si ha una risoluzione costante per tutte le frequenze, dal momento che tale trasformata usa la medesima finestra per l'analisi dell'intero segnale. Se il segnale ha alte frequenze, per intervalli temporali brevi, si dovrebbe usare per la sua analisi una finestra stretta

(ovvero con bande in frequenza più ampie); ma se, al contrario il segnale è anche caratterizzato da basse frequenze per lunghi intervalli temporali, si necessiterebbe di finestre più ampie. Questa variazione nell'ampiezza delle finestre non è possibile con la STFT e tale limite viene superato con le Wavelets.

➤ **Trasformata Wavelets**

La trasformata Wavelet agisce in modo opposto delle STFT, in quanto scompone il segnale in frequenza per poi analizzarlo nel tempo:

$$W(a,b) = \frac{1}{\sqrt{a}} \int x(t) \psi^* \left(\frac{t-b}{a} \right) dt \quad \text{con} \quad x(t) = c_\psi \iint W(a,b) \psi \left(\frac{t-b}{a} \right) da db$$

dove $a > 0$ sono parametri di scala e di traslazione, ψ è la Wavelet di partenza e c_ψ è una costante che dipende da ψ e $W(a,b)$ è la trasformata continua di $x(t)$.

La trasformata Wavelet continua sfrutta finestre temporali che si adattano al segnale e consentono una analisi corretta in tempi brevi, mantenendo una buona potenza nell'analisi in frequenza.

Si può poi ottenere le trasformate Wavelet discrete, discretizzando i parametri a e b . La decomposizione multilivello di $a=2^j$ e $b=2^{jk}$ deriva dall'utilizzo di funzioni a base ortonormale in funzione della scelta di $\psi \rightarrow \psi_{(j,k)}(t) = 2^{-j/2} \psi(2^{-j/2} \psi(2^{-j}(t-k)))$.

La procedura di decomposizione del segnale viene eseguita seguendo lo schema di figura 6.1, usando filtri passa-alto e filtri passa-basso. Ad ogni livello di decomposizione vengono calcolati i coefficienti:

$$y_{high}(k) = x(n)g(2k - n)$$

$$y_{low}(k) = x(n)h(2k - n)$$

dove n rappresenta il livello di decomposizione del segnale.

Il segnale originale può essere ricostruito per passaggi inversi, ricomponendo i coefficienti ottenuti:

$$x(n) = \sum_k (y_{high}(k)g/(2k - n)) + (y_{low}(k)h(2k - n)).$$

Appendice C – Principali Funzioni Matlab impiegate

- $Y=uiigetfile('FilterSpec','DialogTitle')$: mostra una finestra di dialogo denominata 'DialogTitle'.
- $y = fft(x)$: implementa la Trasformata Discreta di Fourier di un vettore di lunghezza x , restituendo un vettore complesso, il cui modulo è lo spettro di potenza del segnale $|F(w)|^2$.
- $p = polyfit(x,y,n)$: trova i coefficienti di un polinomio $p(x)$ di grado n . P si presenta come un vettore riga, di lunghezza $n+1$, costituito dai coefficienti polinomiali in ordine decrescente.

$$p(x) = p_1x^n + p_2x^{n-1} + \dots + p_nx + p_{n+1}$$

- $y = polyval(p,x)$: restituisce la stima di un polinomio di grado n , valutato in funzione di x . L'imput p è un vettore di lunghezza $n+1$, costituito da elementi che sono i coefficienti, in ordine decrescente, del polinomio che deve essere stimato. X può essere una matrice o un vettore; in entrambi i casi, 'polyval' individua un valore di p per ciascun elemento di x .
- $y = detrend(x)$: rimuove dal vettore x la presenza di *trend*, ovvero di un comportamento stazionario. Se x è una matrice, il comando *detrend* rimuove il *trend* da ciascuna colonna. La detrendizzazione generalmente è impiegata nell'ambito di processi FFT.
- $y=csaps(x, y, p)$: fornisce una spline cubica f , approssimante i dati (x,y) , con un parametro di smoothing p . La spline f minimizza:

$$p \sum_{j=1}^n w(j) |y(:, j) - f(x(j))|^2 + (1-p) \int \lambda(t) |D^2 f(t)|^2 dt$$

- $subplot(m,n,p)$: suddivide la finestra grafica in una serie di pannelli rettangolari disposti su m righe e n colonne, visualizzando il diagramma successivo nel p -esimo pannello.
- $M = mean(A)$: restituisce il valor medio degli elementi costituenti l'array A .

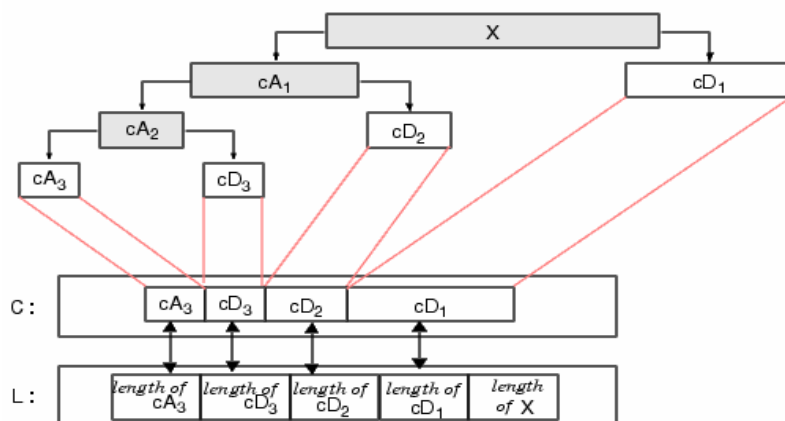
- $s = std(X)$: calcola la deviazione standard di x , dove x è un vettore, applicando la relazione (1):

$$(1) s = \left(\frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 \right)^{\frac{1}{2}}$$

- $y = normpdf(X, MU, SIGMA)$: calcola una funzione di densità di probabilità normale, con media MU e varianza $SIGMA$. Mu e $SIGMA$ possono essere vettori, matrici o array multidimensionali che devono avere la stessa dimensione. $SIGMA$, inoltre, deve essere positivo. La distribuzione di probabilità normale viene calcolata con la seguente formula:

$$y = f(x|\mu, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

- $B = cumsum(A)$: restituisce la somma cumulata di A . Se A è un vettore, questa funzione restituisce un vettore contenente la somma cumulata degli elementi di A . Se A è una matrice, invece, la funzione *cumsum* genera una matrice della stessa dimensione di A , contenente la somma cumulata per ciascuna colonna di A .
- $[C,L] = wavedec(X,N,'wname')$: la funzione *wavedec* restituisce una decomposizione wavelet multilivello 1-D del segnale, al livello N , impiegando una specifica trasformata, identificata con 'wname'. N deve essere intero positivo. L'output della struttura di decomposizione è rappresentato dal vettore C e dal vettore 'contatore' L . Si riporta un esempio di decomposizione al terzo livello:



- $[Ea,Ed] = WENERGY(C,L)$: restituisce Ea che è la percentuale di energia associata all'approssimazione e Ed , il vettore corrispondente alla percentuale di energia associata al dettaglio, inerentemente ad una decomposizione wavelet del tipo $[C, L]$.
- $A = appcoef(C,L,'wname',N)$: calcola i coefficienti di approssimazione a livello N , relativi ad una struttura di decomposizione wavelet del tipo $[C,L]$.
- $D = detcoef(C,L,N)$: estrae i coefficienti di dettaglio, a livello N , inerentemente ad una wavelet con struttura $[C,L]$.
- $X = wrcoef('type',C,L,'wname',N)$: calcola il vettore di ricostruzione dei coefficienti, relativamente ad una struttura di decomposizione wavelet $[C,L]$, al livello N . “*type*” indica se vengono ricostruiti i coefficienti di approssimazione, (“*type*” =”a”), o quelli di dettaglio (“*type*” = “b”). ‘*wname*’ indica il nome della wavelet applicata.
- $THR = thselect(X,TPTR)$: individua il valore di threshold per effettuare il denoising di un segnale x . Il valore di threshold viene stabilito applicando una delle possibili metodologie, $TPTR$, riportate di seguito:

Option	Threshold Selection Rule
'rigrsure'	Selection using principle of Stein's Unbiased Risk Estimate (SURE)
'sqtwolog'	Fixed form threshold equal to $\sqrt{2 \cdot \log(\text{length}(s))}$
'heursure'	Selection using a mixture of the first two options
'minimaxi'	Selection using minimax principle

- $NC = wthcoef('t',C,L,N,T,SORH)$: restituisce i coefficienti ottenuti dalla struttura di decomposizione wavelet del tipo $[C,L]$, sottoposta a thresholding soft (se $SORH='s'$) o hard (se $SORH='h'$). N contiene i livelli di dettaglio che devono essere sottoposti a threshold e T i corrispondenti valori di soglia. N e T devono avere la stessa lunghezza.
- $X = waverec(C,L,'wname')$: ricostruisce il segnale x , sulla base di una decomposizione multilivello wavelet, con struttura $[C,L]$ e tipologia '*wname*'

Bibliografia

- Alyson Fletche, “*Denoising via recursive Wavelet Thresholding*”;
- Stefano Marsili-Libelli, *Dispense del corso di Modellistica dei Sistemi Ambientali*, Anno Accademico 2005-2006;
- Matlab Wavelet ToolBox, The MathWork.Inc;
- Robi Polikar, “*The Wavelet tutorial part I, fundamental concepts & an overview of the Wavelet theory*”, sito: <http://users.rowan.ed-robipolikar>;
- Robi Polikar, “*The Wavelet tutorial part II, fundamentals: the Fourier Transform and the short term Fourier transform*”, sito: <http://users.rowan.ed-robipolikar>;
- Robi Polikar, “*The Wavelet tutorial part III, Multiresolution Analysis & the Continuous Wavelet Transform*”, sito: <http://users.rowan.ed-robipolikar>;
- Robi Polikar, “*The Wavelet tutorial part IV, Multiresolution Analysis: the Discrete Wavelet Transform*”, sito: <http://users.rowan.ed-robipolikar>;
- Rowan University, Roby Polikar, “*The story of Wavelet*”;
- Syracuse University, Department of Electrical and Computer Engineering, T. K. Sarkar, C. Su, R. Adve, M. Salazar-Palma, L. Garcia-Castillo, Rafael R. Boix, “*A tutorial on Wavelets from an electrical engineering perspective, part I: discrete wavelet techniques*”, 1998;
- Sylvain Sardy, Paul Tseng, Andrew Bruce, “*Robust Wavelet Denoising*”, sito: <https://its4www.epfl.ch/journalclub/rwdsard.pdf>;
- The Royal Society, Guy P. Nason and Rainer von Sachs, “*Wavelets in time-series analysis*”, 1999;
- Università degli Studi di Siena, “*Seminario Wavelet*”, sito: www.dii.unisi.it;
- Università degli Studi di Genova, Simone Contradi, Roberto Pesce, “*Wavelet e Denoising delle immagini*”, sito: www.ge.infn.it;
- University of Vermont, Electrical and Computer Engineering Department, Jun Ge and Gagan Mirchandani, “*Wavelets and Denoising*”, October 10, 2003, sito: www.cs.uvm.edu;
- University of Colorado, Christopher Torrence and Gilbert P. Compo. “*A practical guide to Wavelet analysis*”;
- “*Wavelet Denoising*”, sito: <http://klapetek.cz/wdenoise.html>;
- “*Wavelet Denoising*”, sito: <http://www.math.tau.ac.il>;
- William J. Palm III, “*Matlab 6*”, McGraw-Hill, 2001;
- Sito: www.ing.unibs.it;
- Sito: www.vlsilab.polito.it;