



*Università degli studi di Firenze*  
*Facoltà di Ingegneria*

*Corso di Laurea Specialistica in Ingegneria per la  
Tutela dell'Ambiente e del Territorio*

*Anno Accademico 2005/2006*

**TRATTAMENTO DI SERIE DO E CALCOLO  
DEL COEFFICIENTE DI REAREAZIONE**

*Prof. Stefano Marsili-Libelli*

*Studenti:*

*Silvia Pannacci*

*Alessandra Stefani*

---

## INDICE

<b>1</b>	<b>PREMESSA</b> .....	<b>3</b>
<b>2</b>	<b>DENOISING</b> .....	<b>4</b>
2.1	WAVELET .....	5
2.2	SPLINES CUBICHE .....	11
<b>3</b>	<b>ANALISI DEI RESIDUI</b> .....	<b>15</b>
3.1	WAVELET .....	16
3.2	SPLINES CUBICHE .....	21
<b>4</b>	<b>CALCOLO DEL COEFFICIENTE DI REAREAZIONE</b> .....	<b>25</b>

## ***1 PREMESSA***

L'obiettivo del nostro studio è quello di modellare il rateo di ossigenazione naturale del fiume Ombrone utilizzando due serie di Ossigeno Disciolto (DO) che rappresentano la variazione circadiana del DO misurata alla stazione di Carmignano per i periodi di Luglio-Agosto e Settembre. La prima fase del nostro lavoro consiste nel *denoising* delle serie temporali di partenza, ovvero si cerca di ridurre le asperità. Per questo abbiamo utilizzato due insiemi di funzioni a base radiale, le wavelet e le spline cubiche, sfruttate per le loro proprietà approssimanti.

Abbiamo quindi confrontato i risultati ottenuti dai due metodi attraverso l'analisi dei residui (bianchezza, gaussianità e spettro di potenza).

Abbiamo infine utilizzato il segnale del DO liscio per la stima del coefficiente di reazione.

Inizialmente abbiamo derivato il segnale e quindi inserito questi dati nell'equazione di bilancio del DO. Quello che si ricava è l'andamento del coefficiente di reazione in funzione degli istanti di tempo considerati.

## 2 DENOISING

Prima dell'analisi occorre eliminare tutti i comportamenti non stazionari per cui risulta necessario detrendizzare le serie di partenza. Di seguito si riportano a confronto le serie originali e quelle detrendizzate che saranno oggetto della successiva analisi.

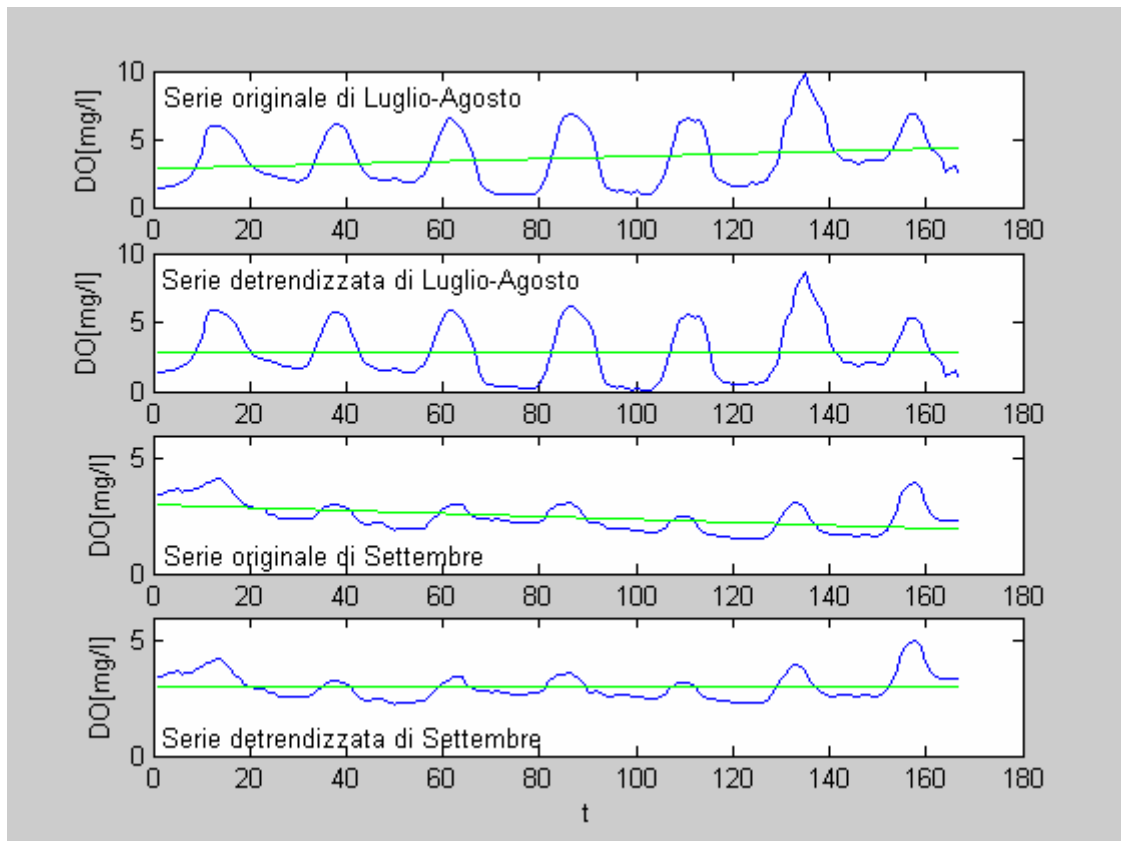


Figura 1. Serie originali e relative serie detrendizzate

## 2.1 Wavelet

I nostri valori di partenza sono una serie discreta di punti rappresentati misure dell'Ossigeno Dissolto in vari istanti di tempo. Per lo studio di queste serie temporali abbiamo cercato di eliminare il rumore senza perdere informazioni. Si è quindi effettuato un denoising mediante l'utilizzo di funzioni a base radiale: le wavelet e le spline cubiche. Mentre le spline sono utili sia per lo smoothing che per l'interpolazione dei dati mancanti, le wavelet permettono di estrarre l'informazione deterministica e consentono anche la derivazione dei segnali filtrati. Abbiamo deciso di svolgere il lavoro in parallelo confrontando ad ogni passo i risultati ottenuti con entrambi i metodi.

Le wavelet rappresentano l'estensione dell'analisi di Fourier. Questa infatti presuppone segnali di durata infinita perciò si perde tutta l'informazione relativa alla scala temporale mentre le wavelet sono funzioni a durata finita nel tempo che utilizzano una trasformazione tempo→scala. In tal modo si può combinare l'analisi nel tempo e nella frequenza attraverso una tecnica di finestre adattative. Ciò permette l'analisi locale di eventi di breve durata senza perdere la potenza dell'analisi in frequenza.

Data una generica wavelet  $\psi(a,b,t)$ , dove  $a$  e  $b$  rappresentano i fattori di scaling e shifting, la trasformata wavelet continua (CWT) è definita come l'integrale del segnale  $s(t)$  moltiplicato per la wavelet scalata:

$$C(a,b) = \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} s(t) \psi\left(\frac{t-b}{a}\right) dt$$

Se la wavelet viene dilatata ( $a \gg 1$ ) contiene prevalentemente basse frequenze e fornisce un'approssimazione globale del segnale. Se invece è compressa ( $a \ll 1$ ) contiene prevalentemente alte frequenze e dà un'immagine dettagliata di una piccola porzione del segnale  $s(t)$ .

Effettuando uno scaling diadico (ad ogni passo dimezza l'ampiezza della wavelet) in funzione della finestra di base di durata  $T_s$ :

$$\left( a^{T_s/2}, a^{T_s/4}, \dots, a^{T_s/2n} \right)$$

si ha la trasformata wavelet discreta (DWT) basata su potenze del 2 della finestra di base  $T_s$ . In tal modo si ha una decomposizione gerarchica del segnale in termini di:

- Approssimazioni: contengono la parte a bassa frequenza del segnale;
- Dettagli: contengono la parte ad alta frequenza del segnale.

Procedendo per dimezzamenti successivi della scala si ottiene una decomposizione multivivello (MD) del segnale in successivi dettagli ed approssimazioni con un dimezzamento dei dati ad ogni dimezzamento della finestra.

Le approssimazioni effettuano uno smoothing del segnale proporzionale al grado dell'approssimazione.

Abbiamo deciso di decomporre le nostre serie con approssimazioni A1, A2 e A3. Dato che il tempo di campionamento delle serie di partenza è di 1 ora, l'approssimazione A1 è basata sulle differenze a 2 ore, la A2 a 4 ore e la A3 a 8 ore. Si riportano di seguito i grafici rappresentanti le approssimazioni e relativi dettagli.

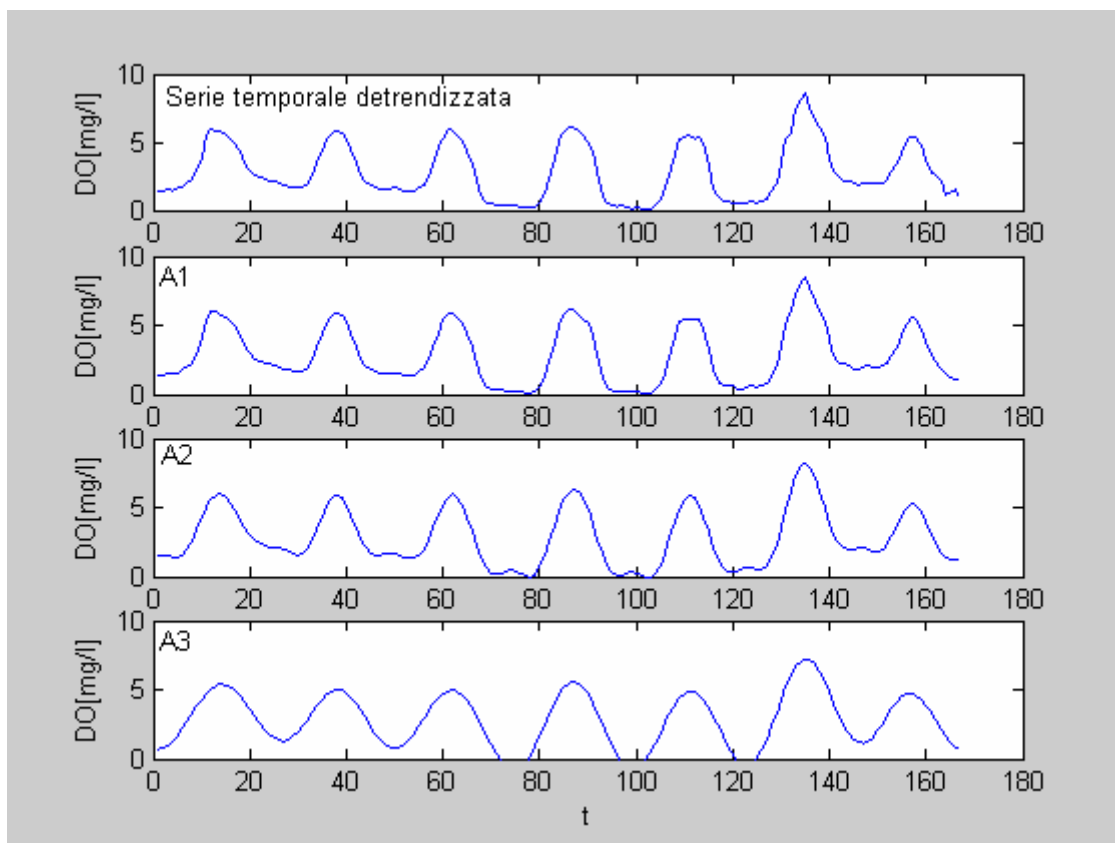


Figura 2. Approssimazione wavelet per la serie di Luglio-Agosto

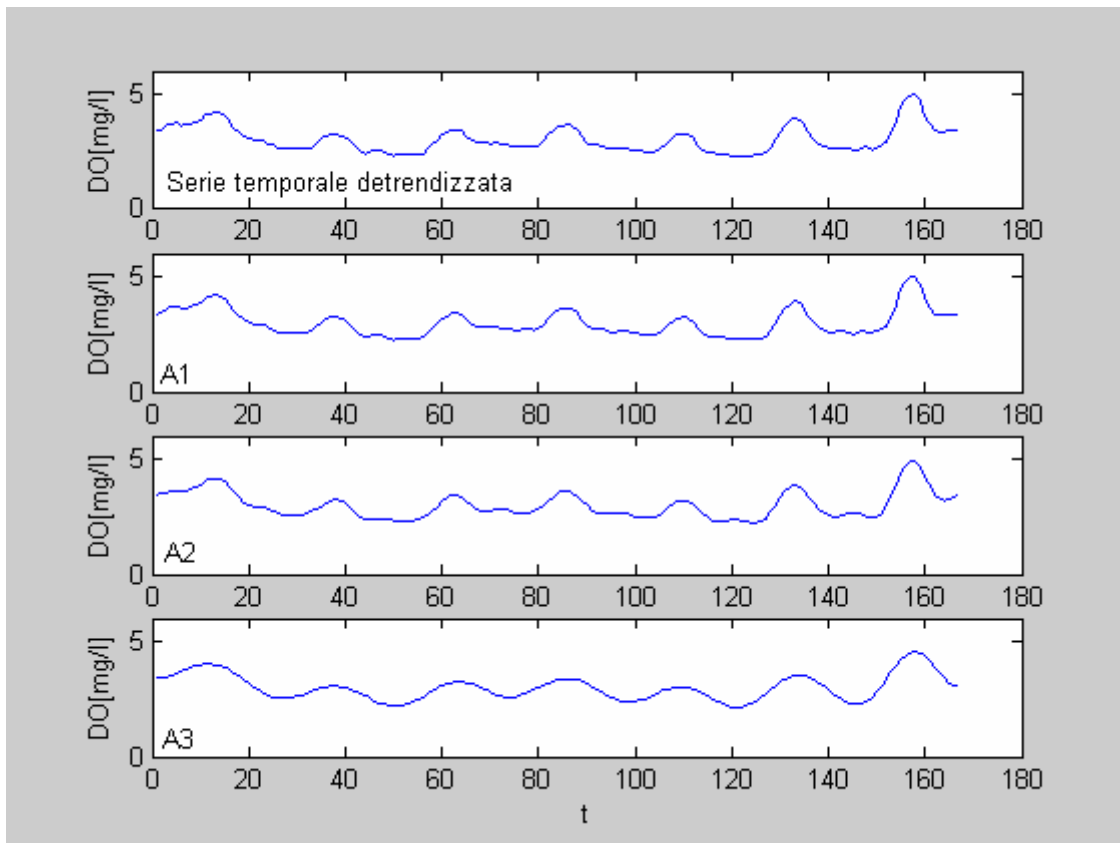


Figura 3. Approssimazioni wavelet per la serie di Settembre

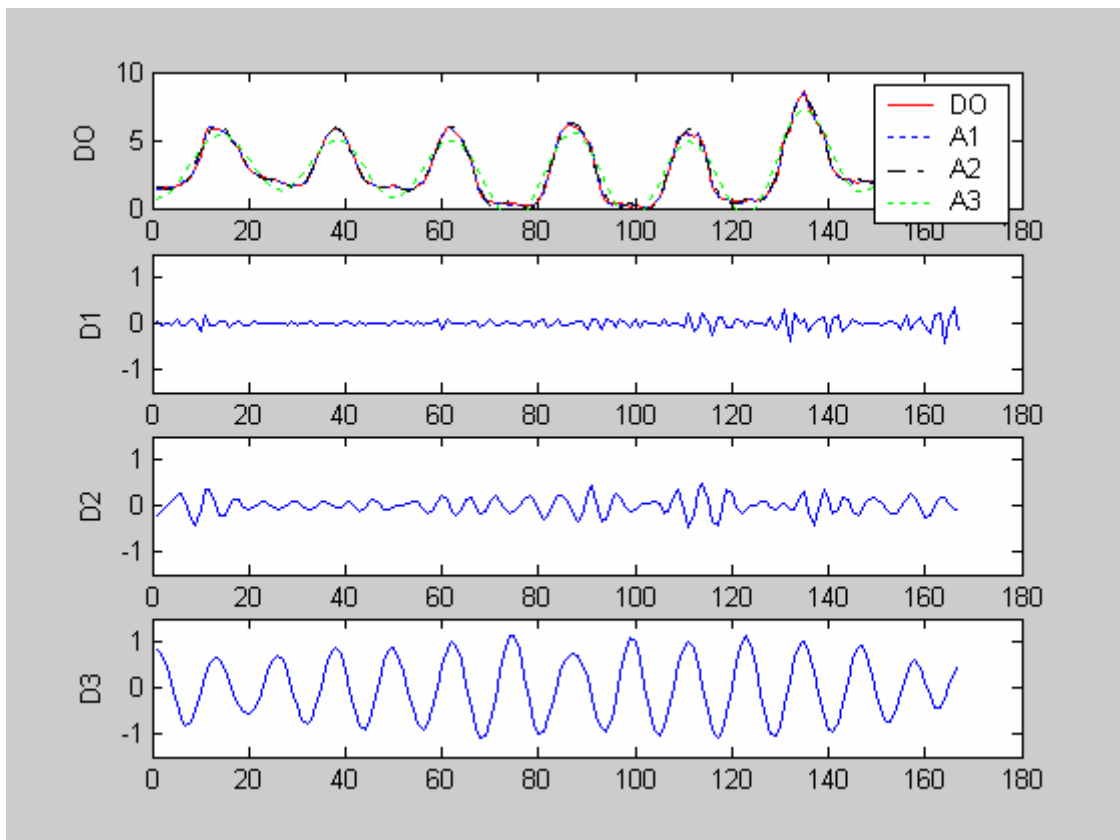


Figura 4. Approssimazioni wavelet e dettagli per la serie di Luglio-Agosto

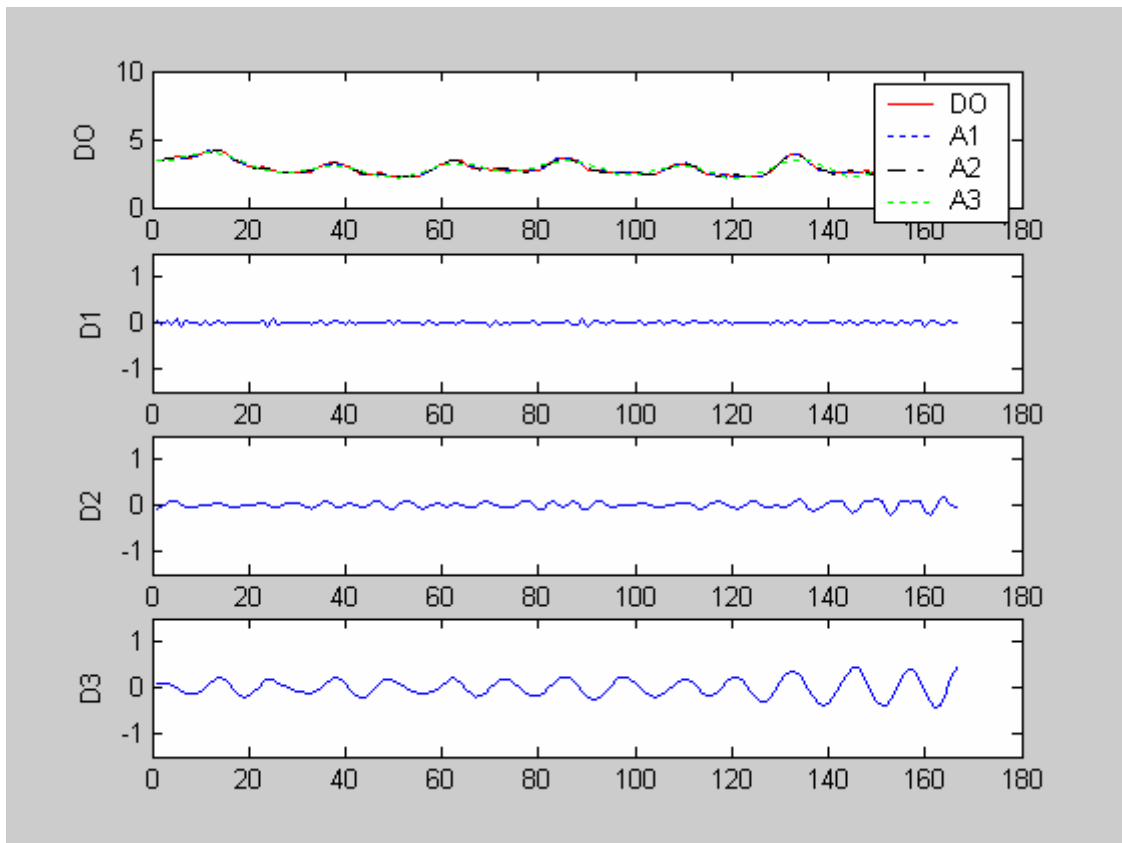


Figura 5. Approssimazioni wavelet e dettagli per la serie di Settembre

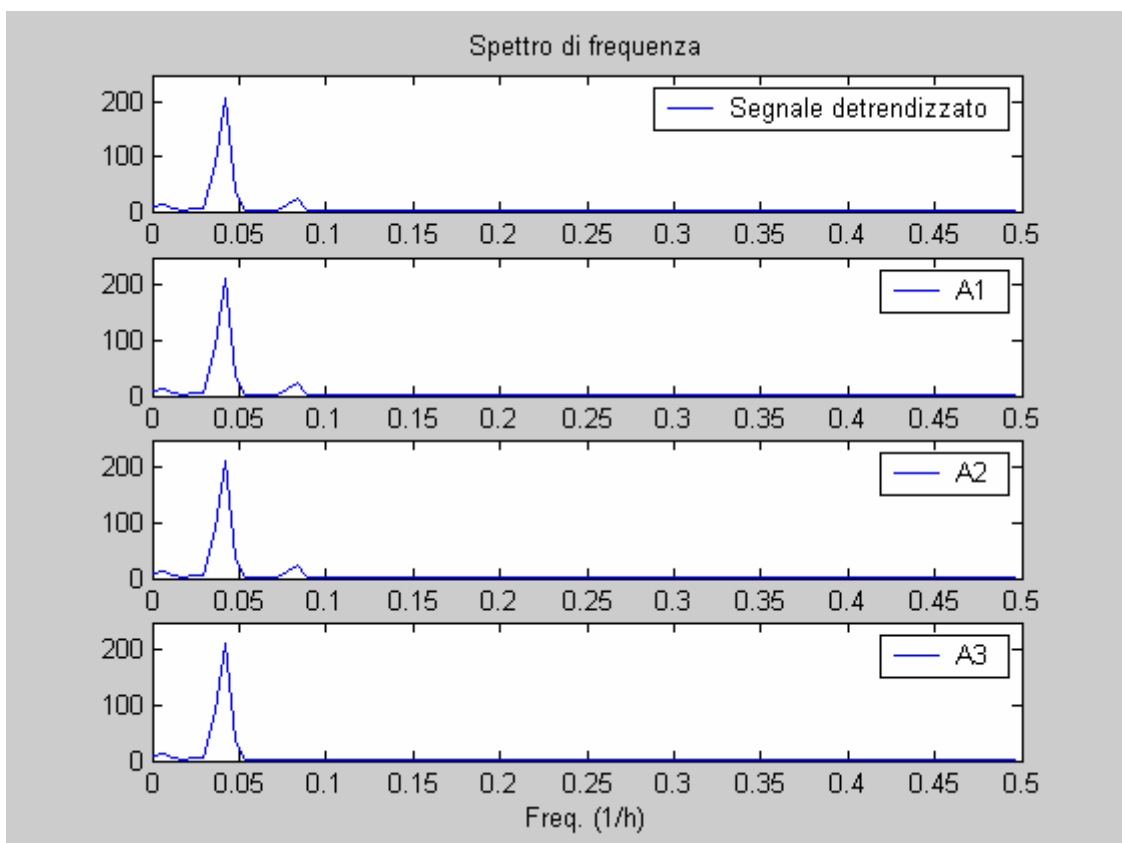


Figura 6. Spettro di frequenza della serie detrendizzata di Luglio-Agosto

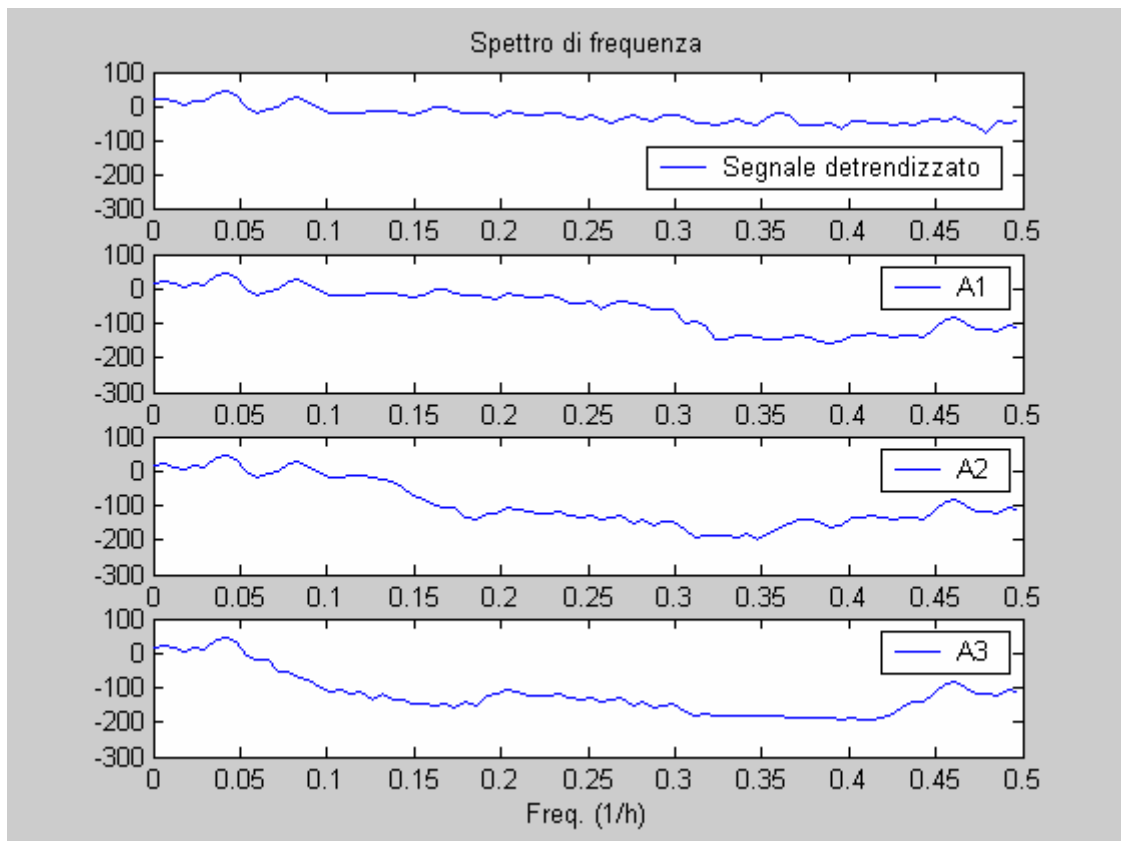


Figura 7. Spettro di frequenza della serie detrendizzata di Luglio-Agosto in decibel

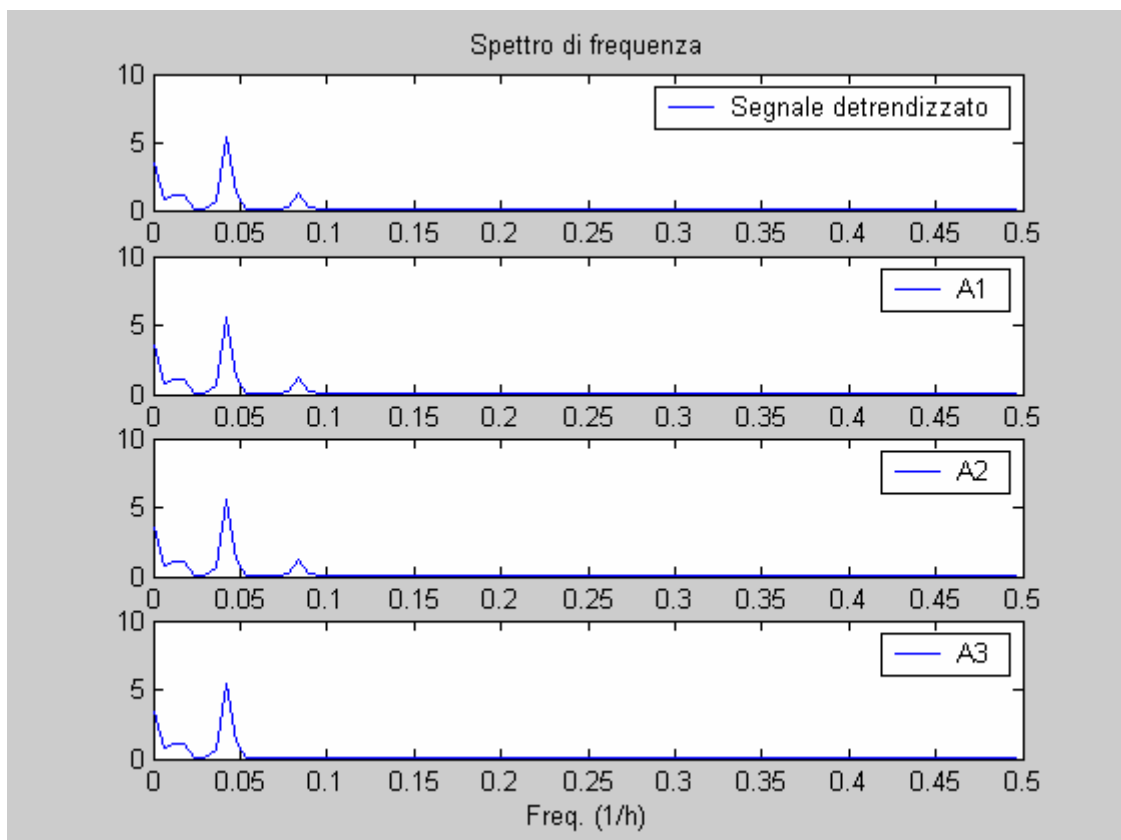
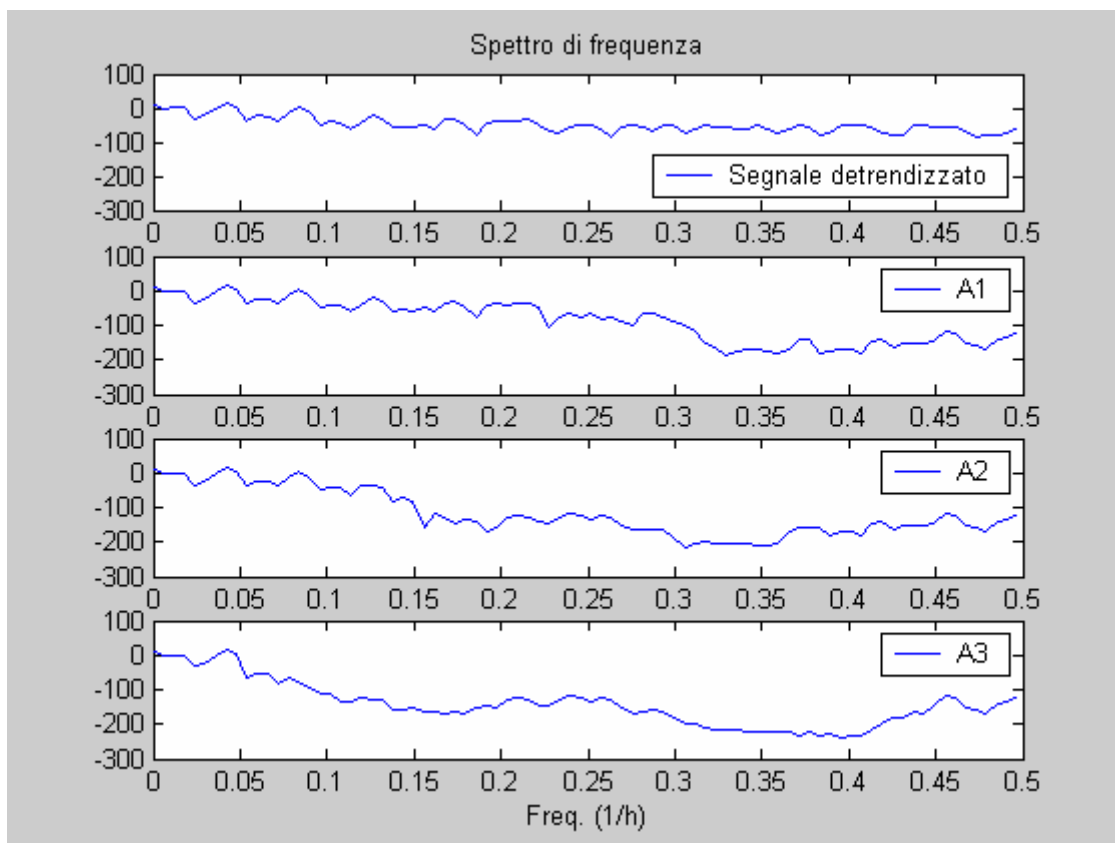


Figura 8. Spettro di frequenza della serie detrendizzata di Settembre



**Figura 9. Spettro di frequenza della serie detrendizzata di Settembre in decibel**

Dall'analisi dei risultati si osserva come l'approssimazione di ordine 3 porti ad un liscio eccessivo della serie di partenza con conseguente perdita di informazioni. Dall'analisi degli spettri di frequenza (riportati anche in decibel) delle approssimazioni a livello 1 e 2 si osserva come in entrambi i casi l'informazione mantenuta è circa la stessa. Quindi, per ridurre il costo computazionale, scegliamo l'approssimazione a livello 1.

## 2.2 Splines cubiche

Le funzioni splines rappresentano una tecnica semplice per il denoising e l'interpolazione dei dati mancanti. Il loro pregio è la continuità delle derivate e permettono l'approssimazione dei dati  $(x,y)$  minimizzando la seguente funzione obiettivo:

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

dove  $R$  rappresenta la scabrezza dei dati.

Il parametro di smoothing  $p$  determina l'importanza relativa fra le due esigenze contrastanti dello smoothing e della fedeltà ai dati.

- Per  $p=0$   $ys$  è la retta ai minimi quadrati;
- Per  $p=1$   $ys$  è l'interpolante naturale ai dati.

La scelta di  $p$  è fortemente dipendente dalla natura dei dati ed è determinabile dall'analisi in frequenza. Infatti nell'analisi di frequenza lo spettro della parte informativa del segnale (deterministico), degli artefatti (deterministici) e del rumore (stocastico) sono ben separati. Lo smoothing consiste nel filtraggio del rumore e degli artefatti.

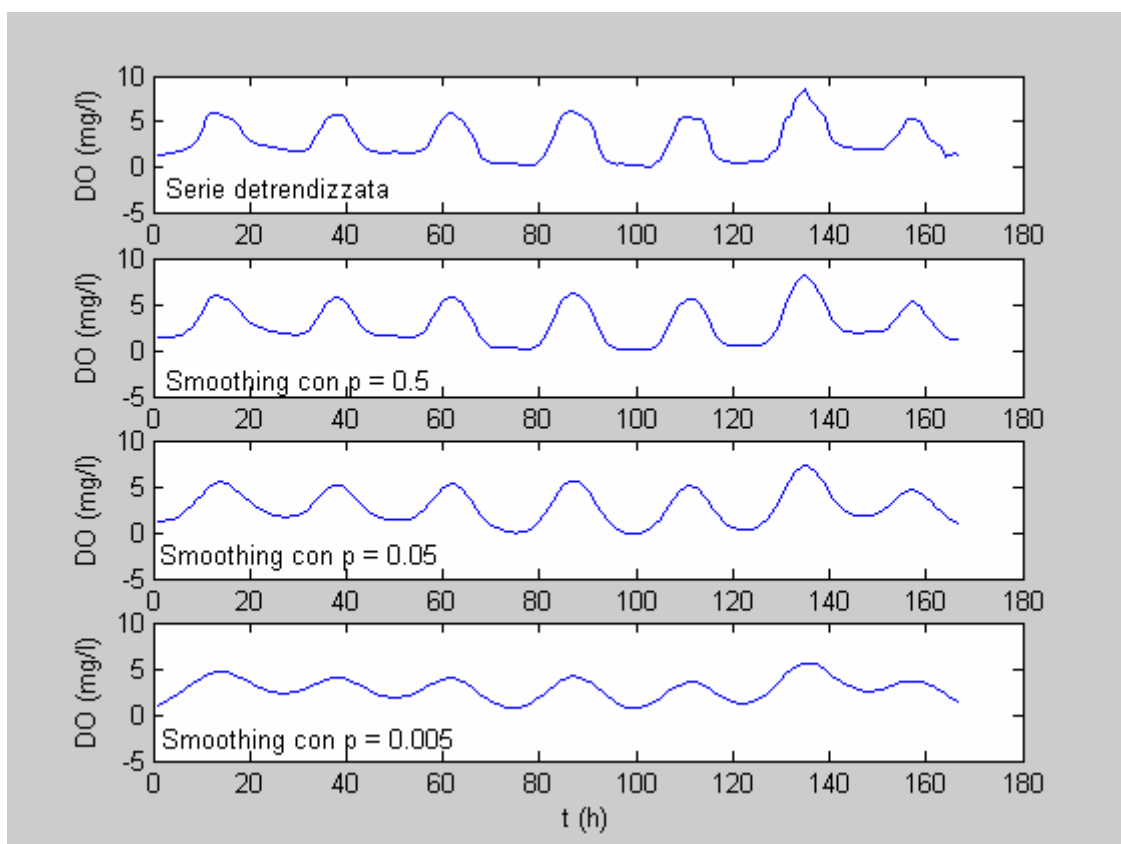


Figura 10. Approssimazione spline per la serie di Luglio-Agosto

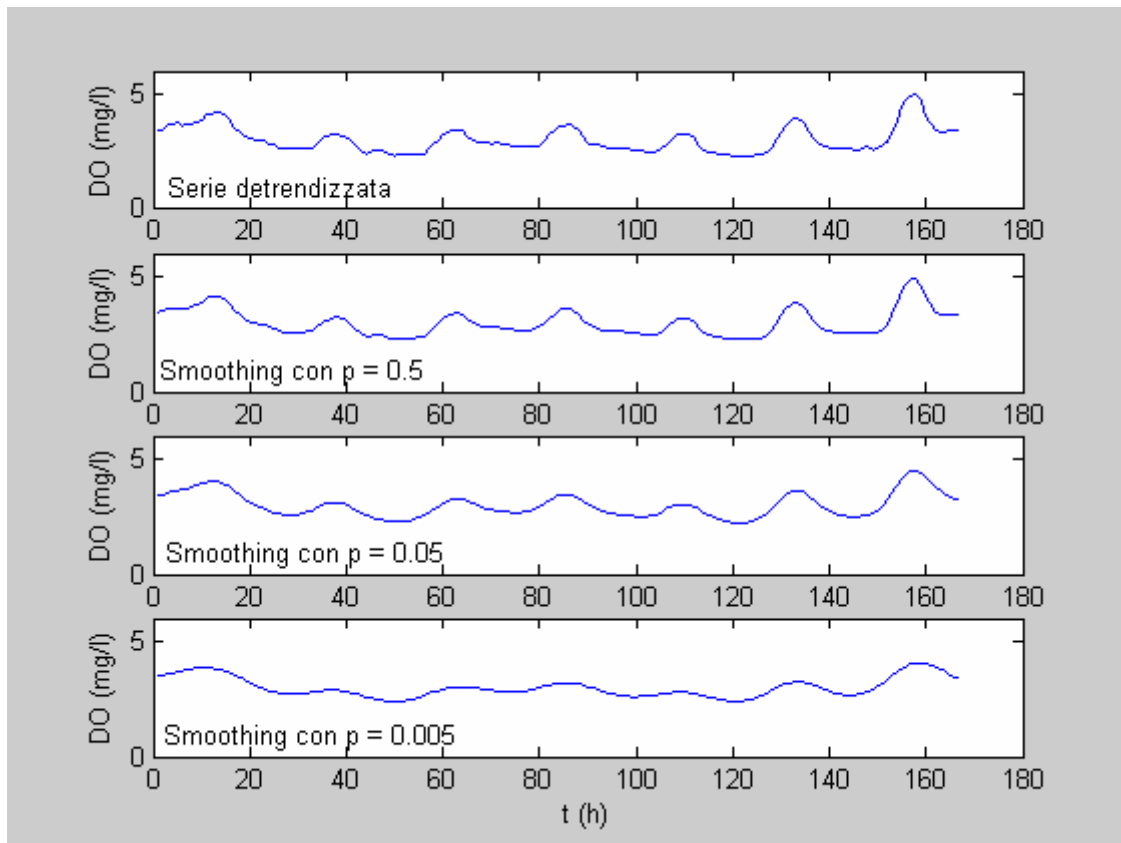


Figura 11. Approssimazioni spline della serie di Settembre

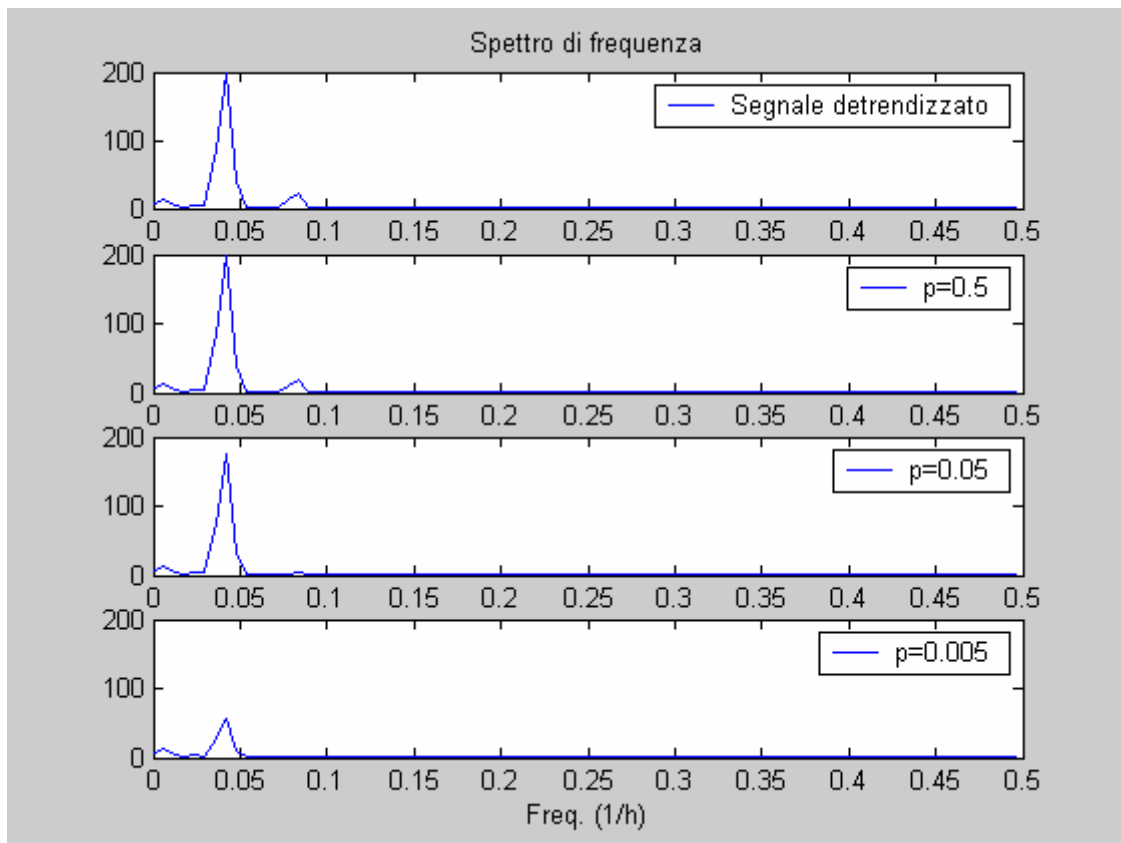


Figura 12. Spettro di frequenza della serie di Luglio-Agosto

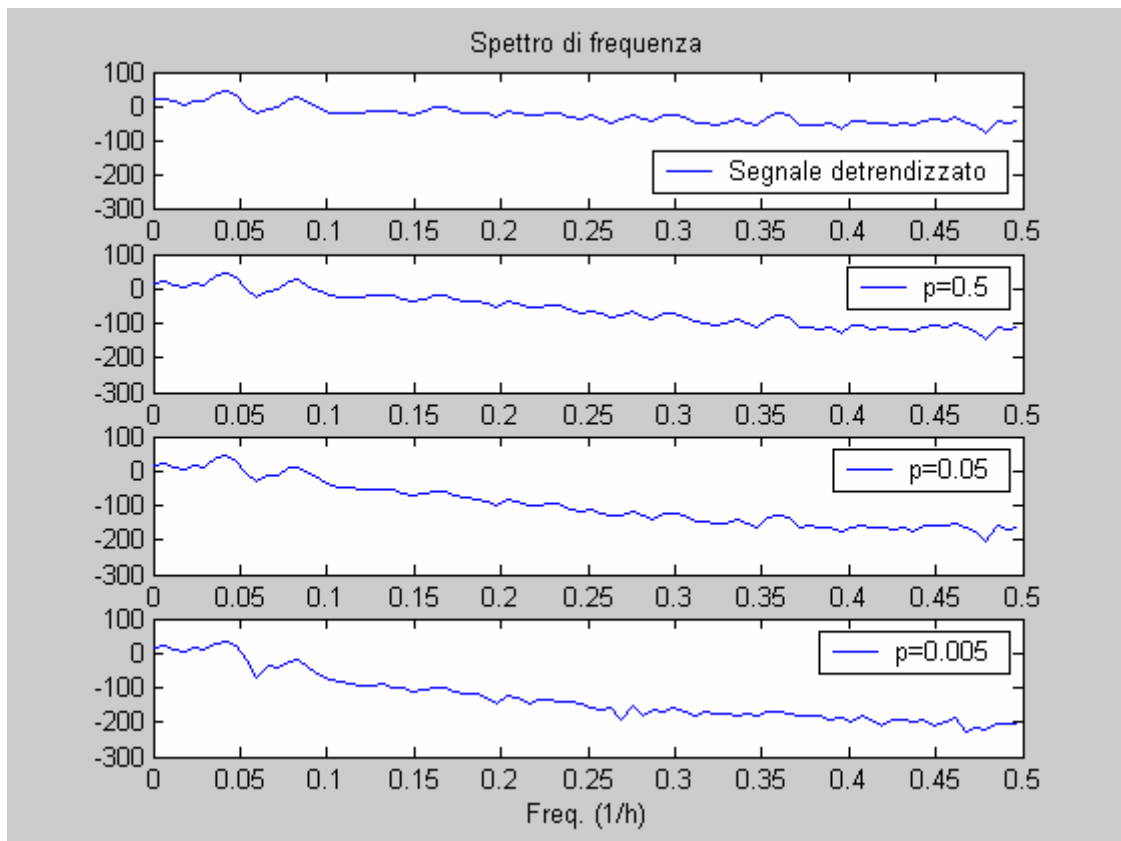


Figura 13. Spettro di frequenza della serie di Luglio-Agosto in decibel

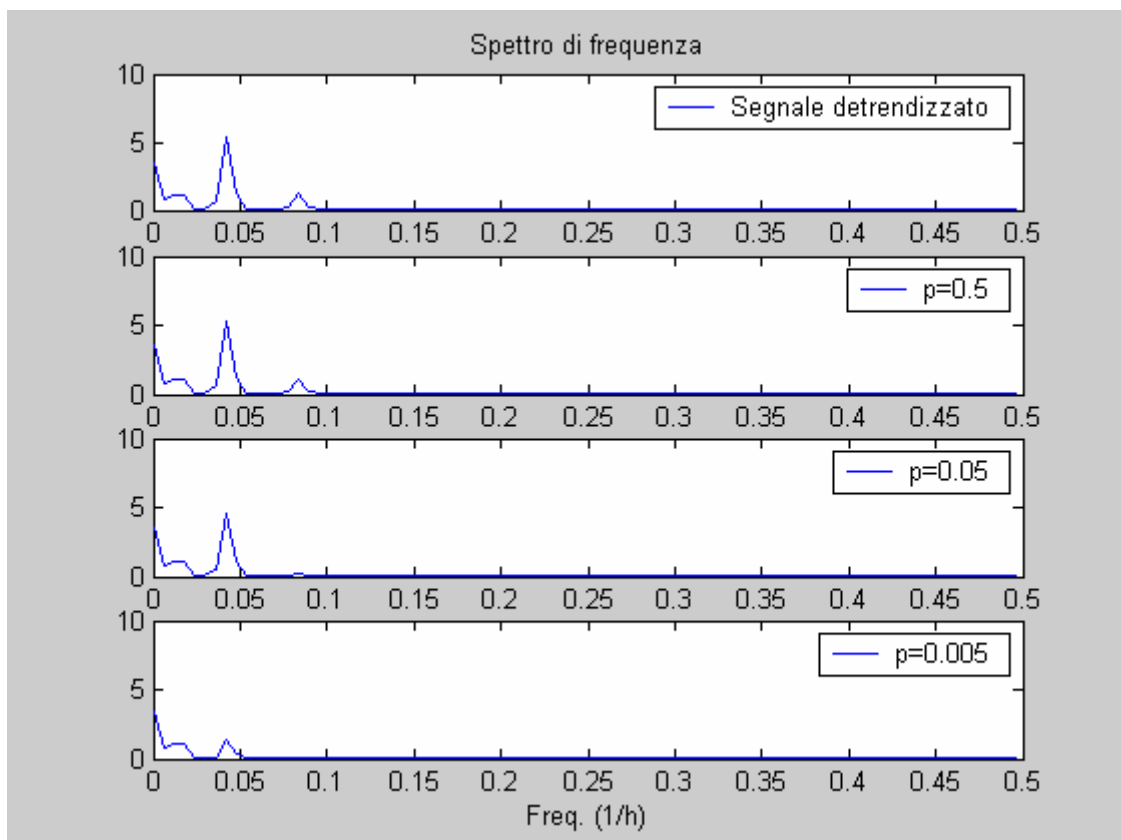
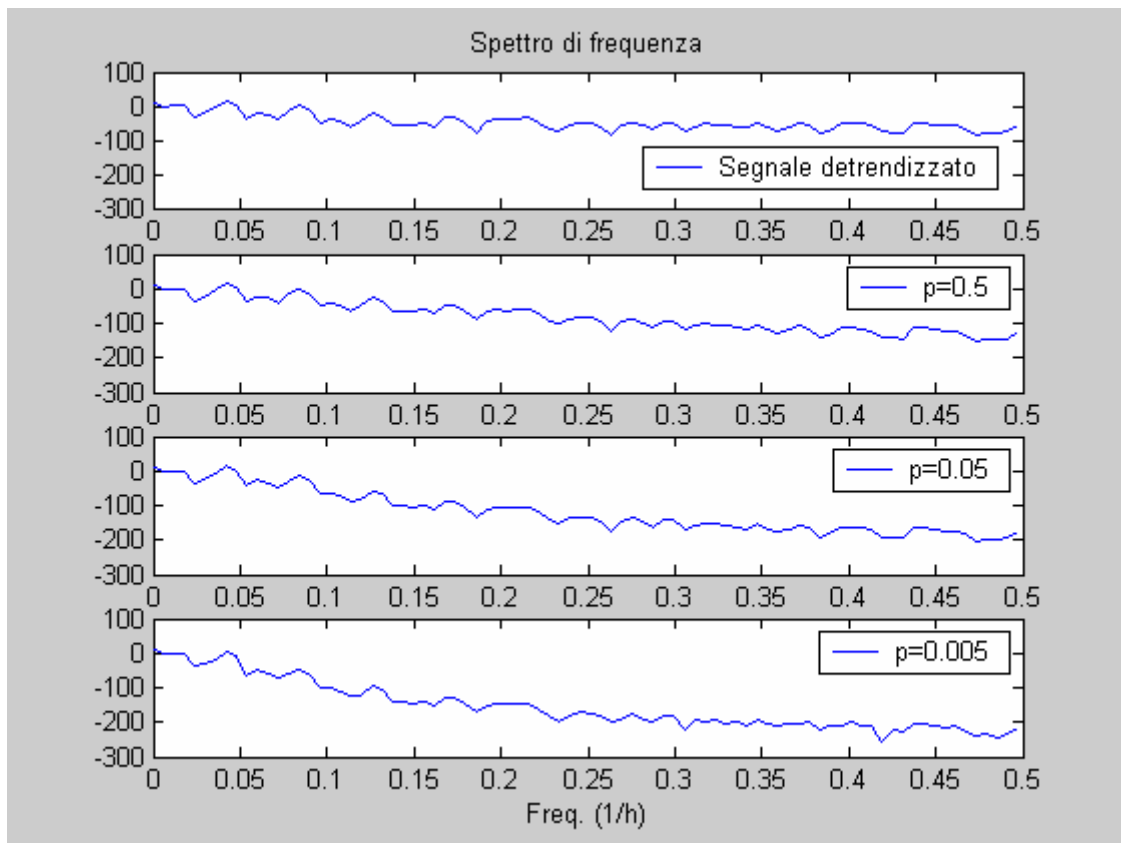


Figura 14. Spettro di ferquenza della serie di Settembre



**Figura 15. Spettro di frequenza della serie di Settembre in decibel**

Dall'analisi delle approssimazioni si ritiene che la spline relativa a  $p=0.5$  sia quella che meglio rappresenta l'andamento dell'Ossigeno Disciolto. Infatti l'analisi di frequenza mostra che per tale valore si mantiene sia il picco alle 24 ore caratteristico del ciclo circadiano che quello alle 12 ore mantenendoci vicini ai dati originali.

### 3 ANALISI DEI RESIDUI

L'analisi dei residui consiste nel confronto tra il segnale originario e quello approssimato (sia con le wavelet che con le spline) attraverso la loro differenza (residui). La prima analisi condotta sui residui è il test di bianchezza sulla loro correlazione. Il valore di autocorrelazione  $\rho$  deve essere contenuto nella fascia di zero statistico per poter dire che i residui non sono correlati:

$$\rho(k) < \frac{1.96}{\sqrt{N}} \quad \forall k > 1$$

L'altra analisi condotta sui residui è quella statistica per mezzo della quale se ne verifica la gaussianità. Abbiamo prima paragonato la distribuzione approssimata dell'istogramma con quella gaussiana e abbiamo poi cercato conferma di questo primo risultato mediante il test di Kolmogoroff- Smirnov. L'istogramma deriva dalla distribuzione dei residui, mentre quella gaussiana è calcolata a partire della media e dalla varianza dei residui stessi. L'integrale delle due distribuzioni dà la statistica cumulata che viene poi utilizzata nel test di Kolmogoroff- Smirnov.

La gaussianità dei residui è una caratteristica necessaria per la bontà dell'approssimazione, mentre la bianchezza non è una proprietà determinante.

### 3.1 Wavelet

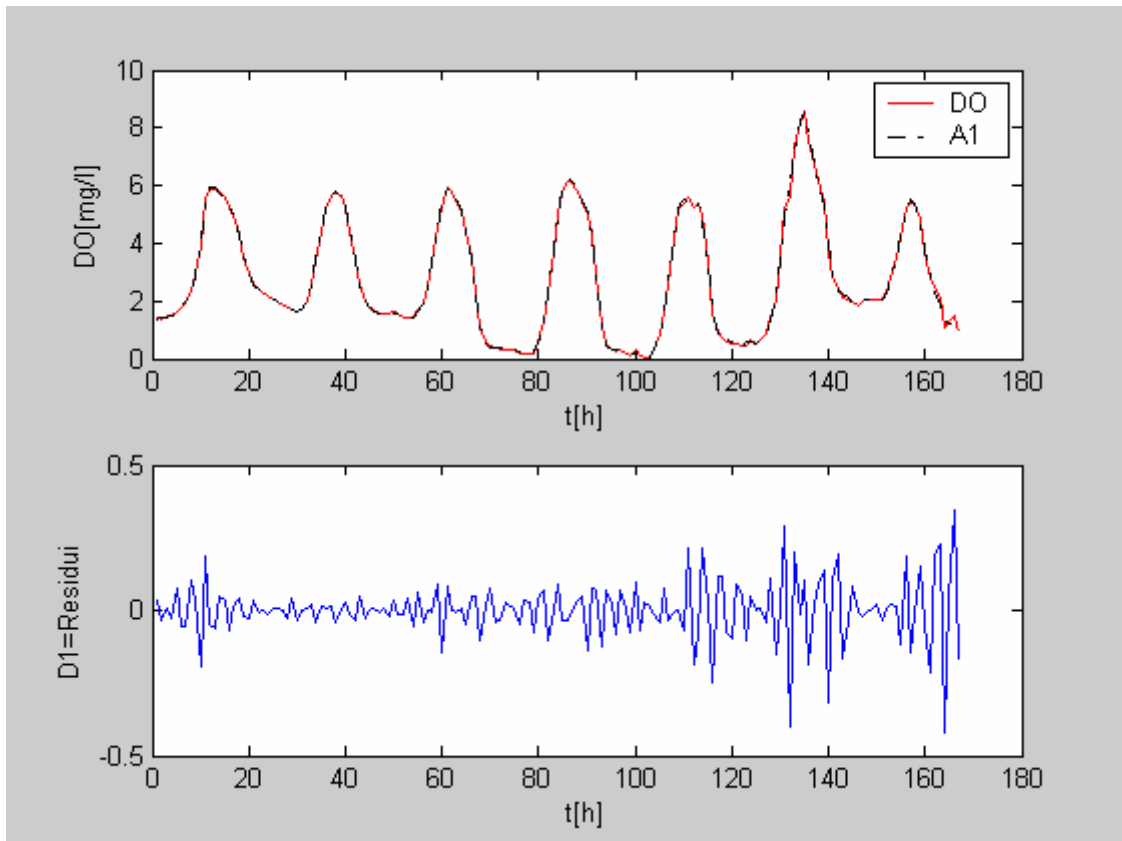


Figura 16. Residui wavelet A1 per la serie di Luglio-Agosto

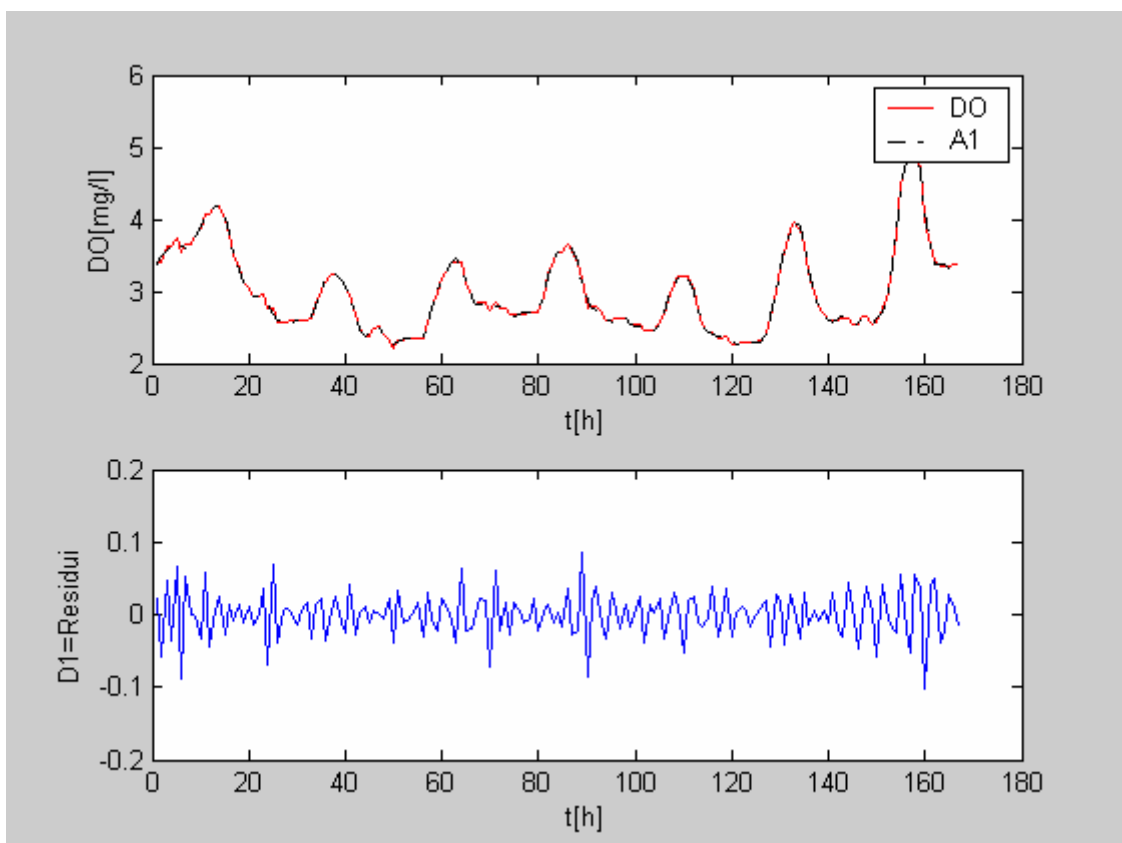


Figura 17. Residui wavelet A1 per la serie di Settembre

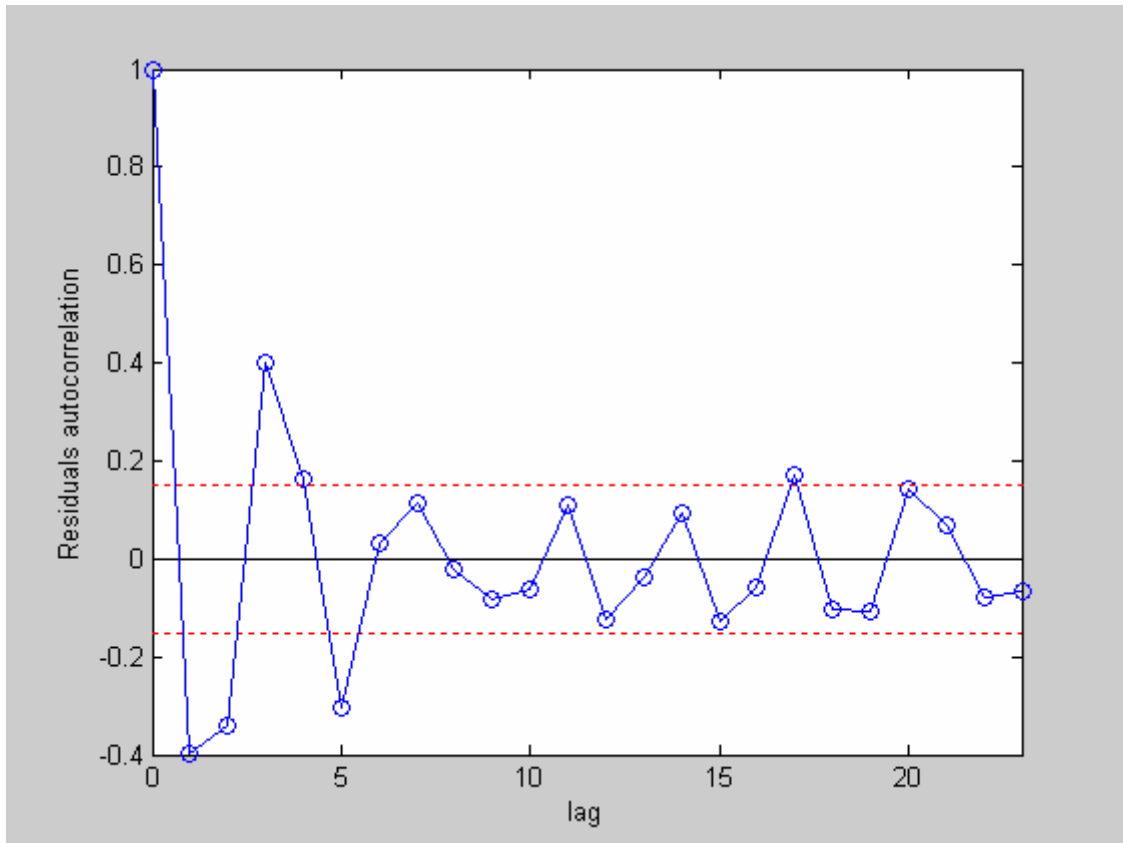


Figura 18. Autocorrelazione dei residui wavelet A1 della serie di Luglio-Agosto

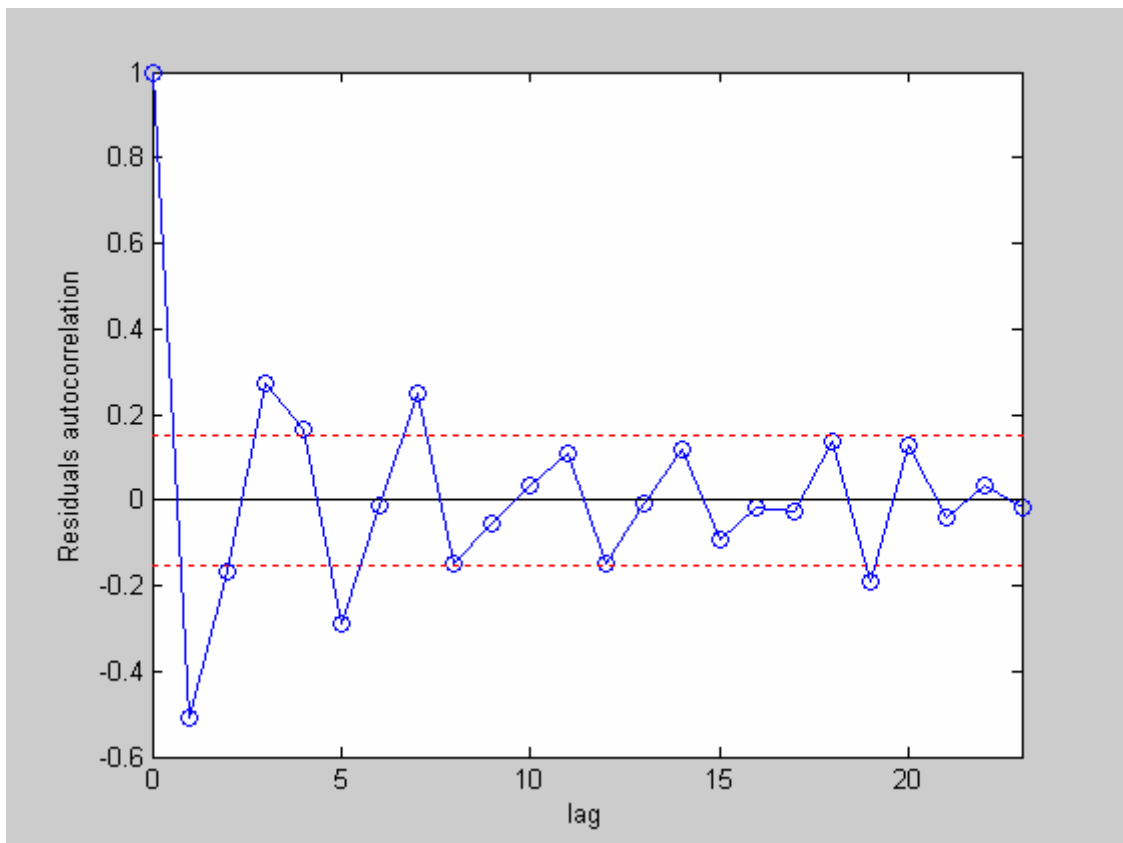
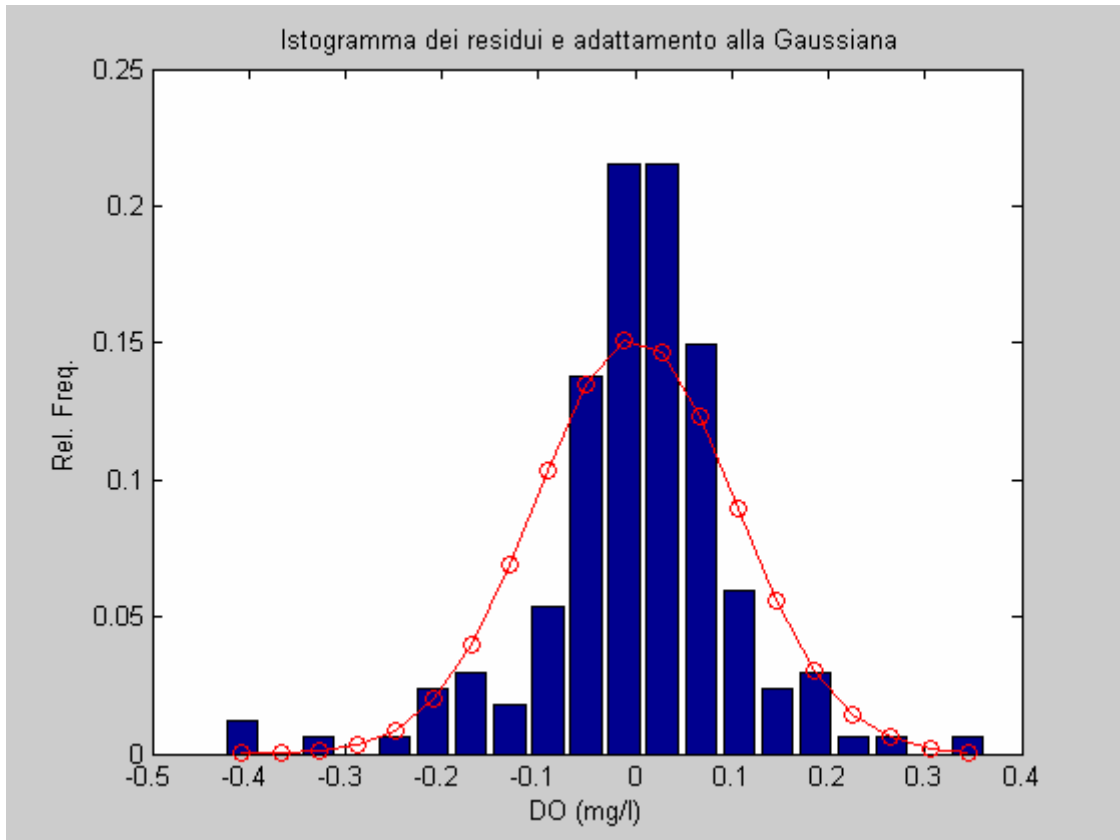


Figura 19. Autocorrelazione dei residui wavelet A1 della serie di Settembre

Dalle figure sopra riportate si osserva che i residui risultano autocorrelati per lag bassi, mentre superata una soglia iniziale rimangono tutti entro la fascia di zero statistico. Si può quindi affermare in buona approssimazione che i dati sono bianchi.



**Figura 20. Istogramma dei residui e andamento Gaussiano per la serie di Luglio-Agosto, approssimazione wavelet A1**

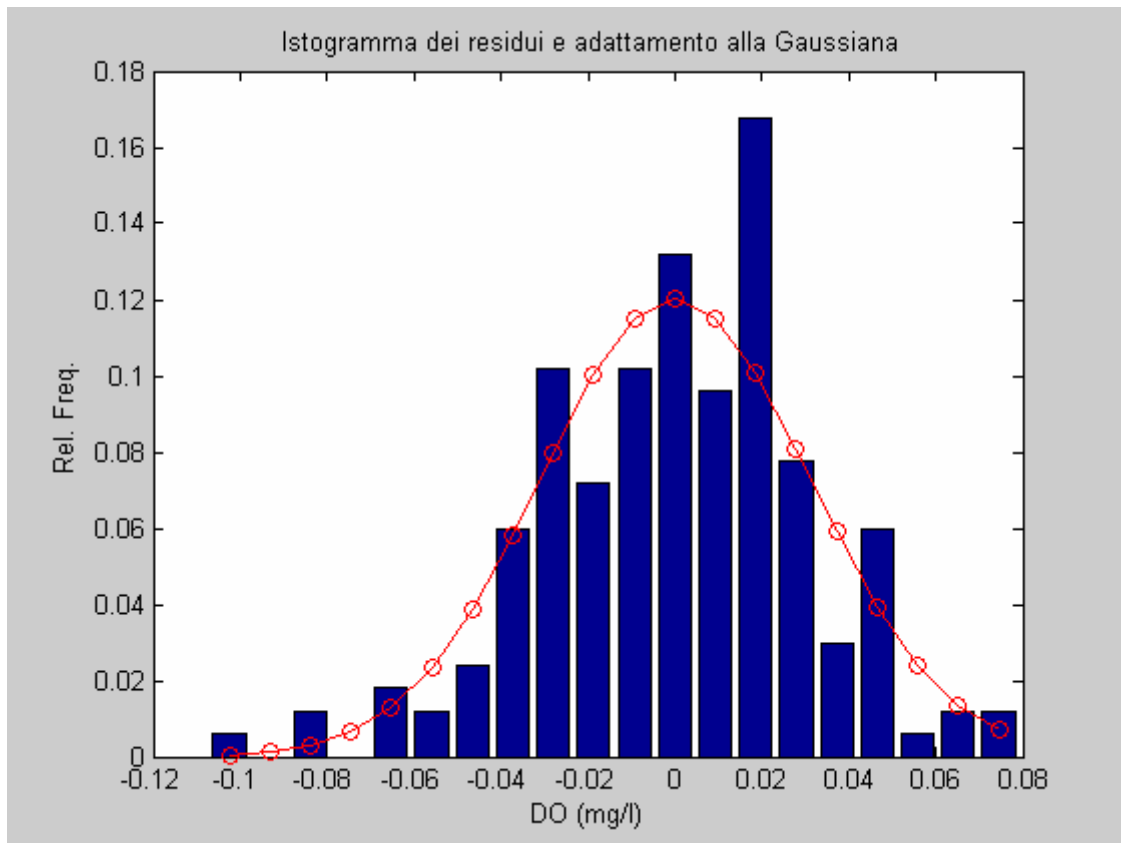


Figura 21. Istogramma dei residui e andamento Gaussiano per la serie di Settembre, approssimazione wavelet A1

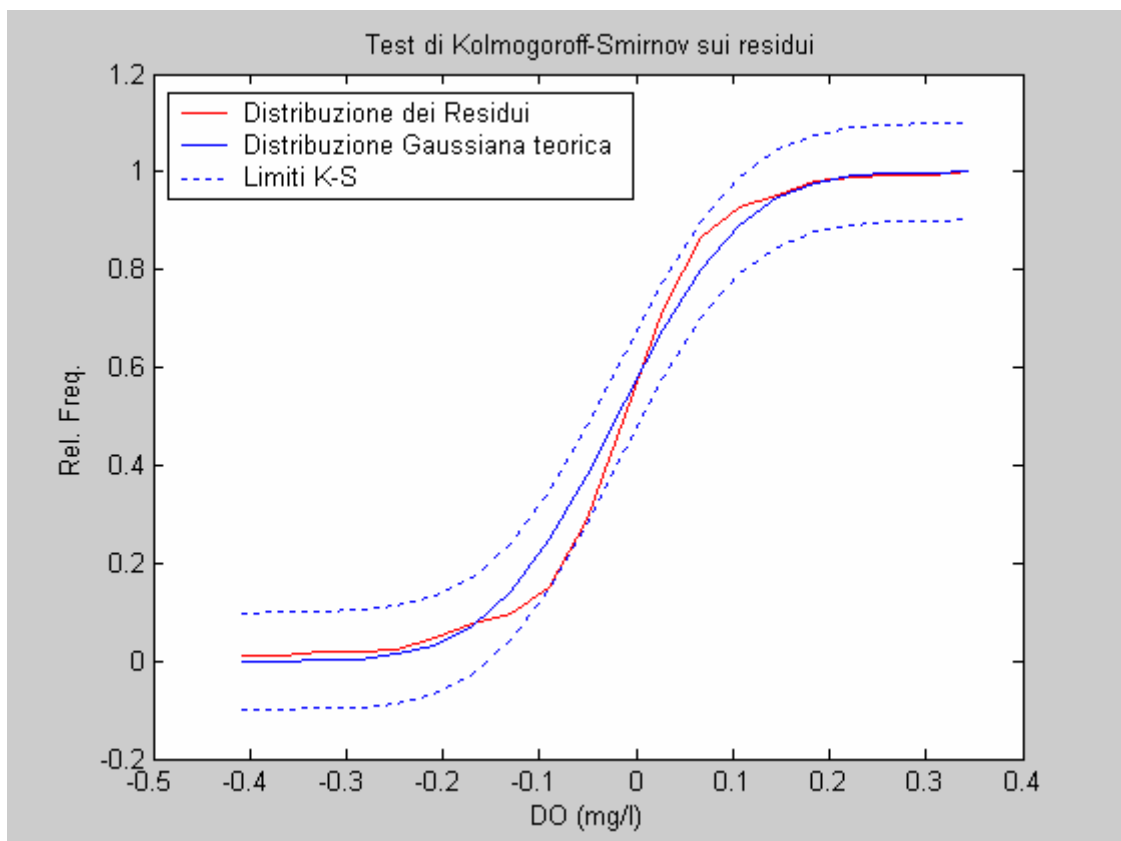
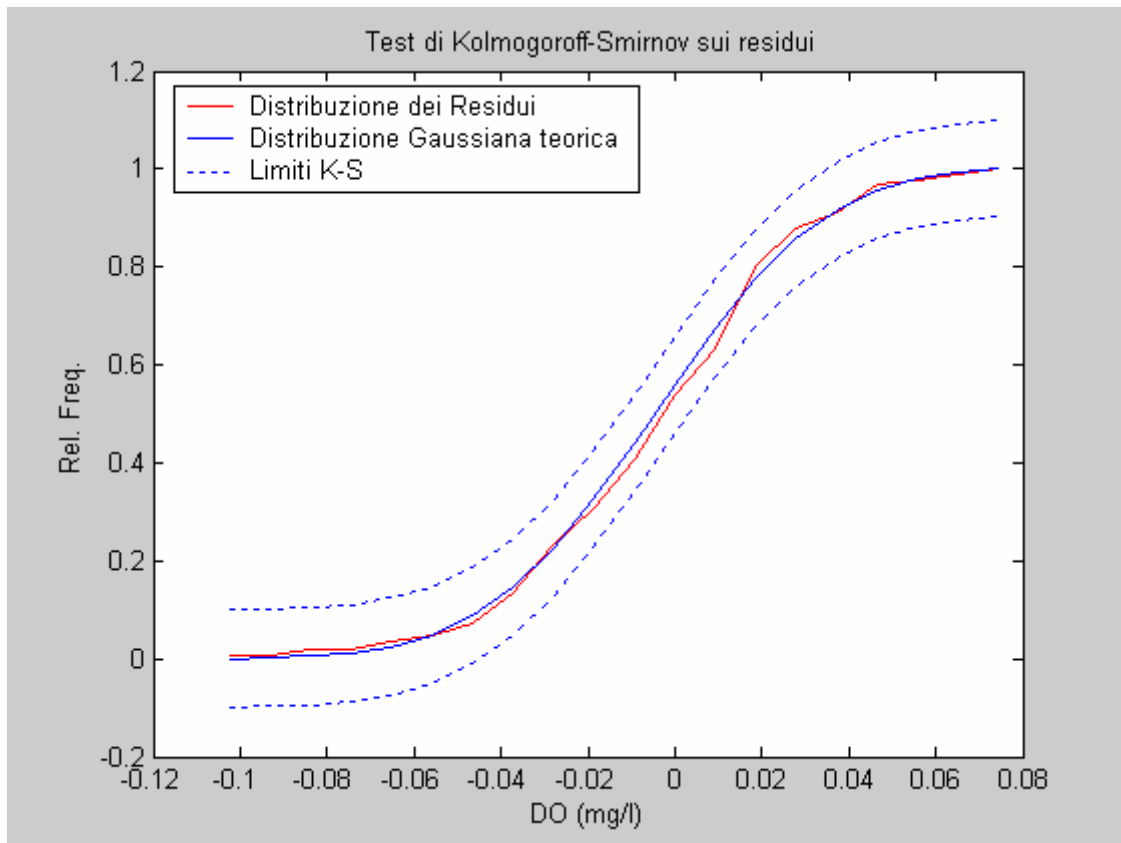


Figura 22. Test Kolmogoroff-Smirnov per la serie di Luglio-Agosto, approssimazione wavelet A1



**Figura 23. Test Kolmogoroff-Smirnov per la serie di Settembre, approssimazione wavelet A1**

Il confronto tra l'istogramma dei residui e l'andamento Gaussiano mostra come i residui possano essere considerati gaussiani; il risultato è confermato anche dal test di Kolmogoroff-Smirnov poiché la distribuzione dei residui rientra all'interno delle bande di confidenza.

Si può quindi concludere che i residui dell'approssimazione wavelet A1 possono ragionevolmente essere considerati Gaussiani e incorrelati.

### 3.2 Splines cubiche

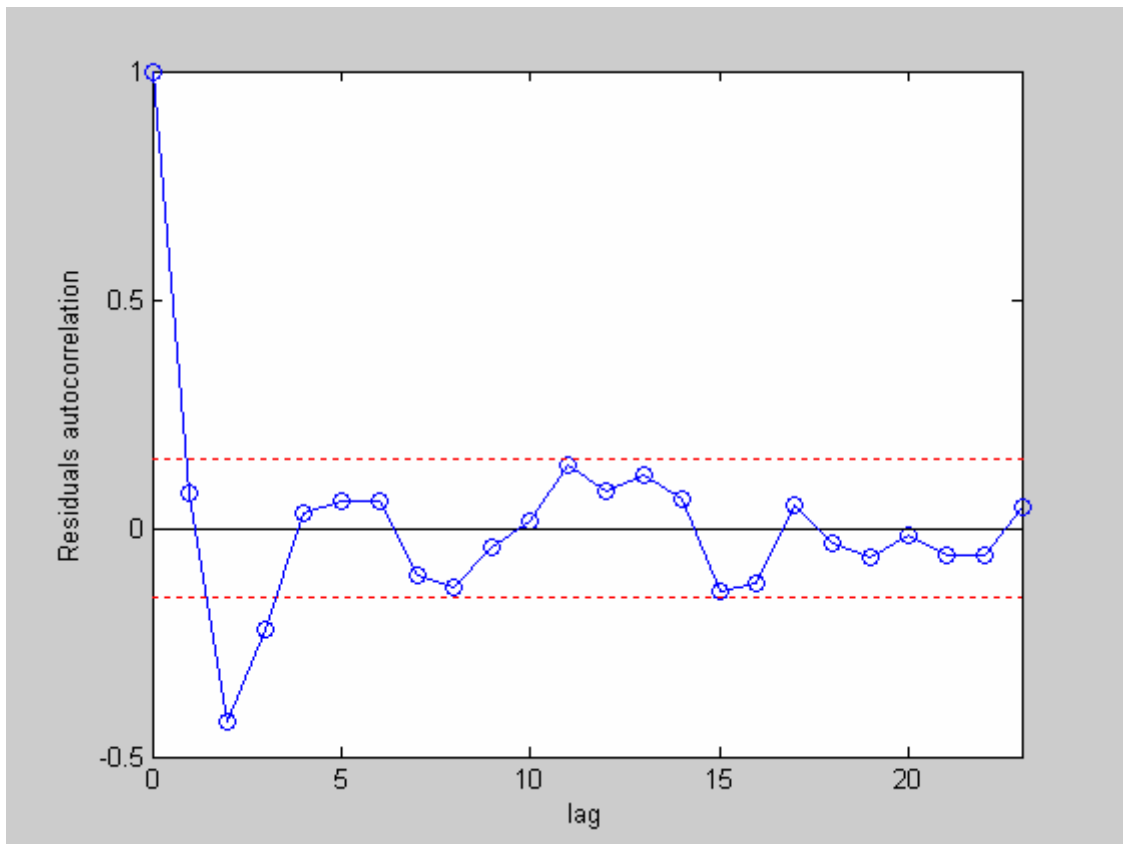


Figura 24. Autocorrelazione dei residui spline ( $p=0.5$ ) della serie di Luglio-Agosto

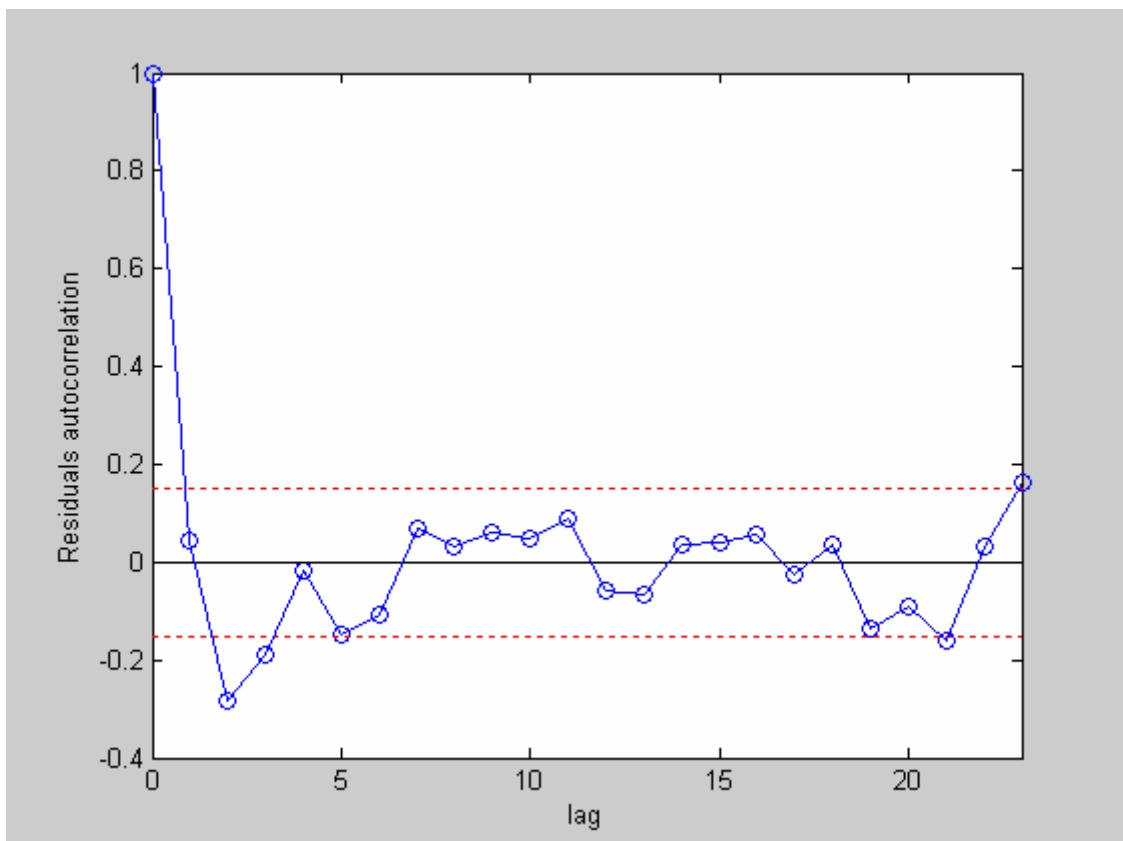


Figura 25. Autocorrelazione dei residui spline ( $p=0.5$ ) della serie di Settembre

Dalle figure sopra riportate si osserva che i residui risultano autocorrelati per lag bassi, mentre superata una soglia iniziale rimangono tutti entro la fascia di zero statistico. Si può quindi affermare in buona approssimazione che i dati sono bianchi.

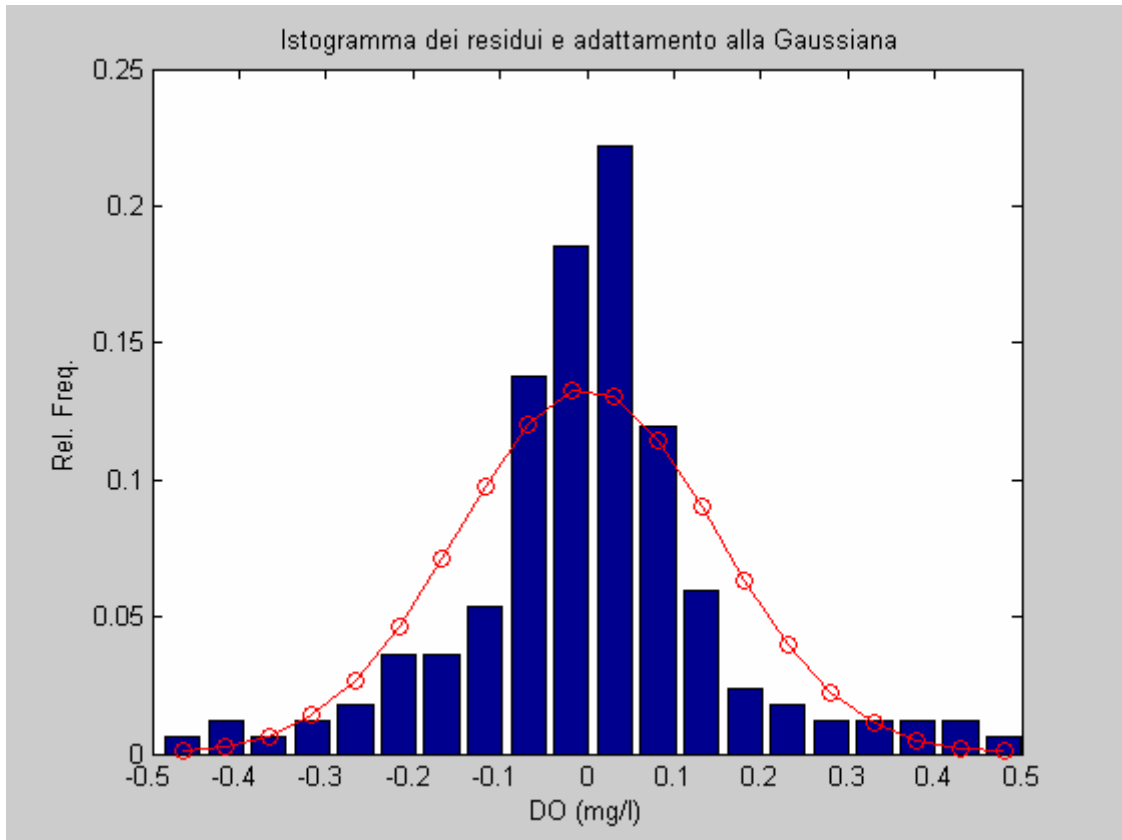


Figura 26. Istogramma dei residui e andamento Gaussiano per la serie di Luglio-Agosto, approssimazione spline  $p=0.5$

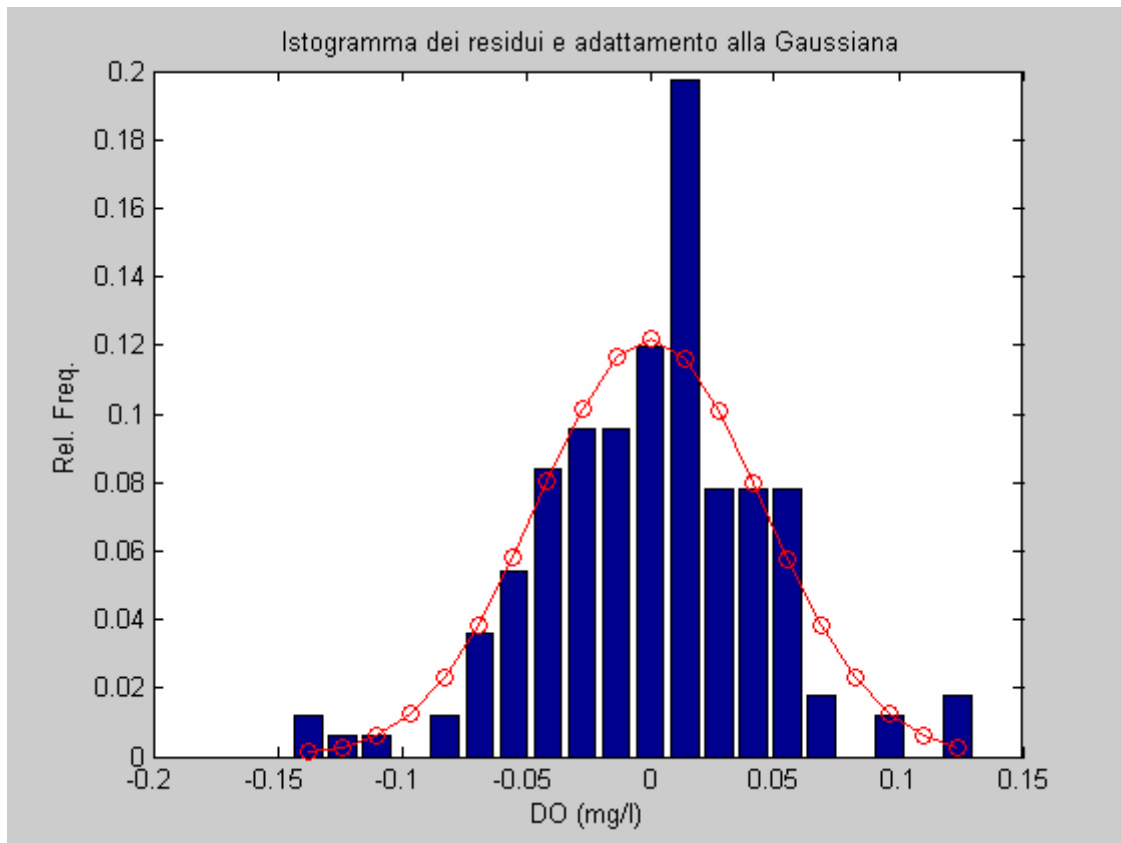


Figura 27. Istogramma dei residui e andamento Gaussiano per la serie di Settembre, approssimazione spline  $p=0.5$

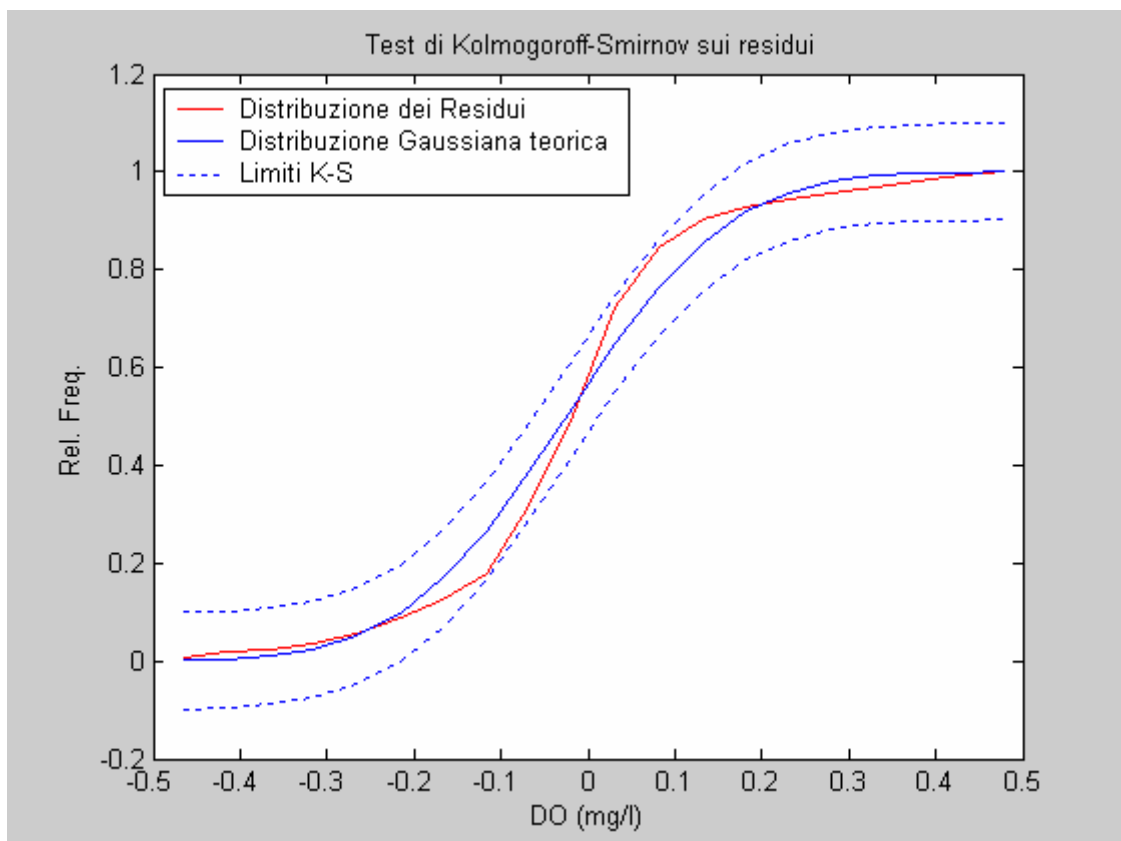
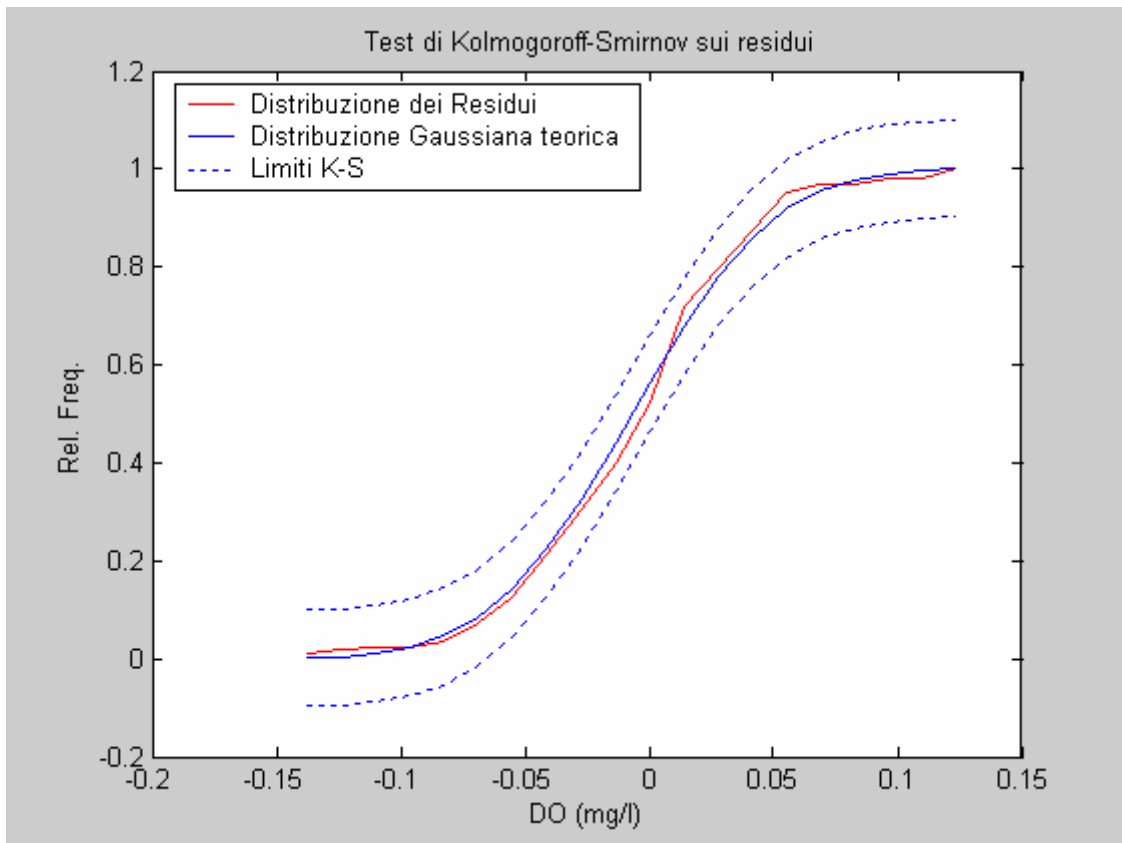


Figura 28. Test Kolmogoroff-Smirnov per la serie di Luglio-Agosto, approssimazione spline  $p=0.5$



**Figura 29.** Test Kolmogoroff-Smirnov per la serie di Settembre, approssimazione spline  $p=0.5$

Il confronto tra l'istogramma dei residui e l'andamento Gaussiano mostra come i residui possano essere considerati gaussiani; il risultato è confermato anche dal test di Kolmogoroff-Smirnov poiché la distribuzione dei residui rientra all'interno delle bande di confidenza.

Si può quindi concludere che i residui dell'approssimazione spline con  $p=0.5$  possono ragionevolmente essere considerati Gaussiani e incorrelati.

## 4 CALCOLO DEL COEFFICIENTE DI REAREAZIONE

La stima dello scambio di Ossigeno tra atmosfera ed ecosistema fluviale è un punto critico per il controllo ed il monitoraggio della qualità dell'acqua. Il punto chiave è la stima del coefficiente di reareaazione  $K$  necessario per la determinazione della capacità autodepurativa del fiume. Tale parametro è inoltre richiesto per il calcolo della respirazione ( $R$ ) e della produzione fotosintetica di Ossigeno ( $P$ ).

In condizioni naturali l'ecosistema è caratterizzato da metabolismo eterotrofo e  $P/R < 1$ . Quando si verifica un aumento della concentrazione di nutrienti la produzione  $P$  è stimolata e questo si traduce in un  $P/R > 1$ .

I metodi per il calcolo della produttività sono legati all'andamento della concentrazione di DO che dipende a sua volta da  $K$ . Una buona stima di questo parametro è quindi fondamentale per lo studio di qualsiasi ecosistema fluviale.

Esistono due metodi per la valutazione di  $K$ . Il primo si basa sullo studio della concentrazione di Ossigeno Disciolto nelle ore notturne, mentre il secondo effettua il confronto tra la radiazione incidente e la produzione di ossigeno nelle ore soleggiate.

Il processo con cui l'ossigeno è scambiato tra l'atmosfera e il corpo idrico può essere vista come la diffusione attraverso un sottile strato d'acqua. Questa teoria porta alla conclusione che la variazione della concentrazione di DO nel tempo è proporzionale alla differenza tra la concentrazione a saturazione e quella presente nel corso d'acqua in un dato istante. Il bilancio totale di ossigeno è inoltre influenzato dai termini di produzione e consumo sopra citati.

La variazione nel tempo della concentrazione di Ossigeno Disciolto in un ecosistema fluviale è rappresentata quindi dalla seguente formula:

$$\frac{dC}{dt} = K(C_{sat} - C) + P - R$$

dove:

$C$  = Concentrazione di Ossigeno Disciolto;

$K$  = Coefficiente di Reareaazione;

$C_{sat}$  = Concentrazione dell'Ossigeno Disciolto a saturazione;

$P$  = Produzione di Ossigeno per fotosintesi;

$R$  = Respirazione (carico di BOD).

Il metodo che abbiamo utilizzato per la stima di  $K$  è quello che si basa sull'analisi della concentrazione nelle ore notturne. Questo richiede due ipotesi abbastanza restrittive:

1-  $R = \text{cost}$  nelle ore notturne;

2-  $K = \text{cost.}$  su periodi di durata limitata (nel nostro caso  $K = \text{cost}$  per una settimana).

Tale metodo consiste nella soluzione dell'equazione differenziale sopra riportata che può essere trattata sia in senso Euleriano che Lagrangiano. Noi abbiamo scelto il secondo approccio in quanto le misure di cui disponiamo sono state prelevate dalla stessa stazione ad istanti diversi.

La concentrazione di saturazione  $C_{sat}$  dipende dalla temperatura secondo la relazione:

$$C_{sat}(T) = 14.6244 - 0.367134 \cdot T + 0.0044972 \cdot T^2$$

Il primo passo che abbiamo compiuto è stato proprio il calcolo di tale concentrazione conoscendo il valore della temperatura ad ogni istante di campionamento. I valori ottenuti sono compresi nell'intervallo [8.29mg/l , 9.19mg/l] per la serie di Settembre e nell'intervallo [7.39mg/l , 8.01mg/l].

Abbiamo quindi deciso di utilizzare, per una prima analisi, i dati originali di concentrazione non trattati. In seguito abbiamo confrontato i risultati con quelli ottenuti mediante l'utilizzo delle approssimazioni wavelet (A1) e spline ( $p=0.5$ ) precedentemente determinate.

A questo punto disponiamo delle misure di concentrazione  $C$  e concentrazione di saturazione  $C_s$  agli istanti di campionamento. Per il calcolo di  $K$  dovremmo conoscere  $R$ ; si ricorre quindi alla calibrazione parametrica. In pratica si sono determinati i valori numerici di  $K$  e  $R$  in modo da minimizzare la differenza fra la risposta del modello ed i dati sperimentali. Ricordiamo che il nostro modello è:

$$\frac{dC}{dt} = K(C_{sat} - C) - R$$

Il funzionale di errore da minimizzare ha la forma:

$$E(P) = \sum_{k=1}^N (y_k(P) - y_k)^T Q_k (y_k(P) - y_k)$$

dove  $y_k(P)$  sono i valori del modello calcolati nei  $k$  punti mentre  $y_k$  sono i dati originali.

$Q_k$  rappresenta l'importanza da dare a ciascun componente in ciascuna misura. Spesso include l'informazione sugli errori di misura per cui, se si suppongono misure incorrelate, è una matrice diagonale costituita dai reciproci delle varianze. Dato che non disponiamo di informazioni sugli errori si pone  $Q_k = I$ .

Ci siamo quindi ricondotte ad un problema di ottimizzazione. Dato che il problema che intendiamo risolvere è dinamico, la soluzione analitica non è disponibile (non sono applicabili metodi basati sul gradiente). Si ricorre allora a metodi di ricerca diretta che si basano sulla conoscenza per punti di  $E(P)$ .

L'algoritmo di ricerca che abbiamo utilizzato è il metodo del semplice: si procede verso il minimo valutando la funzione  $E(P)$  in un numero finito di punti disposti a reticolo (vertici del semplice) scartando ad ogni passo il punto peggiore, nel quale la funzione è massima, e sostituendolo con un nuovo punto che si ritiene più vicino all'ottimo cercato. La scelta del nuovo punto ad ogni passo è effettuata attraverso le operazioni sul semplice: riflessione, espansione, contrazione e riduzione.

Come primo approccio abbiamo studiato il modello con  $C_s$  costante e pari a 8 mg/l. Il modello è stato implementato in *Simulink* e richiamato dal programma.

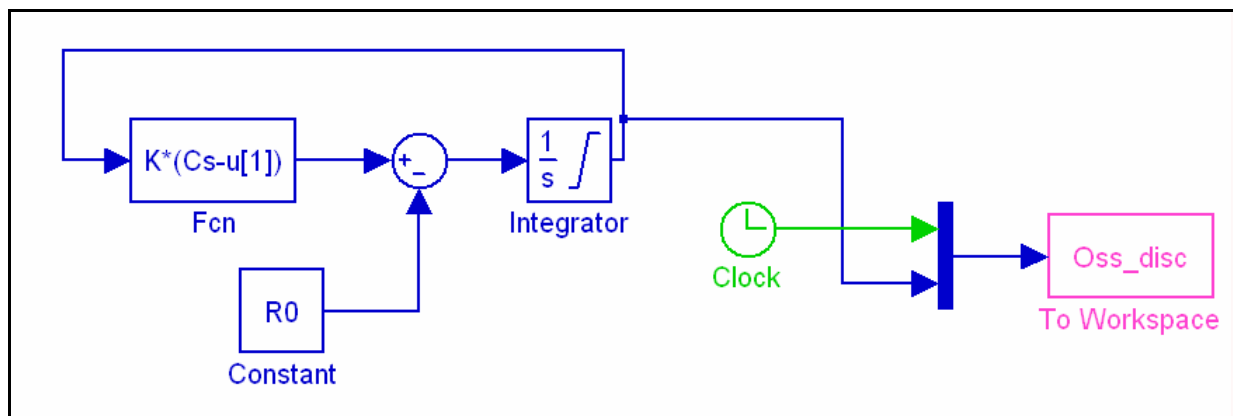


Figura 30. Implementazione del modello con  $C_s$  costante

L'analisi è stata svolta sulle serie originali depurate dai dati diurni (cerchietti rossi) e l'elaboratore ha fornito la soluzione ottima (linea e cerchietti blu) ricavata dalla calibrazione dei parametri  $K$  e  $R$  mediante il metodo del semplice.

Di seguito si riporta anche l'andamento del funzionale di errore all'aumentare del numero di iterazioni. Si osserva una progressiva riduzione fino al raggiungimento di un accettabile valore minimo che soddisfa i criteri di arresto.

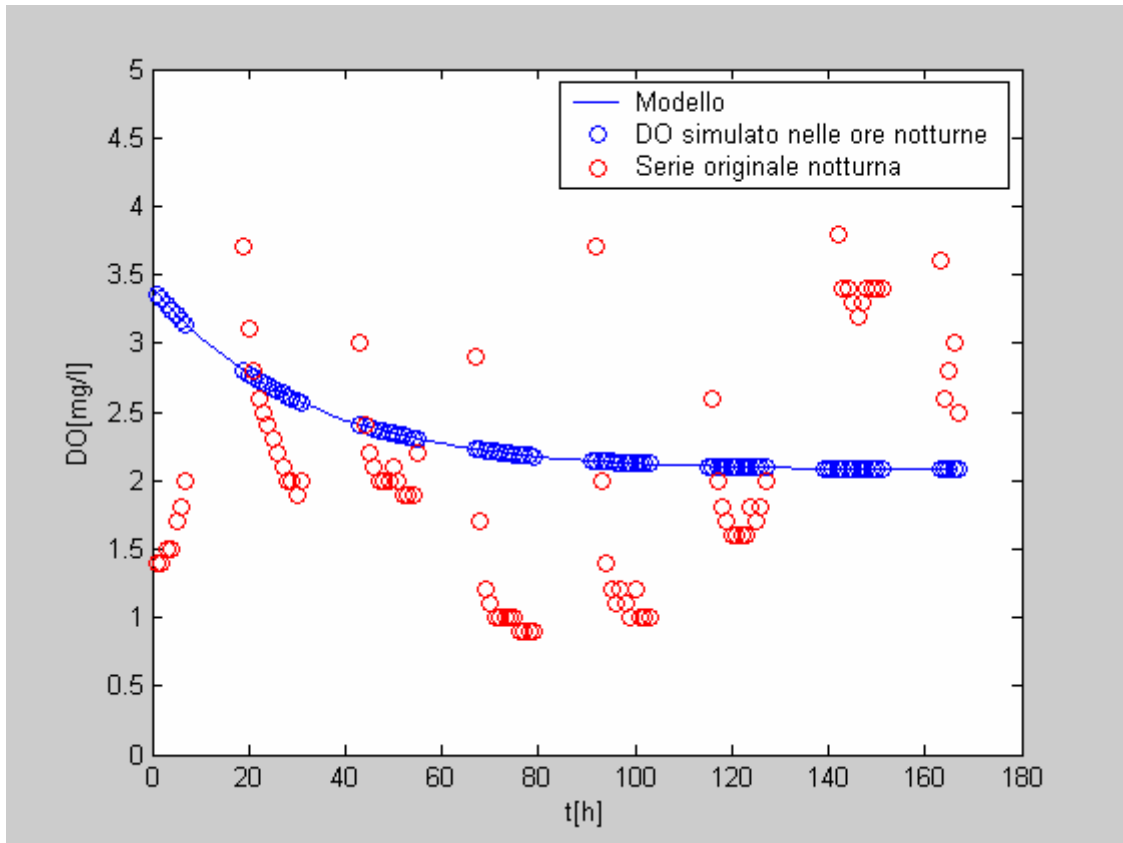


Figura 31. Serie originale Luglio-Agosto notturna e relativo modello, Cs costante

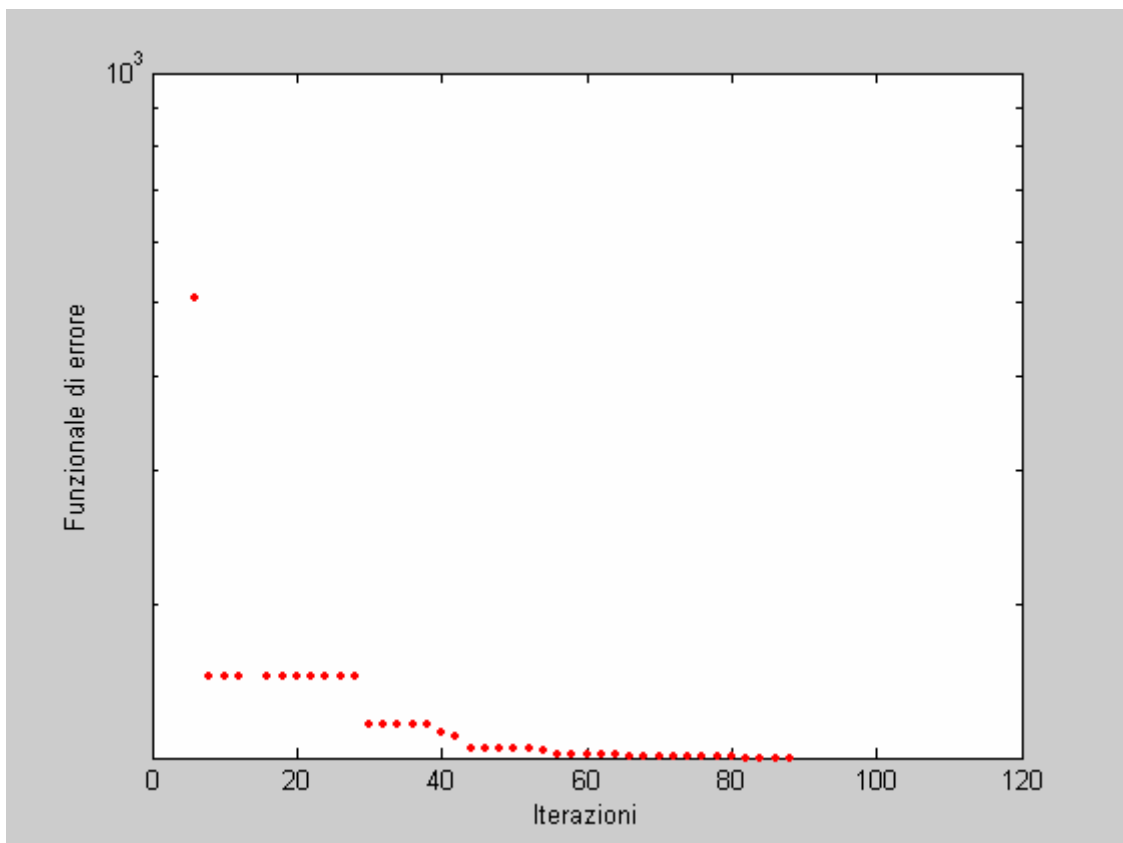


Figura 32. Andamento del funzionale di errore del simplesso

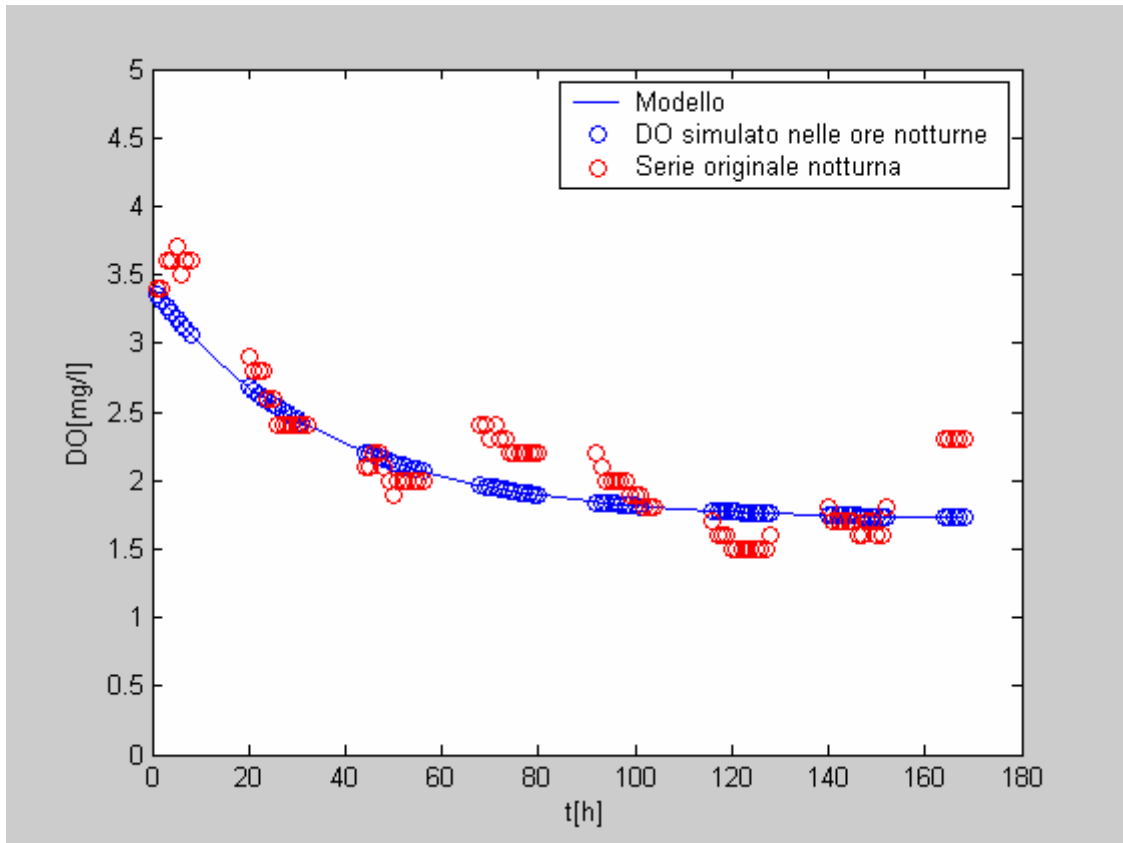


Figura 33. Serie originale Settembre notturna e relativo modello, Cs costante

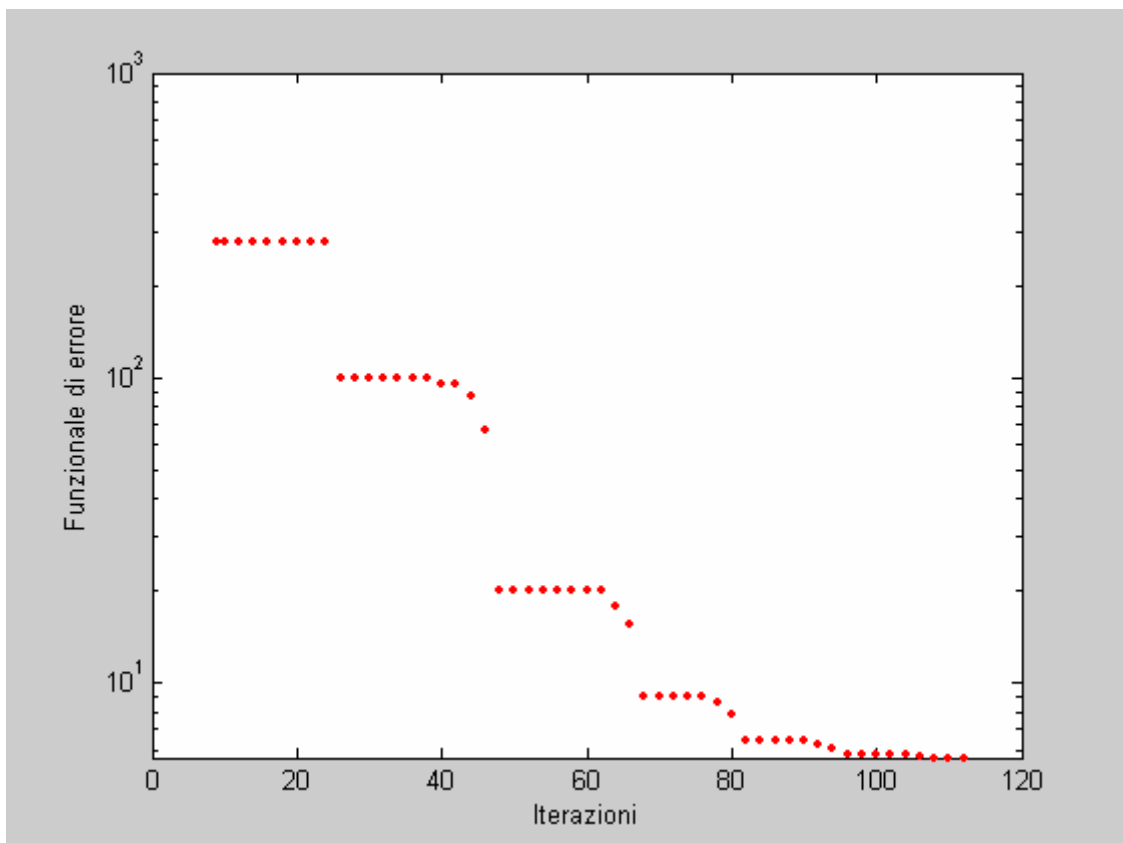


Figura 34. Andamento del funzionale di errore del simplesso

Le operazioni compiute con Cs costante sono state ripetute anche nel caso più realistico di concentrazione di saturazione variabile con la temperatura.

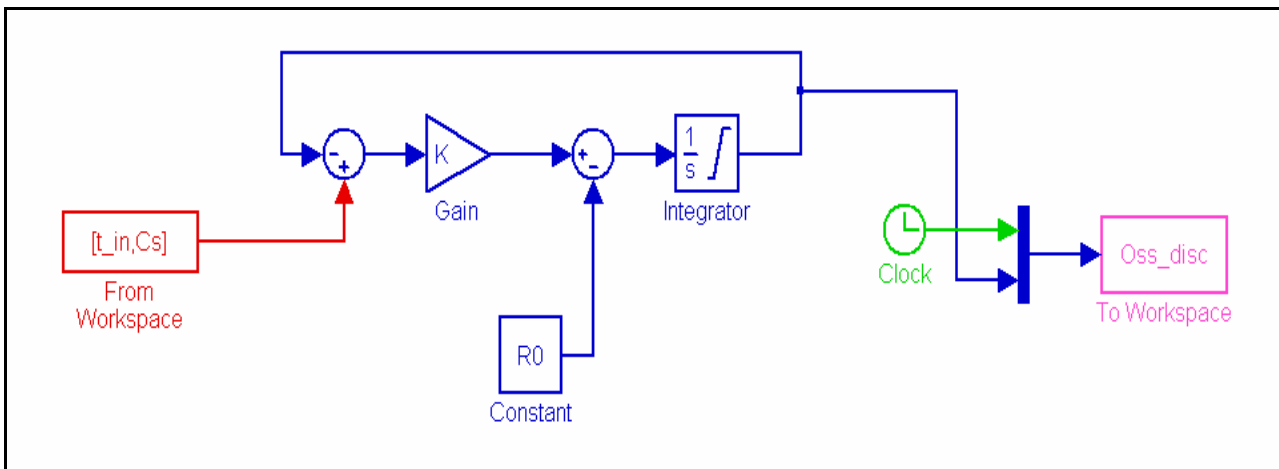


Figura 35. Implementazione del modello con Cs variabile

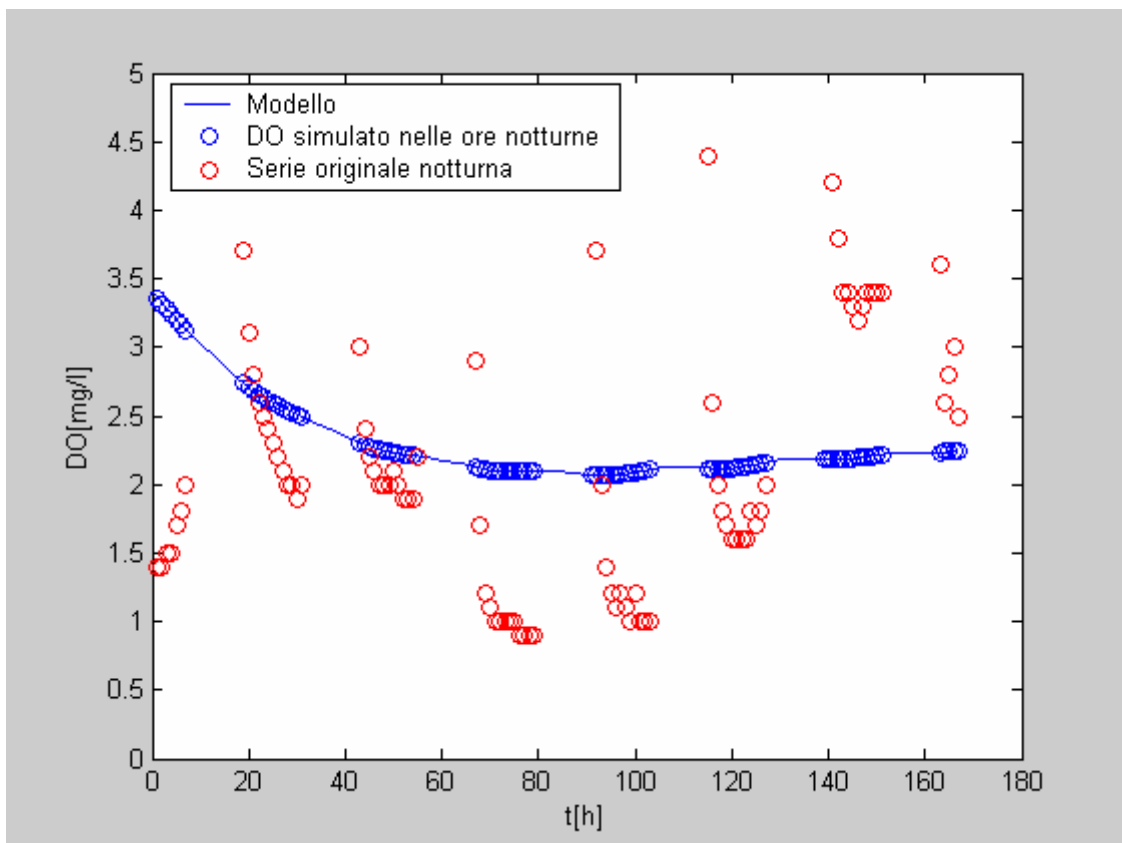


Figura 36. Serie originale Luglio-Agosto notturna e relativo modello, Cs variabile

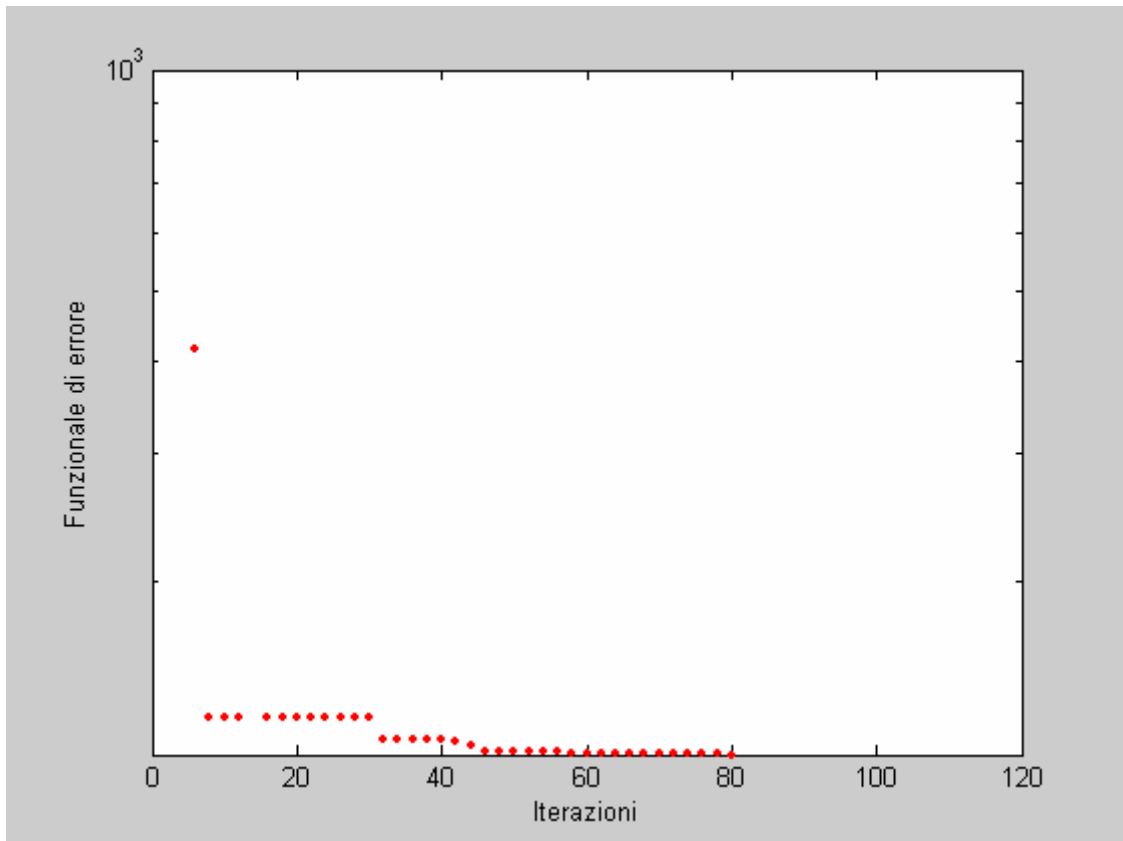


Figura 37. Andamento del funzionale di errore del simplesso

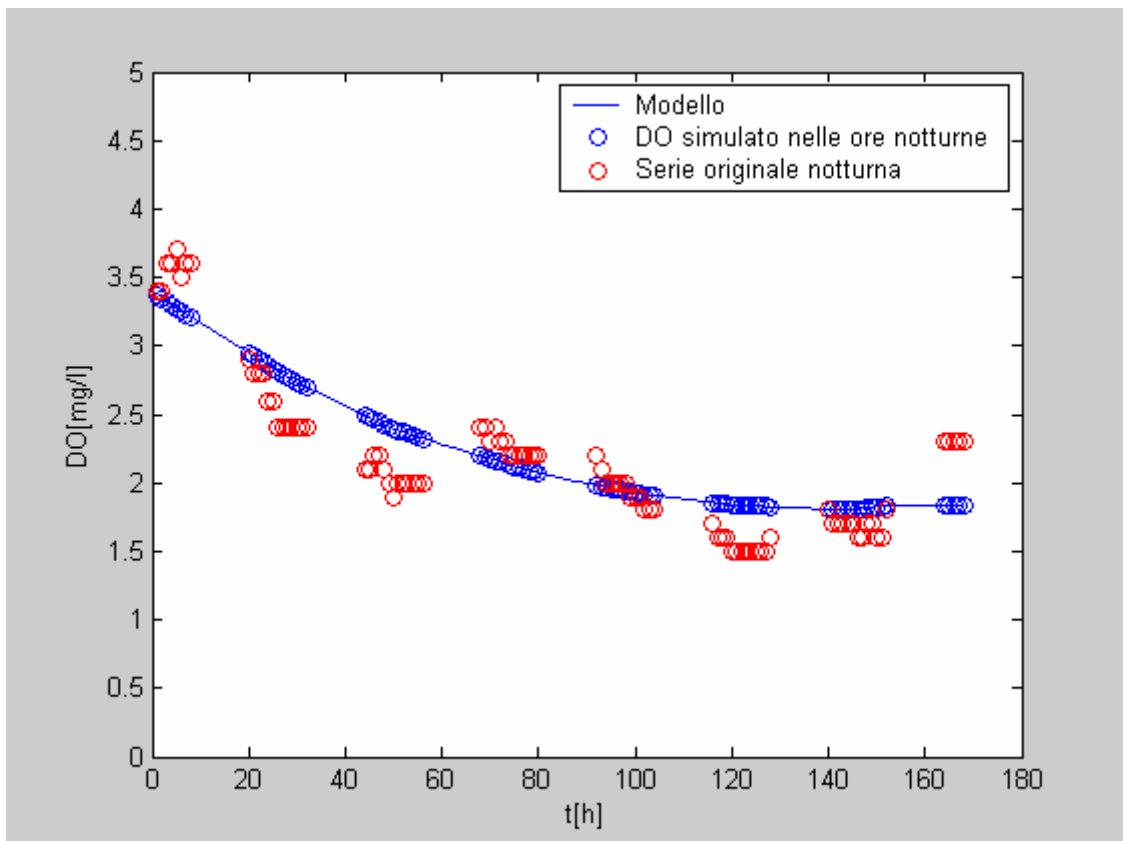


Figura 38. Serie originale Settembre notturna e relativo modello, Cs variabile

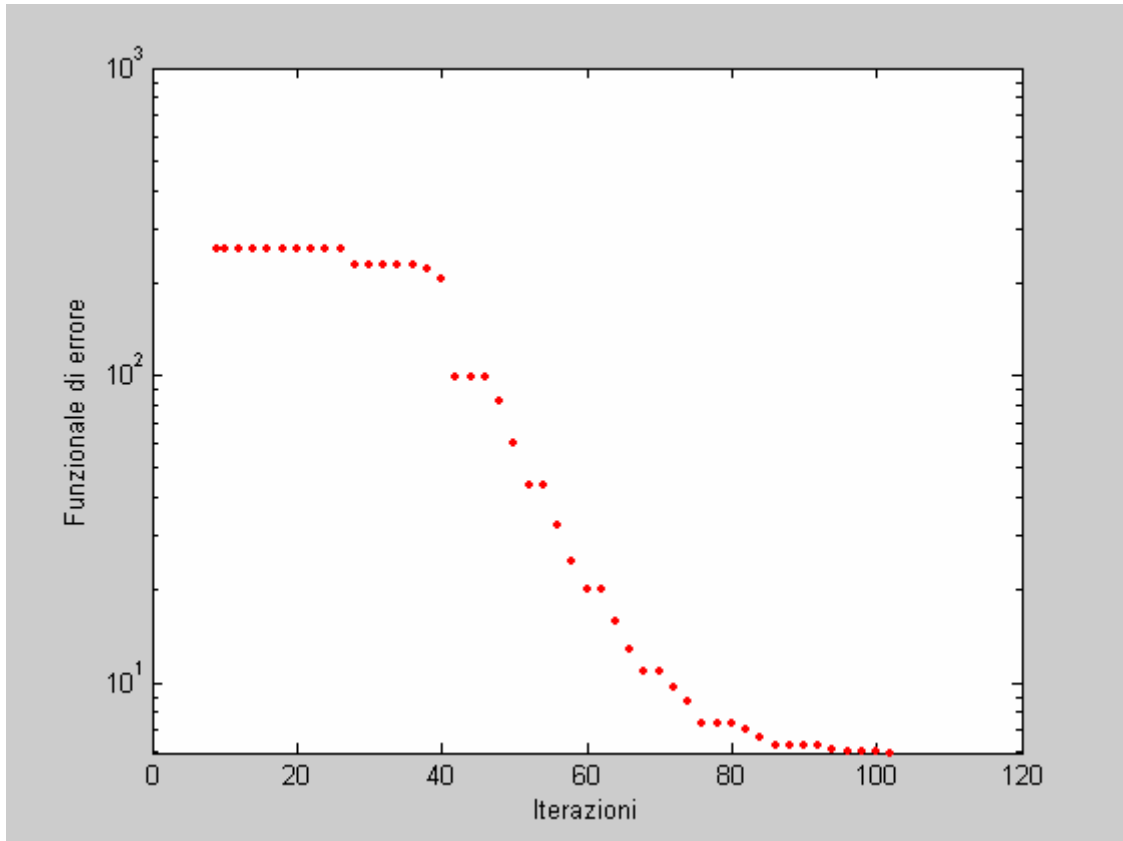


Figura 39. Andamento del funzionale di errore del simplesso

Nella tabella che segue abbiamo riportato i risultati ottenuti dalle quattro simulazioni svolte. Si osserva che il valore di K e R ottenuto nel caso di Cs costante si avvicina molto a quello ottenuto nel caso più reale di Cs variabile. Tuttavia l'approssimazione per la serie di Settembre è meno buona perché il valore fissato di 8 mg/l si discosta maggiormente dalla media reale dei valori.

SIMULAZIONE	Dati originali	
	K	R
1- Luglio-Agosto, Cs cost	0.032	0.189
2- Luglio-Agosto, Cs var	0.034	0.190
3- Settembre, Cs cost	0.028	0.175
4- Settembre, Cs var	0.024	0.165

Abbiamo poi effettuato ulteriori simulazioni inserendo come ingressi al modello i valori di concentrazione ottenuti con le approssimazioni wavelet al livello A1.

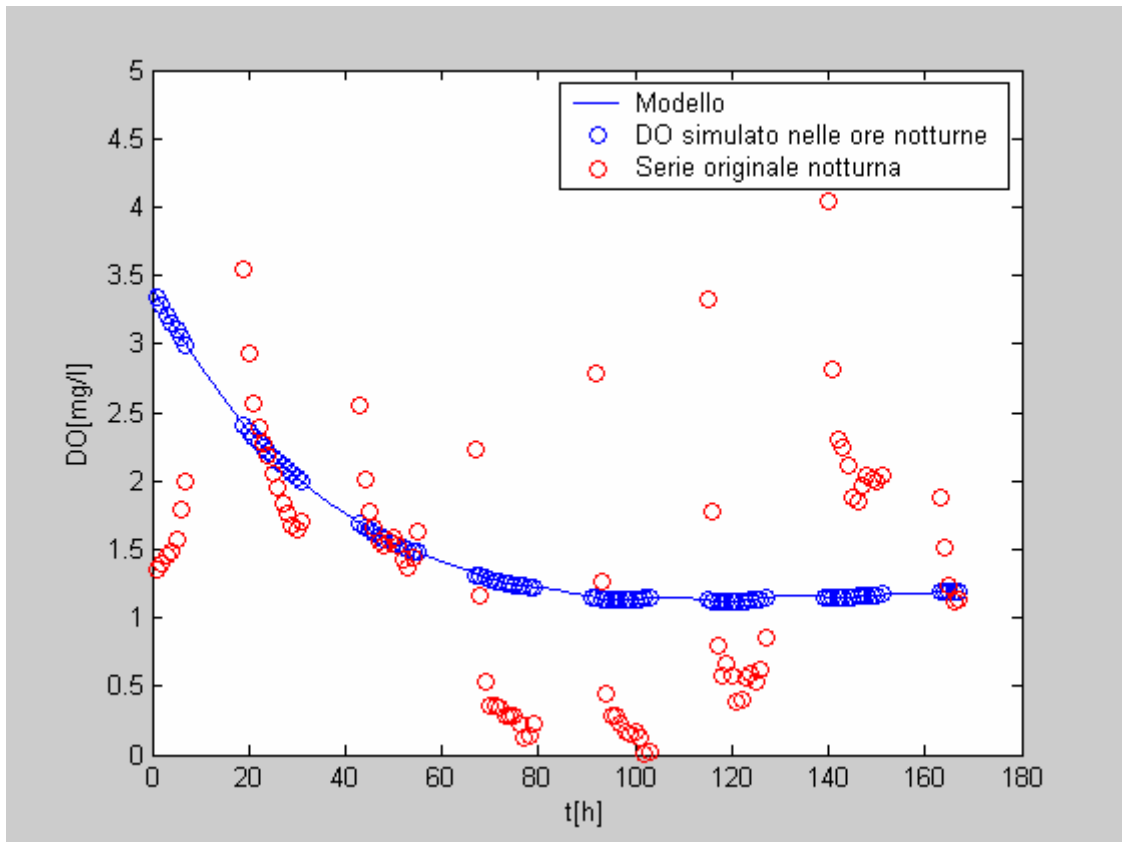


Figura 40. Serie wavelet Luglio-Agosto notturna e relativo modello.

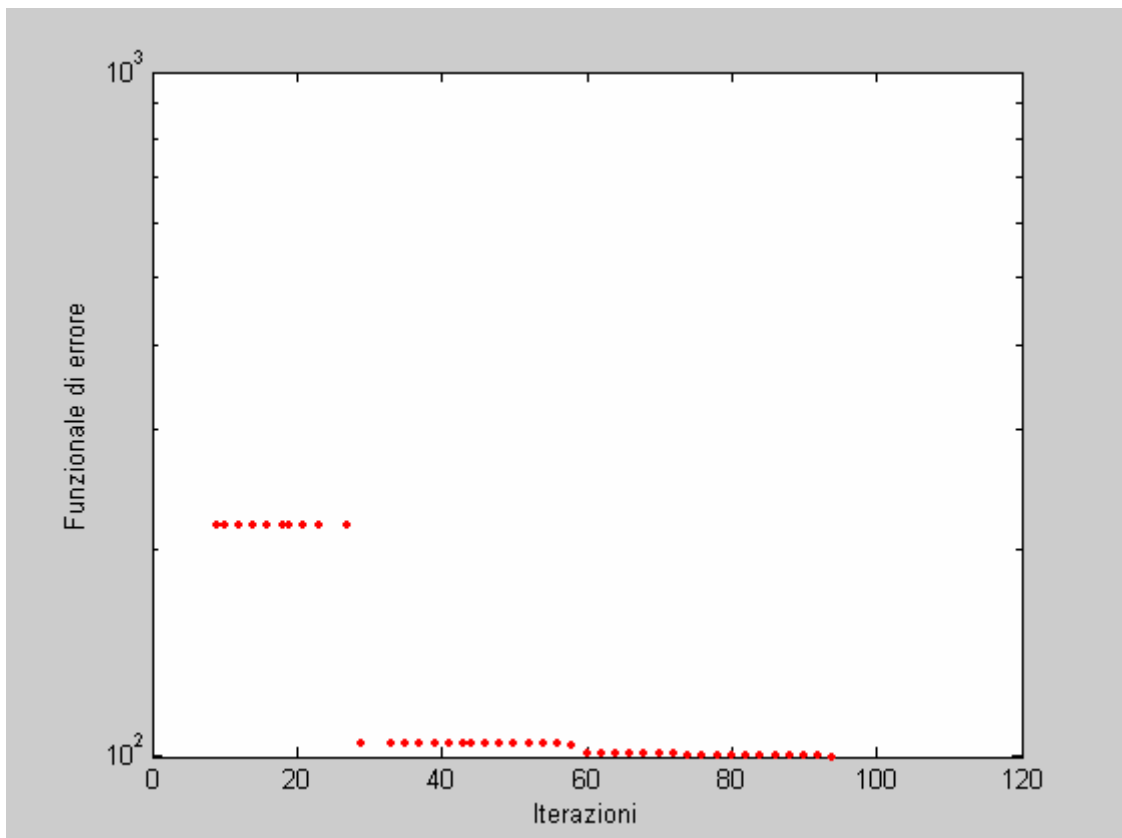


Figura 41. Andamento del funzionale di errore del semplice

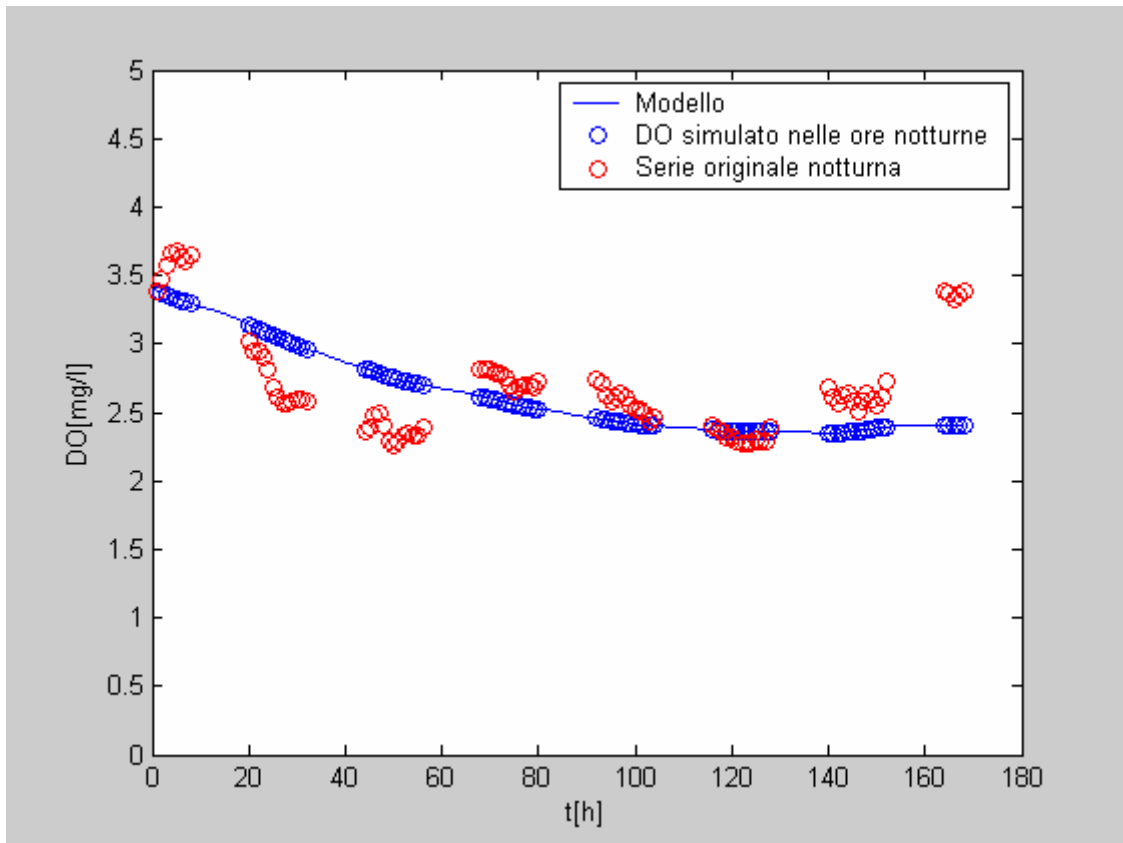


Figura 42. Serie wavelet Settembre notturna e relativo modello.

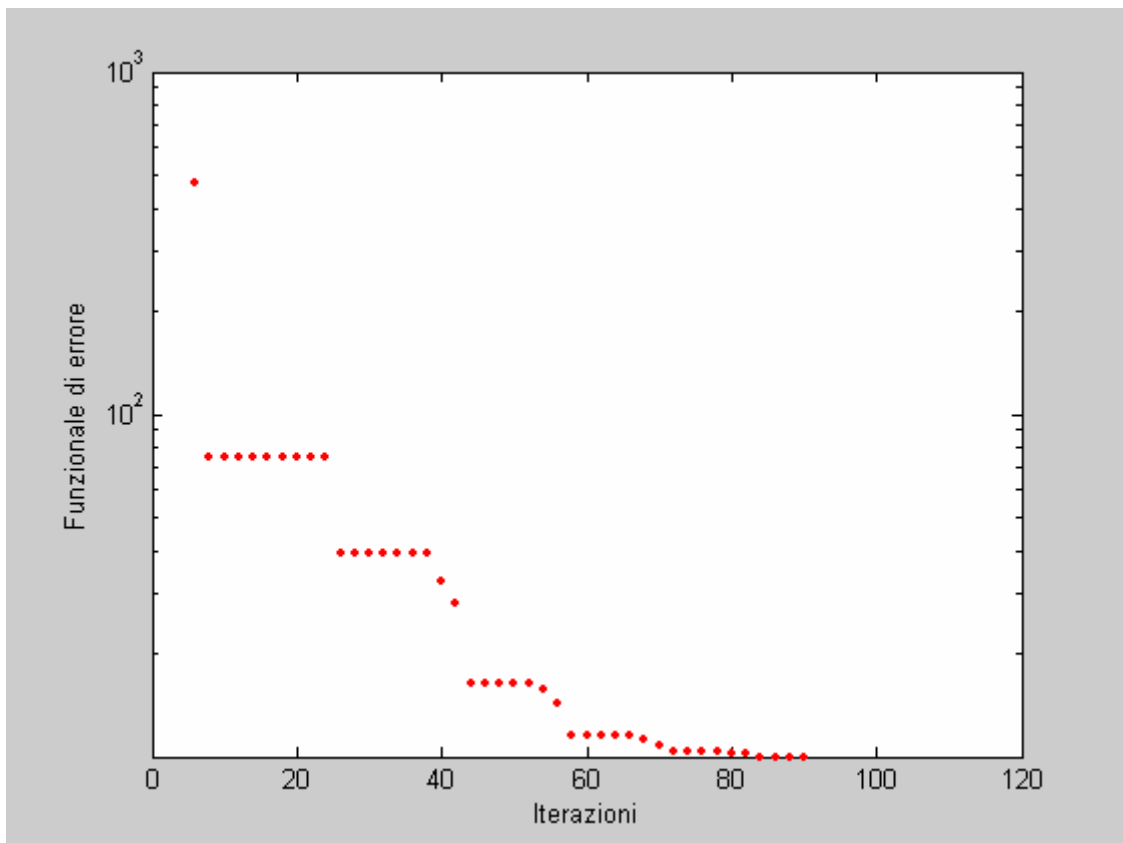


Figura 43. Andamento del funzionale di errore del semplice

I grafici che seguono si riferiscono alle simulazioni svolte con l'approssimazione spline con  $p=0.5$  in ingresso al modello.

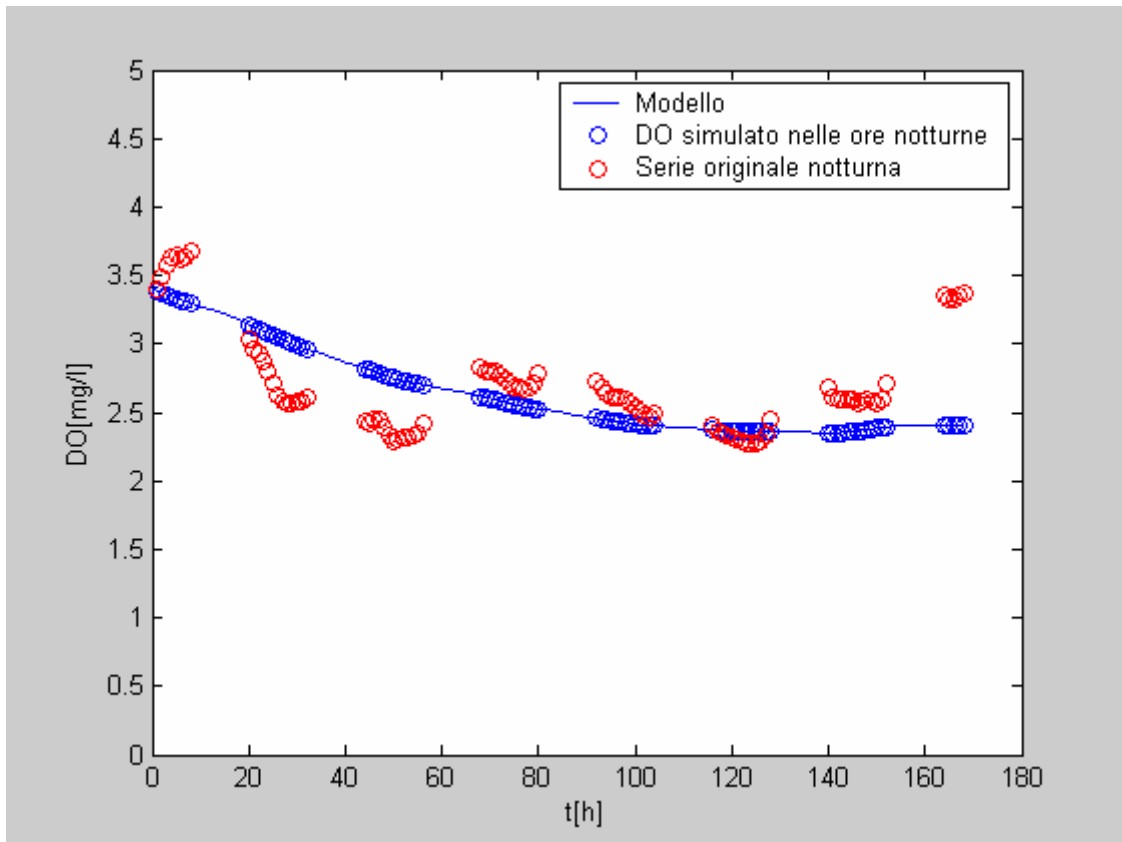


Figura 44. Serie spline Settembre notturna e relativo modello.

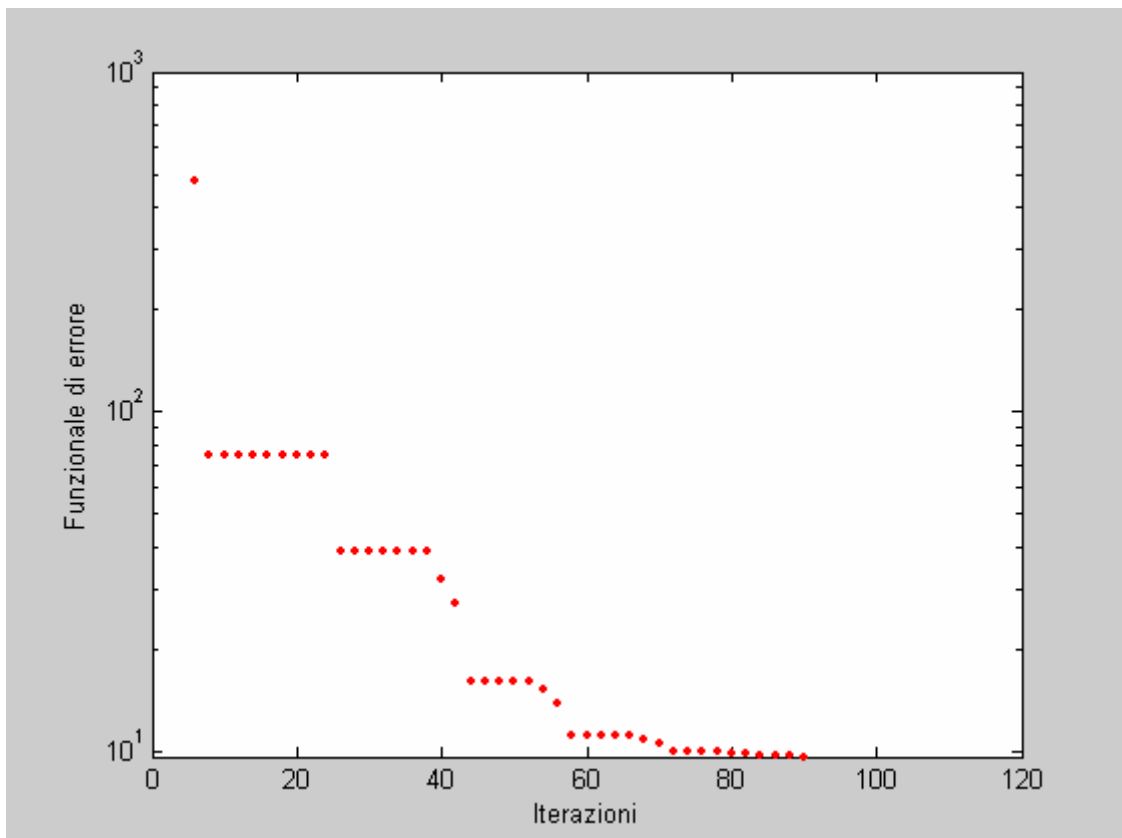


Figura 45. Andamento del funzionale di errore del semplice

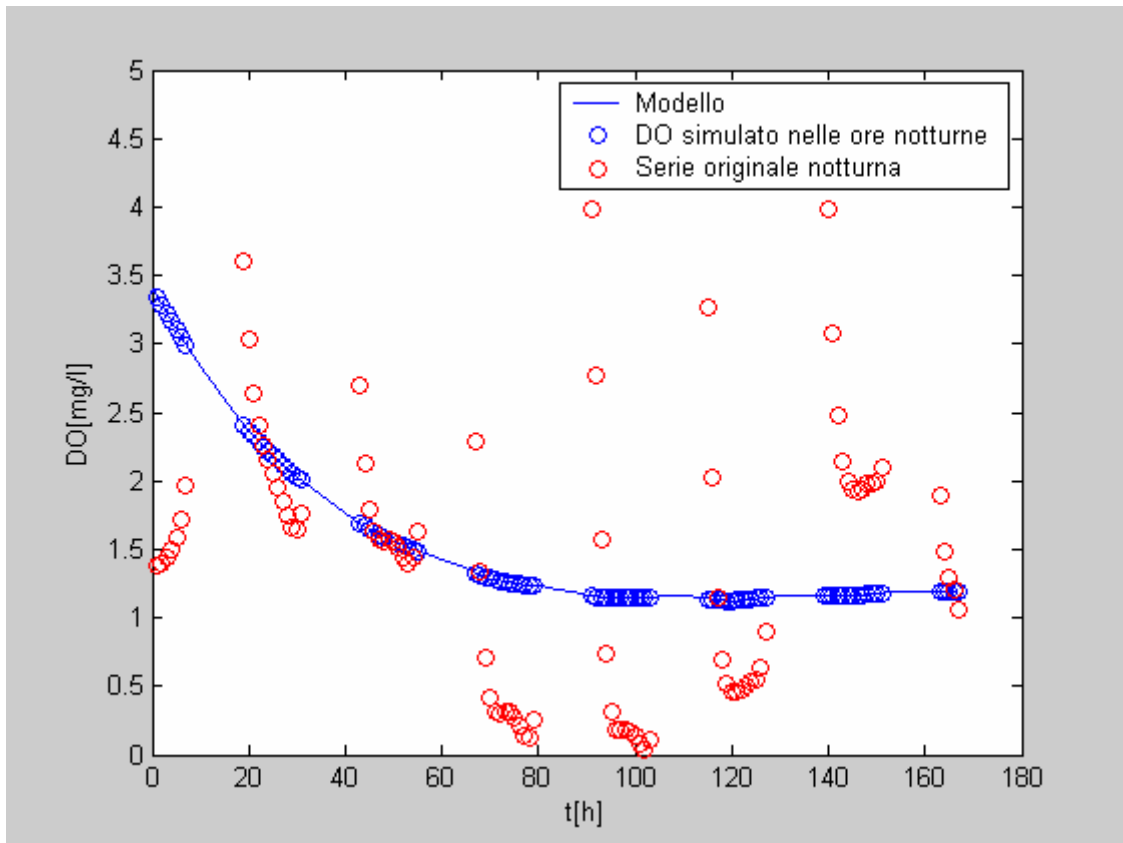


Figura 46. Serie spline Luglio-Agosto notturna e relativo modello.

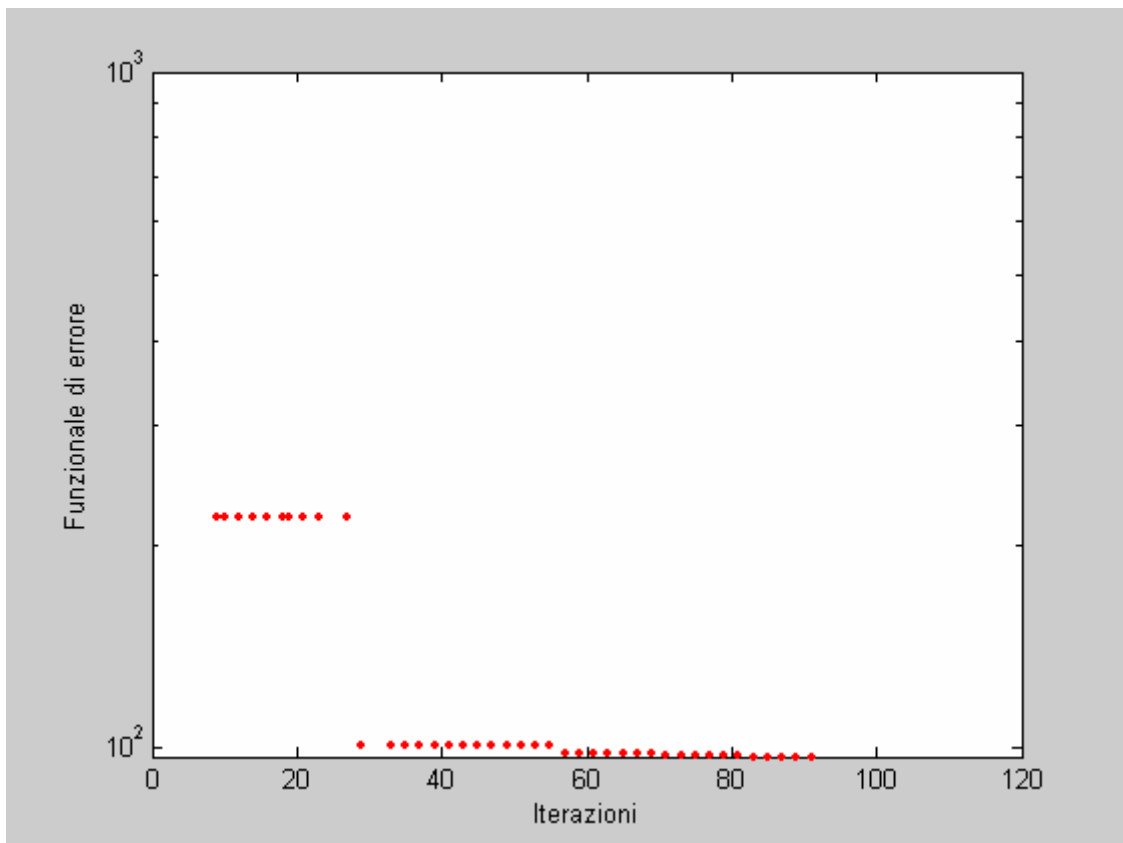


Figura 47. Andamento del funzionale di errore del semplice

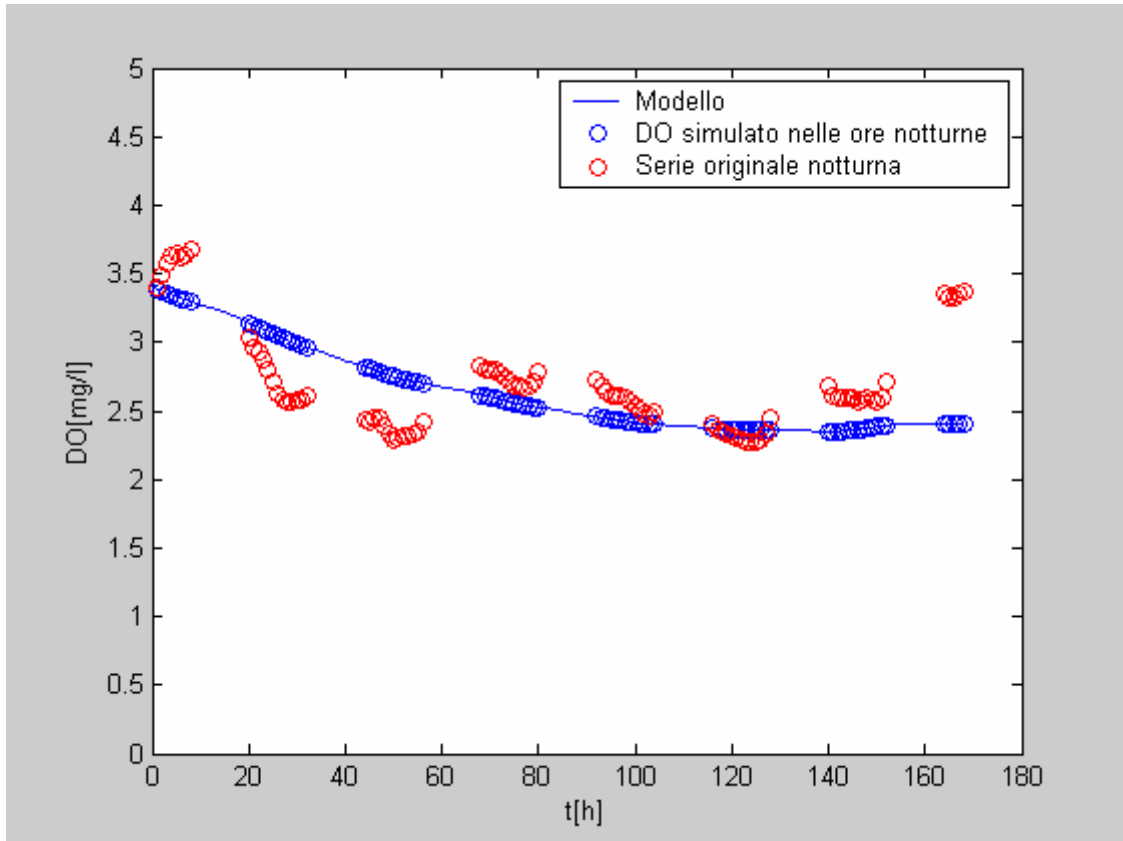


Figura 48. Serie spline Settembre notturna e relativo modello.

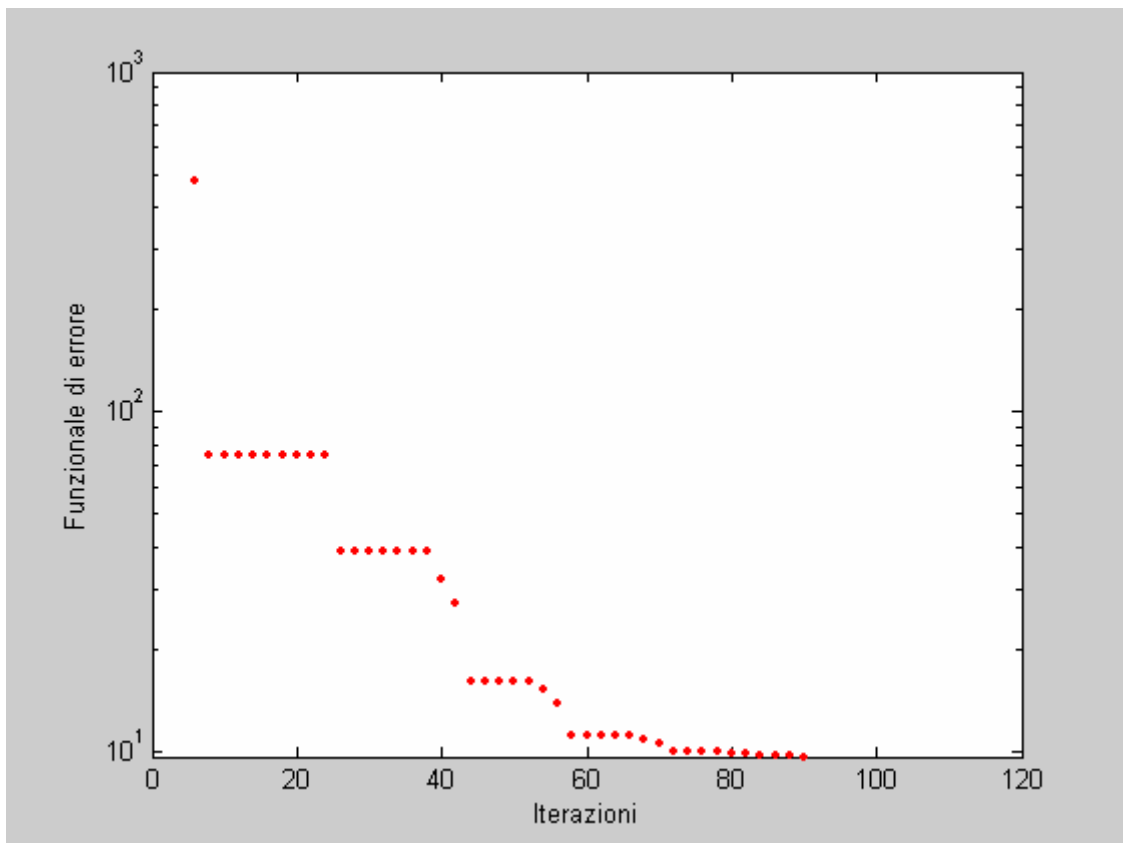


Figura 49. Andamento del funzionale di errore del semplice

SIMULAZIONE	Dati originali		Wavelet		Splines	
	K	R	K	R	K	R
1- Luglio-Agosto	0.032	0.189	0.0279	0.184	0.0275	0.183
2- Settembre	0.034	0.190	0.0292	0.180	0.0297	0.181

Come si evince dai valori riportati in tabella le approssimazioni wavelet e spline sono da considerarsi ugualmente valide dal momento che i risultati della calibrazione sono pressoché equivalenti.

Si nota una differenza un po' più marcata tra i risultati ottenuti con l'utilizzo delle approssimazioni rispetto ai dati originali. Questo può essere spiegato dalla presenza di rumore nella serie di dati originali che probabilmente ha reso più difficoltoso il calcolo delle derivate.

Si può infine concludere che per il trattamento di serie di dati ambientali è preferibile effettuare un'operazione preliminare di denoising per poter lavorare meglio sull'informazione contenuta nei dati, in particolare se si ha a che fare con derivate.