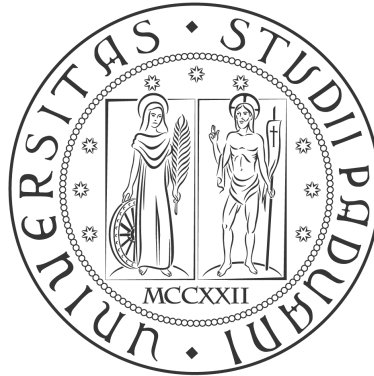


UNIVERSITÀ DEGLI STUDI DI PADOVA

Facoltà di Scienze Statistiche

Corso di Laurea Triennale in Statistica e Gestione delle Imprese



Tesi di Laurea

**Carte di controllo per la sorveglianza
di un contatore gas**

Relatrice: Prof.ssa Giovanna Capizzi

Tutor aziendale: Dott. Francesco Rampazzo

Laureanda: Giulia Angela

Anno Accademico 2009/2010

INDICE

INTRODUZIONE	5
CAPITOLO 1: L'AZIENDA	8
CAPITOLO 2: L' OGGETTO DELL'ANALISI	10
2.1 Tecnica di misura del sensore.....	10
2.2 Alcuni punti della norma (UNI EN 14236)	11
2.3 I vantaggi del contatore intelligente	12
2.4 Strumenti di riferimento.....	12
2.5 Metodo di raccolta dei dati	13
CAPITOLO 3: MODELLI PER L'ANALISI DELLE SERIE STORICHE	14
3.1 Processo Stocastico	14
3.2 Processi Stazionari.....	14
3.3 Funzione di autocorrelazione globale	15
3.4 Funzione di autocorrelazione parziale	15
3.5 L'operatore ritardo	15
3.6 Processo white noise	15
3.7 Processo Moving Average	15
3.8 Processi autoregressivi	16
3.9 Modelli ARIMA.....	16
3.10 Processi stocastici non stazionari	16
3.11 La procedura di Box e Jenkins	18
CAPITOLO 4: CARTE DI CONTROLLO.....	21
4.1 Determinazione della numerosità campionaria	22
4.2 Frequenza di campionamento ottimale	22
4.3 ARL (Average Run Length)	22
4.4 Analisi degli andamenti tipici.....	23
4.5 Stima dei parametri del processo.....	23
4.6 I sottogruppi razionali.....	24
4.7 Le carte di controllo Shewhart	24
4.8 Carta EWMA	26

CAPITOLO 5: APPLICAZIONE AL CASO STUDIO	28
5.0 Scelta dei parametri per le carte di controllo utilizzate	28
5.1 Sensore numero 17	29
5.2 Sensore numero 2	36
5.3 Sensore numero 21	47
5.4 Sensore numero 15	53
5.5 Sensore numero 21A	60
5.6 Sensore numero 22	66
CONCLUSIONI	73
BIBLIOGRAFIA	75
SITI INTERNET CONSULTATI	76

Introduzione

Definire il concetto di qualità non è sicuramente qualcosa di semplice. Questo perché la qualità non è un parametro assoluto, ma è un concetto relativo che può essere spiegato in molti modi. Essa è composta sia da una parte soggettiva, le caratteristiche di un prodotto o servizio volte a soddisfare le aspettative di un cliente, che da una componente oggettiva, conformità alle specifiche tecniche richieste. Lo studioso italiano Fabrizio Colonna nel libro "La Fabbrica nera: aspetti economici della qualità" (1993) definisce la qualità come il rapporto tra le aspettative di un cliente e l'insieme delle funzioni realizzate da un prodotto: più ciò che si aspetta il cliente e il prodotto/servizio si sovrappongono più questo può essere definito come "di qualità".

Tra le definizioni più famose ricordiamo:

"La qualità è l'adeguatezza all'uso"(M.Juran, 1988)

"è la (minima) perdita impartita alla società dal momento che il prodotto stesso lascia la fabbrica" (Taguchi Genichi, 1986)

Applicare il concetto di qualità all'interno di un organizzazione è finalizzato al raggiungimento di diversi obiettivi, tra i quali: dimostrare di saper fornire prodotti o servizi affidabili e rispondenti alle specifiche richieste, aumentare efficacia ed efficienza dei propri processi produttivi, diminuire il numero dei pezzi scarto o che presentano non conformità .

Questo importante concetto non ha sempre avuto questa accezione e si è evoluto con il mutare del mercato e delle esigenze dei clienti. La qualità, intesa nel senso tradizionale, è nata negli anni '20. In questo periodo iniziavano ad essere fondate grandi aziende che necessitavano di un'organizzazione molto più complessa. Il mercato era caratterizzato da grandi volumi e richiesta di prodotti con costi sempre inferiori. È nata, quindi, l'esigenza di sottoporre le variabili del processo produttivo a controlli sempre più rigidi, sono nati, così, i primi metodi statistici per il controllo della qualità.

Nel 1924 Walter Shewhart, che lavorava alla Western Electric, tracciò per la prima volta lo schema di una moderna carta di controllo ed elaborò, inoltre, la teoria delle cause comuni e di quelle speciali di variazione. Le prime sono dovute a moltissime anomalie ben identificabili e sempre presenti. La varianza comune è propria del processo stesso, se in questo sono presenti solo cause comuni di variazione, questo sarà caratterizzato dalla minima variabilità possibile. Le cause speciali, invece, sono accidentali e non prevedibili, e, una volta identificate, devono essere eliminate. Tra le più comuni: errori umani, difettosità degli strumenti, variazioni delle materie prime e modifica delle condizioni ambientali.

È stato il Giappone che per primo iniziò ad applicare un sistema di qualità capace di ridurre in modo sostanziale la possibilità di fare errori. Questo diventò una vera variabile competitiva di successo. Nel 1945 venne pubblicato da un ingegnere americano Armand Feigenbaum il primo articolo sull'applicazione del

metodo *Total Quality Control* sulla rivista *Industrial Quality Control* (1957). Egli spiegò che la responsabilità di riuscire ad ottenere un sistema di qualità adeguato non era solo di un piccolo gruppo di tecnici, ma di tutte le funzioni aziendali che dovevano avere come finalità la soddisfazione del cliente.

Nel 1946 venne fondata la *American Society for Quality Control*, che successivamente prese il nome di ASQ (*American Society for Quality*). Questa organizzazione promosse un nuovo modo di pensare denominato “*statistical thinking*”.

I fondamentali principi sui quali si basa sono:

- Identificare le diverse fasi, tra loro interconnesse, di un processo produttivo poiché queste contribuiscono all’esito del prodotto finale
- Riconoscere l’esistenza della variabilità
- Quantificare e dominare la variabilità per quanto possibile
- Riconoscere che ogni decisione è sempre affetta da un certo grado di incertezza. Le decisioni prese su basi statistiche, però, presentano un’incertezza quantificabile

La presente relazione descrive il progetto svolto durante lo stage nello stabilimento di Padova di SIT La Precisa, all’interno del reparto Ricerca e Sviluppo.

L’obiettivo del lavoro è stato quello di analizzare, con strumenti statistici opportuni, il funzionamento del sensore posto all’interno di un contatore gas. È stato necessario verificare che l’accuratezza di misura del sensore rimanesse sempre all’interno di limiti stabiliti dalla normativa di riferimento (UNI EN 14236). Il fine ultimo di questo lavoro consisteva nel definire i controlli necessari alla certificazione del prodotto.

Il sistema di acquisizione delle misure era supportato da un computer al quale venivano inviati sia i dati provenienti dal sensore che dallo strumento di riferimento. Si decideva una determinata portata espressa in m^3/h e si registravano in contemporanea le misurazioni di entrambi gli strumenti (per maggiori dettagli si veda il capitolo 2 paragrafo 5). Di seguito verrà usato il termine acquisizione per riferirsi a questo processo di raccolta dei dati.

Per verificare la correttezza di un qualsiasi strumento di misura è sempre necessario il confronto con uno strumento di riferimento. In questo caso il parametro utilizzato è stato il rapporto tra misurazioni del sensore e quelle del micromanometro. La presenza di forte correlazione tra le osservazioni ha portato all’adattamento di modelli, come quelli ARIMA, adatti a descriverla.

Il successivo passo compiuto, finalizzato al controllo dello strumento, è stato quello di analizzare i residui dei diversi modelli stimati. Su questi sono stati applicati gli strumenti classici del controllo statistico della qualità (carte di controllo Shewhart e EWMA). Essi hanno permesso di evidenziare eventuali anomalie di ogni singola acquisizione dovute alle più diverse cause. Ad esempio: funzionamento scorretto del sensore

stesso, variazioni di temperatura durante l'acquisizione dei dati, anomalie del programma che acquisisce i dati e instabilità del flusso.

L'organizzazione della tesi è la seguente.

Il capitolo 1 contiene una breve descrizione dell'azienda. Nel capitolo 2 viene introdotto l'oggetto di questa analisi e gli strumenti di riferimento utilizzati durante le acquisizioni. Il capitolo 3 è dedicato ai principali modelli per l'analisi delle serie temporali; particolare attenzione è stata data a test e grafici necessari ad analizzare le principali assunzioni sui residui, la cui analisi risulta fondamentale per il controllo diagnostico del modello scelto. In seguito, nel capitolo 4 saranno introdotti i concetti fondamentali sulla teoria del Controllo Statistico di Qualità (SPC). Verrà dato maggior rilievo alle carte Shewhart \bar{X} e S per sottogruppi razionali (Shewhart, 1931) e alle carte di controllo EWMA (Roberts, 1959) poiché saranno quelle utilizzate nelle analisi del capitolo finale. Nel capitolo 5 vengono messi in pratica tutti i passi descritti precedentemente per la stima dei modelli ARIMA (procedura di Box e Jenkins, 1976). Per ciascun modello vengono verificate le assunzioni fondamentali sui residui. In particolare, vengono utilizzate le due tipologie di carte di controllo applicate ai residui standardizzati dei modelli precedentemente stimati.

Per l'elaborazione dei dati ho utilizzato i software R (versione 2.11.0), Gretl (versione 1.9.0) e Minitab (versione 15).

Capitolo 1

L'azienda

SIT (Società Italiana Tecnomeccanica "La precisa") nasce nel 1953 a Padova, quando iniziava a svilupparsi l'utilizzo del gas per scopi domestici. E' stata fondata dai fratelli Pierluigi e Giancarlo De Stefani. Oggi è un'azienda leader nello sviluppo e nella produzione di sistemi per la sicurezza, il controllo e la regolazione degli apparecchi a gas. Questa è da sempre considerata la missione aziendale.

Fino alla fine degli anni Novanta la Sit è cresciuta internamente attuando una strategia competitiva verso il principale competitor americano. Quando le quote di mercato delle due aziende sono diventate equivalenti, l'azienda ha deciso di modificare le sue modalità di espansione. Da un lato inizia a diversificarsi, tramite acquisizioni di importanti imprese più o meno affini in termini di tecnologia sviluppata. Dall'altro, a partire dal 1997, intraprende il processo di esternalizzazione del processo produttivo. L'obiettivo è quello di creare una gamma di offerta più ampia, in modo da poter offrire sistemi di prodotto sempre più evoluti e complessi. Il fine ultimo consiste nell'avere un portafoglio di aziende leader nel mercato che possano sostenere un intero processo produttivo.

I principali passi compiuti dall'azienda, negli ultimi anni, sono i seguenti:

Nel 1997 apre la filiale in Cina, a Shanghai

Nel 1999 attua l'importante acquisizione di ENCON, azienda olandese produttrice di schede elettroniche per apparecchi a gas.

Nel 2000 apre il nuovo stabilimento di produzione a Monterrey (Messico) e a Shanghai (Cina).

Nel 2001 inizia a diversificarsi con l'acquisizione di OMVL, specializzata in impianti per auto.

Nel 2003 apre una sede produttiva in Argentina.

L'anno successivo avviene l'importante acquisizione del gruppo Natalini, specializzato in elettroventilatori e kit scarico fumi.

Nel 2005 nasce la filiale produttiva a Rovigo. L'esternalizzazione continua con la nascita di una sede produttiva a Brasov (Romania).

L'ultima, un anno fa, a Hoogeveen (Olanda).

Ad oggi Sit è organizzata in tre divisioni: Sit, Natalini e OMVL.

Le principali caratteristiche del gruppo sono:

- Posizione di leadership nel mondo per caldaie domestiche, stufe termostatiche, caminetti alimentati a gas
- 1430 dipendenti di trenta diverse nazionalità
- 170 unità dedicate alla ricerca e sviluppo nelle quali è impegnato il 10% del personale

- 9 siti produttivi(in Italia, Olanda, Romania, Messico e Cina) e 17 sedi commerciali, tra filiali, uffici vendite e agenzie

Sit è presente in numerosi paesi nel mondo e questo le ha permesso di creare relazioni sempre più strette con i propri clienti e fornitori. Inoltre questa strategia ha permesso di creare un know-how comune, in grado di garantire elevati standard di qualità che si estendono in tutti i centri SIT.

Il gruppo produce ogni anno oltre otto milioni di controlli meccanici e un milione di controlli elettronici ai quali si aggiungono sensori, ventilatori e dispositivi per l'alimentazione a gas di autoveicoli. Uno degli obiettivi aziendali è garantire sicurezza, efficienza e precisione anche se si producono grandissimi quantitativi di pezzi. Per questo motivo Sit è stata tra le prime aziende ad ottenere la certificazione ISO 9001 per il sistema di qualità e, dal 2001, è certificata VISION 2000.

Capitolo 2

L' oggetto dell'analisi

Gli attuali contatori gas a pareti deformabili si basano sul continuo svuotamento e riempimento di due contenitori che racchiudono volumi noti di gas detti camere di misura. Viene misurato quante volte questa operazione ciclica viene compiuta tramite un display meccanico.

Per ottenere il vero valore del volume, in condizioni standard, è necessario operare una compensazione in base ai valori di pressione e temperatura.

Nell'ottobre del 2008 l'Autorità dell'Energia ha emanato una direttiva (ARG 155/08), che impone, entro dicembre 2012 la sostituzione di almeno l'80% dei contatori gas attuali con quelli di tipo *smart* che permettono la telelettura e telegestione del contratto.

Il contatore progettato è costituito da: un sensore di misura, una valvola che permette di regolare il flusso, una batteria a lunga durata ed alcuni componenti elettronici per la lettura e l'invio dei dati.

2.1 Tecnica di misura del sensore

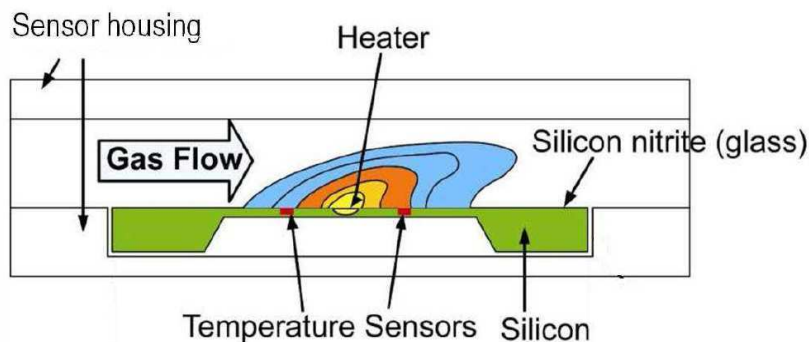


Figura 2.1 schema di funzionamento del sensore

In una sezione della tubatura principale è posta una resistenza elettrica (*heater*). Questa ha la funzione di riscaldare il flusso di gas che scorre. Questo meccanismo altera la temperatura del gas, così che la temperatura a valle risulta più alta della temperatura a monte. La differenza viene misurata tramite un sensore di temperatura. Essa è inversamente proporzionale al flusso di massa secondo la seguente relazione:

$$\frac{d(m)}{d(t)} = \frac{Q}{C_p(T)(T_{valle} - T_{monte})}$$

dove:

Q = calore

$C_p(T)$ = calore specifico a pressione costante

T = temperatura

Ad oggi non esistono normative specifiche per contatori che utilizzano misuratori di flusso massici. La norma europea che più si adatta al contatore progettato è quella per i misuratori di gas domestici ad ultrasuoni (UNI EN 14236) emanata nel 2004 ed aggiornata nel 2007.

Affinchè il contatore possa essere prodotto, commercializzato e certificato conforme per effettuare misurazioni a fini fiscali deve soddisfare i requisiti richiesti dalla MID 2004/22/CE (*Measuring Instrument Device*). Questa, però, non definisce con precisione le specifiche tecniche da rispettare.

Per questo motivo la norma qui utilizzata come riferimento è la UNI EN 14236.

2.2 Alcuni punti della norma (UNI EN 14236)

Consideriamo solo i punti della norma relativi all'accuratezza e agli errori permessi ai contatori gas domestici ad ultrasuoni:

- Quando il valore di Q(portata) è compreso tra Q_{min} (minima portata calcolata in m^3/h per il quale il contatore è stato progettato pari in questo caso a $0,04 m^3/h$) e Q_t (10% di Q_{max}) l'errore medio massimo permesso sul volume misurato è del $\pm 3\%$.
- Quando il valore di Q è compreso tra Q_t e Q_{max} (massima portata calcolata in m^3/h per il quale il contatore è stato progettato pari, in questo caso, a $6 m^3/h$) l'errore medio massimo permesso è del $\pm 1,5\%$

Riassumendo:

Se $[0,04 \leq Q \leq 0,6] m^3/h \rightarrow \epsilon_{max} = 0,03$

Se $[0,6 \leq Q \leq 7,2] m^3/h \rightarrow \epsilon_{max} = 0,015$

Nel primo caso dunque la massima differenza tra gli errori permessa è del 4% mentre nel secondo caso è del 2%.

La norma richiede, inoltre, che vengano effettuati 6 test in aria a ciascuno dei seguenti andamenti di flusso in ordine crescente o decrescente:

Q_{min} , $3Q_{min}$, $5Q_{min}$, $10Q_{min}$, $0,1Q_{max}$, $0,2Q_{max}$, $0,4Q_{max}$, $0,7Q_{max}$, Q_{max} .

La stessa tipologia di test va effettuata anche sui gas della seconda famiglia(gruppo H, L).

La differenza massima tra errore medio calcolato in aria e con ciascuna tipologia di gas per ogni andamento di flusso deve essere del 3% se il valore di Q appartiene al primo caso mentre nel secondo caso deve essere del 1,5%.

Il contatore progettato è in grado di funzionare ad una temperatura minima di $-10^\circ C$ fino a una temperatura massima di $55^\circ C$. La norma impone, inoltre, che il sensore debba attivarsi a $0,01 m^3/h$ (0,25% di Q_{max}) e il contatore debba registrare il passaggio di gas.

E' necessario testare il contatore anche a $7,2 \text{ m}^3 / \text{h}$ (1,2% di Qmax) e controllare che l'errore medio sia compreso nei limiti sopra citati. Quando il contatore funziona in modo normale il periodo di campionamento non deve superare i 2 secondi, a meno che il costruttore non dimostri che un periodo più lungo non comporti una perdita di accuratezza nella misura.

Il contatore analizzato appartiene alla categoria G4 ed è stato progettato per l'uso domestico.

2.3 I vantaggi del contatore intelligente

- Conoscenza diretta e immediata dei propri consumi da parte dell'utente che può anche memorizzare i consumi giornalieri
- Corretta attribuzione dei consumi effettivi al cliente
- Incremento del risparmio energetico da parte dell'utilizzatore
- Rilevazione di eventuali perdite e/o malfunzionamenti da parte del distributore tramite controlli da remoto
- Riconoscimento di diversi tipi di gas ai quali associare diverse tariffe
- Comunicazione immediata al centro dei dati di consumo e di diagnostica
- Interruzione della fornitura ad un cliente moroso
- Mancanza del pericolo dell'effetto di usura che, con il passare degli anni, comporta una minore accuratezza delle misurazioni
- Assenza del rischio che i cambiamenti di volume siano dovuti a variazioni termiche e/o di pressione
- Dimensioni più contenute rispetto ai contatori domestici ad oggi utilizzati

2.4 Strumenti di riferimento

Durante le acquisizioni sono stati utilizzati due strumenti di riferimento. Essi hanno un ruolo fondamentale per determinare e migliorare l'accuratezza del sensore.

1) BIOS DRY CAL è uno strumento di misura del flusso primario. L'errore massimo dichiarato dal produttore è pari a $\pm 0,25\%$. Il tempo impiegato dal pistone ad attraversare una distanza nota all'interno del cilindro è misurato da un processore interno. Il pistone e il cilindro sono fatti rispettivamente di graffite e borosilicato per il loro basso coefficiente di espansione termica e per il basso attrito che provocano. Il flusso di gas viene trasformato in flusso di massa tramite opportuna calibrazione che dipende da temperatura e pressione lette da opportuni sensori. Il tempo che il pistone impiega a muoversi lungo una distanza nota (che implica un volume noto) permette di ottenere la misura della portata (Q) come:

$$Q = V/T = \pi r^2 h/T$$

Dove:

V = volume misurato

T =tempo misurato
R =raggio del cilindro
H =altezza del percorso fatto dal pistone

- 2) Il MICROMANOMETRO è uno strumento molto sensibile anche per la misura di bassissimi differenziali di pressione. Permette inoltre di leggere la temperatura esterna dell'aria, il flusso di volume, la temperatura del flusso e la pressione. L'errore massimo dichiarato dal produttore è, in valore assoluto, pari ad 1.

Il primo strumento di riferimento descritto ha dei tempi di acquisizione molto più elevati rispetto al secondo e al sensore. Per questo motivo la procedura seguita, ai fini di una corretta calibrazione, è la seguente:

- Confronto tra valori del flusso, a portata costante, letti dal BIOS DRY CAL e dal MICROMANOMETRO.
- Comparazione tra misurazioni del MICROMANOMETRO e del sensore

Sul secondo punto si baseranno le analisi esposte nel capitolo 5.

2.5 Metodo di raccolta dei dati

Per ogni sensore è stata fatta un'acquisizione con modalità a gradini. Si impostano i valori dei gradini all'interno di un programma sul computer. Il sistema impiega del tempo a fare in modo che la portata di ogni gradino rimanga stabile e, solo quando questo avviene, inizia le misurazioni. Vengono quindi acquisiti in contemporanea sia i valori del sensore che quelli del micromanometro (strumento usato come riferimento). Vengono registrate 50 misurazioni all'interno di un singolo gradino, una ogni 0,5 secondi.

Ogni acquisizione è composta da una rampa ascendente e da una discendente. Se, ad esempio, la rampa in salita è formata da tre gradini la cui portata è di 1,5; 2; 3 m³/h, allora quella in discesa sarà composta da altri tre gradini di 3; 2; 1,5 m³/h. Questa acquisizione, ad esempio, è composta da 300 osservazioni provenienti dal sensore e altrettante 300 dal micromanometro.

Il parametro analizzato è il rapporto tra i valori di flusso letto dal sensore e quelli misurati dallo strumento di riferimento.

Le osservazioni sono risultate serialmente correlate. In seguito verrà descritta la correlazione con opportuni modelli di serie temporali. L'obiettivo è quello di individuare un modello probabilistico adatto a descrivere il processo generatore dei dati.

Capitolo 3

Modelli per l'analisi delle serie storiche

Nel seguito viene presentato l'approccio "moderno", contrapposto a quello classico, dell'analisi delle serie storiche. L'evoluzione temporale di una variabile presenta sia una componente sistematica, $f(t)$, che una puramente casuale detta stocastica, u_t . Nell'approccio classico viene concentrata l'attenzione su $f(t)$ ed u_t è considerato un processo trascurabile perché a componenti incorrelate ($Cov[u_r, u_s]=0, \forall r, s, r \neq s$). In quello moderno, invece, $f(t)$ viene convenientemente modellata e si analizza la parte residua per verificare la bontà dell'adattamento. Per quest'ultima si ipotizza che $Cov[u_r, u_s] \neq 0$, per qualche $r \neq s$.

3.1 Processo Stocastico

Intuitivamente un processo stocastico è un insieme ordinato di variabili casuali indicizzate secondo il parametro t , detto tempo. La serie storica, quindi, è un campione osservato in un dato istante (una realizzazione) del processo stocastico sottostante (universo di possibili realizzazioni).

In modo rigoroso è necessario definire uno spazio di probabilità (Ω, F, P) , dove Ω è lo spazio di misura degli eventi, F una σ -algebra su Ω e P è una misura di probabilità, e uno spazio parametrico T .

Un processo stocastico è una funzione $X(t, \omega)$ finita a valori reali che, per ogni $t \in T$, è una funzione misurabile di $\omega \in \Omega$ e ω è definito come evento che specifica su Ω quale risultato si verifica per t fissato.

3.2 Processi Stazionari

A livello intuitivo, una serie è stazionaria se, qualora perturbata da shock, ritorna al suo equilibrio di lungo periodo una volta terminati questi effetti. Ovvero se la sua media e varianza non presentano cambiamenti sistematici. La definizione rigorosa è che un processo stocastico si dice stazionario in senso stretto se le distribuzioni di probabilità congiunte rispettivamente di (Y_{t1}, \dots, Y_{tn}) e di $(Y_{t1+s}, \dots, Y_{tn+s})$ non dipendono da s . La condizione di stazionarietà in senso stretto impone vincoli sui momenti di qualsiasi ordine.

La stazionarietà in senso debole, invece, impone condizioni solo sui primi due momenti. In particolare, la media e la varianza sono costanti e la funzione di autocovarianza dipende solo dal ritardo k e non dal tempo t . Quindi:

$$E[Y_t] = \mu \quad \forall t$$

$$Var[Y_t] = \sigma^2 \quad \forall t$$

$$Cov[Y_t, Y_{t+k}] = \gamma_k \quad \forall t, \forall k$$

3.3 Funzione di autocorrelazione globale

La funzione di autocorrelazione (ACF) globale, che misura al variare di k la correlazione che esiste tra Y_t e Y_{t-k} , è pari a:

$$\rho_k = \frac{\text{cov}(Y_t, Y_{t-k})}{\sqrt{\text{var}(Y_t) \text{var}(Y_{t-k})}}$$

Il grafico degli indici di correlazione si chiama correlogramma.

3.4 Funzione di autocorrelazione parziale

La funzione di autocorrelazione parziale (PACF) misura la correlazione tra Y_t e Y_{t-k} al netto della dipendenza lineare con le variabili intermedie:

$$P_k = \text{Corr}(Y_t, Y_{t-k} | Y_{t-1}, \dots, Y_{t-k+1})$$

3.5 L'operatore ritardo

L'operatore ritardo L per la variabile Y al tempo t è definito da: $Ly_t = y_{t-1}$. Iterando l'applicazione di questo operatore k volte si ha $L^k y_t = y_{t-k}$

3.6 Processo white noise

Un processo stocastico il cui t -esimo elemento indicheremo con ε_t si definisce white noise, se:

- $E(\varepsilon_t) = 0 \quad \forall t \in T$
- $V(\varepsilon_t) = \sigma^2 < \infty$
- $\text{Cov}(\varepsilon_t, \varepsilon_{t-k}) = 0 \quad \forall k \neq 0$

È un processo casuale senza relazione tra eventi in diversi istanti nel tempo. Esso è stazionario in senso debole ma non in senso forte.

3.7 Processo Moving Average

Un processo $MA(q)$ è ottenuto come una somma pesata di shock casuali riferiti al passato. Si applica, quindi, un polinomio di ordine q all'operatore ritardo che, a sua volta, è stato applicato ad un white noise.

$$y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} + \theta_3 \varepsilon_{t-3} - \dots - \theta_q \varepsilon_{t-q}$$

quindi, scrivendolo in forma ridotta otteniamo:

$$y_t = 1 + \theta_1 L + \dots + \theta_q L^q$$

Un processo $MA(q)$ è sempre stazionario perché ha media zero e varianza e covarianza non dipendenti dal tempo.

La funzione di autocorrelazione di questo processo si annulla quando k risulta maggiore di q ed ha un andamento che dipende dal valore dei parametri.

3.8 Processi autoregressivi

I modelli lineari che assumono che la componente prevedibile del modello sia una somma pesata dei valori passati e di un white noise sono detti autoregressivi di ordine p , numero di parametri in esso presenti.

$$y_t = \Phi_0 + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \Phi_3 y_{t-3} + \dots + \Phi_p y_{t-p} + \varepsilon_t$$

Il modello si può riscrivere nella seguente forma:

$$1 - \Phi_1 L - \Phi_2 L^2 - \Phi_3 L^3 - \dots - \Phi_p L^p = \varepsilon_t$$

Un processo AR(p) è stazionario se tutte le p radici dell'equazione caratteristica $\Phi(L)=0$ sono, in modulo, maggiori di uno.

Si può inoltre mostrare che la funzione di autocorrelazione globale di un processo AR(p) stazionario tende a zero al divergere di k . Invece l'autocorrelazione parziale è diversa da zero se k è minore o uguale a p e si annulla per k maggiore di p .

3.9 Modelli ARIMA

La combinazione dei due modelli AR(p) e MA(q) porta alla classe dei modelli ARMA(p,q). Questa classe è molto utile quando il modello AR(p) o MA(q) risultano essere sovradimensionati, in questi casi usare un modello ARMA rende la scelta più parsimoniosa. Modelli di questo tipo sono stati proposti nel 1976 da Box e Jenkins.

La formulazione è la seguente:

$$\Phi_p(L)y_t = \theta_q(L)\varepsilon_t + \eta$$

dove η è una costante nota pari alla media del processo.

Un processo ARMA(p,q) è stazionario se tutte le p radici dell'equazione caratteristica associata alla componente AR sono in modulo maggiori di uno.

Per k maggiore di q l'ACF tende ad annullarsi, mentre la PACF tende a zero per k maggiore di p .

Un modello ARIMA(p,d,q) è analogo ad un modello ARMA(p,q) applicato alle differenze di ordine d della serie dei valori.

3.10 Processi stocastici non stazionari

Fino ad ora sono stati considerati esclusivamente processi stazionari. La non stazionarietà è spesso presente in una serie storica e può essere osservata:

- in media per la presenza di un trend lineare o quadratico o per la presenza di stagionalità
- in varianza quando questa varia in funzione del tempo t

Se ci si trova in una di queste situazioni o entrambe è necessario procedere tramite opportune trasformazioni.

Nel primo caso si assume che la non stazionarietà in media sia di tipo omogeneo e, per questo motivo, potrà essere rimossa tramite opportuni operatori differenza.

Se X_t è un processo non stazionario e se $E(X_t - X_{t-k}) = vk$, dove k è la distanza tra X_t e X_{t-k} e v è una costante, risulta utile utilizzare un operatore differenza. Nel caso sia presente un trend lineare è sufficiente un operatore differenza di ordine uno. Se, invece, il trend è di tipo quadratico l'ordine opportuno di differenziazione è il secondo. Nel caso la serie sia dotata di una componente stagionale, si usa un operatore differenza di ordine s pari alla periodicità stagionale.

Nel secondo caso può risultare utile usare una trasformata appartenente alla classe di Box e Cox definita come:

$$T(Y_t) = \begin{cases} \frac{Y_t^\lambda - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \ln(Y_t) & \text{se } \lambda = 0 \end{cases}$$

Il parametro λ può essere stimato tramite il metodo della massima verosimiglianza. Sotto l'assunzione di normalità distributiva dei residui si massimizza, rispetto a λ , la funzione di log-verosimiglianza del processo.

Se si considera l'ipotesi (Bartlett, 1947) secondo cui il legame tra la deviazione standard e la media sia il seguente:

$$\sigma = K\mu^{(1-\lambda)}$$

allora è possibile individuare un valore approssimato per λ tramite una rappresentazione grafica.

Se risulta $\lambda=1$ significa che non è necessaria alcuna trasformazione poiché σ è costante. Se $\lambda=0$ la trasformazione più appropriata è quella logaritmica perché la varianza è proporzionale alla media.

Per stimare il valore di questo parametro e' possibile, ad esempio, utilizzare il metodo grafico di Box - Cox. Si divide la serie in sottogruppi e per ciascuno di essi si calcola la deviazione standard e la media. Dall'andamento dei punti si può stimare un valore approssimato per λ .

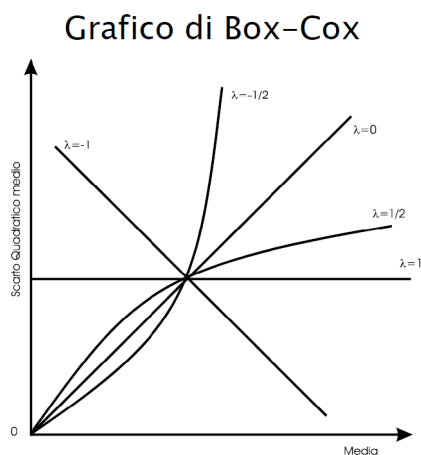


Figura 3.1

Il programma Gretl, utilizzato per le analisi dei dati nel capitolo finale, permette di costruire il range-mean plot. In particolare, viene divisa la serie in piccoli sottogruppi di grandezza arbitraria k . Per ciascuno di essi viene calcolata la media e il range, posti rispettivamente sull'asse delle ascisse e su quello delle ordinate. Quindi ciascun sottogruppo viene rappresentato da un punto sul grafico. Se la varianza della serie è costante ci si aspetta che i range dei sottogruppi siano indipendenti dalle medie degli stessi. Se invece il grafico mostra una retta, stimata con il metodo dei minimi quadrati, vuol dire che la pendenza di essa è significativa almeno al 10%. Il modello di regressione lineare semplice utilizzato ha come variabile dipendente i range campionari dei sottogruppi e come variabile indipendente le rispettive medie. Se il coefficiente angolare risulta positivo ne consegue che la varianza della serie cresce al crescere della media, viceversa se la pendenza è negativa.

3.11 La procedura di Box e Jenkins

Questa è una procedura di tipo iterativo per l'identificazione e l'adattamento di un modello ARIMA. Può essere riassunta nelle quattro seguenti fasi:

1. ANALISI PRELIMINARE che consiste nell'individuare le trasformazioni più appropriate da apportare alla serie affinché risulti rappresentabile per mezzo di un processo stazionario.
2. IDENTIFICAZIONE DEL MODELLO e dell'ordine dei parametri p e q . Il parametro differenza d viene scelto nella fase precedente. Qui si cerca di riconoscere tramite l'andamento della funzione di autocorrelazione empirica la struttura di una funzione di autocorrelazione teorica.
3. STIMA DEI PARAMETRI che si ottiene massimizzando la funzione di verosimiglianza. Valgono i risultati asintotici standard. Gli stimatori sono consistenti e asintoticamente normali.
4. VERIFICA DEL MODELLO che ha come obiettivo quello di garantire che l'identificazione e la stima del modello siano state effettuate in modo adeguato. Verranno analizzati i residui, la bontà dell'adattamento e la significatività dei parametri. Vengono spesso utilizzati i criteri di Akaike e Schwartz (*Asymptotic Information Criterion*, Akaike, 1974; *Schwarz Criterion*, Schwarz, 1978). Questi ci permettono di scegliere tra modelli alternativi. Viene considerato il modello "migliore" quello che presenta il valore più basso di entrambi gli indici. Ricordiamo che il criterio di Schwartz penalizza in misura maggiore l'introduzione di coefficienti nella stima del modello.

Nel seguito, particolare attenzione verrà posta all'analisi dei residui. Questa è molto importante per determinare se la serie può essere adeguatamente descritta dal modello ARIMA ottenuto. Le prime tre fasi dell'adattamento hanno come obiettivo quello di ottenere residui che siano:

- I. casuali
- II. incorrelati e indipendenti
- III. normalmente distribuiti

IV. a varianza costante

I test e le analisi grafiche utilizzate per accertarsi che queste ipotesi siano soddisfatte sono:

- I. Il grafico della serie dei residui che non dovrebbe mostrare alcun tipo di regolarità e i valori dovrebbero oscillare in modo casuale attorno allo zero.
- II. La serie dei residui può essere considerata come una serie storica a sè stante e, per ogni valore di k, e' possibile calcolare $\hat{\rho}_k$.

Se $\hat{\rho}_k \approx 0$ per $k > q$ Bartlett ha dimostrato che vale la seguente approssimazione (Bartlett, 1946):

$$\text{var}(\hat{\rho}_k) \cong \frac{1}{n} \sum_{i=-q}^q \rho_i^2$$

È stato inoltre dimostrato da Quenouille (Box e Jenkins, 1976, p.65) che la varianza di \hat{P}_k si approssima

ad $\frac{1}{n}$. Quindi, sotto l'ipotesi nulla che $\hat{\rho}_k$ sia zero, si ha che

$$\text{var}(\hat{\rho}_k) \cong \frac{1}{T}$$

Si respinge quindi l'ipotesi nulla se

$$\rho_k \notin \left\{ \frac{-1,96}{\sqrt{T}}, \frac{+1,96}{\sqrt{T}} \right\}$$

con un livello di significatività osservato dello 0,05%.

Una statistica test che può essere utilizzata per verificare l'assenza di correlazione e' quella che utilizza il test di tipo Portmanteau sviluppato da Ljung e Box (1978)

La statistica test e' la seguente:

$$Q(m) = n(n+2) \sum_{k=1}^m \frac{1}{n-k} \hat{\rho}_k^2$$

dove m e' il numero di autocorrelazioni prese in esame.

Sotto l'ipotesi nulla che i residui del modello ARMA(p,q) siano realizzazione di un processo white noise, la statistica Q si distribuisce come una variabile casuale χ^2 con m-p-q gradi di libertà.

- III. Poichè per le variabili aleatorie gaussiane, come lo sono i residui, l'incorrelazione implica l'indipendenza, è fondamentale verificare che questi siano distribuiti normalmente.

Esistono diverse possibilità, alcune basate sulla costruzione di opportuni grafici (ad esempio quello "quantile-quantile") ed altre affidate all'uso di particolari test statistici.

Uno strumento spesso utilizzato per verificare la normalità della distribuzione è il Q-Q plot. Esso confronta i quantili della distribuzione campionaria con quelli di pari livello della normale. Se i punti, rappresentanti le coppie di quantili, si dispongono attorno alla retta $y=x$ allora si può dire che i dati

sono approssimativamente gaussiani. La pendenza e l'intercetta forniscono informazioni utili per la stima di μ e σ . Infatti $y = \mu + \sigma x$.

Nelle analisi successive è stato utilizzato il test di Shapiro Wilk. La statistica test W , può assumere valori da 0 a 1, e' così calcolata:

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

e confrontano due stimatori alternativi della varianza. Uno non parametrico al numeratore e la varianza campionaria al denominatore che invece è uno stimatore parametrico (i pesi per la combinazione lineare sono disponibili su apposite tavole).

Particolare attenzione è stata data anche al test di Jarque Bera (1980) che si basa sulla misura sia dell'asimmetria che della curtosi di una distribuzione.

La statistica test è la seguente:

$$JB = \frac{n}{6} \left(S^2 + \frac{(K - 3^2)}{4} \right)$$

dove n è il numero delle osservazioni, S è l'asimmetria e K è la curtosi del campione.

L'ipotesi nulla è un'ipotesi congiunta che sia asimmetria che curtosi siano nulle. Sotto tale ipotesi, la statistica JB è distribuita asintoticamente come una variabile casuale chi quadro con due gradi di libertà.

IV. È infine necessario verificare l'ipotesi di omoschedasticità dei residui.

Essi possono essere definiti tali se la varianza degli errori non è funzione lineare del tempo secondo la seguente relazione:

$$\ln(\sigma_t^2) = a + bt$$

Se i residui sono eteroschedastici la varianza aumenta o diminuisce al variare di t , a seconda del segno di b .

Un metodo grafico per accertarsi della presenza di omoschedasticità si basa sul costruire un diagramma di dispersione che riporti sull'asse delle ascisse i valori previsti dal modello e su quello delle ordinate i residui. I punti rappresentati devono seguire un andamento del tutto casuale. In caso di andamenti sistematici il modello dovrà essere modificato.

Capitolo 4

Carte di controllo

Ogni processo produttivo, per quanto sia stato progettato nel migliore dei modi, presenta sempre una variabilità naturale ineliminabile dovuta a cause dette comuni che agiscono casualmente. Se all'interno del sistema sono presenti esclusivamente errori casuali il processo si definisce sotto controllo statistico.

Esiste invece un'altra famiglia di errori, detti sistematici, che alterano la variabilità naturale. Le cause che generano questi cambiamenti sono dette speciali e, aggiunte a quelle comuni, portano a definire un processo fuori controllo e quindi caratterizzato da un andamento imprevedibile.

La carta di controllo è un importante strumento che serve per monitorare le prestazioni e i parametri del processo e fornisce informazioni utili per migliorarlo.

L'obiettivo principale è quello di individuare la presenza di cause speciali. Si procede dunque ad identificare le cause che hanno agito sul processo. Fatto ciò, si riduce la variabilità riportandola, in caso di fuori controllo, a livelli accettabili.

Una carta di controllo è una visualizzazione grafica di una serie di test statistici utili a verificare se il processo è in controllo. Si estraggono, ad intervalli regolari di tempo, campioni di numerosità n . Si stabilisce la caratteristica da controllare e in base a questa si costruisce la statistica di controllo $W_t = g(x_t)$. Per ogni campione si verifica il seguente sistema d'ipotesi:

H_0 : il processo è in controllo

H_1 : il processo è fuori controllo

Per decidere se accettare o rifiutare l'ipotesi nulla è necessario stabilire una linea centrale (CL) che rappresenta il valor medio della caratteristica di qualità da controllare. Di conseguenza si determinano i limiti di controllo inferiore e superiore (rispettivamente LCL e UCL). Se la statistica test non appartiene all'intervallo (LCL - UCL) si rifiuta l'ipotesi nulla.

Ipotizziamo che la statistica di controllo W abbia valore atteso pari a μ_w e varianza pari a σ_w^2 . I limiti di controllo superiore e inferiore diventano di conseguenza:

$$UCL = \mu_w + k_1 \sigma_w$$

$$LCL = \mu_w - k_2 \sigma_w$$

La linea centrale è rappresentata da μ_w .

Le costanti k_1 e k_2 sono fissate in modo tale che la probabilità che la statistica W non appartenga ai limiti di controllo sia α quando il processo è in controllo. Questo valore è la probabilità di accettare H_1 quando H_0 è vera (errore del primo tipo). Ovvero segnalare un fuori controllo quando il processo in realtà è in controllo.

La probabilità, invece, di non registrare un allarme quando il processo non è più in controllo è definita con β (errore del secondo tipo).

Se si aumenta α i limiti di controllo diventano più stretti e conseguentemente β diminuisce. Non potendo minimizzare contemporaneamente i due parametri si possono applicare due metodiche diverse per determinare i limiti di controllo:

- fissare n e determinare α e β di conseguenza
- fissare α e β e di conseguenza n

4.1 Determinazione della numerosità campionaria

Uno strumento fondamentale a tal fine è la curva operativa caratteristica (OC) che ci permette di sapere qual è la probabilità, β , di non rilevare un determinato shift del parametro del processo a seconda della diversa numerosità campionaria. È una funzione che decresce all'aumentare sia di n che dell'ampiezza dello shift. Ovviamente, più grande sarà il campione più sarà semplice individuare scostamenti del processo.

Se $Y \sim N(\mu_y, \sigma_y^2)$ la statistica di controllo per la media X_t è sotto controllo quando appartiene all'intervallo:

$$\left[\mu_y - Z_{1-\alpha/2} \frac{\sigma_y}{\sqrt{n}} \quad - \quad \mu_y + Z_{1-\alpha/2} \frac{\sigma_y}{\sqrt{n}} \right]$$

Se la media del processo passa da μ_y a μ_{y1} la probabilità beta di non accorgersi di questo cambiamento viene calcolata nel modo seguente:

$$\beta = \Phi \left(\sqrt{n} \frac{\mu_y - \mu_{y1}}{\sigma_y} + Z_{1-\alpha/2} \right) - \Phi \left(\sqrt{n} \frac{\mu_y - \mu_{y1}}{\sigma_y} - Z_{1-\alpha/2} \right)$$

La probabilità β dipende quindi dall'ampiezza del campione n , dall'ampiezza δ dello shift, e da α , intesa come probabilità dell'errore del primo tipo.

4.2 Frequenza di campionamento ottimale

La soluzione ottimale sarebbe quella di campionare frequentemente con numerosità elevata.

Nella pratica si preferisce creare piccoli campioni analizzati frequentemente oppure grandi campioni con frequenza di campionamento più bassa.

4.3 ARL (Average Run Length)

La *Run Length* è una variabile casuale e l'ARL rappresenta il suo valore atteso. Esso è il numero medio di punti che, in media, devono essere riportati sulla carta prima che venga segnalato un fuori controllo.

Se il processo si trova in stato di controllo, con $\alpha = P(\text{falso allarme})$:

$$ARL_0 = \frac{1}{\alpha}$$

Se il processo è fuori controllo, con $\beta = P(\text{mancato allarme})$:

$$ARL_1 = \frac{1}{1 - \beta}$$

Uno schema di controllo ottimale deve segnalare raramente un falso allarme e il più presto possibile un effettivo fuori controllo. Quindi l'ARL deve essere il più grande possibile se il processo è in controllo, mentre è necessario sia il più piccolo possibile se ci si trova in uno stato di fuori controllo.

4.4 Analisi degli andamenti tipici

La presenza di tutti i punti all'interno dei limiti prestabiliti è condizione necessaria, ma non sufficiente per definire un processo in controllo. Quando l'andamento dei punti, pur all'interno dei limiti, non è casuale è sintomo che il processo si sta spostando verso lo stato di fuori controllo.

Per esempio una sequenza crescente o decrescente di punti al di sopra o al di sotto della linea centrale hanno una bassa probabilità di verificarsi e per questo sono da considerarsi sintomo di un fattore specifico che sta agendo.

I principali indizi di non casualità sono:

- *Runs* quando i punti si trovano tutti sopra o sotto la linea centrale
- Tendenze, ovvero una serie di punti con andamento crescente o decrescente
- Presenza di parecchi punti vicini ai limiti di controllo (2 punti su 3 consecutivi cadono oltre i limiti posti a 2σ oppure 4 su 5 oltre i limiti posti a 1σ)
- Stratificazione, tendenza dei punti ad accumularsi attorno alla linea centrale (otto punti consecutivi cadono tutti dalla stessa parte della linea centrale)
- Periodicità dovuta a cause esterne ricorrenti (temperature, turni di lavoro, manutenzione, ecc.)

Tra parentesi ho indicato alcune delle regole decisionali per l'individuazione di comportamenti anomali scritte nel manuale della Western Electric (1956).

4.5 Stima dei parametri del processo

Se media e varianza del processo sono noti, calcolare i limiti di controllo inferiore e superiore risulta immediato.

Molto spesso accade che μ e σ non siano noti ed è quindi necessario stimarli in modo opportuno. Lo stimatore per la media è la media degli m campioni estratti dal processo in controllo.

Se, come spesso accade, anche la variabilità non è nota è necessario stimarla.

I due metodi più comuni per calcolare σ si basano sull'utilizzare:

- i range degli m campioni
- la deviazione standard campionaria

Nel primo caso si tratta di calcolare il range medio degli m campioni di numerosità n che risulta essere uno stimatore del range del processo. Perciò si definisce la variabile $W = \frac{R}{\sigma}$ che ha una distribuzione nota dipendente da n e un valore atteso pari a d_2 . Uno stimatore corretto di σ è definito come il rapporto tra il range medio e d_2 .

Mentre la deviazione standard di W è tabulata in funzione della numerosità campionaria ed è pari a d_3 .

Poiché $R = W\sigma$ allora $\sigma_R = d_3\sigma$. Di conseguenza una stima corretta per $\sigma_R = d_3 \frac{\bar{R}}{d_2}$

Questo è sicuramente un metodo che permette di fare calcoli semplici e veloci.

Purtroppo, però, perde in efficienza relativa con numerosità campionaria maggiori di 10 poiché vengono perse tutte le informazioni contenute nei valori compresi tra valore massimo e minimo.

Nel secondo caso si determina la deviazione standard media di m campioni definita come \bar{S} . Essa ha un valore atteso uguale a $c_4\sigma$ e una deviazione standard pari a $\sigma\sqrt{1-c_4^2}$. Ne consegue che uno stimatore per σ è dal rapporto tra \bar{S} e c_4 .

4.6 I sottogruppi razionali

La raccolta dei dati deve essere fatta in accordo con il concetto di gruppi razionali descritto da Shewhart (1931). Risulta quindi necessario scegliere campioni in modo da rendere massima la probabilità di osservare differenze tra i sottogruppi nel tempo, mentre deve essere minimizzata la probabilità di rilevare differenze all'interno di un campione dovute agli errori sistematici, se questi sono presenti.

I campioni devono essere rappresentativi di tutte le unità prodotte in un determinato intervallo in condizioni di controllo.

E' fondamentale, quindi, che i dati a disposizione siano divisi in campioni nel modo più adeguato possibile. Pensiamo, per esempio, a quali sarebbero le conseguenze se alcuni dati di un campione fossero rilevati all'inizio di una deviazione del processo e le prime osservazioni del campione successivo alla fine dello shift in atto. Probabilmente la variazione avvenuta non verrebbe messa in evidenza.

4.7 Le carte di controllo Shewhart

Questa tipologia di carte di controllo prendono il nome dal loro inventore Walter Shewhart.

Tra il 1920 e il 1945 si svilupparono le prime tecniche statistiche finalizzate al controllo della qualità. Ci si accorse che le verifiche a tappeto su tutti i prodotti finiti, per appurarne un eventuale difettosità, erano troppo costose. Grazie a queste nuove metodiche, si riusciva a stabilire, mentre si produceva, se il processo presentava delle irregolarità.

Le carte di controllo si dividono in due gruppi: per variabili, quando si hanno a disposizione dati misurabili di tipo continuo e per attributi, quando si considerano caratteristiche discrete o non misurabili (ad esempio il numero di difetti).

In particolare le carte per variabili assumono per il loro corretto utilizzo che:

- i dati provengano da una distribuzione normale
- le osservazioni siano indipendenti e identicamente distribuite

E' importante, inoltre, ricordare che le carte Shewhart si definiscono "senza memoria". La statistica di controllo non tiene conto infatti delle informazioni provenienti dai campioni precedenti, ma solo da quello misurato nell'ultimo istante.

Qui di seguito e' stata data particolare attenzione al calcolo dei limiti per la carta S e \bar{X} per il controllo dei sottogruppi. Questa sarà una tra le due tipologie applicate nel capitolo successivo.

Consideriamo m campioni ciascuno di numerosità n e calcoliamo innanzitutto:

$$\bar{S} = \frac{\sum_{i=1}^m s_i}{m}$$

dove

$$s_i = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n - 1)}}$$

Uno stimatore non distorto per σ è $\frac{\bar{S}}{c_4}$ pertanto i parametri della carta S risultano:

$$UCL = \bar{S} + K \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2}$$

$$LC = \bar{S}$$

$$LCL = \bar{S} - K \frac{\bar{S}}{c_4} \sqrt{1 - c_4^2}$$

Di conseguenza calcoliamo i limiti di controllo della carta per monitorare la media del processo:

$$UCL = \bar{\bar{x}} + K \frac{\bar{S}}{c_4 \sqrt{n}}$$

$$LC = \bar{\bar{x}}$$

$$LCL = \bar{\bar{x}} - K \frac{\bar{S}}{c_4 \sqrt{n}}$$

dove $\bar{\bar{x}}$ come la media delle medie dei sottogruppi.

4.8 Carta EWMA

Per poter identificare cambiamenti di piccole dimensioni nella media del processo e' necessario utilizzare una carta di controllo "con memoria". Vengono così chiamate perchè utilizzano sia le informazioni provenienti dai campioni passati che da quelli presenti.

La carta Exponentially Weighted Moving Average (EWMA) è stata introdotta da Roberts (1959) ed e' definita come:

$$z_i = \lambda x_i + (1 - \lambda) z_{i-1}$$

dove λ e' una costante di lisciamento compresa tra 0 e 1 che determina quanto peso assegnare alle osservazioni. Il valore iniziale coincide con il livello medio fissato per il processo.

Il peso più alto viene assegnato all'osservazione corrente e tutte le osservazioni precedenti hanno pesi che decrescono in progressione geometrica.

Ipotizzando che le osservazioni x_i siano indipendenti con varianza costante σ^2 si ottiene che la varianza di z_i e' la seguente:

$$\sigma_{z_i}^2 = \sigma^2 \left(\frac{\lambda}{2 - \lambda} \right) [1 - (1 - \lambda)^{2i}]$$

Di conseguenza i limiti di controllo della carta Ewma sono :

$$UCL = \bar{x} + L\sigma \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2i}]}$$

$$CL = \bar{x}$$

$$LCL = \bar{x} - L\sigma \sqrt{\frac{\lambda}{2 - \lambda} [1 - (1 - \lambda)^{2i}]}$$

La scelta del parametro λ e della costante L dipende dall'ARL in controllo che si vuole ottenere.

Per $i \rightarrow \infty$ i limiti inferiore e superiore convergono a:

$$UCL = \bar{x} + L\sigma \sqrt{\frac{\lambda}{2 - \lambda}}$$

$$LCL = \bar{x} - L\sigma \sqrt{\frac{\lambda}{2 - \lambda}}$$

Se, invece che misure individuali, vengono utilizzati sottogruppi razionali, con media e deviazione standard incognite, i limiti di controllo asintotici diventano:

$$UCL = \bar{\bar{x}} + L\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}}$$

$$CL = \bar{\bar{x}}$$

$$LCL = \bar{\bar{x}} - L\hat{\sigma}\sqrt{\frac{\lambda}{2-\lambda}}$$

$$\text{Dove: } \bar{\bar{x}} = \sum \frac{x_i}{n}, \quad \hat{\sigma} = \frac{\bar{S}}{c_4}$$

Per disegnare una carta EWMA è necessario innanzitutto fissare l'ARL in controllo. I limiti saranno stabiliti in modo che l'ARL fuori controllo risulti il minore possibile per un δ prefissato, con $\delta = \frac{\mu_1 - \mu_0}{\sigma}$.

L'ARL della carta EWMA si può calcolare per mezzo dell'approccio Catene di Markov. Inoltre le tavole di Lucas e Saccucci (1990) e i normogrammi di Crowder (1989) permettono di determinare le diverse coppie delle costanti λ e L che permettono di avere, per un dato ARL in controllo, l'ARL fuori controllo minore possibile per un δ prefissato.

Capitolo 5

Applicazione al caso studio

5.0 Scelta dei parametri per le carte di controllo utilizzate

Le carte di controllo nei paragrafi successivi sono state applicate ai residui standardizzati dei modelli di serie storiche adattati. Dato il valore dell'ARL in controllo si utilizza una funzione in R per calcolare il valore ottimale delle costanti λ e L della carta EWMA. In particolare per un ARL in controllo pari a 500 e per μ e σ posti pari a zero ed uno rispettivamente, il valore ottimo per il parametro λ è risultato pari a 0,0228, mentre L ottimale è pari a 2,3314.

Per le carte di tipo Shewhart è stato scelto K pari a $\pm 3,09$. La probabilità di errore del primo tipo, ovvero α , è pari a 0,002. Questo permette di avere un ARL_0 uguale a 500. Inoltre, poiché i campioni hanno numerosità 50, la costante c_4 utilizzata per la stima di σ è pari a 0,9949.

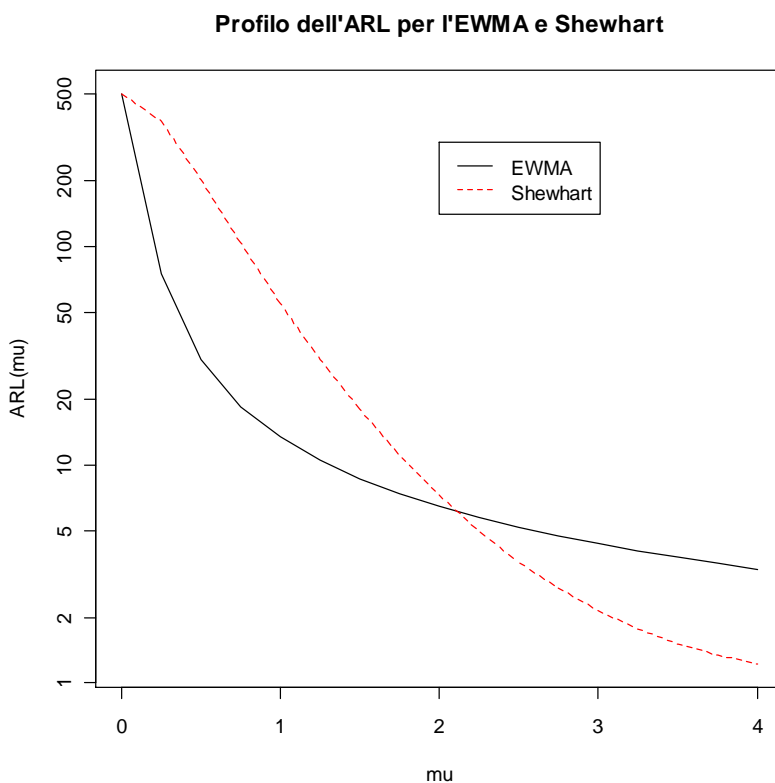


Figura 5.1

La figura 5.1 illustra graficamente la performance nota della carta Shewhart rispetto a quella EWMA. In particolare si osserva che per shift superiori a 2 è la carta Shewhart ad avere una risposta migliore. Quando, invece, l'entità dello shift è inferiore la carta EWMA risponde molto più rapidamente.

5.1 Sensore numero 17

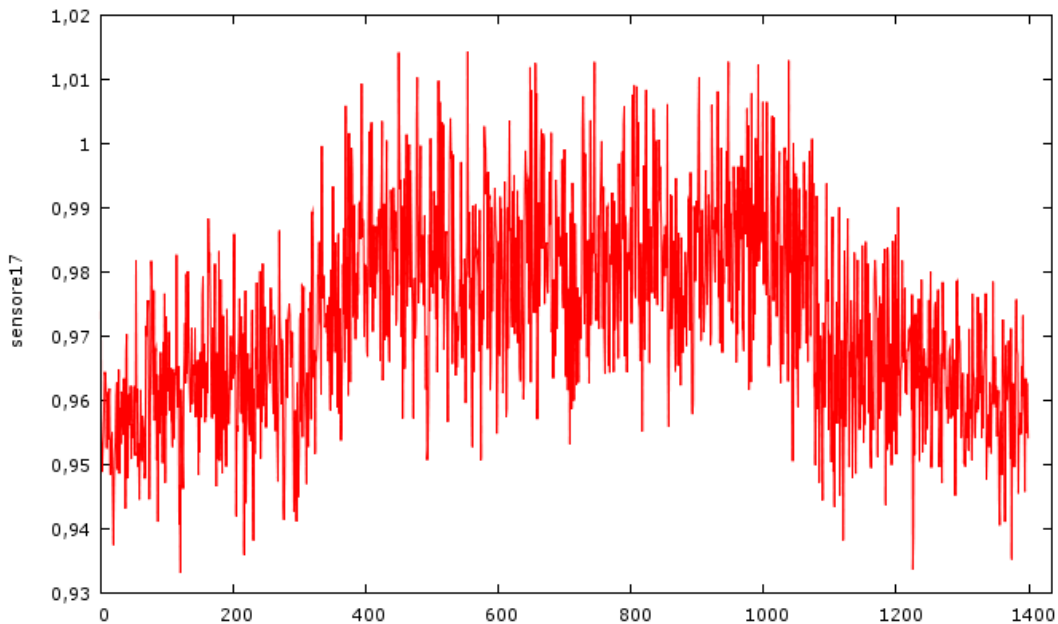


Figura 5.2 grafico della serie del rapporto tra sensore numero 17 e micromanometro

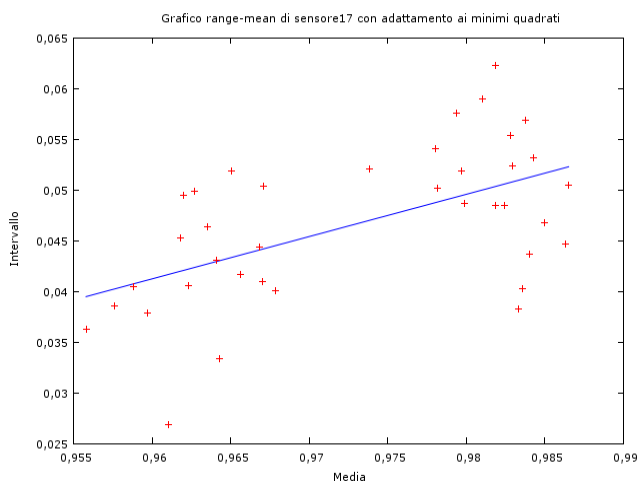


Figura 5.3 grafico range-mean (campione completo) 38 sottogruppi di dimensione 37

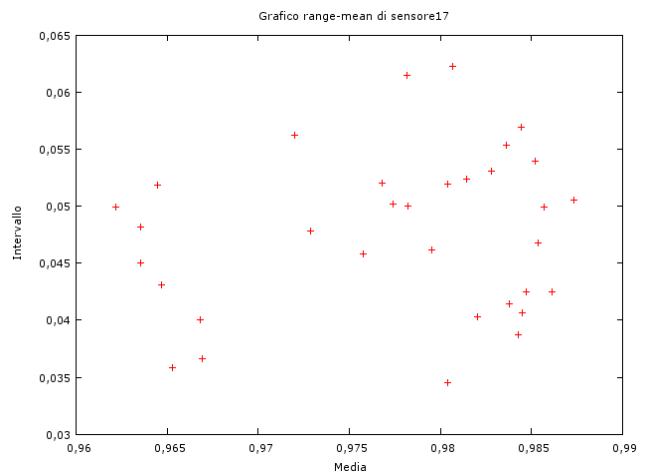


Figura 5.4 grafico range-mean (osservazioni 200-1200) 33 sottogruppi di dimensione 31

Da una prima analisi del grafico della serie si può subito notare come questa non sia stazionaria in media. La stazionarietà in varianza lascia qualche dubbio. Per risolvere il problema della probabile mancanza di stazionarietà in varianza sono state tentate varie strade. Trasformazioni di tipo logaritmico, utilizzo del reciproco e della radice quadrata o l'uso della trasformata di Box-Cox non hanno apportato miglioramenti significativi.

Da un'attenta analisi della figura 5.2 si nota che le prime e le ultime 200 osservazioni sembrano avere una variabilità diversa dalle altre. Queste ipotesi sembrano confermate in particolare dai range-mean plot (figure 5.3-5.4), che costruito sul campione completo mostra come significativa la stima della retta con il metodo dei minimi quadrati e porta al rifiuto dell'ipotesi nulla che la pendenza sia uguale a zero (p-value

pari a 0,00027). Il coefficiente angolare stimato è pari a 0,41 ed è quindi positivo. Questo sembra pertanto indicare che la variabilità aumenti al crescere della media. D'altro canto, se creiamo la stessa tipologia di grafico per le osservazioni dalla 200 alla 1200 del campione, accettiamo in questo caso l'ipotesi che la pendenza sia nulla (il p-value risulta maggiore di 0,3). Sembra quindi che l'ampiezza delle oscillazioni della parte "centrale" della serie sia più omogenea mentre le prime e le ultime 200 osservazioni abbiano una variabilità minore delle altre, ma comunque molto simile tra loro.

La serie, inoltre, è caratterizzata da un trend parabolico. Si è scelto di stimare il trend tramite un modello di regressione lineare. Il modello OLS ha come variabili esplicative trend e trend elevato al quadrato. La variabile dipendente è il rapporto tra le osservazioni ottenute dal sensore e dallo strumento di riferimento. L'obiettivo è quello di rimuovere il trend quadratico che sembra essere presente nella serie. La parabola stimata è la seguente:

$$Y = -6,15530e^{-8}(\text{trend}^2) + 8,87693e^{-5}(\text{trend}) + 0,951075$$

	Coefficiente	Errore Std.	Rapporto t	p-value
costante	0,951075	0,000962620	988,0	0,0000
trend	8,87693e-05	3,17568e-06	27,95	6,40e-137
trend al quadrato	-6,15530e-08	2,19641e-09	-28,02	1,77e-137
R-quadro	0,363157		R-quadro corretto	0,362244

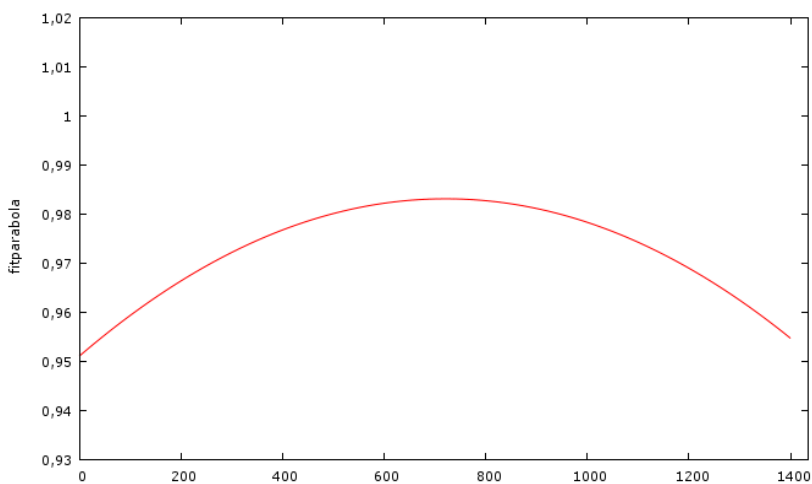


Figura 5.5 parabola stimata con un modello di tipo OLS

Successivamente sono stati sottratti alla serie del rapporto i valori stimati dal modello parabolico.

La variabile ottenuta verrà nominata v1. Sulla nuova serie è stata applicata la procedura di Box e Jenkins per l'identificazione di un opportuno modello ARIMA.

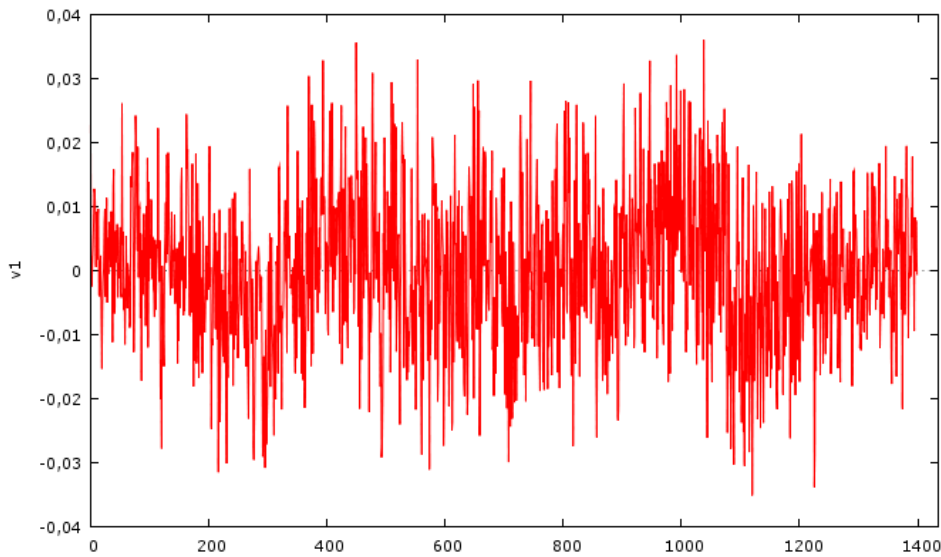


Figura 5.6 grafico della serie detrendizzata

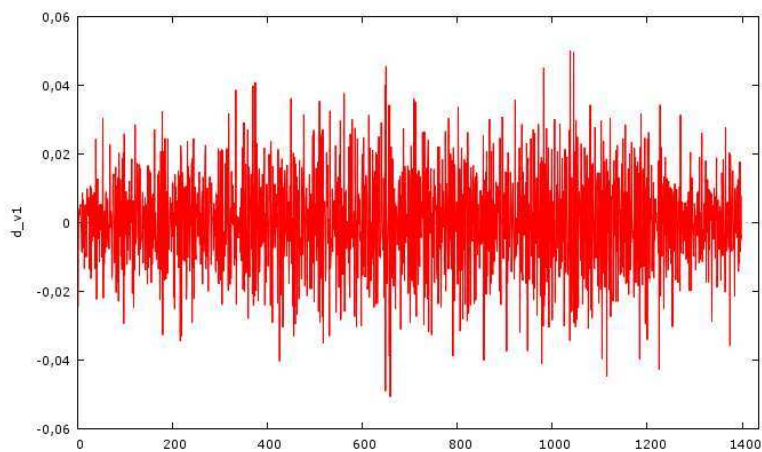


Figura 5.7 grafico della differenze prime di v1

La serie risulta ancora non stazionaria in media, ma una differenziazione di ordine uno sembra essere sufficiente (vedi figura 5.7). Analizziamo il correlogramma totale e parziale delle differenze prime di v1 al fine di identificare gli ordini p e q opportuni.

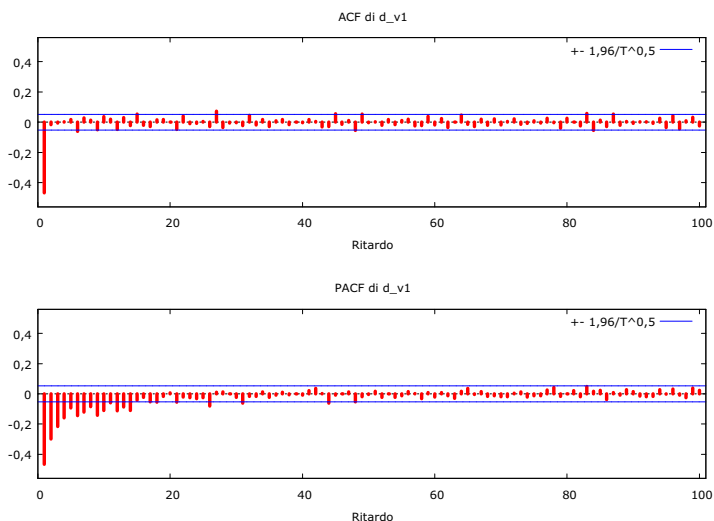


Figura 5.8 stima della funzione di autocorrelazione totale e parziale delle differenze prime di v1

La stima della funzione di autocorrelazione parziale decresce nel tempo e, dopo i primi 13 ritardi significativi, tende ad annullarsi totalmente. Anche $\hat{\rho}_k$ tende a zero, ma ha una velocità di decrescita minore. Procediamo quindi con l'identificazione del modello.

Il modello ARIMA, che si è rivelato il più adatto ha $p=9$, $q=1$ e $d=1$. È un modello vincolato ad avere $\Phi_2, \Phi_3, \Phi_4, \Phi_5, \Phi_7, \Phi_8$ pari a zero. Quindi, di fatto, ha 4 parametri: $\Phi_1, \Phi_6, \Phi_9, \theta_1$

	Coefficiente	Errore Std.	Rapporto z	p-value
Φ_1	0,0917777	0,0284836	3,222	0,0013
Φ_6	-0,0849968	0,0275382	-3,087	0,0020
Φ_9	-0,0560895	0,0272078	-2,062	0,0393
θ_1	-0,943962	0,0102914	-91,72	0,0000

Media var. dipendente	-0,000017	SQM var. dipendente	0,015154
Media innovazioni	-0,000059	SQM innovazioni	0,011475
Log-verosimiglianza	4260,807	Criterio di Akaike	-8511,615
Criterio di Schwarz	-8485,401	Hannan-Quinn	-8501,815

	Reale	Immaginario	Modulo
Radice AR 1	-1,0637	0,8216	1,3440
Radice AR 2	-1,0637	-0,8216	1,3440
Radice AR 3	1,2089	-0,5035	1,3095
Radice AR 4	1,2089	0,5035	1,3095
Radice AR 5	-0,1463	-1,3273	1,3354
Radice AR 6	-0,1463	1,3273	1,3354
Radice AR 7	0,7482	1,2649	1,4697
Radice AR 8	0,7482	-1,2649	1,4697
Radice AR 9	-1,4943	0,0000	1,4943
Radice MA 1	1,0594	0,0000	1,0594

Tutti i coefficienti sono risultati significativi. Le radici della parte autoregressiva sono, in modulo, maggiori di uno. Possiamo quindi definire il processo stazionario.

Verifichiamo l'adeguatezza del modello stimato.

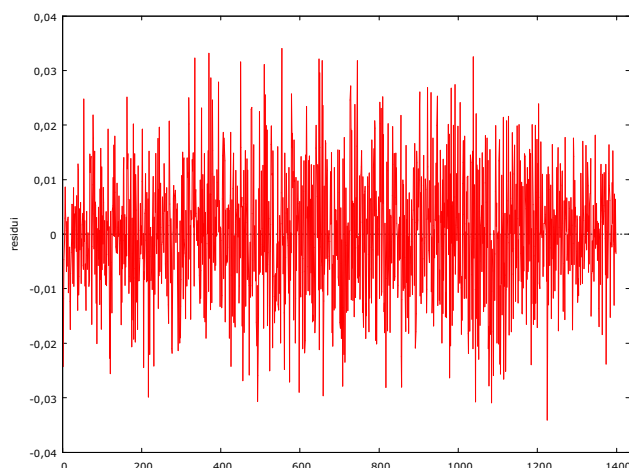


Figura 5.9 grafico della serie dei residui del modello ARIMA(9|1|1) vincolato

La serie dei residui oscilla intorno allo zero. La variabilità dei dati, però, sembra in leggero aumento nel tempo.

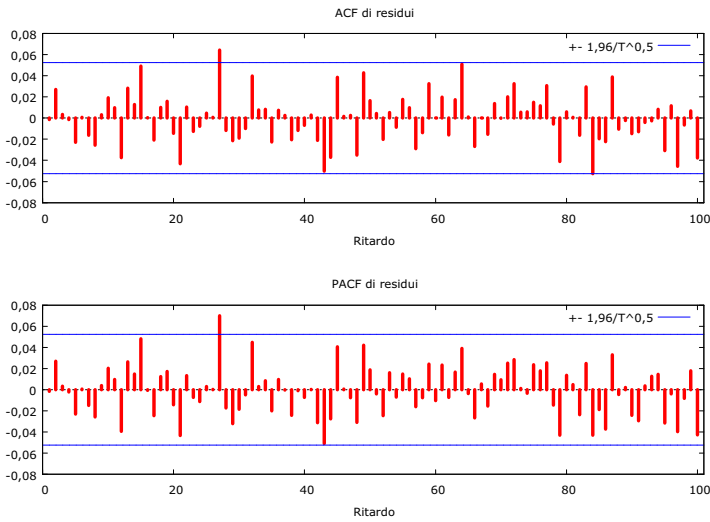


Figura 5.10 stima della funzione di autocorrelazione totale e parziale dei residui del modello

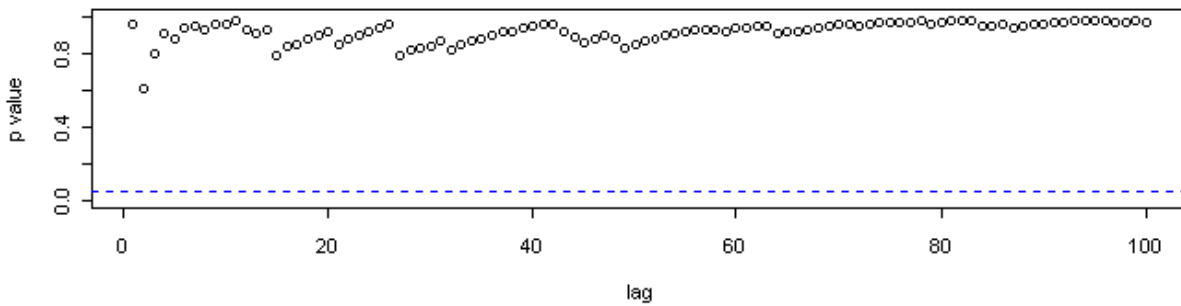


Figura 5.11 p-value della statistica test di Ljung-Box

Solo per tre ritardi su cento considerati la funzione di autocorrelazione globale risulta significativamente diversa da zero. I p-value della statistica di Ljung-Box sono tutti sopra la soglia dello 0,05. Il valore della statistica test, calcolata sui primi 100 ritardi, è pari a 90,6973. Con un livello di significatività osservato maggiore di 0,7 si accetta l'ipotesi nulla che i residui siano incorrelati.

Verifichiamo che i residui siano distribuiti normalmente.

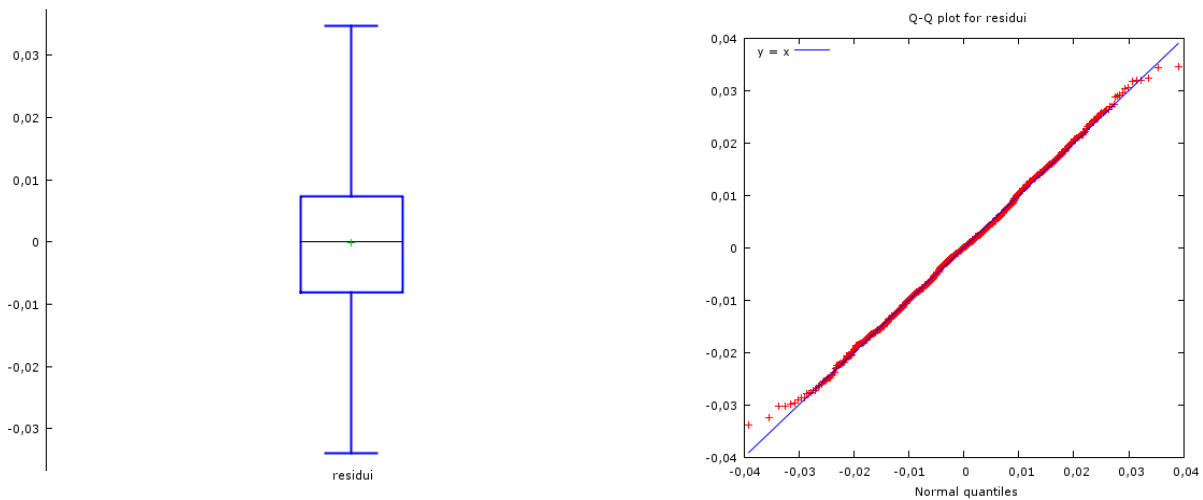


Figura 5.12 box plot e diagramma quantile-quantile dei residui del modello

Test per la normalità effettuati:

W di Shapiro-Wilk = 0,998638, con p-value 0,349938

Test di Jarque-Bera = 1,08448, con p-value 0,581444

Entrambi i p-value ottenuti sono fortemente a favore dell'ipotesi che i residui siano distribuiti in modo normale. Anche se dall'analisi della figura 5.12 i residui sembrano presentare una leggerissima asimmetria destra (indice di asimmetria di Fisher = 0,05).

Controlliamo che i residui siano omoschedastici con il diagramma sottostante.

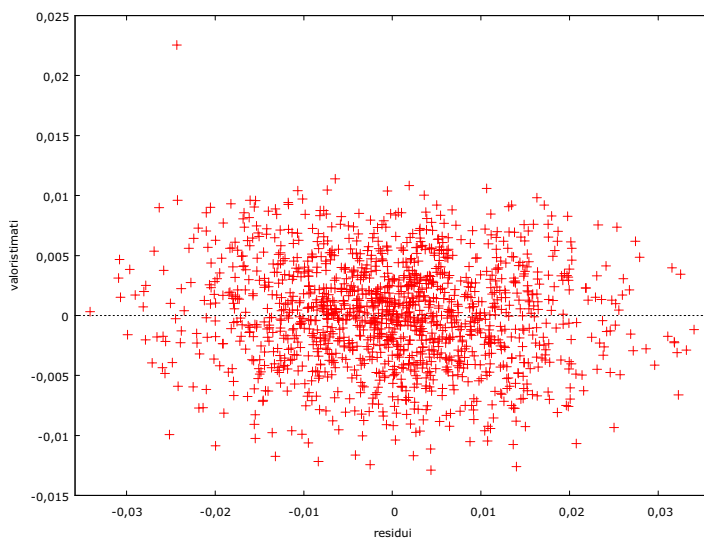


Figura 5.13 diagramma di dispersione dei residui del modello rispetto ai valori stimati dallo stesso

Non sembrano essere presenti andamenti sistematici di alcun tipo.

Disegniamo, infine, le carte di controllo per i residui standardizzati del modello ARIMA stimato precedentemente.

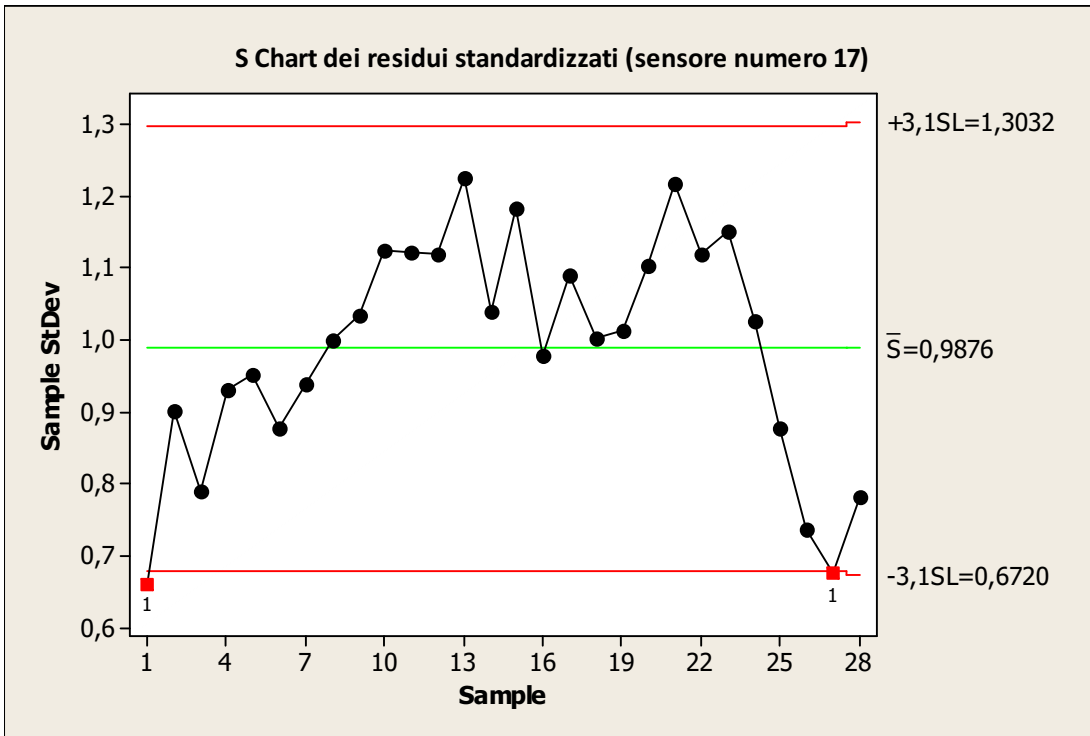


Figura 5.14

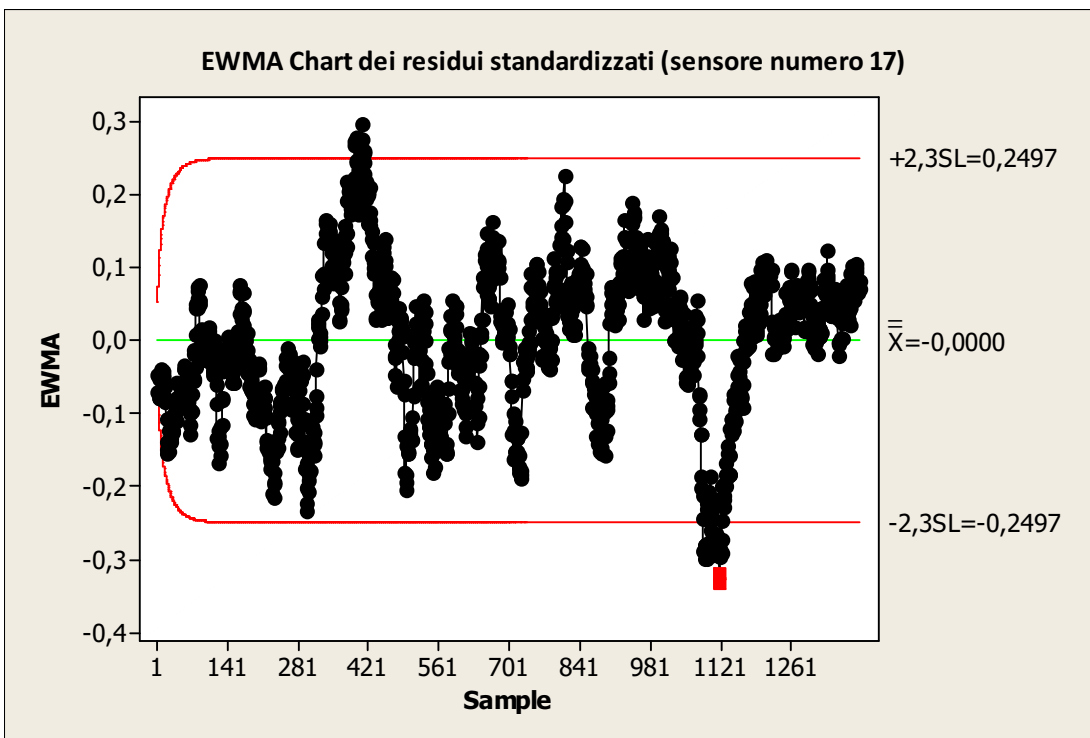


Figura 5.15

La carta S mostra due punti (campione 1 e 27) fuori controllo e, di conseguenza, non ho disegnato quella per le medie. Poiché la variabilità all'interno dei sottogruppi non è stabile, i limiti di controllo usati nella carta per le medie potevano non essere significativi. Non è stato possibile individuare le cause dei fuori controllo e quindi eliminarle per ricalcolare i limiti della carta per le medie.

La variabilità sembra avere un trend crescente per la rampa ascendente (sottogruppi 1-14) e, simmetricamente, un trend decrescente per quella discendente (sottogruppi 15-28). Le sequenze di punti non sono casuali. Le osservazioni dalla 200 alla 1200 (sottogruppi 5-24) sono quasi tutte al di sopra della linea centrale e sembrano avere una variabilità superiore alle altre coerentemente con quello che era apparso dalle analisi precedenti. La carta EWMA presenta i seguenti punti fuori dai limiti di controllo: 393, 396, 408, 409. Questi corrispondono al momento di passaggio dal gradino a $4,5 \text{ m}^3/\text{h}$ a quello a $5 \text{ m}^3/\text{h}$. Il problema sembrerebbe derivante da una leggera instabilità del flusso in un momento di cambiamento. I punti che devono preoccupare maggiormente sono quelli da 1101 al 1124 che si riferiscono ad un flusso pari a $3,5 \text{ m}^3/\text{h}$ nella rampa in discesa.

Sono più numerosi dei precedenti e alcuni tra essi (1119 e 1120) superano notevolmente il limite di controllo inferiore, infatti il valore della statistica test è pari rispettivamente a $-0,3343$ e $-0,3315$.

5.2 Sensore numero 2

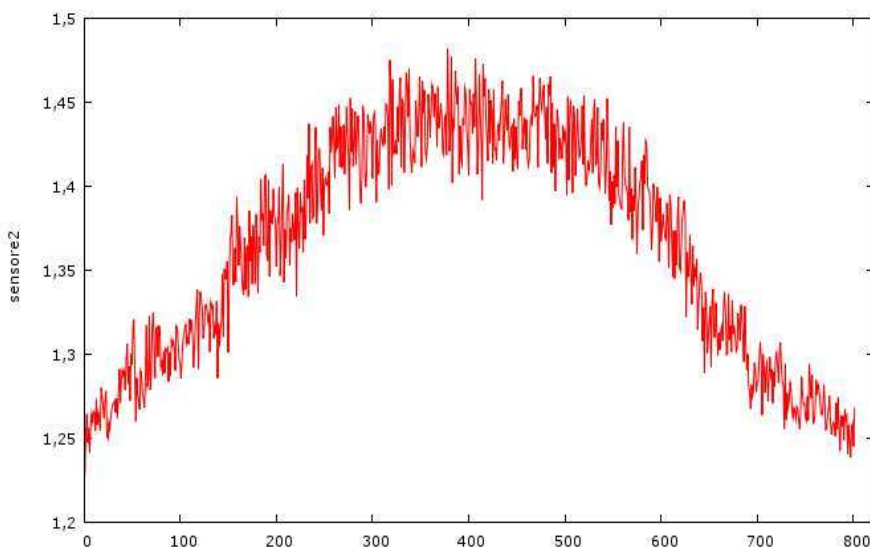


Figura 5.16 serie del rapporto tra il sensore numero 2 e il micromanometro

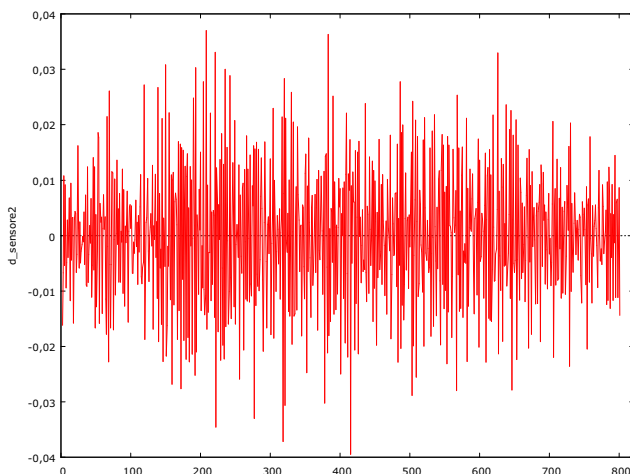


Figura 5.17 serie delle differenze di ordine uno

La serie dei dati provenienti dal secondo sensore è visibilmente non stazionaria in media e presenta un forte trend parabolico. Una differenziazione di ordine unitario (vedi figura 5.17) non sembra sufficiente a stabilizzare la media. Consideriamo, quindi, la serie delle differenze di ordine due rappresentata dal grafico sottostante.

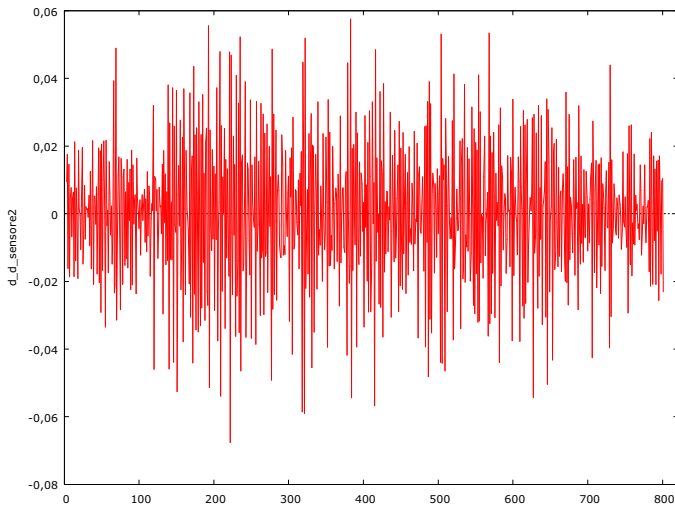


Figura 5.18 serie delle differenze di ordine due

La stabilità della media sembra leggermente migliorata. Verifichiamo se la serie è stazionaria in varianza analizzando il range-mean plot.

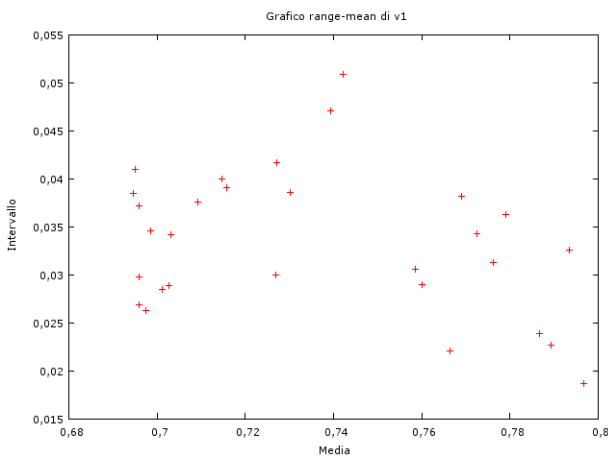


Figura 5.19 grafico range-mean della variabile rapporto 29 sottogruppi di dimensione 28

La pendenza dell'intervallo rispetto alla media è pari a $-0,0628$. Il livello di significatività osservato è di $0,108$. Il p-value risulta appena al di sopra del limite del 10%. Questo porterebbe ad accettare H_0 e quindi che la retta abbia un coefficiente angolare nullo, ovvero a ritenere che l'ampiezza delle oscillazioni rimanga la stessa per tutte le osservazioni. Per decidere gli ordini p e q del modello osserviamo il correlogramma parziale e totale della serie delle differenze di ordine due (figura 5.20). \hat{P}_k è significativamente diversa da zero per i primi 10 ritardi considerati poi tende ad annullarsi, $\hat{\rho}_k$ invece, presenta anche per lag elevati autocorrelazioni fuori dalle bande.

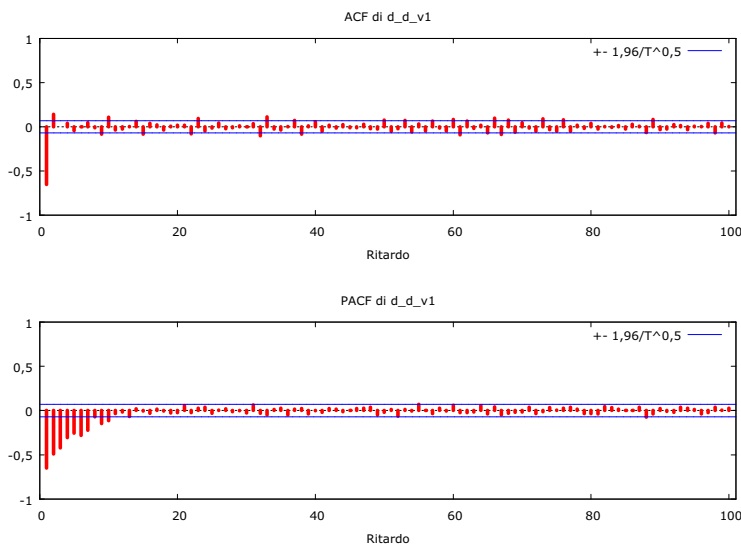


Figura 5.20 correlogramma totale e parziale delle serie delle differenze di ordine due

Il modello ARIMA che meglio è sembrato adattarsi ai dati ha ordine $p=3$, $d=2$ e $q=5$ vincolato, però, ad avere θ_1 e θ_4 uguali a zero.

	Coefficiente	Errore Std.	Rapporto z	p-value
Φ_1	-1,84287	0,0277283	-66,46	0,0000
Φ_2	-1,71752	0,0824747	-20,82	2,58e-096
Φ_3	-0,810456	0,0403480	-20,09	9,66e-090
θ_2	-0,868145	0,0742076	-11,70	1,29e-031
θ_3	-0,859237	0,0862095	-9,967	2,13e-023
θ_5	0,737653	0,0403077	18,30	8,19e-075
Media var. dipendente	2,20e-06	SQM var. dipendente	0,020807	
Media innovazioni	0,000583	SQM innovazioni	0,009129	
Log-verosimiglianza	2613,977	Criterio di Akaike	-5213,954	
Criterio di Schwarz	-5181,170	Hannan- Quinn	-5201,359	

	Reale	Immaginario	Modulo
Radice AR 1	-1,0718	0,0000	1,0718
Radice AR 2	-0,5237	-0,9365	1,0730
Radice AR 3	-0,5237	0,9365	1,0730
Radice MA 1	-1,0764	0,0000	1,0764
Radice MA 2	1,0187	0,0000	1,0187
Radice MA 3	-0,5279	-0,9120	1,0538
Radice MA 4	-0,5279	0,9120	1,0538
Radice MA 5	1,1134	0,0000	1,1134

I parametri utilizzati hanno tutti un p-value minore di 0,001. Tutte le radici della parte AR in modulo sono maggiori di uno.

Proseguiamo quindi con la verifica del modello.

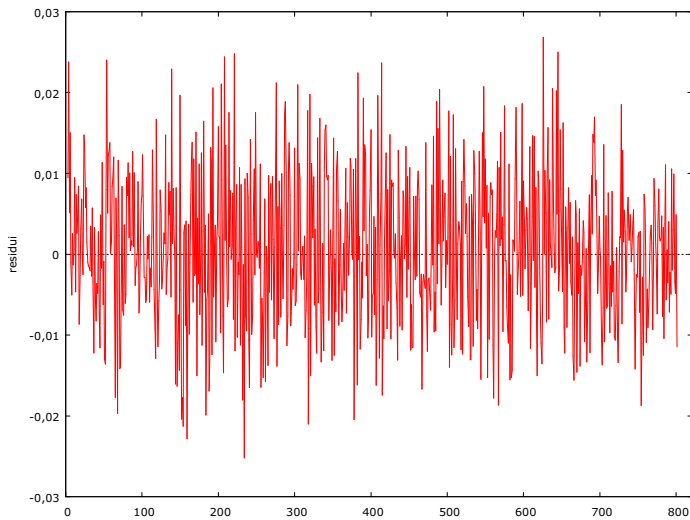


Figura 5.21 grafico della serie dei residui del modello ARIMA(3,2,5) vincolato

Il grafico della serie non mostra andamenti anomali, anche se la variabilità non sembra mantenersi costante.

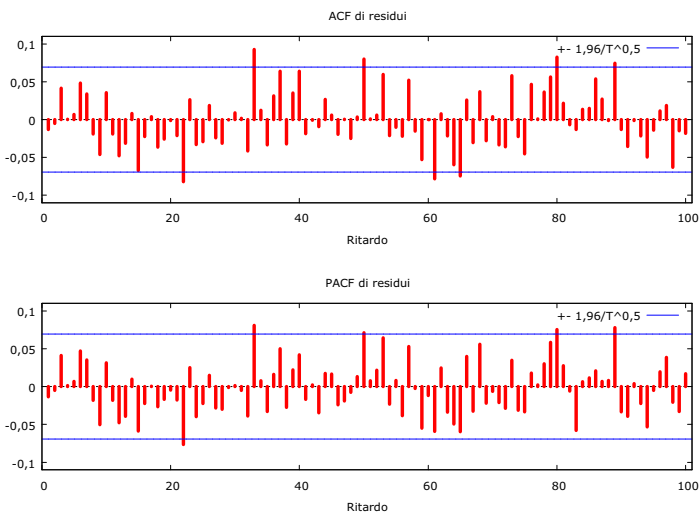


Figura 5.22 stima della funzione di autocorrelazione globale e parziale dei residui

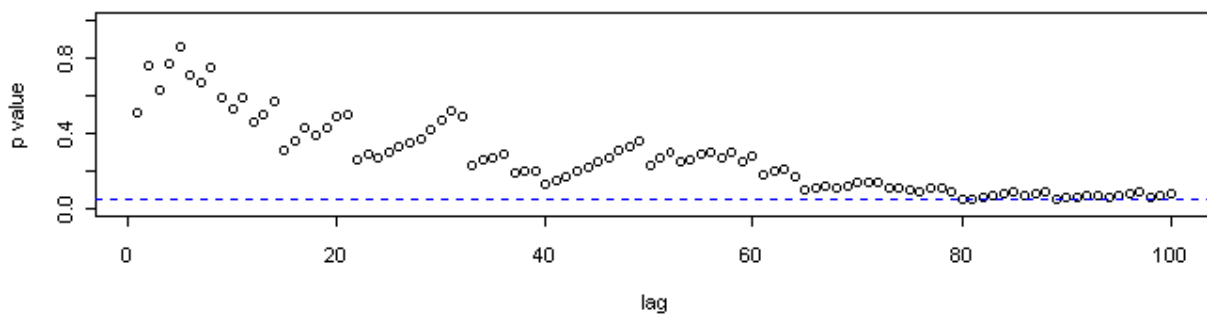


Figura 5.23 p-value della statistica test di Ljung-Box

Dal grafico della funzione di autocorrelazione totale dei residui sembrano essere numerose le osservazioni fuori dalla bande di confidenza al 95% (otto su cento prese in considerazione).

I p-value della statistica di Ljung-Box tendono ad avvicinarsi alla soglia dello 0,05 all'aumentare dei lag. Il valore della statistica test, calcolata sui primi cento ritardi, è pari a 120,94. Il livello di significatività osservato è pari a 0,075 ed è quindi al limite della soglia del 95% per l'accettazione di H_0 . I residui del modello stimato, quindi, presentano ancora della correlazione residua.

Verifichiamo se le innovazioni sono normali.

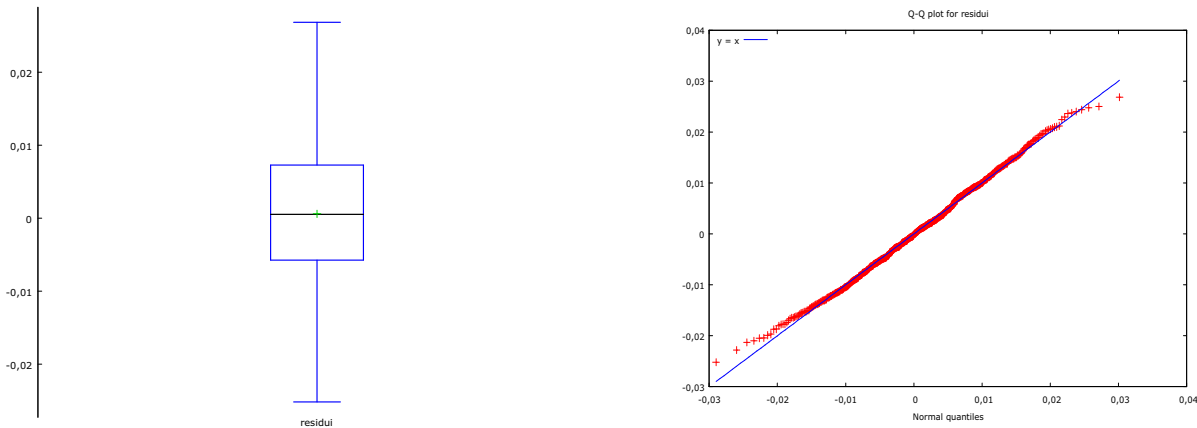


Figura 5.24 box plot e Q-Q plot dei residui

L'applicazione dei test per la verifica della normalità restituiscono i seguenti valori :

W di Shapiro-Wilk = 0,996962, con p-value 0,136801

Test di Jarque-Bera = 4,67737, con p-value 0,0964543

Il p-value del test di normalità di Jarque-Bera risulta al limite della soglia di accettazione (considerando α pari a 0,1). Si nota in particolare, analizzando il diagramma quantile-quantile, una coda destra nei dati (indice di asimmetria di Fisher = 0,1187). I risultati di queste prime analisi sui residui sembrano indicare che il modello utilizzato non riesce a descrivere completamente il processo generatore dei dati.

Il trend parabolico presente nella serie è molto marcato. Proviamo ad eliminarlo e verifichiamo se riscontriamo meno problemi durante il controllo diagnostico. Con la stessa procedura utilizzata per il sensore numero 17, stimiamo la seguente parabola:

$$Y = 7,23053e^{-7} (\text{trend}^2) - 0,000571195 (\text{trend}) + 0,808931$$

	Coefficiente	Errore Std.	Rapporto t	p-value
costante	0,808931	0,00112847	716,8	0,0000
trend	-0,000571195	6,49847e-06	-87,90	0,0000
trend al quadrato	7,23053e-07	7,84614e-09	92,15	0,0000
R-quadro	0,914369	R-quadro corretto	0,914155	

Il coefficiente di determinazione ottenuto è molto alto. Il grafico della nuova serie, ottenuta tramite la differenza tra la variabile rapporto e i valori stimati dal modello parabolico, è quello rappresentato dalla figura sottostante.

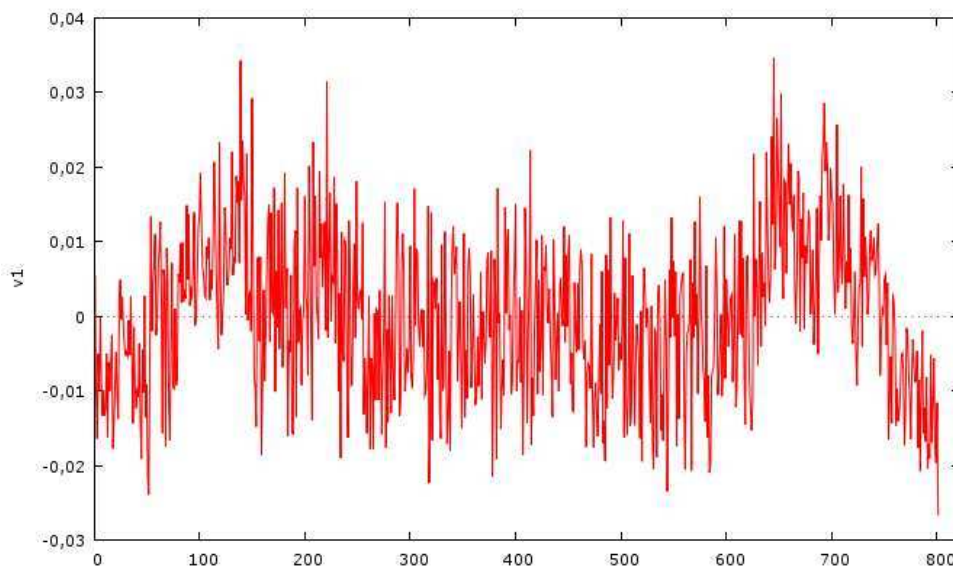


Figura 5.25 grafico della serie detrendizzata

La serie, non ancora stazionaria in media, sembra avere una variabilità più costante. Una differenziazione di ordine unitario apporta comunque un miglioramento soddisfacente (vedi figura 5.26).

Per la corretta identificazione del modello è necessario analizzare i correlogrammi parziali e totali della serie delle differenze (vedi figura 5.27).

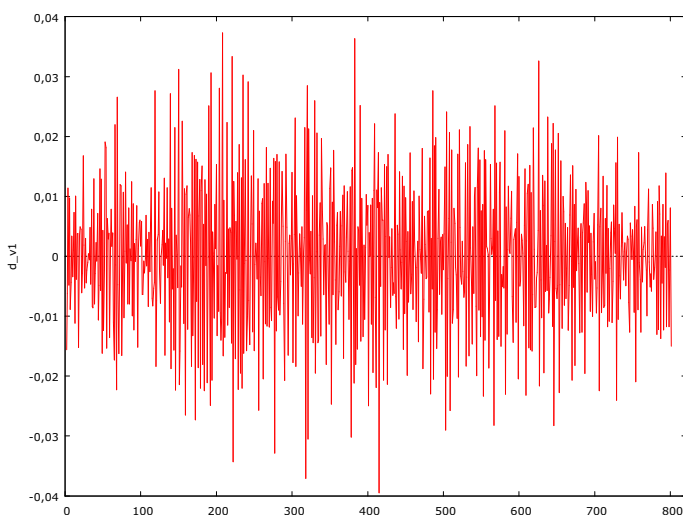


Figura 5.26 grafico della serie delle differenze di v1

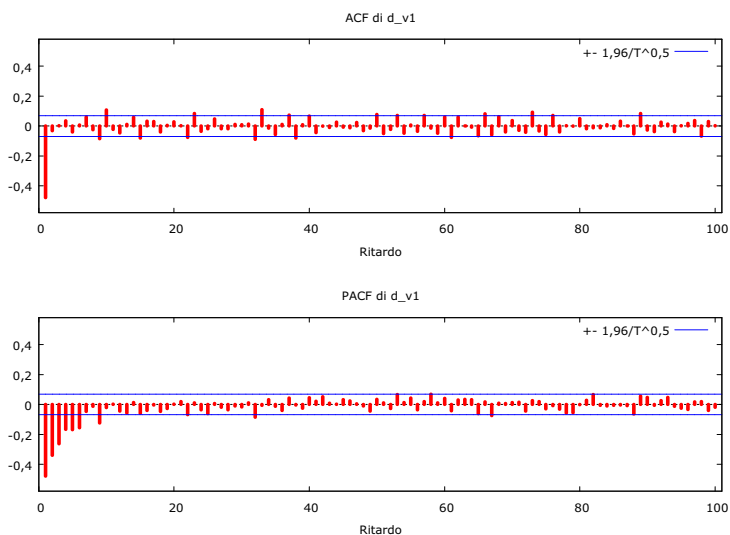


Figura 5.27 stima della funzione di autocorrelazione totale e parziale delle differenze prime di v_1

La stima della funzione di autocorrelazione globale tende a zero molto lentamente. \hat{P}_k , invece, tende ad annullarsi completamente dopo i primi cinque ritardi significativamente diversi da zero.

Stimiamo un modello ARIMA(2,1,3).

	Coefficiente	Errore Std.	Rapporto z	p-value
Φ_1	-0,638364	0,0102314	-62,39	0,0000
Φ_2	-0,985868	0,0136306	-72,33	0,0000
θ_1	-0,227029	0,0249436	-9,102	8,89e-020
θ_2	0,403464	0,0301417	13,39	7,34e-041
θ_3	-0,841824	0,0267051	-31,52	4,21e-218

Media var. dipendente	-0,000040	SQM var. dipendente	0,012097
Media innovazioni	-0,000112	SQM innovazioni	0,009012
Log-verosimiglianza	2631,199	Criterio di Akaike	-5250,397
Criterio di Schwarz	-5222,289	Hannan-Quinn	-5239,599

	Reale	Immaginario	Modulo
Radice AR 1	-0,3238	-0,9537	1,0071
Radice AR 2	-0,3238	0,9537	1,0071
Radice MA 1	1,1470	0,0000	1,1470
Radice MA 2	-0,3339	-0,9613	1,0177
Radice MA 3	-0,3339	0,9613	1,0177

Tutti i coefficienti hanno un p-value minore di 0,001 e sono quindi fortemente significativi.

Entrambe le radici della parte autoregressiva sono appena superiori ad uno in modulo. Inoltre tutti e tre i criteri considerati (Akaike, Schwarz, Hannan-Quinn) sono minori rispetto a quelli ottenuti dal precedente modello.

Procediamo con le analisi diagnostiche del modello scelto.

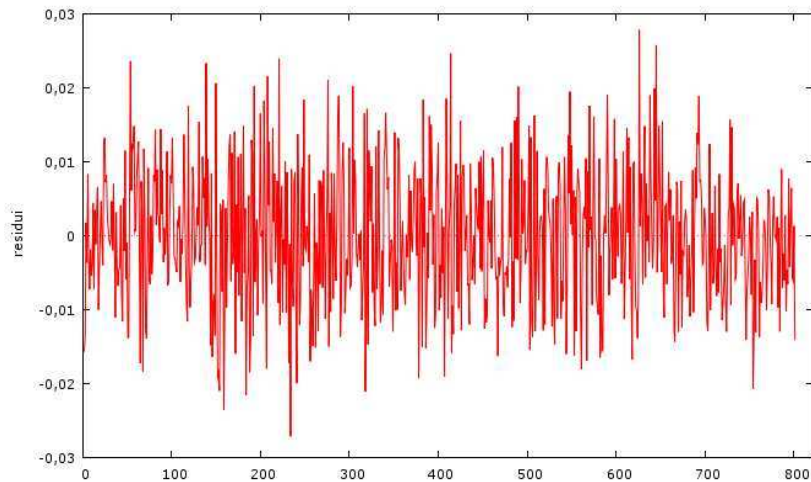


Figura 5.28 serie dei residui del modello ARIMA(213)

La serie dei residui oscilla intorno alla zero e non sembra esserci evidenza di comportamenti anomali.

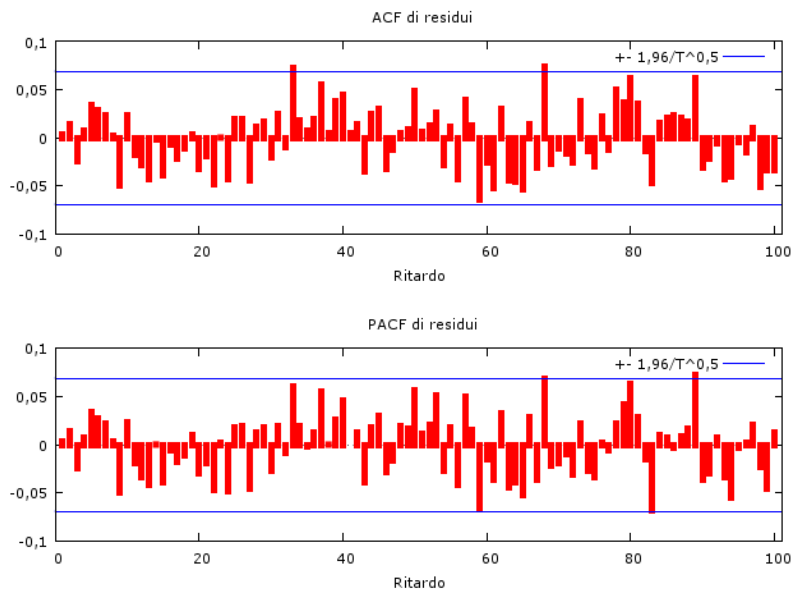


Figura 5.29 stima della funzione di autocorrelazione totale e parziale dei residui

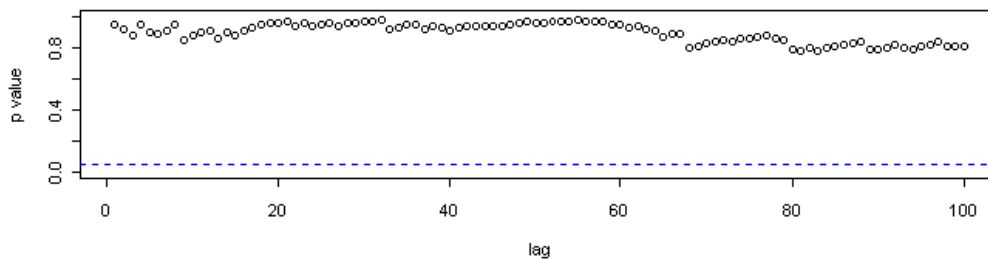


Figura 5.30 p-value della statistica test di Ljung-Box

I livelli di significatività della statistica di Ljung-Box superano ampiamente la soglia del 95%. Le autocorrelazioni parziali e globali stimate dei residui sono significativamente diverse da zero per 4 ritardi

rispetto ai 100 considerati. La statistica di Ljung e Box $Q(100)$ è pari a 87,43 ($p\text{-value}=0,8108$). I residui sono distribuiti come un white noise.

Verifichiamo che la distribuzione delle innovazioni sia di tipo normale.

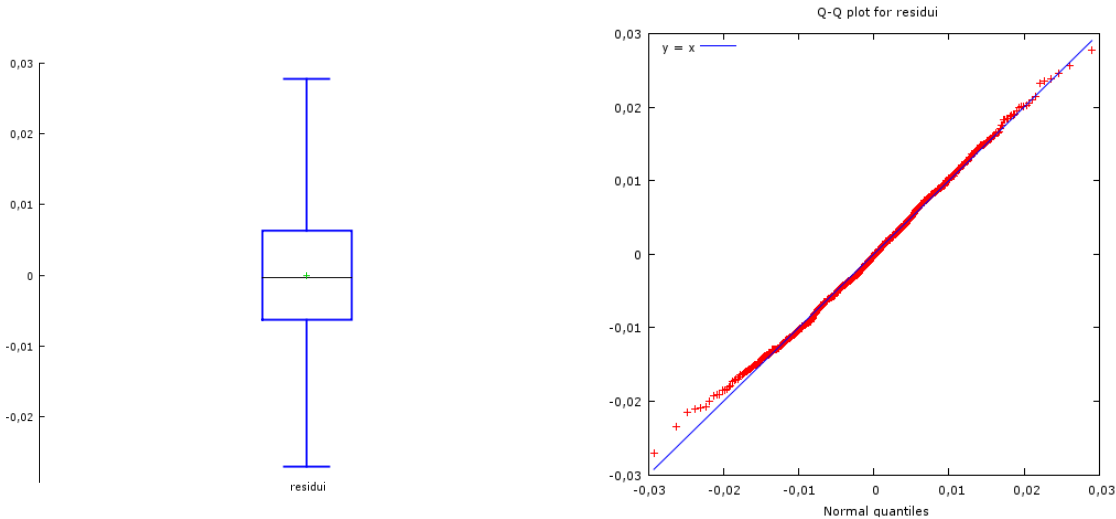


Figura 5.31 box plot e diagramma quantile-quantile dei residui

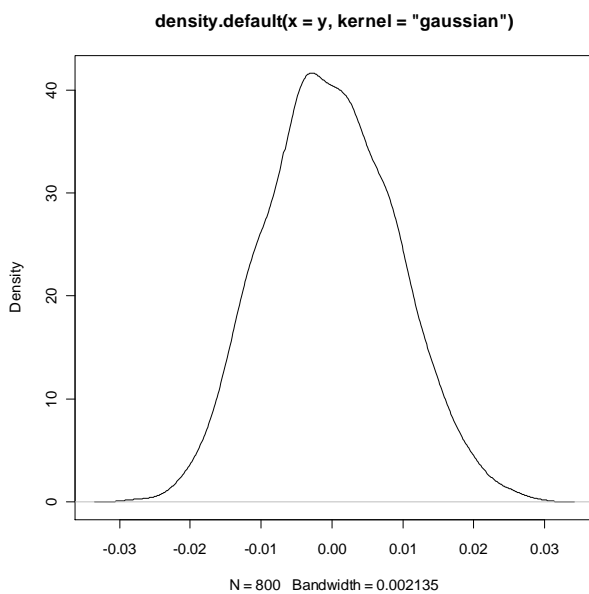


Figura 5.32 funzione di densità dei residui

Si nota una leggera asimmetria destra (indice di asimmetria di Fisher pari a 0,13) e curtosi (indice di curtosi di Fisher pari a 0,26). Le code hanno un maggior peso rispetto alla parte centrale della forma.

Complessivamente, però, i dati sembrano essere distribuiti in modo normale. Infatti le statistiche test per la verifica della normalità sono le seguenti:

W di Shapiro-Wilk = 0,997429, con $p\text{-value}$ 0,245922

Test di Jarque-Bera = 4,45502, con $p\text{-value}$ 0,107796

Entrambi $p\text{-value}$ dei test portano a non rifiutare l'ipotesi nulla che i residui siano normalmente distribuiti.

Controlliamo che questi abbiano varianza costante e non dipendente dal tempo tramite il diagramma di dispersione sottostante.

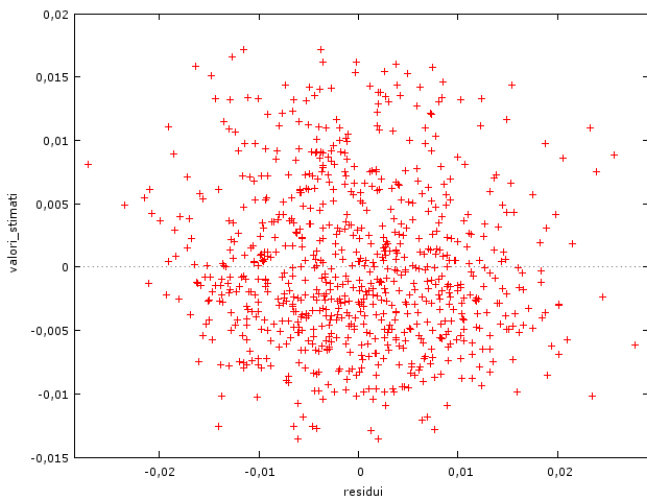


Figura 5.33 diagramma di dispersione dei residui del modello rispetto ai valori stimati dallo stesso

I punti sembrano disporsi in maniera completamente casuale.

Terminiamo l'analisi del sensore numero 2 creando le carte di controllo sui residui del modello scelto, ovvero un ARIMA(2,1,3).

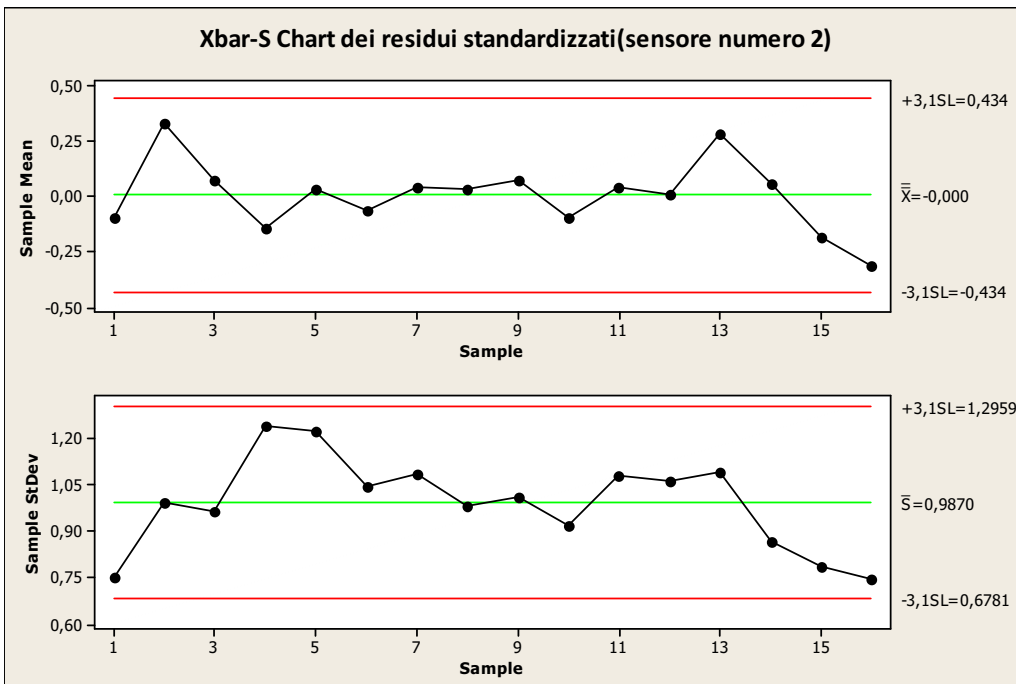


Figura 5.34

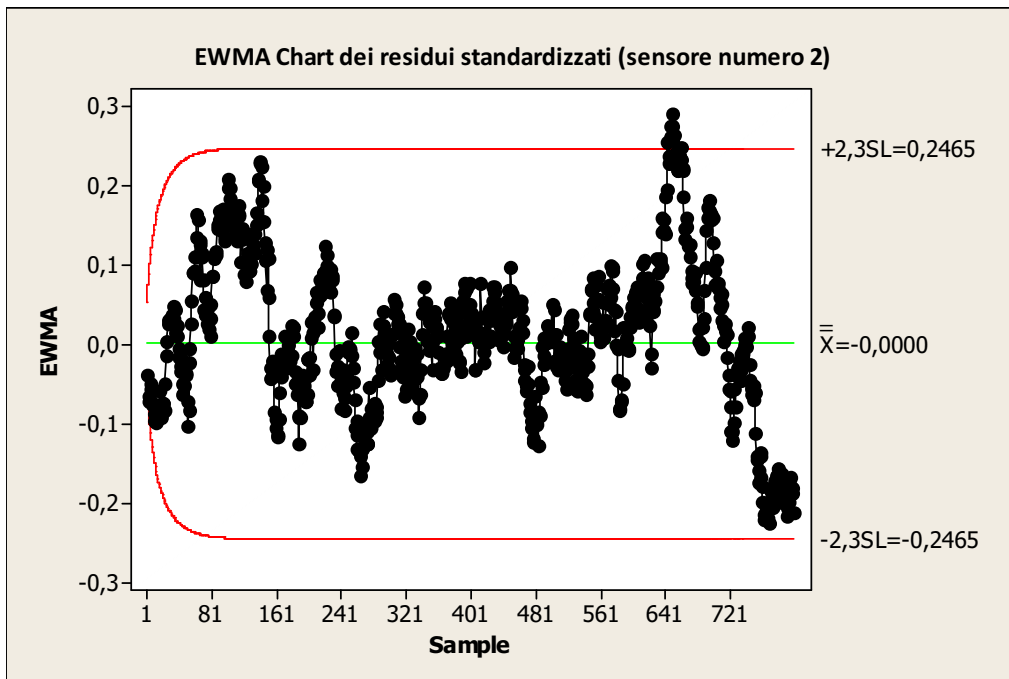


Figura 5.35

Sia la carta per le medie che quella per il controllo della variabilità dei sottogruppi non mostrano alcun fuori controllo. Si nota nella carta per le medie, però, dal sottogruppo 13 in poi una deriva decrescente. Dall'analisi della carta EWMA si può notare che i punti 647, 648, 651, 652 sono fuori dal limite di controllo superiore. Essi si riferiscono agli ultimi punti del gradino 12 e i primi di quello successivo. La portata era di 4 e 4,2 m³/h nella rampa in discesa. Il numero dei punti fuori controllo è limitato e, probabilmente, si tratta di scostamenti dovuti ad instabilità del flusso visto che ci troviamo tra due gradini. Si nota inoltre una deriva decrescente dopo le osservazioni fuori controllo. Non conosciamo la causa che ha provocato ciò, ma possiamo ipotizzare che ci sia stato un cambiamento nelle condizioni con le quali era iniziata l'acquisizione (temperatura o pressione in ingresso dell'aria).

Questa acquisizione, però, era composta da 800 osservazioni. Questo ha permesso di creare solo 16 sottogruppi di numerosità 50 e, di conseguenza, il numero di punti sulle carte di controllo è limitato.

5.3 Sensore numero 21

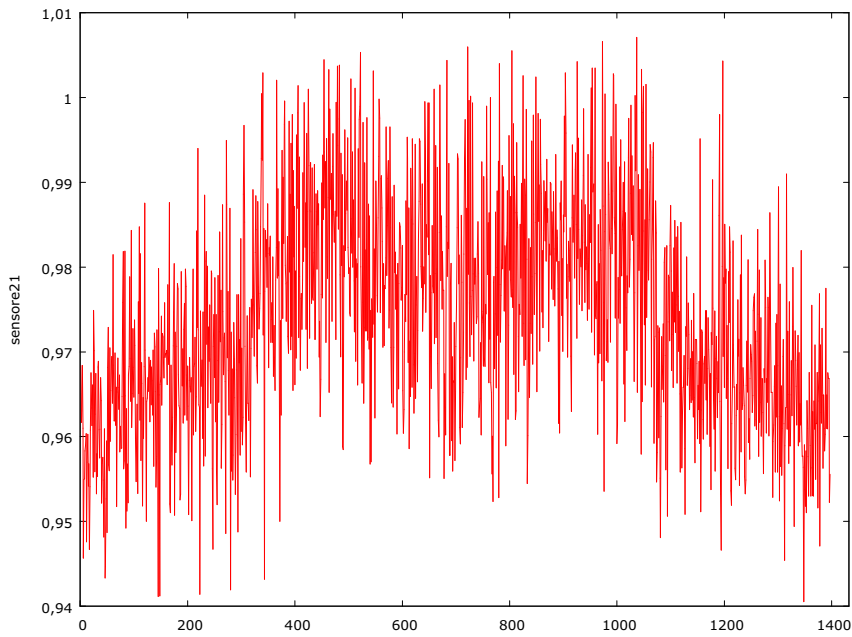


Figura 5.36 grafico della serie del rapporto tra sensore numero 21 e micromanometro

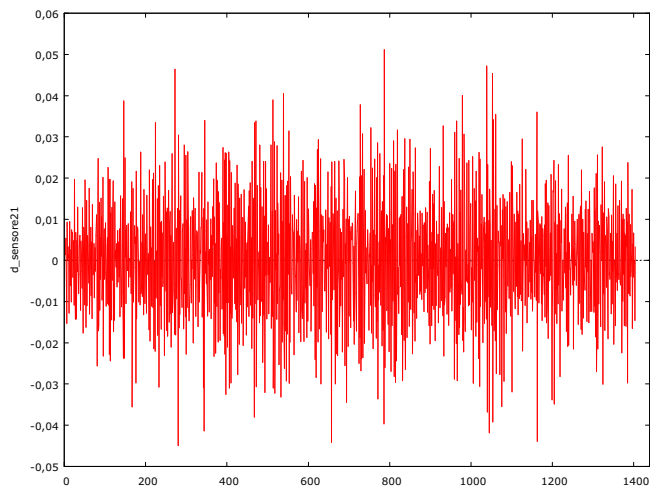


Figura 5.37 grafico della serie delle differenze prime

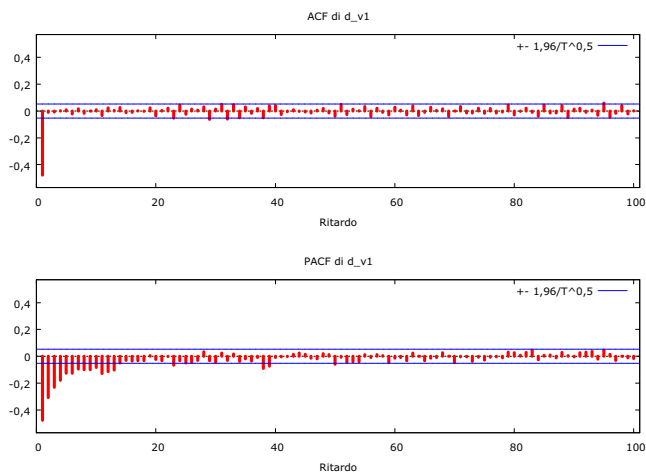


Figura 5.38 stima della funzione di autocorrelazione totale e parziale della serie differenziata

Da una prima analisi del grafico della serie, relativa al sensore 21, essa appare stazionaria in varianza, ma non in media.

Un operatore differenza di ordine uno rende la media della serie più stabile nel tempo (vedi figura 5.37).

Per questa serie il trend parabolico sembra meno evidente.

Il correlogramma parziale (figura 5.38) risulta significativo per i primi 12 ritardi, poi sembra annullarsi.

In quello totale si alternano autocorrelazioni positive e negative fuori dalle bande.

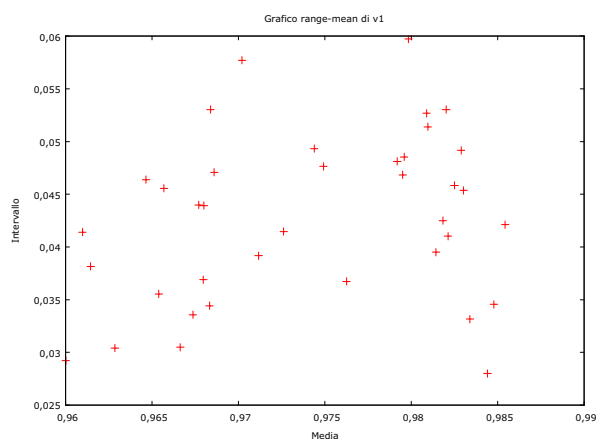


Figura 5.39 grafico range-mean della variabile rapporto 38 sottogruppi di dimensione 37

Il grafico range-mean conferma la stazionarietà in varianza della serie. L'ampiezza delle oscillazioni sembra rimanere costante per tutte le osservazioni. Infatti il p-value, che indica che la pendenza sia nulla, è pari a 0,12. Accettiamo dunque l'ipotesi che la retta stimata con il metodo dei minimi quadrati abbia coefficiente angolare pari a zero.

Il modello che meglio è risultato adattarsi ai dati è un ARIMA(2,1,3).

	Coefficiente	Errore Std.	Rapporto z	p-value
Φ_1	1,75570	0,0113853	154,2	0,0000
Φ_2	-0,983046	0,0122171	-80,46	0,0000
θ_1	-2,67478	0,0196767	-135,9	0,0000
θ_2	2,59909	0,0371983	69,87	0,0000
θ_3	-0,911122	0,0188087	-48,44	0,0000

Media var. dipendente	-4,35e-06	SQM var. dipendente	0,014219
Media innovazioni	0,000033	SQM innovazioni	0,010634
Log-verosimiglianza	4379,615	Criterio di Akaike	-8747,230
Criterio di Schwarz	-8715,756	Hannan-Quinn	-8735,466

	Reale	Immaginario	Modulo
Radice AR 1	0,8930	-0,4688	1,0086
Radice AR 2	0,8930	0,4688	1,0086
Radice MA 1	1,0596	0,0000	1,0596
Radice MA 2	0,8965	-0,4817	1,0177
Radice MA 3	0,8965	0,4817	1,0177

Tutti i coefficienti hanno un p-value prossimo a zero. Le radici della parte autoregressiva sono maggiori di uno in modulo. Il processo può essere definito stazionario.

Verifichiamo l'adeguatezza del modello tramite le analisi diagnostiche.

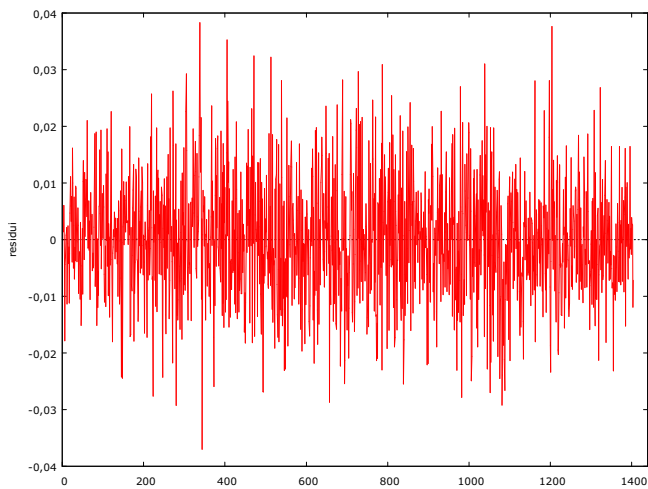


Figura 5.40 serie dei residui del modello ARIMA(213)

L'andamento della serie sembra totalmente casuale.

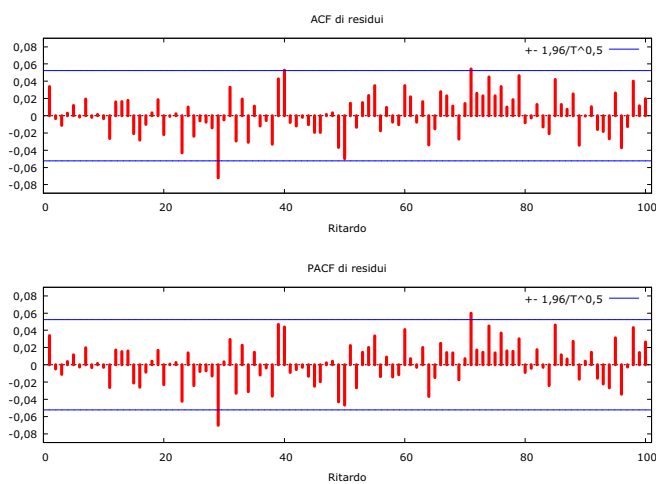


Figura 5.41 stima della funzione di autocorrelazione parziale e totale dei residui

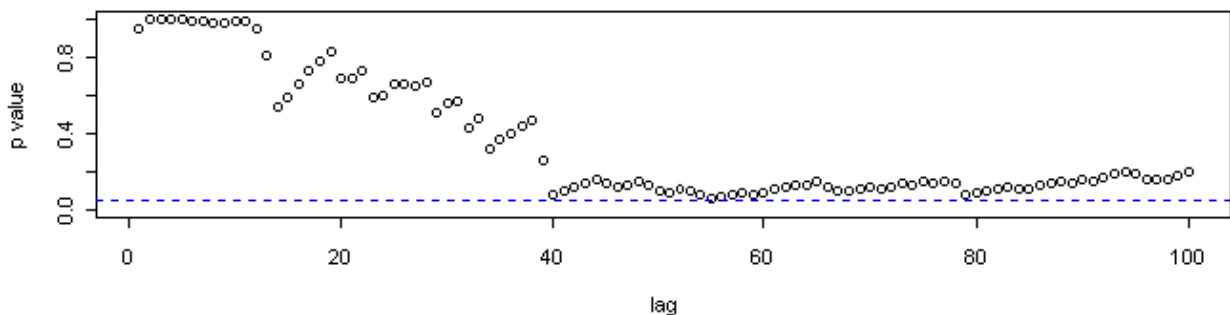


Figura 5.42 p-value della statistica test di Ljung-Box

Dall'analisi dei grafici nella pagina precedente non si evidenzia alcuna correlazione significativa. I residui sembrano generati da un processo white noise. Il test di Ljung-Box porta ad accettare l'ipotesi nulla con un p-value uguale a 0,2453 (statistica test=109,3643).

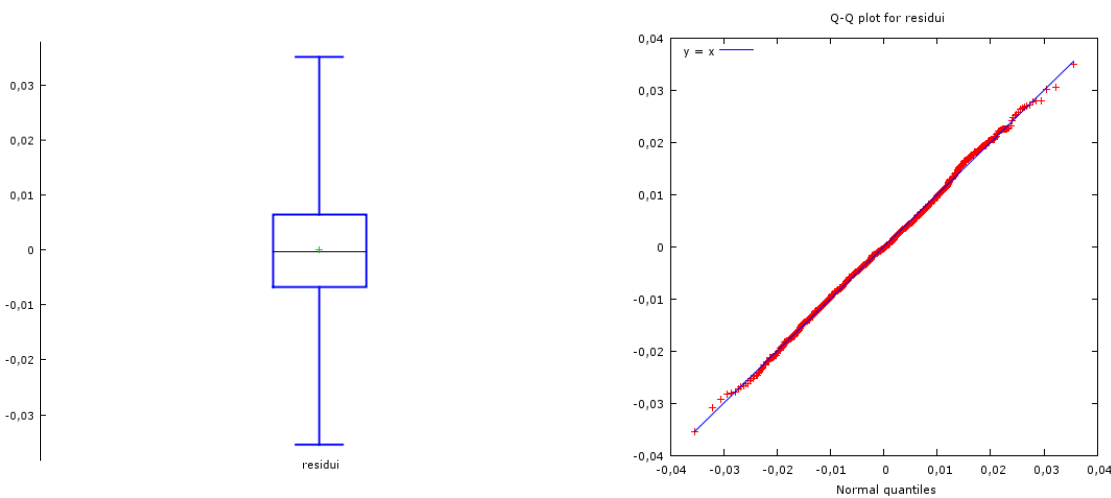


Figura 5.43 box plot e Q-Q plot dei residui

Il box plot mostra come la media (indicata nella figura 5.43 con una croce) sia molto vicina alla mediana. L'ampiezza della scatola appare contenuta e, di conseguenza, lo è la variabilità dei dati. Il diagramma quantile-quantile mostra che i punti sembrano disporsi lungo la retta $y = x$. Gli scostamenti da essa sono minimi. I residui, quindi, sembrano avere un comportamento gaussiano.

Test per la normalità effettuati:

W di Shapiro-Wilk = 0,998412, con p-value 0,221377

Test di Jarque-Bera = 0,541607, con p-value 0,762766

Il livello di significatività osservato di entrambi i test effettuati per la verifica della normalità portano ad accettare l'ipotesi nulla.

L'ultima caratteristica da verificare per i residui è che questi siano omoschedastici. Graficamente i punti sembrano disposti in modo casuale. Non sembra essere presente alcun tipo di sistematicità.

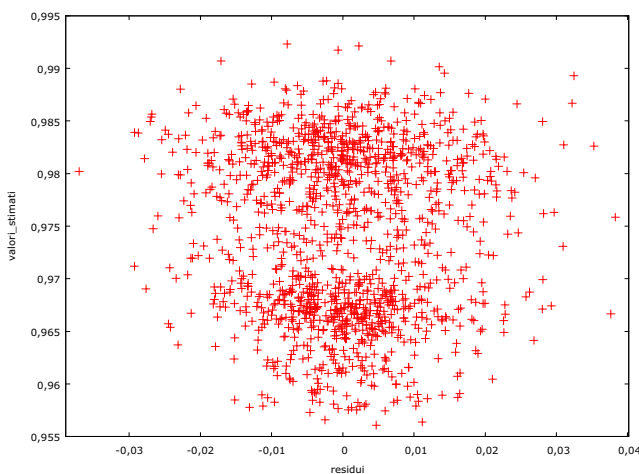


Figura 5.44 diagramma di dispersione dei residui del modello rispetto ai valori stimati dallo stesso

Proseguiamo con il disegno delle carte di controllo applicate ai residui standardizzati del modello scelto.

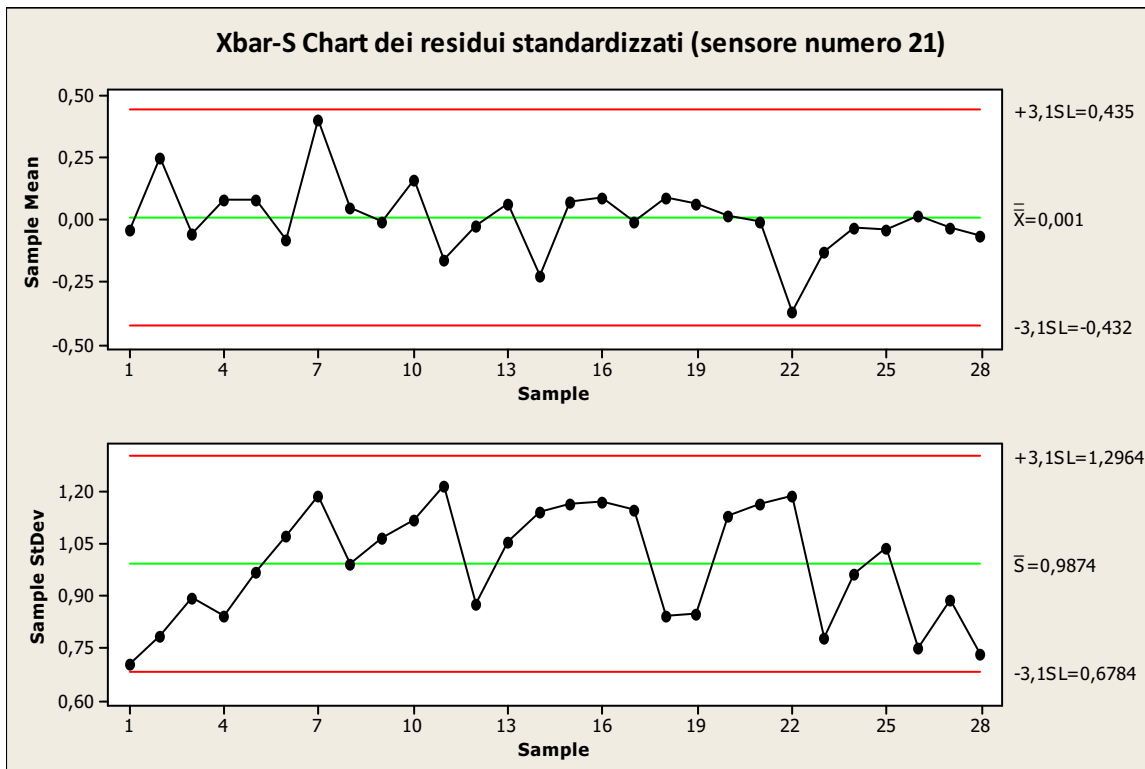


Figura 5.45

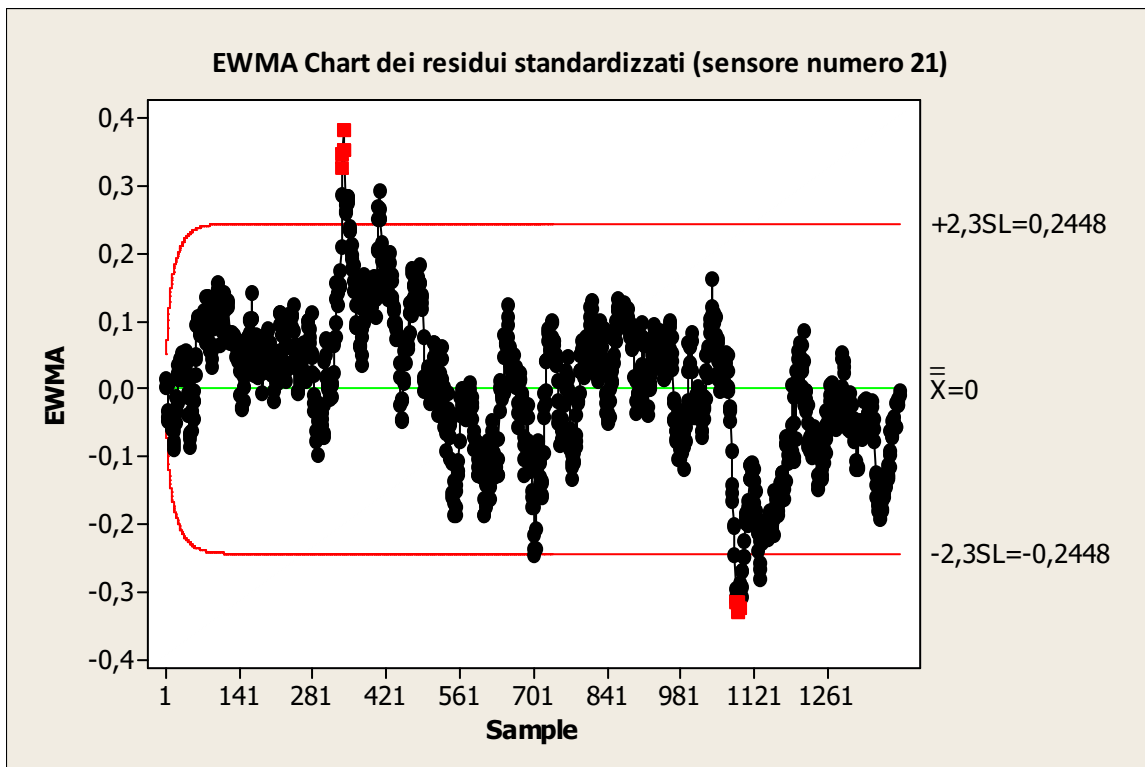


Figura 5.46

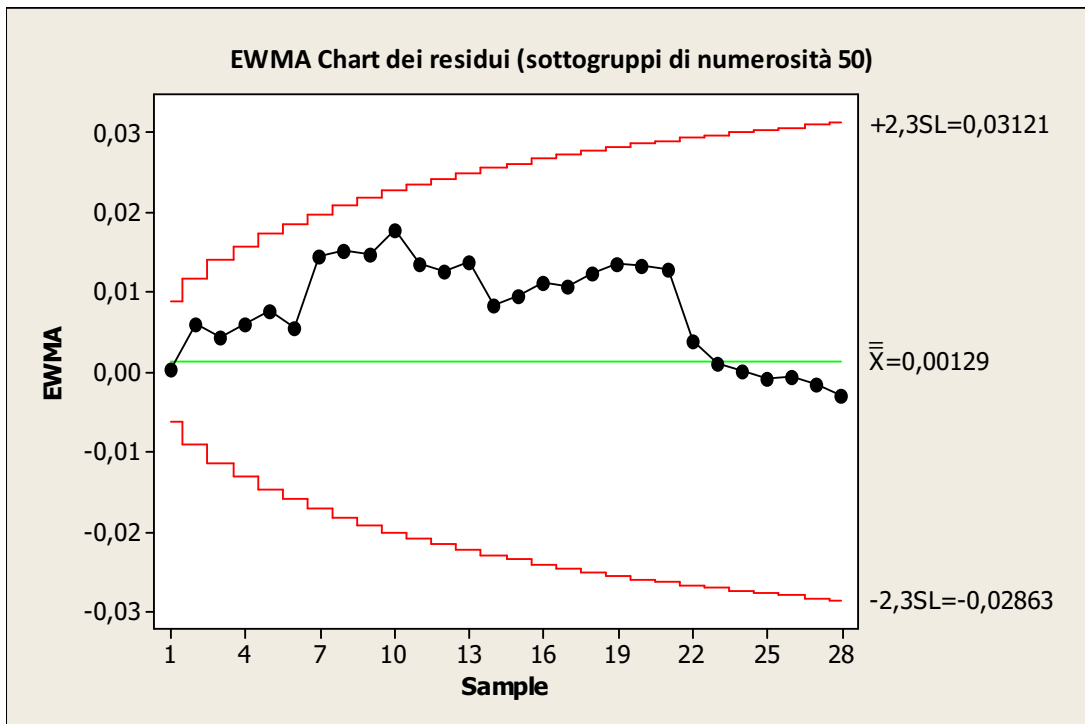


Figura 5.47

La carta Shewhart per la media è in controllo. Si nota, però, una leggera tendenza del parametro di posizione a decrescere nel tempo. La carta, costruita sulla base della deviazione standard, mostra un trend crescente per la rampa in ascesa (sottogruppi 1-14). Sembra esserci una diminuzione della variabilità ogni 4 sottogruppi (campioni 4, 8 e 12). Anche nella rampa in discesa si nota un trend decrescente ed andamenti ciclici.

La carta EWMA presenta i punti dal 336 al 406 oltre i limite di controllo superiore e le osservazioni dalla 1090 alla 1100 fuori dal limite di controllo inferiore. Il primo gruppo di punti corrisponde ad un flusso di 4,5 e 5 m³/h nella rampa in ascesa, mentre per il secondo gruppo il flusso è di 4 m³/h nella rampa in discesa. Osservando la figura 5.47 si nota che dal sottogruppo 23 in poi la media si stabilizza sotto la linea centrale, mentre tutti i punti precedenti, tranne il primo, sono risultati sempre sopra. Non conosciamo le cause che hanno portato alla stima del valor medio più basso, ma ipotizziamo che ciò sia dovuto ad una certa instabilità delle condizioni che hanno caratterizzato l'acquisizione.

5.4 Sensore numero 15

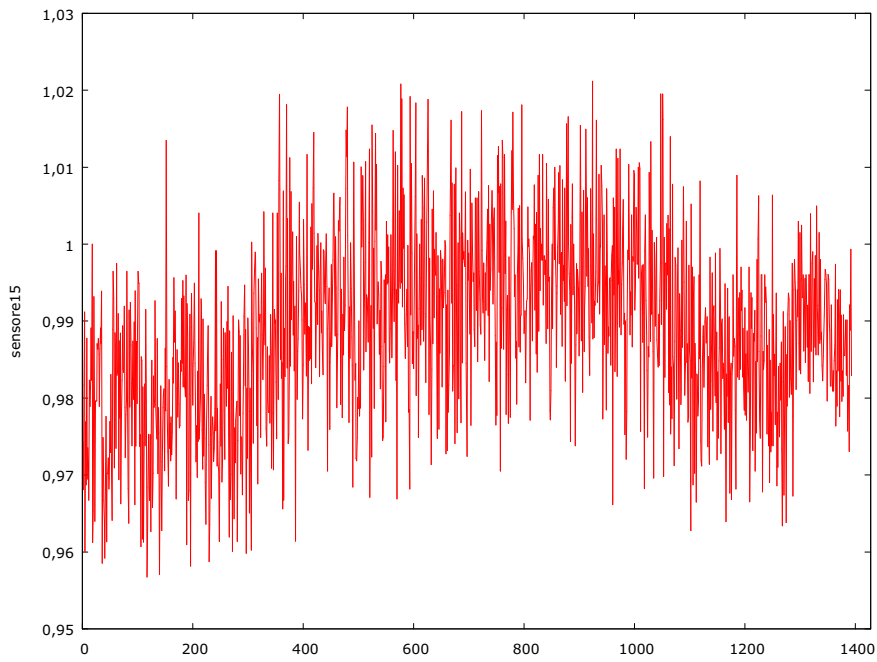


Figura 5.48 serie del rapporto tra sensore numero 15 e micromanometro

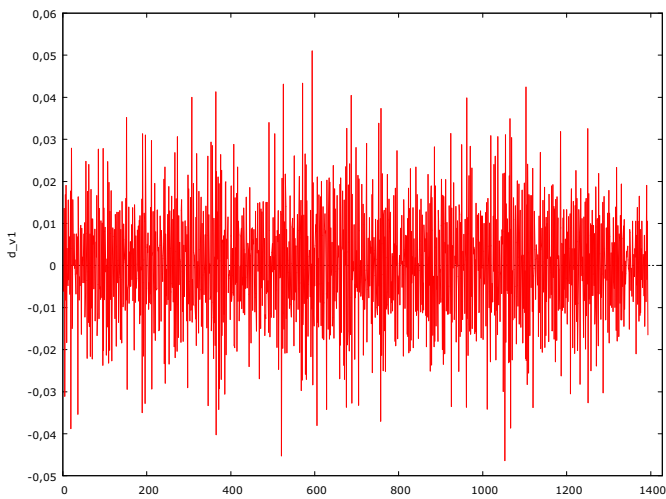


Figura 5.49 serie delle differenze prime

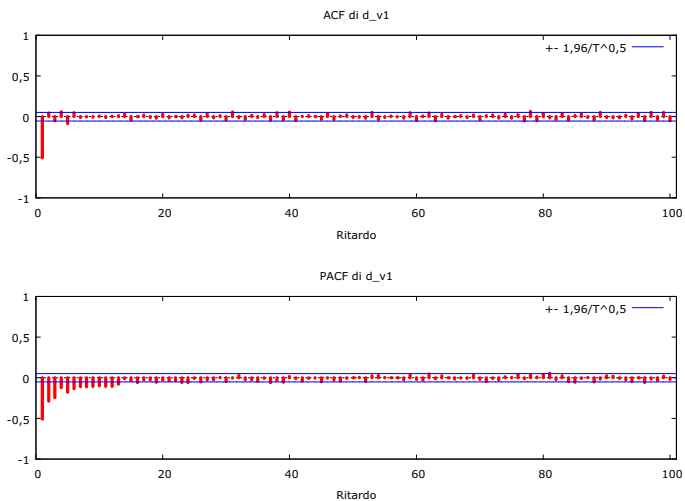


Figura 5.50 stima della funzione di autocorrelazione totale e parziale della serie delle differenze prime

La stima della funzione di autocorrelazione parziale (figura 5.50), dopo i primi 12 ritardi significativamente diversi da zero, sembra annullarsi totalmente.

Una differenziazione di ordine unitario (vedi figura 5.49) sembra essere sufficiente a rendere la serie stabile in media.

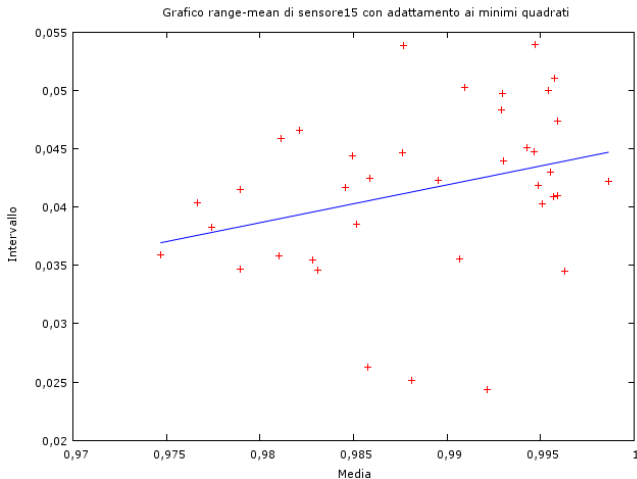


Figura 5.51 grafico range-mean (campione completo) 38 sottogruppi di dimensioni 37

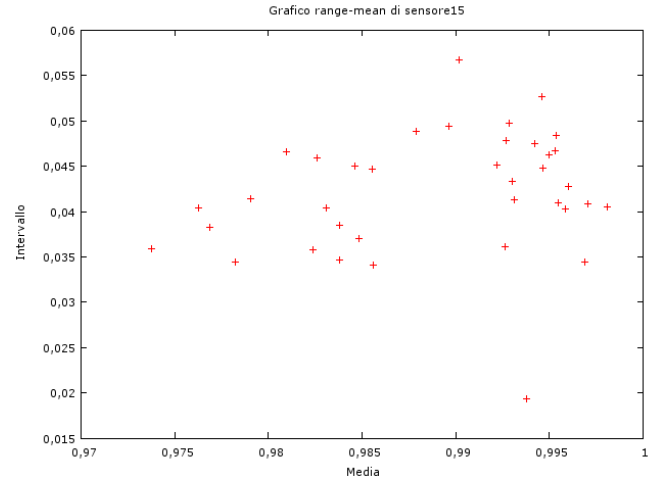


Figura 5.52 grafico range-mean (osservazioni 1-1300) 37 sottogruppi di dimensione 36

Osservando il grafico della serie (figura 5.48) questa sembra non stazionaria in varianza. In realtà sono le ultime 100 osservazioni ad avere una variabilità più contenuta delle altre. Infatti, considerando solo le prime 1300 osservazioni, il grafico range-mean (figura 5.52) non mostra alcuna retta e quindi accettiamo l'ipotesi che la pendenza sia nulla (p-value pari a 0,20). Mentre, se costruiamo la stessa tipologia di grafico considerando il campione completo, il p-value per l'ipotesi nulla che il coefficiente angolare sia zero è pari a 0,052 (figura 5.51). La pendenza della retta stimata è pari a 0,32 ed essendo positiva sembra indicare che la variabilità cresce all'aumentare della media.

Il trend parabolico in questa serie non appare così marcato. Anche se i tre parametri considerati (costante, trend e trend al quadrato) sono tutti significativi, il coefficiente di determinazione non è molto elevato. La parabola stimata è la seguente: $Y = -3,47720e^{-8} (\text{trend}^2) + 5,51557e^{-5} (\text{trend}) + 0,972823$

	Coefficiente	Errore Std.	Rapporto t	p-value
costante	0,972823	0,000849496	1145	0,0000
trend	5,51557e-05	2,81455e-06	19,60	1,07e-075
trend al quadrato	-3,47720e-08	1,95502e-09	-17,79	6,37e-064
R-quadro	0,226354	R-quadro corretto	0,225241	

Proviamo quindi a compiere i passi successivi senza detrendizzare la serie.

Il modello che meglio risulta adattarsi ai dati è risultato essere un ARIMA(5,1,1) vincolato ad avere Φ_3 e Φ_4 nulli.

	Coefficiente	Errore Std.	Rapporto z	p-value
Φ_1	0,0589051	0,0280579	1,890	0,0587
Φ_2	0,0661172	0,0280534	2,357	0,0184
Φ_5	-0,0682132	0,0279008	-2,445	0,0145
θ_1	-0,957208	0,00934516	-102,4	0,0000

Media var. dipendente	6,23e-06	SQM var. dipendente	0,013840
Media innovazioni	0,000166	SQM innovazioni	0,010216
Log-verosimiglianza	4404,264	Criterio di Akaike	-8798,529
Criterio di Schwarz	-8772,336	Hannan-Quinn	-8788,735

	Reale	Immaginario	Modulo
Radice AR 1	1,3961	0,9117	1,6674
Radice AR 2	1,3961	-0,9117	1,6674
Radice AR 3	-1,6764	0,0000	1,6764
Radice AR 4	-0,5579	-1,6835	1,7735
Radice AR 5	-0,5579	1,6835	1,7735
Radice MA 1	1,0447	0,0000	1,0447

Procediamo con la verifica del modello.

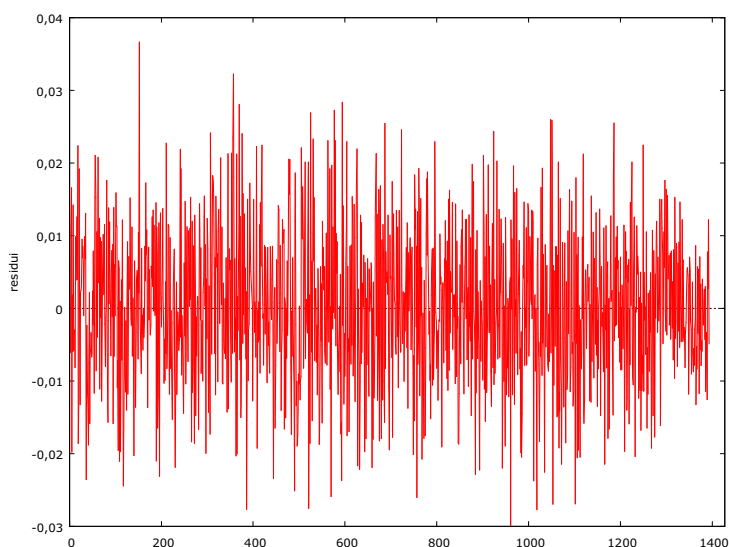


Figura 5.53 serie dei residui del modello ARIMA(511) vincolato

La serie dei residui ha un andamento casuale. Si nota, però, che la variabilità delle ultime 100 osservazioni risulta minore rispetto a quella delle precedenti.

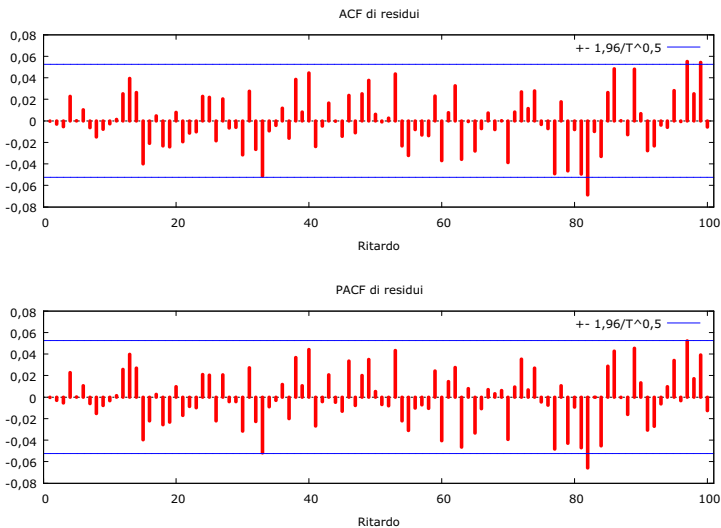


Figura 5.54 correlogramma globale e parziale dei residui

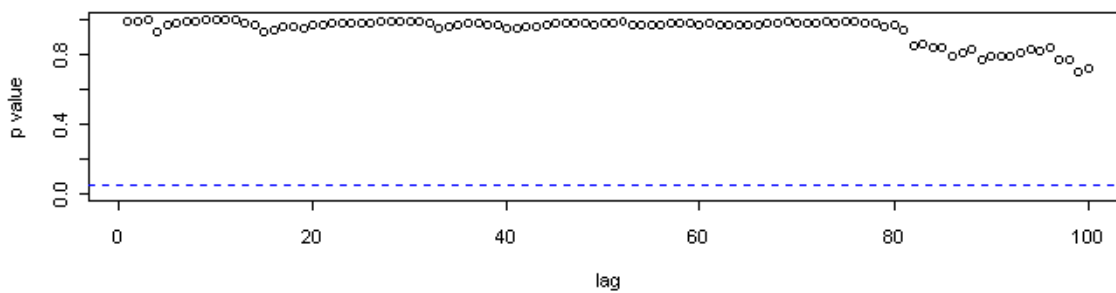


Figura 5.55 p-value della statistica test di Ljung-Box

Il test di Ljung e Box($Q(100) = 91,5766$, $p\text{-value} = 0,7142$) e l'analisi delle correlazioni dei residui portano a ritenere adeguato il modello utilizzato. Non sembra esserci autocorrelazione residua. Infatti, per quattro ritardi su cento considerati, la funzione di autocorrelazione totale stimata è significativamente diversa da zero e i p-value della statistica di Ljung-Box sono tutti ampiamente sopra la soglia del 95%.

Verifichiamo che le innovazioni abbiano una distribuzione normale.

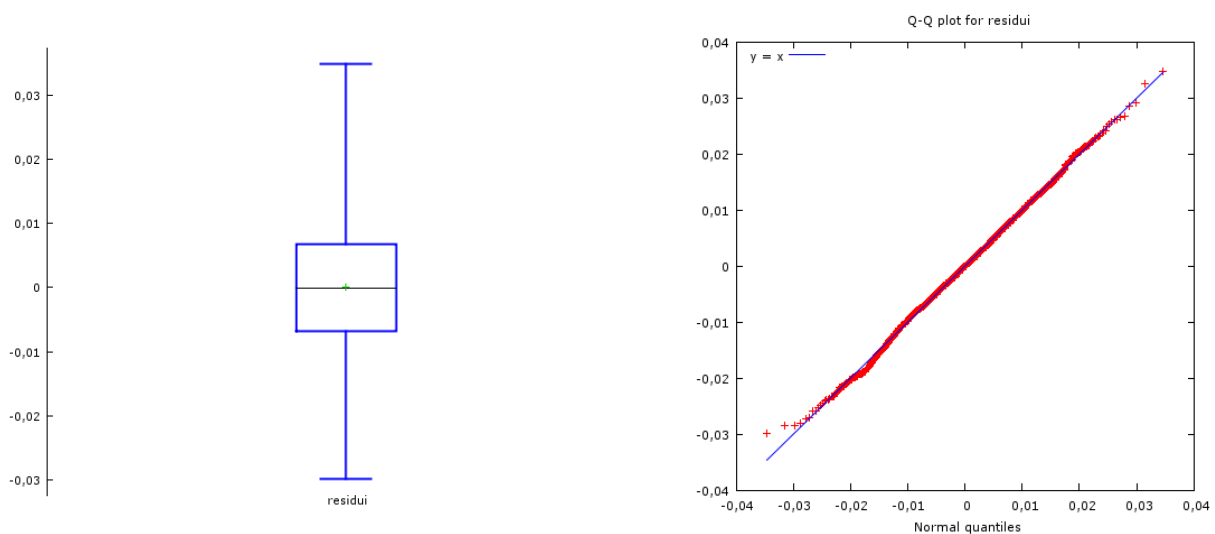


Figura 5.56 box plot e diagramma quantile-quantile dei residui

Dall'analisi della figura 5.56 i residui non sembrano essere caratterizzati da scostamenti dalla normalità. I punti del q-q plot sono perfettamente allineati lungo la retta (indice di asimmetria di Fisher=-0,004). Il box plot mostra come media e mediana siano coincidenti e la lunghezza dei baffi sia quasi la stessa.

I p-value dei test per la normalità

W di Shapiro-Wilk = 0,998919, con p-value 0,581104

Test di Jarque-Bera = 0,049734, con p-value 0,97544

confermano che i residui si distribuiscono in modo normale.

Continuiamo ad analizzare i residui accertandoci che l'ipotesi di omoschedasticità sia soddisfatta con il diagramma di dispersione sottostante.

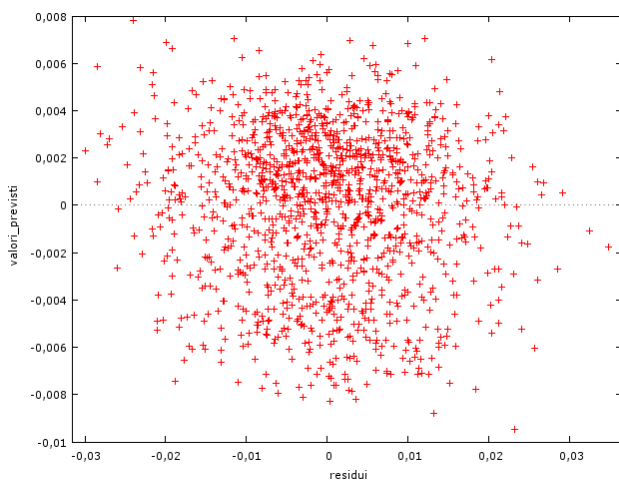


Figura 5.57 diagramma di dispersione dei residui del modello rispetto ai valori stimati dallo stesso

La nuvola di punti si distribuisce in maniera casuale. Non sembra esserci evidenza di eteroschedasticità per i residui.

Procediamo applicando i consueti strumenti per il controllo statistico di processo ai residui standardizzati del modello ARIMA(5,1,1).

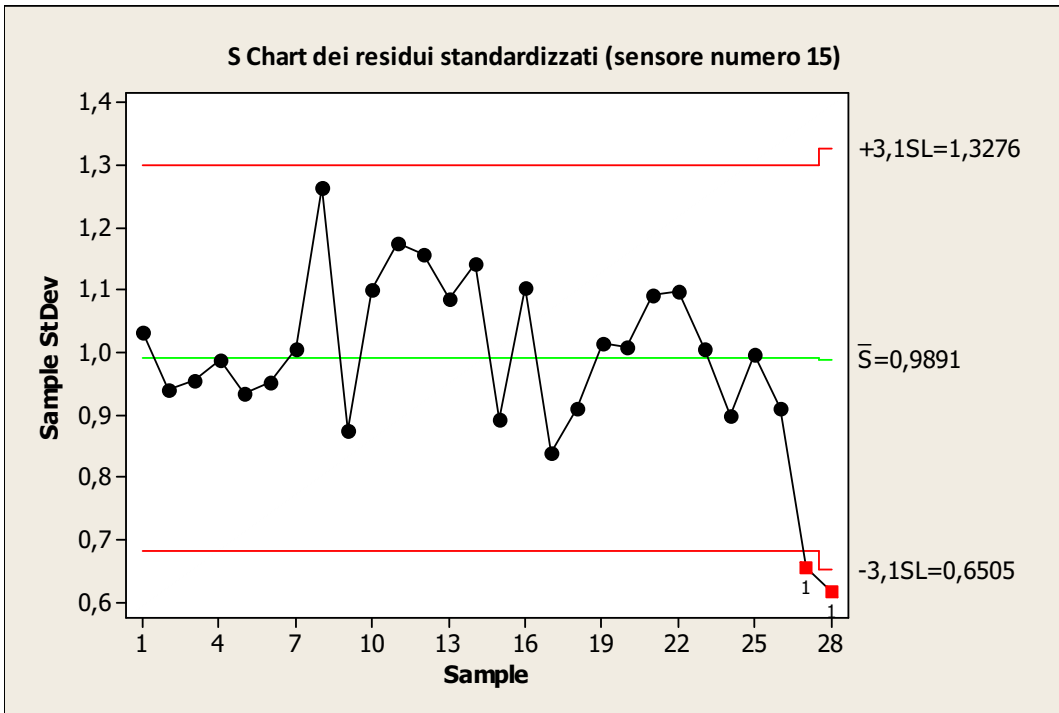


Figura 5.58

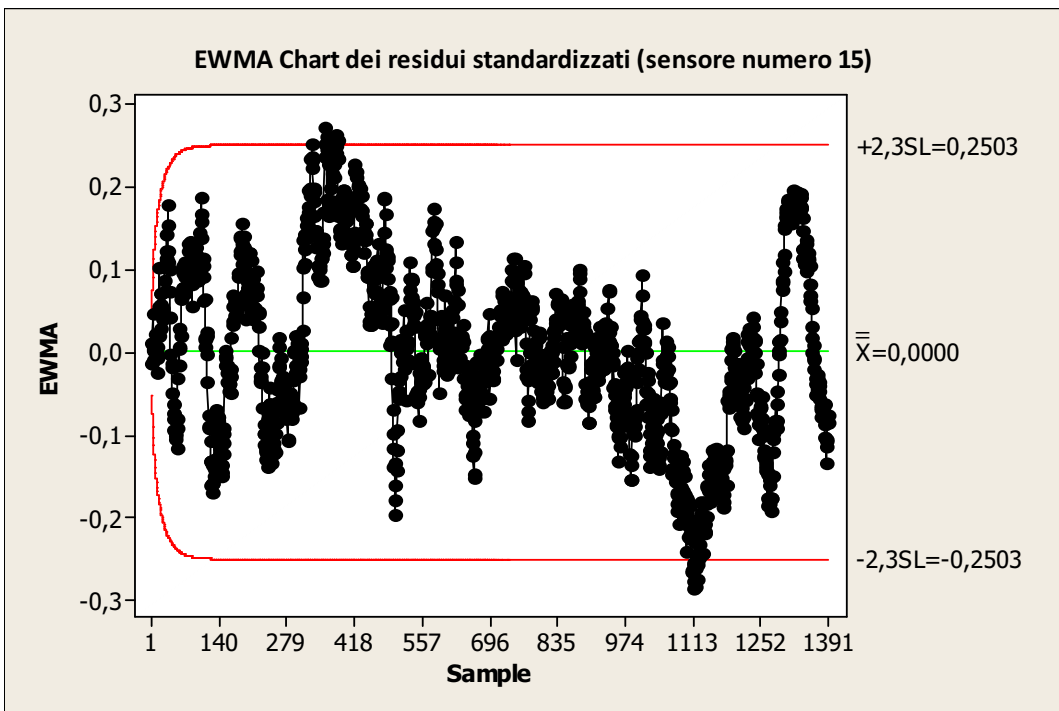


Figura 5.59

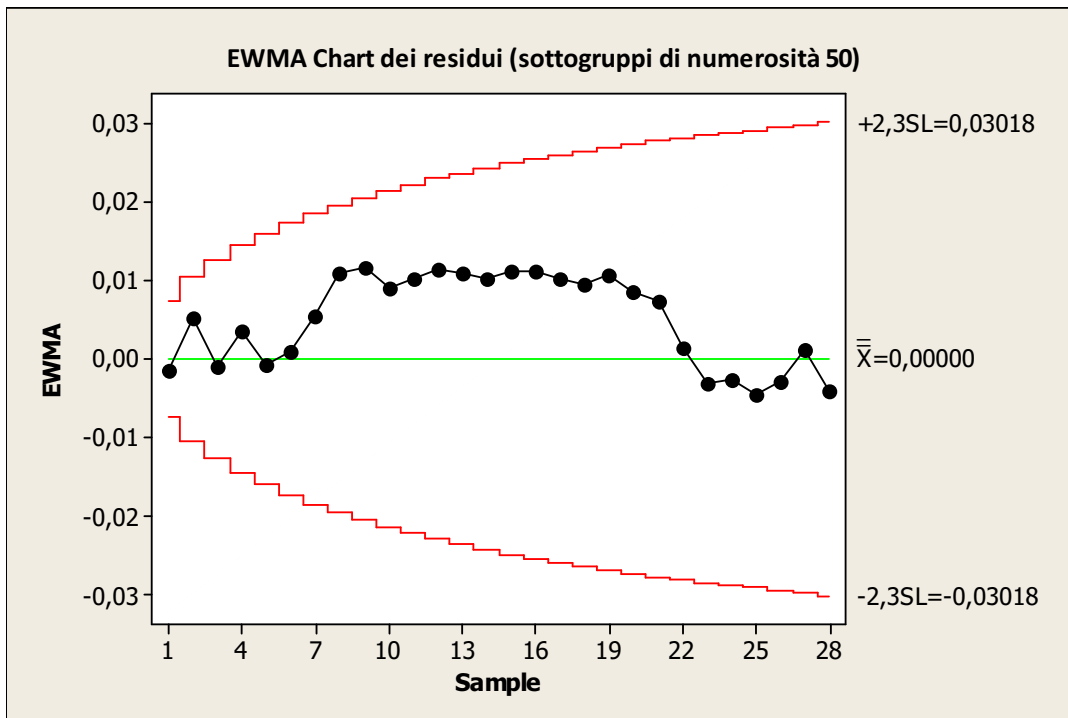


Figura 5.60

La carta di tipo S mostra due fuori controllo per i campioni 27 e 28 (flusso 2 e 1,5 m³/h). Per questo motivo non è stata rappresentata la carta per le medie. La variabilità dei residui sembra avere un crollo per gli ultimi due sottogruppi. Questo risulta coerente con le analisi svolte in precedenza.

La carta EWMA presenta due gruppi di una decina di punti circa ciascuno fuori dai limiti di controllo. Essi corrispondono ad un flusso di 4,5 m³/h nella rampa in ascesa e 3,5 m³/h in quella in discesa. Si nota inoltre una deriva decrescente dopo i primi punti fuori controllo. Osservando la figura 5.60 si osserva che anche per questa acquisizione dall'osservazione 1100 in poi, ovvero dal sottogruppo 22, quasi tutti i punti si stabilizzano sotto la linea centrale. Anche in questo caso non ne conosciamo la causa e possiamo solo fare le stesse ipotesi del caso precedente, ovvero la mancanza di stabilità di temperatura e pressione.

5.5 Sensore numero 21A

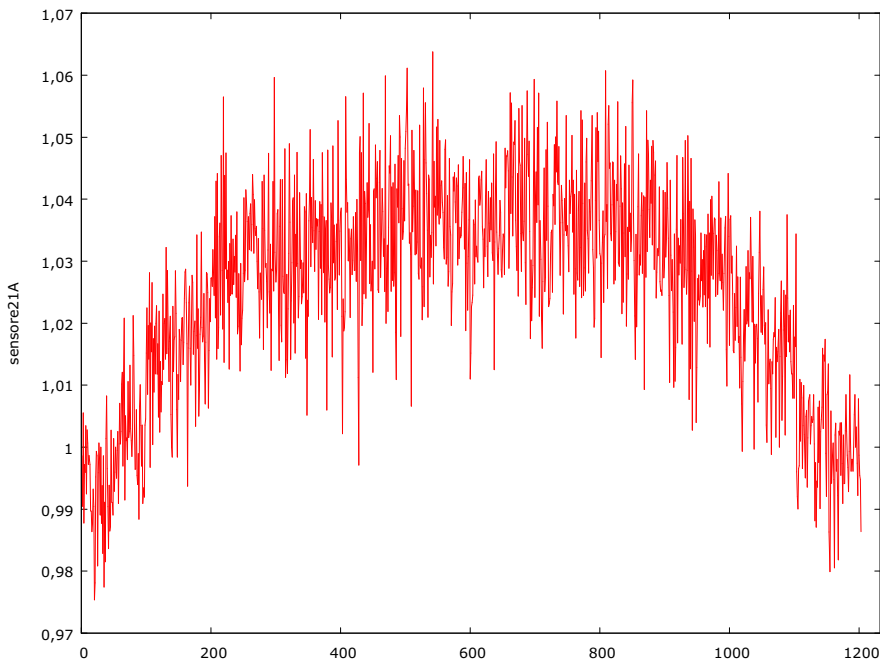


Figura 5.61 serie del rapporto tra sensore numero 21A e micromanometro

La serie è sicuramente non stazionaria in media e la variabilità delle osservazioni non rimane costante. Da una prima analisi della figura 5.61 le prime e le ultime 100 osservazioni mostrano una varianza minore delle altre. I grafici range-mean sottostanti confermano questa ipotesi. Infatti il p-value per l'ipotesi nulla che la pendenza della retta sia zero derivante dalla figura 5.62 è pari a 0,021. La stima del coefficiente angolare è di 0,23 ed essendo positiva indica che la variabilità della serie cresce all'aumentare della media. Se costruiamo la stessa tipologia di grafico considerando solo le osservazioni 100-1100 (figura 5.63) otteniamo un p-value di 0,121. In questo caso, invece, accettiamo l'ipotesi nulla e da ciò deduciamo che l'ampiezza delle osservazioni è rimasta costante.

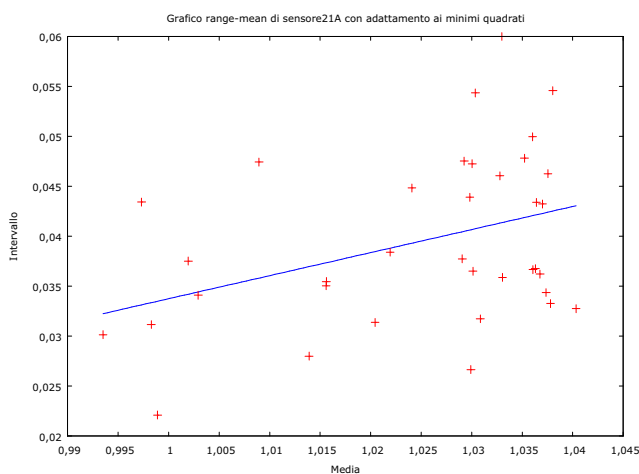


Figura 5.62 range-mean plot (campione completo)
36 sottogruppi di dimensione 34

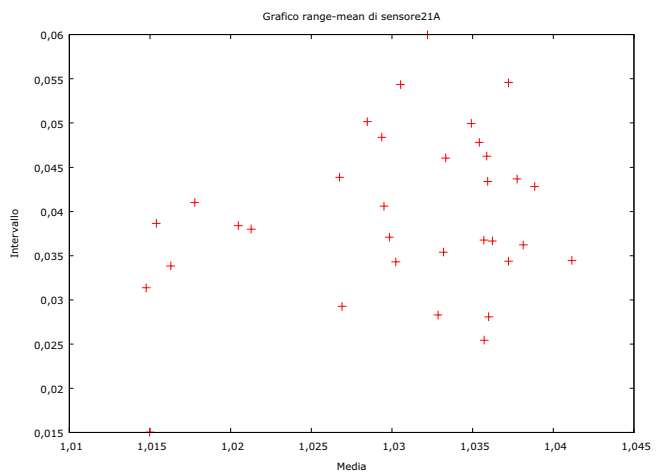


Figura 5.63 range-mean plot (osservazioni 100-1100)
33 sottogruppi di numerosità 31

La serie considerata presenta un trend parabolico abbastanza marcato, proviamo a stimarlo tramite un modello OLS. I risultati sono i seguenti:

	Coefficiente	Errore Std.	Rapporto t	p-value
costante	0,994488	0,000872266	1140	0,0000
trend	0,000148817	3,34602e-06	44,48	4,72e-256
trend al quadrato	-1,21568e-07	2,69098e-09	-45,18	3,74e-261
R-quadro	0,630709	R-quadro corretto	0,630093	

Tutti e tre i coefficienti hanno un p-value minore di 0,001 e risultano quindi fortemente significativi. Il modello sembra soddisfacente visto che il coefficiente di determinazione indica che questo spiega il 63% della variabilità della variabile rapporto.

Sottraiamo, quindi, il trend stimato alla variabile rapporto ed otteniamo v1.

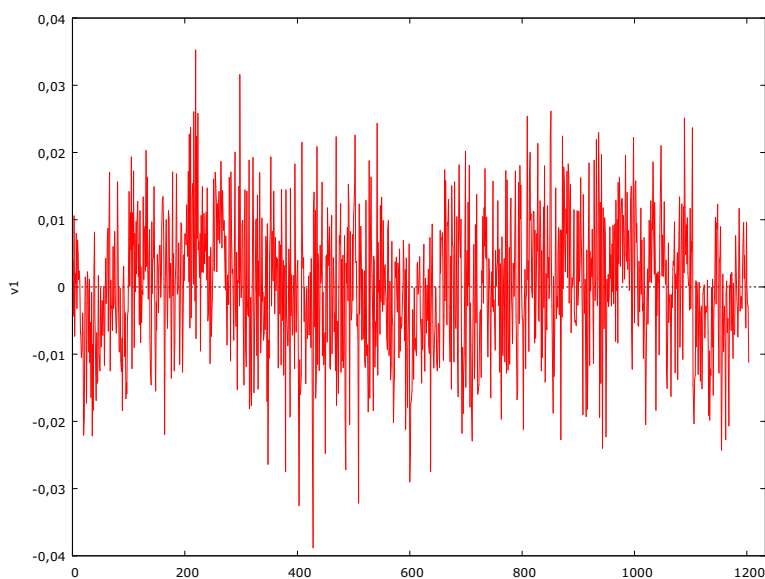


Figura 5.64 grafico della serie detrendizzata

Dal grafico risulta comunque necessaria una differenziazione di ordine uno.

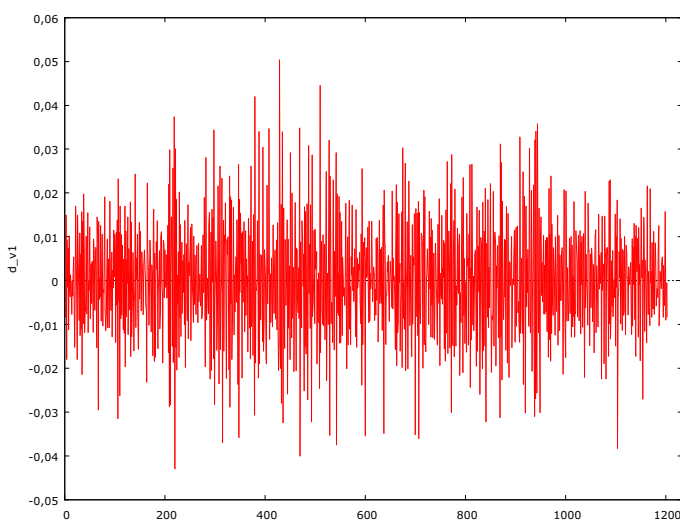


Figura 5.65 grafico della serie differenziata (ordine uno)

La serie sembra essere più omogenea relativamente alla variabilità delle osservazioni.

Al fine di identificare il modello migliore analizziamo il correlogramma parziale e totale della serie differenziata.

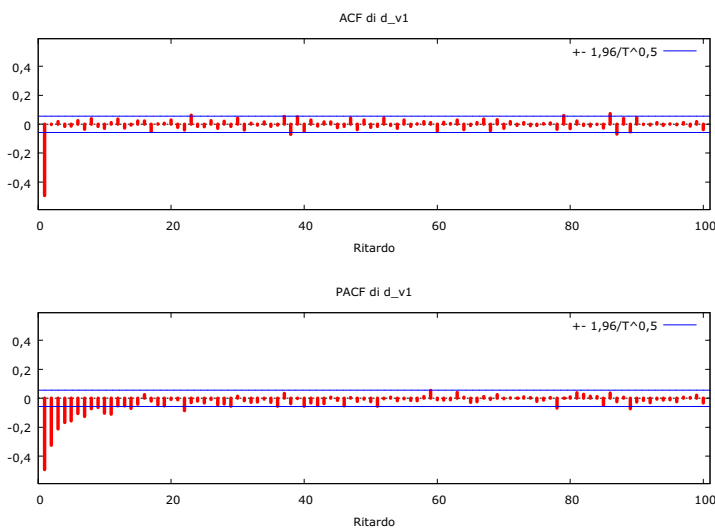


Figura 5.66 correlogramma totale e parziale della serie differenziata

La funzione di autocorrelazione parziale stimata presenta i primi 15 ritardi significativamente diversi da zero, ma poi sembra lentamente annullarsi. Quella totale, invece, presenta un alternarsi continuo di autocorrelazioni positive e negative.

Il modello che meglio sembra adattarsi ai dati è un ARIMA(2,1,3) applicato alla serie detrendizzata.

	Coefficiente	Errore Std.	Rapporto z	p-value
Φ_1	-0,348930	0,00586873	-59,46	0,0000
Φ_2	-0,983708	0,00580233	-169,5	0,0000
θ_1	-0,579138	0,0145134	-39,90	0,0000
θ_2	0,653742	0,00667109	98,00	0,0000
θ_3	-0,945396	0,0147455	-64,11	0,0000

Media var. dipendente	-0,000013	SQM var. dipendente	0,013004
Media innovazioni	-0,000025	SQM innovazioni	0,009670
Log-verosimiglianza	3866,732	Criterio di Akaike	-7721,464
Criterio di Schwarz	-7690,914	Hannan-Quinn	-7709,957

	Reale	Immaginario	Modulo
Radice AR 1	-0,1774	-0,9925	1,0082
Radice AR 2	-0,1774	0,9925	1,0082
Radice MA 1	1,0578	0,0000	1,0578
Radice MA 2	-0,1831	-0,9831	1,0000
Radice MA 3	-0,1831	0,9831	1,0000

Tutti i coefficienti hanno un livello di significatività osservato prossimo a zero. Le radici della parte autoregressiva sono, in modulo, maggiori di uno. Il processo può essere definito stazionario.

Procediamo con la verifica delle assunzioni sui residui.

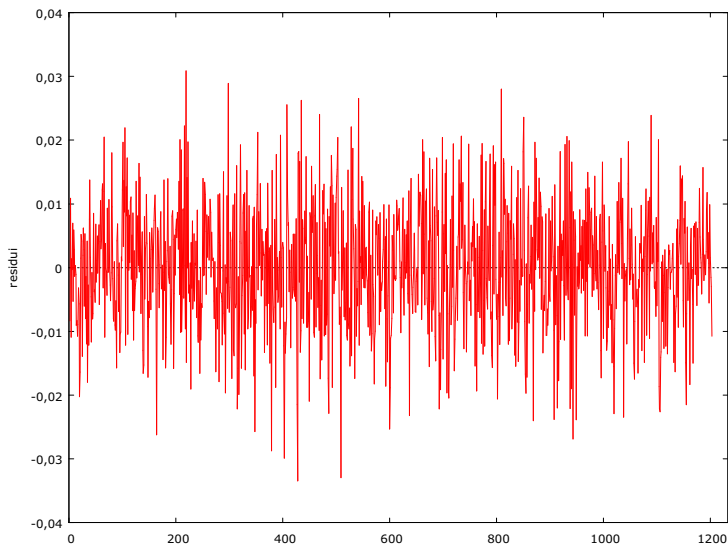


Figura 5.67 serie dei residui del modello ARIMA(213)

La serie dei residui mostra che la variabilità non è molto stabile.

Controlliamo che le innovazioni siano distribuite normalmente.

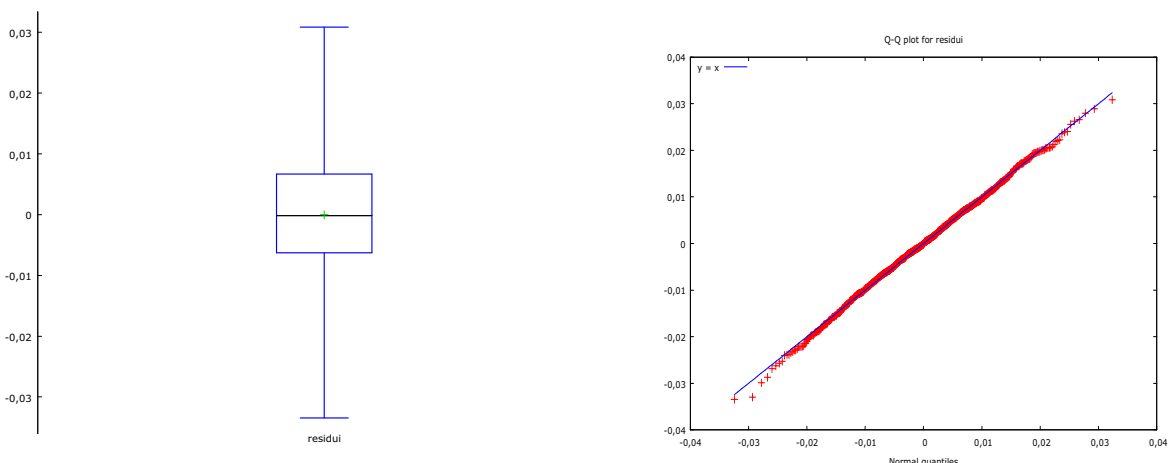


Figura 5.68 box plot e Q-Q plot dei residui del modello

I residui presentano una leggera asimmetria a sinistra (indice di asimmetria di Fisher pari a $-0,084$). Lo si può notare osservando i piccoli scostamenti dalla retta del diagramma quantile-quantile. Oltre a ciò guardando con attenzione il box plot si nota che la media non coincide perfettamente con la mediana e che la mediana non è esattamente a metà della scatola del box plot. Le innovazioni presentano una leggera curtosi (indice di curtosi di Fisher pari a $0,123$) e quindi hanno una forma leptocurtica.

Tuttavia il livello di significatività osservato di entrambi i test sulla normalità, cioè:

W di Shapiro-Wilk = $0,998832$, con p-value $0,634953$

Test di Jarque-Bera = $2,18467$, con p-value $0,335433$

indicano che non può rifiutarsi l'ipotesi che i residui siano normalmente distribuiti.

Procediamo accertandoci che i residui siano incorrelati.

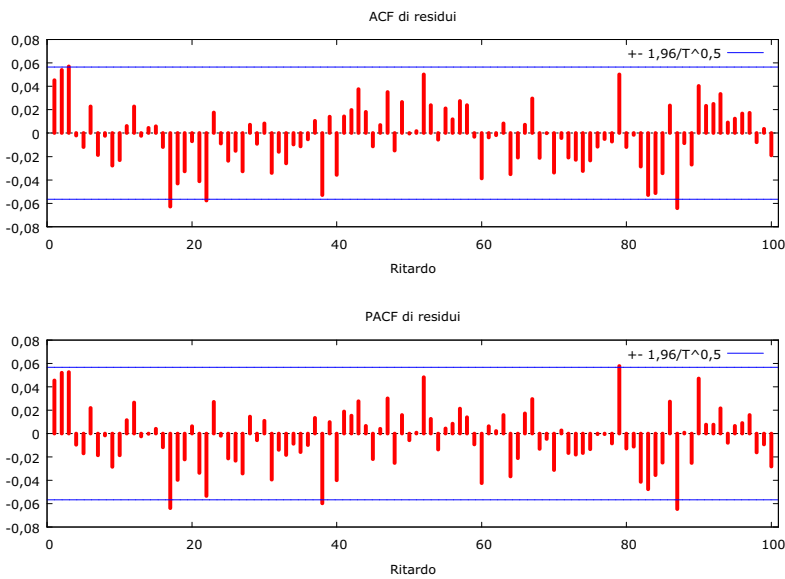


Figura 5.69 stima della funzione di autocorrelazione globale e parziale dei residui

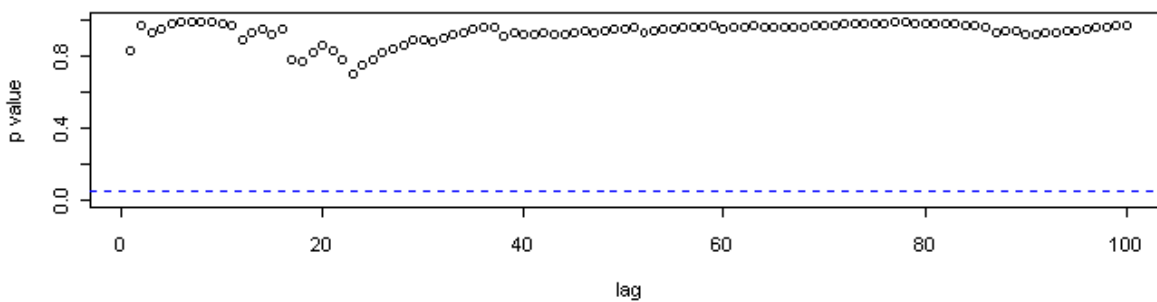


Figura 5.70 p-value della statistica test di Ljung-Box

Per quattro ritardi su cento considerati la stima della funzione di autocorrelazione globale e parziale è risultata fuori dalle bande al 95%. I p-value della statistica di Ljung-Box sono tutti ampiamente sopra la soglia.

La statistica test Q calcolata per i primi 100 ritardi vale 75,386 e, con un p-value maggiore di 0,9 accettiamo l'ipotesi nulla. I residui sembrerebbero generati da un processo white noise.

Costruiamo un diagramma di dispersione che riporta sull'asse delle ascisse i residui del modello e su quello delle ordinate i valori stimati dallo stesso. La forma della nuvola di punti (figura 5.71) sembra indicare l'omoschedasticità dei residui.

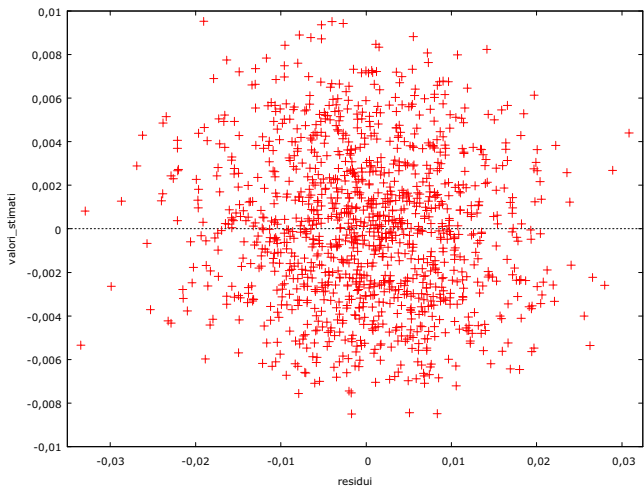


Figura 5.71 diagramma di dispersione dei residui del modello rispetto ai valori stimati dallo stesso

Analizziamo, infine, le carte di controllo costruite sui residui standardizzati del modello precedentemente stimato.

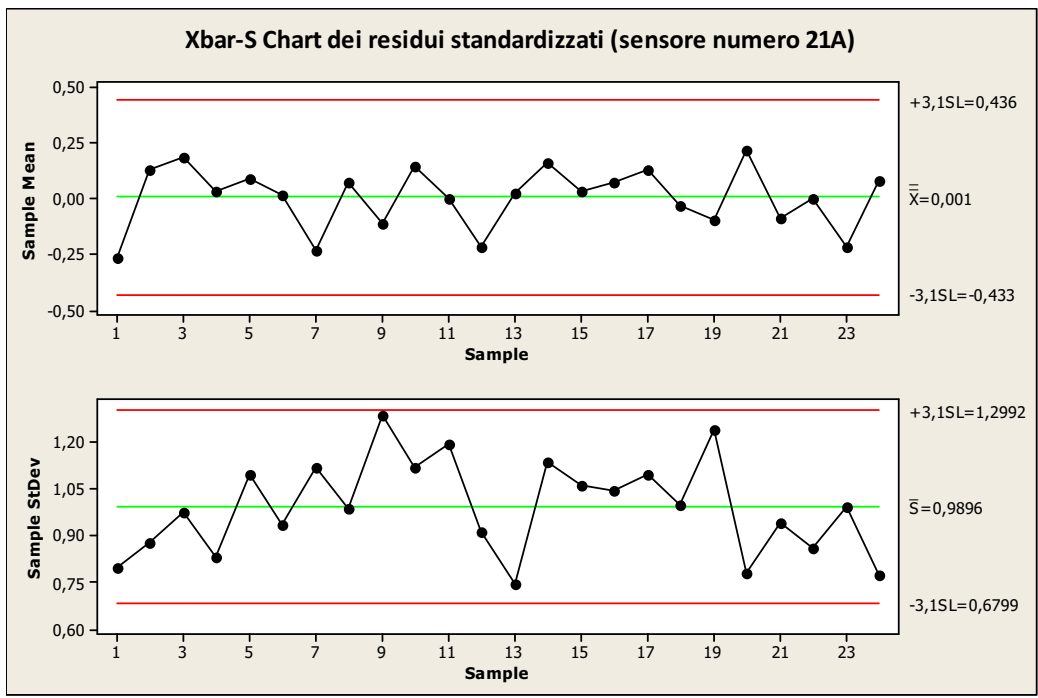


Figura 5.72

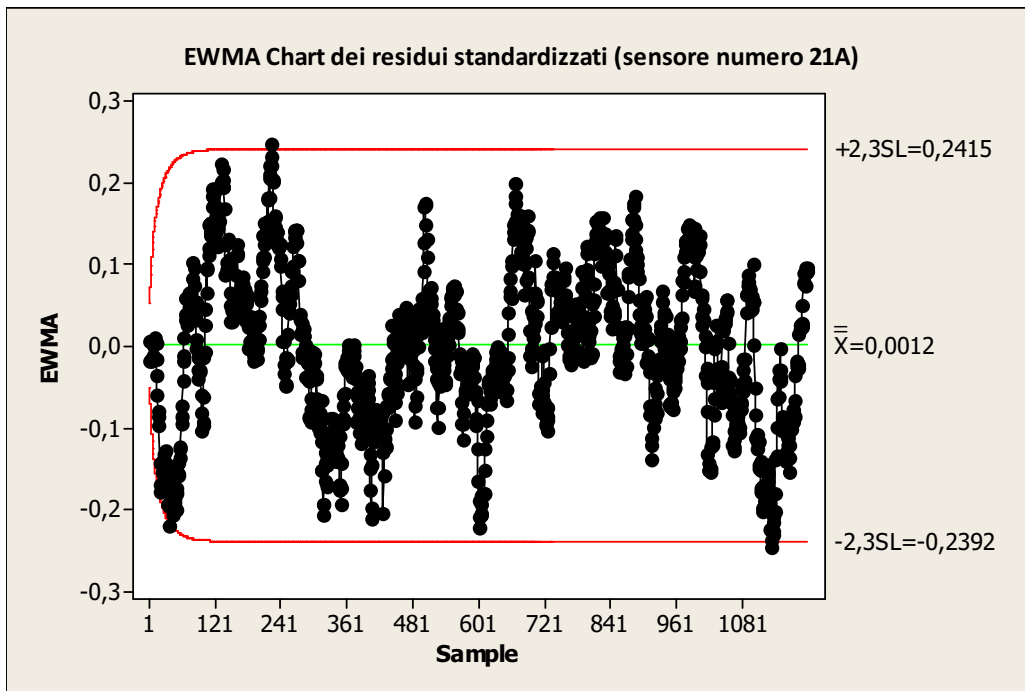


Figura 5.73

Dall'analisi della carta S la variabilità sembra avere un trend crescente fino al sottogruppo 11, si nota, però, un andamento ciclico di continua discesa e risalita. I campioni 12 e 13 hanno una variabilità molto minore degli altri. Il trend nella rampa in discesa (sottogruppi 13-24) è decrescente, coerentemente con quello che si nota per la rampa in salita. La carta Shewhart per le medie non mostra alcun fuori controllo. La carta EWMA mostra un comportamento stabile all'interno dei limiti di controllo.

5.6 Sensore numero 22

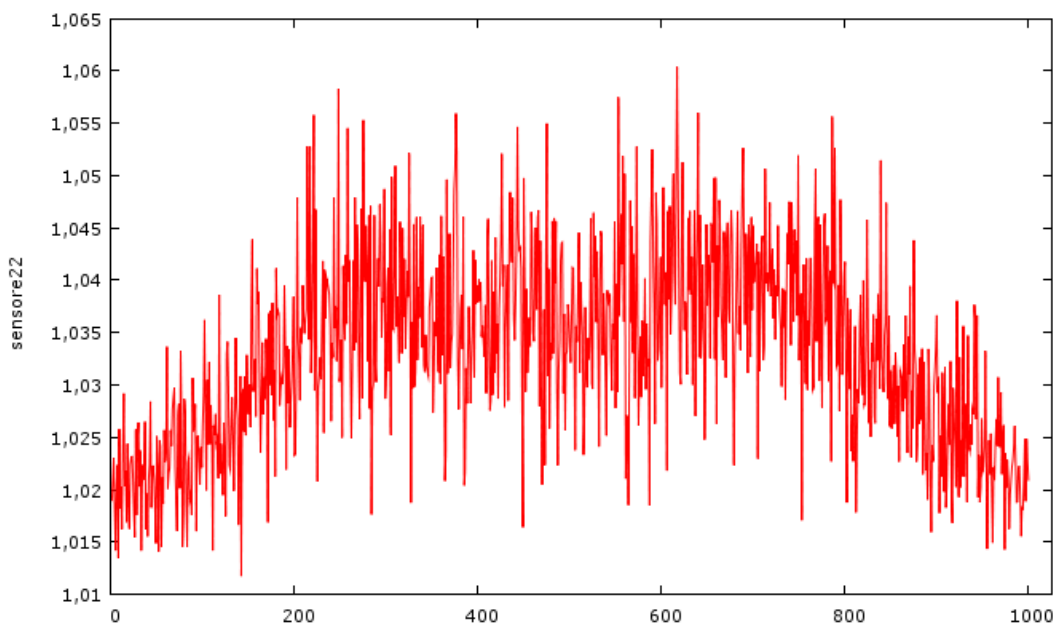


Figura 5.74 serie del rapporto tra sensore numero 22 e micromanometro

Per il sensore 22, la serie è visibilmente non stazionaria in media. Sembra mancare stabilità anche in varianza. Infatti il grafico range-mean costruito sul campione completo (figura 5.75) indica che il p-value per H_0 , che indica che la pendenza della retta sia nulla, è praticamente pari a zero. Il coefficiente angolare stimato è pari a 0,86 ed essendo positivo e fortemente significativo indica che la variabilità della serie aumenta con il crescere della media. Sono, però, le prime e le ultime 200 osservazioni a presentare una variabilità minore rispetto alle altre. La figura 5.76 conferma questa ipotesi poiché il coefficiente angolare della retta, stimata con il metodo dei minimi quadrati, può essere considerato pari a zero (p-value = 0,8879). Ancora una volta sembra che le osservazioni “centrali” della serie siano dotate di una variabilità omogenea tra loro.

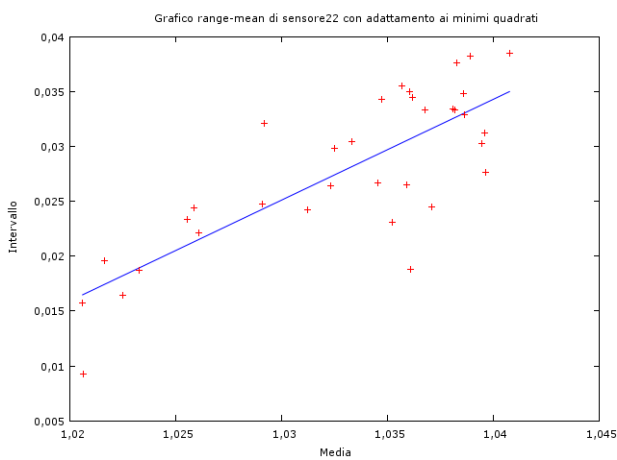


Figura 5.75 range-mean plot(campione completo)
33 sottogruppi di dimensione 31

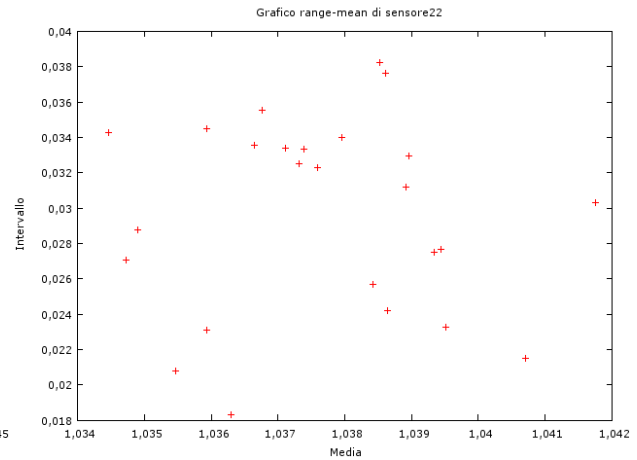


Figura 5.76 range-mean plot(osservazioni 200-800)
25 sottogruppi di numerosità 24

Stimiamo il trend parabolico presente nella serie ed otteniamo quindi la nuova variabile v1.

	Coefficiente	Errore Std.	Rapporto t	p-value
costante	1,01883	0,000658531	1547	0,0000
trend	7,99690e-05	3,03840e-06	26,32	2,40e-116
trend al quadrato	-7,71218e-08	2,93916e-09	-26,24	8,33e-116
R-quadro	0,413125	R-quadro corretto	0,411948	

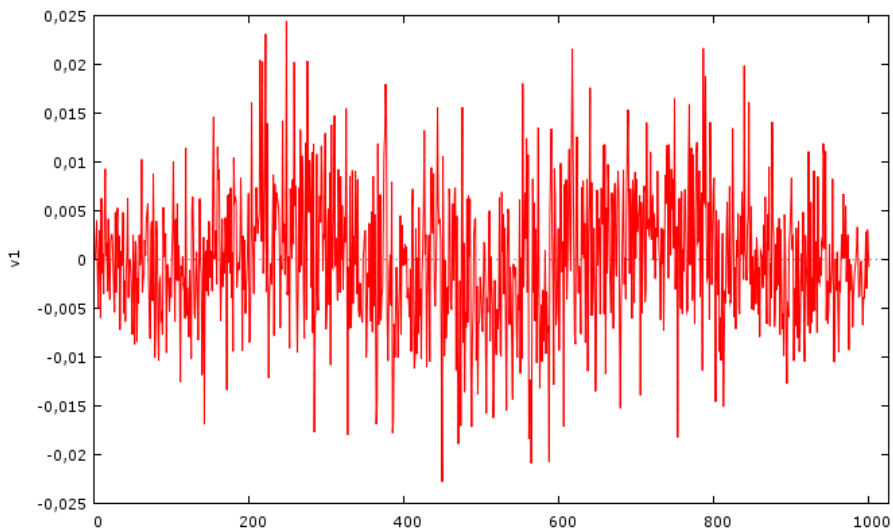


Figura 5.77 grafico della serie detrendizzata

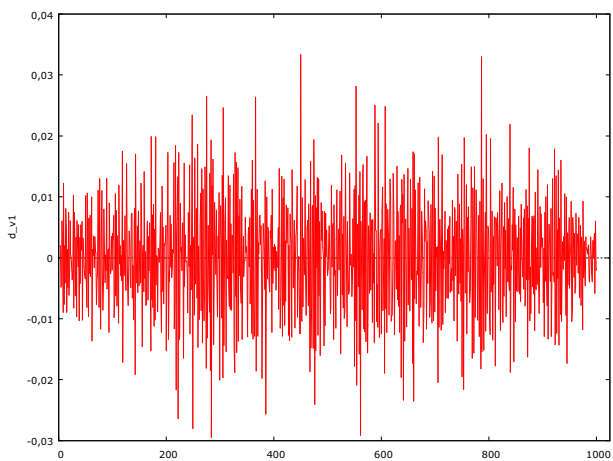


Figura 5.78 grafico delle differenze prime della serie

Sembra necessaria una differenziazione di ordine uno per rendere la serie stabile in media. La variabilità risulta più uniforme.

Per identificare il modello ARIMA opportuno è necessario stimare la funzione di correlazione globale e parziale della serie.

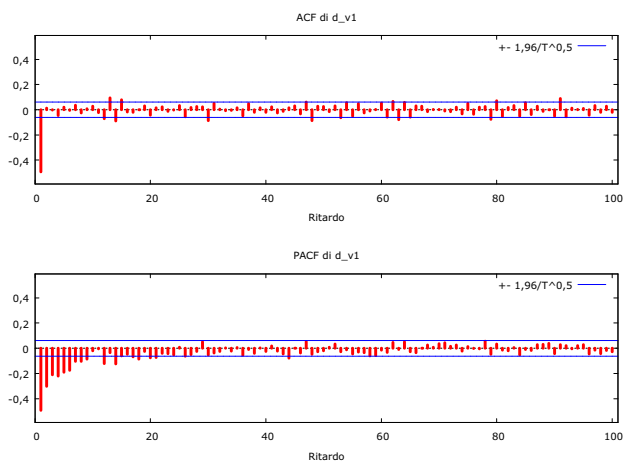


Figura 5.79 correlogramma totale e parziale delle differenze prime di v1

\hat{P}_k , dopo i primi 12 lag significativi, tende a zero.

Anche per questa serie viene scelto un modello ARIMA(2,1,3) vincolato, però, ad avere Φ_1 uguale a zero.

	Coefficiente	Errore Std.	Rapporto z	p-value
Φ_2	-0,676693	0,193478	-3,498	0,0005
θ_1	-0,940675	0,0242056	-38,86	0,0000
θ_2	0,698938	0,194195	3,599	0,0003
θ_3	-0,695168	0,176385	-3,941	8,11e-05

Media var. dipendente	-8,23e-07	SQM var. dipendente	0,009179
Media innovazioni	-0,000042	SQM innovazioni	0,006731
Log-verosimiglianza	3577,172	Criterio di Akaike	-7144,344
Criterio di Schwarz	-7119,810	Hannan- Quinn	-7135,019

	Reale	Immaginario	Modulo
Radice AR 1	0,0000	-1,2156	1,2156
Radice AR 2	0,0000	1,2156	1,2156
Radice MA 1	1,0375	0,0000	1,0375
Radice MA 2	-0,0161	-1,1774	1,1775
Radice MA 3	-0,0161	1,1774	1,1775

Le radici della parte AR sono in modulo maggiori di uno.

Procediamo quindi con l'analisi diagnostica dei residui del modello scelto.

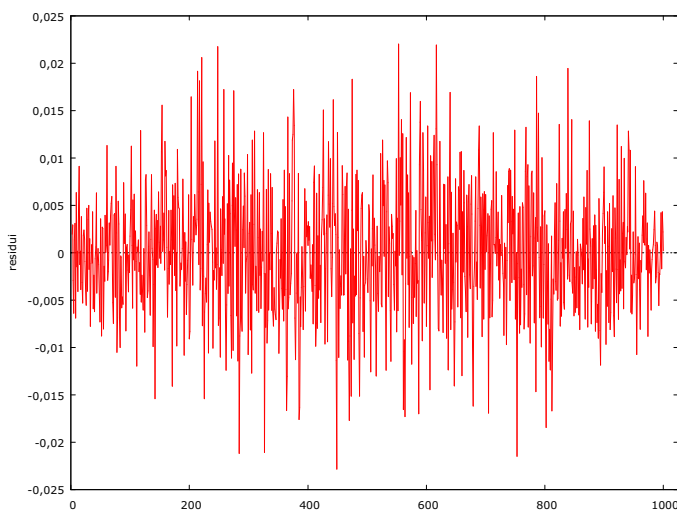


Figura 5.80 serie dei residui del modello ARIMA(213) vincolato

La serie dei residui sembra avere una variabilità leggermente in crescita.

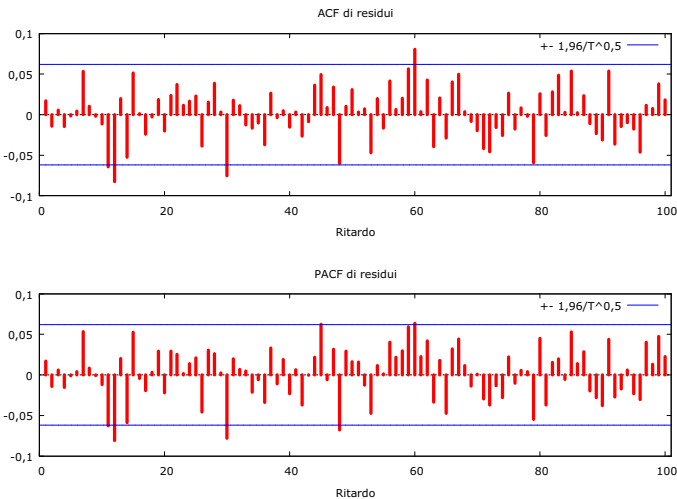


Figura 5.81 stima della funzione di autocorrelazione dei residui

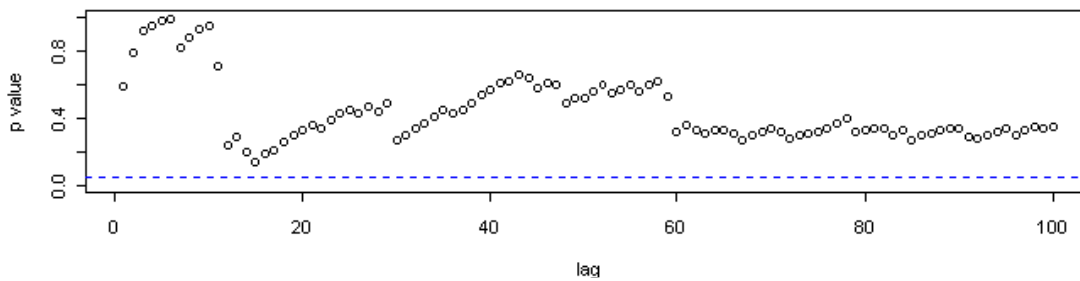


Figura 5.82 p-value della statistica test di Ljung-Box

L'analisi dei residui, per quanto riguarda l'incorrelazione, sembra soddisfacente. Per cinque ritardi rispetto ai cento considerati l'autocorrelazione globale risulta fuori dalle bande di confidenza. Ciò può essere attribuito ad un effetto del caso.

Ulteriore conferma viene data dal test di Ljung e Box sui primi 100 lag che, con un p-value di 0,3645, porta ad accettare l'ipotesi nulla (statistica test = 104,29659).

Continuiamo con la verifica della normalità delle innovazioni.

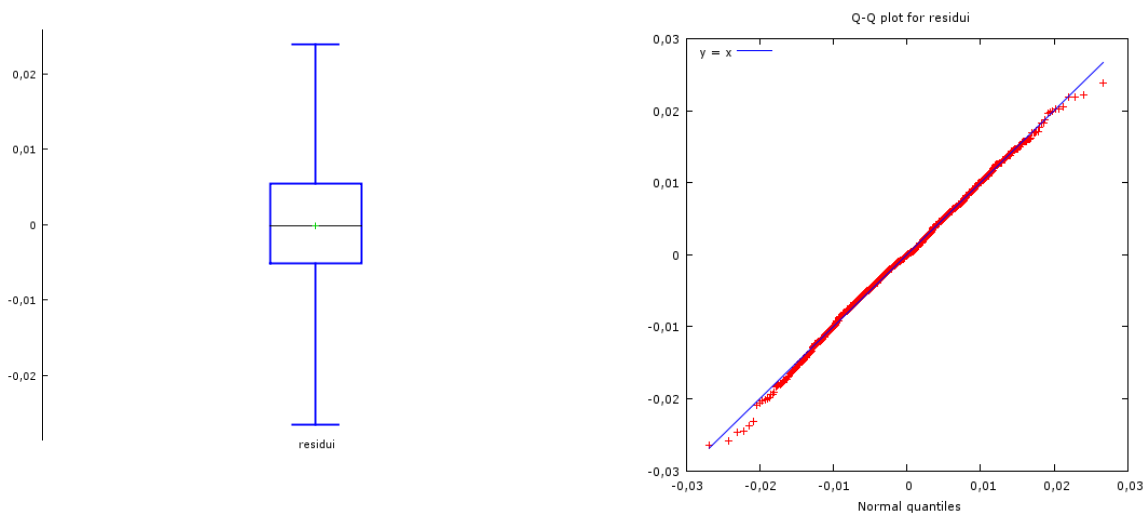


Figura 5.83 box plot e diagramma quantile-quantile dei residui

Test per la normalità effettuati:

W di Shapiro-Wilk = 0,998216, con p-value 0,306112

Test di Jarque-Bera = 2,05858, con p-value 0,357261

I residui sono distribuiti normalmente. I p-value dei due test effettuati lo confermano. È presente un leggera asimmetria destra nei dati (indice di asimmetria di Fisher = 0,093). I residui hanno una forma lievemente leptocurtica (indice di curtosi di Fisher pari a 0,102).

L'ultima verifica da fare sui residui consiste nel controllare che la varianza di essi non aumenti nel tempo.

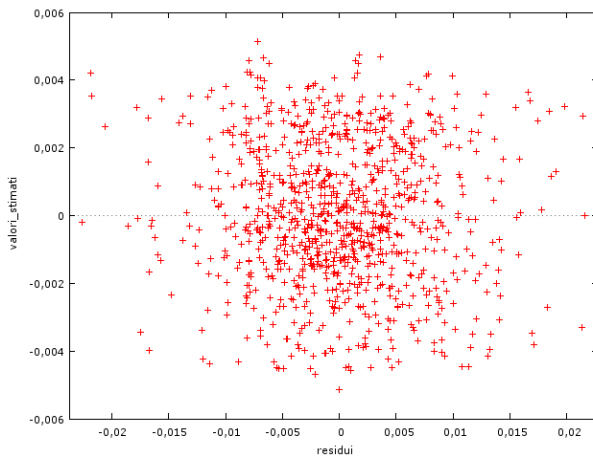


Figura 5.84 diagramma di dispersione dei residui del modello rispetto ai valori stimati dallo stesso

I residui e i valori adattati non sembrano mostrare alcuna correlazione tra loro. Il grafico si presenta come una nuvola di punti ben dispersa e con una variazione uniforme attorno all'asse delle ascisse. La varianza degli errori appare costante.

Proseguiamo, infine, con l'applicazione delle carte di controllo Shewhart e EWMA ai residui standardizzati del modello ARIMA(2,1,3) vincolato ad avere Φ_1 uguale a zero.

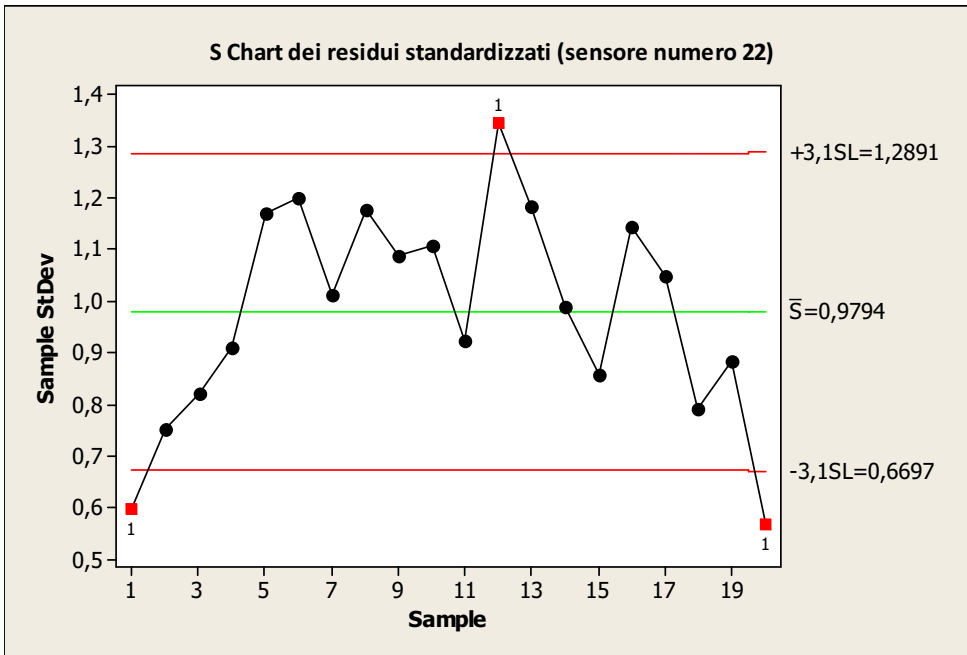


Figura 5.85

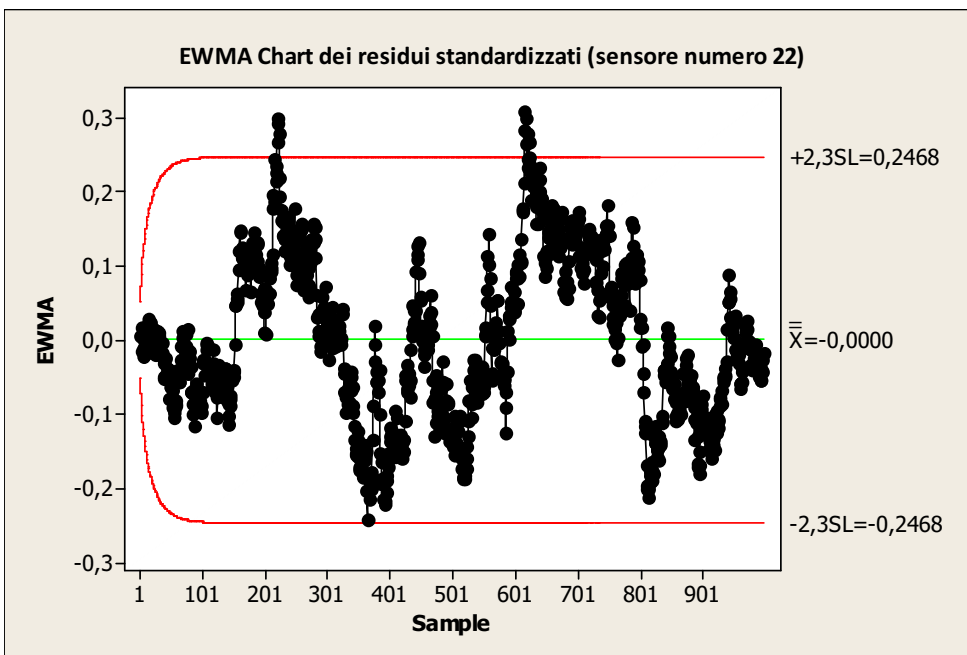


Figura 5.86

La carta S indica tre sottogruppi fuori controllo (1,12,20). Per questo motivo non è stata rappresentata la carta Shewhart per le medie. La rampa in salita (campioni 1-10) e quella in discesa (campioni 11-20) mostrano, rispettivamente, una tendenza ad aumentare e diminuire con il tempo. Questo conferma le osservazioni fatte precedentemente durante le analisi preliminari e la verifica del modello.

La carta EWMA mostra sia le osservazioni 220, 221, 222 che quelle dalla 617 alla 619 fuori controllo. Queste corrispondono rispettivamente a flussi di 3,5 e 5 m³/h. Si nota una deriva decrescente dopo il primo fuori controllo. È importante notare, però, che sono solamente 6 osservazioni (rispetto alle 1000 totali) che superano leggermente i limiti. Complessivamente il sensore 22 sembra in controllo.

Conclusioni

L'obiettivo di questa relazione è quello di individuare un metodo, statisticamente valido, per controllare che ogni sensore misuri correttamente nell'ambito di una produzione in linea di milioni di pezzi.

Si è rivelato necessario confrontare i valori di flusso letti dal sensore con quelli di uno strumento di riferimento. Il parametro utilizzato è, quindi, il rapporto. Il primo passo compiuto è quello di trovare un modello in grado di spiegare l'andamento di questo parametro durante ciascuna acquisizione.

Poiché le osservazioni sono risultate fortemente correlate, convenienti modelli di tipo ARIMA sono stati identificati e adattati. Per poter individuare il miglior modello, in grado di spiegare il processo generatore dei dati, sono state messe in luce le caratteristiche di ogni serie. In realtà molte di queste si sono rivelate comuni tra le diverse serie analizzate. In particolare, tutte presentavano una forte mancanza di stazionarietà in media. Quello che, invece, ha dato origine a dubbi è la stazionarietà in varianza. Molte sembravano avere un'ampiezza delle oscillazioni non costante per tutte le osservazioni. Grazie all'uso del range-mean plot ed all'analisi del grafico della serie è stato messo in evidenza che la variabilità delle prime e delle ultime osservazioni è minore rispetto alle altre, ma comunque apparentemente omogenea tra loro. Invece le osservazioni "centrali" di alcune tra le serie esaminate mostravano una variabilità maggiore, ma uniforme tra loro. Molte tra le serie temporali presentavano un trend parabolico più o meno marcato. Le corrispondenti serie de-trendizzate hanno apportato dei miglioramenti notevoli quali l'ottenimento di una maggior stabilità nella variabilità della serie e un comportamento soddisfacente dei corrispondenti residui.

Le tipologie di modelli ARIMA adattati sono differenti per ogni sensore. Il modello con ordine autoregressivo pari a 2 e ordine media mobile pari a 3 ha dato buoni risultati per molte serie. Spesso sono stati utilizzati modelli ARIMA vincolati ad avere alcuni parametri nulli. Questa varietà di modelli non deve stupire, infatti i sensori analizzati sono ancora dei "pezzi unici" prodotti e calibrati ciascuno da un'azienda esterna. Non sono il risultato di una produzione in serie e, quindi, le differenze possono essere notevoli. Quando la produzione diverrà più consistente potrà essere applicata una stessa tipologia di modello ad esempio per lotto di produzione. Una volta identificati gli ordini p, d, q adeguati i controlli su ogni sensore diventeranno molto più veloci.

Il passo successivo compiuto consisteva nell'analizzare i residui ottenuti dai diversi modelli per verificarne l'adeguatezza.

Infine sulla serie delle innovazioni, per le quali è stata verificata l'assunzione di normalità e identica distribuzione, sono state applicate carte Shewhart ed EWMA per riuscire ad individuare sia piccoli che grandi shift dalla media. La violazione di queste assunzioni comporterebbe uno scorretto funzionamento della carta Shewhart in termini di frequenza troppo elevata di falsi allarmi. Le carte EWMA resistono meglio agli scostamenti dalla normalità, ma non alla mancanza di indipendenza delle osservazioni. Per queste motivazioni sono stati utilizzati i residui per applicare i classici strumenti del controllo statistico della qualità.

Si sono rivelate utili anche le carte di controllo Shewhart per la variabilità, basate sulla deviazione standard. Queste mostravano spesso la presenza di un trend crescente nella rampa in salita e decrescente in quella in discesa, andamenti ciclici o comunque non del tutto casuali.

La carta EWMA ha mostrato che i sensori numero 17, 15 e 21 erano caratterizzati dal maggior numero di punti fuori controllo. Se il numero di punti fuori dai limiti risulta limitato e corrisponde al passaggio da un gradino al successivo, la causa più probabile è l'instabilità del flusso. Le altre tipologie di punti fuori controllo non sono attribuibili a cause note, ma possiamo soltanto ipotizzare siano dovute alla mancanza di stabilità delle condizioni nelle quali sono stati acquisiti i dati. Consideriamo, inoltre, che ogni acquisizione dura in media 30 minuti e non sono quindi così improbabili cambiamenti in pressione e temperatura. Spesso gli ultimi sottogruppi di osservazioni si trovavano al di sotto della linea centrale, sintomo che uno o più fattori hanno agito. Non di rado, inoltre, dopo un fuori controllo si notava una deriva decrescente dei punti. L'unico sensore risultato in totale controllo è il numero 21A. Anche il sensore 22, nonostante la presenza di qualche punto fuori dai limiti, nel complesso può essere definito in controllo.

I possibili miglioramenti che può apportare l'azienda sono relativi alle stabilità delle condizioni di prova. Mantenere temperatura, pressione e flusso costante durante ogni acquisizione è fondamentale perché permetterebbe di distinguere se le anomalie o i punti fuori dai limiti presenti sulle carte di controllo sono dovuti al funzionamento scorretto del sensore o a condizioni dell'esperimento troppo poco stabili.

Bibliografia

- Bartlett M. S. (1946), On the theoretical specification of sampling properties of autocorrelated time series, *Journal of the Royal Statistical Society*, pp. 27-41
- Bartlett M. S. (1947), The use of transformation, *Biometric Bulletin* 3, pp. 39-52.
- Box G. E. P., Jenkins G. M. (1976), *Time series Analysis, Forecasting and control*, Holden-Day
- Capizzi G. , (dispense del corso) *Metodi statistici per il controllo della qualità*
- Colonna F. (1993), *La fabbrica nera :Aspetti economici della qualità*, Edizioni Tecna
- Crowder S. W. (1989), Designing exponentially weighted moving average schemes, *Journal of Quality Technology* 21
- Di Fonzo T. Lisi F. (2005), *Serie storiche economiche*, Carrocci editore
- Douglas C. Montgomery (2006), *Controllo statistico della qualità*, McGraw-Hill
- Jarque C. M., Bera A. K. (1980), Efficient tests for normality homoskedasticity and serial independence of regression residuals, *Economics Letters*, 6, pp.255-259
- Juran J. M. (1988), *Juran's Quality Control Handbook* 4th edition, J.M. Juran e Frank M. Gryna
- Ljung G. M., Box G. E. P. (1978), On a measure of lack of fit in time series models, *Biometrika*, 65, pp.297-303
- Lucas J. M. e Saccucci M. S. (1990), Exponentially weighted moving average control schemes: Properties and enhancements, *Technometrics* 32
- Masser W.J. (1957), The Quality Manager and Quality Costs, *Industrial Quality Control*, 14, pp. 5-8
- Piccolo D. (2004), *Statistica per le decisioni*, Il Mulino
- Roberts S. W. (1959), Control chart tests based on geometric moving averages, *Technometrics* 1
- Shewhart W. A. (1931), *Economic Control of Quality of Manufactured Product*, Van Nostrand Reinhold Co., (Republished in 1981 by the *American Society for Quality Control*)
- Taguchi G. (1986), *Introduction to Quality Engineering*, Asian productivity organization UNIPUB White Plains
- Western Electric (1956), *Statistical Quality Control Handbook*, Western Electric Corporation

Siti internet consultati

- Manuali in linea del programma R (scaricabili dal sito www.r-project.org)
- Manuale del programma Gretl (http://gretl.sourceforge.net/manual_it.pdf)
- www.sitgroup.it
- www.wikipedia.org
- www.asq.org
- www.qualitiamo.org

Ringraziamenti

Grazie a tutti i miei parenti e, in particolare, ai miei genitori che mi hanno sostenuto economicamente e moralmente durante questi anni di studio. Li ringrazio per aver creduto in me.

Ringrazio mio padre per essersi seduto sopra il telecomando e per aver reso impossibile l'uso del tasto volume.

Grazie a mia madre perché, dopo essere diventata un'assidua frequentatrice del sito di statistica, è quasi sempre riuscita a sapere i risultati degli esami prima di me.

Grazie a mia sorella Elena perché, senza le sue severe ripetizioni di matematica, avrei ancora dei problemi a scomporre i polinomi di secondo grado!

Ringrazio la professoressa Giovanna Capizzi per aver risposto con pazienza a tutti i miei dubbi durante lo svolgimento di questa tesi. La ringrazio per la disponibilità e per la passione che mette nel fare con professionalità il suo lavoro.

Ringrazio il professor Tommaso Di Fonzo. I suoi consigli si sono rivelati estremamente preziosi per lo svolgimento di questa tesi.

Grazie a tutti i colleghi per aver fatto in modo che questo stage si svolgesse in un ambiente sereno e gioioso.

Ringrazio Marco Ernesto per aver allietato le mie giornate in ufficio.

Grazie a Francesco Rampazzo, il mio tutor aziendale, per avermi dato la possibilità di lavorare su un progetto davvero interessante.

Grazie a tutti i miei compagni di facoltà che hanno reso questi tre anni indimenticabili.

Li ringrazio per avermi sopportato in preda al panico prima di ogni esame, ma soprattutto per aver capito l'importanza del cd della facoltà.

In ordine alfabetico perché so che siete permalosì: Ale, Beg, Beppe, Cami, Checco, Chiere, Cri, Francesco, Giorgina, Giò, Grig, Helena, Indy, Mary, Orio, Paola, Ricky, Rigo, Vero.

Grazie a Giulia perché, ogni volta che esco con lei, mi ricorda quanto sia fantastica la complicità tra amiche.

Grazie alla mia morosetta perché è riuscita a sopportarmi a letto con lei per tutta la vacanza in Croazia. La storia purtroppo è finita male e ci siamo tradite vicendevolmente!

Grazie a tutti i miei futuri avvocati: Bea, Elsa, Ire, Davide, Giulia e Sassa.

Grazie ai miei attuali e futuri ingegneri di fiducia: Fiore, Castro, Fav e Ale.

Grazie a Michele, il mio consulente informatico, in grado di risolvere ogni problema; e non solo quelli relativi al mio pc. Grazie per tutto quello che ha fatto e fa per me. Lo ringrazio, soprattutto, per aver cercato di capire il significato di livello di significatività osservato ad improbabili ore notturne.