

L'elaboratore elettronico nello studio delle serie temporali

Filtraggio numerico ai minimi quadrati

*(Electronic computer in the study of time series.
Numerical filtering to the last squares)*

R. PASQUALI LARI - P. RANDI (*)

Ricevuto il 24 Gennaio 1972

RIASSUNTO. — In questo lavoro viene discussa la opportunità di utilizzare il filtraggio numerico ai minimi quadrati nello studio delle serie temporali. Si dà la possibilità ai ricercatori di poter utilizzare questa tecnica anche senza specifiche competenze sulla programmazione di calcolatori elettronici. Il ricercatore può altresì ristrutturare a piacere una parte del programma descritto per introdurre condizioni che gli permettano di sfruttare completamente la sua serie di dati tenendo conto delle caratteristiche dei laboratori da cui essi dati provengono.

SUMMARY. — During this work we are concerned with the expedience to make use of numerical filtering to the last squares in the study of time series. The users can utilize this technique also without specific knowledge about programming of electronic computers. The users can also manipulate at pleasure a part of program in order to make the most of their data series, being known the features of the laboratories which the data come from.

1. - TECNICA DI CALCOLO E POSSIBILITÀ.

Il ricercatore può trovarsi nella condizione di avere una serie di dati molto grande e voler vederne le caratteristiche globali, ad esempio estraendo da essa informazioni relative alla esistenza di particolari bande di frequenza. Alcune tecniche numeriche relative a questa impostazione del problema, prendono il nome di filtraggio numerico (2, 3, 4, 6, 7, 8, 9) e ne sono note di diverso tipo e con diverse finalità:

(*) Centro di Calcolo Interuniversitario dell'Italia Nord-Orientale, Casalecchio del Reno, Bologna (Italia).

gli autori hanno scelto di effettuare il calcolo dei coefficienti del filtro numerico come descritto da (2) integrato da (1).

Le tecniche esposte permettono di estrarre bande di frequenza utilizzando i coefficienti nel numero minore possibile e con una risposta in frequenza estremamente vicina a quella ideale; naturalmente essa è in funzione del predetto numero di coefficienti.

Ad ogni modo, poiché il problema a volte impone una separazione tra due tipi di informazioni altrimenti indistinguibili, la tecnica seguita dà quei coefficienti che sono stimati i migliori sotto le condizioni richieste.

Le relazioni su cui si basa il calcolo dei coefficienti sono:

— per il filtro di tipo passa-basse frequenze sotto la condizione che la somma dei coefficienti sia 1, si ha:

$$g_k = \frac{1}{\pi} \left(Si \pi \frac{2k+1}{T} - Si \pi \frac{2k-1}{T} - \frac{2}{2N+1} Si \pi \frac{2N+1}{T} \right) + \frac{1}{2N+1}$$

— per il tipo di filtro passa-banda sotto la condizione che la somma dei coefficienti sia 0, si ha:

$$g_k = \frac{1}{\pi} \left(Si \pi \frac{2k+1}{T} - Si \pi \frac{2k+1}{t} - Si \pi \frac{2k-1}{t} + Si \pi \frac{2k-1}{T} \right) - \frac{2}{\pi(2N+1)} \left(Si \pi \frac{2N+1}{t} - Si \pi \frac{2N+1}{T} \right)$$

— per il tipo di filtro passa-alte frequenze sotto la condizione che la somma dei coefficienti sia 0, si ha:

$$g_0 = 1 - \frac{2}{\pi} Si \pi \frac{\pi}{T} + \frac{2}{(2N+1)\pi} Si \pi \frac{2N+1}{T} - \frac{1}{2N+1}$$

e

$$g_k = \frac{1}{\pi} \left(Si \pi \frac{2k-1}{T} - Si \pi \frac{2k+1}{T} \right) + \frac{2}{(2k+1)\pi} Si \pi \frac{2k+1}{T} - \frac{1}{2N+1}$$

Per la discussione sulla scelta da imporre alla somma dei coefficienti si rimanda al lavoro (2) già citato.

Per rendere accessibile il programma di calcolo a qualsiasi ricercatore per qualsiasi problema di tale natura, gli autori hanno concentrato in poche schede parametri tutte quelle informazioni che sono necessarie al calcolatore per la elaborazione dei dati. Vengono eseguiti in successione: calcolo dei coefficienti del filtro, risposta del filtro alle frequenze

interessate, valutazione statistica dei dati filtrati sulla base della deviazione standard dei dati in ingresso, calcolo della serie residua e preparazione dei dati da graficare. La graficazione è ottenibile mediante un programma (5) che fa parte della biblioteca del Centro di Calcolo.

I dati in ingresso vengono quindi trattati a seconda delle opzioni espresse dai ricercatori (v. descrizione delle schede parametri più sotto) si da costituire un insieme omogeneo: le assenze di dati, inevitabili in lunghe registrazioni per mancanza di funzionamento delle apparecchiature o alterazione delle stesse, possono essere o sostituite col valore della media della serie o con una interpolazione lineare fra gli estremi validi della interruzione (in questo caso la coincidenza fra il valore estremo finale calcolato e quello reale viene fissato dal ricercatore, in forma percentuale, a piacere), o con qualsiasi altra forma che il ricercatore ritenga opportuno. In questo ultimo caso, il ricercatore dovrà però scriversi una subroutine adatta col nome di CHIUS e con i parametri descritti più sotto che sostituisca quella inserita nella versione standard del programma.

I 6 parametri della subroutine CHIUS devono avere il seguente significato:

- P_1 reale - valore dell'ultimo dato precedente al gap
- P_2 reale - valore del primo dato successivo al gap
- P_3 intero - numero d'ordine nella serie elaborata del dato il cui valore è P_1
- P_4 intero - numero d'ordine nella serie elaborata del dato il cui valore è P_2
- P_5 reale - percentuale d'errore (vedi CHIUSX nella sezione «Utilizzazione del programma»)
- P_6 reale - parametro di uscita: matrice dei dati da sostituire all'interno del gap (dimensione massima 500).

La subroutine viene chiamata all'interno del programma ogni qual volta si incontra un gap.

2. - DESCRIZIONE TECNICA DEL PROGRAMMA

Il programma consta di un MAIN e di 5 subroutines, che sono:

COEFIL, SINT, RDWR, CHIUS, MOVE.

Il MAIN legge i parametri necessari alla elaborazione, li controlla e dispone le chiamate alle subroutines a seconda dei problemi che sono posti ed infine sovrintende alle stampe ed alla preparazione dei dati su nastro da utilizzarsi per la graficazione.

COEFIL: calcola i coefficienti del filtro numerico con le modalità richieste dal ricercatore; effettua il controllo sulla somma degli stessi ⁽²⁾ e calcola la risposta del filtro alle frequenze richieste.

SINT: è chiamata da COEFIL per il calcolo dei coefficienti del filtro numerico ⁽¹⁾.

RDWR: prepara i dati della serie temporale che debbono essere elaborati: elimina dati ritenuti non validi e provvede alla sostituzione degli stessi nelle forme previste.

CHIUS: viene chiamata da RDWR quando è richiesta una interpolazione lineare sostitutiva di dati mancanti o ritenuti non validi dal ricercatore. L'interpolazione è giudicata ottima quando la differenza fra il primo dato valido, successivo ai dati mancanti, e quello calcolato dalla interpolazione sono uguali a meno del valore dichiarato dal ricercatore (v. « Utilizzazione del programma »).

MOVE: ripristina le condizioni di lavoro, mantenendo la sequenza temporale, in quanto il programma elabora una serie qualunque di dati, però a gruppi di 2000 per volta.

3. - UTILIZZAZIONE DEL PROGRAMMA.

Le schede contenenti i parametri e tutte le altre informazioni necessarie alla elaborazione della serie dei dati sono 7 e debbono essere *sempre* presenti.

Le prime 4 sono relative al calcolo dei coefficienti del filtro e alla risposta del filtro alle frequenze di prova. Le ultime 3 sono legate al processo di elaborazione dei dati.

Altre schede sono richieste se si verifica la condizione $GAPIX \neq 0$ come è scritto nella nota relativa a questo parametro nella 5ª scheda parametri.

1^a scheda - descrizione del tipo di filtro.
 Formato di lettura (213, 3F10.4.)

COLONNA	NOME DELLA VARIABILE	DESCRIZIONE
1-3	KTF	se = 1 filtro passa basso se = 2 filtro passa alto se = 3 filtro passa banda
4-6	KCOE	numero dei coefficienti del filtro. Essendo simmetrico, occorre indicare soltanto quanti termini stanno da una parte rispetto al termine centrale; se si vuole un filtro di 21 termini KCOE sarà = 10.
7-16	TG	valore del periodo su cui viene fatta la finestra.
17-26	TP	valore del periodo a cui viene chiusa la finestra (questo valore è $\neq 0$ solo quando $KTF=3$).
27-36	SG	valore a cui la somma dei coefficienti deve essere uguagliata. In generale: se $KTF = 2$ o 3 $SG = 0$ se $KTF = 1$ $SG = 1$

2^a scheda - delimita il campo dei test di fiducia.
 Formato = (13)

COLONNA	NOME DELLA VARIABILE	DESCRIZIONE
1-3	NT	è il numero di periodi su cui si fa il test.

3ª scheda - elenco di periodi su cui fare il test di fiducia.
Formato = (8E10.2)

COLONNA	NOME DELLA VARIABILE	DESCRIZIONE
1-n	T	valore dei periodi su cui si chiede il test (questa scheda andrà ripetuta fino ad esaurimento dei periodi richiesti, mettendone 8 per scheda).

4ª scheda - chiude la prima serie di operazioni. Essa *deve* essere una scheda senza perforazioni.

5ª scheda - scelta dei tipi di calcolo e deviazione standard della serie.
Formato = (11, F10.0)

COLONNA	NOME DELLA VARIABILE	DESCRIZIONE
1	IPAR	se = 1 viene calcolata solo la serie filtrata se \neq 1 viene calcolata anche la serie residua
2-9	SIGMA	deviazione standard della serie originale.

6ª scheda - contiene i parametri di lettura e di elaborazione della serie.
Formato = (A3, 212, 11, 1X, 12, 2X, 12, 2X, 11, 1X, F10.0.)

COLONNA	NOME DELLA VARIABILE	DESCRIZIONE
1-3	IDISP	deve essere XXX
4-5	NN	numero dei dati per record (max. 80)
6-7	MT	numero di unità logica di lettura dei dati se = 5 vengono lette schede se = 66 i dati sono su un filo di qualunque nome dichiarato sulla scheda LGO (v. descrizione dell'INPUT più sotto)

Segue 6ª scheda.

COLONNA	NOME DELLA VARIABILE	DESCRIZIONE
8	IZERX	<p>se = 0 gli zeri della serie in INPUT sono reali</p> <p>se \neq 0 gli zeri (o l'assenza di perforazione) della serie in INPUT sono dati mancanti e verranno sostituiti nel corso del programma</p>
10-11	GAPIX (*)	<p>se = 0 non ci sono gaps in lettura</p> <p>se \neq 0 indica il numero di volte che si debbono saltare dati in lettura, dati che verranno sostituiti successivamente (max. 99)</p>
14-15	GAPEX (*)	<p>se = 0 la successione dei dati ottenuta dopo la lettura e l'eventuale sostituzione (GAPIX \neq 0) è completa</p> <p>se \neq 0 la successione ottenuta è incompleta. se ad esempio GAPEX = 1 questo significa che esiste un gruppo di dati fisicamente mancanti; la serie letta va integrata per ripristinare la sequenza temporale</p>
18	CHIUSX	<p>se = 0 nel gap viene inserita la media generale della serie</p> <p>se \neq 0 nel gap vengono inseriti i dati elaborati mediante interpolazione lineare usufruendo del metodo di chiusura</p>
20-29	PERCX	differenza percentuale richiesta dal metodo di chiusura

(*) N.B. - A questi due parametri è associato un certo numero di schede che devono seguire la scheda parametri successiva a questa. Su ciascuna scheda compaiono due valori interi, con formato (2I10), rappresentanti il numero d'ordine, nella sequenza temporale da elaborare, dei dati validi immediatamente esterni ad un gaps da integrare. Relativamente ad un GAPIX = 1 comparirà una scheda contenente ad esempio 25 nelle colonne 9-10 e 42 nelle colonne 19-20; questo significa che in lettura i dati dal 26.mo al 41.mo non sono significativi e debbono essere opportunamente sostituiti. Queste schede debbono essere ordinate per valori crescenti dei limiti dei gaps.

7^a scheda - contiene il formato di lettura dei dati.

Formato = (8A10)

COLONNA	NOME DELLA VARIABILE	DESCRIZIONE
1-80	IFORM	viene scritto il formato di lettura nel normale linguaggio Fortran, comprese le parentesi esterne

4. - DESCRIZIONE DELL'INPUT E OUTPUT.

Il programma è a disposizione dei ricercatori sotto forma di compilato ed è richiamabile mediante la seguente sequenza di schede:

JOB.

REQUEST CIC1.

REWIND CIC1.

COPYN (A, LGO, CIC1)

LGO.

7/8/9

FILTRO, 6, CIC1

7/8/9

⋮

schede parametri e dati

6/7/8/9

L'uscita su tabulato è costituita dalle seguenti informazioni:

- a) tipo di filtro e sue caratteristiche
- b) coefficienti del filtro
- c) test di fiducia
- d) valori della serie filtrata
- e) deviazione standard della serie filtrata
- f) serie residua (se richiesta).

RISPOSTA DEI FILTRI

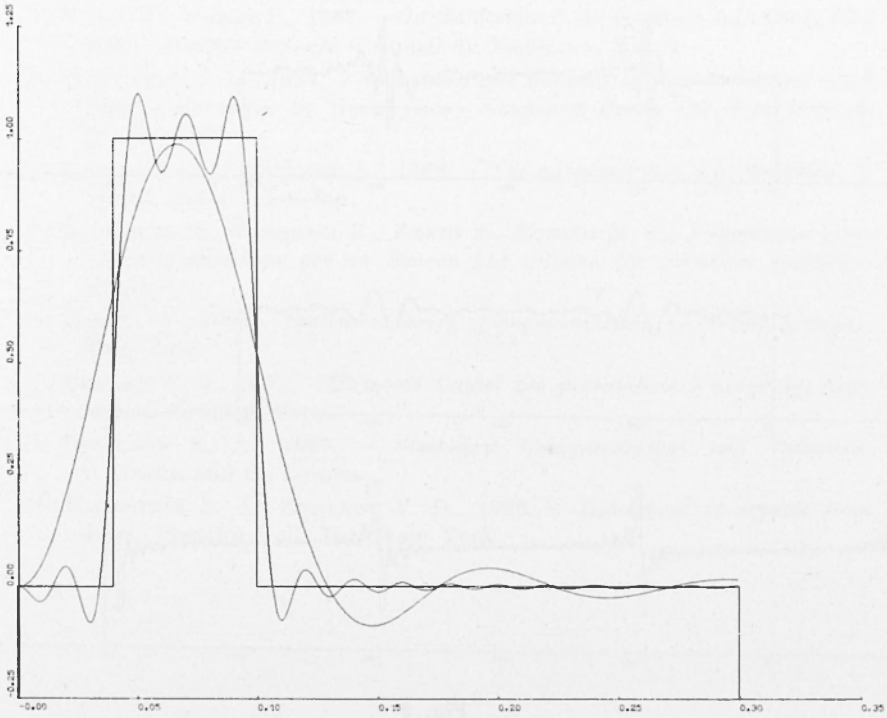


Fig. 1

SERIE RESIDUA / SERIE FILTRATA / SERIE ORIGINALE

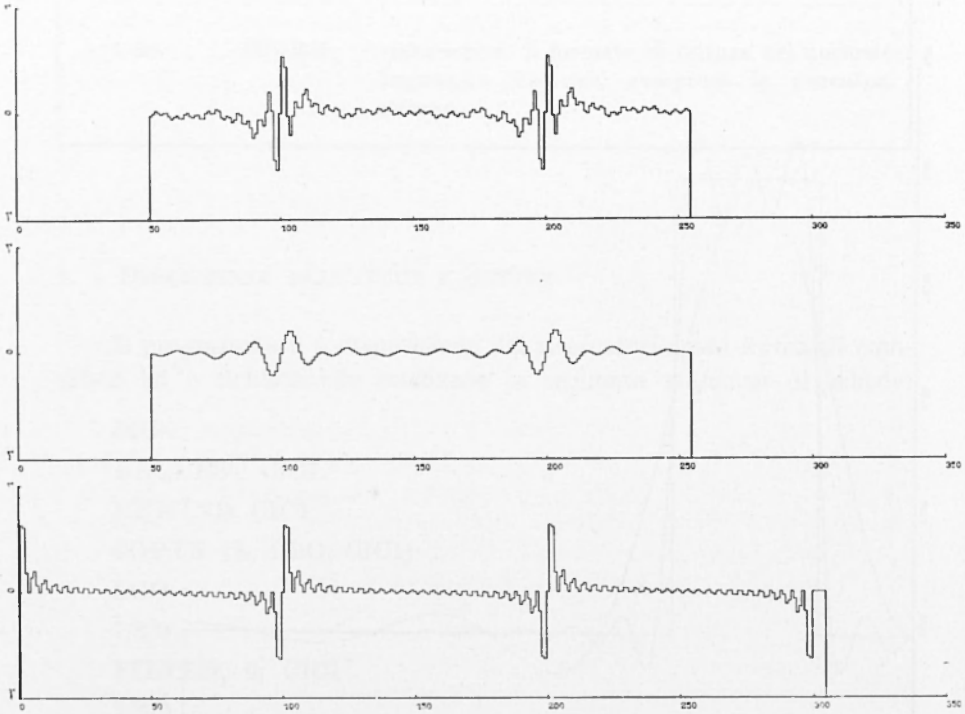


Fig. 2

Nelle figure 1 e 2 si mostrano risultati di elaborazioni ottenute utilizzando il lavoro già citato ⁽⁵⁾.

BIBLIOGRAFIA

- (1) DALL'OLIO L., GUERRIERI V., RANDI P., 1969. - *Coefficients and frequency response computing for every numerical filter*. « Lettere al Nuovo Cimento », **1**, pp. 271-278.
 - (2) GALLI M., RANDI P., 1967. - *On the design of the optimum numerical filter with a prefixed response*. « Annali di Geofisica », **XX**, 4.
 - (3) HOLLOWAY J. L., 1958. - *Smoothing and filtering of time series and space fields*. « Advances in Geophysics » Academic Press. Inc. New York, **4**, pp. 351-389.
 - (4) KENDALL M. G., STUART A., 1969. - *The advanced theory of Statistics*. C. Griffin and C., London.
 - (5) LANZARINI M., PASQUALI R., RANDI P., RONCHI M. E., *Programma generale di graficazione per un Benson 111 off/line* (di prossima pubblicazione).
 - (6) LEE Y. W., 1960. - *Statistical theory of communication*. « J. Wiley & Sons ». New York.
 - (7) PEYRET B. G., 1964. - *Éléments Calcul des probabilités à usage des ingénieurs*. Eyrolles, Paris.
 - (8) ROBINSON E. A., 1967. - *Statistical Communication and Detection*. C. Griffin and C., London.
 - (9) WAINSTEIN L. A., ZUBAKOV V. D., 1962. - *Extraction of signals from noise*. Prentice-Hall, Inc. New York.
-