

Modelli parametrici lineari per serie idrologiche giornaliere

Salvatore Grimaldi

*Dipartimento di Idraulica Trasporti e Strade,
Università degli Studi di Roma "La Sapienza".
E-mail: salvatore.grimaldi@uniroma1.it*

Summary: The aim of this paper is to describe, and solve in some cases, the problems that arise in modelling hydrological daily time series. The preliminary analysis, the identification and simulation steps of the traditional procedure are studied in depth. The effects of Box and Cox transformation, the problem of deseasonalization and the initial parameter estimation of a fractional model, are highlighted. The revised procedure was applied to time series of daily flows of Tevere.

Keywords: Linear Parametric Models, Fractional Model, Autoregressive Metric

1. Introduzione

I fenomeni idrologici d'interesse tecnico-scientifico che necessitano di un'indagine supportata da strumenti di tipo statistico, sono molteplici. La complessità dei processi fisici alla base di tali fenomeni, infatti, non permette in alcuni casi di sviluppare sistemi di previsione o di gestione di natura completamente deterministica.

Dalla complessa genesi degli eventi meteorici e dalle conseguenti trasformazioni che si verificano all'interno del ciclo idrologico deriva l'esigenza di disporre di uno strumento che prescindendo dalla fisica del problema ma che tragga informazioni esclusivamente dal dato misurato.

In tale ambito, la modellazione parametrica delle serie storiche ha avuto, e tuttora riveste, un importante ruolo nelle analisi di tipo idrologico.

La possibilità di identificare una componente dinamica che caratterizzi un segnale ha permesso lo sviluppo di procedure di simulazione e di previsione del segnale stesso.

Nella gestione delle risorse idriche sono numerose le applicazioni volte ad accertare la possibilità di dare origine a serie sintetiche simili a quella reale disponibile, al fine di poter far riferimento a scenari statisticamente “simili”.

Altrettanto numerose sono le applicazioni finalizzate a verificare la possibilità di operare previsioni a breve termine, utili per la gestione in tempo reale di sistemi complessi.

Si pensi ad esempio alla regolazione di invasi artificiali o di strutture idrauliche simili, alle numerose applicazioni sulla trasformazione afflussi-deflussi, all'analisi dei dati per la gestione delle falde, alla generazione di curve di durata di deflussi superficiali, alla simulazione di serie di temperatura, all'analisi del regime delle portate di sorgente e a tanti altri tipi di applicazioni.

Per poter svolgere tali analisi, le procedure tradizionali (Hipel e McLeod, 1994; Piccolo, 1990a; Brockwell e Davis, 1987; Box e Jenkins, 1976), nate per lo più per applicazioni in ambito idrologico con serie annuali, sono state modificate per poter essere adeguate alla modellazione di serie registrate a scala di aggregazione sempre più piccola.

Il presente lavoro ha la finalità di analizzare le problematiche metodologiche emerse nell'analisi di serie idrologiche giornaliere e di proporre alcune miglioramenti procedurali. Vengono infatti analizzate le singole fasi della metodologia classica evidenziando quali sono i problemi da risolvere ed eventualmente le possibili soluzioni per alcuni di essi.

Le serie giornaliere, oltre a fornire con maggior dettaglio l'informazione desiderata, statisticamente mostrano il vantaggio, di poter usufruire di un campione più numeroso rendendo più consistenti le stime dei parametri del modello identificato. Tali serie però presentano l'inconveniente di una maggiore variabilità che riduce

ulteriormente l'assolvimento delle ipotesi di base dei processi stocastici. E' infatti più significativa la presenza di andamenti eteroschedastici e di distribuzioni non Gaussiane. Ciò ha indotto ad approfondire, come descritto nel paragrafo 2, la reale necessità dell'utilizzo delle trasformazioni di Box e Cox (1964), e gli effetti che tale maggiore variabilità produce nella destagionalizzazione di tali serie.

Al diminuire della scala temporale di osservazione di una grandezza idrologica si constata un'estensione della struttura di autocorrelazione del segnale ad essa associato. Ciò ha reso necessario in alcuni casi, considerare i modelli frazionari o a "lunga memoria". In letteratura sono numerosi i contributi che hanno approfondito le problematiche che si possono incontrare nelle applicazioni di tali modelli in ambito idrologico (Hosking, 1984; Montanari et al., 1997; Montanari et al., 2000; Grimaldi, 2000). A tal proposito nel paragrafo 3 viene descritta e testata una recente tecnica di stima iniziale dei parametri dei modelli frazionari, basata sulla metrica autoregressiva (Corduas, 1999; Corduas, 2000), che si pensa debba essere considerata per poter effettuare una veloce modellazione delle serie e per poter impostare procedure automatiche di identificazione.

Nel quarto paragrafo viene infine illustrato un esempio di modellazione su una serie di portate del fiume Tevere per verificare l'applicabilità delle procedure descritte.

2. Le analisi preliminari

Nel caso di serie di tipo idrologico a scala di aggregazione giornaliera l'attenzione principale, nella fase delle analisi preliminari, va rivolta a due precisi aspetti: la normalizzazione e la destagionalizzazione.

Riguardo al primo aspetto la maggiore variabilità presente in questa tipologia di serie esalta le problematiche riguardo l'ipotesi sulla distribuzione rispetto ai casi tradizionali di serie annuali o mensili.

Come è noto, in tali circostanze si è soliti far riferimento alla trasformazione di Box e Cox:

$$Y_t = (X_t^\lambda - 1) / \lambda; \text{ ove per } \lambda = 0, Y_t = \log(X_t) \quad (1)$$

Nel caso di serie idrologiche tale classe di trasformazioni non produce sempre risultati coerenti. Infatti in un precedente lavoro (Grimaldi, 2000) è stato verificato come per la serie delle portate medie giornaliere del fiume Tevere, registrate presso la stazione di Roma-Ripetta, il valore di $\lambda = -0.75$, che rende simmetrica la distribuzione della serie trasformata, consente una buona modellabilità della serie, ma risulta completamente inutile in fase di simulazione poiché le serie generate presentano valori divergenti derivanti dalle trasformazioni inverse.

La disponibilità di tecniche di stima che prescindono, almeno asintoticamente, dal tipo di distribuzione che caratterizza la serie analizzata (Taqqu e Teverovsky, 1998) ha suggerito di provare ad accettare la non gaussianità della stessa. In letteratura sono presenti contributi (Montanari et al., 1997, Montanari et al., 2000) che mostrano la modellabilità delle serie con tale approccio ma anche in questo caso le serie sintetiche, ovvero simulate, presentano distorsioni, sicuramente meno importanti rispetto al caso precedente, ma comunque inaccettabili fisicamente (presenza valori negativi).

Esempi pratici applicativi descritti nell'ultimo paragrafo mostrano come la migliore soluzione rimanga quella "classica", e cioè operare la semplice trasformazione logaritmica ove questa sia strettamente necessaria. Bisogna però sottolineare che mentre da una parte si ottiene un miglioramento nella modellazione assieme alla certezza nell'aver valori positivi nelle serie sintetiche generate (trasformazione inversa sempre positiva), dall'altra si opera comunque un'alterazione della struttura di correlazione e si ottengono in alcune delle serie simulate delle distorsioni non accettabili per i valori estremi delle stesse.

Il secondo aspetto da approfondire, come già detto, in caso di analisi di serie giornaliere è il problema della stagionalità. Infatti i fenomeni idrologici sono caratterizzati da una componente stagionale di tipo deterministico dovuta alla rotazione della terra intorno al sole, comportando nel caso delle serie giornaliere una periodicità di periodo

365¹. In particolare ciò si riflette in un andamento periodico dei livelli medi e degli scarti quadratici medi che rende la serie non stazionaria.

Per ristabilire la condizione di stazionarietà, per una serie idrologica X_t^t con t indice temporale progressivo ($t=1, \dots, N$ con N numerosità della serie) e \mathbf{t} giorno ($\mathbf{t}=1, \dots, 365$) dell'anno di osservazione, si procede definendo due componenti periodiche $m_{\mathbf{t}}$ e $n_{\mathbf{t}}$ tali che la serie Y_t^t risultante dalla relazione:

$$Y_t^t = \frac{X_t^t - m_{\mathbf{t}}}{n_{\mathbf{t}}} \quad (2)$$

non risulti più affetta da nessun tipo di stagionalità.

Nei casi studio presenti in letteratura si utilizza usualmente una tecnica che denomineremo nel seguito *Metodo Classico*, valida sia per la media sia per la varianza. Prima di tutto si suddividono le osservazioni disponibili in 365 sottoserie, ognuna comprendente i dati osservati in ogni singolo giorno dell'anno. Quindi si procede al calcolo del valore della media e della varianza per ognuna delle sottoserie, ottenendo così delle serie di 365 dati che rappresentano il valore di ciascuna statistica lungo il periodo annuale.

Analiticamente ciò è esprimibile con le seguenti relazioni:

$$\mathbf{m}_{\mathbf{t}} = n^{-1} \sum_{i=1}^n X_{\mathbf{t}+365(i-1)}^t \quad \mathbf{s}_{\mathbf{t}}^2 = (n-1)^{-1} \sum_{i=1}^n [X_{\mathbf{t}+365(i-1)}^t - \mathbf{m}_{\mathbf{t}}]^2 \quad (3)$$

con n numero di anni disponibili dalle osservazioni.

Sostituendo nella (2) i valori ottenuti nella (3), $\mathbf{m}_{\mathbf{t}}$ e $\mathbf{s}_{\mathbf{t}}$ si ottiene la destagionalizzazione desiderata.

Tale metodo applicato a serie giornaliere, mostra alcuni problemi legati, come detto, alla eccessiva variabilità mostrata da questo tipo di serie. Infatti, riferendoci ancora alla serie del Tevere misurata dal 1930 al 1983, si constata che le due componenti calcolate (cfr. Figure 1 e 2), $\mathbf{m}_{\mathbf{t}}$ e $\mathbf{s}_{\mathbf{t}}$ mostrano una variabilità non più accettabile sia

¹ E' prassi nelle applicazioni di tipo idrologico correggere la reale periodicità non intera (365.25) scartando i dati registrati il 29 febbraio.

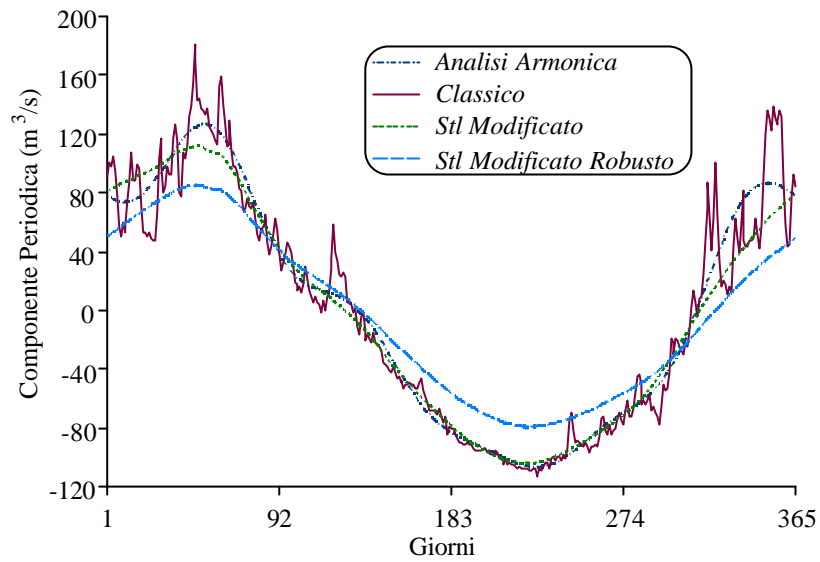


Figura 1. Componente periodica nella media. Confronto tra quattro metodi di stima (Grimaldi e Montanari, 2000).

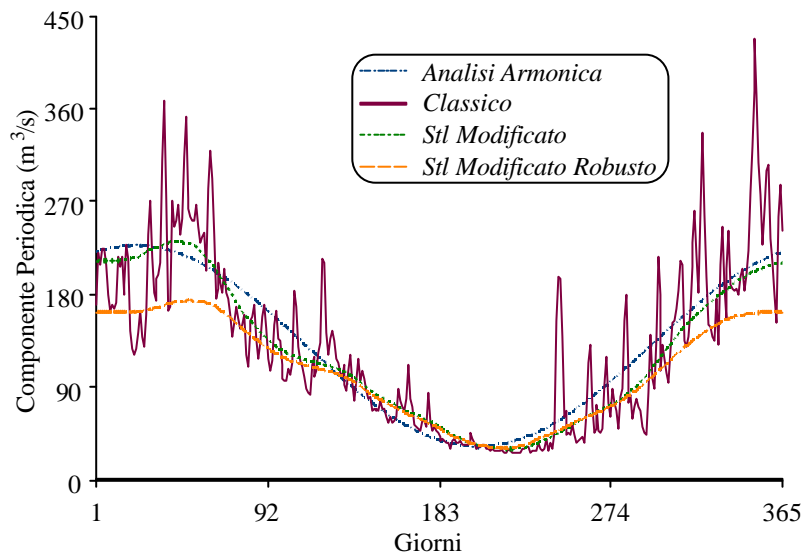


Figura 2. Componente periodica negli scarti quadratici medi. Confronto tra quattro metodi di stima (Grimaldi e Montanari, 2000).

fisicamente sia statisticamente. Difatti, esse incorporano una parte di rumore che quindi è considerato come una componente fissa da non modellare.

Per ovviare a tale problema si può far riferimento a diversi metodi che prevedono un'operazione di "smoothing" della componente periodica. Per le applicazioni svolte in questa memoria si è utilizzato il metodo denominato *STL Modificato Robusto* (per la descrizione e il confronto con gli altri metodi si veda Grimaldi e Montanari, 2000 e Grimaldi, 2001). Come si può constatare sempre dalle Figure 1 e 2 esso fornisce delle componenti più congruenti alla procedura di modellazione.

Ulteriore elemento da considerare nel trattamento delle serie giornaliere rispetto a quelle mensili per quanto concerne la destagionalizzazione è la maggiore significatività delle componenti periodiche degli scarti quadratici medi. La reale necessità di rimuovere entrambe le componenti periodiche si può evidenziare empiricamente. Infatti, a titolo di esempio per la serie del Tevere nei due casi: singola depurazione, doppia depurazione, se si confrontano le distribuzioni delle serie generate e originarie (cfr. Figura 3) si nota come effettivamente sia necessario depurare la serie di entrambe le componenti (Grimaldi, 2001).

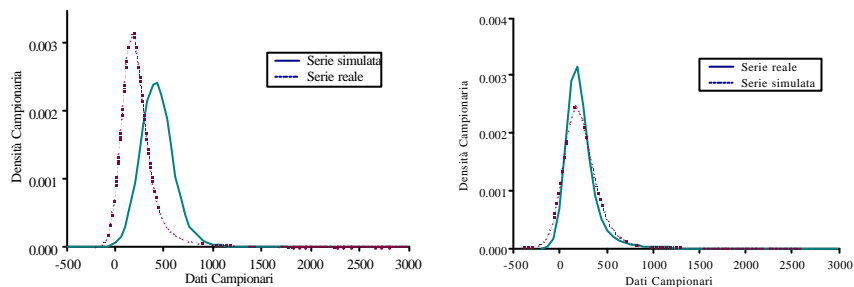


Figura 3. Confronto tra la distribuzione di una serie simulata e una serie osservata. a) Nel caso di analisi di una serie depurata della componente periodica della media. b) Nel caso di un'analisi di una serie depurata di entrambe le componenti periodiche.

3. I modelli frazionari misti - un nuovo metodo di stima iniziale

Come premesso nell'introduzione, le serie osservate giornalmente mostrano in molti casi una struttura di correlazione diversa da quella che è stata introdotta per l'analisi delle serie annuali.

Si è infatti constatato come in alcuni casi solo l'utilizzo dei modelli frazionari misti permette la modellazione di alcune serie giornaliere (Montanari et al., 1997; Grimadi et al., 1999).

A tal fine ricordiamo in breve che $X_t \sim \text{FARIMA}(p,d,q)$ stazionario e invertibile, è un processo con formulazione

$$\mathbf{f}(B) \nabla^d X_t = \mathbf{q}(B) a_t \quad (4)$$

dove B è l'operatore "backward" tale che $B^k X_t = X_{t-k}$, ed inoltre:

$$\mathbf{f}(B) = 1 - \mathbf{f}_1 B - \mathbf{f}_2 B^2 - \dots - \mathbf{f}_p B^p,$$

$$\mathbf{q}(B) = 1 - \mathbf{q}_1 B - \mathbf{q}_2 B^2 - \dots - \mathbf{q}_q B^q,$$

$$\nabla^d = (1 - B)^d \text{ con } -0.5 < d < 0.5,$$

mentre $a_t \sim WN(0, \mathbf{S}^2)$ è una successione di variabili casuali incorrelate di media zero ed omoschedastiche.

La procedura per costruire un modello di tale classe, (Montanari et al. 1997), risulta essere piuttosto articolata nella fase di identificazione nel definire i valori iniziali dei parametri del modello frazionario misto. Essa non comporta alcun aggravio per l'analisi di una singola realizzazione, anzi ne permette una piena comprensione di tutte le trasformazioni, ma risulta inadeguata nel caso si voglia analizzare un numero elevato di serie. La procedura per l'identificazione consiste infatti nelle seguenti fasi:

a) identificazione della lunga memoria con l'applicazione di più metodi di stima iniziale;

b) definizione di un valore di d medio;

c) filtraggio della serie analizzata con il valore di d calcolato: $Y_t = (1 - B)^d X_t$;

d) stima del modello per la serie Y_t depurata del termine di lunga memoria, con le tecniche tradizionali dei modelli a breve memoria.

In questo modo si ottengono i valori iniziali dei parametri AR e MA da utilizzare nella fase di stima.

Occorre poi sottolineare che i principali metodi per l'identificazione e la stima iniziale del parametro di Hurst, e quindi dell'esponente d , sono metodi grafici che producono valori molto approssimati del parametro stimato e che necessitano di un'analisi visiva dei risultati. Inoltre tale approccio presuppone che il valore del parametro d sia costante al variare dei modelli stimati, mentre come riportato negli esempi dell'ultimo paragrafo si nota proprio una significativa variabilità del parametro d al variare del modello ipotizzato.

Per questo motivo si ritiene utile prendere in esame il metodo di stima iniziale proposto in Corduas (1999) e Corduas (2000) per il quale effettueremo una più ampia verifica in relazione alle applicazioni in campo idrologico .

Tale metodo si basa su una caratteristica emersa comparando le strutture di correlazione di modelli a breve memoria e lunga memoria e sapendo che nonostante i primi non riescano a descrivere pienamente la dinamica di una serie a lunga memoria possono comunque fornire una buona approssimazione di quest'ultima (Lawrance e Kottegoda, 1977; Hosking 1984, Tiao and Tsay, 1994).

Empiricamente si può constatare che effettivamente tale "somiglianza" sussiste; infatti se si considerano le configurazioni FARIMA(1,0.35,0) con $f=0.5$ e ARMA(2,1) con $f_1=1.741$ e $f_2=-0.746$ e $q=0.944$ si può constatare come le funzioni di autocorrelazione globali risultino molto simili (cfr. Figura 4), così come gli andamenti delle densità spettrali (cfr. Corduas, 2000).

Tiao and Tsay (1994) richiamano il contributo di Lee (1991) ove viene dimostrato che i parametri di un modello ARMA(1,1) sono consistenti per una procedura di stima adattiva finalizzata alla previsione di una serie modellabile con un FARMA(0,d,0).

Empiricamente si può estendere tale similitudine ad una situazione più generale assumendo che sussista una certa somiglianza tra un FARMA(p,d,q) e un ARMA(p+1,q+1) i cui ordini massimi delle componenti autoregressive e a media mobile vengono aumentati di 1; infatti dall'espressione generale (4) del modello frazionario FARMA(p,d,q) sapendo che la differenziazione frazionaria:

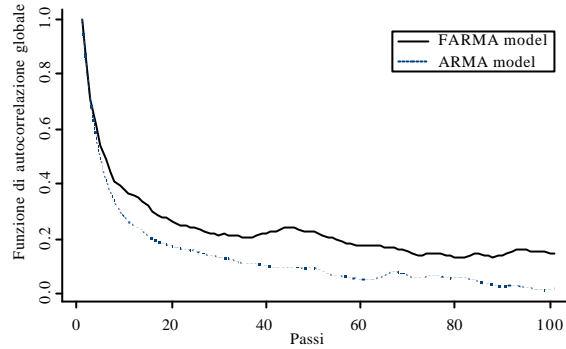


Figura 4: Andamenti delle funzioni di autocorrelazioni delle serie generate dai modelli FARMA(1,d,0) e ARMA(2,1).

$$\nabla^d z_t \approx w_t = \frac{(1 - \tilde{q}B)}{(1 - \tilde{f}B)} Y_t \quad (5)$$

è simile ad un modello ARMA(1,1) si ha che effettivamente la (4) può considerarsi simile alla configurazione seguente:

$$\mathbf{f}(B) \frac{(1 - \tilde{q}B)}{(1 - \tilde{f}B)} Y_t = \mathbf{q}(B) a_t \quad (6)$$

e cioè ad un ARMA(p+1,q+1).

Quindi l'idea alla base del metodo proposto in questo lavoro consiste nel determinare i parametri di un generico modello frazionario misto FARMA(p,d,q) imponendo la condizione che siano quelli che lo rendono più "simile" a quello corrispondente con breve memoria ARMA(p+1,q+1).

Per poter sviluppare un approccio di questo tipo occorre quindi che ci sia la disponibilità di un indice, funzione dei parametri, che quantifichi tale "similitudine". Tale indice si concretizza con l'introduzione del concetto di metrica tra modelli lineari parametrici (Piccolo, 1984; Piccolo, 1990b) che permette di utilizzare la seguente relazione per calcolare la distanza tra due processi X_t e Y_t dei quali sono noti le successioni dei pesi delle rappresentazioni infinite autoregressive $\{\mathbf{p}_x\}$ e $\{\mathbf{p}_y\}$ dei modelli parametrici lineari ad essi associate:

$$d(X_t, Y_t) = \sqrt{\sum_{j=1}^{\infty} (\mathbf{p}_{jx} - \mathbf{p}_{jy})^2} \quad (7)$$

Praticamente, data una serie da analizzare, se si volesse determinare la stima dei valori iniziali dei parametri di un modello FARMA(1,d,2) si dovrebbe stimare prima il modello ARMA(2,3) e poi minimizzare la (7) esprimendo una delle due rappresentazioni infinite in funzione dei parametri del modello frazionario.

Volendo generalizzare la procedura, questa si può sintetizzare in 4 passi:

- stima del modello ARMA con ordini maggiorati di uno rispetto a quello del FARMA corrispondente,
- definizione della rappresentazione autoregressiva infinita del modello ARMA stimato,
- definizione della rappresentazione autoregressiva infinita di un modello FARMA funzione dei parametri d , \mathbf{f} e \mathbf{q} ,
- determinazione del valore dei parametri d , \mathbf{f} e \mathbf{q} che minimizzano la (7).

Per quanto riguarda la rappresentazione autoregressiva infinita, si può procedere stabilendo l'equivalenza tra la formulazione generale di un modello ARMA

$$\mathbf{q}(B)^{-1} \mathbf{f}(B) z_t = a_t$$

e la sua formulazione puramente autoregressiva

$$\mathbf{p}(B) z_t = a_t$$

ottenendo con facili passaggi la seguente relazione ricorsiva:

$$\mathbf{p}_j = \mathbf{f}_j + \sum_{k=1}^q \mathbf{q}_k \mathbf{p}_{j-k} \quad (8)$$

In essa il termine \mathbf{f}_j sussisterà solo per $1 < j < p$.

La definizione della rappresentazione autoregressiva infinita per un modello frazionario è molto simile a quella dei modelli a breve memoria e si ottiene come si può evincere dalla (4) moltiplicando il polinomio (8) con quello ottenibile dal solo operatore differenza frazionario. Quest'ultimo può essere espresso come segue:

$$p_k^f = \sum_{k=0}^{\infty} \frac{G(d+1)}{G(k+1)G(d-k+1)} (-1)^k \quad (9)$$

Per verificare che effettivamente tale strumento possa risultare utile per la stima di modelli frazionari la procedura proposta è stata applicata ad un certo numero di serie sintetiche con parametri noti, constatando l'errore che si commette con tale stima iniziale.

Utilizzando la funzione *arima.fracdiff.sim* presente nel pacchetto software S-Plus si sono generate insiemi di 100 serie di numerosità pari a 10000, con distribuzione gaussiana.

In particolare si è fatto riferimento a tre insiemi di serie storiche generate da: processi frazionari semplici FARMA(0,d,0) e da processi caratterizzati da un solo parametro della componente a breve memoria: FARMA(1,d,0) e FARMA(0,d,1).

Nel primo caso sono stati variati i possibili valori del coefficiente d con un passo di 0.1 considerando (0.1, 0.2, 0.3, 0.4), mentre negli altri due casi si è fissato un valore di d pari a 0.35 e sono stati variati i parametri f e q in modo da formare sei classi (-0.8, -0.5, -0.3, 0.3, 0.5, 0.8).

Alle serie generate è stata quindi applicata la procedura descritta. I valori medi dei parametri stimati, ed il relativo valore dello scarto quadratico medio ottenuto, sono riportati nelle Tabelle 1, 2 e 3.

Come si può constatare i risultati confermano l'efficacia del metodo proposto mostrando degli errori modesti, ribadendo per altro i risultati ottenuti in alcune simulazioni poco dissimili presenti in Corduas (2000).

Tabella 1: Applicazione della tecnica stima iniziale proposta a serie sintetiche gaussiane di tipo FARIMA(0,d,0).

d	Media (\bar{d})	Scarto Q.M. (\bar{d})
0.1	0.094	0.010
0.2	0.193	0.009
0.3	0.290	0.011
0.4	0.380	0.010

Tabella 2: Applicazione della tecnica di stima iniziale proposta a serie sintetiche gaussiane di tipo FARMA(1,0.35,0).

Media(\bar{d})	S.Q.M. (\bar{d})	f	Media(\bar{f})	S.Q.M. (\bar{f})
0.365	0.011	-0.8	-0.82	0.011
0.346	0.014	-0.5	-0.507	0.014
0.332	0.012	-0.3	-0.282	0.012
0.364	0.03	0.3	0.283	0.030
0.593	0.144	0.5	0.283	0.144
0.378	0.045	0.8	0.778	0.050

Tabella 3: Applicazione della tecnica di stima iniziale proposta a serie sintetiche gaussiane di tipo FARMA(0,0.35,1).

Media(\bar{d})	S.Q.M. (\bar{d})	q	Media(\bar{J})	S.Q.M. (\bar{J})
0.285	0.014	-0.8	-0.818	0.006
0.297	0.014	-0.5	-0.536	0.012
0.307	0.017	-0.3	-0.339	0.016
0.356	0.026	0.3	0.311	0.029
0.367	0.037	0.5	0.52	0.037
-0.006	0.016	0.8	0.43	0.016

4. Un'applicazione sulla serie delle portate liquide del fiume Tevere

In questo paragrafo viene presentata la modellazione di una parte della serie storica disponibile delle portate medie giornaliere misurate presso la stazione di Roma-Ripetta. Precisamente si considerano solamente i primi 15 anni di osservazioni per limitare gli effetti sulla serie dell'avvenuta antropizzazione sul bacino idrografico, che ha comportato il manifestarsi di fenomeni di non-stazionarietà (Grimaldi, 2001).

La serie in esame presenta l'andamento riportato in Figura 5 e la struttura di correlazione descritta dalla funzione di autocorrelazione globale riportata in Figura 6a. Come previsto dalle analisi preliminari il primo passo da fare è ridurre la non-gaussianità della serie (cfr. Tabella 4) applicando la trasformazione logaritmica. Gli effetti di tale applicazione sono riportati sempre nella Tabella 4 mentre nella Figura 6b viene riportata la nuova funzione di autocorrelazione globale. Si può notare come la trasformazione produce una certa normalizzazione della serie assieme ad una ovvia variazione della struttura di correlazione.

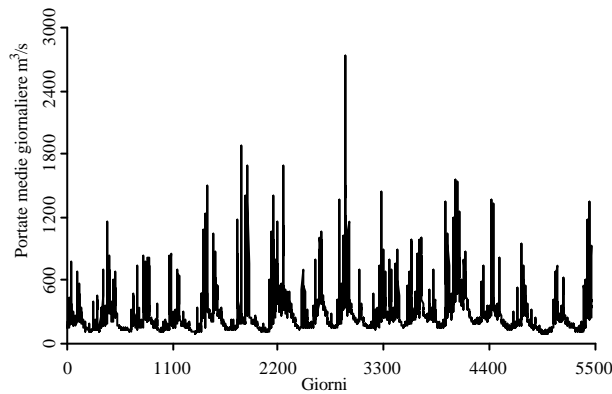


Figura 5. Serie delle portate medie giornaliere del Tevere misurate dal 1930 al 1944.

Tabella 4. Valori di alcuni parametri statistici stimati sulla serie osservata e sulla serie trasformata logaritmicamente.

<i>SERIE</i>	<i>Media</i>	<i>Varianza</i>	<i>Asimmetria</i>	<i>Curtosi</i>	<i>Minimo</i>	<i>Massimo</i>
<i>Osservata</i>	269.5	39515.8	3.62	23.53	92	2730
<i>Trasformata</i>	5.43	0.266	1.003	4.16	4.52	7.91

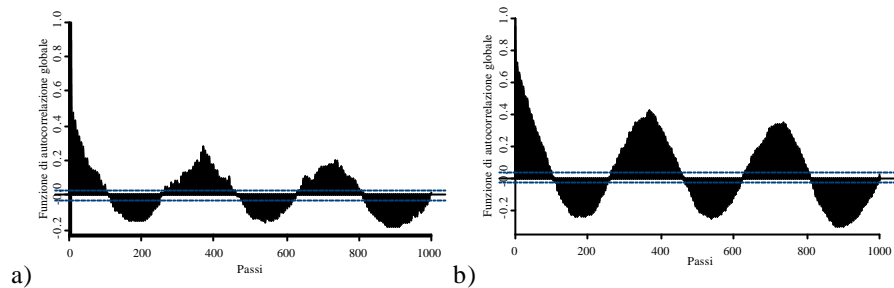


Figura 6. Funzioni di autocorrelazione globale stimate sulla serie osservata (a) e sulla serie trasformata logaritmicamente (b).

Passo successivo da eseguire sulla serie trasformata è la depurazione della componenti stagionali nella media e negli scarti.

Sono state quindi calcolate le componenti periodiche con il metodo *STL Modificato Robusto* (cfr. Figura 7) e si è depurata le serie in esame. Gli effetti di tale approccio sono ben visibili nella funzione di autocorrelazione globale riportata in Figura 8 in cui si nota sia l'assenza di periodicità sia un decadimento della stessa struttura di correlazione che lascia presumere la necessità di un modello frazionario misto.

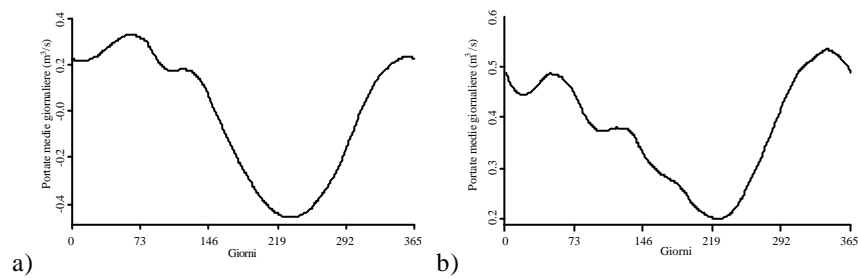


Figura 7. Componenti periodiche nella media e nella varianza stimate sulla serie logaritmica con il metodo *STL-Modificato-Robusto*, Grimaldi e Montanari (2000).

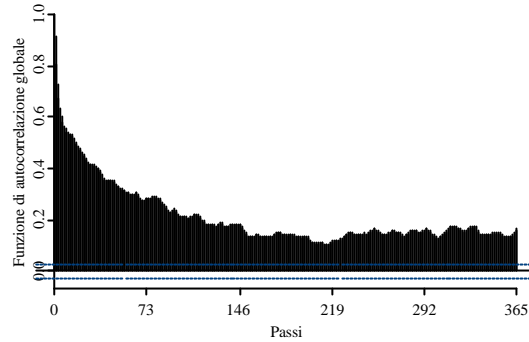


Figura 8: Funzione di autocorrelazione globale stimata sulla serie destagionalizzata.

Per testare in maniera esaustiva il metodo di stima iniziale descritto nel terzo paragrafo vengono di seguito riportati (Tabella 5) i risultati della stima iniziale eseguita con la tecnica tradizionale, con la tecnica di stima preliminare confrontando i valori con quelli ottenuti con la tecnica di stima definitiva (metodo della massima verosimiglianza applicato con l'approssimazione suggerita da Whittle (1953)). Si può notare come effettivamente le stime oltre ad essere vicine al valore definitivo dei parametri presentino comunque un valore del parametro di Hurst variabile con la configurazione. La procedura tradizionale avrebbe mostrato invece un valore fisso del parametro H condizionando negativamente le configurazioni più complesse. Utilizzando l'indice BIC e verificando le successive simulazioni si è identificato il modello "ottimo":

$$(1 - B)^{0.485} X_t = (1 + 0.612B + 0.122B^2) a_t$$

La verifica più efficiente che il modello stimato sia adeguato ad una ipotetica utilizzazione è verificarne le caratteristiche di alcune serie simulate.

Per eseguire la simulazione in maniera tradizionale si procede definendo la rappresentazione infinita a medie mobili:

$$X_t = \frac{\mathbf{q}(B)}{\mathbf{f}(B)(1 - B)^d} a_t = \sum_{s=0}^{\infty} \mathbf{y}_s a_{t-s}$$

Tabella 5. Analisi delle principali configurazioni ipotizzate. Stima iniziale e stima definitiva dei parametri. Test di Portmanteau e Indice BIC.

FARMA (<i>p d q</i>)		<i>H</i>	<i>f</i> ₁	<i>f</i> ₂	<i>q</i> ₁	<i>q</i> ₂	NOTE
1, <i>d</i> , 1	<i>iniz</i>	0.924	0.293	-	-0.374	-	<i>Q</i> =175.3 <i>c</i> =121 <i>BIC</i> =-4425
	<i>fin</i>	0.941	0.285		-0.375		
2, <i>d</i> , 0	<i>iniz</i>	0.959	0.639	-0.186	-	-	<i>Q</i> =199.8 <i>c</i> =121 <i>BIC</i> =-4412
	<i>fin</i>	0.991	0.596	-0.193			
0, <i>d</i> , 2	<i>iniz</i>	0.956	-	-	-0.653	-0.155	<i>Q</i> =185.8 <i>c</i> =121 <i>BIC</i> =-4422
	<i>fin</i>	0.985			-0.612	-0.122	

e generando delle serie sintetiche da un processo WN Gaussiano a_t . La distribuzione empirica dei residui risulta spesso lontana dalla Gaussiana. Nelle serie idrologiche giornaliere in cui la numerosità si presenta elevata (15 anni osservazioni: circa 5000 dati) si ritiene utile far riferimento al metodo descritto da Hipel e McLeod (1994) che suggerisce di generare le serie dei residui sintetici tramite una semplice permutazione dei residui ottenuti dal modello ottimo. In questo modo si otterranno n serie casuali con la medesima distribuzione empirica della serie dei residui ottenuta dal modello stimato.

Procedendo in tal modo sono state generate alcune serie sintetiche e confrontate con la serie originale per le quali in Figura 9 viene proposto il confronto tra le distribuzioni, in Figura 10 il confronto tra le funzioni di autocorrelazione globale e nella Tabella 6 sono confrontati gli indicatori statistici più significativi.

Si può notare come i risultati ottenuti dalle simulazioni sono incoraggianti essendo esse utili per qualsivoglia tipo di applicazione come anche testimoniato da un confronto visivo tra una delle serie simulate (cfr. Figura 11) e la serie reale (cfr. Figura 5).

In fase di generazione si è constatata una non trascurabile influenza della trasformazione inversa del logaritmo sui valori estremi della serie. Ciò è riconducibile al fatto che la trasformazione esponenziale amplia sensibilmente la varianza del segnale generato. Infatti un valore della serie è determinato sia dai parametri del filtro applicato e quindi dalla struttura di correlazione del segnale sia dalla variabilità imposta dalla varianza dei residui ottenuti. Tale varianza viene alterata significativamente dalla trasformazione inversa potendo produrre dei valori molto dissimili dalla serie originaria. Per poter quindi disporre di uno strumento definitivamente affidabile occorrerà analizzare con cura il problema della normalizzazione eseguita nel corso delle analisi preliminari esaminando con attenzione gli effetti delle trasformazioni sulle serie.

Tabella 6. Valori di alcuni parametri statistici stimati sulla serie osservata e sulle serie sintetiche simulate.

<i>SERIE</i>	<i>Media</i>	<i>Varianza</i>	<i>Asimmetria</i>	<i>Curtosi</i>	<i>Minimo</i>	<i>Massimo</i>
<i>Osservata</i>	269.5	39515.8	3.62	23.53	92	2730
<i>Simul. n°1</i>	257.9	26168.8	3.97	37.3	64.1	2829.9
<i>Simul. n°2</i>	264.3	30006	3.95	36.15	69.05	3207.4
<i>Simul. n°3</i>	257.1	26621.2	3.81	29.56	64.44	2298.8
<i>Simul. n°4</i>	263.2	32254.5	4.24	38.01	49.99	2720.6

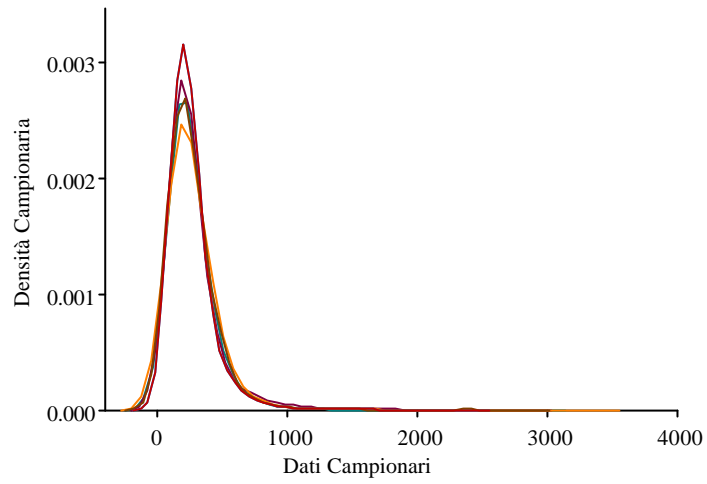


Figura 9. Confronto tra le funzioni di distribuzione della serie osservata e di quattro serie sintetiche simulate.

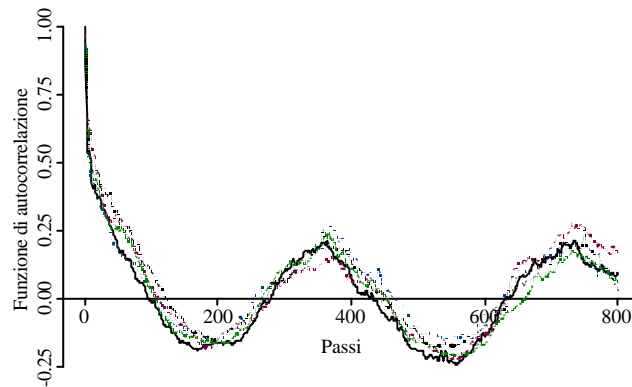


Figura 10. Confronto tra le funzioni di autocorrelazioni globali stimate sulla serie osservata e su quattro serie sintetiche simulate.

Tabella 6. Valori di alcuni parametri statistici stimati sulla serie osservata e sulle serie sintetiche simulate.

SERIE	Media	Varianza	Asimmetria	Curtosi	Minimo	Massimo
Osservata	269.5	39515.8	3.62	23.53	92	2730
Simul. n°1	257.9	26168.8	3.97	37.3	64.1	2829.9
Simul. n°2	264.3	30006	3.95	36.15	69.05	3207.4
Simul. n°3	257.1	26621.2	3.81	29.56	64.44	2298.8
Simul. n°4	263.2	32254.5	4.24	38.01	49.99	2720.6

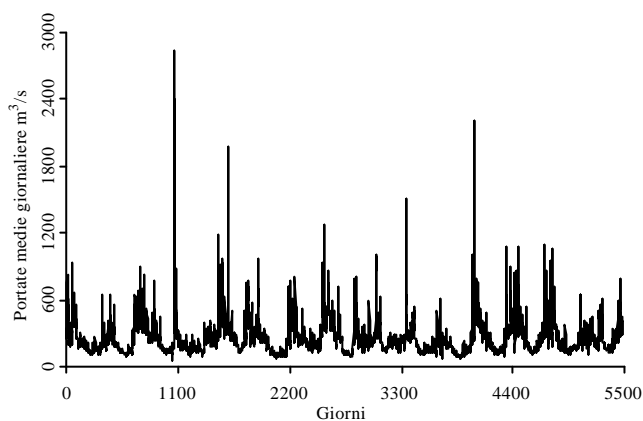


Figura 11. Esempio di serie sintetica simulata

5. Considerazioni finali.

In questo lavoro vengono evidenziati le problematiche che si incontrano applicando i modelli parametrici alle serie idrologiche osservate a scala giornaliera. La maggiore variabilità presente in questo tipo di serie rende necessarie alcune modifiche alla procedura tradizionale per la costruzione dei modelli parametrici lineari.

In particolare sono stati suggeriti accorgimenti relativamente alle analisi preliminari e alla fase di identificazione di tali modelli. Inoltre, viene evidenziata l'inadeguatezza delle trasformazioni di Box e Cox in fase applicativa a causa della distorsione che esse manifestano nella fase di trasformazione inversa.

Confermata la necessità di dover far riferimento a modelli frazionari misti per una corretta parametrizzazione delle strutture di correlazione delle serie giornaliere, è stato descritto e ulteriormente testato un nuovo metodo di stima iniziale basato su una metrica autoregressiva.

L'applicazione presentata, eseguita sulla serie storica delle portate media giornaliere del Tevere, mostra, nonostante la procedura necessiti ulteriori perfezionamenti, dei risultati soddisfacenti.

Ringraziamenti: Il presente lavoro è stato parzialmente finanziato dal CNR-GNDCL.

Riferimenti Bibliografici

- Box G. E. P., Cox D. (1964) An analysis of transformation (with discussion), *Journal of the Royal Statistical Society*, B, 26, 211-246
- Box G. E. P., Jenkins G. M. (1976) *Time Series Analysis : Forecasting and Control*, Holden-Day, San Francisco, California
- Brockwell P.J., Davis R.A (1987) *Time Series: Theory and Methods*, Springer Verlag, New York
- Corduas M. (1999) Modelli a differenze frazionarie: uno studio su dati idrologici, *Quaderni di Statistica*, vol. 1, 83-99
- Corduas M. (2000) A preliminar estimator for ARFIMA models, *XIV COMPSTAT Conference, Utrecht (NL)*
- Grimaldi S., Napolitano F., Ubertini L. (1999) "Extreme value flood analysis with a seasonal stochastic model: a case study", *Proceedings of the IASTED, May 5-8 – Philadelphia, USA*, in stampa

- Grimaldi S. (2000) Stochastic long-memory modelling of seasonal hydrologic time series: a case study, VIII *International Symposium on Stochastic Hydraulics-ISSH-Beijing*, Luglio, 25-28
- Grimaldi S., Montanari A., Su alcune tecniche di stima della componente periodica stagionale presente nelle variabili idrologiche giornaliere”, XXVII *Convegno di Idraulica e Costruzioni Idrauliche*, Genova, Settembre
- Grimaldi S. (2001) Modellazione parametrica di serie idrologiche giornaliere, *Tesi di Dottorato*, Manoscritto non pubblicato.
- Hipel K.W., McLeod A.I. (1994) *Time Series Modelling of Water Resources and Environmental Systems*, Elsevier Science
- Hosking J. R. M. (1984) Modeling persistence in hydrological time series using fractional differencing, *Water Resources Research*, 20, 1989-1908
- Lawrance A.J., Kottegoda N.T. (1977) Stochastic modelling of riverflow time series, *Journal of the Royal Statistical Society, A*, 140, 1-47
- Lee S.A (1991) Long memory time series. *PhD paper*, Graduate School of Business, Chicago
- Montanari A., Rosso R., Taquu M.S. (1997) Fractionally differenced ARIMA models applied to hydrologic time series: identification, estimation and simulation, *Water Resources Research*, 33, 1035-1044
- Montanari A., Rosso R., Taquu M. S. (2000) A seasonal fractional ARIMA model applied to the Nile River monthly flows at Aswan, *Water Resources Research*, 36, 1249- 1265
- Piccolo D. (1984) Una topologia per la classe ARIMA, *Statistica*, 44, 47-59
- Piccolo D. (1990a) *Introduzione all'analisi delle serie storiche*, La Nuova Italia Scientifica, Roma
- Piccolo D. (1990b) A distance measures for classifying ARIMA models, *Journal of Time Series Analysis*, 11, 153-164
- Taqqu M.S., Teverovsky V. (1998) Estimating long-range dependence in finite and infinite variance series, in a *Practical Guide to Heavy Tails: Statistical Techniques for Analysing Heavy Tailed*

- Distributions*, edited by R. Adler, R. Feldman, and M.S. Taqqu,
Birkhauser, Boston, Mass
- Tiao G. C., Tsay R. (1994) Non linear and adaptive modelling in time
series, *Journal of Forecasting*, 13, 338-344
- Whittle P. (1953) Estimation and information in stationary time series,
Arkiv för Matematik, 2, 423-434.