

DOTTORATO DI RICERCA IN
TECNOLOGIE E SISTEMI DI LAVORAZIONE
XIV CICLO

POLITECNICO
DI MILANO

UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

UNIVERSITÀ
DEGLI STUDI
DI PAVIA

UNIVERSITÀ
DEGLI STUDI
DI LECCE

IL CONTROLLO DI QUALITÀ PER PROCESSI
MANIFATTURIERI TRAMITE L'IMPIEGO DI UN
ALGORITMO NEURALE BASATO SULLA
TEORIA DELLA RISONANZA ADATTATIVA

RELATORE
PROF. QUIRICO SEMERARO

COORDINATORE
PROF. TULLIO TOLIO

CANDIDATO
ING. MASSIMO PACELLA

DOTTORATO DI RICERCA IN
TECNOLOGIE E SISTEMI DI LAVORAZIONE
XIV CICLO



POLITECNICO
DI MILANO



UNIVERSITÀ
DEGLI STUDI
DI BRESCIA



UNIVERSITÀ
DEGLI STUDI
DI PAVIA



UNIVERSITÀ
DEGLI STUDI
DI LECCE

IL CONTROLLO DI QUALITÀ PER PROCESSI
MANIFATTURIERI TRAMITE L'IMPIEGO DI UN
ALGORITMO NEURALE BASATO SULLA
TEORIA DELLA RISONANZA ADATTATIVA

RELATORE
PROF. QUIRICO SEMERARO

COORDINATORE
PROF. TULLIO TOLIO

CANDIDATO
ING. MASSIMO PACELLA

Indice

1	Cap.1. Introduzione	9
1.1	Il controllo della qualità.....	9
1.1.1	Cause non assegnabili di variabilità.....	9
1.1.2	Cause assegnabili di variabilità.....	10
1.1.3	Il controllo statistico di processo	11
1.1.4	Pattern sulle carte di controllo	12
1.1.5	La violazione delle ipotesi delle carte di controllo	13
1.2	Reti neurali per il controllo della qualità	16
1.3	Motivazioni e obiettivi della tesi	17
1.3.1	Motivazioni della tesi.....	18
1.3.2	Obiettivi della tesi.....	18
1.3.3	Struttura del lavoro di tesi.....	20
2	Cap. 2. Analisi della letteratura.....	22
2.1	Applicazioni delle reti neurali nel controllo qualità	22
2.2	Classi di applicazione delle reti neurali	23
2.2.1	Identificazione del cambiamento strutturale.....	24
2.2.2	Identificazione dei pattern sulle carte di controllo.....	26
2.2.3	Previsione delle osservazioni sul processo	27
2.3	Costruzione di una rete neurale per il controllo di qualità.....	28
2.3.1	Numero di livelli hidden	30
2.3.2	Numero di nodi della rete	30
2.3.3	Funzioni di trasferimento.....	32
2.4	Addestramento della rete neurale	34
2.4.1	Trasformazione dei dati in ingresso.....	34
2.5	Addestramento della rete neurale	37
2.5.1	Frequenza di presentazione del training set.....	38
2.5.2	Presentazione dei pattern di addestramento.....	39
2.5.3	Inclusione o esclusione di dati in controllo	39
2.5.4	Rapporto segnale/rumore.....	40
2.5.5	Selezione della finestra di dati.....	40
2.6	Prestazioni e testing	43
2.6.1	Misure speciali di valutazione delle prestazioni.....	45
2.6.2	Confronti fra i differenti lavori	46
3	Cap. 3. Basi teoriche e modelli di riferimento	49
3.1	Analisi di un processo.....	49
3.1.1	Il modello del processo	50
3.2	Modelli di pattern.....	52
3.2.1	Variazione sistematica	52
3.2.2	Ciclo.....	52

3.2.3	Shift	53
3.2.4	Mixture	53
3.2.5	Trend	54
3.3	La carta di controllo Shewhart	55
3.3.1	Costruzione della carta di controllo	55
3.3.2	Carta di controllo di Shewhart per il modello di riferimento	57
3.4	La carta di controllo CUSUM	57
3.4.1	Carta di controllo CUSUM per il modello di riferimento	59
3.5	Misure di prestazione	59
3.5.1	Esempi di applicazione delle carte di controllo	60
3.5.2	Calcolo errori Tipo I e Tipo II per la carta Shewhart	62
3.5.3	Discussione	64
3.6	Correlazione temporale dei dati	66
3.6.1	Terminologia	66
3.6.2	Modelli ARIMA(p, d, q)	67
3.6.3	Il modello AR di ordine 1	69
3.6.4	Il modello del processo AR di ordine 1	69
3.7	Carte di controllo per dati auto-correlati	71
3.7.1	Carta di controllo di Shewhart per il modello di riferimento	72
3.7.2	Calcolo errori Tipo I e Tipo II per la carta Shewhart sui residui ...	72
3.7.3	Discussione	74
3.7.4	La carta CUSUM per i residui	79
3.8	Finestra di analisi	79
4	Cap. 4. Reti neurali per il controllo di qualità	82
4.1	Overview sulle reti neurali	82
4.1.1	Struttura di una rete neurale	82
4.1.2	Algoritmi di apprendimento	84
4.1.3	Addestramento supervisionato	84
4.2	Applicazione di una rete supervisionata	85
4.2.1	Scelta della rete neurale supervisionata	85
4.2.2	L'insieme di dati di addestramento	86
4.2.3	Addestramento della rete neurale supervisionata	87
4.2.4	Risultati sperimentali	88
4.2.5	Considerazioni sull'uso delle reti supervisionate	89
4.3	Adaptive Resonance Theory	89
4.3.1	ART1	90
4.3.2	Una sintesi dell'algoritmo ART1	91
4.4	Fuzzy ART	95
4.4.1	Nomenclatura	95
4.4.2	Algoritmo Fuzzy ART	96
4.4.3	Interpretazione geometrica dell'algoritmo	99
4.4.4	Training PLP - Perfectly Learned Patterns	102
4.4.5	Testing di una rete ad addestramento PLP	104
5	Cap. 5. Il sistema di controllo qualità Fuzzy ART	107

5.1	L'algoritmo ART per il monitoraggio della qualità.....	107
5.2	Applicazione dell'algoritmo Fuzzy ART	107
5.3	I parametri della rete Fuzzy ART	109
5.3.1	Scelta dei parametri della rete neurale.....	110
5.4	Applicazione della rete per il controllo di qualità.....	111
5.5	Fasi operative della rete neurale	113
5.5.1	Training.....	113
5.5.2	Testing	113
5.6	Modello Matematico.....	114
5.6.1	Modello Statistico	115
5.6.2	Studio di casi specifici	118
5.7	Discussione	124
5.7.1	Alcune proprietà statistiche sull'errore di Tipo I.....	126
5.8	Generalizzazione del modello.....	130
5.8.1	Addestramento su target naturale e non naturale.....	131
6	Cap. 6. Risultati sperimentali: dati indipendenti	134
6.1	Analisi delle prestazioni.....	134
6.2	Comparazione delle prestazioni.....	139
6.2.1	Confronto dei risultati	140
6.3	Miglioramento delle prestazioni della rete neurale.....	142
6.3.1	Analisi delle prestazioni.....	143
6.4	Il pattern di tipo trend	148
7	Cap. 7. Risultati sperimentali: dati dipendenti	152
7.1	Analisi caso AR -0.8	152
7.1.1	Comparazione delle prestazioni caso AR -0.8	156
7.1.2	Miglioramento delle prestazioni della rete neurale.....	158
7.2	Analisi caso AR -0.2	160
7.2.1	Comparazione delle prestazioni caso AR -0.2	164
7.2.2	Miglioramento delle prestazioni della rete neurale.....	165
7.3	Analisi caso AR 0.2	168
7.3.1	Comparazione delle prestazioni caso AR 0.2	171
7.3.2	Miglioramento delle prestazioni della rete neurale.....	172
7.4	Analisi caso AR 0.8	175
7.4.1	Comparazione delle prestazioni caso AR 0.8	178
7.4.2	Miglioramento delle prestazioni della rete neurale.....	179
8	Conclusioni e sviluppi futuri	184
9	Bibliografia.....	188
A	App. A. La metodologia di sperimentazione	192
A.1	Stima delle prestazioni.....	192

A.2	Analisi dei risultati della simulazione	192
A.3	Relazioni di stima della prestazioni	193
A.3.1	Alcuni vantaggi del metodo batch mean	195
A.3.2	Confronto	195
B	App. B. Risultati sperimentali analitici	198
B.1	Descrizione	198
B.2	Dati indipendenti	199
B.2.1	Errori di Tipo I	199
B.2.2	Errori di Tipo II	200
B.2.3	Confronto carta CUSUM	201
B.2.4	Confronto carta CUSUM, pattern di tipo trend.....	201
B.3	Dati dipendenti AR 1° ordine fattore-0.8	202
B.3.1	Errori di Tipo I	202
B.3.2	Errori di Tipo II	203
B.3.3	Confronto carta CUSUM	204
B.4	Dati dipendenti AR 1° ordine fattore-0.2	204
B.4.1	Errori di Tipo I	204
B.4.2	Errori di Tipo II	205
B.4.3	Confronto carta CUSUM	206
B.5	Dati dipendenti AR 1° ordine fattore+0.2	207
B.5.1	Errori di Tipo I	207
B.5.2	Errori di Tipo II	207
B.5.3	Confronto carta CUSUM	208
B.6	Dati dipendenti AR 1° ordine fattore+0.8	209
B.6.1	Errori di Tipo I	209
B.6.2	Errori di Tipo II	210
B.6.3	Confronto carta CUSUM	210
C	App. C. Addestramento sull'uscita del processo.....	213
C.1	Lo scenario della fase di addestramento	213
C.2	Il modello di riferimento	213
C.2.1	Costruzione dell'esperimento.....	214
C.3	Risultati sperimentali.....	216
C.3.1	Errore di Tipo I.....	217
C.3.2	Errore di Tipo II	220
C.3.3	Considerazioni.....	222
C.4	Miglioramento delle prestazioni.....	222
C.5	Funzione errore di Tipo I e di Tipo II	223

Indice delle figure

Figura 1.1 – Dati normali, indipendenti e identicamente distribuiti.....	15
Figura 1.2 – Dati normali e dipendenti: auto-regressivi con fattore negativo.....	15
Figura 1.3 – Dati normali e dipendenti: auto-regressivi con fattore positivo	15
Figura 3.1 – Modello generale di simulazione del processo	51
Figura 3.2 – Modello semplificato di simulazione del processo.....	51
Figura 3.3 – Pattern variazione sistematica.....	52
Figura 3.4 – Pattern ciclo	53
Figura 3.5 – Pattern shift.....	53
Figura 3.6 – Pattern mixture	54
Figura 3.7 – Pattern trend.....	55
Figura 3.8 – Comparazione degli effetti di pattern non naturali sulle carte di controllo	61
Figura 3.9 – Errore di Tipo II per carta di controllo Shewhart.....	65
Figura 3.10 – Modello generale di simulazione del processo AR di ordine 1	70
Figura 3.11 – Modello semplificato di simulazione del processo AR di ordine 1	70
Figura 3.12 – Errore di Tipo II per carta di controllo Shewhart. AR(-0.8)	76
Figura 3.13 – Errore di Tipo II per carta di controllo Shewhart. AR(-0.2)	76
Figura 3.14 – Errore di Tipo II per carta di controllo Shewhart. AR(+0.2).....	77
Figura 3.15 – Errore di Tipo II per carta di controllo Shewhart. AR(+0.8).....	77
Figura 3.16 – Errore di Tipo II per carta di controllo Shewhart. Pattern shift.....	78
Figura 3.17 – Errore di Tipo II per carta di controllo Shewhart. Pattern Variazione Sistematica	78
Figura 3.18 – Errore di Tipo II per carta di controllo Shewhart. Pattern Ciclo	78
Figura 3.19 – Errore di Tipo II per carta di controllo Shewhart. Pattern Mixture	78
Figura 4.1 – Schematizzazione dell’algoritmo ART	92
Figura 4.2 – Rappresentazione geometrica nel caso di uno spazio bidimensionale	96
Figura 4.3 – Schematizzazione della rete neurale Fuzzy ART.....	97
Figura 4.4 – Interpretazione geometrica di un cluster della rete Fuzzy ART.....	100
Figura 4.5 – Interpretazione geometrica fase di verifica	101
Figura 5.1 – Modello del sistema di controllo qualità.....	111
Figura 5.2 – Valore atteso del MAD al variare dell’entità di disturbo.....	119
Figura 5.3 – Deviazione standard del MAD al variare dell’entità del disturbo per differenti finestre	119
Figura 5.4 – Valore atteso del MAD al variare dell’entità di disturbo per un ciclo di periodo 4	120
Figura 5.5 – Deviazione standard del MAD al variare dell’entità del disturbo per differenti finestre per un ciclo di periodo 4.	121
Figura 5.6 – Distribuzione di frequenza della misura MAD su 10000 dati raggruppati in finestre da 1, 10, 50 e 75 punti.	125
Figura 5.7 – Limiti superiori per gli errori di Tipo I e di Tipo II ottenuti dalla disuguaglianza di Čebyšev. Limite $l=3$; Window Size $M=1,2,10,50$	128
Figura 5.8 – Limiti superiori per gli errori di Tipo I e di Tipo II ottenuti dalla disuguaglianza di Čebyšev. Limite $l=6$; Window Size $M=1,2,10,50$	129
Figura 6.1 – Stimatori puntuali degli errori di Tipo I (ordinate) e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)	136
Figura 6.2 – Stimatori puntuali degli errori di Tipo II (ordinate) per shift di una unità di deviazione standard, e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione).....	138
Figura 6.3 – Stimatori puntuali degli errori di Tipo II (ordinate) per shift espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).....	148
Figura 6.4 – Stimatori puntuali degli errori di Tipo II (ordinate) per trend con tassi (slope) espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).....	150
Figura 7.1 – Distribuzione densità statistica MAD per finestre di 75 osservazioni.....	152
Figura 7.2 – Stimatori puntuali degli errori di Tipo I (ordinate) su dati AR(-0.8) e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione).....	154
Figura 7.3 – Distribuzione densità statistica MAD per finestre di 75 osservazioni.....	155
Figura 7.4 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(-0.8) per shift di una unità di deviazione standard, e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)	156
Figura 7.5 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(-0.8) per shift espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).....	160
Figura 7.6 – Distribuzione densità statistica MAD per finestre di 75 osservazioni.....	161

Figura 7.7 – Stimatori puntuali degli errori di Tipo I (ordinate) su dati AR(-0.2) e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)	162
Figura 7.8– Distribuzione densità statistica MAD per finestre di 75 osservazioni	162
Figura 7.9 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(-0.2) per shift di una unità di deviazione standard, e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione).....	164
Figura 7.10 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(-0.2) per shift espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).....	168
Figura 7.11 – Distribuzione densità statistica MAD per finestre di 75 osservazioni	168
Figura 7.12 – Stimatori puntuali degli errori di Tipo I (ordinate) su dati AR(0.2) e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)	169
Figura 7.13– Distribuzione densità statistica MAD per finestre di 75 osservazioni	170
Figura 7.14 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(0.2) per shift di una unità di deviazione standard, e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione).....	171
Figura 7.15 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(0.2) per shift espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).....	175
Figura 7.16 – Distribuzione densità statistica MAD per finestre di 75 osservazioni	175
Figura 7.17 – Stimatori puntuali degli errori di Tipo I (ordinate) su dati AR(0.8) e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)	176
Figura 7.18– Distribuzione densità statistica MAD per finestre di 75 osservazioni	177
Figura 7.19 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(0.8) per shift di una unità di deviazione standard, e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione).....	178
Figura 7.20 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(0.8) per shift espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).....	182
Figura C 1 – Schematizzazione risultati ANOVA per errore di Tipo I.....	219
Figura C 2 – Schematizzazione risultati ANOVA per errore di Tipo II	221
Figura C 3 – Schematizzazione errore di Tipo I per 15 insiemi di addestramento (caso $M=75$, $d=7500$, $\rho=0.85$).....	222
Figura C 4 – Funzioni errore di Tipo I e di Tipo II (sinistra) e derivate parziali rispetto al paramero di vigilanza ($d=100$, $M=10$).....	224
Figura C 5 – Funzioni errore di Tipo I e di Tipo II (sinistra) e derivate parziali rispetto al paramero di vigilanza ($d=1000$, $M=10$).....	225
Figura C 6 – Funzioni errore di Tipo I e di Tipo II (sinistra) e derivate parziali rispetto al paramero di vigilanza ($d=100$, $M=25$).....	226
Figura C 7 – Funzioni errore di Tipo I e di Tipo II (sinistra) e derivate parziali rispetto al paramero di vigilanza ($d=1000$, $M=25$).....	227
Figura C 8 – Funzioni errore di Tipo I e di Tipo II (sinistra) e derivate parziali rispetto al paramero di vigilanza ($d=7500$, $M=75$).....	228

Indice delle tabelle

Tabella 2.1 – Classificazione della letteratura in base alle tematiche di riferimento.....	24
Tabella 2.2 – Classificazione della letteratura in base alle classi di applicazione.....	28
Tabella 2.3 – Classificazione della letteratura in base alla struttura della rete neurale implementata.....	33
Tabella 2.4 – Classificazione della letteratura in base al tipo di algoritmo di codifica dei dati in ingresso alla rete neurale.....	37
Tabella 2.5 – Classificazione della letteratura in base al tipo e ai parametri di addestramento della rete neurale.....	43
Tabella 2.6 – Classificazione della letteratura in base ai criteri di prestazione utilizzati.....	47
Tabella 4.1 – Parametri della rete neurale MLP-BP.....	86
Tabella 4.2 – Confronto fra Shewhart-CUSUM (risultati simulazione riportati da Cheng 1995, tab. 7) e rete neurale (risultati simulazione).....	89
Tabella 5.1 – Parametri di implementazione di una rete neurale Fuzzy ART.....	109
Tabella 5.2 – Parametri di una rete Fuzzy ART ad addestramento PLP.....	111
Tabella 6.1 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale (risultati di simulazione).....	135
Tabella 6.2 – Stimatori puntuali e intervallari dell'errore di Tipo II generati dalla rete neurale (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard.....	137
Tabella 6.3 – Confronti fra le prestazioni della rete neurale e della carta CUSUM (risultati di simulazioni).....	141
Tabella 6.4 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale (risultati di simulazione).....	144
Tabella 6.5 – Stimatori puntuali e intervallari dell'errore di Tipo II generati dalla rete neurale (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard.....	145
Tabella 6.6 – Confronti fra le prestazioni della rete neurale e della carta CUSUM (risultati di simulazioni).....	146
Tabella 7.1 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale su dati AR(-0.8) (risultati di simulazione).....	153
Tabella 7.2 – Stimatori puntuali e intervallari dell'errore di Tipo II generati dalla rete neurale su dati AR(-0.8) (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard.....	155
Tabella 7.3 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(-0.8) (risultati di simulazioni).....	157
Tabella 7.4 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(-0.8) (risultati di simulazioni).....	158
Tabella 7.5 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale su dati AR(-0.2) (risultati di simulazione).....	161
Tabella 7.6 – Stimatori puntuali e intervallari dell'errore di Tipo II generati dalla rete neurale su dati AR(-0.2) (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard.....	163
Tabella 7.7 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(-0.2) (risultati di simulazioni).....	165
Tabella 7.8 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(-0.2) (risultati di simulazioni).....	166
Tabella 7.9 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale su dati AR(0.2) (risultati di simulazione).....	169
Tabella 7.10 – Stimatori puntuali e intervallari dell'errore di Tipo II generati dalla rete neurale su dati AR(0.2) (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard.....	170
Tabella 7.11 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(0.2) (risultati di simulazioni). La rete neurale Fuzzy ART è addestrata al riconoscimento della sola media naturale del processo.....	172
Tabella 7.12 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(0.2) (risultati di simulazioni).....	173
Tabella 7.13 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale su dati AR(0.8) (risultati di simulazione).....	176
Tabella 7.14 – Stimatori puntuali e intervallari dell'errore di Tipo II generati dalla rete neurale su dati AR(0.8) (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard.....	178
Tabella 7.15 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(0.8) (risultati di simulazioni).....	179
Tabella 7.16 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(0.8) (risultati di simulazioni).....	180
Tabella B 1– Errori di Tipo I 50 run 2000 punti. Osservazioni indipendenti.....	200
Tabella B 2 – Errori di Tipo II 50 run 2000 punti. Osservazioni indipendenti.....	200
Tabella B 3 – Confronto NN/SPC 50 run 2000 punti. Osservazioni indipendenti.....	201
Tabella B 4 – Confronto NN/SPC 50 run 2000 punti (pattern trend).....	202
Tabella B 5 – Errori di Tipo I 50 run 2000 punti. Osservazioni AR(-0.8).....	203
Tabella B 6 – Errori di Tipo II 50 run 2000 punti. Osservazioni AR(-0.8).....	203

Tabella B 7 – Confronto NN/SPC 50 run 2000 punti. Osservazioni AR(-0.8).....	204
Tabella B 8 – Errori di Tipo I 50 run 2000 punti. Osservazioni AR(-0.2).....	205
Tabella B 9 – Errori di Tipo II 50 run 2000 punti. Osservazioni AR(-0.2).....	206
Tabella B 10 – Confronto NN/SPC 50 run 2000 punti. Osservazioni AR(-0.2).....	206
Tabella B 11 – Errori di Tipo I 50 run 2000 punti. Osservazioni AR(+0.2).....	207
Tabella B 12 – Errori di Tipo II 50 run 2000 punti. Osservazioni AR(+0.2).....	208
Tabella B 13 – Confronto NN/SPC 50 run 2000 punti. Osservazioni AR(+0.2).....	209
Tabella B 14 – Errori di Tipo I 50 run 2000 punti. Osservazioni AR(+0.8).....	209
Tabella B 15 – Errori di Tipo II 50 run 2000 punti. Osservazioni AR(+0.8).....	210
Tabella B 16 – Confronto NN/SPC 50 run 2000 punti. Osservazioni AR(+0.8).....	211
Tabella C 1 – Parametri rete neurale.....	215
Tabella C 2 – Parametri esperimento.....	215
Tabella C 3– Fattori dell’esperimento di simulazione.....	216
Tabella C 4– Notazione esperimento.....	216
Tabella C 5– Tabella ANOVA errore di Tipo I.....	217
Tabella C 6– Errore di Tipo I e variabilità.....	218
Tabella C 7– Tabella ANOVA errore di Tipo II.....	220
Tabella C 8– Errore di Tipo I e variabilità.....	220
Tabella C 9– Errore di Tipo I e variabilità (caso M=75, d=7500).....	222
Tabella C 10– Errore di Tipo II e variabilità (caso M=75, d=7500).....	223

Capitolo 1

Introduzione

1.1 Il controllo della qualità

Un processo è un insieme di cause, condizioni, strumenti e procedure applicati ripetutamente su un insieme d'ingressi per ottenere uno o più prodotti in uscita. Gli ingressi possono includere materiali o informazioni; le uscite possono comprendere prodotti o servizi. La risposta di un processo produttivo, manifatturiero o continuo, è solitamente influenzata da una certa componente di *variabilità*. Il controllo della *qualità* è una serie di procedure e di strumenti, utili ad analizzare, controllare, e a ridurre tale variabilità con l'obiettivo di contenerla entro limiti ritenuti accettabili.

La relazione esistente tra i concetti di *variabilità* e *qualità* può essere espressa efficacemente dalla seguente relazione estratta da Montgomery (2000): *quality is inversely proportional to variability*. Questa definizione indica il contributo potenziale che metodi di analisi, quali per esempio i metodi statistici, possono dare al controllo della qualità: essendo esistente una componente di variabilità all'interno di un processo, diviene necessario adottare sistemi che permettono di trattare l'incertezza in modo esplicito ed analitico.

La più diffusa teoria del controllo della qualità è basata sull'idea che due sorgenti di variabilità possono influenzare la risposta di un processo: *cause non assegnabili* (ovvero cause comuni al naturale funzionamento del processo) e *cause assegnabili* (ovvero cause speciali o non riconducibili ad un naturale funzionamento del processo). Da questa distinzione fra due categorie separate e indipendenti di variabilità trae origine tutta la teoria del controllo di qualità per processi produttivi (manifatturieri o continui) che, in molti casi pratici, può essere estesa anche ad altri diversi tipi di processo.

1.1.1 Cause non assegnabili di variabilità

Le variazioni, della risposta di un processo, sono attribuite a cause non assegnabili quando le loro sorgenti sono considerate intrinseche al processo stesso. Questa variabilità nella risposta del processo è il risultato di una miriade di cambiamenti non rilevanti, che possono verificarsi in un comune funzionamento dello stesso. Per esempio, vibrazioni ed attrito, alterazioni della rugosità superficiale in un organo della macchina utensile, modifiche di forma per una parte o per un utensile, sono tipiche cause non assegnabili di variabilità per processi di lavorazione meccanica. In generale, le cause non assegnabili di variabilità, non possono essere rimosse, oppure non è conveniente rimuoverle perché esse richiederebbero un approfondito riesame del processo. Rimuovere le cause non assegnabili di variazione, richiede generalmente una nuova progettazione delle procedure e dei sistemi utilizzati. Ciò può comportare una serie d'investimenti consistenti, e un insieme d'azioni che possono influenzare l'intero sistema produttivo. Si tratta, in pratica, di scelte decisionali che non si limitano a modificare specifiche fasi operative ma che possono influenzare l'intera gestione del

sistema produttivo: in genere tali scelte trascendono dalle competenze degli operatori addetti alla conduzione operativa del processo.

Un processo è considerato in uno stato *naturale*, quando solo cause non assegnabili influenzano la variabilità della relativa risposta. Nell'ambito della teoria del controllo qualità, un processo in uno stato naturale è anche detto in *controllo statistico*. In particolare, si ha il controllo statistico di un processo quando la variabilità che caratterizza le uscite è, in qualche modo, *prevedibile* in funzione ad una serie di osservazioni raccolte in passato. Il concetto di controllo statistico per un processo è alla base dei metodi del controllo qualità. Un processo in controllo statistico non implica che la sua risposta è priva di variabilità, o che in ogni caso, questa variabilità è sufficientemente piccola: il controllo statistico significa semplicemente che è possibile stabilire un comportamento atteso per il processo in relazione di quello che è stato osservato nel passato.

Ad esempio, nei processi manifatturieri, il controllo statistico si traduce, molto spesso, nel riuscire a stabilire dei limiti di variabilità sulle osservazioni, ed un valore atteso circa la percentuale di uscite del processo che rientreranno tra tali limiti. La prevedibilità di un processo, in uno stato di controllo statistico, è la base per strumenti quali le carte di controllo che possono essere utilizzate per distinguere situazioni nelle quali solo cause non assegnabili influenzano il processo, da situazioni nelle quali sussistono anche delle cause assegnabili.

Nel caso in cui, le variazioni dovute a cause non assegnabili sono anche sufficientemente contenute rispetto ai limiti di variabilità specificati per la risposta del processo, e se quest'ultimo è tarato in modo che la risposta media coincida con un *target* di riferimento, allora il processo è detto *capace* di rispondere alle specifiche di produzione.

1.1.2 Cause assegnabili di variabilità

Le cause assegnabili di variabilità sono quelle che derivano da fattori estranei al processo. Per esempio, nuovi metodi, materiali grezzi, cambiamenti nelle macchine, negli strumenti di misura, o nei turni della mano d'opera, sono comuni sorgenti di variazione assegnabile nei processi manifatturieri. In genere, poiché uno degli obiettivi del controllo di qualità è di *ridurre la variabilità* nella risposta del processo, è necessario eliminare tutte le variazioni dovute a cause assegnabili. In tale situazione il processo è detto in uno stato non naturale, e poiché la variabilità non naturale si somma a quella naturale, molto spesso il processo diviene non più in controllo, in altre parole non più prevedibile, e in alcuni casi, non capace di rispondere alle specifiche. Molto spesso, il miglioramento della qualità di un processo è ottenuto identificando andamenti non naturali di variazione della risposta, e rimuovendo le relative cause assegnabili.

Alcuni strumenti di controllo della qualità mirano ad analizzare dettagliatamente la variabilità delle risposte di un processo. Il miglioramento della qualità, tuttavia, non può prescindere dalla capacità di associare stati di variazione non naturale del processo con le relative cause assegnabili. Ciò può essere fatto solamente attraverso un'approfondita conoscenza del processo come, ad esempio, quella che può avere un operatore addetto alla conduzione dello stesso. A differenza, infatti, delle cause non assegnabili di variabilità, la rimozione di quelle assegnabili riguarda specifiche modifiche delle fasi operative e quindi, l'identificazione di cause assegnabili, è tipicamente di competenza degli operatori addetti alla conduzione e alla supervisione del processo. In genere, poiché una causa assegnabile è una sorgente esterna al processo, la sua rimozione può avvenire senza rivedere l'intero processo.

È importante osservare com'è indispensabile riuscire a discriminare situazioni nelle quali solo cause naturali di variazione stanno influenzando il processo, da situazioni nelle quali esistono anche delle cause speciali di variazione. Intervenire su un processo stabile al fine di compensare variazioni dovute a cause naturali, come se esse fossero invece delle cause speciali, può non condurre a nessun tipo di miglioramento e, al contrario, può comportare perdite in termini di costi e di tempo.

1.1.3 Il controllo statistico di processo

Il controllo statistico di processo (Statistical Process Control – SPC), è una metodologia, basata su differenti tecniche di tipo statistico, il cui obiettivo è di migliorare continuamente la qualità (Stuart *et al.* 1996, Montgomery 2000). Lo scopo principale dell'utilizzo di tecniche SPC è di identificare la presenza di qualunque disturbo estraneo al processo quanto prima possibile, e di esplorare eventuali azioni correttive. Fra le tecniche SPC, le Carte di Controllo sono considerate gli strumenti più potenti per identificare cause assegnabili di variazione nel processo. Walter A. Shewhart ha proposto, nei primi anni venti dello scorso secolo, il concetto di carta di controllo che, nel tempo, si è continuato ad evolvere migliorando progressivamente le prestazioni. Ad oggi, la carta di controllo, è tra gli strumenti più diffusamente utilizzati per il monitoraggio della qualità, in particolare nei processi manifatturieri, sia per identificare variazioni non naturali, sia per fornire indicazioni riguardanti le loro cause. Attraverso una rappresentazione visuale dei dati di processo, infatti, una carta di controllo può fornire segnali riguardanti eventuali problemi derivanti da uno stato di variabilità non naturale.

Una carta di controllo è basata sull'idea che, se un processo è in uno stato di controllo statistico, allora le sue uscite possono essere previste attraverso l'uso di un opportuno modello probabilistico. Nella loro proposta originale le carte di controllo corrispondono ad un diagramma temporale delle singole uscite del processo, o di una grandezza statistica calcolata su un campione di esse, ad esempio: la media, l'intervallo di variazione o la deviazione standard campionaria. Le carte di controllo forniscono un utile strumento di rappresentazione visuale delle caratteristiche del processo e, in genere, ciò che le differenzia da un comune diagramma temporale di dati, è un insieme di metodi per identificare cambiamenti o tendenze. Tipicamente un operatore, responsabile del monitoraggio del processo può identificare una o più osservazioni riportate sulla carta di controllo (o sulle carte di controllo) tali da segnalare una condizione di variazione derivante da cause assegnabili. Ad esempio, dei limiti di controllo sono assegnati alla variabilità dei dati. I punti tracciati sul diagramma sono confrontati con tali limiti, che indicano l'ampiezza di variabilità prevista unicamente per le sole cause non assegnabili. L'ampiezza di tali limiti è tale che, continuando ad osservare punti all'interno degli stessi, è ragionevole supporre che il processo sia in uno stato di controllo statistico. Un punto al di fuori di tali limiti, segnala una potenziale situazione di fuori controllo, tuttavia, a causa della naturale variabilità dei dati, può esistere una certa probabilità che alcune osservazioni superino i limiti della carta anche quando il processo è in una condizione di controllo statistico. Nella loro proposta originale, le carte di controllo tradizionali di Shewhart, associano la presenza di cause assegnabili esclusivamente al superamento di tali limiti, rendendo il monitoraggio del processo semplicemente legato all'osservazione del valore più recente.

I limiti di una carta di controllo non devono essere confusi con limiti di specifica del processo o con il target di riferimento per la produzione. I limiti di controllo sono semplicemente una previsione della variabilità, che si può manifestare sulle osservazioni del processo, in funzione delle sole cause non assegnabili (naturali) di

variazione. Se la variabilità legata a cause non assegnabili di variazione, è relativamente ampia, allora tutti i prodotti del processo potrebbero essere all'esterno dei limiti di specifica anche se essi rientrano nei limiti di controllo. Analogamente, la presenza di cause assegnabili di variazione non implica necessariamente che le specifiche dello stesso non siano più soddisfatte dal prodotto in uscita, ma semplicemente che esiste una qualche sorgente di variazione che determina delle uscite del processo diversamente variabili da quanto è possibile attribuire alle sole cause naturali di variazione.

1.1.4 Pattern sulle carte di controllo

In generale, il tradizionale criterio di Shewhart, basato solo sul valore della più recente osservazione e il suo confronto con i limiti di controllo, può essere insufficiente per segnalare specifiche cause assegnabili di variabilità. L'utilità delle carte di controllo sta proprio nella possibilità di poter visualizzare graficamente la distribuzione temporale dei dati osservati e, conseguentemente, nell'identificare altre possibili cause di disturbo. Alcune importanti informazioni sul processo possono provenire proprio dall'osservazione dei *pattern* (ovvero degli andamenti) che i dati osservati presentano nel tempo. La presenza di un pattern non naturale, in altre parole un andamento non prevedibile in funzione delle sole cause naturali di variazione, costituisce (oltre al tradizionale superamento dei limiti di controllo) un fattore d'allarme. Un'approfondita conoscenza delle caratteristiche del processo, inoltre, fa sì che le informazioni derivate dall'identificazione di un pattern possono consentire di riconoscere più velocemente le cause che hanno determinato la situazione di fuori controllo al fine di rimuoverle. Ne consegue che il riconoscimento dei pattern è essenziale nell'utilizzo delle carte di controllo per il miglioramento della qualità di un processo.

Alcune strategie, dirette ad identificare pattern non naturali sulle carte di controllo, possono potenziare l'efficacia di tale strumento. Solitamente queste strategie hanno il vantaggio di utilizzare le informazioni della carta di controllo più efficientemente. Un esempio è rappresentato dai test di zona o regole di sequenza proposte da Western Electric (in *Statistical Quality Control Handbook* 1958). I test di zona si basano sull'idea di scomporre l'intervallo di variazione delle osservazioni del processo (ovvero l'intervallo di valori compreso fra limite inferiore e superiore di controllo) in diverse zone. La segnalazione di un eventuale allarme non dipenderà solo dal superamento dei limiti di controllo dell'ultima osservazione, ma deriverà anche dalla distribuzione fra le zone di scomposizione dell'intervallo di variazione, di un certo numero fra le più recenti osservazioni. Si tratta, in altri termini, di un metodo utile per includere nella decisione di un allarme non solo la più recente osservazione ma l'intero andamento (pattern) di una certa sequenza (finestra) di osservazioni.

Dal punto di vista statistico, le regole di sequenza sono basate sulla considerazione che una predefinita serie d'osservazioni (un run) ha una scarsa probabilità di verificarsi in una distribuzione di punti indipendenti, e identicamente distribuiti attorno ad un valore di media costante. Il problema delle regole di sequenza è che la presenza contemporanea di differenti criteri di segnalazione del fuori controllo, ha l'effetto di accrescere l'errore di Tipo I fino a valori inaccettabili. L'errore di Tipo I è la probabilità di identificare erroneamente un fuori controllo, quando non esistono delle cause assegnabili. Per esempio, supponendo di considerare s criteri multipli ed indipendenti di fuori controllo e indicato con α_i la probabilità di avere una condizione di allarme sul i^{esimo} criterio dato il sistema in controllo, si ha che la probabilità di commettere un errore di Tipo I può essere calcolata attraverso la seguente equazione (1.1).

$$P\{\text{allarme} \mid \text{sistema in controllo}\} = 1 - \prod_{i=1}^s (1 - \alpha_i) \quad (1.1)$$

Generalmente, l'identificazione e l'analisi di pattern non naturali, richiedono una notevole esperienza e capacità di discriminazione da parte dell'operatore. Conseguentemente, negli ambienti manifatturieri automatizzati, esiste la necessità di sviluppare un sistema di riconoscimento automatico dei pattern nel quale i dati di processo possono essere automaticamente raccolti ed analizzati on-line.

1.1.5 La violazione delle ipotesi delle carte di controllo

Una caratteristica importante delle carte di controllo, ed in genere tutti gli strumenti SPC, è che esse sono basate su specifici modelli probabilistici della variabilità del processo. Un modello probabilistico prevede di rappresentare le singole risposte del sistema, o una grandezza statistica caratterizzante un campione di risposte, con una distribuzione probabilistica definita. I parametri di tale distribuzione possono essere stimati, con opportuno livello di confidenza, utilizzando tecniche d'inferenza statistica sulla base di un campione di risposte utilizzato proprio per progettare il sistema di controllo. L'utilizzo delle carte di controllo richiede la verifica di una serie di ipotesi sulle quali si fonda la validità del relativo modello probabilistico. In genere, un'ipotesi comune è quella di considerare le osservazioni del processo normalmente distribuite attorno ad un valor medio costante nel tempo. Un'altra ipotesi fondamentale è quella di poter considerare le risposte, per uno stato naturale del processo, distribuite identicamente e in modo indipendente nel tempo. Entrambe le ipotesi sono frequentemente violate nella pratica, richiedendo di adottare delle modifiche impegnative sulla progettazione o sull'uso della carta di controllo stessa. La violazione delle ipotesi, infatti, può comportare notevoli degni delle prestazioni della carta di controllo che, generalmente, si manifestano in un numero elevato di falsi allarmi.

1.1.5.1 La non normalità

L'effetto della non normalità dei dati raccolti dal processo sulle prestazioni della carta di controllo è stato ampiamente documentato in letteratura (Montgomery 2000). Molti autori, infatti, hanno studiato l'effetto della non normalità, sulle prestazioni della carta di controllo, in termini di falsi allarmi segnalati per uno stato naturale di funzionamento. In genere, la soluzione proposta dagli autori è di adottare un'appropriata elaborazione dei dati osservati in modo da ricondurli ad una distribuzione *approssimativamente* normale. Ad esempio, la scelta di un opportuno campione d'osservazioni e la rappresentazione della relativa media campionaria può risultare già sufficiente per soddisfare la condizione di normalità almeno in modo approssimato grazie al *Central Limit Theorem* (Montgomery 2000), e quindi, per costruire una tradizionale carta di controllo di Shewhart sulla media del processo.

Diverse altre elaborazioni dei dati osservati, possono consentire di ottenere metodi di controllo sicuramente più robusti alle violazioni dell'ipotesi di normalità. Ad esempio, un'appropriata carta di controllo a media mobile con pesi esponenzialmente decrescenti (Exponentially Weigthed Moving Average – EWMA) si dimostra poco sensibile a violazioni della normalità della popolazione di osservazioni raccolte sul processo. Montgomery (2000), ad esempio, raccomanda l'utilizzo della carta EWMA come una procedura di controllo non parametrica ovvero priva di ipotesi riguardanti la distribuzione delle osservazioni.

1.1.5.2 La dipendenza temporale delle osservazioni

Con la crescita dell'automazione, e la presenza d'operazioni a flusso continuo, i processi sono controllati con sempre maggiore frequenza. In questi casi, le osservazioni utilizzate per il controllo di qualità, sono solitamente dipendenti (ovvero statisticamente auto-correlate) nel tempo. La dipendenza dei dati nel tempo può causare dei rilevanti degradi nelle prestazioni delle carte di controllo: in molti casi, l'adeguatezza delle carte di controllo tradizionali per identificare cause assegnabili è completamente invalidata. Le proprietà di lunghezza dei run delle carte di controllo sono notevolmente influenzate dall'auto-correlazione, in questi casi l'Average Run Length (ARL), quando il processo è in uno stato di controllo, può essere molto più breve di quanto previsto. Usando dati auto-correlati, le carte di controllo presentano, infatti, un gran numero di segnali di fuori controllo i quali risultano prodotti dalla struttura di auto-dipendenza fra le uscite del processo, piuttosto che da reali cause assegnabili presenti in esso (Montgomery 2000).

In molti casi la correlazione dei dati non può essere rimossa senza ridisegnare completamente il processo. Tale dipendenza deve essere considerata una caratteristica intrinseca, e dunque, una fra le cause non assegnabili di variabilità. Per controllare questo particolare tipo di processo non è possibile adottare il tradizionale approccio di Shewhart, ma è necessario adottare dei metodi di controllo che tengono in conto esplicitamente della componente di dipendenza temporale.

Un comune approccio, per risolvere questo problema, è di modellare i dati auto-correlati attraverso modelli di serie temporali di tipo Auto-Regressive Integrated Moving-Average (ARIMA) (Box *et al.* 1994). Tali modelli possono descrivere una classe estesa di processi manifatturieri (Box *et al.* 1994, Montgomery 2000). Se il modello di serie temporale è sufficientemente accurato, allora è possibile filtrare l'auto-correlazione temporale in modo che le osservazioni (dette residui) siano indipendenti, consentendo così l'applicazione delle tecniche di controllo SPC (Montgomery 2000). Tuttavia, le carte di controllo basate sui residui potrebbero non consentire di identificare ogni possibile causa non naturale di variazione. I residui, infatti, sono versioni filtrate dei dati iniziali del processo e, in alcuni casi, la conseguenza che un pattern non naturale può avere sui residui potrebbe non avere la stessa forma e ordine di grandezza che esso ha sui dati originali. I residui, invece, potrebbero avere dei cambiamenti che sono strettamente dipendenti dalla natura e dai parametri del modello ARIMA. In alcuni casi, lo specifico modello ARIMA può ridurre, o cancellare completamente, l'effetto che un disturbo non naturale sui dati originali ha sui residui utilizzati per costruire la carta di controllo.

1.1.5.3 La violazione dell'ipotesi di indipendenza: alcuni esempi

Ipotizzando unicamente cause naturali di variazione, il diagramma temporale delle osservazioni in uscita ad un generico processo può apparire simile a quelli riportati nelle successive figure. La Figura 1.1 riporta il caso di osservazioni non correlate: non esistendo memoria nel processo, un'osservazione non può essere influenzata da quelle realizzate nel passato. Si può pensare che il processo fluttui in modo casuale attorno ad una media. La Figura 1.2 riporta dei punti che sono negativamente correlati: un'osservazione sotto la media tende ad essere seguita da una sopra la media e viceversa. I punti raffigurati nella Figura 1.3 sono, invece, positivamente correlati: se un'osservazione è in uno dei due settori in cui l'intervallo di variabilità è diviso dalla media, allora è molto probabile che anche il punto successivo si manterrà nello stesso lato. Questo tipo di andamento, che secondo alcuni autori (Box *et al.* 1994) è quello più

diffuso nella pratica, è caratterizzato da sequenze di punti tutti al di sopra o al di sotto della media.

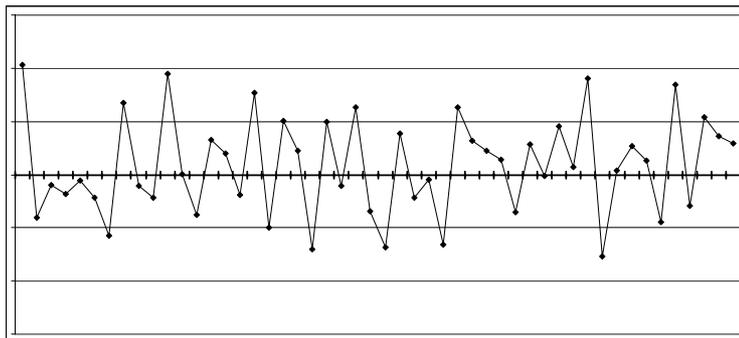


Figura 1.1 – Dati normali, indipendenti e identicamente distribuiti

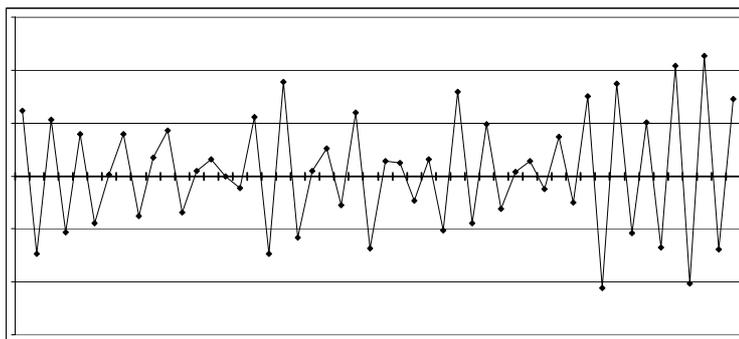


Figura 1.2 – Dati normali e dipendenti: auto-regressivi con fattore negativo

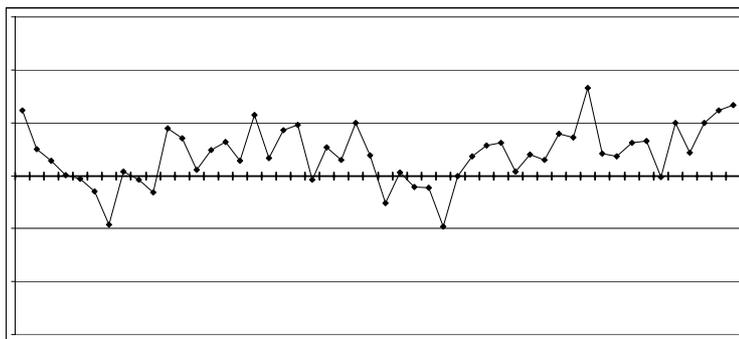


Figura 1.3 – Dati normali e dipendenti: auto-regressivi con fattore positivo

Altri tipi di processi potrebbero manifestare, oltre ad effetti di memoria fra le osservazioni, anche caratteristiche di *non stazionarietà* della media. In questi particolari processi si richiede anche un controllo *attivo*, in altre parole una continua regolazione degli ingressi in funzione dell'uscita, in modo da garantire che quest'ultima non si discosti eccessivamente dal target di riferimento. In questo caso, le carte di controllo sono utilizzate per monitorare il processo regolato al fine di identificare cause speciali di variazione. In molti casi, sia per processi stazionari, sia per quelli non stazionari, il controllo del processo richiede di adottare particolari funzioni di regolazione. Queste procedure rientrano solitamente nell'ambito de Engineering Process Control (EPC).

1.2 Reti neurali per il controllo della qualità

Da più parti si evidenzia la necessità di avere strumenti automatici di supporto al riconoscimento e alla discriminazione di condizioni non naturali di variabilità. Tali strumenti troverebbero, inoltre, una collocazione ottimale nell'ambito dei sistemi automatici di raccolta dei dati, elaborazione e analisi on-line che si diffondono sempre più negli ambienti produttivi ad elevato grado d'automazione (Swift e Mize 1995).

In anni relativamente recenti, le *reti neurali artificiali* (o similmente *reti neurali*, come saranno più semplicemente indicate nel testo), che presentano capacità d'apprendimento e d'auto-organizzazione, sono state efficacemente utilizzate nel riconoscimento di pattern, nel clustering (vale a dire nell'ordinamento in gruppi di dati), e nel controllo di qualità. Tali strumenti si possono dimostrare molto utili per le funzioni di controllo di qualità, grazie alla loro abilità di elaborare grandi moli di dati in tempo reale, e alla loro capacità di trattare il rumore e l'incertezza dei dati. Inoltre, le reti neurali non richiedono assunzioni a priori circa le distribuzioni di probabilità dei dati esaminati: ciò le rende particolarmente efficaci per la soluzione di molti problemi di controllo qualità.

La prima applicazione delle reti neurali al controllo di qualità, riportata in letteratura, è stata proposta da Pugh (1989, 1991). Una rete a addestramento supervisionato di tipo *Back-Propagation* (BP) è stata utilizzata per identificare shift della media. Le prestazioni della rete, in termini d'errore di Tipo I, erano all'incirca le stesse di una carta di controllo Shewhart con limiti 2-sigma. Hwang e Hubele (1991) hanno usato un classificatore analogo per identificare sei predefiniti pattern non naturali: tendenza, ciclo, stratificazione, variazione sistematica, mixture e shift. Hwang e Hubele (1993), in seguito, hanno descritto un progetto sperimentale per studiare l'effetto del numero dei neuroni intermedi, del numero di zone di codifica e del loro effetto cumulato sulle prestazioni di riconoscimento. Guo e Dooley (1992) hanno utilizzato una rete neurale ad addestramento BP per identificare shift sia nella media, sia nella varianza di un processo. Chang e Aw (1996) hanno adottato un approccio combinato, rete neurale BP e un criterio fuzzy, non solo per riconoscere la presenza di pattern non naturali in carte di controllo, ma anche per poterne stimare i parametri essenziali. Analogamente, Guh e Hsieh (1999) hanno proposto un sistema di controllo costituito da differenti reti neurali BP sia per riconoscere pattern non naturali, sia per poterne stimare i parametri principali. Guh e Tannock (1999), hanno ulteriormente sviluppato un approccio su rete neurale BP per il riconoscimento di pattern sovrapposti.

In quasi tutti i lavori pubblicati, l'architettura neurale utilizzata è la rete a perceptrone multiplo (Multi Layer Perceptron – MLP) con addestramento Back – Propagation (BP). Tuttavia, anche diverse altre reti neurali a addestramento supervisionato sono state proposte in letteratura per il controllo qualità. La rete neurale modulare (MNN), e la rete Radial Basis Function (RBF), sono tra queste architetture. La rete MNN è stata proposta da Cheng (1997) ed è stata confrontata ad una rete neurale ad addestramento BP. La rete neurale RBF è stata proposta da Cook and Chiu (1998) per riconoscere shift della media in parametri auto-correlati di processi manifatturieri. Leger *et al.* (1998), hanno esplorato l'utilizzo di un metodo SPC (la carta a somma cumulata CUSUM) e delle reti neurali RBF, per controllare le variazioni di una specifica caratteristica di un processo. Confronti diretti fra le prestazioni della rete BP, e della rete RBF, per compiti di riconoscimento dei pattern sulle carte di controllo, sono riportati in letteratura (West *et al.* 1999).

La motivazione comune a tutte le ricerche in questo campo, deriva dal fatto che, così come l'utilizzo dei calcolatori e l'elaborazione on-line dei dati nell'ambito

produttivo diventano sempre più diffusi, così è necessario avere anche a disposizione degli algoritmi automatici di analisi dei dati raccolti al fine di identificare condizioni di variabilità attribuibili a cause assegnabili. L'obiettivo è: affrontare il problema del controllo della qualità con maggiore efficacia e, ovviamente, con superiore efficienza.

1.3 Motivazioni e obiettivi della tesi

L'analisi della letteratura indica che le reti neurali offrono ottime capacità di supporto al controllo della qualità, e in particolare all'identificazione dei pattern sulle carte di controllo. Tuttavia, la maggior parte delle applicazioni proposte in letteratura è stata costruita sulla base di assunzioni teoriche che in numerosi casi industriali possono non essere accettabili. Due ipotesi ricorrenti sono le seguenti: 1) si suppone che l'uscita del processo, nel suo stato naturale di funzionamento, sia modellabile da osservazioni indipendenti e identicamente distribuite secondo una distribuzione normale; 2) si suppone che si abbia un ben definito insieme di pattern non naturali noti che debbano essere identificati. Entrambe le ipotesi sono sfruttate per riprodurre dati che mimano sia il vero funzionamento naturale del processo, sia ogni possibile andamento non naturale che può essere osservato all'uscita dello stesso. Quest'insieme di dati è essenziale per addestrare il sistema di monitoraggio supervisionato allo scopo di distinguere sequenze di osservazioni naturali da sequenze di dati non naturali. Tuttavia, in molti casi industriali reali tali ipotesi non possono essere confermate poiché l'andamento non naturale di un processo può non essere caratterizzato da predefiniti pattern, pertanto, i modelli matematici non sono noti a priori, oppure non possono essere formulati. Inoltre, in molti casi pratici i dati di processo sono dipendenti nel tempo e non sono normalmente distribuiti. Ciò è particolarmente vero in operazioni manifatturiere a prodotto continuo comprendenti lavorazioni di processi chimici, metallurgici o, in genere, di fonderia.

Solo recentemente sono apparsi in letteratura metodi, basati su algoritmi neurali, efficaci per monitorare processi produttivi anche nel caso in cui le uscite prodotte violano l'ipotesi di indipendenza temporale. Ad esempio, Cook e Chiu (1998) e Cook *et al.* (2001), hanno proposto una rete neurale supervisionata rispettivamente di tipo RBF e BP, per riconoscere shift della media e della varianza, relativi a processi manifatturieri correlati e modellati attraverso schemi ARIMA. Analogamente, West *et al.* (1999) hanno comparato le prestazioni di approcci SPC con quelli di due reti neurali, per controllare processi manifatturieri che esibiscono sia auto-correlazione, sia mutua-correlazione. Entrambi gli approcci, sebbene si siano dimostrati essere un sostanziale miglioramento rispetto ai metodi SPC, presentano ancora due evidenti limitazioni: 1) un metodo neurale di tipo supervisionato richiede di identificare un adeguato modello di auto-correlazione che approssima adeguatamente i dati di processo, ma questo può essere un compito non banale in molti casi; 2) l'approccio su rete neurale è un approccio supervisionato che, per motivi di apprendimento, richiede di modellare sia l'andamento naturale che l'andamento non naturale; ma ciò implica che il sistema risulterà essere tarato unicamente all'identificazione di una classe limitata di pattern non naturali, escludendo, potenzialmente la capacità di individuare altri tipi di variazioni che non sono modellabili a priori.

In definitiva si può osservare che le reti neurali proposte in letteratura, sebbene rappresentino un deciso miglioramento rispetto ai test SPC, sono dipendenti da una serie d'ipotesi (la disponibilità di una predefinita modellazione matematica sia degli

andamenti naturali, sia di quelli non naturali) che pongono ancora molte difficoltà al loro impiego pratico in ambito industriale.

1.3.1 Motivazioni della tesi

L'analisi sopra esposta riguardante le tecniche del controllo della qualità, pone all'attenzione le seguenti osservazioni.

- L'applicazione dei metodi SPC, impone la verifica di definite ipotesi in merito al comportamento del processo in uno stato naturale che sono spesso violate nella pratica. Ciò ne riduce notevolmente il campo di applicabilità.
- Dai primi anni '90 sono stati proposti da diversi ricercatori vari approcci alternativi alle tecniche SPC. Tali approcci sono basati sull'uso di algoritmi neurali di analisi delle risposte di un processo: le prestazioni presentati da questi algoritmi hanno superato di gran lunga quelle dei metodi SPC.
- Molti degli algoritmi neurali presentati in letteratura, tuttavia, hanno sfruttato i vantaggi dell'intelligenza artificiale unicamente per migliorare le prestazioni di analisi in definiti contesti applicativi caratterizzati da una serie di vincoli che in pratica possono anche non essere accettabili. A volte l'uso della rete neurale ha persino ristretto l'ambito di applicazione a casi nei quali sia possibile preventivare con esattezza quale sia l'effetto sulla variabilità del processo dalla presenza di cause assegnabili.
- Poco lavoro è stato compiuto per sfruttare le capacità di generalizzazione tipica degli algoritmi neurali con lo scopo di estendere il controllo di qualità in situazioni nelle quali sono violate le ipotesi di applicabilità dei metodi SPC. In altre parole, poco lavoro è stato compiuto nella direzione di avere metodi di controllo che sono completamente liberi da modelli probabilistici e ipotesi matematiche, sia per quanto riguarda lo stato di variabilità naturale del processo, sia ovviamente, per quanto riguarda eventuali stati non naturali di variazione.
- Appare evidente che è necessario avere metodi di controllo automatici liberi da modelli e ipotesi (*model free*) e che possano essere applicati qualunque siano le caratteristiche del processo (dati dipendenti nel tempo o indipendenti, distribuzione normale delle osservazioni, conoscenza o meno dell'effetto che cause assegnabili possono avere sulle osservazioni, etc.).

1.3.2 Obiettivi della tesi

Lo scopo del presente lavoro, è di presentare un nuovo algoritmo su rete neurale che fornisca un supporto automatico per il monitoraggio della qualità in un generico processo produttivo. In particolare si intende proporre un sistema automatico di analisi dei dati, utilizzati per il controllo di qualità, in grado di discriminare condizioni naturali di variabilità da condizioni non naturali, senza la necessità di assumere ipotesi a priori circa il funzionamento del processo, né nel suo stato di variabilità naturale, né per ogni eventuale stato di variazione non naturale.

Lo scopo principale è di progettare una procedura automatica di controllo dello stato di un processo, che non richieda ipotesi circa modelli di distribuzione probabilistica dei dati naturali, gli eventuali modelli di auto-correlazione degli stessi, o modelli matematici relativi i possibili pattern non naturali. Quest'obiettivo riveste interesse sia da un punto di vista teorico, ma anche e soprattutto da un punto di vista pratico, visto gli ovvi vantaggi che esso comporta in ambito industriale.

Poiché gli obiettivi posti riguardano lo sviluppo di un sistema per il monitoraggio della qualità privo di modello (*model free*), in altre parole privo di informazioni a priori circa quelle che possono essere le possibili distribuzioni naturali e non naturali dei dati, l'approccio neurale che deve essere adottato dovrà essere di tipo non supervisionato. In particolare, l'approccio neurale analizzato nel presente lavoro è basato sulla teoria della Risonanza Adattativa (Adaptive Resonance Theory – ART), e sarà illustrato attraverso l'uso di una carta di controllo sulla media per misure singole.

L'idea è di utilizzare un sistema, basato su un algoritmo neurale non supervisionato, rispondente alla teoria di apprendimento ART, che autonomamente identifichi la variabilità naturale di un processo, in altre parole le differenti strutture (o cluster) di output naturali con lo scopo di poter segnalare quando tali andamenti hanno subito un cambiamento. L'adozione di un algoritmo non supervisionato, come una rete ad apprendimento ART, implica inoltre un ulteriore vantaggio rispetto agli algoritmi supervisionati: quello di poter apprendere in modo cumulativo e progressivo. Quest'aspetto permette una maggiore efficacia di applicazione in tutti quei casi industriali nei quali non è possibile stabilire a priori un insieme definito da utilizzare nell'addestramento dell'algoritmo neurale. Una rete non supervisionata può imparare da un insieme limitato di addestramento, ma può ulteriormente perfezionare tale apprendimento on-line durante la fase di monitoraggio.

Un simile approccio è stato proposto in precedenza da Al-Ghanim (1997), il quale ha presentato una rete neurale ART1 (una versione binaria dell'apprendimento ART) in grado di discriminare andamenti naturali, da non naturali, ed è stato provato su un processo con uscite normali, indipendenti e identicamente distribuite. Il metodo di Al-Ghanim richiede che durante l'addestramento, la rete neurale ART operi unicamente con dati naturali del processo. La rete ART attraverso un metodo auto-organizzativo è, infatti, in grado di partizionare l'insieme di dati naturali presentati in una serie di separati sottoinsiemi (definiti cluster) di dati. La funzione eseguita da un approccio basato sul clustering, quando la rete neurale si confronta con nuovi input, è semplicemente quella di produrre una risposta indicante il cluster cui si conclude che il pattern appartenga (qualora tale cluster esista). Vale a dire che, se la rete è stabile, la funzione del sistema neurale non indicherà il tipo di pattern non naturale che si stima essere presente, ma essa fornirà unicamente un'indicazione se, l'ingresso rientra in uno dei cluster naturali formati durante l'addestramento, oppure, nel caso contrario, se un eventuale cambiamento strutturale nell'output del processo si stima sia avvenuto poiché l'ingresso non è riconosciuto appartenente ai cluster formati durante l'addestramento. Dunque, lo scopo del sistema a rete neurale proposto è duplice: 1) in modo addestramento: scoprire le strutture caratteristiche (ovvero i cluster essenziali) dei dati naturali di processo; 2) in modo operativo: fornire un'indicazione del cambiamento nell'andamento del processo se esso è avvenuto.

Il lavoro descritto nel presente testo sarà strutturato nelle seguenti fasi.

1. In primo luogo, saranno stabilite le basi teoriche per il controllo della qualità e i modelli di riferimento per il riconoscimento di pattern sulle carte di controllo. Questi modelli costituiranno la base per la definizione dei casi di studio di riferimento e le misure di prestazione per il confronto dei diversi metodi di controllo della qualità.
2. In seguito, sarà presentato e analizzato in dettaglio l'algoritmo ART, individuando i parametri di configurazione per la sua applicazione come sistema di monitoraggio della qualità, e ne sarà approfondita l'interpretazione matematico-statistica con riferimento ad un definito modello di studio.

3. Saranno quindi costruiti dei casi di studio indicativi di processi privi di memoria, in altre parole che producono dati indipendenti nel tempo, sul quale le prestazioni delle reti ART potranno essere confrontate con le prestazioni dei tradizionali metodi SPC. In particolare saranno utilizzati come riferimento, sia le prestazioni di una carta di controllo di Shewhart che carte a somma cumulata (CUSUM).
4. Sarà esteso il caso di analisi da quello di processi privi di memoria e quello di processi che presentano uscite dipendenti nel tempo. In particolare saranno analizzati processi stazionari modellati attraverso modelli AR del primo ordine.

Il risultato ultimo del lavoro è di risolvere le limitazioni degli approcci presentati in letteratura, con lo scopo di progettare uno strumento neurale automatico capace di migliorare l'efficienza del controllo di qualità, e che possa essere concretamente utilizzato in reali applicazioni industriali.

1.3.3 Struttura del lavoro di tesi

Il presente testo è stato organizzato secondo la struttura di seguito descritta.

Nel capitolo 2 è condotta una più approfondita analisi della letteratura inerente all'applicazione delle reti neurali per il controllo di qualità. Tale analisi è svolta utilizzando una serie di differenti criteri di classificazione tra i quali rientrano non solo aspetti implementativi (come per esempio l'architettura neurale adottata, l'algoritmo d'addestramento o le funzioni d'attivazione impiegate), ma caratteristiche più generali degli approcci presentati in letteratura e gli obiettivi di miglioramento delle tecniche di controllo qualità che i diversi ricercatori hanno, fino alla redazione del testo presente (2002), raggiunto.

Nel capitolo 3 sono discussi in dettaglio le basi teoriche delle tecniche di monitoraggio statistico di processi manifatturieri e i modelli del processo produttivo di riferimento. Il capitolo ha un duplice obiettivo: da un lato proporre una breve sintesi delle tecniche di controllo che rientrano nell'ambito della definizione Statistical Process Control (SPC) e che saranno utilizzate nel corso del testo per la definizione di benchmark di riferimento; dall'altro è di introdurre le modellazioni matematiche che costituiscono la base per le procedure di simulazione Monte Carlo, utilizzate per riprodurre processi di riferimento sia in condizioni naturali, sia in condizioni non naturali di variabilità.

Il capitolo 4, introduce il lettore nell'ambito delle tecniche neurali per il monitoraggio dei dati e al loro impiego nell'ambito del controllo di qualità. In particolare è presentato il funzionamento generale di una rete a addestramento supervisionata con un esempio di applicazione ad uno specifico caso di controllo. In seguito, si discuteranno in dettaglio gli algoritmi neurali ad apprendimento non supervisionato con particolare riferimento alla teoria della risonanza adattativa (Adaptive Resonance Theory – ART), e ne sarà analizzata in dettaglio un'interpretazione geometrica del funzionamento. Utilizzando la descrizione dell'algoritmo neurale ART, inoltre, è proposta una procedura d'implementazione e di addestramento di una particolare rete neurale per l'impiego al monitoraggio della qualità di un generico processo.

Nel capitolo 5, sulla base della descrizione della matematica dell'algoritmo ART svolta nella precedente sezione, sarà definita una procedura neurale di monitoraggio dello stato di un processo evidenziandone tutti i parametri di influenza. Tale procedura sarà poi analizzata in dettaglio definendo le relazioni matematiche, e statistiche, che legano i parametri dell'algoritmo neurale con le prestazioni attese nel riconoscimento di variazioni dello stato di variabilità di un processo.

Nel capitolo 6, sono riportati i risultati sperimentali (ottenuti per mezzo della simulazione Monte Carlo) dell'applicazione dell'algoritmo neurale al caso di monitoraggio di un processo manifatturiero. Il processo manifatturiero di riferimento è anch'esso riprodotto per mezzo della tecnica Monte Carlo in accordo ai modelli presentati e discussi nel capitolo 3. In particolare, i risultati presentati nel capitolo fanno riferimento al caso di un processo privo di memoria, per il quale si suppone di poter modellare le uscite naturali come una sequenza di osservazioni casuali, indipendenti e identicamente distribuite. I risultati sperimentali, da un lato, forniscono una validazione alle relazioni sulle prestazioni attese dalla rete e introdotte nel capitolo 5, dall'altro, confrontano le prestazioni ottenute con quelle realizzabili attraverso l'impiego di tecniche SPC.

Nel capitolo 7 è rivista l'applicazione della procedura neurale al caso di un processo le cui uscite non sono modellabili come realizzazioni indipendenti. In particolare, attraverso l'utilizzo della simulazione Monte Carlo, saranno analizzate le misure di prestazione per il monitoraggio di un processo manifatturiero nel caso (tipicamente tra i più comuni) in cui le uscite possano essere modellare secondo uno schema Auto-Regressivo (AR) del primo ordine. Anche in questo caso, le prestazioni dell'algoritmo neurale saranno confrontate con quelle ottenibili dai metodi SPC per il monitoraggio di sequenze di dati auto-correlati.

Dopo aver tracciato le conclusioni, e riportata la bibliografia di riferimento, sono presentati tre capitoli di appendice che completano il lavoro presentato in alcuni particolari aspetti. Il primo (appendice A) riporta la metodologia di simulazione e quella di analisi statistica, adottate nel lavoro per la valutazione delle prestazioni e la comparazione dei risultati. Nel secondo (appendice B) sono presentati con dettaglio analitico i principali risultati discussi nel testo fornendo, inoltre, un accenno alla configurazione dell'ambiente di simulazione attraverso cui sono stati ottenuti. Infine, nel terzo (appendice C) si riportano e di discutono ulteriori risultati sperimentali ottenuti utilizzando la rete neurale in un particolare scenario di addestramento.

Capitolo 2

Analisi della letteratura

In questa sezione è approfondita l'analisi della letteratura inerente all'applicazione delle reti neurali per il controllo di qualità. L'analisi è condotta attraverso una serie di differenti criteri di classificazione dei lavori presentati. Tra questi criteri rientrano non solo aspetti implementativi, come per esempio l'architettura neurale adottata, l'algoritmo d'addestramento o le funzioni d'attivazione impiegate, ma anche altri a carattere più generale come ad esempio gli obiettivi di miglioramento delle tecniche di controllo statistico di processo, cui i diversi autori mirano attraverso l'utilizzo di tecniche neurali. Con l'analisi presentata in questa sezione, s'intende ampliare ed aggiornare il review sull'uso delle tecniche neurali per il controllo di processo, proposto da Zorriassantine e Tannock (1998).

2.1 Applicazioni delle reti neurali nel controllo qualità

Il motivo principale dell'uso delle reti neurali per il controllo di qualità di processo risiede nella necessità di automatizzare con tecniche informatiche il monitoraggio dei sistemi produttivi. Inoltre, le tecniche neurali forniscono ulteriori possibilità di migliorare la capacità di controllo grazie all'accurata modellazione di processo che esse permettono di raggiungere, indipendentemente dall'uso di modelli matematici e/o statistici.

La prima applicazione di una rete neurale, per il controllo di qualità di processi manifatturieri riportata in letteratura, è apparsa negli ultimi anni '80. In particolare Pugh nel 1989 proponeva un sistema di controllo di processo basato sull'uso di una rete neurale a addestramento supervisionato, con un articolo pubblicato su *Computers and Industrial Engineering*. Un miglioramento della rete neurale, è stato poi formulato dallo stesso autore con un articolo apparso nel 1991 su *International Journal of Production Research*, nel quale inoltre, è stato presentato un confronto con metodi SPC.

Da allora, l'impiego delle reti neurali per il controllo di qualità di processi manifatturieri, è proseguito, e si è sviluppato, per tutto il decennio successivo fino alla data di redazione del presente testo (2002). L'analisi della letteratura che è condotta in questa sezione, abbraccia tale periodo e descrive tutta una serie di differenti approcci che variano in relazione alle architetture utilizzate, agli algoritmi implementati, alle ipotesi formulate etc.

Ad esclusione degli articoli presentati a varie conferenze, le principali riviste scientifiche sulle quali sono stati proposti i lavori che rientrano nell'abito dell'analisi condotta in questa sezione comprendono:

- *Computers and Industrial Engineering*, su cui, come già detto, è stato pubblicato il primo lavoro proposto in letteratura.
- *Journal of Intelligent Manufacturing*, nel quale è stato proposto nel 1998 il primo lavoro di review della applicazione delle reti neurali per il controllo di qualità di processi manifatturieri (Zorriassantine e Tannock 1998)
- *IIE Transactions*, nel quale sono stati proposti alcuni studi di progettazione delle reti neurali come, ad esempio, la prima applicazione delle reti neurali per il controllo di processi che presentano correlazione temporale (Cook e Chiu 1998).

- *International Journal of Production Research*, che riporta il maggior numero di applicazioni e, indubbiamente, quelle a maggiore approfondimento sperimentale. Altre riviste sulle quali sono apparsi sporadicamente dei lavori inerenti all'uso delle reti neurali per il controllo di qualità per processi manifatturieri sono le seguenti.
- *Artificial Intelligence in Engineering*.
- *Decision Support Systems*.
- *International Journal of Machine Tools & Manufacture*.
- *Mathematical Computing Modelling*.
- *Omega, International Journal of Management Science*.

2.2 Classi di applicazione delle reti neurali

Fondamentalmente si possono distinguere tre principali classi per l'applicazione delle reti neurali al controllo di qualità nei processi manifatturieri.

1. Identificazione del cambiamento strutturale di un processo attraverso una simulazione di una carta di controllo. Questa categoria potrebbe essere ulteriormente suddivisa in relazione a due aspetti non esclusivi.
 - Dettaglio di informazione richiesto alla rete neurale sullo stato del processo. Si possono distinguere due sotto-obiettivi
 - Semplice identificazione dello stato del processo (modificato o non modificato).
 - Valutazione e stima dell'entità della modifica che è stata identificata nel processo.
 - Grandezze monitorate dalla rete neurale. Come per una carta di controllo anche per le reti neurali sono state proposte architetture per:
 - Monitorare esclusivamente la media del processo.
 - Monitorare esclusivamente la varianza del processo.
 - Monitorare sia la media, sia la varianza del processo.
2. Identificazione di un preciso pattern caratterizzante un segnale speciale di disturbo su i dati del processo. A sua volta questa categoria potrebbe essere suddivisa in relazione al dettaglio di informazione che è fornito dal sistema di controllo su base neurale.
 - Semplice identificazione di uno o più pattern (eventualmente sovrapposti).
 - Identificazione e stima dei parametri dei pattern identificati.
3. Previsione degli andamenti futuri delle osservazioni del processo.

Reti neurali supervisionate per il controllo di qualità
Dati indipendenti nel tempo
<i>Controllo della media</i>
<i>Identificazione pattern non naturale</i>
CHENG, C. S., 1995. CHENG, C. S., 1997. CHENG, C. S. e CHENG, S. S., 2001. GUO, Y. e DOOLEY, K. J., 1992. GUH, R. S., ZORRIASSATINE, F. TANNOCK, J. D. T., e O'BRIEN C., 1999. HWARNG, H. B. e HUBELE, N. F., 1993. LEGER R. P., GARLAND, W. J. e POEHLMAN, W. F. S., 1998. PERRY, M. B., SPOERRE, J. K. e VELASCO, T., 2001. PUGH, G. A., 1989.

PUGH, G. A., 1991. REDDY, D. C., GHOSH, K. e VARDHAN, V.A., 1998. PHAM, D. T., e OZTEMEL, E., 1994. HWRANG, H. B., 1995a. HWRANG, H. B., 1995b.
<i>Identificazione di più pattern sovrapposti</i>
GUH, R. S. e TANNOCK, J. D. T., 1999a. GUH, R.-S. e TANNOCK, J. D. T., 1999b.
<i>Identificazione e stima dei parametri del pattern non naturale</i>
CHANG, S. I. e AW, C. A., 1996. GUH, R. S. e HSIEH, Y. C., 1999. WANG, T.-Y. e CHEN, L.-H., 2002.
<i>Controllo della media e della variabilità</i>
CHANG, S. I. e HO, E. S., 1999. SMITH, A. E., 1994. DEDEKAYOĞULLARI, I. e BURNAK, N., 1999.
Dati temporalmente correlati
COOK, D. F. e CHIU C. C., 1998. COOK, D. F., ZOBEL, C. W. e NOTTINGHAM, Q. J., 2001. WEST, D. A., MANGIAMELI, P. M. e CHEN, S. K., 1999.
Progettazione della rete neurale per il controllo di qualità
HWRANG, H. B. e HUBELE, N. F., 1993. PHAM, D.T. e SAGIROGLU, S., 2001.
Reti neurali non supervisionate per il controllo di qualità
AL-GHANIM, A., 1997. HWRANG, H. B. AND CHONG, C. W., 1995.
Euristiche per il controllo della qualità
AL-GHANIM, A. e KAMAT, S. J., 1995. AL-GHANIM, A. e LUDEMAN L. C., 1997. CHENG, C.-S. e HUBELE, N. F., 1996. PHAM, D. T. e WANI, M. A., 1997.

Tabella 2.1 – Classificazione della letteratura in base alle tematiche di riferimento

2.2.1 Identificazione del cambiamento strutturale

Nell’ambito di questa categoria rientra ad esempio il primo lavoro presentato in letteratura da Pugh (1989, 1991). Si tratta di un sistema basato sull’uso di una rete neurale per la simulazione di una carta di controllo sulla media. Pugh ha dimostrato che l’impiego della rete neurale permette di ottenere prestazioni simili in termini di errore di Tipo I a quello di una carta di controllo, ma contemporaneamente, di ottenere prestazioni migliori in termini di errore di Tipo II se confrontati con quelli di una carta Shewhart a limiti 2σ . Essenzialmente, per ogni possibile ingresso, l’output prodotto dalla rete di Pugh indica se la media del processo è rimasta inalterata oppure se il processo manifesta un aumento o una diminuzione della media.

Analogamente Smith (1994), evidenziando che le reti neurali sono in grado di manipolare più input e output contemporaneamente, propone un sistema per modellare simultaneamente sia una carta di controllo sulla media, sia una carta di controllo sulla varianza. Il singolo valore di uscita della rete neurale è interpretato come un’indicazione che nessun cambiamento della media o della varianza del processo è stato identificato, oppure che è stata identificata un’alterazione della media o della varianza. Comparando le prestazioni della rete neurale con quelle analoghe di una carta di controllo a limiti 3σ , Smith ha verificato nel proprio lavoro che la rete neurale presenta delle prestazioni comparabili a quelle di una carta di controllo di Shewhart nell’identificazione di grandi

shift della media o della varianza, e che inoltre, essa presenta delle prestazioni migliori nell'identificazione di shift di piccola ampiezza. Il lavoro proposto da Smith rafforza ulteriormente i risultati presentati da Pugh, e soprattutto dimostra l'applicabilità delle reti neurali per il monitoraggio di processo. Altri successivi lavori come ad esempio quelli di Guo e Dooley (1992), Cheng (1995), Chang e Aw (1996) appartengono a questo tipo di utilizzo delle reti neurali per il monitoraggio di processo.

In questa categoria di sistemi di monitoraggio, solo l'approccio di Chang e Aw (1996) mira a identificare anche l'ampiezza dello shift attraverso un approccio basato sulla teoria dei Fuzzy Set. Gli altri approcci, al contrario, mirano esclusivamente a identificare shift nella media o nella varianza.

Tutti gli articoli citati in precedenza, basando l'identificazione di cambiamenti strutturali sull'impiego di tecniche supervisionate, richiedono un addestramento che utilizzi sia andamenti naturali sia non naturali. Come già evidenziato, quest'ultimi sono stati sempre caratterizzati in letteratura come shift (ovvero aumenti o diminuzioni costanti) della media, della varianza, o di entrambi. La rete dunque è costruita per riconoscere cambiamenti strutturali del processo solo se questi possono essere rappresentati da aumenti o diminuzioni della media o della varianza, ma rimane del tutto non preventivabile il comportamento che essa avrà nel momento in cui il processo presenta un cambiamento della struttura che non sia semplicemente uno shift. In un recente lavoro Perry *et al.* (2001) evidenziano tale limite indica la necessità di addestrare la rete a riconoscere come non naturali pattern che rispondono alle regole run rules Western Electric. Si potrebbe affermare che, mentre i lavori citati in precedenza mirano ad utilizzare reti neurali per simulare il comportamento di una carta tradizionale di Shewhart, il lavoro di Perry *et al.* (2001) è inteso a migliorarne le prestazioni attraverso un algoritmo neurale che simuli le regole di controllo Western Electric.

Nell'ambito della categoria di applicazioni delle reti neurali per l'identificazione di un cambiamento strutturale dello stato del processo rientra anche il lavoro di Al-Ghanim (1997), che tuttavia si differenzia dai precedenti per l'impiego di un algoritmo neurale non supervisionato. Nel lavoro proposto da Al-Ghanim, l'impiego di un'architettura non supervisionata consente di limitare l'addestramento su un'unica categoria d'ingressi, in particolare quella naturale. Una simile applicazione mira a addestrare la rete a riconoscere la struttura naturale di variazione del processo, mentre, qualunque altro dato che non possa essere associato a tale struttura naturale, sarebbe indicato dalla rete come un segnale di un cambiamento nello stato del processo.

2.2.1.1 Dati indipendenti e dati correlati nel tempo

Un interessante impiego delle reti neurali che è emerso in tempi relativamente recenti riguarda il monitoraggio di sistemi produttivi che presentano dati di uscita correlati nel tempo. Si tratta di un'applicazione interessante tenuto conto delle ricadute industriali che essa può avere in molti processi produttivi reali, ad esempio quelli a produzione continua. La prima applicazione di reti neurali per processi che presentano correlazione temporale è stata proposta da Cook e Chiu (1998). Si trattava di una applicazione per il monitoraggio della media di un processo correlato per il quale si ipotizzava noto un modello della serie rientrante nella categoria ARIMA di Box *et al.* (1994). Analogamente West *et al.* (1999) hanno proposto un sistema neurale per il monitoraggio di processi multi-variati, caratterizzati sia da auto-correlazione (secondo un modello AR del primo ordine) che da mutua correlazione. In seguito Cook *et al.* (2001) proponevano un'applicazione per il monitoraggio della varianza di dati correlati nel tempo.

Si ritiene che questo campo di applicazione delle reti neurali sia quello nel quale si potrà assistere ad ulteriori ed interessanti sviluppi nei prossimi anni. In particolare, le proprietà delle reti neurali di modellare autonomamente sistemi complessi potranno essere efficacemente sfruttate per monitorare processi per i quali le tecniche SPC presentano diversi limiti. Particolare attenzione dovrà poi essere riposta nell'uso delle reti neurali per il controllo di serie di dati correlati nel tempo indipendentemente dall'uso di modelli di correlazione (ovvero applicazioni *model free*).

2.2.2 Identificazione dei pattern sulle carte di controllo

Il principio utilizzato in questo caso è simile a quello della categoria precedente, nel senso che la rete neurale è addestrata a riconoscere condizioni naturali di variazione da altre che non possono essere tali. In questo caso però, l'insieme di condizioni non naturali è costituito da un gruppo più ampio di possibili pattern che può manifestarsi sulla carta di controllo. Pertanto, la rete neurale non è più addestrata a simulare il funzionamento di una carta di controllo, ma essa è impiegata a riconoscere predeterminati pattern di variazione in serie temporali di dati. Tali pattern devono essere rappresentativi degli effetti che cambiamenti strutturali nello stato del processo possono avere sui dati di output.

Hwarng e Hubele (1992) per esempio utilizzano otto differenti pattern: trend, stratificazione, cicli, variazioni sistematiche, mixture, e sudden shift, sulla base dell'ipotesi che questi tipi di pattern fossero quelli di maggiore interesse da identificare sulle carte di controllo. Hwarng (1995a) ha suggerito che l'impiego delle reti neurali per l'identificazione di predeterminati pattern sui dati di un processo, può essere implementato utilizzando sia un sistema d'identificazione speciale (ovvero limitato su un particolare tipo di pattern), sia un classificatore generale (ovvero tale da identificare una famiglia estesa di possibili pattern). Hwarng (1995b), ha indicato che un sistema di riconoscimento speciale potesse avere prestazioni migliori rispetto ad uno di tipo generale, e che dovesse essere quello da utilizzare nell'ipotesi di conoscere il tipo di disturbo che si presenta più frequentemente in un determinato processo. Tuttavia, un unico sistema speciale di riconoscimento si potrebbe rilevare inadeguato siccome esso non sarebbe in grado di identificare altri tipi di pattern sulle carte di controllo. Pertanto, altri autori, propongono di utilizzare sistemi generali di riconoscimento che presentano maggiori garanzie di identificazione su un più ampio spettro di possibili condizioni di variabilità non naturale. I lavori che rientrano in questa categoria sono diversi partendo da quello di Hwarng del 1991. Per esempio, Cheng (1997), Reddy *et al.* (1998), Guh *et al.* (1999), Guh e Tannock (1999a, 1999b), Guh e Hsieh (1999), Perry *et al.* (2001).

Alcuni di questi lavori si differenziano per l'informazione fornita dalla rete neurale. Come già detto, questa si può limitare a fornire unicamente indicazione che un tipo particolare di pattern è stato rilevato, oppure può indicare informazioni relative alla natura del tipo di pattern riconosciuto (per esempio una stima dell'ampiezza). Nella prima categoria rientrano i lavori di Hwarng e Hubele (1992), Cheng (1997), Reddy *et al.* (1999) Perry *et al.* (2001). Nella seconda quelli di Guh *et al.* (1999), Guh e Tannock (1999a, 1999b), Guh e Hsieh (1999). In particolare i diversi lavori di Guh mirano anche ad identificare pattern sovrapposti oltre che alle caratteristiche principali che li contraddistinguono.

Come già indicato in letteratura, un appropriato sistema di riconoscimento dei pattern dovrebbe soddisfare i seguenti requisiti.

- Riconoscere una classe da un insieme di classi di pattern per le quali la rete neurale è stata addestrata.
- Identificare le classi di pattern non note come classi outlier.

- Identificare classi ambigue e segnalarle.

Le metodologie che sono state trovate nella letteratura, tuttavia, non riescono a soddisfare appieno gli ultimi due requisiti. In particolare l'impiego di una rete neurale ad addestramento supervisionato, presenta un comportamento prevedibile solo se in fase operativa si presentano pattern simili a quelli impiegati in addestramento. Ciò implica che a priori non può essere prevedibile il comportamento della rete se confrontata con pattern per i quali non ha avuto nessun tipo di addestramento. Solo alcuni lavori basati sull'uso di reti neurali non supervisionate (Hwarng e Chong 1995 hanno proposto l'utilizzo della rete ART) possono consentire di identificare classi non note o eventuali condizioni ambigue di riconoscimento. Ciò deriva dalla diversa natura di funzionamento dell'algoritmo essendo esso basato su una funzione di memorizzazione e di richiamo di quanto appreso in fase di addestramento. In tal caso, pattern non inclusi nell'addestramento possono essere immediatamente identificati della rete e classificati di conseguenza.

2.2.3 Previsione delle osservazioni sul processo

Smith (1994) ha proposto l'utilizzo delle reti neurali per preventivare i valori futuri osservati sul processo. Questa capacità si basa essenzialmente sull'abilità della rete neurale di poter preventivare valori futuri sulla base di una serie di dati storici. Quest'approccio, tuttavia, ha avuto poco seguito in letteratura (Kang e Park 2000). Ciò è legato principalmente al comune approccio nel controllo di processo che prevede un reset del processo, a seguito di una ricerca e rimozione delle cause assegnabili, allorché è segnalato un allarme da parte del sistema di controllo.

Le considerazioni sopra esposte sono, infine, riepilogate nella seguente Tabella 2.2 che riporta una classificazione dei principali lavori presentati in letteratura sulla base della classe di applicazione.

<i>Autore</i>	<i>Classe</i>	<i>Dettaglio</i>	<i>Grandezze Monitorate</i>	<i>Note</i>
Pugh (1989, 1991)	Cambio Struttura	Identificazione	Media	
Guo e Dooley (1992)	Cambio Struttura	Identificazione	Media + Varianza	
Hwarng e Hubele (1993a, 1993b)	Ric. Pattern	Identificazione	8 pattern	
Smith (1994)	Cambio Struttura	Identificazione	Media + Varianza	
Pham e Oztemel (1994)	Ric. Pattern	Identificazione	6 pattern	
Cheng (1995)	Cambio Struttura	Identificazione	Media	
Hwarng e Chong (1995)	Ric. Pattern	Identificazione	9 pattern	
Chang e Aw (1996)	Cambio Struttura	Stima Valori	Media	Approccio Fuzzy
Al-Ghanim (1997)	Cambio Struttura	Identificazione	Media	Approccio non supervisionato
Cheng (1997)	Ric. Pattern	Identificazione	7 pattern	
Cook e Chiu (1998)	Cambio Struttura	Identificazione	Media	Dati Correlati
Reddy <i>et al.</i> (1998)	Ric. Pattern	Identificazione	5 pattern	
Leger <i>et al.</i> (1998)	Cambio Struttura	Identificazione	Media	Sistema Integrato con CUSUM
Chang e Ho (1999)	Cambio Struttura	Stima Valori	Varianza	
Dedekayogullari e Burnak (1999)	Cambio Struttura	Identificazione	Media + Varianza	
Guh <i>et al.</i> (1999)	Ric. Pattern	Identificazione	6 pattern	
Guh e Tannock (1999a,	Ric. Pattern	Stima Valori	6 pattern	Pattern Sovrapposti

1999b)				
Guh e Hsieh (1999)	Ric. Pattern	Stima Valori	6 pattern	
West <i>et al.</i> (1999)	Cambio Struttura	Identificazione	Media	Proc. Multivariato
Kang e Park (2000)	Cambio Struttura	Identificazione	Media	Proc. Multivariato
Cheng e Cheng (2001)	Cambio Struttura	Identificazione	Media	Distribuzione Esponenziale
Cook <i>et al.</i> (2001)	Cambio Struttura	Identificazione	Varianza	Dati Correlati
Perry <i>et al.</i> (2001)	Ric. Pattern	Identificazione	5 pattern	Simulazione regole Run Rules.
Pham e Sagiroglu (2001)	Ric. Pattern	Identificazione	6 pattern	
Wang e Cheng (2002)	Cambio Struttura	Stima Valori	Media	Proc. Multivariato

Tabella 2.2 – Classificazione della letteratura in base alle classi di applicazione

2.3 Costruzione di una rete neurale per il controllo di qualità

La scelta di un'architettura neurale adatta al controllo di qualità di un processo può, in molti casi, non essere un compito facile perché diversi sistemi neurali potrebbero soddisfare una determinata applicazione. Hwarng e Hubele (1993a 1993b) e Cheng (1997) forniscono nei loro lavori delle utili idee di partenza, ma ciononostante, non esistono in letteratura dei metodi sistematici per determinare la migliore topologia e la più vantaggiosa struttura neurale per un determinato approccio.

Architetture neurali comunemente utilizzate sono: Multi-Layer Perceptron (MLP), Radial Basis Function (RBF), Learning Vector Quantization (LVQ), Adaptive Resonance Theory (ART) e Kohonen Self-Organizing Maps (SOM). Altre reti meno comuni includono le reti modulari Modular Neural Network (MNN), nonché le reti ricorrenti Recurrent Neural Network (RNN). Assieme alle diverse topologie di rete neurale, esistono diversi algoritmi di apprendimento. I fattori che caratterizzano gli algoritmi di addestramento delle reti neurali sono: la velocità di convergenza, l'adattabilità ad addestramenti on-line o off-line, l'adattabilità a rappresentare funzioni di tipo lineare o non lineare fra input e output.

Uno studio sugli algoritmi di apprendimento è stato recentemente proposto da Pham e Segiroglu (2001), comprendendo alcuni tra i più comuni per la rete MLP quali: il ben noto Back Propagation (BP), Quick-Pro (QP), Delta-Bar-Delta (DBD), Extended Delta-Bar-Delta (EDBD). I risultati pubblicati hanno confermato che l'algoritmo BP è quello più adatto, per l'addestramento di una rete neurale al riconoscimento dei pattern sulle carte di controllo qualora si utilizzi una rete MLP. In effetti, le reti MLP, addestrate attraverso l'algoritmo d'apprendimento BP, rappresentano fin dalla prima pubblicazione in letteratura (1989) la scelta più comune per la costruzione di sistemi di controllo della qualità (sia sistemi d'identificazione cambiamento di struttura, sia sistemi di riconoscimento pattern). Una possibile spiegazione di tale caratteristica è legata al fatto che le reti MLP a addestramento BP sono state applicate con successo in diversi problemi di riconoscimento pattern in svariati altri settori. Tuttavia, queste reti soffrono di diverse limitazioni quali la lentezza dell'addestramento (ovvero la convergenza richiede un numero elevato d'iterazioni ed un insieme di training sufficientemente ampio), e l'incapacità di poter eseguire un addestramento di tipo on-line. Hwarng e Chong (1995) hanno suggerito nel loro lavoro l'utilizzo della rete ART. Gli autori, infatti, evidenziano che tale tipo di rete da un lato richiede un insieme di addestramento decisamente ridotto se confrontato ad un'equivalente rete BP, dall'altro presenta prestazioni equivalenti, o migliori in alcuni casi. Pham e Oztemel (1994) giungono ad una conclusione simile, indicando che la loro rete modificata LVQ

presenta migliori capacità di riconoscimento rispetto ad una rete BP. Nel loro lavoro Chang e Ho (1999), evidenziano nuovamente il limite della rete BP derivante dalla necessità di disporre di un adeguato insieme di addestramento. Tale limite si rende particolarmente evidente nel caso in cui, dovendo utilizzare dati prodotti da un sistema reale, non si dispone di un numero di esempi di addestramento sufficiente per la costruzione di una rete MLP BP. La soluzione proposta dagli autori consiste nell'utilizzare tecniche di *bootstrapping*, che permettono, partendo da un insieme limitato di dati reali osservati dal processo, di ottenere un insieme di addestramento adeguatamente ampio per il training della rete MLP BP.

Recentemente sono state proposte in letteratura sistemi neurali per il controllo di processo basati sull'utilizzo di reti Radial Basis Function (RBF). Ad esempio Cook e Chiu (1998) utilizzano tale architettura per il controllo di dati di processo che presentano una correlazione temporale. Analogamente West *et al.* (1999) propongono l'utilizzo di reti RBF per il monitoraggio di processi multivariati che presentano correlazioni temporali. Nel loro lavoro, inoltre, è fornita un'utile comparazione fra le prestazioni di una rete MLP BP con una RBF nel monitoraggio di un processo. Leger *et al.* (1998) propongono uno schema combinato carta CUSUM e rete neurale per il monitoraggio di processo. Da una comparazione delle prestazioni di una rete MLP BP, con una RBF, gli autori concludono che quest'ultima presenta i maggiori vantaggi per la specifica applicazione descritta nel loro lavoro.

Per la soluzione di determinati problemi, in alcuni casi, potrebbe essere più utile incorporare un numero di differenti tecniche piuttosto che una sola. Per esempio, Cheng (1997) ha valutato combinazioni di differenti moduli MLP posti in parallelo per realizzare un sistema ibrido di riconoscimento pattern sulle carte di controllo. Nel lavoro presentato da Cheng (1997) è stato inoltre verificato che tale combinazione fornisce dei risultati migliori rispetto ad un semplice sistema MLP. Ciò è stato attribuito alla capacità di una verifica incrociata dei risultati ottenuti dai singoli moduli. Altre reti neurali multi modulo (MNN) sono inoltre state adottate da Hwarng (1995b) dimostrando di possedere capacità migliori rispetto a quelle di una rete MLP BP. Sistemi modulari composti da architetture eterogenee sono inoltre identificabili in letteratura. Per esempio Kang e Park (2000) utilizzano un sistema di monitoraggio composito basato sia sull'uso di reti neurali ad addestramento non supervisionato (precisamente le reti SOM) che reti neurali ad addestramento supervisionato (MLP BP). Altri esempi di sistemi modulari multi rete sono forniti dai diversi lavori che propongono identificazione dei pattern e stima dei principali parametri (Guh e Tannock 1999a, 1999b). In tal caso, una delle reti costituenti il sistema risolve il problema della identificazione del pattern non naturale, mentre altri moduli neurali risolvono il problema della stima dei parametri. Analogamente Chang e Ho (1999) propongono un sistema modulare di identificazione del cambiamento strutturale dello stato di un processo (attraverso identificazione di cambiamenti della varianza) composto di due moduli neurali. Il primo modulo, era addestrato per identificare cambi nella varianza dei dati osservati, il secondo, per identificare l'entità della variazione.

Si possono identificare tre classi fondamentali di reti modulari: reti composte da moduli indipendenti che lavorano parallelamente su una determinata porzione dello spazio degli ingressi (come ad esempio quella proposta da Cheng 1997). Reti composte da moduli indipendenti ma che lavorano tutti parallelamente sullo stesso ingresso (ad esempio Hwarng 1995b). Reti composte da moduli che lavorano sequenzialmente su differenti stadi di elaborazione del problema. Tuttavia, le reti modulari possono presentare dei problemi. Il più importante è legato alla complessità di tali reti, che rende gli algoritmi di apprendimento di difficile soluzione computazionale, poiché questa

richiede la ricerca di un ottimo globale (nella scelta dei pesi di connessione), piuttosto che un ottimo locale ad una singola rete.

In ultima analisi, le carte di controllo possono essere considerate come una rappresentazione grafica di una serie temporale di dati. Una delle architetture, proposte nella letteratura relativa agli algoritmi neurali, per il monitoraggio e la classificazione di serie temporali è la rete ricorrente RNN (rete di Elman e/o Jordan). Le reti ricorrenti, a differenza delle reti MLP presentano delle connessioni in modo da apprendere non soltanto in funzione degli ingressi (e uscite richieste) ma anche in relazione alla sequenza degli ingressi (e uscite). La differenza fra la rete Jordan e la rete Elman è che nel primo caso per ogni neurone del livello di output esiste un unico neurone memoria (*state unit*) che riporta nelle successive fasi di addestramento la risposta della rete nei livelli intermedi (*hidden*), nel caso della rete di Elman esiste un *context unit* dedicato ad ogni coppia (nodo *hidden*, nodo *output*). Inoltre la rete di Elman consente collegamenti “laterali” ovvero tra neuroni dello stesso livello. Le reti ricorrenti (Jordan o Elman) private dell’elemento di memoria (o *delay*), ovvero delle *state units* (Jordan) o *context units* (Elman), divengono a tutti gli effetti delle reti MLP. Si tratta in pratica di reti adeguate al riconoscimento (supervisionato) di serie di dati. Pur essendo una rete di questo tipo adeguata al monitoraggio di serie temporali, ad ora non è stato identificato in letteratura nessun lavoro che ne studiasse l’uso per il controllo di processi manifatturieri.

Nelle sezioni seguenti sono valutati diversi aspetti generali di progettazione di una rete neurale, e le diverse soluzioni che sono state identificate in letteratura.

2.3.1 Numero di livelli hidden

Molti tipi di reti neurali consistono di tre layer fondamentali di nodi: 1 strato di *input*, 1 di *output* e 1 (o più) *hidden*. Mentre gli strati di *input* e *output* sono generalmente legati alla specifica applicazione che s’intende risolvere, il numero di livelli *hidden*, e il numero di neuroni che ne fanno parte, costituiscono una scelta progettuale più generale, svincolata dal contesto applicativo, ma importantissima per le prestazioni della rete neurale.

Guo e Dooley (1992) evidenziano che non c’è nessun modo standard per decidere sul numero di layer, e di nodi, intermedi che possa essere generalizzato a qualunque classe di problema. Nel loro lavoro è indicato che l’uso dei metodi euristici può essere sufficiente per determinare il numero ottimo di nodi nei livelli *hidden*, e che un numero di strati pari ad uno o al più due, può essere adeguato per la maggior parte dei problemi. Gli stessi autori giustificano tale scelta evidenziando che una struttura neurale molto complessa potrebbe richiedere un insieme di addestramento molto esteso. S’incorre, in altri termini, nel problema dell’*over-training*, ovvero nella capacità della rete di apprendere dettagliatamente gli esempi di addestramento, ma nell’incapacità di poter generalizzare quest’apprendimento su generici dati di testing. In quasi tutti gli articoli presentati in letteratura, i metodi usati per l’identificazione del numero ottimo di livelli *hidden* sono tutti classificabili come metodi euristici di tipo *trial and error*.

2.3.2 Numero di nodi della rete

Ogni layer utilizzato nella rete neurale è costituito da un determinato numero di nodi (neuroni). Di seguito sono esaminati i metodi proposti in letteratura per la scelta del numero di neuroni per ogni singolo strato: *input*, *hidden* e *output*.

2.3.2.1 Nodi dello strato input

Il numero di neuroni dello strato di ingresso è direttamente legato al numero di segnali che s'intende far elaborare contemporaneamente dalla rete neurale: ognuno di questi segnali rappresenterà una variabile di ingresso. Nell'ambito dell'uso delle reti neurali per il controllo di qualità di un processo, questo numero è influenzato da alcune scelte come per esempio lo schema di codifica utilizzato per i dati in ingresso (per esempio una rappresentazione binaria richiede un numero d'ingressi alla rete, per ognuna delle variabili di input, pari al numero di bit utilizzati per la codifica). In letteratura sono state proposte scelte differenti ed il numero di ingressi varia tra 5 e 60 (vedi Tabella 2.3). Cheng (1995), critica il lavoro di Hwang (1992) per l'utilizzo di un numero elevato di neuroni in ingresso legato all'impiego di una codifica distribuita di tipo binario. Hwang, infatti, utilizza una codifica binaria in cui ogni input è codificato attraverso 7 bit. In tal caso una finestra di ingresso di 8 dati richiedeva una rete neurale con 56 ingressi. Un simile approccio è stato poi adottato anche da Al-Ghanim (1997).

2.3.2.2 Nodi degli strati hidden

Come per il numero di livelli hidden, anche il numero di neuroni per ognuno di tali strati è generalmente il risultato di una scelta empirica. Guo and Dooley (1992), affermano che, anche per il numero di neuroni dello strato hidden, non esistono dei metodi standard di scelta ottima che possano essere applicati in generale.

In genere aumentare il numero di neuroni negli strati intermedi porta da un lato ad un miglioramento delle capacità di generalizzazione della rete neurale, ma dall'altro ad un aumento della complessità computazionale e dei tempi necessari di addestramento. Hwang and Hubele (1993), dopo aver eseguito un piano fattoriale per valutare gli effetti del numero di nodi hidden (3 livelli: 15, 20 e 30), e del numero di zone di codifica degli ingressi (3 livelli: 7, 9 e 11) concludono affermando che aumenti del numero di neuroni nello strato hidden non comportano miglioramenti delle prestazioni ma unicamente incrementi dei tempi di addestramento. È necessario rilevare, tuttavia, che una simile conclusione vale per la particolare architettura che è stata presa in considerazione (MLP BP) e per il particolare tipo di pattern che è stato utilizzato nelle prove pubblicate nel lavoro di Hwang e Hubele (1993). Ciononostante, una simile conclusione è stata diffusamente condivisa nella letteratura (ad esempio Cheng 1997) poiché in genere nessun autore ha fornito indicazioni su metodologie di progettazione di reti neurali per il controllo di qualità, limitandosi a fornire indicazioni sulle prestazioni ottenute con una definita architettura e topologia.

Anche, nei lavori più recenti gli autori non propongono metodologie di progettazione, ma si limitano a fornire i risultati di metodi euristici o di tipo *trial and error*. Per esempio, Chang e Ho (1999) sperimentano tre architetture differenti per ognuno dei due moduli neurali componenti il sistema proposto. Le tre architetture differiscono unicamente per il numero di neuroni del livello intermedio che le compone. Gli autori si limitano a identificare l'architettura ottima in termini di prestazioni ottenute da quelle esaminate ma non forniscono ulteriori apprendimenti su una metodologia generale di costruzione. Analogamente Cook e Chiu (1998), riguardo alla rete RBF, valutano le prestazioni di un gruppo di reti che differisce per il numero di neuroni dello strato hidden. Dal loro lavoro appare che, per una rete RBF l'aumento del numero di neuroni dello strato hidden consente di ottenere un miglioramento delle prestazioni della rete in termini di riconoscimento di dati naturali e non naturali. Tuttavia, un numero eccessivamente elevato di neuroni hidden non produce alcun vantaggio ma peggiora unicamente la complessità computazionale della rete neurale.

2.3.2.3 Nodi dello strato output

Generalmente il numero di neuroni nello strato di uscita della rete è dipendente al modo in cui s'intende codificare la risposta dell'algoritmo neurale. Molto comune nella letteratura, è la scelta di associare un neurone in output per ogni particolare tipo di pattern che s'intende identificare nei dati d'ingresso.

Un importante fattore di progettazione è l'identificazione di un valore adatto di attivazione (ovvero un valore cut-off). Il valore di cut-off è necessario per trattare la natura inesatta degli output forniti dalla rete neurale. Questo valore è utilizzato per convertire i dati continui in uscita alla rete neurale (che dipendono dalla scelta della funzione di attivazione, vedi sezione seguente) in valori fissi che possano condurre ad una decisione. Tale decisione è ottenuta, appunto, comparando l'uscita della rete con il valore di soglia. Cheng (1997) indica che questo valore dipende dalla natura del processo e suggerisce l'utilizzo di un valore elevato quando l'analisi di un segnale di fuori controllo si rivela particolarmente costosa.

Nel caso di reti neurali progettate non per riconoscere specifici pattern, ma semplicemente per identificare cambiamenti di struttura nell'uscita del processo il numero di neuroni in uscita si riduce generalmente a solo uno. Hwarng (1995a) utilizza un approccio incrementale per determinare il numero ottimo di pattern da poter identificare in uscita ad una rete neurale, e dunque il numero di neuroni ottimo in uscita alla rete. L'autore conclude che ridurre il numero di pattern da identificare in una rete neurale, non necessariamente comporta sempre un aumento delle prestazioni della rete.

Per quanto riguarda i sistemi neurali di riconoscimento dei pattern, il numero di output presentato dalla rete corrisponde al numero di differenti classi di pattern che si intende identificare (Cheng 1997). Tuttavia, ciò non è sempre vero: per esempio è possibile addestrare una rete a riconoscere molti pattern di tipo ciclo a varie frequenze ed ampiezze, e poi verificare se un pattern di tipo ciclo (o un pattern di diversa natura) è stato identificato dalla rete, senza avere informazioni sulle ampiezze o sulle caratteristiche del particolare pattern. Altri esempi sono forniti da reti neurali nelle quali un unico neurone di uscita codifica differenti tipi di informazioni. Questo, per esempio, è ciò che accade nel lavoro di Smith (1994) nel quale un solo nodo è utilizzato per fornire indicazioni su tutti i tipi di pattern riconosciuti dalla rete neurale (shift nella media o nella varianza). In altre reti neurali, come quelle riportate da Pugh (1989, 1991), Smith (1994) ed anche Chang e Aw (1996) un unico neurone di uscita è stato utilizzato per identificare da tre fino a dieci classi distinte di andamenti non naturali.

2.3.3 Funzioni di trasferimento

Le funzioni di trasferimento influenzano notevolmente il funzionamento di una rete neurale (Hwarng 1992). Una funzione di trasferimento (o funzione di attivazione) è utilizzata come una funzione di soglia per poter limitare l'ingresso di un determinato nodo ad un predefinito range. Un altro importante beneficio è che essa può filtrare piccoli segnali, allo scopo di evitare che essi abbiano degli effetti sulle prestazioni della rete neurale nel riconoscimento di determinati pattern.

La funzione sigmoideale, è stata quella più comunemente utilizzata nelle applicazioni neurali per il controllo di processo (Hwarng e Hubele 1993). D'altro canto, Cheng (1997) afferma che una funzione di trasferimento a tangente iperbolica consente un'identificazione più accurata dei cambiamenti nella media del processo.

Generalmente, ad ogni layer della rete neurale è assegnato un tipo di funzione di trasferimento per tutti i nodi del layer. Tuttavia, è possibile configurare ogni singolo nodo con una funzione di trasferimento che non necessariamente è la stessa per tutti i

nodi dello strato. I possibili effetti di questo metodo non uniforme, non sono stati ancora valutati in letteratura.

La seguente Tabella 2.3 riporta la classificazione di alcuni lavori presentati in letteratura sulla base della struttura della rete neurale implementata.

<i>Autore</i>	<i>Arch. Neurale</i>	<i>Strati Hid.</i>	<i>Neuroni Input-Hidden-Output</i>	<i>Funzione di Trasferimento</i>	<i>Window Size</i>
Pugh (1989, 1991)	MLP BP	1	5-4-1	Sigmoid	5
Guo e Dooley (1992)	MLP BP	2	16-48-4-2	Sigmoid	5
Hwang e Hubele (1993a, 1993b)	MLP BP	1	56-30-10	Sigmoid	8
Smith (1994)	MLP BP	2	(13/3)-10-10-1 (13/3)-3-3-1	Sigmoid	10
Pham e Oztemel (1994)	LVQ	1	60-36-3	Non applicabile	60
Cheng (1995)	MLP BP	1	17-9-1	Tang. iperbolica	16
Hwang e Chong (1995)	ART1 map	1	56-28-9	Non applicabile	8
Chang e Aw (1996)	MLP BP	2	5-8-5-1	Sigmoid	5
Al-Ghanim (1997)	ART1	2	(70/35)-18-0	Non applicabile	(10/5)
Cheng (1997)	MNN/MLP BP	1	Tre reti (16-12-5) + (16-4-3) controllo	Tang. iperbolica	16
Cook e Chiu (1998)	RBF	1	5-4-1; 5-6-1; 5-8-1; 5-10-1; 5-12-1	Gaussian kernel function	2+3
Reddy <i>et al.</i> (1998)	MLP BP (Delta Bar Rule)	2	5-13-10-1 25-13-10-7	Tang. iperbolica	5 25
Leger <i>et al.</i> (1998)	MLP BP/RBF	1	5-15-7	Sigmoid / Gaussian kernel function	5
Chang e Ho (1999)	MLP BP	2	5-12-12-1 + 20-27-27-5	Sigmoid	5
Dedekayogullari e Burnak (1999)	MLP BP	1	8-9-1	Tang. iperbolica	5
Guh <i>et al.</i> (1999)	MLP BP	2	24-18-18-(7 o 2)	Sigmoid	24
Guh e Tannock (1999a, 1999b)	MLP BP	1	56-35-4 + tre 56-35-1	Sigmoid	56
Guh e Hsieh (1999)	MLP BP	1	56-35-4	Sigmoid	56
West <i>et al.</i> (1999)	MLP BP/RBF	1	6-6-2	Sigmoid / Gaussian kernel function	6
Kang e Park (2000)	MLP BP/SOM	1	5-5-1 + 4-4-4	Sigmoid	5
Cheng e Cheng (2001)	MLP BP	1	32-17-1	Sigmoid	32
Cook <i>et al.</i> (2001)	MLP BP	1	Non riportato	Sigmoid	Non rip.
Perry <i>et al.</i> (2001)	MLP BP	2	64-34-14-4 + 64-34-14-5	Sigmoid	64
Pham e Sagiroglu (2001)	MLP BP, QP, DBD, EDBD	1 o 2	60-40-(20)-6	Sigmoid	60
Wang e Cheng (2002)	MLP BP	2	12-12-12-2	Tang. Iperbolica	5

Tabella 2.3 – Classificazione della letteratura in base alle struttura della rete neurale implementata

2.4 Addestramento della rete neurale

Le reti neurali per il controllo di qualità devono essere addestrate a riconoscere andamenti naturali da quelli non naturali. Molti dei dati utilizzati per l'addestramento, indicati in letteratura, sono stati prodotti attraverso l'uso della simulazione computerizzata con tecniche Monte Carlo. Tuttavia, come indicato da Cheng (1997), se un pattern di interesse non può essere espresso per mezzo di equazioni matematiche, allora i dati devono essere collezionati dal sistema reale.

2.4.1 Trasformazione dei dati in ingresso

Ogni sistema di riconoscimento dei pattern su una serie di dati dovrebbe usare il maggior numero possibile di informazioni che possano essere computate al fine di identificare i possibili andamenti (Bishop 1995). D'altro canto, per motivi di efficienza i dati di addestramento dovrebbero essere ridotti ad una quantità gestibile riportando unicamente i dati che possano essere significativi per il training.

In tutti i lavori presentati in letteratura, i dati in ingresso alla rete sono sottoposti ad una serie di codifiche ed elaborazioni prima di essere presentati effettivamente alla rete neurale. Il principale obiettivo della serie di codifiche cui sottoporre i dati d'ingresso è quello partizionare lo spazio di variabilità in classi, e di rappresentare ogni possibile ingresso in funzione della classe di appartenenza. Le funzioni di codifica sono rivolte, da un lato, a massimizzare le similarità dei pattern all'interno di una medesima classe, e dall'altro, a minimizzare quelle tra i pattern in differenti classi. In altri termini, le trasformazioni dei dati in ingresso consentono di ridurre la variabilità degli input della rete neurale, facilitando in questo modo l'apprendimento e la velocità di convergenza dell'algoritmo di training.

La codifica dei dati di ingresso ad una rete neurale è ritenuta indispensabile per garantire il raggiungimento di un'opportuna capacità di generalizzazione. Le caratteristiche degli input della rete sono comunemente definite features. Le features possono essere o selezionate, o estratte dal processo in esame (Bishop 1995). In particolare, nell'ambito dell'applicazione delle reti neurali per il monitoraggio di processo, è possibile classificare i lavori proposti in letteratura in due principali approcci.

- *Input feature semplice*: in tal caso l'input della rete è basato sull'uso di una singola feature del processo. Tale feature coincide normalmente con i dati grezzi di uscita del processo.
- *Input feature mista*: in tal caso l'input del sistema è caratterizzato da più feature (omogenee o eterogenee). Tali feature rappresentano solitamente una combinazione fra i dati grezzi del sistema, ed alcune elaborazioni statistiche, come per esempio medie campionarie calcolate per gruppi successivi di dati.

Diversi autori propongono di utilizzare come input alla rete i soli dati ottenuti dal processo, eventualmente elaborati attraverso le funzioni riportate di seguito (Pugh 1989, 1991, Cheng 1995, 1997, Perry *et al.* 2001). Altri autori indicano la necessità di utilizzare elaborazioni statistiche assieme ai dati grezzi della rete (Guo e Dooley 1992, Smith 1994, Wang e Cheng 2002). Tipiche elaborazioni utilizzate sono rappresentate da medie campionarie di sotto-gruppi di osservazioni. Cook e Chiu (1998) propongono, ad esempio di fornire in ingresso al sistema neurale proposto per il monitoraggio di dati correlati nel tempo cinque input: 1) il valore attuale di uscita del processo all'istante t , 2) il valore presentato all'istante precedente $t-1$, 3) la previsione del valore all'istante

attuale ottenuta attraverso la conoscenza di un opportuno modello AR del primo ordine, e sulla base del valore all'istante precedente, 4) la differenza dei valori tra l'istante attuale t e quello previsto dal modello AR(1) (ovvero il residuo all'istante t), 5) la differenza fra il valore osservato all'istante t e quello osservato nell'istante precedente $t-1$.

Smith (1994) propone un sistema neurale che accetta in ingresso un gruppo di dieci punti ottenuti dal processo, e di tre misure statistiche ad essi associati (media, deviazione standard e range). Chang e Ho (1999), al fine di poter identificare modifiche della varianza di un processo, forniscono in ingresso alla rete neurale cinque dati che rappresentano gli scarti in valore assoluto fra cinque osservazioni successive del processo e la loro media campionaria.

Una volta prescelto il tipo di feature in ingresso alla rete, il passo successivo è, come già accennato, quello di codificare l'input prima di fornirlo in ingresso alla rete. Le più importanti elaborazioni e trasformazioni dei dati adottate in letteratura sono: 1) la standardizzazione, 2) zoning, 3) scaling, 4) l'uso di una rappresentazione continua (contrapposta a quella binaria). Nelle sezioni seguenti ogni singola codifica è brevemente discussa.

2.4.1.1 Standardizzazione

La standardizzazione dei dati in ingresso consente a differenti valori di media e deviazione standard di essere elaborati da una stessa rete neurale, senza bisogno di effettuare dei nuovi addestramenti. Alcuni ricercatori non accennano per nulla all'uso della standardizzazione poiché esprimono la componente aleatoria secondo una distribuzione normale di media 0 e deviazione standard 1. Tuttavia, la fase di standardizzazione dei dati d'ingresso è ritenuta fondamentale. Se i dati non sono standardizzati allora la rete potrebbe essere in grado di riconoscere pattern unicamente rispetto ad una determinata media e deviazione standard del processo. La standardizzazione consente di utilizzare una stessa rete anche per processi che presentano differenti medie e/o deviazioni standard.

2.4.1.2 Zoning

Questa procedura consiste nel partizionare l'insieme di variabilità dell'ingresso (solitamente standardizzato) in sottoinsiemi disgiunti. Come regola comune, ogni dato rientrante in una zona caratterizzata da un limite superiore e da un limite inferiore è codificata con un unico valore che rappresenta di tale zona (per esempio il valore medio centrale). Hwang (1993a, 1993b), ad esempio, suddivide l'insieme di variabilità dei valori in ingresso in sette zone, se un dato standardizzato ricade in una di queste zone esso è codificato con un codice binario univoco. Ogni osservazione standardizzata è associata in una di queste 7 zone. Ovviamente tutti i valori che ricadono in una medesima zona avranno lo stesso codice di rappresentazione. Cheng (1995) indica che un numero di zone eccessivamente ridotto può comportare perdita di informazione importante per l'identificazione di determinati tipi di pattern. Pertanto nel primo lavoro propone di suddividere l'intervallo di variabilità dei dati in ingresso standardizzati in 16 zone contigue di ampiezza 0.5. Successivamente lo stesso autore (Cheng 1997) propone di suddividere l'insieme di variabilità in 32 zone di ampiezza 0.5. Guh e Tannock (1999a, 1999b) suddividono gli ingressi in 61 intervalli di ampiezza 0.25.

Una codifica basata sull'uso di un numero ridotto di zone, consente una rapida convergenza dell'insieme di addestramento, ma provoca una ridotta sensibilità della rete al riconoscimento di variazioni minori dell'ampiezza delle zone di codifica. Tuttavia, se il numero di zone è eccessivamente ampio allora il clustering dei dati ingresso in classi

potrebbe essere sostanzialmente irrilevante (Hwarng e Hubele 1993) con la conseguenza che l'algoritmo di addestramento potrebbe non ammettere convergenza.

Il lavoro di Hwarng a Hubele inoltre, ha dimostrato che il numero di zone scelto per la codifica, riveste un'importanza statisticamente rilevante sulle prestazioni della rete neurale MLP con algoritmo di apprendimento BP. In particolare è stato dimostrato che l'utilizzo di 7 zone di codifica permette di ottenere prestazioni migliori rispetto ad una codifica di 9 o 11 zone in quanto permette alla rete di segnalare un numero minore di falsi allarmi. Questo valore è stato ottenuto attraverso varie comparazioni fra diverse configurazioni della rete BP.

2.4.1.3 Scaling

I valori d'ingresso della rete neurale potrebbero assumere qualunque valore, tuttavia, limitare la loro variabilità all'interno di un range prefissato produce un migliore addestramento della rete neurale. Lo scaling è la trasformazione che in genere è adottata da diversi autori al fine modificare i dati di ingresso per avere variabilità all'interno di un predeterminato intervallo di variazione. Lo scaling è generalmente imposto sulla base del particolare tipo di funzione di attivazione utilizzata nella rete neurale. Il vantaggio dell'uso dello scaling sui dati di ingresso è principalmente legato all'attenuazione dell'effetto degli outliers, e alla trasformazione dei dati in modo che l'insieme di variabilità sia compreso in opportuni intervalli. Le trasformazioni scaling sono tipicamente trasformazioni lineari da un intervallo di definizione ad un intervallo di arrivo (tipicamente [0,1] oppure [-1,1] in relazione alle funzioni di attivazione utilizzate dalla rete neurale). Tutti i punti al di fuori dell'intervallo di definizione sono riportati nei valori estremi dell'intervallo di definizione.

Per esempio, Cheng (1995) trasforma i dati d'ingresso, i cui valori possono variare da $-\infty$ a $+\infty$, in modo che essi ricadano all'interno dell'intervallo -7σ a $+7\sigma$. Altri esempi di scaling dei dati d'ingresso sono riportati negli articoli di Hwarng e Hubele (1993a, 1993b).

2.4.1.4 Uso della codifica continua/binaria

In letteratura, gli ingressi e le uscite di una rete neurale potrebbero essere espressi come valori continui (reali) o al contrario con formati discreti di tipo binario. Smith (1994) per esempio, utilizza una codifica continua sia per i dati d'ingresso sia per il risultato in uscita della rete neurale. Quest'ultimo, infatti, è codificato in modo tale da variare all'interno dell'intervallo [0,1] dove valori prossimi a 0.1 sono intesi a codificare shift della media, valori prossimi a 0.9 shift della varianza e valori prossimi a 0.5 andamenti naturali dei dati.

D'altro canto Hwarng (1992) propone una procedura di conversione dei dati osservati sul processo in valori binari sulla base della convinzione che ciò possa consentire di catturare maggiori informazioni sulla struttura dei dati rispetto all'utilizzo diretto dei dati originali. Tuttavia, come evidenziato da diversi autori il formato binario che è in grado di produrre solo due decisioni (0 o 1) potrebbe essere meno flessibile rispetto ad una codifica continua (o comunque scalata su un numero sufficientemente ampio di intervalli). D'altro canto, la fine risoluzione dei dati che può essere fornita dall'uso di una codifica continua potrebbe risultare in reti neurali molto più complesse che richiedono insiemi di addestramento molto ampi.

La seguente Tabella 2.4 riporta una classificazione della letteratura in base al tipo di algoritmo di codifica dei dati di ingresso alla rete neurale.

<i>Autore</i>	<i>Input feature</i>	<i>Standar.</i>	<i>Zoning</i>	<i>Scaling</i>	<i>Codifica</i>
Pugh (1989, 1991)	Semplice	Si	Si	Si	Binaria
Guo e Dooley (1992)	Mista	No	No	No	Continua
Hwang e Hubele (1993a, 1993b)	Semplice	Si	Si	Si	Binaria
Smith (1994)	Mista	No	No	No	Continua
Pham e Oztemel (1994)	Semplice	Si	No	Si	Continua
Cheng (1995)	Semplice	No	Si	Si	Continua
Hwang e Chong (1995)	Semplice	Si	Si	Si	Binaria
Chang e Aw (1996)	Semplice	No	No	No	Continua
Al-Ghanim (1997)	Semplice	Si	Si	Si	Binaria
Cheng (1997)	Semplice	Si	Si	Si	Continua
Cook e Chiu (1998)	Mista	No	No	No	Continua
Reddy <i>et al.</i> (1998)	Mista	No	No	No	Continua
Leger <i>et al.</i> (1998)	Semplice	Si	No	No	Continua
Chang e Ho (1999)	Mista	No	No	No	Continua
Dedekayogullari e Burnak (1999)	Mista	No	No	Si	Continua
Guh <i>et al.</i> (1999)	Mista	No	No	No	Continua
Guh e Tannock (1999a, 1999b)	Semplice	Si	Si	Si	Continua
Guh e Hsieh (1999)	Semplice	Si	Si	Si	Continua
West <i>et al.</i> (1999)	Semplice	No	No	Si	Continua
Kang e Park (2000)	Mista	No	No	No	Continua
Cheng e Cheng (2001)	Semplice	No	No	Si	Continua
Cook <i>et al.</i> (2001)	Mista	No	No	No	Continua
Perry <i>et al.</i> (2001)	Semplice	Si	Si	Si	Binaria
Pham e Sagiroglu (2001)	Semplice	Non riportata	Non riportata	Non riportata	Non riportata
Wang e Cheng (2002)	Mista	No	No	Si	Continua

Tabella 2.4 – Classificazione della letteratura in base al tipo di algoritmo di codifica dei dati in ingresso alla rete neurale

2.5 Addestramento della rete neurale

L'ampiezza richiesta all'insieme di dati di addestramento, e le sue implicazioni in termini di tempo di raggruppamento e preparazione dei dati, sono gli aspetti principali riguardanti il training delle reti neurali. L'ampiezza dell'insieme di addestramento, inoltre, è influenzata da fattori quali l'algoritmo di addestramento, l'architettura della rete neurale, la natura del problema di controllo che s'intende risolvere (per esempio identificazione di cambiamenti nella media e/o nella varianza del processo). La questione più dibattuta in letteratura riguarda se l'insieme di addestramento deve includere esempi di pattern naturali, oltre ai non naturali e, per questi ultimi, quali tipi di ampiezze di disturbo utilizzare.

Nella maggior parte dei lavori presentati in letteratura i diversi ricercatori hanno tentato di includere il maggior numero possibile di pattern non naturali nell'insieme di addestramento e tali da coprire il più ampio range di ampiezze. Gli esperimenti di Cheng (1995), infatti, dimostrano che l'utilizzo di piccoli insiemi di training può causare risultati non desiderati sulle prestazioni della rete. Tuttavia, training set eccessivamente ampi possono determinare una ridotta capacità della rete di riuscire ad identificare pattern di piccola ampiezza. D'altro canto, l'utilizzo di ampi insiemi di addestramento può comportare una maggiore complessità dell'intera fase di training.

Una possibilità di ridurre l'ampiezza dell'insieme di addestramento della rete neurale consiste nella scelta di suddividere il carico di addestramento fra differenti reti neurali. Per esempio, questo è ciò che è stato proposto da Hwarng (1995) che ha presentato un sistema neurale specializzato al riconoscimento di un particolare tipo di pattern (invece che più pattern). In modo simile la rete neurale modulare proposta da Cheng (1997) è organizzata in modo tale che ognuna di tre differenti reti neurali (definite appunto *local expert network*) sia responsabile unicamente di un determinato sottoinsieme dei dati di addestramento.

È stato suggerito in letteratura che un metodo per ottenere un gruppo di dati sufficientemente ampio per costruire un insieme di addestramento è di adottare un approccio incrementale fino a che un aumento dell'insieme di addestramento non produce più alcun beneficio ulteriore sulle prestazioni della rete neurale. In effetti, un simile approccio è stato adottato da Hwarng e Hubele (1993) per identificare un adeguato set di addestramento sulla base degli errori di Tipo I e di Tipo II prodotti dalla rete neurale. Gli stessi autori, inoltre, raccomandano di costruire opportunamente l'insieme di addestramento anche sulla base delle prestazioni attese dalla rete. Per esempio, poiché il pattern di tipo ciclo è quello per il quale la rete presenta le prestazioni peggiori, gli autori raccomandano di sbilanciare il numero di esempi di addestramento più su il pattern ciclo, piuttosto che su altri tipi di pattern come ad esempio il trend. Quest'osservazione, però, si deve considerare limitata alla sola rete MLP BP visto che, con altre architetture, come per esempio la rete ART (Hwarng e Chong 1995, Al-Ghanim 1997), è richiesto un insieme di addestramento decisamente più ridotto (al limite un esempio per ogni pattern) in modo da raggiungere prestazioni comparabili con quelle di una rete MLP BP.

2.5.1 Frequenza di presentazione del training set

Nell'ambito della procedura di addestramento di una rete neurale supervisionata, è necessario stabilire il numero di epoche di training da adottare. Tale parametro, che in pratica coincide con il numero di volte in cui il training set è presentato alla rete neurale, varia tra i diversi lavori presentati in letteratura. Per esempio, Pham e Oztemel (1994) indicano che il loro sistema modificato LVQ termina l'addestramento dopo 20 epoche, mentre, per ottenere risultati analoghi da una rete Back Propagation è necessario proseguire l'addestramento per un numero di epoche pari ad almeno 10 volte quelle necessarie per LVQ. Hwarng e Chong (1995) hanno dimostrato che, a differenza della rete MLP BP, un modello basato sull'architettura ART richiede un numero di epoche decisamente più contenuto. Al limite, è possibile che ogni esempio sia presentato alla rete un'unica volta.

Un ulteriore importante aspetto dell'addestramento delle reti neurali è evitare i problemi noti in letteratura come *undertraining* e *overtraining*. Il fenomeno di *undertraining* può verificarsi quando la capacità della rete è insufficiente e può accettare unicamente training set di ampiezza limitata. Ciò risulta in un gran numero di errori sia in fase di training (difficoltà di convergenza dell'algoritmo di addestramento) che in

fase di testing (difficoltà di generalizzazione della rete addestrata). Al contrario, si ha overtraining quando, l'utilizzo di un ampio insieme di addestramento provoca come risposta una buona capacità della rete di riconoscere i pattern utilizzati in addestramento, ma una limitata capacità di generalizzazione. Secondo alcuni autori, talune architetture, come la rete SOM, non soffrono di tale inconveniente come per esempio le reti BP.

2.5.2 Presentazione dei pattern di addestramento

Gli insiemi di addestramento possono essere classificati a due livelli. In primo luogo si possono creare insiemi di addestramento costituiti da differenti famiglie di pattern. In secondo luogo, all'interno di ogni famiglia è possibile identificare dei specifici sottoinsiemi in funzione della scelta fatta per i parametri relativi. Per esempio Cheng (1997) utilizza cinque differenti famiglie di pattern per l'addestramento della rete. Nell'ambito di alcune di tali famiglie, inoltre, sono utilizzati 9 differenti livelli del parametro di ampiezza. Ad ognuno di questi livelli, pattern di addestramento possono essere presentati in due diversi modi: sequenzialmente o random. Ad esempio, Guo e Dooley (1992) hanno indicato che usare un campionamento uniforme, ovvero un training in cui si alternano in modo random pattern di differenti famiglie (in questo caso shift della media e della varianza), migliora la capacità di identificazione della rete neurale.

In modo simile, Hwarng e Hubele (1993b) indicano che la selezione casuale dell'insieme di addestramento all'interno di ogni classe di pattern produce una rapida convergenza dell'algoritmo di addestramento. Gli autori, attribuiscono quest'aspetto alla possibilità che l'adattamento dei pesi sia molto più pronto rispetto al caso in cui si presentano tutti i pattern di una classe in successione. Inoltre, Hwarng e Hubele hanno indicato che una modifica dei pesi delle connessioni dopo ogni presentazione di pattern presenti nella lista di addestramento, generalmente produce una più veloce convergenza rispetto invece al caso di cambio improvviso di tutti i pesi della rete dopo l'intera presentazione di tutto l'insieme di addestramento.

2.5.3 Inclusione o esclusione di dati in controllo

Cheng (1995) critica gli approcci di controllo, basato sull'uso di reti neurali, nel quale l'addestramento è limitato solo a dati in fuori controllo. Cheng giustifica tale appunto indicando che, solitamente, un reale sistema sottoposto a controllo parte da una condizione di variabilità naturale, e dunque la rete deve essere in grado di riconoscere questa condizione altrettanto bene come quella di variazione non naturale.

Normalmente, durante una tipica sessione di analisi di un processo attraverso una carta di controllo, gli output del processo (per esempio le medie campionarie) sono esaminati in modo continuo. Quando un pattern non naturale si sovrappone nel processo, solitamente la rete è chiamata a valutare finestre di dati nei quali coesistono sia output naturali che non naturali. Hwarng e Hubele (1993a) indicano, infatti, che la loro rete neurale presenta le migliori prestazioni nel riconoscimento del pattern shift quando, quando lo shift della media è posto nel punto medio della finestra di controllo.

Infine, anche se si può condividere l'idea proposta da Cheng di utilizzare addestramenti che comprendono sia dati di variazione naturale che dati di variazione non naturale, in realtà in letteratura non è fornita alcuna risposta chiara sul numero di pattern naturali da utilizzare per l'addestramento. Cheng (1995) suggerisce semplicemente di utilizzare l'ARL di uno schema di controllo comparabile per dedurre il numero adeguato di pattern naturali nell'insieme di addestramento.

2.5.4 Rapporto segnale/rumore

Come indicato da Cheng (1997) molti dei sistemi di riconoscimento dei pattern sulle carte di controllo sono basati su un modello del sistema nel quale l'uscita monitorata è rappresentata come la somma di due elementi: un componente di variazione naturale (o rumore) e un segnale speciale di disturbo. Il primo, rappresenta le variazioni dovute a cause non assegnabili (naturali) sul processo, il secondo è legato alla presenza di cause non naturali ovvero assegnabili. La componente naturale di variazione, è solitamente rappresentata come una successione di valori IID distribuiti normalmente con deviazione standard σ .

Se il rapporto segnale/rumore è piccolo allora il segnale speciale di disturbo può essere drasticamente distorto ad un limite tale da non poter essere più distinguibile dal rumore naturale. Hwang (1995a, 1995b) indica che è rilevante identificare pattern non naturali solo se questi mantengono una forma riconoscibile anche dopo essere stati contaminati da rumore naturale. A supporto di quest'osservazione c'è il fatto che se il pattern non modifica sostanzialmente l'andamento naturale, è possibile che il processo continui ad essere capace di produrre entro determinati limiti di specifica. In altri termini, se il segnale speciale di disturbo è talmente debole rispetto al rumore da non provocare una sostanziale modifica della variabilità, allora può essere conveniente evitare indicazioni di falsi allarmi perché la capacità del processo di produrre in specifica non ne risulta sostanzialmente modificata.

Cheng (1995) critica l'approccio di Hwang nel quale sono considerati solo elevati rapporti segnali/rumore. La critica è fondata sul fatto che, limitare l'utilizzo delle reti neurali al solo riconoscimento di elevati rapporti segnale/rumore, rende le prestazioni di queste ultime sostanzialmente simili a quelle di una normale carta di controllo. Secondo diversi autori (Cheng 1995 e Smith 1994) le carte di controllo sono sufficientemente adeguate a identificare ampie variazioni della media (ovvero elevate variazioni segnale rumore).

L'impiego di reti neurali risulta giustificato solo se si possono migliorare le prestazioni di metodi SPC tradizionali sia con riferimento alle comuni carte di controllo, ma anche con riferimento ai metodi SPC più avanzati tipo carte CUSUM e EWMA. In altri termini è utile poter disporre di un metodo che permetta di identificare segnali speciali di disturbo sia nel caso di elevati rapporti segnale/rumore ma anche nel caso di piccoli rapporti.

2.5.5 Selezione della finestra di dati

Le osservazioni raccolte in uscita ad un processo, costituiscono una successione temporale di valori che devono essere utilizzati per fornire un'indicazione sul suo stato. Se si considera il funzionamento di una carta di controllo della media, la segnalazione di una condizione di fuori controllo è unicamente basata sul confronto dell'ultima osservazione prodotta, con limiti di variabilità. In tal caso si può considerare che, la finestra sulla sequenza dei dati utilizzati per valutare lo stato del processo, è limitata ad un singolo punto. Nel caso d'utilizzo delle reti neurali è possibile sfruttare le capacità computazionali della rete per dedurre una conclusione sullo stato del processo sulla base di una finestra più ampia d'osservazioni. L'ampiezza di questa finestra è cruciale per le prestazioni della rete neurale ed è chiamata *Window Size* (WS).

La determinazione della WS si basa, fondamentalmente, su considerazioni inerenti la capacità di riconoscimento che ha la rete neurale sia in termini di tempo, che di capacità discriminatoria. Evidentemente una rete neurale con una piccola ampiezza di finestra reagirà in maniera molto rapida, ma nello stesso tempo porterà ad un maggiore

numero di falsi allarmi. D'altro canto, una rete neurale con un'ampia finestra avrà una migliore capacità discriminativa ma reagirà molto più lentamente. L'ampiezza della finestra è determinata in accordo alla probabilità teorica di commettere un errore di prima specie (o analogamente al ARL) per ogni pattern da identificare, tuttavia in letteratura, non si trova alcun approccio "sistematico" al problema, ma sono adottati unicamente valutazioni empiriche.

Per le applicazioni di riconoscimento pattern sulle carte di controllo, Hwarng (1992) indica che l'ampiezza della finestra di dati deve essere tale da garantire un adeguato livello di errori di Tipo I. In genere, maggiore è l'ampiezza della finestra di punti utilizzata per il controllo del processo, minore è la probabilità di commettere errori di Tipo I.

Analizzando le pubblicazioni presentate in letteratura, non si evidenzia un forte consenso fra i differenti ricercatori in merito alla scelta della WS. Per esempio, un primo lavoro di Hwarng e Hubele (1993a) ha indicato che un'ampiezza della finestra pari a 8 punti si può considerare ragionevolmente ampia per identificare i principali tipi di pattern. Una ragione per la quale non sono considerate grandi di WS è legata al lungo tempo di attesa che sarebbe richiesto prima di poter ottenere dalla rete una classificazione sullo stato del processo. D'altro canto, altri ricercatori come Oztemel (1993) hanno proposto delle window size di ampiezza ben superiore (56 punti) indicando prestazioni superiori per il riconoscimento di differenti tipi di pattern. Lo stesso autore critica il lavoro di Hwarng (1993), poiché considera che una finestra di soli 8 osservazioni, sia del tutto insufficiente per individuare il vero stato del processo.

Hwarng e Hubele (1993b) suggeriscono che una tecnica statistica come un piano fattoriale 3^2 può costituire un utile strumento per determinare la dimensione ottima della WS per un'applicazione di riconoscimento dei pattern sulle carte di controllo attraverso l'uso di reti neurali. Ovviamente un importante fattore che deve essere notato è che la sperimentazione di Hwarng e Hubele è stata condotta con riferimento ad uno specifico insieme di pattern da identificare e con specifici parametri (ampiezze, cicli et.). Per esempio il lavoro di Cheng (1997) indica che incrementare l'ampiezza della finestra della rete oltre il valore 16 non comporta alcun reale beneficio sulle prestazioni della rete e che, invece, produce solo un aumento delle dimensioni della rete stessa. In realtà, però, questo risultato non deve essere considerato valido in generale per qualunque tipo di rete. È necessario, quindi, approfondire la reale influenza della finestra di addestramento attraverso studi più approfonditi.

Uno studio sistematico circa l'ampiezza ottimale della WS è stato condotto da Guh e Tannock (1999b) con lo scopo di valutare la relazione fra WS e ARL in controllo (ovvero errore di prima specie). Lo studio è stato condotto utilizzando i medesimi dati di test e facendo riferimento a reti con una topologia simile, (ad esclusione del numero di neuroni nel livello di input). Dai risultati è stato evidenziato che l'ampiezza ottimale della finestra è 16 che comporta, (fissando un valore di cut-off θ pari a 0.8) un valore di ARL di 814 su dati normali standardizzati, indipendenti e identicamente distribuiti (coerente con quello indicato da Cheng, 800, che utilizza una W.S. pari a 16); un valore W.S. pari a 8 (θ pari a 0.8) comporterebbe un livello ARL più basso, pari circa a 100.

Il problema della scelta della WS risulta poi maggiormente complicato se si considerano quelle applicazioni delle reti neurali nelle quali s'intende discriminare i diversi tipi di pattern, ovvero, nelle quali è necessario evitare di confondere pattern che presentano caratteristiche simili. A questo fine Cheng (1997) suggerisce che l'unica possibile soluzione è di disporre di una finestra sufficientemente ampia da includere un numero tale di punti da identificare correttamente il pattern: ad esempio con pochi punti un ciclo di ampio periodo potrebbe essere confuso con un trend. Hwarng (1992), ha

proposto un approccio con finestra ad ampiezza variabile in modo da adeguare in maniera incrementale l'ampiezza al tipo particolare di problema. Purtroppo in nessuno dei lavori successivi è stato indicato come possa essere implementato un algoritmo neurale ad ampiezza di finestra variabile.

In breve, appare evidente come ad oggi, la tendenza sia quella di utilizzare finestra di elevata ampiezza allo scopo di ottenere prestazioni di riconoscimento sempre maggiori. Per esempio, Hwarng (1995) dopo aver utilizzato una finestra di 8 punti suggerisce che risulta più adeguato l'utilizzo di una finestra a 16 punti per migliorare le prestazioni di riconoscimento in special modo per pattern come il ciclo. Analogamente Cheng (1997) conclude che un numero maggiore di osservazioni è necessario per discriminare tra differenti tipi di pattern non naturali.

2.5.5.1 Fixed e Moving Window

Un'altra caratteristica della finestra di dati è legata al fatto che essa può essere fissa (*fixed window*) o mobile (*moving window*). Le due possibilità sono riferite al caso di utilizzo on-line di una rete neurale già addestrata, se indichiamo con m la WS, allora la rete riceve il primo input con le prime m osservazioni campionarie che derivano dal processo osservato, pertanto: si avvia il monitoraggio della carta con la rete neurale solo al m^{esimo} dato osservato. Successivamente, la finestra è ottenuta facendo riferimento agli ultimi $m - p$ campioni della finestra precedente e le p più recenti osservazioni generate dal processo. Nel caso *fixed window* risulta $p = m$, ovvero non esiste sovrapposizione di punti fra due finestre successive. Nel caso *moving window*, risulta $p < m$ (tipicamente $p = 1$) ovvero esiste sovrapposizione di $m - p$ punti fra due finestre successive.

Appaiono evidenti i vantaggi e gli svantaggi delle due soluzioni: il metodo della finestra mobile è sicuramente più efficiente nel senso che ha più possibilità di evidenziare per primo la presenza di un pattern. Il metodo della finestra fissa pur non essendo efficiente è, quantomeno, più efficace poiché, qualora evidenziato un pattern, il riconoscimento del tipo (ovvero della classe) è più attendibile.

Sostanzialmente tutti gli autori fanno riferimento all'utilizzo del metodo della finestra mobile durante le fasi di utilizzo di una rete addestrata. Per contro, nelle fasi di training e di testing sono generate finestre di m osservazioni (pattern campioni) utilizzando una modalità di analisi a finestra fissa.

La seguente Tabella 2.5 riporta una classificazione della letteratura sulla base dei parametri utilizzati per la fase di addestramento della rete neurale.

<i>Autore</i>	<i>Window Size</i>	<i>Training Set (Medio)</i>	<i>Parametri del training (min:incremento:MAX)</i>
Pugh (1989, 1991)	5	100	
Guo e Dooley (1992)	5	1000	0.25 e 0.5:0.5:2.0 (media) 1.25 e 1.5:0.5:3.0 (var.)
Hwarng e Hubele (1993a, 1993b)	8	48	0.20:0.05:0.35 (trend) (SD 0:0.025:0.175) 1.50:0.50:2.50 (ciclo+sys var.) (SD 0:0.05:0.25) -0.20:0.20:0.20 (strat.) (SD 0.10:0.05:0.20) 1.75:0.25:2.25 (mix) (SD 0.10:0.20:0.50) 1.50:0.50:2.50 (shift) (SD 0.20:0.20:0.60)
Smith (1994)	10	500	1 e 3 (media) 1, 2 e 3 (varianza)
Pham e Oztemel (1994)	60	83	
Cheng (1995)	16	2300 - 2500	1.0:0.25:6.0

			0.01:0.02:0.5(trend)
Hwarng e Chong (1995)	8	1 - 9	Non applicabile
Chang e Aw (1996)	5	100	-4.0:1:0:4.0
Al-Ghanim (1997)	(10/5)	31-98	0: Solo naturali
Cheng (1997)	16	500	1.0:0.25:0.3 0.2:0.025:0.5 (trend)
Cook e Chiu (1998)	2+3	123 - 227	1.0:0.5:2.0
Reddy <i>et al.</i> (1998)	5 25	100	-1.0:xx:1.0 (media) 0.25:xx:1.0 (varianza) -0.04:xx:0.04 (cicli)
Leger <i>et al.</i> (1998)	5	Non riportato	Non riportato
Chang e Ho (1999)	5	120	1.0:1.0:5.0 (varianza)
Dedekayogullari e Burnak (1999)	5	375	-3.0:3.0:3.0 (media) 1.0:2.0:3.0 (varianza)
Guh <i>et al.</i> (1999)	24	810 600 (nat.)	-3.0:0.25:-1.0;1.0:0.25:3.0 (shift) -0.26:0.02:-0.10;0.10:0.02:0.26 (trend) 1.0:0.25:3.0 (ciclo 8 e mixture)
Guh e Tannock (1999a, 1999b)	56	1120 560 (ciclo) 420 (nat.)	-2.5:0.25:-1.0; 1.0:0.25:2.5 (shift) -0.22:0.02:-0.10; 0.10:0.02:0.22 (trend) 1.0:0.25:2.5 (ciclo)
Guh e Tannock (1999b)	56	1120 560 (ciclo) 420 (nat.)	-7.0:1.0:7.0 (shift) -0.14:0.02:-0.14 (trend) 0.5:0.5:7.0 (ciclo)
Guh e Hsieh (1999)	56	960 480(nat.)	-7.0:1.0:7.0 (shift) -0.14:0.02:-0.14 (trend) 0.5:0.5:7.0 (ciclo)
West <i>et al.</i> (1999)	6	Non riportato	Non riportato
Kang e Park (2000)	5	Non riportato	Non riportato
Cheng e Cheng (2001)	32	1500	Non riportato
Cook <i>et al.</i> (2001)	Non rip.	Non rip.	Non rip.
Perry <i>et al.</i> (2001)	64	250	0.1:0.05:0.3 (trend) 1.0:0.5:3.0 (ciclo, mixture) 1.0:0.25:2.0 (var. sis.)
Pham e Sagiroglu (2001)	60	83	0.2:xx:0.5
Wang e Cheng (2002)	5		-4.0:1.0:4.0

Tabella 2.5 – Classificazione della letteratura in base al tipo e ai parametri di addestramento della rete neurale

2.6 Prestazioni e testing

I criteri di misura delle prestazioni dei sistemi neurali per il controllo di qualità di processi variano notevolmente fra i diversi lavori proposti in letteratura. Questa è, infatti, un'area dove una maggiore standardizzazione potrebbe fornire una migliore comparabilità tra i differenti lavori, nonché criteri generali di benchmarking.

In diversi lavori, i sistemi di controllo neurale sono sottoposti ad una comparazione con i metodi statistici SPC che comprendono carte di controllo X, R, S, carte per il monitoraggio di piccole variazioni della media quali CUSUM ed EWMA, schemi

combinati Shewhart CUSUM, nonché carte per il controllo multivariato. In genere le comparazioni con i metodi di controllo statistico sono condotte per mezzo delle comuni misure di prestazione indicate nella letteratura SPC (Montgomery 2000), ad esempio Average Run Length (ARL), errori di Tipo I ed errori di Tipo II.

Cheng (1995) indica che per compiere degli studi comparativi tra due sistemi di controllo, ad esempio un metodo neurale ed uno SPC, oppure due diversi sistemi neurali, è necessario mantenerne simili gli errori di Tipo I. Ciò equivale a dire: tarare i due sistemi a confronto in modo da mantenere comparabili i valori di ARL ottenuti in corrispondenza di dati naturali generati da un processo in uno stato di controllo statistico. Questa condizione è indispensabile per confrontare efficacemente le prestazioni della rete neurale con quelle di una carta di controllo. Il confronto può essere poi condotto attraverso la comparazione dei valori di errore di Tipo II (o analogamente di ARL) nel monitoraggio di una sequenza di dati non naturali generati da un processo in uno stato di fuori controllo statistico.

Per esempio Cheng (1995) stima, attraverso la simulazione, i valori di ARL per ogni valore di ampiezza di cambiamento della media di processo. Il valore di prestazione è calcolato su 10000 run indipendenti di simulazione. La lunghezza dei run, (RL) è stata calcolata dall'istante in cui è simulato il primo valore della sequenza non naturale di dati, fino all'istante in cui il sistema di controllo segnala un andamento non naturale dei dati di processo. Ciò equivale a conteggiare in differenti run di simulazione il numero di osservazioni che intercorrono tra avvio del fuori controllo e indicazione di un segnale di allarme da parte del sistema di controllo.

Chang e Aw (1996) applicano un criterio simile per comparare il loro schema Neural-Fuzzy con una carta di controllo sulla media, e una CUSUM. Chang e Aw (1996), hanno dimostrato che il loro sistema presenta ARL più bassi rispetto ai metodi SPC tradizionali, e dunque è in grado di poter identificare più velocemente eventuali modifiche dello stato del processo.

Un aspetto che accomuna la maggior parte dei lavori presentati in letteratura è l'utilizzo di simulazione Monte Carlo, essa è usata per la generazione dei run sui quali effettuare le prove di valutazione, e di comparazione, dei sistemi proposti. Tuttavia, una scarsa attenzione si è rilevata in quasi tutti i lavori presentati in letteratura sull'indicazione delle scelte sperimentali effettuate. Queste comprendono l'indicazione del tipo di generatore di numeri casuali utilizzato per la simulazione Monte Carlo, nonché l'indicazione del seme adottato per la generazione dei diversi run. Simili indicazioni permetterebbero più facilmente una riproduzione dei risultati presentati nei differenti lavori. Solo Cheng (1995, 1997, 2001) accenna alla necessità di un cambio del seme nella generazione delle sequenze di dati utilizzate in fase di training e in fase di testing. Questa condizione è indispensabile per garantire una corretta valutazione della capacità di generalizzazione della rete neurale.

Un altro limite rilevabile sulla letteratura proposta riguarda la scarsa attenzione all'impiego di tecniche statistiche per il confronto delle prestazioni. Le misure di prestazione, infatti, sono ottenute da simulazioni al calcolatore, pertanto esse rappresentano delle stime degli effettivi valori che caratterizzano il sistema di controllo nella specifica condizione operativa simulata. I risultati pubblicati sono quindi caratterizzati da un errore di stima che dovrebbe essere trattato con metodi statistici. Per esempio, sarebbe opportuno riportare stimatori intervallari della misura di prestazione (piuttosto che uno stimatore puntuale), ed utilizzare tecniche statistiche per il confronto delle prestazioni fra i differenti sistemi.

Alcuni ricercatori (Cheng 1995, 2001, Guh e Tannock 1999a 1999b) riportano la deviazione standard degli stimatori puntuali utilizzati per valutare le prestazioni. Essa

fornisce un'idea della variabilità riscontrata sullo stimatore puntuale tra i differenti run di simulazione. Guo e Dooley (1992), con lo scopo di verificare la significatività statistica delle misure ottenute dalla simulazione, e per confrontare le prestazioni di una rete neurale con metodi SPC, propongono l'utilizzo di un test statistico basato sull'uso della *t di Student* al fine di verificare l'ipotesi di effettiva differenza fra le prestazioni ottenute con quelle di riferimento. In particolare gli autori hanno misurato le prestazioni in termini di tasso di errore, ovvero il numero di classificazioni errate in run di 1000 dati, per una specifica categoria di andamento non naturale (ovvero una specifica ampiezza di cambiamento per la media e la varianza). Considerando i valori di riferimento ottenibili da un metodo SPC, in particolare la carta di controllo per misure individuali a range mobile, l'impiego del test statistico basato sulla *t di Student* ha consentito di verificare la significatività del miglioramento delle prestazioni di controllo ottenute dalla rete neurale.

2.6.1 Misure speciali di valutazione delle prestazioni

Hwarng e Hubele (1993) suggeriscono che mentre la misura ARL, o analogamente gli errori di Tipo I e di Tipo II, rappresentano delle buone misure per le prestazioni di un sistema di monitoraggio SPC, esse possono essere insufficienti per valutare le prestazioni di un sistema neurale di riconoscimento dei pattern sulle carte di controllo. Hwarng e Hubele (1993) giustificano una simile affermazione indicando che, mentre una carta di controllo è diretta a segnalare semplicemente la presenza di un fuori controllo, un algoritmo su base neurale deve essere in grado di fornire informazioni sul tipo di pattern non naturale che è stato identificato sui dati, nonché, in alcuni casi, fornire informazioni sulle caratteristiche del tipo di pattern che è stato identificato.

In particolare secondo Hwarng e Hubele (1993), l'indice ARL è inappropriato per la misura delle prestazioni di un sistema di riconoscimento dei pattern sulle carte di controllo fondamentalmente per due motivi.

1. Non fornisce alcun'indicazione sulla corretta rilevazione del tipo *di pattern* la cui presenza ha generato la situazione di fuori controllo. In altri termini ARL non contiene alcun'informazione circa la capacità di discriminazione della rete neurale.
2. Il sistema di riconoscimento è stato progettato con l'obiettivo di classificare differenti pattern simultaneamente e non una generica situazione di "fuori controllo".

Da ciò segue la necessità di definire delle nuove misure. In particolare Hwarng e Hubele (1993) hanno proposto una serie di diversi indici per valutare le prestazioni di un algoritmo di riconoscimento dei pattern.

ROT (Rate Of Target) caratterizza quanto frequentemente la rete neurale è capace di identificare un pattern target. Quest'indice quantifica il potere discriminante di un algoritmo di riconoscimento dei pattern definendo la frequenza con cui il *pattern target* è identificato come primo pattern non naturale nella sequenza di dati.

Dal momento che è possibile identificare diversi altri pattern, prima che quello target sia riconosciuto, l'informazione riguardante il primo riconoscimento (anche se diverso dal target) è riportata nell'indice FDP (First Detected Pattern). L'obiettivo, ovviamente, è di avere come primo target riconosciuto proprio il target pattern (FDTP) ma questa non è sempre una situazione possibile dal momento che molti pattern potrebbero assomigliare ad altri in differenti aspetti. Il numero di volte percentuali in cui, durante la simulazione, il target pattern coincide con il FDT è il ROT. Il ROT assume un valore compreso fra 0 e 1 e quanto più alto è il valore assunto, tanto migliori sono le prestazioni del sistema di riconoscimento.

Analogamente il RONT (Rate Of Non Target) identifica la frequenza con cui la rete neurale identifica pattern non corrispondenti al target. Da notare, tuttavia, che il termine “non-target” è differente dal termine “falso allarme”. Non-target pattern si riferisce, infatti, a qualsiasi altro pattern diverso dal target. Il fatto che la rete identifichi un pattern differente non significa che la rete identifica un pattern incorrettamente (si pensi ad un pattern ciclo che potrebbe essere confuso con un pattern mixture).

ATPRL (Average Target Pattern Run Length), caratterizza la velocità attraverso cui la rete neurale è in grado di identificare il target in una sequenza di dati. ATPRL è definito come il numero medio di osservazioni richiesto prima che FDP sia identificato dalla rete neurale. Ovviamente ATPRL non è una misura esatta del ARL perché i run nei quali FDP non è il target sono esclusi.

Conseguentemente un valore relativo, e più ragionevole, del ARL è dato dal ARLIDX (Average Run Length InDeX) definito come il rapporto fra ATPRL/ROT. ARLIDX fornisce una misura migliore ai fini di un confronto relativo.

Un ulteriore indice è rappresentato da AMPRL (Average Multiple Pattern Run Length) ovvero il numero medio di osservazioni richieste prima che qualunque pattern (uguale o diverso dal target) sia identificato. Si tratta di una misura molto interessante quando il processo è in controllo (simulando il riconoscimento su pattern naturali). In questo caso (processo in controllo) un ampio valore di AMPRL corrisponde ad un basso tasso di falsi allarmi.

Altri autori estendono il set di indici proposti in precedenza in funzione di specifici problemi. Ad esempio Guh e Tannock (1999b) propongono la costruzione di una rete neurale per il riconoscimento multiplo di pattern utilizzando altri indici specifici quali:

ABPRL (Average Both Pattern Run Length), ovvero il numero medio di campioni richiesti prima che due pattern sovrapposti siano identificati. ABTPRL (Average Both Target Pattern Run Length), il numero medio per identificare correttamente entrambi i pattern target quando questi sono anche i primi pattern identificati. ROBT (Rate Of Both Target), ovvero la frequenza con cui i pattern target sono per primi correttamente identificati. ABRLIDX (Average Both Run Length Index) il rapporto tra ABTPRL su ROBT.

In generale si può affermare, come riportato dalla Tabella 2.6 sottostante, che mentre i sistemi proposti per identificare cambiamenti di struttura nello stato di un processo utilizzano i tradizionali criteri di misura SPC, i sistemi proposti quali riconoscitori di pattern, utilizzano una serie particolare di misure ideate per valutare le capacità di riconoscimento.

2.6.2 Confronti fra i differenti lavori

Per quanto riguarda le comparazioni fra i lavori di diversi ricercatori, si può evidenziare che solo Cheng (1997) confronta i risultati ottenuti dal suo sistema, con quelli di studi in precedenza pubblicati da altri ricercatori. In genere, infatti, in tutti gli altri lavori non è dato nessun accenno di confronto fra i risultati ottenuti con quelli precedentemente pubblicati. Per esempio, Al-Ghanim (1997) propone un confronto fra un metodo neurale non supervisionato basato sull'uso della rete ART1, con un comune metodo neurale supervisionato basato sull'uso di una rete MLP BP. Tuttavia, non è fatto riferimento a nessuno dei lavori precedentemente pubblicati pur esistendo un'estesa letteratura sull'utilizzo di reti neurali supervisionate basate su MLP BP per il monitoraggio di processo.

Questa ridotta capacità di confronto tra i differenti lavori proposti in letteratura nasce, tuttavia, dalla notevole difficoltà di replicare le sperimentazioni pubblicate nei differenti articoli che, come già evidenziato in precedenza, deriva dalla mancanza di

informazioni dettagliate sullo schema sperimentale implementato. Un'altra difficoltà deriva dalla mancanza di misure comuni di valutazione delle prestazioni tra i differenti lavori pubblicati che, in molti casi, rende difficile non solo il confronto fra le prestazioni, ma persino la stessa interpretazione dei dati pubblicati.

<i>Autore</i>	<i>Comparazione SPC</i>	<i>Criteri di Prestazioni</i>
Pugh (1989, 1991)	Carta di controllo X	ARL, Errori di Tipo I e di Tipo II
Guo e Dooley (1992)	Carta di controllo IMR (Misure individuali a range mobile)	Errore di Tipo II, t di student per confronto a coppie
Hwang e Hubele (1993a, 1993b)	Carta di controllo X	ARL modificato
Smith (1994)	Carta di controllo X-R	Percentuale di errore
Pham e Oztemel (1994)	Carta di controllo X	Percentuale di precisione
Cheng (1995)	Carta combinata Shewart CUSUM	ARL
Hwang e Chong (1995)	Carta di controllo X	ARL modificato
Chang e Aw (1996)	Carta di controllo X e CUSUM	ARL
Al-Ghanim (1997)	Nessuna	Errore di Tipo I, Percentuale di identificazione
Cheng (1997)	Nessuna	ARL modificato, Percentuale di identificazione
Cook e Chiu (1998)	Carta di controllo X	ARL, Percentuale di identificazione
Reddy <i>et al.</i> (1998)	Carta di controllo X-R	Percentuale di identificazione
Leger <i>et al.</i> (1998)	Carta di controllo CUSUM	Errore di Tipo I, Percentuale di identificazione
Chang e Ho (1999)	Carta di controllo R, CUSUM, EWMA	Percentuale di identificazione, ARL
Dedekayogullari e Burnak (1999)	Carta di controllo X-S, X-R	Errore di Tipo II
Guh <i>et al.</i> (1999)	Carta combinata Shewart CUSUM Carta CUSUM	ARL
Guh e Tannock (1999a, 1999b)	Nessuna	ARL modificato, Percentuale di identificazione
Guh e Hsieh (1999)	Nessuna	Percentuale di identificazione, Errore di Tipo II
West <i>et al.</i> (1999)	Carta di controllo X, EWMA	ARL
Kang e Park (2000)	Nessuna	Percentuale di identificazione
Cheng e Cheng (2001)	Carta CUSUM esponenziale	Errore di Tipo II
Cook <i>et al.</i> (2001)	Nessuna	Percentuale di identificazione
Perry <i>et al.</i> (2001)	Nessuna	Percentuale di identificazione
Pham e Sagiroglu (2001)	Nessuna	Percentuale di identificazione
Wang e Cheng (2002)	Carta multivariata T ²	Errore di Tipo I, Percentuale di identificazione

Tabella 2.6 – Classificazione della letteratura in base ai criteri di prestazione utilizzati

Capitolo 3

Basi teoriche e modelli di riferimento

In questo capitolo sono discusse in dettaglio le basi teoriche delle tecniche di monitoraggio statistico di processi manifatturieri, e sono presentati i modelli di riferimento per il processo utilizzato come caso di studio. L'obiettivo è duplice: da un lato proporre una breve sintesi delle tecniche di controllo che rientrano nell'ambito della definizione Statistical Process Control (SPC) e che saranno utilizzate per la definizione di benchmark di riferimento; dall'altro è di introdurre le modellazioni matematiche che costituiscono la base per le procedure di simulazione Monte Carlo, utilizzate per riprodurre processi di riferimento sia in condizioni naturali, sia non naturali.

3.1 Analisi di un processo

Un processo è definito in controllo statistico quando è possibile prevedere l'andamento futuro di determinate grandezze caratteristiche, sulla base di una serie di osservazioni realizzate nel passato. Le carte di controllo sono utilizzate quali strumenti per segnalare se un processo è in controllo o fuori controllo statistico (Montgomery 2000). Con questo scopo, una serie di campioni dei parametri del processo, utilizzati per il monitoraggio della qualità, sono raccolti nel tempo e utilizzati per fornire un'indicazione sul suo stato. Da questo punto in poi, nel presente lavoro, si considereranno processi caratterizzati da un *singolo* parametro di qualità ovvero da una misura mono-dimensionale.

Ciò che caratterizza il funzionamento di un sistema per il monitoraggio della qualità (sia carta di controllo, sia rete neurale) è la distinzione fra variazioni naturali e non naturali. Le prime sono connaturate al processo, mentre le seconde sono generate da fattori esterni che devono essere eliminati: un sistema di controllo dovrebbe riuscire a distinguere questi due tipi di variazione utilizzando la serie di osservazioni prodotte all'uscita del processo. L'obiettivo di una carta di controllo, così come di un algoritmo neurale, è di analizzare l'andamento temporale del parametro caratteristico della qualità del processo, per identificare cambiamenti del suo stato quanto più velocemente possibile, e senza generare troppi falsi allarmi. Il funzionamento di un sistema di monitoraggio della qualità, carta di controllo o rete neurale, può essere considerato come un test statistico ripetuto. L'ipotesi nulla H_0 e l'ipotesi alternativa H_1 possono essere formulate come segue.

H_0 : il processo è in uno stato naturale;

H_1 : il processo è in uno stato non naturale;

Come per qualunque test statistico, possono essere commessi errori di Tipo I e di Tipo II. Gli errori di Tipo I sono chiamati *falsi allarmi*. Gli errori possono essere interpretati rispettivamente come segue.

Errore di Tipo I: sono adottate delle azioni correttive sebbene il processo sia in uno stato naturale (in controllo statistico).

Errore di Tipo II: nessun'azione correttiva è adottata sebbene il processo sia un uno stato non naturale (fuori controllo statistico).

Al fine di valutare le prestazioni del sistema di monitoraggio della qualità basato su un algoritmo neurale, e per confrontarle con quelle di un equivalente metodo SPC, è necessario stabilire il modello del processo. Tale modello servirà sia per riprodurre l'andamento naturale del processo (ovvero una sequenza di dati in controllo), sia un insieme di andamenti non naturali dello stesso (ovvero una sequenza di dati fuori controllo). Nella parte restante del paragrafo si esaminerà in dettaglio il modello del processo utilizzato nelle simulazioni descritte nei capitoli successivi. In particolare, il modello del processo sarà riprodotto utilizzando il generatore di numeri pseudo-casuali dell'ambiente MATLAB (Vattulainen *et al.* 1995). Nell'appendice B sono riportati alcuni dettagli di configurazione del generatore di numeri pseudo-casuali di MATLAB.

3.1.1 Il modello del processo

Gli andamenti non naturali delle osservazioni utilizzate per il controllo della qualità, possono essere modellati secondo uno schema *lineare*, in altre parole sovrapponendo una sequenza di punti generati dal processo in uno stato naturale ad una sequenza di punti di disturbo generati da una particolare funzione del tempo chiamata *pattern*. La sequenza di punti generati dal processo in uno stato naturale è simulata attraverso l'uso di un modello probabilistico. Il pattern, vale a dire la sequenza di punti che si sovrappone alle osservazioni per effetto di cause speciali, è considerato una serie temporale (deterministica o aleatoria) caratterizzata da un insieme di predefiniti parametri. Quest'approccio di modellazione lineare è quello più diffusamente utilizzato in letteratura (Hwarng e Hubele 1993, Al-Ghanim 1997, Cheng 1997).

Il modello probabilistico, che descrive l'andamento naturale delle osservazioni del processo, è caratterizzato da una media nominale, più una parte di variazione aleatoria. Le variazioni naturali del processo sono realisticamente modellate come una sequenza di osservazioni identicamente distribuite in modo normale attorno ad una media (che supponiamo nota e indichiamo con μ) e con una predefinita deviazione standard (che supponiamo anche essa nota e indichiamo con σ) che modella sia la variabilità naturale inerente al processo, sia quella legata agli strumenti di misura utilizzati per monitorarne l'uscita.

Sia t l'indice temporale che rappresenta valori di tempo discreto ai quali delle osservazioni sono campionate all'uscita del processo: $t = 1, 2, \dots$. Definiamo una serie di dati $\{X_t\}$ un insieme di osservazioni ordinate nel tempo. Si farà riferimento, unicamente, a serie di tipo discreto e si supporrà che le osservazioni siano tratte ad uguali intervalli di tempo. Una serie temporale di dati $\{X_t\}$ può essere considerata come un insieme di osservazioni di qualche processo: deterministico o stocastico.

Indichiamo con $\{X_t\}$ la serie di dati derivante dalla componente naturale di variazione del processo, con $\{S_t\}$ la serie di dati caratterizzante il pattern speciale di variazione e con $\{Y_t\}$ la serie di dati all'uscita del processo che consideriamo standardizzata prima che sia analizzata dal sistema di controllo. Allora valgono le seguenti relazioni per $\forall t = 1, 2, \dots$

$$\begin{aligned} X_t &= \mu + \varepsilon_t = \mu + \sigma n_t & n_t &\approx NID(0,1) \\ S_t &= \varphi \sigma n_t \\ Y_t &= \frac{X_t + S_t - \mu}{\sigma} \end{aligned} \tag{3.1}$$

Dove, in particolare è stato indicato con φ l'ampiezza del segnale speciale di disturbo (che consideriamo costante nel tempo) espressa in unità di deviazione standard, $\{n_t\}$ è una serie temporale di osservazioni indipendenti e identicamente distribuite in modo normale di media 0 e deviazione standard 1. Le relazioni riportate in equazione (3.1) possono essere schematizzate nel seguente diagramma a blocchi.

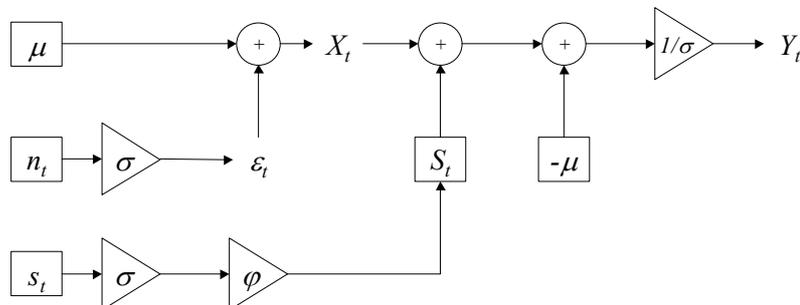


Figura 3.1 – Modello generale di simulazione del processo

Introducendo la serie temporale $\{Z_t\}$ come la serie temporale standardizzata delle uscite del processo dovute alla componente naturale di variazione, e ridefinendo la serie $\{S_t\}$ legata alla componente non naturale, allora si può semplificare il precedente modello nella seguente forma equivalente.

$$\begin{aligned}
 X_t &= \mu + \varepsilon_t = \mu + \sigma n_t & n_t &\approx NID(0,1) \\
 Z_t &= \frac{X_t - \mu}{\sigma}; S_t = \varphi s_t & & \\
 Y_t &= Z_t + S_t & &
 \end{aligned}
 \tag{3.2}$$

Il precedente modello può essere poi ulteriormente semplificato nello schema seguente che, da questo punto in poi, sarà considerato il modello di riferimento per un processo privo di memoria (ovvero con serie di uscite indipendenti statisticamente).

$$\begin{aligned}
 n_t &\approx NID(0,1); Z_t = n_t; S_t = \varphi s_t & & \\
 Y_t &= Z_t + S_t & &
 \end{aligned}
 \tag{3.3}$$

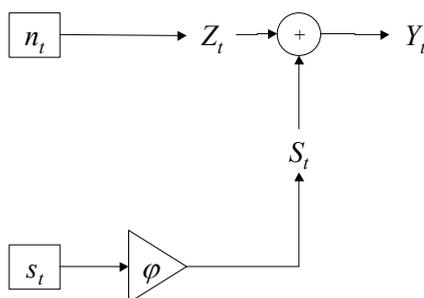


Figura 3.2 – Modello semplificato di simulazione del processo

Conseguentemente, l'ipotesi nulla può essere formulata all'istante t come segue:

$$H_0 : Y_t = Z_t$$

e l'ipotesi alternativa come

$$H_1 : Y_t = Z_t + S_t$$

3.2 Modelli di pattern

Allo scopo di simulare il processo in uno stato non naturale, di seguito sono descritti alcuni modelli particolari per la serie di dati $\{s_t\}$. Tali modelli sono quelli più diffusamente trattati in letteratura, e saranno utilizzati anche in questo testo per la costruzione dei casi di studio.

3.2.1 Variazione sistematica

Un pattern di variazione sistematica (systematic variation) si manifesta su una carta di controllo quando una sequenza di n punti posti al di sopra della linea centrale è sempre seguita da una sequenza di m punti posti al di sotto della linea centrale (e viceversa). Generalmente, in questi pattern le osservazioni sulla carta di controllo si posizionano in maniera quasi simmetrica rispetto alla linea centrale. Il caso più semplice, quello considerato nel seguito, è quello in cui $n = m = 1$, ovvero il caso in cui il pattern è costituito da punti alternanti sopra e sotto il livello di media.

Per quanto riguarda le cause di tale pattern si possono citare: differenze fra diverse linee di produzione, dalle quali, in rotazione, sono osservati i campioni, oppure differenze tra turni di produzione, o ancora differenze tra campioni o misure su gli stessi, ed altre ancora. La variazione sistematica è stata modellata come segue:

$$s_t = (-1)^t \quad \forall t = 1, 2, \dots \quad (3.4)$$

Di seguito, è rappresentato l'effetto di un pattern di variazione sistematica su una carta di controllo (ampiezza pari a due unità di deviazione standard).

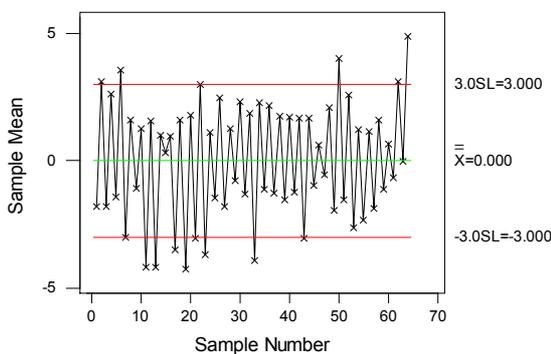


Figura 3.3 – Pattern variazione sistematica

3.2.2 Ciclo

Un ciclo, è un pattern che presenta aumenti e diminuzioni della media del processo nel tempo che si alternano in maniera periodica e regolare. Questo tipo di pattern deriva dalla presenza di condizioni di alternanza, ad esempio turni degli operatori, variazioni sistematiche ambientali, fluttuazioni del voltaggio o della pressione o qualunque altra variazione di tipo periodico. Andamenti ciclici della media di processo possono essere riconosciuti da una serie di picchi separati da una sequenza d'avvallamenti. Il ciclo è stato modellato come segue.

$$s_t = \cos\left(\frac{2\pi t}{\Omega}\right) \quad \forall t = 1, 2, \dots \quad (3.5)$$

dove Ω rappresenta il periodo del ciclo. La seguente figura riporta un pattern di tipo ciclo con un valore di ampiezza pari a 2 unità di deviazione standard e un periodo pari a $\Omega = 12$.

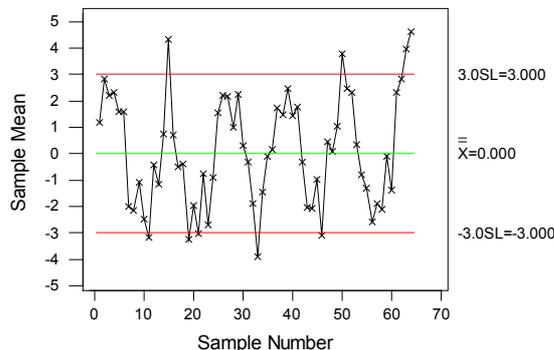


Figura 3.4 – Pattern ciclo

Si osservi in particolare che se $\Omega = 2$, allora un pattern di tipo ciclo è equivalente ad un pattern di variazione sistematica.

3.2.3 Shift

Il pattern shift si ha quando la variazione della media del processo si mantiene costante nel tempo. In genere questo tipo di pattern può essere causato da una brusca variazione dei parametri influenzanti il processo, ad esempio una modifica dei parametri macchina (si pensi ai pre-settaggi di una macchina utensile), una differenza tra i tipi di materiali utilizzati, l'introduzione di un nuovo operatore macchina, l'adozione di diversi materiali, cambiamenti nei metodi d'ispezione e altri ancora. Uno shift può essere definito come una variazione istantanea o improvvisa della media di processo ed è stato modellato come segue.

$$s_t = \begin{cases} 0 & t < \tau \\ 1 & t \geq \tau \end{cases} \quad \forall t = 1, 2, \dots \quad (3.6)$$

dove τ ($\tau \geq 1$) rappresenta l'indice temporale di applicazione dello shift. La seguente figura rappresenta uno shift ($\tau = 16$) di ampiezza pari a 2 unità di deviazione standard.

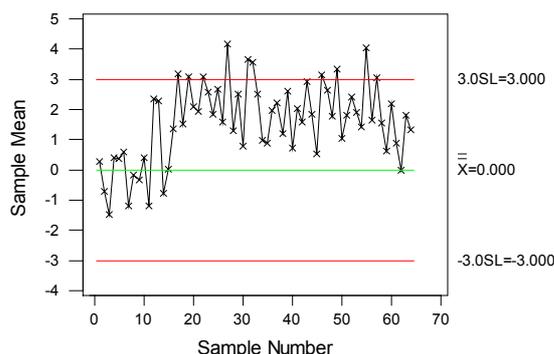


Figura 3.5 – Pattern shift

3.2.4 Mixture

Una mixture si ottiene quando i valori osservati per ogni singolo campione derivano da distribuzioni differenti con medie opposte, tipicamente l'una superiore e l'altra inferiore alla linea centrale della carta di controllo. In una carta di controllo, un pattern mixture si manifesta quando i punti tendono ad accumularsi vicino il limite

superiore o inferiore con assenza di variazioni vicine al valore medio. Le cause più comuni derivano da diverse sorgenti da cui sono campionati i dati: ad esempio differenti fornitori di un prodotto, differenti macchine di lavorazione o linee di assemblaggio etc. Altre cause possono essere legate a continui presetting delle macchine eseguiti dagli operatori per rispondere a condizioni di fuori controllo. Un pattern mixture può essere modellato come segue.

$$s_t = m_t; m_t \in \{-1,1\} \quad \forall t = 1,2,\dots \quad (3.7)$$

Dove m_t è lo stato di una catena di Markov binaria a valori varianti nell'insieme $\{-1,1\}$ (dove non diversamente indicato si supponrà lo stato iniziale $m_1 = 1$). Nel resto del lavoro, inoltre, si indicherà con $p_c = P\{m_t \neq m_{t-1} | m_{t-1}\}$ la probabilità che all'istante di indice $t > 1$ lo stato della catena di Markov cambi, e si supponrà che $P\{m_t = m_{t-1} | m_{t-1}\} = 1 - p_c$ sia la probabilità che nell'istante di tempo $t > 1$ lo stato della catena di Markov non cambi. La seguente figura riporta un pattern di tipo mixture ($p_c = 0.4$) generato con ampiezza paria a 2 unità di deviazione standard.

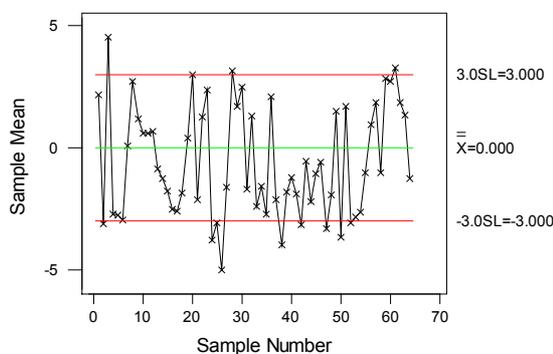


Figura 3.6 – Pattern mixture

Si osservi in particolare che

- Se $p_c = 1$ e $m_1 = -1$, allora il pattern mixture è equivalente ad un pattern di variazione sistematica.
- Se $p_c = 0$ e $m_1 = 1$, allora il pattern mixture è equivalente ad un pattern shift ($\tau = 1$).

3.2.5 Trend

Un altro tipo di pattern che comunemente è considerato in letteratura è il pattern di tipo trend: esso è caratterizzato da un graduale e continuo aumento, o diminuzione, della media del processo. Le tipiche cause per tale tipo di pattern, in processi manifatturieri di tipo meccanico, possono essere: usura utensile, fatica dell'operatore, deterioramento dell'attrezzatura di lavoro etc. Diverse procedure sono state considerate per tenere "sotto controllo" tale causa. Ad esempio, se il tasso di usura dell'utensile è noto o, quantomeno, può essere identificato, è possibile definire dei valori limite al trend oltre i quali risulta necessario sostituire l'utensile (Montgomery 2000).

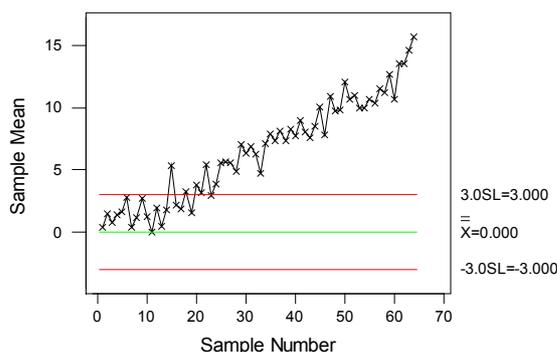


Figura 3.7 – Pattern trend

Una possibile modellazione del pattern trend può essere fornita dalla seguente relazione.

$$s_t = s \cdot t \quad \forall t = 1, 2, \dots \quad (3.8)$$

Dove in particolare, è stato indicato con s (*slope*) il tasso di variazione della media per istante di osservazione. In pratica, uno *slope* di valore $0 < s \leq 1$ implica che la media del processo è variata di una unità di deviazione standard (in aumento o in diminuzione a seconda del segno di s) dopo ogni gruppo di circa $\frac{1}{s}$ osservazioni.

3.3 La carta di controllo Shewhart

In precedenza è stata discussa la necessità di identificare la presenza di cause non naturali di variazione, e la necessità di disporre di strumenti che permettano di discriminare tali variazioni da quelle naturali. Il più semplice tra tali strumenti, e sicuramente quello più diffusamente utilizzato, è la carta di controllo di Shewarth. Essa è basata sull'idea che se un processo è in uno stato di controllo statistico allora le relative uscite risultano prevedibili sulla base di un opportuno modello probabilistico.

Utilizzando un insieme di osservazioni è possibile stabilire dei limiti alla variabilità, e determinare la probabilità che le future osservazioni rientrino o meno fra questi limiti. Un punto fuori tali limiti è considerato un segnale di allarme ed è indicativo del fatto che cause diverse, da quelle ritenibili naturali, stanno influenzando il processo. Tuttavia, proprio per la natura aleatoria delle osservazioni esiste una minima probabilità che un segnale di fuori controllo sia emesso mentre il processo è statisticamente in controllo: in tal caso si parla di un falso allarme. Una carta di controllo deve essere sensibile sufficientemente per identificare l'influenza sulla variabilità del processo di cause assegnabili, ma deve essere anche in grado di non generare eccessivi falsi allarmi. In pratica, l'ampiezza dei limiti posti sulla carta dovrà essere decisa opportunamente per bilanciare i due effetti.

Nella seguente sezione sono fornite le basi teoriche per la costruzione di una carta di controllo di Shewarth che sarà, poi, applicata al modello di processo riportato in precedenza.

3.3.1 Costruzione della carta di controllo

Sia dato un certo processo e supponiamo di considerare k campioni di n osservazioni, denotiamo la j^{esima} osservazione dell' i^{esimo} campione con X_{ij} $i = 1 \dots k; j = 1 \dots n$. Supponiamo che tutte le osservazioni siano identicamente e indipendentemente distribuite all'interno del campione i con la funzione F_i . Inoltre

assumiamo che le osservazioni in differenti campioni siano indipendenti. Si può osservare che, le funzioni di distribuzione F_1, F_2, \dots, F_k sono utilizzate per modellare l'effetto di cause naturali di variazione del processo. Si assume cioè che all'interno di ogni campione la distribuzione non cambi, ma si assume che essa possa cambiare da campione a campione. Il controllo del processo per l'identificazione di cause assegnabili può essere formalizzato dal seguente test di ipotesi:

$$H_0 : F_1 = F_2 = \dots = F_k \equiv F_0$$

contro l'ipotesi alternativa

$$H_1 : \exists p, q \in \{1, 2, \dots, k\} \ni F_p \neq F_q$$

In molti casi, si assume che le distribuzioni F_1, F_2, \dots, F_k siano note eccetto che per un gruppo di parametri (come per esempio la media e la varianza attesa). Le carte di controllo sono, in genere, costruite per ognuno di questi parametri. Un segnale di fuori controllo in uno di questi diagrammi è il segnale della presenza di una causa speciale di variazione.

Consideriamo una carta di controllo per distribuzioni dipendenti da un unico parametro $F_i(\eta_i)$. Sia $T_i = T(X_{i1}, \dots, X_{in})$ uno stimatore del parametro η_i basato sul campione i^{esimo} . Inoltre siano M_k e V_k delle statistiche costruite sulla base dei k campioni dove $M_k = M(X_{11}, \dots, X_{1n}, \dots, X_{k1}, \dots, X_{kn})$ è uno stimatore consistente della media di T_i sotto l'ipotesi H_0 e $V_k = V(X_{11}, \dots, X_{1n}, \dots, X_{k1}, \dots, X_{kn})$ è uno stimatore consistente di T_i sotto l'ipotesi H_0 . In una carta di controllo, le realizzazioni di T_i sono rappresentate rispetto all'indice i . Questi valori sono poi confrontati con i limiti di controllo nella forma:

$$L = M_k + c(n, k, p)V_k \quad (3.9)$$

Dove $c(n, k, p)$ è il p^{esimo} quantile della distribuzione $\frac{T_i - M_k}{V_k}$. Per i due limiti di controllo inferiore e superiore (Lower Control Limit – LCL e Upper Control Limit – UCL) della carta di controllo, due differenti costanti $c(n, k, p_{LCL})$ e $c(n, k, p_{UCL})$ devono essere rispettivamente determinate. Nelle situazioni dove si suppone siano noti i valori di media e/o di varianza della distribuzione T_i , allora questi devono essere utilizzati invece degli stimatori M_k e V_k e in tal caso $c(n, p)$ risulterà essere indipendente da k .

Solitamente nell'utilizzo della carta di controllo di Shewhart si assume che F_i sia normalmente distribuita con media μ_i e varianza σ_i^2 . La media e/o la varianza delle osservazioni possono cambiare nel tempo per la presenza di cause assegnabili di variazione. Con queste ipotesi il processo è in controllo se e solo se esistono due valori di media e varianza tali che:

$$\begin{cases} \mu_i = \mu \\ \sigma_i = \sigma \end{cases} \quad \forall i = 1 \dots k; \quad (3.10)$$

Questa è la ragione per la quale, in molti casi, un processo produttivo è monitorato utilizzando due carte di controllo, una per la media del processo e l'altra per la deviazione standard.

Per semplicità, e sulla base di motivazioni economiche, Shewhart propone di utilizzare: $c(n, k, p_{LCL}) = -3$ e $c(n, k, p_{UCL}) = 3$. Questi valori si dimostrano utili nella maggior parte dei casi pratici. Inoltre, assumendo di testare l'ipotesi nulla

$H_0 : \mu_1 = \mu_2 = \dots = \mu_k \equiv \mu$ nel caso di varianza nota e costante σ^2 , allora assumendo una media campionaria $T_i = \frac{X_{i1} + \dots + X_{in}}{n}$ allora i limiti della carta di controllo saranno rispettivamente: $LCL = \frac{\mu - 3\sigma}{\sqrt{n}}$ e $UCL = \frac{\mu + 3\sigma}{\sqrt{n}}$. Questo implica avere un valore $p_{LCL} = (1 - p_{UCL}) = 0.00135$.

3.3.2 Carta di controllo di Shewhart per il modello di riferimento

Con riferimento al modello di processo descritto dall'equazione (3.3 e rappresentato dalla Figura 3.2, la costruzione della carta di controllo è notevolmente semplificata. Dalle ipotesi del modello del processo in uno stato naturale (ovvero ipotizzando vera l'ipotesi H_0), si ricava che $E[Y_t] = 0, VAR[Y_t] = 1$ per ogni fissato istante di tempo t . Inoltre, ipotizzando vera l'ipotesi H_0 , si ricava che le osservazioni $\{Y_t\}$ sono indipendenti e identicamente distribuite in modo normale. La conoscenza di tali parametri consente di impostare la carta senza necessità della fase preparatoria di stima dei parametri del processo. Analizzando, inoltre, risposte singole risulta $n = 1$. Con queste ipotesi i limiti di controllo della carta risulteranno i seguenti

$$UCL = E[Y_t] + c(1, p_{UCL}) \cdot \sqrt{VAR[Y_t]} = c(1, p_{UCL}) \quad (3.11)$$

$$LCL = E[Y_t] + c(1, p_{LCL}) \cdot \sqrt{VAR[Y_t]} = c(1, p_{LCL}) \quad (3.12)$$

Posto $l > 0$ e imponendo $c(1, p_{UCL}) = -c(1, p_{LCL}) = l$, ipotizzando di fissare un opportuno livello di lunghezza media dei run prima che sia segnalato un fuori controllo (Average Run Length – ARL) quando il processo è in uno stato naturale, la progettazione dei limiti è semplicemente ottenuta ponendo $l = \left| Z_{\frac{1}{2 \cdot ARL}} \right|$ ovvero uguale al

quantile $\frac{1}{2 \cdot ARL}$ della distribuzione standard normale.

3.4 La carta di controllo CUSUM

La carta di controllo di Shewhart utilizza sempre e solo l'ultimo campione per analizzare l'intero processo. Questo tipo di approccio è privo di memoria in quanto le precedenti osservazioni non influenzano la capacità di identificare in futuro un segnale di fuori controllo. Grandi variazioni nella media possono essere facilmente identificate attraverso una carta di controllo di Shewhart, tuttavia essa non è sensibile a piccoli cambiamenti.

Nelle carte di controllo di tipo EWMA (Exponentially Weighted Moving Average), per esempio, si introduce un effetto di memoria in quanto il processo è analizzato utilizzando una media pesata tra l'osservazione attuale con tutte quelle precedenti. Il peso associato alla più recente osservazione ha il valore più alto rispetto ai pesi associati alle osservazioni precedenti. La sequenza dei pesi utilizzati, inoltre, diminuisce esponenzialmente quanto più aumenta la distanza fra osservazione attuale e quelle che la precederono.

Analogamente, la carta di controllo CUSUM nella sua forma originaria (Montgomery 2000) introduce un effetto di memoria perché utilizza una somma non pesata di tutte le osservazioni precedenti. Definendo con μ_0 un livello medio target per il processo in analisi, è possibile rappresentare su un diagramma una somma cumulata delle osservazioni del tipo:

$$S_t = \sum_{i=0}^t (X_i - \mu_0) \quad (3.13)$$

Se la reale media del processo all'istante di tempo t è paria a $\mu_t > \mu_0$ allora l'insieme dei valori $\{S_t\}$ avrà un andamento crescente nel tempo. D'altro canto se risulta $\mu_t < \mu_0$ la sequenza dei valori $\{S_t\}$ avrà un andamento decrescente nel tempo. Rappresentando la somma cumulata su un diagramma temporale, i cambiamenti della media potranno essere visibili come trend di crescita o di decrescita delle osservazioni. Nel caso di test mono-laterale e supponendo di essere interessati a cambiamenti positivi della media di $\{X_t\}$, allora è possibile utilizzare il seguente criterio per la segnalazione di un allarme:

$$S_t - \min_{0 \leq i \leq t} S_i \geq h \quad (3.14)$$

fissato un certo valore della variabile h . Questo approccio è equivalente a monitorare la seguente grandezza:

$$S'_t = \max(S'_{t-1} + X_t, 0) \quad (3.15)$$

Un segnale di fuori controllo potrà essere emesso al tempo t se risulta $S'_t \geq h$. Nel caso, invece, di un test bilaterale, ovvero se si è interessati sia a variazioni positive della media sia a variazioni negative, un segnale di fuori controllo è dato se risulta verificata almeno una delle seguenti condizioni:

$$S_t - \min_{0 \leq i \leq t} S_i \geq h, \max_{0 \leq i \leq t} S_i - S_t \geq h' \quad (3.16)$$

Dove h e h' sono due predefinite costanti. La carta di controllo CUSUM è divenuta particolarmente popolare dopo la proposta di uno schema di analisi dei punti fuori controllo definito $V - mask$. Successivamente è stato dimostrato che l'uso della procedura $V - mask$ su una carta CUSUM è equivalente al monitoraggio di due grandezze:

$$SH_t = \max\{0; SH_{t-1} + Z_t - k\} \quad (3.17)$$

$$SL_t = \max\{0; SL_{t-1} - Z_t - k\} \quad (3.18)$$

con un limite orizzontale a distanza h dall'asse dei tempi. La variabile $Z_t = \frac{X_t - \mu_0}{\sigma_x}$

denota la variabile standardizzata della serie di osservazioni. Un segnale di allarme è generato se risulta $SH_t \geq h$ oppure se risulta $SL_t \geq h$. La costante k è chiamata valore di riferimento, mentre h è l'intervallo di decisione. La carta di controllo CUSUM, monitorata come indicato attraverso le grandezze SH_t e SL_t , è anche nota attraverso l'acronimo DIS (Decision Interval Schema). Le relazioni fra i parametri d e θ , di un approccio $V - mask$ e i parametri k e h , di un equivalente approccio DIS, sono riportati in letteratura (Montgomery 2000).

Quando si disegna una carta CUSUM con approccio di analisi DIS è conveniente disegnare entrambi gli accumulatori in un grafico. In genere, per fare ciò, si preferisce

rappresentare il valore dell'accumulatore inferiore cambiato di segno e confrontato con il valore di $-h$. I parametri k e h possono essere scelti per ottimizzare alcune misure di prestazioni d'identificazione. Valori tipici di k variano tra 0.25 e 1.5 in funzione della grandezza dello shift della media che si è interessati a identificare. Nella rimanente parte del lavoro si farà esclusivamente riferimento all'approccio DIS per la costruzione delle carte di controllo CUSUM.

3.4.1 Carta di controllo CUSUM per il modello di riferimento

Con riferimento al modello di processo descritto dall'equazione (3.3 e rappresentato dalla Figura 3.2, la costruzione della carta di controllo CUSUM risulta sensibilmente semplificata. Dal modello di processo si ricava che $E[Y_t] = 0, VAR[Y_t] = 1$, supponendo vera l'ipotesi H_0 . Pertanto il target di riferimento della carta di controllo si pone pari a $\mu_0 = 0$. Sempre considerando vera l'ipotesi H_0 , si ricava che le osservazioni $\{Y_t\}$ sono indipendenti e identicamente distribuite in modo normale. Conseguentemente, le due sommatorie dello schema DIS si riducono alle seguenti.

$$SH_t = \max\{0; SH_{t-1} + Y_t - k\} \quad (3.19)$$

$$SL_t = \max\{0; SL_{t-1} - Y_t - k\} \quad (3.20)$$

In particolare, per dati non correlati e normali, indipendenti e identicamente distribuiti, se è importante identificare una determinata ampiezza di variazione della media (φ in termini di unità di deviazione standard), allora una regola diffusamente utilizzata è di scegliere $k = \frac{\varphi}{2}$. Il valore di h può essere ottenuto in funzione del livello di ARL desiderato per lo stato di controllo statistico (Montgomery 2000).

3.5 Misure di prestazione

Un sistema di monitoraggio della qualità, come ad esempio una carta di controllo oppure un algoritmo neurale, è progettato con lo scopo di eseguire una discriminazione binaria fra classi di dati naturali e classi di dati non naturali. Pertanto, allo scopo di misurare le prestazioni di un sistema di controllo, dovranno essere esaminate due importanti caratteristiche.

La prima di queste è la capacità del sistema di controllo di modellare le cause naturali (non assegnabili) senza creare errori di Tipo I. Gli errori di Tipo I rappresentano falsi allarmi, generati erroneamente dal sistema di controllo su un processo in uno stato naturale. La misura degli errori di Tipo I è fornita dal parametro $\alpha = P\{H_1|H_0\}$, vale a dire il valore atteso della probabilità che il sistema di controllo emetta un allarme quando il sistema controllato è in uno stato naturale. Questa caratteristica può essere calcolata per mezzo di modelli statistici, oppure potrebbe essere stimata per mezzo della simulazione Monte Carlo.

Il sistema di controllo deve, inoltre, essere in grado di identificare velocemente condizioni non naturali di variazione nei dati di uscita di un processo. La misura degli errori di Tipo II è fornita dal parametro $\beta = P\{H_0|H_1\}$, ovvero il tasso atteso di segnali di non allarme, emessi dal sistema di controllo quando il processo monitorato presenta delle cause speciali di variazione. Anche questo parametro può essere valutato per

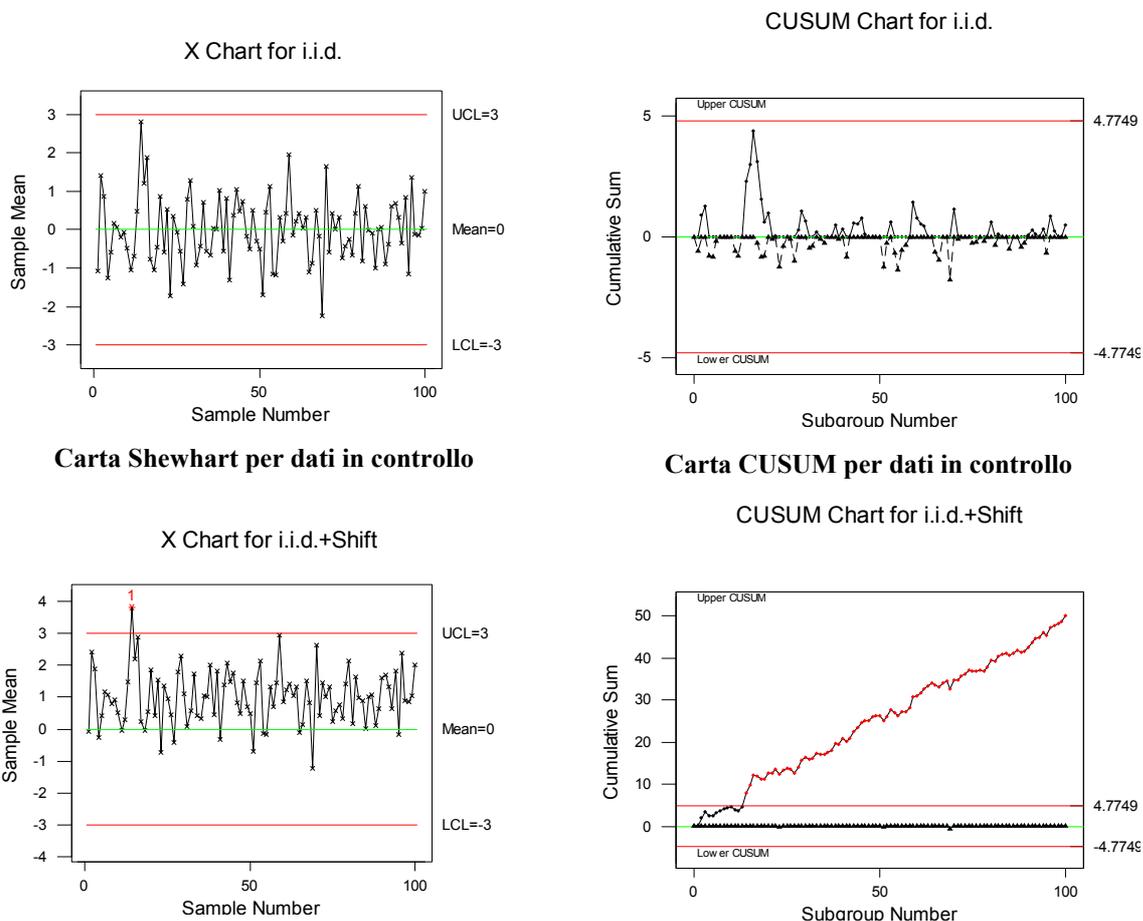
mezzo di modelli statistici, oppure può essere stimato attraverso l'uso della simulazione Monte Carlo. In ogni caso, la misura β , o gli stimatori puntuali e intervallari ottenuti dalla simulazione Monte Carlo, saranno dipendenti dalle ipotesi assunte nella modellazione degli effetti sulla uscita del processo della presenza di cause assegnabili di variazione.

Generalmente, l'obiettivo di un qualunque sistema di controllo è quello di identificare rapidamente i cambiamenti nello stato di un processo nel modo più rapido possibile (ovvero contenendo il tasso di errori di Tipo II), senza però provocare un numero eccessivamente elevato di falsi allarmi (ovvero contenendo il tasso di errori di Tipo I).

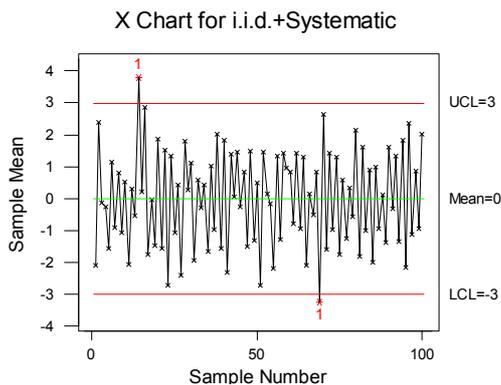
3.5.1 Esempi di applicazione delle carte di controllo

Prima di esaminare dettagliatamente le prestazioni delle carte di controllo al modello di riferimento, in questa sezione si esaminano alcuni semplici esempi di applicazione.

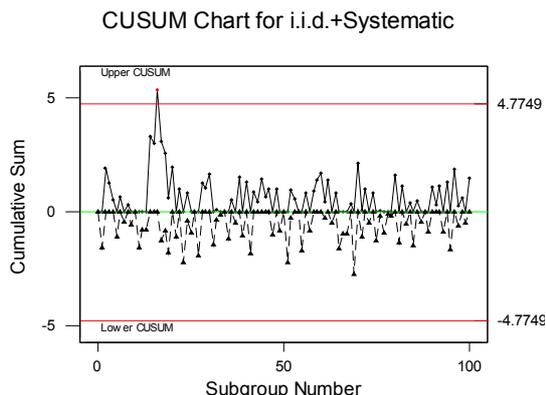
In particolare, nella Figura 3.8 sono state riportate le carte di controllo Shewhart e CUSUM per 100 osservazioni ($\{Y_t\}_{t=1,2,\dots,100}$) simulate dal modello di riferimento della equazione (3.3). La prima coppia di carte di controllo (Shewhart – CUSUM) si riferisce al caso di processo in uno stato naturale di variazione. Le successive quattro coppie si riferiscono ad altrettanti stati di variazione non naturale. Le situazioni di fuori controllo sono state simulate attraverso quattro specifici tipi di pattern (rispettivamente shift $\tau = 1$, variazione sistematica, ciclo $\Omega = 4$ e mixture $p_c = 0.4$) di ampiezza pari a 1 unità di deviazione standard ($\phi = 1$).



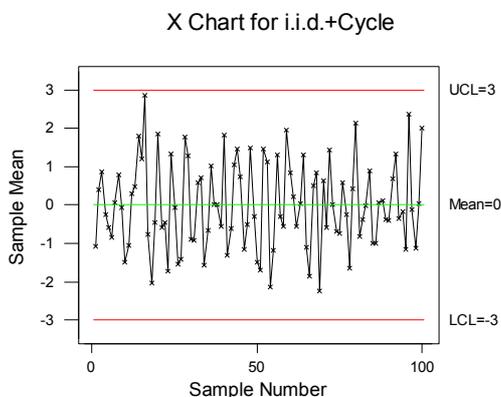
Carta Shewhart per shift



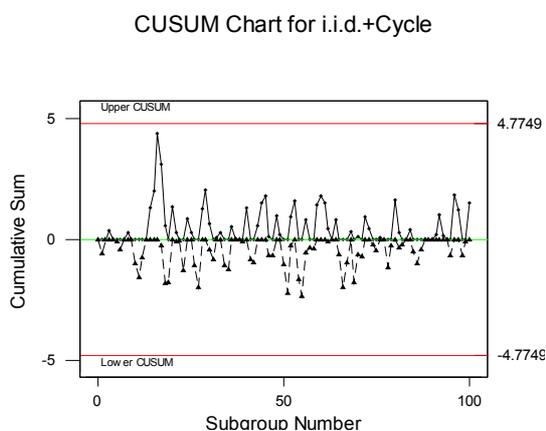
Carta CUSUM per shift



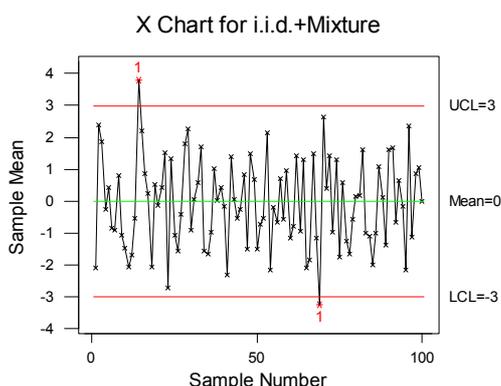
Carta Shewhart per variazione sistematica



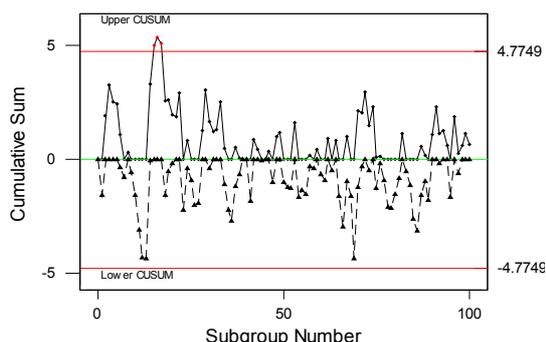
Carta CUSUM per variazione sistematica



Carta Shewhart per ciclo



CUSUM Chart for i.i.d.+Mixture



Carta Shewhart per mixture

Carta CUSUM per mixture

Figura 3.8 – Comparazione degli effetti di pattern non naturali sulle carte di controllo

La carta Shewhart è stata impostata con valore $l = 3$ (valore atteso ARL circa 370 equivalente ad un tasso di errori di Tipo I di circa 0.27%) mentre la carta CUSUM è stata impostata con $k = 0.5$ e $h = 4.7749$. Il valore $k = 0.5$ è stato scelto in modo da ottimizzare le prestazioni della carta CUSUM nell'identificare variazioni della media di 1 unità di deviazione standard. Il valore $h = 4.7749$ è stato scelto in modo da ottenere un valore di ARL in controllo pari a quello della carta di Shewhart (ARL di circa 370, errore di Tipo I di circa 0.27%).

Nel caso del processo in controllo, si può osservare che né la carta di Shewhart, né la carta CUSUM, segnala alcun allarme sulle 100 osservazioni realizzate. Nel caso di shift, invece, la carta di Shewhart segnala un solo allarme su 100 dati, mentre la carta CUSUM segnala un allarme per tutte le osservazioni successive alla 10°. L'utilizzo di un approccio *Fast Initial Response (FIR)* avrebbe comportato una rilevazione della situazione di fuori controllo anche a partire da campioni precedenti al 10° (Montgomery 2000).

La carta CUSUM, dunque, appare identificare efficacemente lo shift del processo, tuttavia essa non fornisce nessun ulteriore vantaggio, se confrontata con la carta di Shewhart, per gli altri tipi di disturbo. Infatti, per quanto riguarda gli ulteriori pattern di disturbo si può notare come entrambe le carte non presentano nessuna significativa capacità di discriminazione dalla condizione naturale di variabilità. Ciò evidenzia i limiti dei due metodi SPC nell'identificazione di generiche situazioni di fuori controllo.

Nelle sezioni seguenti saranno esaminate in dettaglio le prestazioni attese per una carta di controllo di Shewhart applicata al modello di riferimento.

3.5.2 Calcolo errori Tipo I e Tipo II per la carta Shewhart

Considerando il modello del processo di riferimento descritto dalla equazione (3.3), e ipotizzando uno stato naturale di variazione con osservazioni indipendenti, e identicamente distribuite secondo il modello statistico $n_t \approx NID(0,1)$, è possibile osservare che per ogni fissato istante di tempo $t = 1,2,\dots$, dato il segnale eventuale di disturbo S_t , allora anche la variabile aleatoria Y_t (l'uscita del processo all'istante di tempo t) risulta distribuita normalmente attorno ad un valore medio ($E[Y_t]$) con una varianza pari a 1 ($VAR[Y_t] = 1$).

In particolare, valgono le seguenti relazioni sui valori medi attesi in un generico istante di tempo $t = 1,2,\dots$

$$E[Y_t] = E[n_t] + E[S_t] = E[S_t] \tag{3.21}$$

$$E[Y_t] = 0 \quad t = 1,2,\dots \tag{3.22}$$

se il processo è in uno stato naturale di variazione

$$E[Y_t] = \varphi \quad t = \tau + i \quad i = 0,1,2,\dots \tag{3.23}$$

se il pattern non naturale è modellato da uno shift

$$E[Y_t] = \begin{cases} -\varphi & t = (2i + 1) \quad i = 0,1,2,\dots \\ \varphi & t = (2i) \quad i = 1,2,\dots \end{cases} \tag{3.24}$$

se il pattern non naturale è modellato da una variazione sistematica.

$$E[Y_t] = \begin{cases} 0 & t = (2i + 1) \\ -\varphi & t = (4i + 2) \\ \varphi & t = (4i + 4) \end{cases} \quad i = 0,1,2,\dots \tag{3.25}$$

se il pattern è un ciclo di periodo 4

$$E[Y_t] = \begin{cases} -\varphi & m_t = -1 \\ \varphi & m_t = 1 \end{cases} \quad t = 1, 2, \dots \quad (3.26)$$

se il pattern è modellato come una mixture

Sulla base delle precedenti relazioni, di seguito sono calcolati sia gli errori di Tipo I, sia gli errori di Tipo II attesi da una carta di controllo di Shewhart.

3.5.2.1 Errore di Tipo I

Nel caso in cui il processo è in controllo statistico, si avrà un allarme quando l'attuale osservazione supera i limiti di controllo della carta Shewhart. Il valore atteso dell'errore di Tipo I in un generico istante di tempo $t \geq 1$, può essere espresso in formula come segue.

$$\alpha_t = 1 - P\{-l \leq n_t \leq l\} \quad t = 1, 2, \dots \quad (3.27)$$

3.5.2.2 Errore di Tipo II: pattern shift

Supponiamo di considerare uno shift di ampiezza φ a partire dall'istante τ . L'errore di Tipo II, atteso per ogni istante di tempo $t \geq \tau$, è dipendente dall'ampiezza del parametro φ ed è espresso in formula dalla seguente relazione.

$$\begin{aligned} \beta_t(\varphi) &= P\{-l \leq n_t + \varphi \leq l\} & t = \tau + i; & \quad i = 0, 1, 2, \dots \\ \beta_t(\varphi) &= P\{-l - \varphi \leq n_t \leq l - \varphi\} & t = \tau + i; & \quad i = 0, 1, 2, \dots \end{aligned} \quad (3.28)$$

3.5.2.3 Errore di Tipo II: pattern variazione sistematica

Consideriamo un pattern variazione sistematica di ampiezza φ . L'errore di Tipo II atteso in un generico istante di tempo $t \geq 1$ può essere espresso in formula come segue.

$$\begin{aligned} \beta_t(\varphi) &= \begin{cases} P\{-l \leq n_t - \varphi \leq l\} & t = (2i + 1) & i = 0, 1, 2, \dots \\ P\{-l \leq n_t + \varphi \leq l\} & t = (2i) & i = 1, 2, \dots \end{cases} \\ \beta_t(\varphi) &= \begin{cases} P\{-l + \varphi \leq n_t \leq l + \varphi\} & t = (2i + 1) & i = 0, 1, 2, \dots \\ P\{-l - \varphi \leq n_t \leq l - \varphi\} & t = (2i) & i = 1, 2, \dots \end{cases} \\ \beta_t(\varphi) &= \begin{cases} P\{-l - \varphi \leq n_t \leq l - \varphi\} & t = (2i + 1) & i = 0, 1, 2, \dots \\ P\{-l - \varphi \leq n_t \leq l - \varphi\} & t = (2i) & i = 1, 2, \dots \end{cases} \\ \beta_t(\varphi) &= P\{-l - \varphi \leq n_t \leq l - \varphi\} & t = 1, 2, \dots \end{aligned} \quad (3.29)$$

3.5.2.4 Errore di Tipo II: pattern ciclo

Consideriamo un pattern Ciclo di ampiezza φ e periodo pari a 4. In tale caso, il valore atteso dell'errore di Tipo II dipenderà oltre che dall'ampiezza φ del disturbo anche dall'istante di osservazione. In particolare valgono la seguente formulazione per istanti di tempo $t \geq 1$.

$$\begin{aligned}
 \beta_i(\varphi) &= \begin{cases} P\{-l \leq n_t \leq l\} & t = (2i+1) \\ P\{-l \leq n_t - \varphi \leq l\} & t = (4i+2) \\ P\{-l \leq n_t + \varphi \leq l\} & t = (4i+4) \end{cases} \quad i = 0,1,2,\dots; \\
 \beta_i(\varphi) &= \begin{cases} P\{-l \leq n_t \leq l\} & t = (2i+1) \\ P\{-l + \varphi \leq n_t \leq l + \varphi\} & t = (4i+2) \\ P\{-l - \varphi \leq n_t \leq l - \varphi\} & t = (4i+4) \end{cases} \quad i = 0,1,2,\dots; \\
 \beta_i(\varphi) &= \begin{cases} P\{-l \leq n_t \leq l\} & t = (2i+1) \\ P\{-l - \varphi \leq n_t \leq l - \varphi\} & t = (4i+2) \\ P\{-l - \varphi \leq n_t \leq l - \varphi\} & t = (4i+4) \end{cases} \quad i = 0,1,2,\dots; \\
 \beta_i(\varphi) &= \begin{cases} P\{-l \leq n_t \leq l\} & t = (2i+1) \\ P\{-l - \varphi \leq n_t \leq l - \varphi\} & t = (2i) \end{cases} \quad i = 0,1,2,\dots;
 \end{aligned} \tag{3.30}$$

3.5.2.5 Errore di Tipo II: pattern mixture

Consideriamo un pattern mixture di ampiezza φ . Il valore atteso per l'errore di Tipo II in un generico istante di tempo $t \geq 1$ può essere espresso come somma di probabilità di eventi indipendenti.

$$\begin{aligned}
 \beta_t(\varphi) &= \left(P\{-l \leq n_t - \varphi \leq l, m_t = 1\} + P\{-l \leq n_t + \varphi \leq l, m_t = -1\} \right); & t = 1,2,\dots \\
 \beta_t(\varphi) &= \left(P\{-l \leq n_t - \varphi \leq l\}P\{m_t = 1\} + P\{-l \leq n_t + \varphi \leq l\}P\{m_t = -1\} \right); & t = 1,2,\dots \\
 \beta_t(\varphi) &= \left(P\{-l + \varphi \leq n_t \leq l + \varphi\}P\{m_t = 1\} + P\{-l - \varphi \leq n_t \leq l - \varphi\}P\{m_t = -1\} \right); & t = 1,2,\dots \\
 \beta_t(\varphi) &= \left(P\{-l - \varphi \leq n_t \leq l - \varphi\}P\{m_t = 1\} + P\{-l - \varphi \leq n_t \leq l - \varphi\}P\{m_t = -1\} \right); & t = 1,2,\dots \\
 \beta_t(\varphi) &= P\{-l - \varphi \leq n_t \leq l - \varphi\} \{P\{m_t = 1\} + P\{m_t = -1\}\}; & t = 1,2,\dots \\
 \beta_t(\varphi) &= P\{-l - \varphi \leq n_t \leq l - \varphi\} & t = 1,2,\dots
 \end{aligned} \tag{3.31}$$

3.5.3 Discussione

Per quanto riguarda l'errore di Tipo I di una carta di controllo di Shewhart è possibile osservare che esso è indipendente dall'istante di osservazione. Pertanto, l'errore di Tipo I per una carta di Shewhart è dato dalla seguente relazione.

$$\alpha = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \alpha_i = 1 - P\{-l \leq n_t \leq l\} \tag{3.32}$$

Per quanto riguarda il tasso di errori di Tipo II per pattern di tipo variazione sistematica e mixture, si può osservare che anch'essi risultano indipendenti dall'istante di osservazione. Risulta inoltre che per tutti gli istanti successivi a quello di avvio di un pattern shift (τ), ovvero per ogni $t \geq \tau$, anche il tasso atteso di errori di Tipo II per uno shift è equivalente a quello del pattern di variazione sistematica e di mixture, ed anch'esso è indipendente dall'istante di osservazione $t \geq \tau$. Per i pattern di tipo shift

(con $\tau = 1$), variazione sistematica e mixture, il tasso di errore di Tipo II per una ampiezza φ è pari alla seguente quantità.

$$\beta(\varphi) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \beta_i(\varphi) = P\{-l - \varphi \leq n_t \leq l - \varphi\} \tag{3.33}$$

Nel caso di pattern ciclo di periodo pari a 4, invece, il tasso di errore di Tipo II è dipendente dall'istante di osservazione. In generale il valore medio temporale del tasso di errori di Tipo II è pari alla seguente quantità.

$$\begin{aligned} \beta(\varphi) &= \lim_{t \rightarrow \infty} \frac{1}{t} \left(\sum_{i=0}^{\lfloor t/2 \rfloor} \beta_{2i+1}(\varphi) + \sum_{i=1}^{\lfloor t/2 \rfloor} \beta_{2i}(\varphi) \right) = \\ \beta(\varphi) &= \lim_{t \rightarrow \infty} \frac{1}{t} \left(\sum_{i=0}^{\lfloor t/2 \rfloor} P\{-l \leq n_t \leq l\} + \sum_{i=1}^{\lfloor t/2 \rfloor} P\{-l - \varphi \leq n_t \leq l - \varphi\} \right) = \\ \beta(\varphi) &= \frac{P\{-l \leq n_t \leq l\} + P\{-l - \varphi \leq n_t \leq l - \varphi\}}{2} \end{aligned} \tag{3.34}$$

Dove con la notazione $\lfloor \cdot \rfloor$, si è indicato il primo intero inferiore o uguale al suo argomento.

Il seguente grafico rappresenta l'andamento dell'errore atteso di Tipo II ottenibile da una carta di controllo Shewhart ($l = 3$) per i quattro tipi di pattern descritti in precedenza. L'andamento dell'errore di Tipo II è stato rappresentato in funzione dell'ampiezza del disturbo misurata in unità di deviazione standard (φ).

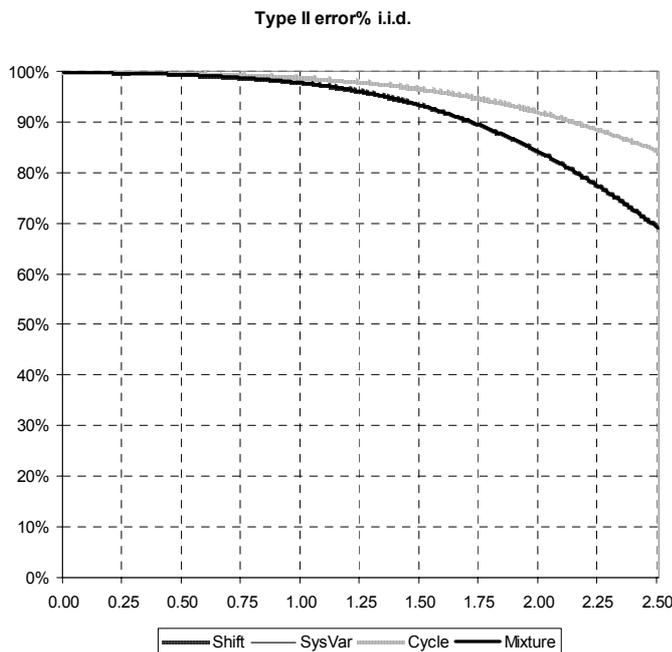


Figura 3.9 – Errore di Tipo II per carta di controllo Shewhart

Si può notare che, a parità di ampiezza φ , la carta di controllo presenta un più elevato errore di Tipo II nella segnalazione del pattern di tipo ciclo rispetto agli altri tipi di pattern. Si può osservare che, le prestazioni di riconoscimento ottenibili da una carta di controllo di Shewhart sono molto limitate. L'errore di Tipo II è, infatti, di circa 85% per il ciclo, e 70% per gli altri tipi di pattern, quando il disturbo è pari a due volte e mezzo la deviazione standard del processo.

3.6 Correlazione temporale dei dati

Nelle precedenti sezioni è stata formulata implicitamente l'ipotesi che le misure utilizzate per controllare il processo sono distribuite indipendentemente. Quest'ultima è una fra le ipotesi più comuni nella maggior parte della letteratura SPC. Tuttavia, essa non può essere giustificata in tanti casi pratici di controllo.

In letteratura si possono trovare alcune proposte per estendere il tradizionale criterio della carta di Shewhart al caso di processi le cui uscite sono dipendenti nel tempo (Box *et al.* 1994, Montgomery 2000). Tutti gli approcci sono basati su modelli matematici attraverso cui esprimere la dipendenza temporale delle informazioni. Ad esempio, i modelli di serie temporale noti come ARIMA di parametri (p, d, q) sono diffusamente utilizzati nei problemi di modellazione dell'auto-correlazione per serie temporali di dati.

Alcuni approcci presentati in letteratura, per adattare i metodi SPC al caso di dati temporalmente auto-correlati, mirano a modificare il calcolo dei limiti delle carte di controllo in funzione del desiderato valore di ARL in controllo e dei parametri del modello ARIMA utilizzato. Questo tipo di approccio, tuttavia, non sarà considerato nel presente lavoro; un altro approccio, che ha ricevuto molta attenzione in letteratura, sarà discusso in questa sezione. Esso è basato sull'idea che, un processo può essere controllato utilizzando invece dei dati direttamente osservati i *residui*, in altre parole gli errori di previsione fra i valori originali e quelli previsti dal modello di serie temporale. In genere se avviene un cambiamento nella media del processo controllato, questo si traduce in una corrispondente alterazione della media dei residui. Inoltre se il modello della serie di dati è adeguato allora i residui saranno approssimativamente non correlati. Questo metodo appare particolarmente pratico poiché, una volta identificato il modello dell'auto-correlazione, esso non richiede una nuova elaborazione del modello probabilistico del processo, ma semplicemente di calcolare gli errori fra quanto previsto dal modello, e quanto effettivamente osservato sul processo. Questi errori di previsione (o residui) essendo approssimativamente non correlati, potranno poi essere direttamente analizzati con i metodi SPC sviluppati per dati indipendenti.

In questa sezione saranno richiamati brevemente i modelli di auto-correlazione temporale ARIMA. Prima di introdurre i modelli ARIMA sarà indicata la terminologia adottata.

3.6.1 Terminologia

Sia data una serie temporale di dati $\{X_t\}$. Si farà riferimento, unicamente, a serie di tipo discreto e si supporrà che le osservazioni siano tratte ad uguali intervalli di tempo. Consideriamo un insieme generico di osservazioni $X_{i+j_1}, \dots, X_{i+j_r}$ estratti dalla serie temporale. La serie di dati è definita *strettamente stazionaria*, se è possibile identificarne una distribuzione congiunta di probabilità che non risulta dipendente dall'indice i qualunque siano gli indici j_1, \dots, j_r , e qualunque sia r , in caso contrario essa è definita *non stazionaria*. Se, invece, la definizione di stazionarietà stretta vale solo per $r \leq p$ dove p è un intero positivo allora il processo è detto *stazionario di ordine p* . In particolare la stazionarietà di ordine 2 è detta *stazionarietà debole*. Una stazionarietà debole implica semplicemente che:

$$\begin{cases} E(X_t) = \mu \\ \text{Cov}(X_t, X_{t-k}) = \gamma_k \end{cases} \quad \forall t = 1, 2, \dots \quad (3.35)$$

Ovvero la media è non dipendente dal tempo e l'auto-covarianza di lag (ritardo) k (intero) è sempre costante cioè anch'essa indipendente dal tempo. In particolare la auto-covarianza di lag 0 è la varianza del processo che, quindi, risulta costante per un processo stazionario e in particolare per un processo debolmente stazionario. Si definisce coefficiente di auto-correlazione al lag k :

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \frac{\gamma_k}{\sigma_X^2} \quad (3.36)$$

La serie di valori $\{\rho_k\}$ $k=1,2,\dots$ è definita la funzione di auto - correlazione della serie di dati (AutoCorrelation Function – ACF). Essa descrive il grado di mutua dipendenza, tra i valori della serie temporale a differenti intervalli di ritardo (lag). La funzione d'auto-correlazione parziale (PACF) è una misura della correlazione, ed è utilizzata per identificare l'estensione della relazione tra il valore corrente di una variabile, e i precedenti valori della stessa, mantenendo gli effetti per tutti gli altri ritardi di tempo costanti. Una stima delle due funzioni, ACF e PACF, può essere di notevole supporto per l'identificazione del più appropriato modello matematico della serie temporale.

3.6.2 Modelli ARIMA(p, d, q)

Un approccio comune per trattare con processi auto-correlati, è la procedura di modellazione attraverso l'uso di modelli **Auto-Regressive Integral Moving-Average** (ARIMA) di parametri (p, d, q) (Box *et al.* 1994). I modelli ARIMA sono utilizzati per descrivere diversi processi auto-correlati, e la loro applicabilità, come strumento SPC per processi auto-correlati, è largamente riconosciuta (Montgomery 2000). Una serie di dati $\{X_t\}$ è generata da un processo ARIMA (p, d, q) del modello di Box *et al.* (1994) se risulta:

$$\phi_p(B)(1-B)^d(X_t - \mu) = \Theta_q(B)\varepsilon_t \quad \varepsilon_t \approx NID(0, \sigma^2) \quad (3.37)$$

dove

$$\phi_p(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p \quad (3.38)$$

$$\Theta_q(B) = 1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q \quad (3.39)$$

Le due equazioni precedenti sono polinomi in B, di grado p e q rispettivamente, B è l'operatore backward shift (ovvero $B: BX_t = X_{t-1}$), p, d, q sono indici interi, e X_t denota il valore osservato della serie temporale all'istante di tempo t . $\phi_p(B)$ e $\Theta_q(B)$ sono indicati come i polinomi AR e MA rispettivamente. ε_t è un processo normale i.i.d. di media uguale a zero e varianza σ^2 , t è un indice temporale. Se indichiamo con \hat{X}_t la previsione per l'istante t ottenuta da un modello ARIMA adeguato, allora i residui $E_t = X_t - \hat{X}_t$ si comporteranno come variabili aleatorie indipendenti e identicamente distribuite (Box *et al.* 1994, Montgomery 2000). La presenza di cause assegnabili di variazione, perciò, può essere identificata analizzando l'andamento dei residui delle previsioni $\{E_t\}$ tracciati in una carta di controllo di Shewhart. Se l'uscita del processo non presenta cambiamenti nella media, o nella struttura d'auto-correlazione, allora i residui esibiranno, approssimativamente, una media di zero e una non rilevante auto-correlazione ad ogni lag.

Dalla equazione (3.37) è possibile ricavare il valore della serie temporale all'istante di tempo t .

$$X_t = \mu + \frac{\Theta_q(B)\varepsilon_t}{\phi_p(B)(1-B)^d} \quad \varepsilon_t \approx NID(0, \sigma^2) \quad (3.40)$$

Definendo con $\{S_t\}$ la serie temporale di un pattern sull'uscita del processo allora la reale serie controllata è rappresentabile come segue.

$$Y_t = \frac{X_t - \mu + S_t}{\sigma_x} \quad (3.41)$$

Sostituendo il valore dell'espressione (3.40) nella equazione (3.41) si ottiene la seguente relazione.

$$\sigma_x Y_t = \frac{\Theta_q(B)\varepsilon_t}{\phi_p(B)(1-B)^d} + S_t \quad \varepsilon_t \approx NID(0, \sigma^2) \quad (3.42)$$

La serie temporale dei residui del processo denotati con E possono essere rappresentati come segue.

$$E_t = \frac{\phi_p(B)(1-B)^d}{\Theta_q(B)} Y_t \quad (3.43)$$

E quindi utilizzando l'espressione (3.42) nella equazione (3.43) si ottiene la seguente espressione per i residui.

$$E_t = \frac{\varepsilon_t}{\sigma_x} + \frac{\phi_p(B)(1-B)^d}{\sigma_x \Theta_q(B)} S_t \quad \varepsilon_t \approx NID(0, \sigma^2) \quad (3.44)$$

Il valore medio della serie dei residui dipende dall'entità del disturbo, dal tipo di modello ARIMA e dalla forma della funzione di disturbo.

$$E[E_t] = \frac{1}{\sigma_x} E \left[\frac{\phi_p(B)(1-B)^d}{\Theta_q(B)} S_t \right] \quad (3.45)$$

Essendo

$$S_t = \varphi \sigma_x s_t \quad (3.46)$$

Allora si ha che

$$E[E_t] = \varphi E \left[\frac{\phi_p(B)(1-B)^d}{\Theta_q(B)} s_t \right] \quad (3.47)$$

Quest'ultima espressione è significativa del fatto che dato un generico segnale di disturbo su un processo (s_t) di ampiezza φ questo modificherà la media dei residui per

un fattore pari a φ moltiplicato la media della serie temporale $\left\{ \frac{\phi_p(B)(1-B)^d}{\Theta_q(B)} s_t \right\}$

quest'ultima è da considerarsi l'effetto filtrato attraverso il modello ARIMA del pattern non naturale sui residui. Quest'osservazione avrà delle importanti conseguenze sulle prestazioni delle carte di controllo sui residui per l'identificazione di determinati tipi di pattern.

Un modello ARIMA che si è dimostrato particolarmente utile nella pratica è il modello AR di ordine 1. Questo modello, che è il più semplice fra i modelli ARIMA(p, d, q), sarà analizzato nel dettaglio, e servirà come caso di studio per valutare le prestazioni dell'algoritmo neurale in caso di dipendenza temporale dei dati.

3.6.3 Il modello AR di ordine 1

Un processo $\{X_t\}$ è detto un modello auto-regressivo del primo ordine se esso è generato dalla seguente equazione:

$$X_t = \mu + \phi(X_{t-1} - \mu) + \varepsilon_t; \quad \varepsilon_t \approx NID(0, \sigma^2) \quad (3.48)$$

Dove X_t è il valore della serie temporale all'istante di tempo $t = 1, 2, \dots$, μ è la media della serie di dati, ε_t è un termine d'errore distribuito secondo una normale, e ϕ è il coefficiente d'auto-regressione ristretto a variare nell'intervallo aperto $]-1, 1[$. Le successive osservazioni della serie X_t sono correlate siccome è:

$$Cov(X_t, X_{t-k}) = \phi^k \gamma_0 = \phi^k \sigma_X^2 \quad \forall t, k \quad (3.49)$$

e dove risulta che

$$\sigma_X^2 = Var(X_t) = \frac{\sigma^2}{1 - \phi^2} \quad (3.50)$$

Un modello AR di ordine 1 è un modello strettamente stazionario dal momento che è stato posto nella definizione $|\phi| < 1$ (Box *et al.* 1994). Ovviamente si ha che:

$$E[X_t] = \mu \quad \forall t \quad (3.51)$$

I modelli auto regressivi del primo ordine, si rivelano utili quando dei disturbi influenzano non solo l'attuale risposta del sistema, ma hanno anche (con pesi esponenzialmente decrescenti) un effetto su quelle future.

3.6.4 Il modello del processo AR di ordine 1

Indichiamo con $\{X_t\}$ la serie di dati derivante dalla componente naturale di variabilità del processo, con $\{S_t\}$ la serie di dati caratterizzante il pattern speciale di variazione e con $\{Y_t\}$ la serie di dati all'uscita del processo che consideriamo standardizzata prima che sia analizzata dal sistema di controllo. Allora valgono le seguenti relazioni ad ogni generico valore dell'indice temporale $t = 1, 2, \dots$

$$\begin{aligned} X_t &= \mu + \phi(X_{t-1} - \mu) + \varepsilon_t = \mu + \phi(X_{t-1} - \mu) + \sigma n_t \quad n_t \approx NID(0, 1) \\ \sigma_X &= \frac{\sigma}{\sqrt{1 - \phi^2}}; S_t = \phi \sigma_X S_t; \\ Y_t &= \frac{X_t + S_t - \mu}{\sigma_X} \end{aligned} \quad (3.52)$$

Che possono essere schematizzate nel seguente diagramma a blocchi.

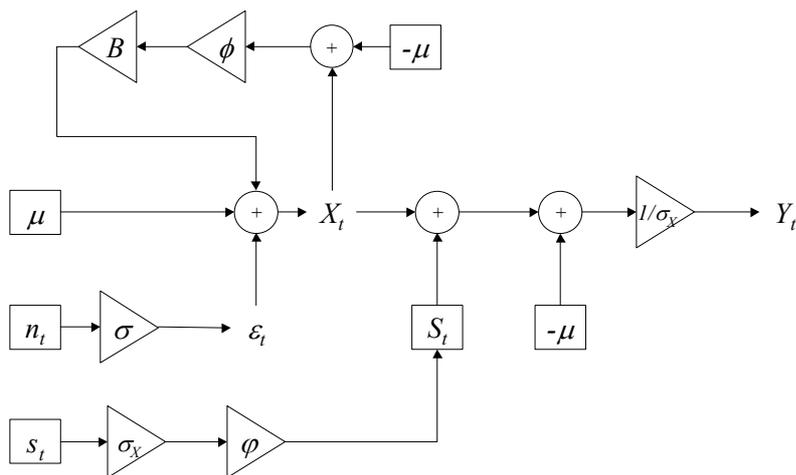


Figura 3.10 – Modello generale di simulazione del processo AR di ordine 1

Le relazioni dell'equazione (3.52) sono equivalenti alle seguenti dov'è rinominata la serie di dati caratterizzante il pattern non naturale di variazione.

$$X_t = \mu + \phi(X_{t-1} - \mu) + \varepsilon_t = \mu + \phi(X_{t-1} - \mu) + \sigma n_t \quad n_t \approx NID(0,1);$$

$$\sigma_X = \frac{\sigma}{\sqrt{1-\phi^2}}; Z_t = \frac{X_t - \mu}{\sigma_X}; S_t = \phi s_t \tag{3.53}$$

$$Y_t = Z_t + S_t$$

Le relazioni della equazione (3.53) possono ulteriormente essere riscritte come segue.

$$n_t \approx NID(0,1)$$

$$\sigma_X = \frac{\sigma}{\sqrt{1-\phi^2}}; Z_t = \phi Z_{t-1} + \frac{\sigma n_t}{\sigma_X} = \phi Z_{t-1} + n_t \sqrt{1-\phi^2}; S_t = \phi s_t; \tag{3.54}$$

$$Y_t = Z_t + S_t$$

Il corrispondente diagramma a blocchi è il seguente.

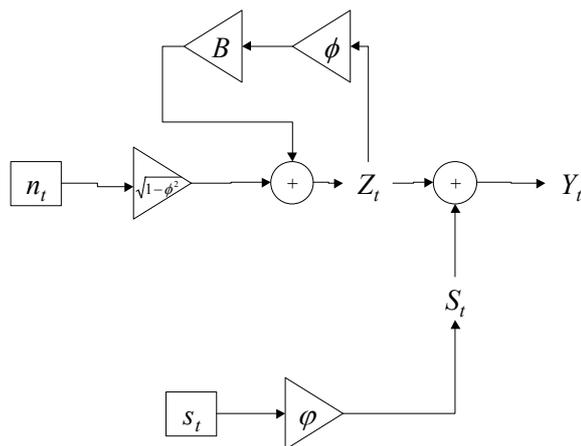


Figura 3.11– Modello semplificato di simulazione del processo AR di ordine 1

Conseguentemente, l'ipotesi nulla può essere formulata all'istante t come segue:

$$H_0 : Y_t = Z_t$$

e l'ipotesi alternativa come

$$H_1 : Y_t = Z_t + S_t$$

3.7 Carte di controllo per dati auto-correlati

Consideriamo un modello AR di ordine 1 e parametro ϕ come descritto nella precedente sezione. Posto $Z_0 = 0$ e denotando il residuo all'istante t attraverso E_t , l'equazione per generare E_t può essere scritta come segue.

$$E_t = Y_t - \phi Y_{t-1} = n_t \sqrt{1 - \phi^2} + S_t - \phi S_{t-1} \quad n_t \approx NID(0,1) \quad (3.55)$$

Che può essere riscritta nella seguente forma equivalente.

$$\begin{aligned} E_t &= \rho_t + S_t - \phi S_{t-1} & \rho_t &\approx NID(0, \sigma_\rho^2) \\ \sigma_\rho^2 &= (1 - \phi^2) \end{aligned} \quad (3.56)$$

Il valore atteso del residuo E_t in un generico istante di tempo $t \geq 1$, può essere espresso attraverso le seguenti relazioni.

$$E[E_t] = \sqrt{1 - \phi^2} E[n_t] + E[S_t] - \phi E[S_{t-1}] = E[S_t] - \phi E[S_{t-1}] \quad (3.57)$$

$$E[Y_t] = 0 \quad t = 1, 2, \dots \quad (3.58)$$

se il processo è in uno stato naturale di variazione

$$E[E_t] = \begin{cases} 0 & t < \tau \\ \phi & t = \tau \\ (1 - \phi)\phi & t > \tau \end{cases} \quad t = 1, 2, \dots \quad (3.59)$$

se il pattern non naturale è modellato da uno shift

$$E[Y_t] = \begin{cases} -(1 + \phi)\phi & t = (2i + 1) & i = 0, 1, 2, \dots \\ (1 + \phi)\phi & t = (2i) & i = 1, 2, \dots \end{cases} \quad (3.60)$$

se il pattern non naturale è modellato da una variazione sistematica.

$$E[E_t] = \begin{cases} 0 & t = 1 \\ -\phi & t = (4i + 2) \\ \phi\phi & t = (4i + 3) & i = 0, 1, 2, \dots, n \dots \\ \phi & t = (4i + 4) \\ -\phi\phi & t = (4i + 5) \end{cases} \quad (3.61)$$

se il pattern è un ciclo di periodo 4

$$E[E_t] = \begin{cases} \phi(1 - \phi) & m_t = 1; m_{t-1} = 1 \\ \phi(1 + \phi) & m_t = 1; m_{t-1} = -1 \\ -\phi(1 + \phi) & m_t = -1; m_{t-1} = 1 \\ -\phi(1 - \phi) & m_t = -1; m_{t-1} = -1 \end{cases} \quad t = 1, 2, \dots \quad (3.62)$$

se il pattern è modellato come una mixture

Si osservi, anzitutto, che la carta di controllo dovrà essere applicata sui residui dei dati standardizzati. Noto il modello AR di ordine 1, vale a dire noto il valore di ϕ , supposto

il processo in uno stato di controllo ovvero supponendo $\phi = 0$, i residui presentano una deviazione standard pari a $\sigma_\rho = \sqrt{1-\phi^2}$. Ne consegue che i limiti della carta di controllo dovranno essere calcolati utilizzando tale deviazione standard.

3.7.1 Carta di controllo di Shewhart per il modello di riferimento

Con riferimento al modello descritto dalle equazioni (3.54 e (3.56, e rappresentato nella Figura 3.11, la costruzione della carta di controllo sui residui risulta sensibilmente semplificata. Dalle ipotesi del modello, ipotizzando vera l'ipotesi H_0 (processo in uno stato naturale), si ricava $E[E_t] = 0, VAR[E_t] = (1-\phi^2)$, così come si ricava che le osservazioni $\{E_t\}$ sono indipendenti e identicamente distribuite in modo normale. La conoscenza di tali parametri consente di impostare la carta senza necessità della fase preparatoria di stima dei parametri del processo. Analizzando, inoltre, risposte singole risulta $n = 1$. Con queste ipotesi i limiti di controllo della carta risulteranno

$$UCL = E[E_t] + c(1, p_{UCL}) \cdot \sqrt{VAR[E_t]} = c(1, p_{UCL}) \sqrt{1-\phi^2} \quad (3.63)$$

$$LCL = E[E_t] - c(1, p_{LCL}) \cdot \sqrt{VAR[E_t]} = -c(1, p_{LCL}) \sqrt{1-\phi^2} \quad (3.64)$$

Come nel caso di dati indipendenti, imponendo $c(1, p_{UCL}) = -c(1, p_{LCL}) = l$, e ipotizzando di fissare un opportuno livello di lunghezza media dei run prima che sia segnalato un fuori controllo (Average Run Length – ARL) quando il processo è in uno stato normale, la progettazione dei limiti è semplicemente ottenuta uguagliando l a

$\left| Z_{\frac{1}{2ARL}} \right|$ ovvero uguale al quantile $\frac{1}{2ARL}$ della distribuzione standard normale.

3.7.2 Calcolo errori Tipo I e Tipo II per la carta Shewhart sui residui

Nel seguito sono riportati i valori attesi per una carta di controllo di Shewhart calcolata sui residui di un processo descritto dal modello delle equazioni (3.54 e (3.56, e rappresentato nella Figura 3.11.

3.7.2.1 Errore di Tipo I

Nel caso in cui il processo è in controllo statistico, si avrà un allarme quando l'attuale osservazione supera i limiti di controllo della carta Shewhart. Il valore atteso dell'errore di Tipo I in un generico istante di tempo $t \geq 1$, può essere espresso in formula come segue.

$$\alpha_t = 1 - P\{-l\sigma_\rho \leq \rho_t \leq l\sigma_\rho\} \quad t = 1, 2, \dots \quad (3.65)$$

3.7.2.2 Errore di Tipo II: pattern shift

Supponiamo di considerare uno shift di ampiezza ϕ a partire dall'istante τ . L'errore di Tipo II, atteso per ogni istante di tempo $t \geq \tau$, è espresso dalla seguente relazione.

$$\begin{aligned} \beta_t(\phi, \phi) &= P\{-l\sigma_\rho \leq \rho_t + \phi(1-\phi) \leq l\sigma_\rho\}, & t = \tau + i; i = 1, 2, \dots \\ \beta_t(\phi, \phi) &= P\{-l\sigma_\rho - \phi(1-\phi) \leq \rho_t \leq l\sigma_\rho - \phi(1-\phi)\} & t = \tau + i; i = 1, 2, \dots \end{aligned} \quad (3.66)$$

3.7.2.3 Errore di Tipo II: pattern variazione sistematica

Consideriamo un pattern variazione sistematica di ampiezza φ . L'errore di Tipo II atteso in un generico istante di tempo $t \geq 1$ può essere espresso in formula come segue.

$$\begin{aligned}
 \beta_t(\varphi, \phi) &= \begin{cases} P\{-l\sigma_\rho \leq \rho_t - \varphi(1+\phi) \leq l\sigma_\rho\} & t = (2i+1), i = 0,1,\dots \\ P\{-l\sigma_\rho \leq \rho_t + \varphi(1+\phi) \leq l\sigma_\rho\} & t = (2i), i = 1,2,\dots \end{cases} \\
 \beta_t(\varphi, \phi) &= \begin{cases} P\{-l\sigma_\rho + \varphi(1+\phi) \leq \rho_t \leq l\sigma_\rho + \varphi(1+\phi)\} & t = (2i+1), i = 0,1,\dots \\ P\{-l\sigma_\rho - \varphi(1+\phi) \leq \rho_t \leq l\sigma_\rho - \varphi(1+\phi)\} & t = (2i), i = 1,2,\dots \end{cases} \quad (3.67) \\
 \beta_t(\varphi, \phi) &= \begin{cases} P\{-l\sigma_\rho - \varphi(1+\phi) \leq \rho_t \leq l\sigma_\rho - \varphi(1+\phi)\} & t = (2i+1), i = 0,1,\dots \\ P\{-l\sigma_\rho - \varphi(1+\phi) \leq \rho_t \leq l\sigma_\rho - \varphi(1+\phi)\} & t = (2i), i = 1,2,\dots \end{cases} \\
 \beta_t(\varphi, \phi) &= P\{-l\sigma_\rho - \varphi(1+\phi) \leq \rho_t \leq l\sigma_\rho - \varphi(1+\phi)\} \quad t = 1,2,\dots
 \end{aligned}$$

3.7.2.4 Errore di Tipo II: pattern ciclo

Consideriamo un pattern Ciclo di ampiezza φ e periodo pari a 4. In tale caso, il valore atteso dell'errore di Tipo II dipenderà oltre che dall'ampiezza φ del disturbo, anche dall'istante di osservazione. In particolare valgono la seguente formulazione per istanti di tempo $t \geq 1$.

$$\begin{aligned}
 \beta_t(\varphi, \phi) &= \begin{cases} P\{-l\sigma_\rho \leq \rho_t \leq l\sigma_\rho\} & t = 1 \\ P\{-l\sigma_\rho \leq \rho_t - \varphi \leq l\sigma_\rho\} & t = (4i+2) \\ P\{-l\sigma_\rho \leq \rho_t + \phi\varphi \leq l\sigma_\rho\} & t = (4i+3) \quad i = 0,1,2,\dots; \\ P\{-l\sigma_\rho \leq \rho_t + \varphi \leq l\sigma_\rho\} & t = (4i+4) \\ P\{-l\sigma_\rho \leq \rho_t - \phi\varphi \leq l\sigma_\rho\} & t = (4i+5) \end{cases} \\
 \beta_t(\varphi, \phi) &= \begin{cases} P\{-l\sigma_\rho \leq \rho_t \leq l\sigma_\rho\} & t = 1 \\ P\{-l\sigma_\rho + \varphi \leq \rho_t \leq l\sigma_\rho + \varphi\} & t = (4i+2) \\ P\{-l\sigma_\rho - \phi\varphi \leq \rho_t \leq l\sigma_\rho - \phi\varphi\} & t = (4i+3) \quad i = 0,1,2,\dots; \\ P\{-l\sigma_\rho - \varphi \leq \rho_t \leq l\sigma_\rho - \varphi\} & t = (4i+4) \\ P\{-l\sigma_\rho + \phi\varphi \leq \rho_t \leq l\sigma_\rho + \phi\varphi\} & t = (4i+5) \end{cases} \quad (3.68) \\
 \beta_t(\varphi, \phi) &= \begin{cases} P\{-l\sigma_\rho \leq \rho_t \leq l\sigma_\rho\} & t = 1 \\ P\{-l\sigma_\rho - \varphi \leq \rho_t \leq l\sigma_\rho - \varphi\} & t = (4i+2) \\ P\{-l\sigma_\rho - \phi\varphi \leq \rho_t \leq l\sigma_\rho - \phi\varphi\} & t = (4i+3) \quad i = 0,1,2,\dots; \\ P\{-l\sigma_\rho - \varphi \leq \rho_t \leq l\sigma_\rho - \varphi\} & t = (4i+4) \\ P\{-l\sigma_\rho - \phi\varphi \leq \rho_t \leq l\sigma_\rho - \phi\varphi\} & t = (4i+5) \end{cases} \\
 \beta_t(\varphi, \phi) &= \begin{cases} P\{-l\sigma_\rho \leq \rho_t \leq l\sigma_\rho\} & t = 1 \\ P\{-l\sigma_\rho - \varphi \leq \rho_t \leq l\sigma_\rho - \varphi\} & t = 2i \quad i = 1,2,\dots; \\ P\{-l\sigma_\rho - \phi\varphi \leq \rho_t \leq l\sigma_\rho - \phi\varphi\} & t = 2i+1 \end{cases}
 \end{aligned}$$

3.7.2.5 Errore di Tipo II: pattern mixture

Consideriamo un pattern mixture di ampiezza ϕ . Il valore atteso per l'errore di Tipo II in un generico istante di tempo $t \geq 1$ può essere espresso come somma di probabilità di eventi indipendenti. In particolare, nelle successive relazioni, si è indicato con p_u e p_c rispettivamente la probabilità di non cambiamento di stato e di cambiamento di stato della catena di Markov utilizzata nella modellizzazione. Si osservi che, incidentalmente, il tasso di allarmi non dipende da tali probabilità.

$$\begin{aligned}
 \beta_t(\phi, \phi) &= \left(\begin{array}{l} P\{-l\sigma_\rho \leq \rho_t + \phi(1-\phi) \leq l\sigma_\rho, m_t = 1, m_{t-1} = 1\} + \\ P\{-l\sigma_\rho \leq \rho_t + \phi(1+\phi) \leq l\sigma_\rho, m_t = 1, m_{t-1} = -1\} + \\ P\{-l\sigma_\rho \leq \rho_t - \phi(1-\phi) \leq l\sigma_\rho, m_t = -1, m_{t-1} = 1\} + \\ P\{-l\sigma_\rho \leq \rho_t - \phi(1+\phi) \leq l\sigma_\rho, m_t = -1, m_{t-1} = -1\} \end{array} \right); \\
 \beta_t(\phi, \phi) &= \left(\begin{array}{l} P\{-l\sigma_\rho \leq \rho_t + \phi(1-\phi) \leq l\sigma_\rho\} p_u P\{m_{t-1} = 1\} + \\ P\{-l\sigma_\rho \leq \rho_t + \phi(1+\phi) \leq l\sigma_\rho\} p_c P\{m_{t-1} = -1\} + \\ P\{-l\sigma_\rho \leq \rho_t - \phi(1-\phi) \leq l\sigma_\rho\} p_c P\{m_{t-1} = 1\} + \\ P\{-l\sigma_\rho \leq \rho_t - \phi(1+\phi) \leq l\sigma_\rho\} p_u P\{m_{t-1} = -1\} \end{array} \right); \\
 \beta_t(\phi, \phi) &= \left(\begin{array}{l} P\{-l\sigma_\rho - \phi(1-\phi) \leq \rho_t \leq l\sigma_\rho - \phi(1-\phi)\} p_u P\{m_{t-1} = 1\} + \\ P\{-l\sigma_\rho - \phi(1+\phi) \leq \rho_t \leq l\sigma_\rho - \phi(1+\phi)\} p_c P\{m_{t-1} = -1\} + \\ P\{-l\sigma_\rho + \phi(1-\phi) \leq \rho_t \leq l\sigma_\rho + \phi(1-\phi)\} p_c P\{m_{t-1} = 1\} + \\ P\{-l\sigma_\rho + \phi(1+\phi) \leq \rho_t \leq l\sigma_\rho + \phi(1+\phi)\} p_u P\{m_{t-1} = -1\} \end{array} \right); \tag{3.69} \\
 \beta_t(\phi, \phi) &= \left(\begin{array}{l} P\{-l\sigma_\rho - \phi(1-\phi) \leq \rho_t \leq l\sigma_\rho - \phi(1-\phi)\} (p_u + p_c) P\{m_{t-1} = 1\} + \\ P\{-l\sigma_\rho - \phi(1+\phi) \leq \rho_t \leq l\sigma_\rho - \phi(1+\phi)\} (p_c + p_u) P\{m_{t-1} = -1\} \end{array} \right); \\
 \beta_t(\phi, \phi) &= \left(\begin{array}{l} P\{-l\sigma_\rho - \phi(1-\phi) \leq \rho_t \leq l\sigma_\rho - \phi(1-\phi)\} P\{m_{t-1} = 1\} + \\ + P\{-l\sigma_\rho - \phi(1+\phi) \leq \rho_t \leq l\sigma_\rho - \phi(1+\phi)\} P\{m_{t-1} = -1\} \end{array} \right); \\
 \beta_t(\phi, \phi) &= \frac{1}{2} \left(\begin{array}{l} P\{-l\sigma_\rho - \phi(1-\phi) \leq \rho_t \leq l\sigma_\rho - \phi(1-\phi)\} + \\ + P\{-l\sigma_\rho - \phi(1+\phi) \leq \rho_t \leq l\sigma_\rho - \phi(1+\phi)\} \end{array} \right) \quad t = 1, 2, \dots
 \end{aligned}$$

3.7.3 Discussione

Per quanto riguarda l'errore di Tipo I di una carta di controllo di Shewhart applicata ai residui di un processo auto-correlato, è possibile osservare che esso è indipendente dall'istante di osservazione. È immediato ricavare che il tasso medio temporale dell'errore di Tipo I per una carta di Shewhart è dato dalla seguente relazione.

$$\begin{aligned}
 \alpha &= \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \alpha_i = 1 - P\{-l\sigma_\rho \leq \rho_i \leq l\sigma_\rho\}; \\
 \alpha &= 1 - P\{-l \leq n_t \leq l\}
 \end{aligned} \tag{3.70}$$

Per quanto riguarda il tasso di errori di Tipo II per pattern di tipo Variazione Sistemática e Mixture, si può osservare che anch'essi risultano indipendenti dall'istante di osservazione. Risulta inoltre che anche il tasso atteso di errori di Tipo II per uno Shift è indipendente dall'istante di osservazione per ogni $t \geq \tau$.

Per i pattern di tipo Shift (con $\tau = 1$), la media temporale del tasso di errore di Tipo II per una ampiezza ϕ è pari alla seguente quantità.

$$\beta(\phi, \phi) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \beta_i(\phi, \phi) = P\{-l\sigma_\rho - \phi(1-\phi) \leq \rho_i \leq l\sigma_\rho - \phi(1-\phi)\}$$

$$\beta(\phi, \phi) = P\left\{-l - \phi \frac{1-\phi}{\sqrt{1-\phi^2}} \leq n_i \leq l - \phi \frac{1-\phi}{\sqrt{1-\phi^2}}\right\} \quad (3.71)$$

Analogamente, per un pattern di Variazione Sistemática risulta che

$$\beta(\phi, \phi) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \beta_i(\phi, \phi) = P\{-l\sigma_\rho - \phi(1+\phi) \leq \rho_i \leq l\sigma_\rho - \phi(1+\phi)\}$$

$$\beta(\phi, \phi) = P\left\{-l - \phi \frac{1+\phi}{\sqrt{1-\phi^2}} \leq n_i \leq l - \phi \frac{1+\phi}{\sqrt{1-\phi^2}}\right\} \quad (3.72)$$

Per un pattern di tipo Mixture si ha che

$$\beta(\phi, \phi) = \lim_{t \rightarrow \infty} \frac{1}{t} \sum_{i=1}^t \beta_i(\phi, \phi) = \frac{1}{2} \left(P\{-l\sigma_\rho - \phi(1-\phi) \leq \rho_i \leq l\sigma_\rho - \phi(1-\phi)\} + \right. \\ \left. + P\{-l\sigma_\rho - \phi(1+\phi) \leq \rho_i \leq l\sigma_\rho - \phi(1+\phi)\} \right)$$

$$\beta(\phi, \phi) = \frac{1}{2} \left(P\left\{-l - \phi \frac{1-\phi}{\sqrt{1-\phi^2}} \leq n_i \leq l - \phi \frac{1-\phi}{\sqrt{1-\phi^2}}\right\} + \right. \\ \left. + P\left\{-l - \phi \frac{1+\phi}{\sqrt{1-\phi^2}} \leq n_i \leq l - \phi \frac{1+\phi}{\sqrt{1-\phi^2}}\right\} \right) \quad (3.73)$$

Ed infine, nel caso di pattern Ciclo di periodo pari a 4 il valore medio temporale del tasso di errori di Tipo II è dato dalla seguente relazione (trascurando il contributo per $t = 1$).

$$\beta(\phi, \phi) = \lim_{t \rightarrow \infty} \frac{1}{t} \left(\sum_{i=0}^{\lfloor t/2 \rfloor} \beta_{2i+1}(\phi) + \sum_{i=1}^{\lfloor t/2 \rfloor} \beta_{2i}(\phi) \right) =$$

$$\beta(\phi, \phi) = \lim_{t \rightarrow \infty} \frac{1}{t} \left(\sum_{i=0}^{\lfloor t/2 \rfloor} P\{-l\sigma_\rho - \phi\phi \leq \rho_i \leq l\sigma_\rho - \phi\phi\} + \right. \\ \left. + \sum_{i=1}^{\lfloor t/2 \rfloor} P\{-l\sigma_\rho - \phi \leq \rho_i \leq l\sigma_\rho - \phi\} \right) =$$

$$\beta(\phi, \phi) = \frac{P\{-l\sigma_\rho - \phi\phi \leq \rho_i \leq l\sigma_\rho - \phi\phi\} + P\{-l\sigma_\rho - \phi \leq \rho_i \leq l\sigma_\rho - \phi\}}{2} \quad (3.74)$$

$$\beta(\phi, \phi) = \frac{1}{2} \left(P\left\{-l - \phi \frac{\phi}{\sqrt{1+\phi^2}} \leq n_i \leq l - \phi \frac{\phi}{\sqrt{1+\phi^2}}\right\} + \right. \\ \left. + P\left\{-l - \phi \frac{1}{\sqrt{1+\phi^2}} \leq n_i \leq l - \phi \frac{1}{\sqrt{1+\phi^2}}\right\} \right)$$

Dove con la notazione $\lfloor \cdot \rfloor$, si è indicato il primo intero inferiore o uguale al suo argomento.

I seguenti grafici riportano l'andamento dell'errore atteso di Tipo II ottenibile da una carta di controllo Shewhart ($l=3$) per i quattro tipi di pattern descritti in precedenza. L'andamento dell'errore di Tipo II è stato rappresentato in funzione dell'ampiezza del disturbo misurata in unità di deviazione standard (ϕ) per quattro valori del fattore ϕ del modello AR di primo ordine ($\phi = \{-0.8, -0.2, 0.2, 0.8\}$).

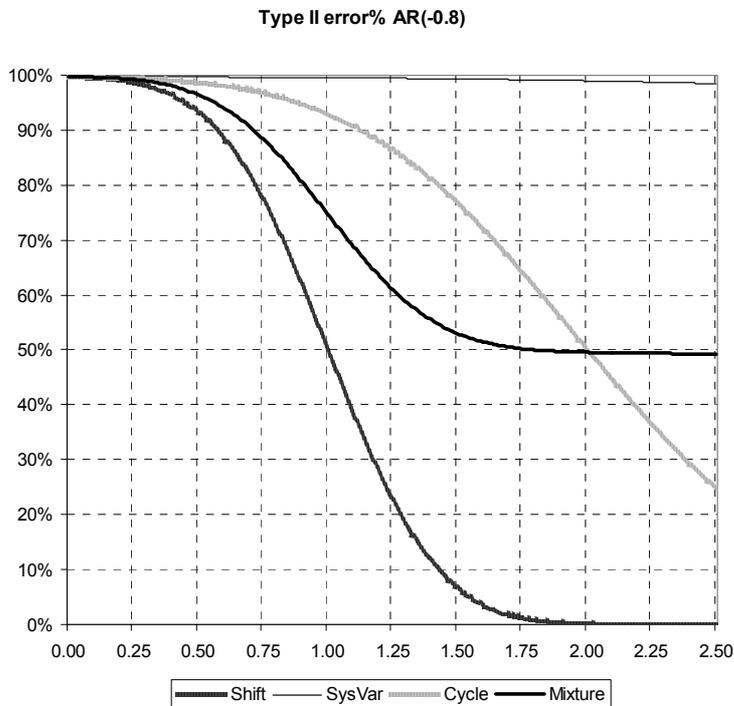


Figura 3.12– Errore di Tipo II per carta di controllo Shewhart. AR(-0.8)

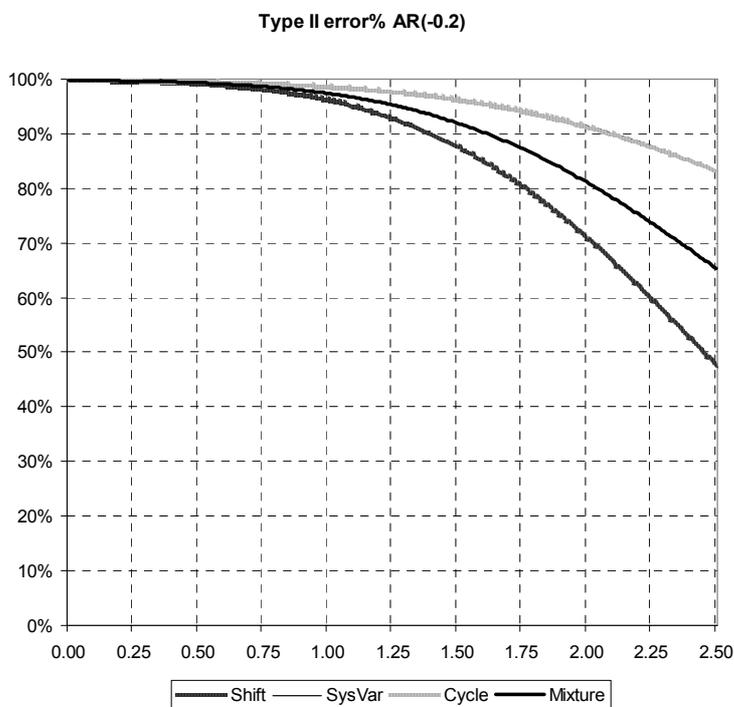


Figura 3.13– Errore di Tipo II per carta di controllo Shewhart. AR(-0.2)

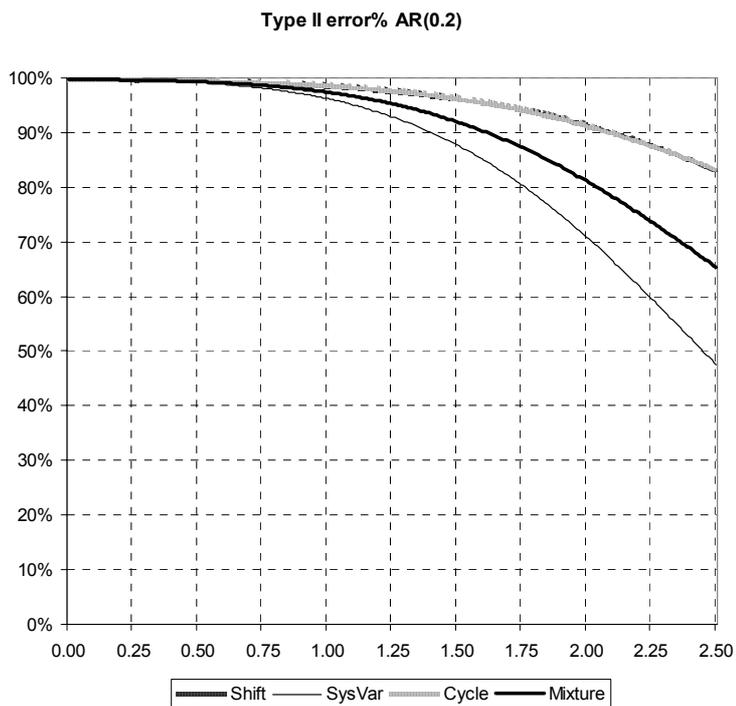


Figura 3.14– Errore di Tipo II per carta di controllo Shewhart. AR(+0.2)

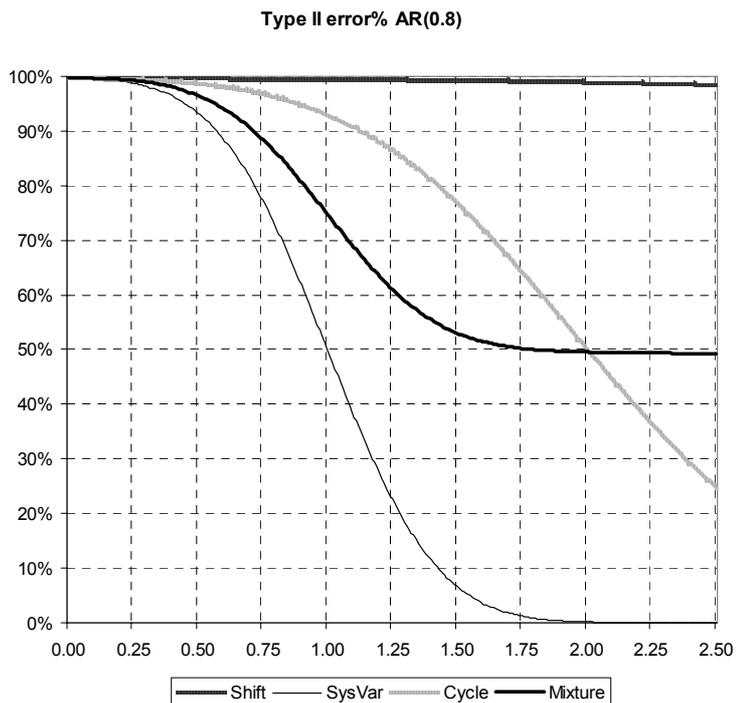


Figura 3.15– Errore di Tipo II per carta di controllo Shewhart. AR(+0.8)

Nei seguenti diagrammi sono state riportate le superfici che descrivono i tassi di errore di Tipo II per quattro tipi di pattern (shift, variazione sistematica, ciclo e mixture rispettivamente) presentati da una carta di controllo di Shewhart ($l = 3$) al variare dell'ampiezza del disturbo ($\phi \in [0,2]$, asse con etichetta "S") e del fattore di auto-correlazione AR del primo ordine ($\phi \in]-1,1[$, asse con etichetta "AR")

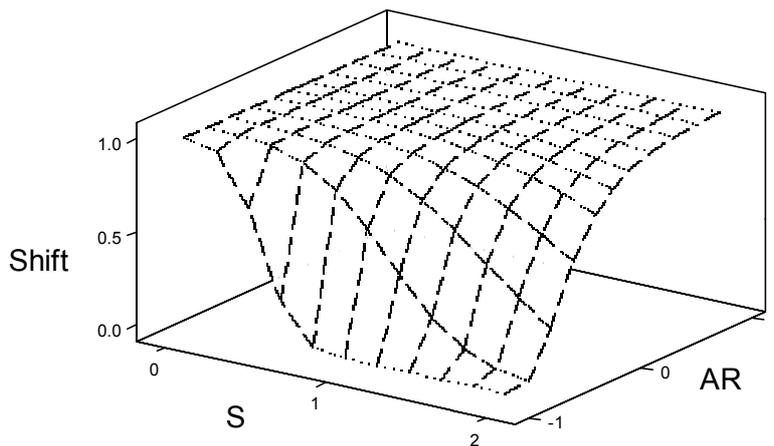


Figura 3.16 – Errore di Tipo II carta di controllo Shewhart. Pattern shift

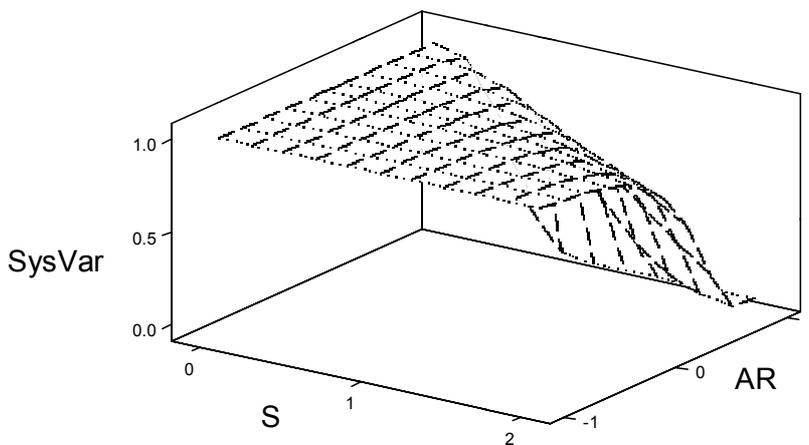


Figura 3.17 – Errore di Tipo II carta di controllo Shewhart. Pattern Variazione Sistemica

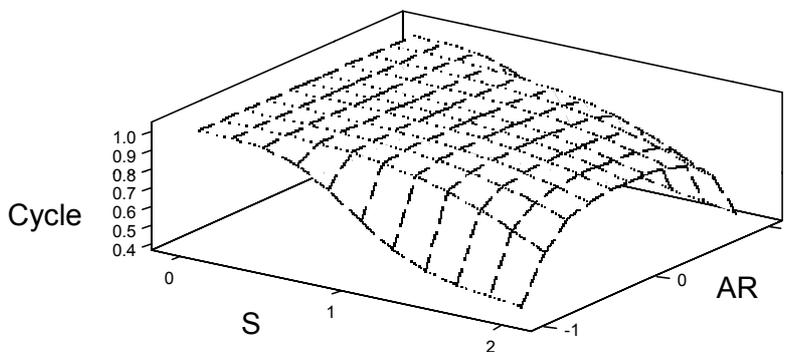


Figura 3.18 – Errore di Tipo II carta di controllo Shewhart. Pattern Ciclo

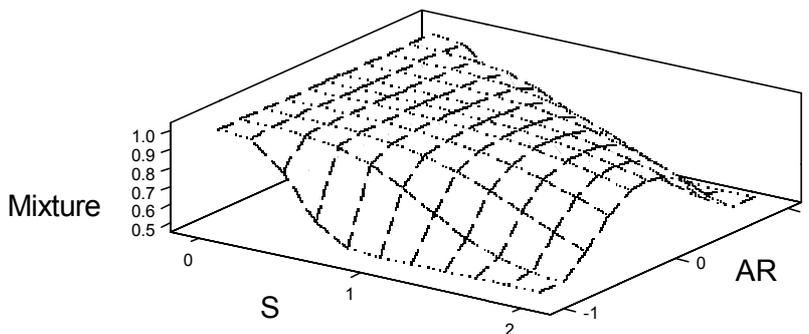


Figura 3.19 – Errore di Tipo II carta di controllo Shewhart. Pattern Mixture

Si può notare come la presenza dell'auto-correlazione dei dati possa degradare le prestazioni della carta di controllo nel riconoscimento di uno specifico tipo di pattern (ad esempio shift nel caso di processi positivamente auto-correlati, oppure variazioni sistematiche nel caso di processi negativamente auto-correlati), o al contrario, le possa migliorare notevolmente (ad esempio shift nel caso di processi negativamente auto-correlati, oppure variazioni sistematiche nel caso di processi positivamente auto-correlati). Nel caso, invece, di pattern di tipo ciclo e mixture, le prestazioni migliorano con la presenza di una forte correlazione fra i dati del processo indipendentemente dal tipo (positiva o negativa). In quest'ultimi due casi, tuttavia, il tasso di errore non risulta essere inferiore rispettivamente al 40% (ciclo) e al 50% (mixture) per disturbi di ampiezza pari a 2 unità di deviazione standard.

3.7.4 La carta CUSUM per i residui

In questa sezione esamineremo l'applicazione della carta di controllo CUSUM per un processo con uscite correlate del tipo AR di ordine 1 di parametro $|\phi| < 1$. Con riferimento al modello descritto dalle equazioni (3.54 e (3.56, e rappresentato nella Figura 3.11, la carta CUSUM sarà costruita utilizzando i residui:

$$E_t = Y_t - \phi Y_{t-1} \quad (3.75)$$

che nel caso di processo in controllo potrà essere riscritto come:

$$\begin{aligned} E_t &= \rho_t & \rho_t &\approx NID(0, \sigma_\rho^2) \\ \sigma_\rho^2 &= (1 - \phi^2) \end{aligned} \quad (3.76)$$

Denotiamo gli accumulatori della carta CUSUM sui residui (applicando l'approccio DIS) con $SH_{e,t}$ e $SL_{e,t}$ allora risulterà:

$$SH_{e,t} = \max(0, SH_{e,t-1} + E_t - K) = \max(0, SH_{e,t-1} + E_t - k\sqrt{1 - \phi^2}) \quad (3.77)$$

$$SL_{e,t} = \max(0, SL_{e,t-1} - E_t - K) = \max(0, SL_{e,t-1} - E_t - k\sqrt{1 - \phi^2}) \quad (3.78)$$

Un segnale di fuori controllo è generato qualora all'istante t sarà

$$SL_{e,t} > H \Rightarrow SL_{e,t} > h\sqrt{1 - \phi^2} \quad (3.79)$$

oppure se

$$SH_{e,t} > H \Rightarrow SH_{e,t} > h\sqrt{1 - \phi^2} \quad (3.80)$$

3.8 Finestra di analisi

Nelle applicazioni on-line per il monitoraggio della qualità, la valutazione dello stato del processo avviene solitamente attraverso l'analisi di una predefinita finestra di M successivi dati della sequenza temporale di osservazioni (Cheng 1995, 1997, 2001).

Indicando con $\{Y_t\}$ la sequenza temporale di osservazioni prodotte dal processo e indicando con $\tau \geq M$ l'indice temporale dell'ultimo punto rientrante nella finestra (definito *istante di applicazione*), allora $\tau - M + 1$ sarà l'indice temporale del primo punto rientrante nella finestra di osservazioni costituita dal seguente insieme di dati.

$$\{Y_{\tau-M+i}\}_{i=1,2,\dots,M} = \{Y_{\tau-M+1}, Y_{\tau-M+2}, \dots, Y_{\tau}\} \quad (3.81)$$

Esistono due possibili approcci di costruzione di una finestra di analisi: 1) a finestra fissa e 2) finestra mobile. I due approcci si differenziano per l'istante di applicazione della finestra di osservazione. In particolare, indicando con $\{\tau_n\}$ la sequenza dei successivi istanti di applicazione della finestra di osservazioni tale che $M \leq \tau_1 < \tau_2 < \dots < \tau_n < \dots$ allora la costruzione delle differenti osservazioni può essere ottenuta ponendo

$$\tau_{n+1} = \tau_n + p \quad p \in \{1, 2, \dots, j, \dots\} \quad (3.82)$$

In particolare, si ha un approccio a finestra fissa di ampiezza M quando $p = M$, mentre si ha un approccio a finestra mobile di ampiezza M se risulta che $p = 1$. In pratica con l'approccio a finestra fissa si prevede di suddividere la sequenza temporale di osservazioni del processo in sotto sequenze disgiunte di M osservazioni e di utilizzare ognuna di queste sotto-sequenze come un campione di osservazioni del processo dalle quali valutare lo stato del processo. Il secondo approccio prevede di utilizzare finestre di punti sovrapposte di $M - 1$ osservazioni.

Utilizzare un approccio di controllo a finestra fissa presenta lo svantaggio di richiedere che siano collezionati M osservazioni successive del processo prima che un segnale di allarme possa essere emesso.

Capitolo 4

Reti neurali per il controllo qualità

In questo capitolo, sono introdotte le tecniche neurali per il monitoraggio dei dati con particolare riferimento al loro impiego nell'ambito del controllo di qualità. In particolare è discusso il funzionamento generale di una rete a addestramento supervisionato con un esempio di applicazione ad uno specifico caso di controllo. In seguito, sono presentati in dettaglio gli algoritmi neurali ad apprendimento non supervisionato con particolare riferimento alla teoria della risonanza adattativa (Adaptive Resonance Theory – ART), di cui è discussa un'interpretazione geometrica del funzionamento. Utilizzando la descrizione dell'algoritmo neurale ART, inoltre, è proposta una procedura d'implementazione e di addestramento di una particolare rete neurale per il monitoraggio di un generico processo.

4.1 Overview sulle reti neurali

I maggiori vantaggi nell'utilizzo delle reti neurali rispetto ai tradizionali approcci SPC nel controllo di qualità sono: 1) l'elevata flessibilità, 2) l'elevata velocità di calcolo. Il primo vantaggio, corrisponde alla possibilità di poter utilizzare gli algoritmi neurali indipendentemente da ipotesi di modelli probabilistici e matematici del processo. L'elevata velocità di calcolo rende gli algoritmi neurali appropriati per sistemi di controllo in tempo reale on-line.

4.1.1 Struttura di una rete neurale

Una rete neurale può essere vista come un sistema composto di molti semplici elementi computazionali che operano in parallelo, la cui funzione dipende dalla struttura della rete, dall'insieme di connessioni e dall'elaborazione effettuata dagli elementi di calcolo. Una rete neurale, quindi, è formata da un elevato numero di elementi computazionali, detti neuroni, connessi tra loro, che elaborano informazioni in parallelo, per risolvere determinati problemi (Burke e Inizio 1997).

Ogni neurone svolge una funzione molto semplice: somma i segnali ricevuti in input ed utilizza il valore ottenuto per generare un output in base ad una particolare funzione matematica. In pratica, ogni neurone opera solo su informazioni locali in modo asincrono, esso deve elaborare un'informazione elementare originata da un'informazione globale ricevendo in input uno o più segnali ed emettendo in output un unico risultato. Generalmente, si assume che i segnali in input ai neuroni siano combinati linearmente mediante una serie di pesi, che rappresentano l'efficienza delle connessioni attraverso cui i segnali sono trasportati. La combinazione input/output, cioè la funzione di trasferimento realizzata dalla rete neurale, non è programmata ma è ottenuta tramite il processo di apprendimento.

Una generica rete neurale a singolo strato (layer) può essere caratterizzata da M input e N neuroni, ognuno dei quali risulta connesso a tutti gli input mediante M connessioni. Ogni connessione è dotata di un proprio peso: l'insieme dei pesi costituisce una matrice \underline{W} ad M righe ed N colonne, il cui generico elemento W_{ij} è il peso della connessione tra l'input di indice $i \in \{1, 2, \dots, M\}$ e il neurone di indice $j \in \{1, 2, \dots, N\}$.

Denotando con $\underline{I} = (I_1, I_2, \dots, I_M)$ il vettore degli ingressi alla rete allora l'output può essere descritto mediante un vettore:

$$\underline{y} = [y_1, y_2 \dots y_N] = \underline{f}(\underline{I} \cdot \underline{W}) \quad (4.1)$$

Avendo indicato con \underline{f} una funzione di trasferimento vettoriale definita *funzione di attivazione*.

Una rete neurale, inoltre, può essere caratterizzata da più strati successivi di neuroni. Tipicamente in ogni neurone di uno strato convergono gli output di tutti i neuroni dello strato precedente. L'architettura di una rete neurale, può anche essere distinta in base alla presenza o meno di feedback nelle connessioni fra neuroni di differenti strati. In generale si hanno due tipi di architetture.

- *Feed-Forward*, in questo caso non ci sono connessioni inverse, cioè collegamenti dai neuroni da uno strato verso quelli precedenti, e la rete non tiene traccia dei precedenti valori dell'output quando va ad attivare i suoi neuroni (un esempio è il *perceptrone*);
- *Feed-Back*, in questo caso sono presenti delle connessioni dai neuroni di uno strato, a quelli degli strati precedenti, con cui è possibile tenere traccia degli stati precedenti della rete. Questo tipo di rete è detto *rete ricorrente*.

Esiste una differenza di comportamento notevole tra le reti Feed-Forward e le reti ricorrenti: mentre l'output delle prime può essere calcolato subito dopo l'applicazione dell'input, le seconde presentano in genere un'evoluzione che passa attraverso una sequenza di *stati* della rete, e che in certi casi può anche portare all'*instabilità* della rete stessa.

Un altro parametro da analizzare è l'*addestramento* della rete. Lo scopo della fase di addestramento di una rete neurale è di trovare una distribuzione dei pesi delle connessioni in grado di produrre l'output desiderato (o almeno, un output coerente) quando è applicato un dato input. La procedura di addestramento comincia, in generale, con la scelta di un insieme di campioni di ingresso, il cosiddetto *training set*, e con l'inizializzazione dei pesi della rete con determinati valori (in alcuni casi casuali); prosegue poi applicando l'uno dopo l'altro, tutti i vettori del training set modificando, al tempo stesso, i pesi delle connessioni in base ad un algoritmo (definito *algoritmo di apprendimento* o di *training*) predeterminato e indipendente dal problema specifico che la rete è chiamata a risolvere. Durante il processo di training, e dopo un numero adeguato di presentazioni degli ingressi utilizzati per l'addestramento, i pesi convergono verso una distribuzione di valori che produrranno gli output desiderati.

Gli algoritmi di addestramento si dividono in due classi: *supervisionati* e *non supervisionati*. Nel metodo supervisionato, esiste un supporto esterno che di volta in volta indica alla rete qual è l'output desiderato. La rete si modifica di conseguenza in modo che, dopo un certo numero d'iterazioni, diventa capace di produrre autonomamente l'output corretto per ogni input. Generalmente, in questo caso, la rete si addestra mediante gli esempi che contengono l'input e il *target* richiesto (ovvero l'output desiderato). Questo metodo prevede che ciascun vettore del training set sia associato con il corrispondente vettore di output desiderato, in modo che nel complesso essi costituiscano una *coppia di addestramento*. Ad ogni applicazione di un input del training set, l'output della rete è calcolato e confrontato con il target corrispondente; l'errore o la differenza tra i due vettori di output è quindi utilizzata per stabilire l'entità delle modifiche da apportare ai valori dei pesi delle connessioni all'interno della rete. La procedura è iterata più volte, ogni volta utilizzando tutti i vettori del training set, fino a quando la differenza tra il target e l'uscita attuale prodotta dalla rete non si riduce ad un valore ritenuto accettabile: a questo punto l'addestramento può considerarsi concluso.

Nei casi d'apprendimento non supervisionato la rete impara scoprendo regolarità negli ingressi senza nessun supporto esterno di esempi di funzionamento desiderato. In questo caso la rete apprende autonomamente senza nessun supporto ulteriore. Ovviamente, affinché l'apprendimento attuato dalla rete neurale sia coerente a quanto desiderato, è necessario che questo avvenga in un ambiente definito.

4.1.2 Algoritmi di apprendimento

La fase di *training* consiste nella presentazione di un insieme di vettori con il fine di calcolare i valori dei pesi di tutte le connessioni della rete. L'algoritmo di apprendimento è la procedura utilizzata per stabilire le modifiche da apportare ai singoli pesi in funzione dell'ingresso attuale e dell'uscita prodotta dalla rete. I valori iniziali dei pesi della rete sono assegnati in modo da variare in un piccolo range tipicamente $[0,1]$ oppure $[-1,1]$, è possibile che i valori dei pesi siano assegnati in modo casuale. Nell'apprendimento supervisionato ogni pattern di addestramento è composto da una coppia (vettore d'input, vettore risposta desiderata), mentre in quello non supervisionato ogni pattern è composto dai soli vettori d'input facenti riferimento ad una prestabilita situazione.

L'aggiornamento dei pesi delle connessioni è calcolato aggiungendo una variazione al valore del peso assunto in precedenza, in pratica:

$$W_{i,j}^{new} = \Delta W_{i,j} + W_{i,j}^{old} \quad (4.2)$$

l'apprendimento procede in modo iterativo (nel senso che ogni pattern è presentato alla rete più volte) e graduale (nel senso che si aggiunge solo una parte della modifica sui pesi e non l'intera modifica) per limitare il problema che le nuove conoscenze vadano a modificare completamente ciò che è stato già appreso. La velocità d'apprendimento può essere regolata tramite il *tasso di apprendimento* (*learning rate*), indicato con β , che indica la percentuale di variazione applicata ai pesi, ovvero l'equazione (4.2) può essere riscritta come segue:

$$W_{i,j}^{new} = \beta \Delta W_{i,j} + (1 - \beta) W_{i,j}^{new} \quad 0 \leq \beta \leq 1. \quad (4.3)$$

Terminata la fase di apprendimento è possibile registrare i valori dei pesi (ciò consiste nel memorizzare la rete addestrata), calcolare la risposta della rete e analizzarla rispetto a dei vettori di test. La fase di *testing* consiste nel calcolare l'attivazione dei neuroni della rete senza modificare i pesi della stessa e nell'utilizzare dei pattern d'input nuovi, studiando così il comportamento della rete di fronte a stimoli nuovi, e verificando se la rete è in grado di *generalizzare*. Questa capacità di estrapolare da un apprendimento una capacità di trattare nuovi dati è uno dei principali vantaggi delle reti neurali che le distingue rispetto ai sistemi simbolici tradizionali.

4.1.3 Addestramento supervisionato

Di seguito è riportata la sequenza di passi eseguiti in un addestramento supervisionato di una rete neurale.

1. Impostare i valori iniziali dei pesi delle connessioni della rete.
2. Fornire un vettore di input alla rete che appartiene all'insieme di input usati per il training della rete stessa.
3. Calcolare il vettore di output.
4. Confrontare il vettore di output ottenuto al passo precedente, con l'output desiderato, e calcolare l'errore.

5. Correggere i pesi della connessione in modo tale che il vettore di output, la volta successiva, sia più vicino all'output desiderato.
6. Ripetere i passi 2 e 3 per far convergere l'output della rete in funzione di un determinato criterio di convergenza.

Quest' algoritmo generale d'addestramento supervisionato ha diverse implementazioni, principalmente distinte dal modo con cui sono cambiati i pesi delle connessioni durante l'addestramento stesso. Alcuni di questi algoritmi sono: Back-Propagation (BP) e l'algoritmo LVQ (Learning Vector Quantization). L'addestramento supervisionato è molto utile per risolvere problemi come la classificazione, per risolvere alcuni comportamenti prescritti, quando sono noti i patterns o le features.

4.2 Applicazione di una rete supervisionata

In questa sezione sarà brevemente descritta l'applicazione di una rete neurale supervisionata per il controllo di qualità. Lo scopo dell'applicazione è di addestrare una rete neurale a identificare alterazioni dello stato di variabilità naturale di un processo.

L'utilizzo di un algoritmo neurale supervisionato richiede di impostare una fase di addestramento che comprenda esempi relativi a tutte le classi di dati con le quali si intende classificare lo spazio di variabilità degli ingressi. Per quanto riguarda l'applicazione di una rete neurale a addestramento supervisionato per il controllo di qualità di un processo, poiché l'obiettivo è di discriminare lo spazio di variabilità degli ingressi in due classi (ingressi naturali e non naturali), è evidente che in fase di training è necessario fornire sia esempi di pattern naturali, sia esempi di pattern non naturali.

In altri termini, la scelta di un sistema supervisionato implica un'assunzione a priori circa il tipo (o i tipi) di pattern non naturale che si richiede sia identificato dalla rete sui dati prodotti dal processo.

In questo paragrafo è descritta l'applicazione di una rete neurale supervisionata per la realizzazione di un sistema specializzato al riconoscimento di un particolare pattern non - naturale: un trend di crescita o di decrescita delle osservazioni del processo.

4.2.1 Scelta della rete neurale supervisionata

Esistono differenti tipi di architetture neurali supervisionate che possono essere applicate al controllo di qualità. In questa sezione, sarà discussa in modo particolare una rete di perceptron a più livelli (Multi-Layer Perceptron - MLP) addestrata per mezzo dell'algoritmo supervisionato Back-Propagation. Questo tipo di rete neurale supervisionata è stata applicata con successo in differenti applicazioni di controllo della qualità. Tali applicazioni comprendono sia, riconoscimenti di uno o più pattern, sia stima dei parametri principali dei pattern identificati sui dati (capitolo 2).

Lo svantaggio della rete MLP-BP, è legato alla laboriosa fase di addestramento che, oltre a non essere cumulativa (ovvero un'estensione dell'insieme di addestramento richiederebbe ri-addestrare la rete anche sul precedente insieme di training), implica in genere un numero molto elevato d'iterazioni per poter raggiungere le prestazioni desiderate. Al contrario, invece, il processo di *recall*, vale a dire la fase operativa dell'algoritmo neurale, è sufficientemente rapido. Poiché l'addestramento può essere realizzato off-line, questa proprietà ha reso la rete MLP-BP quella diffusamente adottata in letteratura per scopi di controllo di qualità.

Altri tipi di reti neurali supervisionate possono essere adottate per il controllo di qualità. Alcuni esempi di utilizzo della rete neurale a connessioni Feed-Forward e a addestramento Radial Basis Function (RBF) sono già stati presentati in alcune

applicazioni (capitolo 2). Nessun accenno, invece, è stato riscontrato in letteratura sull'uso di reti ricorrenti per applicazioni di controllo della qualità. Le reti ricorrenti presentano connessioni di tipo Back-Forward: questa caratteristica topologica, associata ad opportune funzioni di ritardo nella propagazione dei segnali di ingresso, conferisce proprietà di memoria. Il principale vantaggio, atteso da una rete ricorrente per il controllo di qualità, è di poter sfruttare le proprietà di memoria al fine di ridurre l'ampiezza della finestra (la *window size*) di punti in ingresso (anche alla sola osservazione attuale). Tuttavia, alcuni studi preliminari, hanno dimostrato che una rete ricorrente (in particolare la rete di Elman) non è in grado di generalizzare bene nelle regioni dello spazio di valori per gli ingressi per i quali non sono disponibili dati di addestramento. Conseguentemente, al fine di ottenere prestazioni comparabili con quelle di una rete Feed-Forward come una rete MLP-BP, risulta necessario estendere notevolmente l'ampiezza dell'insieme di addestramento in modo aumentare la copertura della regione di variabilità degli ingressi. Di conseguenza, si è deciso di implementare una rete MLP-BP per monitorare lo stato di un generico processo manifatturiero.

Al fine di assicurare una rapida convergenza dell'algoritmo di addestramento Back-Propagation, e per poter ottenere dalla rete neurale le prestazioni desiderate in fase operativa, è necessario stabilire una serie di differenti parametri di configurazione. In genere, e come diffusamente risulta fatto nelle applicazioni riportate in letteratura, i valori di questi parametri sono determinati empiricamente in funzione della specifica applicazione. I principali parametri della rete MLP-BP utilizzata in questa sezione sono riportati nella seguente Tabella 4.1. Tali parametri sono stati decisi dopo alcune simulazioni preliminari nelle quali è stata osservata la loro influenza sulle prestazioni della rete in termini sia di errori di Tipo I sia di errori di Tipo II.

<i>Parametro</i>	<i>Tipo/Valore</i>
Coefficiente di momento	0.4
Coefficiente di apprendimento	0.15
Numero di nodi hidden	24
Numero di nodi in ingresso	32
Funzione di attivazione hidden	Tang. Hp.

Tabella 4.1 – Parametri della rete neurale MLP-BP

La rete neurale esamina 32 punti consecutivi e produce un risultato che indica quale dei particolari tipi di pattern utilizzati in addestramento (quello naturale e lo specifico pattern non naturale) è stato riconosciuto nei dati. Lo strato di uscita della rete neurale è usato per codificare la risposta consistente nella classificazione dell'ingresso come naturale o come non naturale. Lo strato intermedio contiene 24 neuroni. Si è osservato, infatti, che aumentare il numero di nodi intermedi non comporta nessun effettivo miglioramento delle prestazioni, ma al contrario, si traduce in un aumento del tempo totale per il raggiungimento della condizione di convergenza dell'addestramento. La funzione di attivazione utilizzata per lo strato hidden è la tangente iperbolica, mentre nessun tipo di trasformazione è stata utilizzata per lo strato di uscita della rete neurale. L'intera architettura, inoltre, è stata codificata per mezzo del *ToolBox NNET* nell'ambiente operativo MATLAB (MATHWORKS).

4.2.2 L'insieme di dati di addestramento

L'utilizzo di una rete neurale a addestramento supervisionato richiede alcune assunzioni riguardanti il processo monitorato. Questo studio è basato sull'ipotesi che una sequenza *naturale* di dati possa essere modellata attraverso una serie di punti,

distribuiti in modo identico e indipendente con una distribuzione normale di nota media e deviazione standard. Si ipotizza, inoltre, che una sequenza di punti *non naturali* possa essere modellata come un pattern di tipo trend (capitolo 3). Queste due ipotesi sono fondamentali per generare gli insiemi di dati da utilizzare nell'addestramento della rete neurale. Per una rete neurale di tipo supervisionato, infatti, i dati da utilizzare per il training (generati eventualmente attraverso l'uso della simulazione Monte Carlo) devono necessariamente comprendere un numero sufficiente di esempi per ognuna delle classi nelle quali la rete dovrà classificare l'input in fase operativa.

I dati d'ingresso utilizzati in fase di addestramento sono poi stati codificati secondo l'equazione (4.4 in modo da garantire una più rapida convergenza della procedura di training. La procedura di codifica utilizzata divide il range di ogni variabile in ingresso in 64 zone di ampiezza pari a 0.25 unità di deviazione standard, ad ognuna di queste zone è poi associato un numero intero rappresentativo.

$$z = \begin{cases} +8; & \text{if } y > \mu + 8\sigma \\ -8; & \text{if } y < \mu - 8\sigma \\ 0; & \text{if } y = \mu \\ w \left\lfloor \left| \frac{y - \mu}{w\sigma} \right| \right\rfloor \left\lfloor \frac{y - \mu}{|y - \mu|} \right\rfloor; & \text{otherwise} \end{cases} \quad (4.4)$$

μ è la media naturale del processo

σ è la deviazione standard del processo

y è uno degli ingressi mono-dimensionali del processo di codifica

z è la corrispondente codifica dell'ingresso

w è l'ampiezza delle zone di codifica espressa in termini di unità di deviazione standard (in questo studio posto pari a $w = 0.25$)

$\lfloor \cdot \rfloor$ è un operatore che restituisce il più grande intero minore o uguale al suo argomento

4.2.3 Addestramento della rete neurale supervisionata

Durante la fase di addestramento della rete neurale, i pattern d'ingresso e i corrispondenti output-target, appartenenti ad un predeterminato insieme di addestramento, sono presentati in modo casuale alla rete neurale. In particolare, il valore di target per l'output della rete è stato fissato pari a 1 per codificare un trend positivo di crescita delle osservazioni in ingresso, il valore -1 è stato usato per codificare un trend negativo di decrescita delle osservazioni in ingresso, mentre il valore nullo dell'uscita è stato utilizzato per codificare una distribuzione naturale di valori in ingresso.

L'uscita prodotta dalla rete neurale in fase di testing risulta un numero variabile con continuità all'incirca nell'intervallo $[-1,1]$ (ma non è escluso che la rete produca anche valori esterni all'intervallo indicato). Pertanto se in fase operativa la rete produce per un determinato ingresso un'uscita prossima a 0, questa sarà interpretata come indicazione che l'ingresso è considerato naturale, se al contrario, l'uscita della rete è prossima al valore 1 (-1) l'ingresso sarà interpretato come non naturale poiché è stata identificata la presenza di un pattern trend di crescita positivo (decrescita negativa).

I risultati sperimentali descritti nella sezione seguente fanno riferimento ad un addestramento della rete neurale basato sull'uso di un insieme di 8000 vettori, di 32 punti, che includono 4000 pattern naturali, 2000 trend positivo e 2000 di trend negativo. I tassi, in termini di unità di deviazione standard, di crescita (decrescita) per i trend

utilizzati in fase di addestramento sono stati scelti nell'intervallo $[0.025, 0.5]$ ($[-0.5, 0.025]$) con passo pari a 0.025. Pertanto, l'insieme di addestramento comprende 20 classi di trend di crescita e 20 classi di trend di decrescita. Ogni classe di trend, inoltre, è costituita da 100 vettori da 32 dati generati per mezzo della simulazione Monte Carlo.

La fase di addestramento è stata bloccata dopo 1000 iterazioni con un errore di approssimazione dei target di addestramento (Mean Square Error – MSE) ridotto a circa 0.0036.

4.2.4 Risultati sperimentali

Le prestazioni dell'algoritmo neurale supervisionato sopra descritto sono state valutate attraverso l'uso della simulazione Monte Carlo. In particolare, la misura di prestazione considerata è la lunghezza media dei run prima dell'emissione di un allarme: Average Run Length – ARL (Montgomery 2000). Il benchmark di riferimento è rappresentato dalla carta Shewhart–CUSUM con cui si intendono individuare sia piccole che grandi variazioni della media del processo. In particolare, lo schema mono-laterale è ottimizzato per il riconoscimento di una variazione della media in un'unica direzione (o crescita o decrescita). Il valore de ARL della rete è stato stimato su 1000 run di simulazione ognuno consistente di 1032 sequenze di dati dalle quali sono state ricavate finestre di 32 punti consecutivi. E' stato assunto che tutti i dati nella prima finestra di dati (ovvero i primi 32 punti generati dal processo) fossero naturali.

Le uscite prodotte dalla rete neurale, inoltre, sono state valutate sulla base della scelta di un valore di cut-off ($\theta > 0$). In particolare, ogni valore prodotto dalla rete superiore al valore di cut-off θ è considerato un segnale della presenza di un pattern trend di crescita. Analogamente ogni uscita della rete di valore inferiore a $-\theta$ è considerato un segnale della presenza di un trend negativo.

Nella seguente Tabella 4.2 sono riportati i confronti (per diversi valori di crescita del trend) tra le misure de ARL per il benchmark SPC di riferimento (Shewhart-CUSUM $k = 0.25, h = 8$) e i risultati della rete neurale. Le prestazioni dello schema SPC sono state ottenute dalla letteratura ed in particolare dalle simulazioni riportate in Cheng (1995). La colonna con etichetta " ν ", riporta i rapporti tra i valori de ARL ottenuti dalla rete neurale e i corrispondenti valori de ARL dello schema Shewhart-CUSUM. La colonna con etichetta " ρ " riporta l'analogia misura ricavata da Cheng (1995).

Se il processo è in uno stato di fuori controllo, allora un valore di ν minore o uguale a 1 indica che la rete neurale proposta possiede prestazioni migliori rispetto a quelle presentate dallo schema Shewhart-CUSUM. Inoltre, le prestazioni dello schema proposto possono essere confrontate a quelle riportate da Cheng (1995) comparando le due misure ν e ρ . Infine, per rendere bilanciata la comparazione fra l'algoritmo neurale e le misure di prestazione dello schema Shewhart-CUSUM, il valore del parametro di cut-off è stato scelto in modo da garantire un livello di ARL in controllo fra i due sistemi approssimativamente simile ($\theta = 0.7$).

<i>slope</i>	<i>CUSUM</i>	<i>ARL</i>	<i>dev.St. RL</i>	<i>ν</i>	<i>ρ</i>
0	721	727.8	49.8	1.01	1.00
0.0001	471	299.6	19.5	0.64	0.83
0.0005	260	195.6	15.0	0.75	0.80
0.001	185	119.1	11.4	0.64	0.81
0.002	129	96.7	10.1	0.75	0.81

0.005	76.8	66.8	9.6	0.87	0.84
0.01	51.9	41.2	6.8	0.79	0.84
0.02	35.2	29.6	5.6	0.84	0.84
0.05	21.5	18.7	2.6	0.87	0.82
0.1	14.9	12.4	2.2	0.83	0.82
0.5	6.88	5.7	1.1	0.83	0.82
1	4.93	4.0	0.8	0.81	0.86
3	2.86	2.6	0.7	0.91	0.97

Tabella 4.2 – Confronto fra Shewhart-CUSUM (risultati simulazione riportati da Cheng 1995, tab. 7) e rete neurale (risultati simulazione)

Colonna “p” valori ricavati da Cheng (1995) tab. 7.

4.2.5 Considerazioni sull’uso delle reti supervisionate

L’utilizzo della rete neurale ha consentito di ottenere risultati migliori rispetto a quelli del metodo SPC. Inoltre, la rete neurale proposta, poiché specializzata al riconoscimento di uno specifico pattern non naturale, ha consentito di migliorare le prestazioni rispetto ad una rete non specializzata come quella proposta da Cheng (1995). Quest’ultima, infatti, è stata progettata per il riconoscimento di un ulteriore pattern oltre a quello di trend (il pattern shift).

I risultati ottenuti confermano che l’utilizzo di un sistema di monitoraggio basato su rete neurale supervisionata, specializzato al riconoscimento di un solo specifico andamento non naturale, migliora le prestazioni di un sistema SPC. Tuttavia, lo svantaggio principale di tale tipo di approccio, deriva dal fatto che questo sistema ha un’applicabilità pratica molto limitata. Ciò deriva dalla necessità di effettuare a priori un’ipotesi su tutti i possibili andamenti non naturali che possono influenzare il sistema monitorato. A differenza dei metodi SPC, infatti, la rete neurale proposta non ha nessuna preventivabile risposta nel caso in cui una variazione dello stato del processo si manifesti con un pattern diverso rispetto a quello per il quale la rete è stata addestrata (ovvero il trend).

Un ulteriore svantaggio, che è associato all’utilizzo di una rete neurale supervisionata, è legato all’impossibilità di ottenere un addestramento in fasi successive e di tipo cumulativo. In particolare è necessario predisporre tutti gli esempi di pattern, naturali e non-naturali, in un’unica lista di addestramento da utilizzare in una fase di training non scindibile. Ciò procura un addestramento molto oneroso in termini computazionali.

Per sviluppare un algoritmo che permetta di superare gli svantaggi sopra indicati, si rivela indispensabile indagare su una diversa famiglia di reti neurali: quelle a addestramento non supervisionato. Nelle sezioni seguenti, è descritto l’algoritmo non supervisionato Adaptive Resonance Theory (ART). Tale algoritmo, sarà in seguito analizzato in dettaglio, con l’obiettivo di costruire un efficiente sistema di supporto al monitoraggio della qualità.

4.3 Adaptive Resonance Theory

La teoria Adaptive Resonance Theory (ART), è stata introdotta da Grosseberg e Carpenter come modello matematico per la descrizione delle funzioni fondamentali dell’intelligenza biologica: apprendimento, memoria a breve e a lungo termine, riconoscimento di pattern (Pao 1989 e Hagan 1996). Questa teoria ha condotto ad una serie di differenti modelli tra cui ART1, ART2, Fuzzy ART e Fuzzy ARTMAP (Marriot

e Harrison 1995, Koufakou *et al.* 2001). In questo testo una particolare attenzione sarà dedicata allo studio della rete Fuzzy ART, e al suo impiego per il monitoraggio di qualità di un processo.

Le reti ART si comportano come classificatori di pattern basati su un apprendimento di tipo *competitivo* (Haykin 1999). L'apprendimento competitivo è una strategia d'addestramento che realizza un'operazione di raggruppamento (*clustering*), basata sull'ottimizzazione di una determinata funzione (ad esempio la distanza tra vettori di uno spazio n -dimensionale). Una rete a addestramento competitivo accetta un ingresso, e lo classifica in funzione della distanza che esso presenta dalle categorie memorizzate. La categoria, fra quelle memorizzate, alla quale il pattern d'ingresso si avvicina di più, è quella scelta come categoria di appartenenza. Un ingresso, può essere riconosciuto appartenente ad una categoria, ed una sola fra quelle memorizzate, oppure al contrario, potrebbe non essere riconosciuto associabile a nessuna delle categorie formatesi durante l'addestramento.

Gli ingressi presentati durante l'addestramento, vale a dire i vettori che devono essere raggruppati nelle diverse categorie, sono ricevuti al livello inferiore per essere poi trasmessi ai nodi di quello superiore attraverso connessioni bottom-up. Ogni nodo del livello superiore, definito anche strato competitivo della rete, memorizza le caratteristiche (*template* o *prototipo*) per la specifica categoria appresa. Se la differenza tra il pattern d'ingresso e il *template* di una delle categorie memorizzate (per esempio la distanza) rientra all'interno di una determinata tolleranza (*test di vigilanza*), il pattern d'ingresso sarà classificato nella medesima categoria. In tal caso, le caratteristiche critiche del pattern in ingresso saranno incluse nel prototipo della categoria memorizzata. In altri termini, le categorie in precedenza memorizzate, sono in fase di addestramento, modificate attraverso pesi adattativi per includere le caratteristiche sia dei precedenti, sia degli attuali pattern d'ingresso (*plasticità della rete*). Nel livello superiore, il nodo con la minima distanza tra il pattern d'ingresso, e la categoria-prototipo cui è associato, è definito il vincitore della competizione, e il pattern d'ingresso è assunto come un membro del *cluster* rappresentato dal nodo vincitore. La modifica dei pesi non è consentita se l'errore non è all'interno della specifica tolleranza. Se, infatti, il pattern d'ingresso non corrisponde a nessuna delle categorie precedentemente apprese, esso sarà memorizzato come una nuova categoria (*stabilità della rete*).

Il test di vigilanza è finalizzato a verificare che il pattern d'ingresso richiami, con un determinato grado di similitudine, il prototipo del gruppo vincitore. Il grado di similitudine è governato da un attributo fondamentale del funzionamento di una rete ART: il *parametro di vigilanza*. Un'alta vigilanza impone un criterio di conformità molto rigoroso, che come conseguenza suddivide l'insieme degli input in categorie molto dettagliate. Al contrario, un basso valore di vigilanza tollera grandi errori, provocando di conseguenza, categorie molto grezze.

Ci sono due classi generali di reti ART: le ART binarie e le ART analogiche. ART1 è utilizzabile unicamente per classificare input binari, mentre la rete Fuzzy ART è impiegabile per pattern analogici. Dato che l'algoritmo Fuzzy ART è una generalizzazione di quello ART1, un approfondito review di quest'ultima architettura è fornito nelle seguenti sezioni.

4.3.1 ART1

La rete neurale ART1 è costituita, essenzialmente, da due principali elementi: il sistema di *attenzione* e quello di *orientamento*. Nel sottosistema d'attenzione, sono elaborati e memorizzati pattern ritenuti familiari con quelli in precedenza appresi

(*plasticità della rete*). Il sottosistema d'orientamento, invece, fornisce la funzione di controllo necessaria per l'apprendimento e per la classificazione dei pattern di ingresso. Il sottosistema d'orientamento, è in grado di azzerare l'attività del sottosistema d'attenzione qualora siano presentati pattern non familiari, evitando in pratica che nuovi apprendimenti possano provocare una cancellazione di quelli effettuati in precedenza (*stabilità della rete*).

Due strati di neuroni $F1$ (strato inferiore definito *livello di comparazione*) e $F2$ (strato superiore definito *di riconoscimento*), interamente collegati attraverso link pesati di tipo bottom-up e top-down, costituiscono il sottosistema d'attenzione. I pesi di tipo bottom-up e top-down tra $F1$ e $F2$ possono essere aggiornati in modo adattativo come risposta a pattern d'ingresso: ciò consiste proprio nell'addestramento della rete. I pesi rappresentano la memoria a lungo termine (Long Term Memory – LTM), perché essi rimangono nella rete per un periodo esteso.

In termini molto semplificati, il livello di comparazione ($F1$), che riceve pattern d'ingresso esterni alla rete, agisce come un estrattore di caratteristiche essenziali. Lo strato di riconoscimento ($F2$) ricevendo input interni alla rete dal livello di comparazione, funziona come un classificatore di categorie. L'applicazione di un singolo vettore d'ingresso, determina un'attività neurale che porta alla creazione di un pattern-prototipo in entrambi i livelli $F1$ e $F2$. Questi pattern sono noti come la memoria a breve termine (Short Term Memory – STM), perché permangono nel sistema unicamente durante l'applicazione dello specifico ingresso. L'attività neurale dei nodi $F2$ può rinforzare quella dei nodi in $F1$ grazie alle connessioni top-down (vale a dire che può condurre ad una corrispondenza fra i pattern presenti in $F1$ e $F2$). Lo scambio d'informazioni, sia bottom-up sia top-down, prosegue fino ad una *risonanza* nell'attività neurale della rete (in pratica, fino all'identificazione della corrispondenza ricercata). Il sottosistema d'orientamento genera un segnale d'annullamento verso $F2$ quando il pattern d'ingresso bottom-up e il pattern di template (ovvero il prototipo) top-down non corrispondono in accordo al parametro di vigilanza. In base al parametro di vigilanza, se è identificato un errore tra pattern in $F1$ e template generato da $F2$, allora il nodo vincitore del livello competitivo sarà de-attivato e si procederà alla ricerca di un ulteriore cluster rappresentativo per l'ingresso attuale. Se tale procedura non permetterà di individuare nessun cluster rappresentativo per l'ingresso attuale, in altri termini, una volta che è identificato che il pattern d'ingresso è nuovo, la rete modificherà la propria struttura attraverso l'immediata memorizzazione delle novità in pesi aggiuntivi.

Il segnale d'annullamento, se inviato, fermerà qualunque attività che avviene in $F2$, preparando $F2$ a memorizzare il pattern in ingresso come una nuova categoria. Se, al contrario, il segnale d'annullamento non è trasmesso, il prototipo in precedenza codificato, e associato con il nodo di categoria che rappresenta il migliore confronto con l'attuale ingresso, sarà aggiornato con il nuovo pattern. In altri termini, la rete cambierà i pesi del nodo vincitore in modo da comprendere anche le caratteristiche essenziali del nuovo pattern appreso.

4.3.2 Una sintesi dell'algoritmo ART1

Denotiamo un nodo nel livello $F1$ attraverso l'indice $i \in \{1, 2, \dots, M\}$ e un nodo in $F2$ attraverso l'indice $j \in \{1, 2, \dots, N\}$. Ogni nodo i nel livello $F1$ è connesso attraverso un peso bottom-up con ogni nodo j nello strato $F2$: questo peso è denotato con W_{ij} . Inoltre, ogni nodo j nel livello $F2$ è connesso con un peso top-down con ogni nodo i nel livello

$F1$; questo peso è denotato con w_{ji} . Il vettore, i cui elementi sono uguali ai pesi top-down che partono dal nodo j nel livello $F2$ è rappresentato da \underline{w}_j (il *template*). Si ha che $\underline{w}_j = (w_{j1}, w_{j2} \dots w_{jM})$ per $j = 1, 2 \dots N$. Il vettore, dei pesi bottom-up, convergenti verso il nodo j nel livello $F2$, è designato come \underline{W}_j . In tal caso si ha che $\underline{W}_j = (W_{ij}, W_{2j} \dots W_{Mj})$ per $j = 1, 2 \dots N$. I valori iniziali dei pesi bottom-up e top-down sono indicati con $W_{ij}^{(0)}$ e $w_{ji}^{(0)}$ rispettivamente. Si utilizzerà, inoltre, la notazione $|\underline{x}|$ per denotare l'ampiezza del vettore \underline{x} definita come la somma dei suoi elementi, ovvero: $|\underline{x}| = \sum_i x_i$. Denotiamo il vettore d'ingresso alla rete costituito da M componenti binarie (la rete ART1 richiede unicamente ingressi binari) con la seguente notazione.

$$\underline{I} = (I_1, I_2, \dots, I_M); \quad I_i = 0, 1 \forall i = 1, 2 \dots M; \tag{4.5}$$

Denotiamo il vettore dei valori in uscita allo strato $F1$, detto l'attività neurale di $F1$, con la seguente notazione.

$$\underline{x} = (x_1, x_2, \dots, x_M); \quad x_i = 0, 1 \forall i = 1, 2 \dots M; \tag{4.6}$$

Nella fase iniziale dell'algoritmo (passo 1) è posto $\underline{x} = \underline{I}$ ma nelle successive fasi il valore di \underline{x} varia per effetto dell'attività di elaborazione realizzata nel livello $F1$. Il suo valore è dunque svincolato dall'ingresso che, invece, rimane costante durante tutta l'elaborazione dell'algoritmo ART. L'architettura della rete neurale ART1 è illustrata nella Figura 4.1.

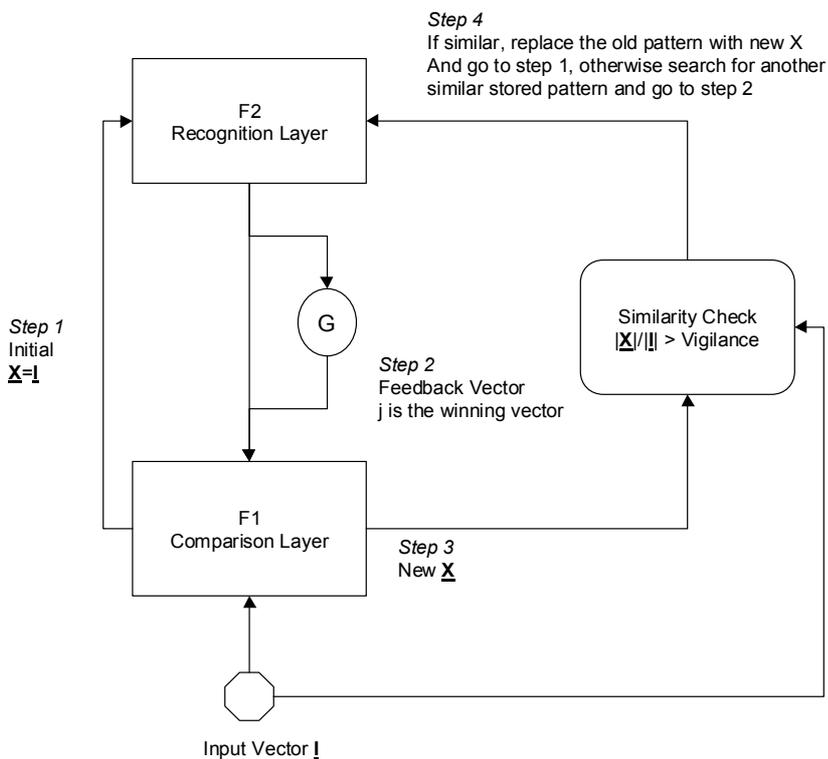


Figura 4.1 – Schematizzazione dell'algoritmo ART

Nel livello di comparazione $F1$, ogni unità d'elaborazione (di indice $i \in \{1, 2 \dots M\}$) ha tre ingressi: 1) un segnale di feedback dal livello di riconoscimento (ovvero un componente w_{ji} di uno specifico \underline{w}_j), 2) un segnale di guadagno G , 3) il corrispondente elemento fra gli M del vettore d'ingresso \underline{I} . Il segnale di guadagno G è zero se nessun

elemento del vettore \underline{w}_j è 1. L'uscita, dei neuroni nel livello di comparazione, è pari a 1 se almeno due dei suoi tre input è pari a 1. Ciò crea la cosiddetta *regola dei due terzi* che determina nella rete ART1 la capacità di prendere una decisione. Attraverso l'implementazione della regola dei due terzi, se G è pari a 0, l'uscita del livello di comparazione è semplicemente l'operazione logica di AND (indicata con il simbolo \wedge) di ogni componente d'ingresso \underline{I} e \underline{w}_j , risultando nel seguente vettore:

$$\underline{x} = \underline{I} \wedge \underline{w}_j \text{ dove } \underline{I} \wedge \underline{w}_j = (I_1 \cdot w_{j1}, \dots, I_i \cdot w_{ji} \dots I_M \cdot w_{jM}). \quad (4.7)$$

4.3.2.1 Primo passo

Inizialmente, il vettore binario \underline{I} è applicato al livello di comparazione e lo attraversa senza alcun cambiamento, divenendo, semplicemente il vettore binario \underline{x} . Non avviene nessun tipo d'elaborazione poiché G è posto uguale a 1, e tutti gli elementi del vettore \underline{w}_j sono nulli. Il vettore \underline{x} poi, diviene l'ingresso del livello di comparazione $F2$. I pesi corrispondenti a questi ingressi formano il vettore \underline{W}_j per il j^{esimo} neurone nel livello di riconoscimento. Gli elementi del vettore \underline{W}_j sono, solitamente, inizializzati ($W_{ij}^{(0)}$) a valori molti piccoli (prossimi a zero). Il risultato calcolato dalla rete nel livello $F2$ per ogni nodo j è ottenuto come prodotto scalare tra \underline{W}_j e \underline{x} (ovvero ampiezza del vettore di prodotto logico AND, in formula $|\underline{W}_j \wedge \underline{x}|$). Il neurone con il più alto valore diviene quello vincitore della competizione.

4.3.2.2 Secondo passo

Supponendo di indicare con j l'indice del nodo vincitore, allora il vettore \underline{w}_j è inoltrato indietro verso il livello di comparazione. Il valore di G è posto pari a zero se nessun elemento di \underline{w}_j è posto 1, conseguentemente l'applicazione della regola dei due terzi fornisce l'operazione logica di AND degli elementi dei vettori \underline{w}_j e \underline{I} risultando in un nuovo valore per \underline{x} ($\underline{x} = \underline{I} \wedge \underline{w}_j$).

4.3.2.3 Terzo passo

Assumiamo che il nodo j in $F2$ sia stato attivato attraverso questo processo. L'attivazione del nodo j in $F2$, indica che esso può essere considerato come il candidato potenziale dalla rete ART1 per rappresentare il pattern in ingresso. L'esame del rapporto $\frac{|\underline{I} \wedge \underline{w}_j|}{|\underline{I}|} = \frac{|\underline{x}|}{|\underline{I}|}$ consente di valutare la conformità di questo nodo. Se questo rapporto è più piccolo di un determinato fattore ρ , allora il nodo j è considerato non appropriato a rappresentare il pattern d'ingresso e, come conseguenza, esso è annullato (ovvero de-attivato). Il fattore ρ è chiamato *parametro di vigilanza*, ed esso assume valori nell'intervallo $[0,1]$. Il processo di de-attivazione è realizzato dal sottosistema d'orientamento ed in particolare dal nodo d'annullamento (*reset node*). Se avviene un annullamento, un altro nodo in $F2$ (differente dal nodo j) è scelto per rappresentare il pattern d'ingresso; l'annullamento del nodo precedente permane per l'intera presentazione del pattern. Il processo precedente è ripetuto fino a che non è trovato un nodo appropriato in $F2$, oppure fino a che non sono stati esaminati tutti i nodi in $F2$.

4.3.2.4 Quarto passo

Assumendo che il nodo j è scelto per rappresentare l'ingresso specifico, il corrispondente vettore dei pesi top-down \underline{w}_j^{old} diventa uguale a \underline{w}_j^{new} . Assumendo, inoltre, un apprendimento veloce (ovvero parametro di apprendimento posto a valore 1), i pesi assumono il valore:

$$\underline{w}_j^{new} = \underline{I} \wedge \underline{w}_j^{old} \quad (4.8)$$

Inoltre, il corrispondente vettore di pesi bottom-up \underline{W}_j^{old} diventa uguale a \underline{W}_j^{new} dove:

$$\underline{W}_j^{new} = \frac{\underline{I} \wedge \underline{w}_j^{old}}{\alpha + |\underline{I} \wedge \underline{w}_j^{old}|} \quad (4.9)$$

Dove α , è un parametro caratteristico dell'algoritmo (definito *parametro di scelta*) assumente valori non negativi. Tenuto conto che gli ingressi possono assumere unicamente valori binari (0 o 1), e che risulta $\underline{x} = \underline{I} \wedge \underline{w}_j$, allora le precedenti equazioni (4.8 e 4.9) possono essere riscritte rispettivamente come segue.

$$w_{ji} = \begin{cases} 1 & \text{if } i : x_i = 1 \\ 0 & \text{otherwise} \end{cases} \quad (4.10)$$

$$W_{ij} = \begin{cases} \frac{1}{\alpha + |\underline{x}|} & \text{if } i : x_i = 1 \\ 0 & \text{otherwise} \end{cases} \quad (4.11)$$

O, equivalentemente, come segue:

$$\underline{w}_j = \underline{x} \quad (4.12)$$

$$\underline{W}_j = \frac{\underline{x}}{\alpha + |\underline{x}|} = \frac{\underline{w}_j}{\alpha + |\underline{w}_j|} \quad (4.13)$$

Cioè risulta che, i pesi del template \underline{w}_j del nodo vincitore j che corrispondono a 1 in \underline{x} , vengono posti a valore 1, mentre gli altri pesi sono forzati a 0, incrementando ulteriormente le possibilità che vettori simili siano identificati dal particolare neurone.

Si può osservare che i pesi bottom-up, sono usati per scegliere le categorie, mentre, i pesi top-down, sono usati per determinare quanto accuratamente la categoria prescelta corrisponde al pattern d'ingresso. Se la corrispondenza non è sufficiente per l'attuale livello di vigilanza, allora il sistema d'orientamento annulla la categoria.

4.3.2.5 Osservazione

Un dato nodo j nel livello $F2$ assume il seguente ingresso dallo strato $F1$:

$$t_j = |\underline{W}_j \wedge \underline{x}| \text{ ovvero } t_j = \sum_{i=1}^M W_{ij} x_i \quad (4.14)$$

O equivalentemente

$$t_j = \frac{1}{\alpha + |\underline{w}_j|} \sum_{i=1}^M w_{ij} x_i = \frac{1}{\alpha + |\underline{w}_j|} \sum_{i=1}^M w_{ij} I_i w_{ij} = \frac{1}{\alpha + |\underline{w}_j|} \sum_{i=1}^M w_{ij} I_i \quad (4.15)$$

Dato che i vettori w_{ij} e I_i possono assumere solo valori di 0 e 1, i termini della somma della equazione (4.15) è semplicemente il numero di 1 che i vettori \underline{w}_j e \underline{I} hanno nelle posizioni corrispondenti. Pertanto, l'equazione (4.14) può essere riscritta come segue:

$$t_j = \frac{|\underline{I} \wedge \underline{w}_j|}{\alpha + |\underline{w}_j|} \quad (4.16)$$

In altri termini, i vettori top-down attivano un processo di competizione, fra i nodi $F2$, che eventualmente porta all'attivazione di un singolo neurone in $F2$, vale a dire il nodo che riceve il massimo valore bottom-up in ingresso da $F1$. La scelta della categoria è fatta attraverso la selezione del neurone in $F2$ con il massimo valore in t_j . Pertanto, la funzione di scelta può essere ridefinita come:

$$T_j(\underline{I}) = \frac{|\underline{I} \wedge \underline{w}_j|}{\alpha + |\underline{w}_j|} \quad (4.17)$$

Per piccoli valori di α , $T_j(\underline{I})$ è approssimativamente una misura del numero di 1 che si sovrappongono tra \underline{I} e \underline{w}_j , in altre parole è la frazione di 1 in \underline{w}_j che sono anche in \underline{I} .

4.4 Fuzzy ART

L'algoritmo neurale Fuzzy ART rappresenta una generalizzazione delle equazioni precedenti; la generalizzazione è ottenuta per mezzo dell'utilizzo della teoria dei Fuzzy Set (Georgeopoulos *et al.* 1996, 1999, Blume e Esener 1997, Burwick and Joublin 1998, Kim *et al.* 2001). Gli ingressi della rete possono assumere valori continui variabili tra 0 e 1. Analogamente ogni connessione presente all'interno della rete assumerà valori compresi nell'intervallo $[0,1]$. Questa caratteristica dell'algoritmo richiede di utilizzare sui generici dati di ingresso una preventiva pre-elaborazione che permetta di rappresentare gli input attraverso numeri appartenenti all'intervallo $[0,1]$.

In questa sezione, sarà presentato in dettaglio l'algoritmo neurale Fuzzy ART, con particolare riferimento alla rete proposta per l'applicazione di monitoraggio della qualità discussa nei successivi capitoli. Prima di introdurre l'algoritmo Fuzzy ART è necessario premettere la nomenclatura utilizzata nel seguito.

4.4.1 Nomenclatura

Sia $\underline{x} = [x_1, \dots, x_i, \dots]$ un generico vettore di uno spazio n -dimensionale i cui elementi siano tutti compresi nell'intervallo $[0,1]$. Definiamo con la notazione $|\underline{x}|$ l'ampiezza del vettore \underline{x} intesa come la somma dei propri componenti presi in valore assoluto, ovvero: $|\underline{x}| = \sum_i |x_i|$.

Sia $\underline{y} = [y_1, \dots, y_i, \dots]$ un ulteriore vettore del medesimo spazio n -dimensionale (con componenti compresi all'interno dell'intervallo $[0,1]$). Definiamo con la notazione $\underline{x} \wedge \underline{y}$ il vettore il cui i^{esimo} componente è il minimo fra i^{esimo} componente del vettore \underline{x} e l' i^{esimo} componente del vettore \underline{y} . Ovvero: $\underline{x} \wedge \underline{y} = [\min(x_1, y_1), \dots, \min(x_i, y_i), \dots]$. L'operatore \wedge è definito il minimo nella teoria dei Fuzzy Set.

Definiamo l'operatore di massimo (\vee) fra due vettori $\underline{x}, \underline{y}$ come il vettore tale che $\underline{x} \vee \underline{y} = [\max(x_1, y_1), \dots, \max(x_i, y_i), \dots]$.

Infine definiamo come distanza fra due vettori $\underline{x}, \underline{y}$, la quantità: $dis(\underline{x}, \underline{y}) = |\underline{x} \vee \underline{y}| - |\underline{x} \wedge \underline{y}|$. Applicando la definizione degli operatori di massimo e minimo è immediato verificare che $dis(\underline{x}, \underline{y}) = \sum_i |x_i - y_i|$.

Nella seguente Figura 4.2 è fornita una rappresentazione geometrica dei concetti sopra esposti con riferimento al caso semplificato di uno spazio a due dimensioni.

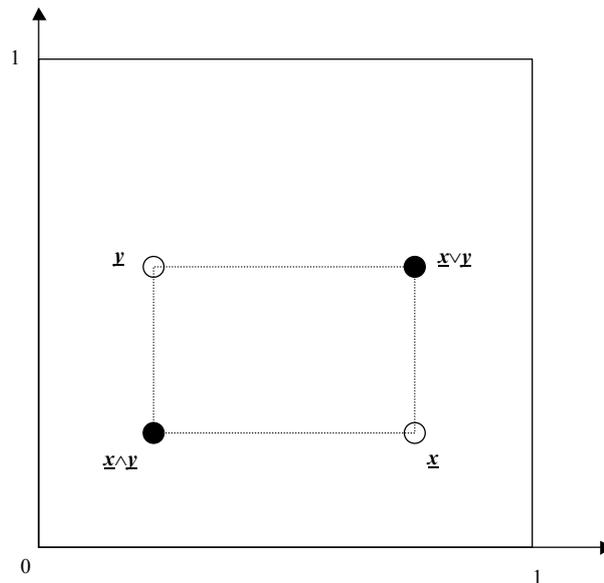


Figura 4.2 – Rappresentazione geometrica nel caso di uno spazio bidimensionale

In particolare, nella Figura 4.2 sono stati rappresentati (nel caso esemplificativo bidimensionale) i valori di massimo e di minimo fra i due punti riportati nel grafico $\underline{x}, \underline{y}$. Il valore di distanza $dis(\underline{x}, \underline{y}) = |\underline{x} \vee \underline{y}| - |\underline{x} \wedge \underline{y}|$ equivale a metà del perimetro del rettangolo tratteggiato nella stessa figura (la somma delle lunghezze di due lati non paralleli). Questo è la minima distanza per congiungere i punti $\underline{x}, \underline{y}$ seguendo le direzioni degli assi cartesiani.

4.4.2 Algoritmo Fuzzy ART

L'architettura di rete Fuzzy ART, che è proposta nel presente lavoro per il monitoraggio della qualità, è schematicamente rappresentata dalla seguente Figura 4.3. Essa consiste di due principali sottosistemi, quello di attenzione e di orientamento. Il sottosistema di attenzione è composto da tre strati di neuroni denotati come $F0, F1$ e $F2$. Il sottosistema di orientamento, invece, è composto di un singolo nodo definito nodo di reset il cui output influisce sui nodi del livello $F2$. L'architettura è stata implementata utilizzando l'ambiente software NeuralWorks Professional II Plus versione 5.40 (NeuralWare).

Nello strato $F0$ è eseguita una pre-elaborazione del vettore di M componenti in ingresso $\underline{I} = [I_1, I_2, \dots, I_{M-1}, I_M]$. Questa pre-elaborazione, chiamata *Complement Coding* trasforma un vettore di ingresso di dimensione M , (i cui componenti sono compresi nell'intervallo $[0,1]$), nel seguente vettore di uscita di dimensione $2M$.

$$\underline{I}^c = (\underline{I}, \underline{I} - \underline{I}) = [I_1, I_2, \dots, I_M, 1 - I_1, 1 - I_2, \dots, 1 - I_M] \quad (4.18)$$

Il vettore in uscita $\underline{I}_i^c = (\underline{I}, \underline{I} - \underline{I})$ costituisce l'ingresso per livello successivo di neuroni $F1$. Il numero di neuroni nello strato $F0$ corrisponde alla dimensione del vettore in ingresso (ovvero M), mentre il numero di nodi nel livello $F1$ è pari a $2M$.

Nella notazione seguente sarà utilizzato l'indice j per indicare un generico neurone del livello $F2$. Il valore dell'indice j , quindi, assume valori interi varianti fra 1 e il numero di neuroni scelti in fase di progettazione per lo strato $F2$. Il numero di neuroni dello strato $F2$ rappresenta un parametro di progettazione della rete.

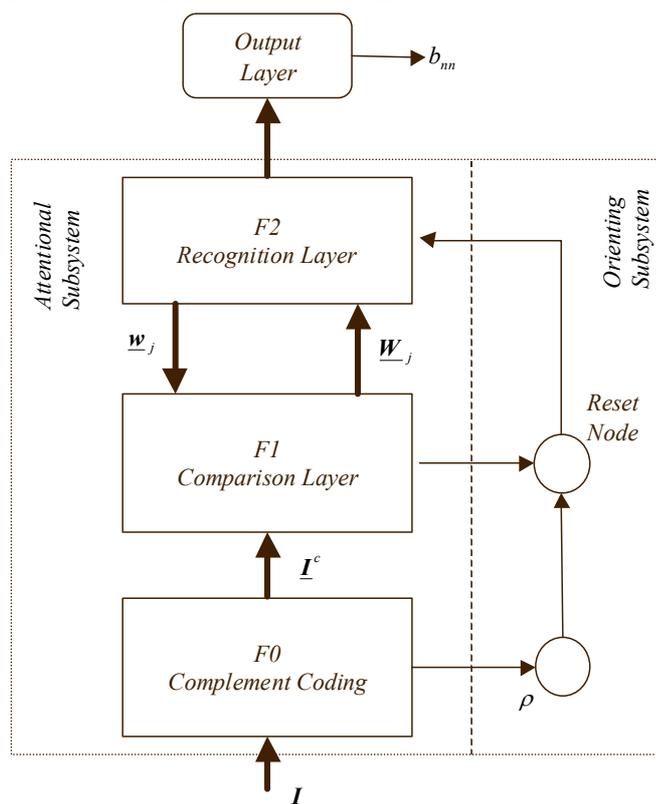


Figura 4.3 – Schematizzazione della rete neurale Fuzzy ART

I due strati di neuroni $F1$ e $F2$ sono connessi da link pesati sia di tipo bottom-up (ovvero diretti dai nodi del livello inferiore $F1$ verso quelli del livello superiore $F2$), sia da link di tipo top-down (ovvero diretti dai nodi del livello superiore $F2$ a quelli del livello inferiore $F1$). Il vettore di $2M$ componenti uguali rispettivamente ai pesi top-down che connettono il nodo j^{esimo} dello strato $F2$ ad ognuno dei $2M$ neuroni dello strato $F1$ saranno indicati con la notazione \underline{w}_j . Inoltre, il vettore, i cui componenti sono uguali ai pesi bottom-up che convergono nel neurone di indice j nello strato $F2$, è indicato con \underline{W}_j (vedi Figura 4.3). I valori iniziali dei pesi top-down e bottom-up, per tutte le connessioni fra i nodi dello strato $F1$ e $F2$, sono uguali rispettivamente a 1 e a $\frac{1}{\alpha + 2M}$ (α è il parametro di scelta e può assumere un qualunque valore non negativo). I valori iniziali dei pesi sono quelli che presentano le connessioni della rete prima che ne sia avviata la fase di addestramento.

Uno strato aggiuntivo di neuroni è stato incluso nell'architettura neurale proposta (Figura 4.3). Esso consiste di un unico neurone che fornisce in uscita un segnale binario

(indicato con $b_{m,t}$). Il valore assunto dal neurone di uscita (0 o 1) dipende dal neurone vincitore del livello $F2$. In particolare, l'associazione fra i neuroni e valori di uscita, dipende da un'informazione di mapping che è fornita in fase di addestramento.

Assumiamo che un pattern \underline{I} (i cui componenti sono compresi nell'intervallo $[0,1]$) sia presentato all'ingresso della rete neurale. Il corrispondente pattern \underline{I}^c prodotto dall'elaborazione di *Complement Coding* effettuata nel livello $F0$, attraversa il livello $F1$ e produce ingressi bottom-up che influiscono sui nodi del livello $F2$. L'ingresso bottom-up, per ogni singolo neurone j del livello $F2$, è rappresentato dalla seguente equazione.

$$T_j(\underline{I}^c) = \frac{|\underline{I}^c \wedge \underline{w}_j|}{\alpha + |\underline{w}_j|} \quad (4.19)$$

Questa funzione è definita Funzione di Scelta (*Choice function*), essa è dipendente dal parametro di scelta α che può assumere un qualunque valore non negativo.

Un nodo j dello strato $F2$ è definito *uncommitted* se tutti i pesi dei link top-down sono uguali ai valori iniziali (ovvero uguali a 1), altrimenti il nodo è definito *committed*. Pertanto, nel caso che un nodo j sia *uncommitted* allora il relativo valore della funzione di scelta, si riduce al seguente:

$$T_j(\underline{I}^c) = \frac{|\underline{I}^c|}{\alpha + 2M} \quad (4.20)$$

Gli input bottom-up avviano un processo di competizione, fra i neuroni del livello $F2$, che porta all'attivazione di un singolo neurone in particolare quello che riceve il massimo valore di ingresso bottom-up dallo strato $F1$. In altri termini sarà scelto il neurone di indice j tale da verificare la seguente condizione

$$j = \arg \max_j T_j(\underline{I}^c) \quad (4.21)$$

Avendo indicato con J l'insieme degli indici dei neuroni nello strato $F2$. Assumiamo che un generico nodo di indice j risulti vincitore nello strato $F2$ (*winner-takes-all*) ciò implica che questo neurone può essere considerato un potenziale candidato a rappresentare il pattern di ingresso. La corrispondenza dell'associazione fra ingresso e il nodo vincitore della competizione, è valutata attraverso l'analisi del seguente rapporto.

$$\frac{|\underline{I}^c \wedge \underline{w}_j|}{|\underline{I}^c|} \quad (4.22)$$

Se tale rapporto è inferiore al parametro di vigilanza, ovvero se $\frac{|\underline{I}^c \wedge \underline{w}_j|}{|\underline{I}^c|} < \rho$, allora il

nodo corrispondente è considerato non appropriato e dunque è resettato (de-attivato). Il parametro ρ è il parametro di vigilanza e assume valori compresi nell'intervallo $[0,1]$. La verifica di vigilanza sopra indicata è eseguita nel sottosistema di orientamento della rete, e l'eventuale de-attivazione del nodo j è eseguita dal nodo di reset (vedi Figura 4.3). Se è eseguito un reset allora la procedura di ricerca di un nodo rappresentativo, è ripetuta escludendo i nodi che sono stati valutati in precedenza. La procedura prosegue fino all'identificazione di un nodo adeguato (ovvero tale da superare il test di vigilanza), oppure, fino alla completa valutazione di tutti i nodi del livello $F2$.

Durante la fase di addestramento, alla procedura di ricerca e di valutazione del più adeguato fra i cluster presenti nella rete per rappresentare l'input attuale segue una fase di modifica delle connessioni top-down. In particolare, durante la fase di addestramento, una volta che un ingresso risulta associato ad un determinato cluster della rete, le connessioni corrispondenti sono opportunamente modificate in modo da includere le caratteristiche del vettore in ingresso. Quando un nodo è assegnato per la prima volta, è utilizzata un'equazione di rapido assegnamento, mentre quando un nodo è già stato assegnato in precedenza, sarà utilizzata un'equazione di ri-codifica lenta. Se il nodo j è stato scelto per rappresentare lo specifico ingresso, durante la fase di addestramento il corrispondente vettore dei pesi top-down \underline{w}_j^{old} diviene uguale a \underline{w}_j^{new} . L'equazione di rapido assegnamento e di ri-codifica lenta sono rispettivamente le seguenti.

- $\underline{w}_j^{new} = \underline{I}^c$ quando un nodo in $F2$ è assegnato per la prima volta;
- $\underline{w}_j^{new} = \beta(\underline{I}^c \wedge \underline{w}_j^{old}) + (1 - \beta)\underline{w}_j^{old} \quad 0 \leq \beta \leq 1$. quando un nodo è stato già assegnato. (4.23)

Piccoli valori di β (tasso di apprendimento), fanno sì che il sistema basi i suoi risultati su una media di lungo periodo della propria esperienza, mentre valori di β prossimi a 1 consentono adattamenti in ambienti rapidamente mutevoli.

Il parametro del tasso di apprendimento assicura che il nuovo vettore di pesi sia una combinazione convessa tra il vettore risultante in $F1$, e il template della categoria attesa nel livello $F2$. Per $\beta = 1$, condizione nota come Fast-Commit-Fast-Recode (FCFR), i vettori risultanti da $F1$ vengono a sostituire direttamente gli esemplari di categoria presenti.

Assumendo che il nodo j sia stato scelto per rappresentare l'uscita specifica, il corrispondente vettore di pesi bottom-up \underline{W}_j^{old} diventa uguale a \underline{W}_j^{new} dove

$$\underline{W}_j^{new} = \frac{\underline{I}^c \wedge \underline{w}_j^{old}}{\alpha + |\underline{I}^c \wedge \underline{w}_j^{old}|} \quad (4.24)$$

Può essere notato che, assumendo un tasso di ri-codifica β uguale a 1 e supponendo che $I_i = 0, 1 \forall i = 1, 2, \dots, M$; (ovvero non esistono valori incerti), allora il modello della rete Fuzzy ART diviene uguale al modello matematico della rete ART1.

4.4.3 Interpretazione geometrica dell'algoritmo

In questa sezione si richiama brevemente l'interpretazione geometrica dell'algoritmo Fuzzy ART. Per ulteriori dettagli il lettore è rimandato ai lavori di Georgeopoulos *et al.* (1996, 1999).

In particolare, sia \underline{w}_j il vettore a $2M$ componenti corrispondenti ai pesi top-down relativi al nodo j^{esimo} del livello $F2$ di una rete Fuzzy ART. Supponiamo che durante la fase di addestramento i componenti del vettore \underline{w}_j siano stati modificati attraverso l'algoritmo Fuzzy ART per codificare p vettori dell'insieme addestramento. Indicando con:

$$\underline{I}^c(1) = (\underline{I}(1), \underline{I} - \underline{I}(1)), \underline{I}^c(2) = (\underline{I}(2), \underline{I} - \underline{I}(2)), \dots, \underline{I}^c(p) = (\underline{I}(p), \underline{I} - \underline{I}(p))$$

l'insieme dei p vettori codificati nel nodo j , allora si può dimostrare che \underline{w}_j può essere riscritto in funzione dell'insieme dei vettori in ingresso secondo la seguente relazione.

$$\underline{w}_j = \underline{I}^c(1) \wedge \underline{I}^c(2) \wedge \dots \wedge \underline{I}^c(p) = \left(\bigwedge_{i=1}^p \underline{I}(i), \bigwedge_{i=1}^p (\underline{I} - \underline{I}(i)) \right) = \left(\bigwedge_{i=1}^p \underline{I}(i), \underline{I} - \left\{ \bigvee_{i=1}^p \underline{I}(i) \right\} \right)$$

ovvero come segue

$$\underline{w}_j = (\underline{u}_j, \underline{I} - \underline{v}_j) \text{ dove } \underline{u}_j = \bigwedge_{i=1}^p \underline{I}(i) \text{ e dove } \underline{v}_j = \bigvee_{i=1}^p \underline{I}(i) \quad (4.25)$$

Pertanto, ogni template dello spazio a $2M$ dimensioni \underline{w}_j è univocamente rappresentato da una coppia di vettori dello spazio ad M dimensioni $(\underline{u}_j, \underline{v}_j)$. I due vettori $(\underline{u}_j, \underline{v}_j)$ rappresentano rispettivamente il minimo e il massimo, secondo la definizione della teoria dei Fuzzy Set, tra tutti gli ingressi che la rete ha associato al cluster j^{esimo} .

Nel seguito, la regione dello spazio a M dimensioni associata al generico template j , sarà rappresentata attraverso la notazione $R_j = \{\underline{u}_j, \underline{v}_j\}$ ovvero dall'iperrettangolo di tutti i punti $\underline{x} = [x_1, \dots, x_i, \dots, x_M]$ tali che risulti la seguente relazione.

$$\begin{aligned} \underline{x} &= [x_1, \dots, x_i, \dots, x_M]; \underline{u}_j = [u_{j1}, \dots, u_{ji}, \dots, u_{jM}]; \underline{v}_j = [v_{j1}, \dots, v_{ji}, \dots, v_{jM}] \\ \underline{x} \in R_j = \{\underline{u}_j, \underline{v}_j\} &\Leftrightarrow \forall i = 1, 2, \dots, M : 0 \leq u_{ji} \leq x_i \leq v_{ji} \leq 1 \end{aligned} \quad (4.26)$$

Nella seguente Figura 4.4 è fornita una rappresentazione geometrica esemplificativa nel caso bidimensionale.

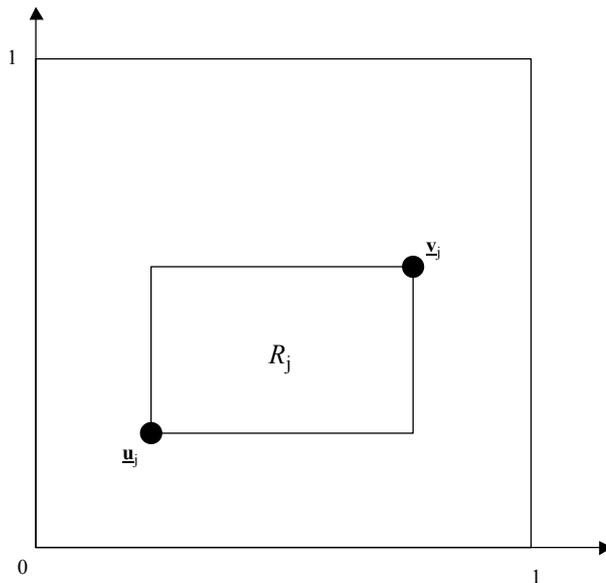


Figura 4.4 – Interpretazione geometrica di un cluster della rete Fuzzy ART

Caso semplificato bidimensionale, neurone j.

Consideriamo ora un ingresso $\underline{I} = [I_1, I_2, \dots, I_M]$ ed il corrispondente vettore ottenuto dalla funzione di *Complement Coding*.

$$\underline{I}_c = [I_1, I_2, \dots, I_M, 1 - I_1, 1 - I_2, \dots, 1 - I_M] \quad (4.27)$$

Supponiamo che sia identificato il template j come nodo rappresentativo: $\underline{w}_j = [w_1, \dots, w_M, w_{M+1}, \dots, w_{2M}]$, è già stato osservato che il template può essere riscritto come segue.

$$\underline{w}_j = (\underline{u}_j, \underline{I} - \underline{v}_j) = [u_1, \dots, u_M, 1 - v_1, \dots, 1 - v_M] \quad (4.28)$$

Definiamo come ampiezza dell'iperrettangolo $R_j = \{\underline{u}_j, \underline{v}_j\}$ associato al template j la seguente quantità: $|R_j| = |\underline{v}_j - \underline{u}_j|$ ovvero $|R_j| = \sum_{i=1}^M |v_{ji} - u_{ji}|$. Nel caso esemplificativo bidimensionale illustrato in precedenza, questa corrisponde alla metà del perimetro del rettangolo $R_j = \{\underline{u}_j, \underline{v}_j\}$ ovvero alla somma di due lati non paralleli. Definiamo, inoltre, con la notazione $R_j \oplus \underline{I}$ l'iperrettangolo ottenuto da R_j composto con il vettore a M dimensioni \underline{I} , come l'iperrettangolo i cui estremi sono rappresentati da:

$$R_j \oplus \underline{I} = \{\underline{I} \wedge \underline{u}_j, \underline{I} \vee \underline{v}_j\} \quad (4.29)$$

L'ampiezza dell'iperrettangolo $R_j \oplus \underline{I}$ è quindi pari a

$$|R_j \oplus \underline{I}| = |(\underline{I} \vee \underline{v}_j) - (\underline{I} \wedge \underline{u}_j)| = |\underline{I} \vee \underline{v}_j| - |\underline{I} \wedge \underline{u}_j| \quad (4.30)$$

La Figura 4.5 fornisce una rappresentazione grafica della notazione sopra esposta.

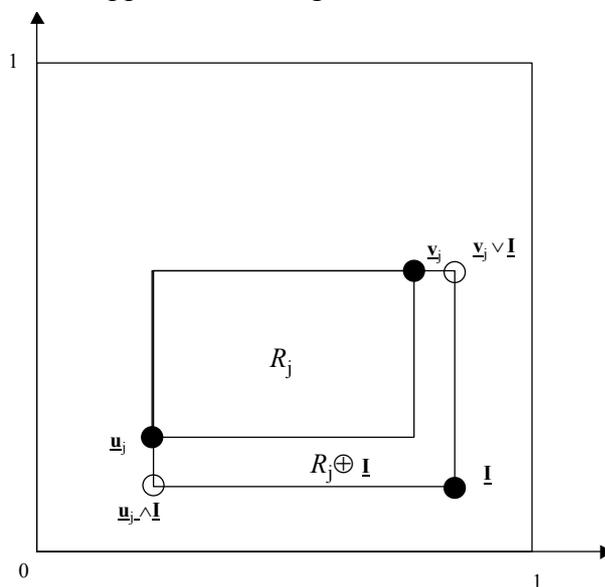


Figura 4.5 – Interpretazione geometrica fase di verifica

Alla luce delle precedenti relazioni, considerando il test $\frac{|\underline{I}^c \wedge \underline{w}_j|}{|\underline{I}^c|} \geq \rho$ attraverso cui il

sistema di orientamento verifica l'adeguatezza di rappresentazione dell'ingresso attuale con il j^{esimo} template, allora risulta che essendo:

$$|\underline{I}^c| = |(\underline{I}, \underline{I} - \underline{I})| = |\underline{I}| + M - |\underline{I}| = M \quad (4.31)$$

$$\begin{aligned} |\underline{I}^c \wedge \underline{w}_j| &= |(\underline{I}, \underline{I} - \underline{I}) \wedge (\underline{u}_j, \underline{I} - \underline{v}_j)| = |(\underline{I} \wedge \underline{u}_j, \underline{I} - \underline{I}^c \wedge \underline{I} - \underline{v}_j^c)| = \\ &= |(\underline{I} \wedge \underline{u}_j), \underline{I} - (\underline{I} \vee \underline{v}_j)| = |\underline{I} \wedge \underline{u}_j| + M - |\underline{I} \vee \underline{v}_j| = M - |R_j \oplus \underline{I}| \end{aligned} \quad (4.32)$$

Allora l'ingresso $\underline{I} = [I_1, I_2, \dots, I_M]$ è riconosciuto appartenente al cluster j^{esimo} se e solo se risulta:

$$\frac{|\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j|}{|\underline{\mathbf{I}}^c|} \geq \rho \Rightarrow |\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j| \geq M\rho \Rightarrow M - |\mathbf{R}_j \oplus \underline{\mathbf{I}}| \geq M\rho \Rightarrow$$

$$|\mathbf{R}_j \oplus \underline{\mathbf{I}}| \leq M(1 - \rho)$$
(4.33)

Considerando la schematizzazione grafica riportata nella precedente Figura 4.5 è evidente che, una volta scelto il template rappresentato dall'iperrettangolo $\mathbf{R}_j = \{\underline{\mathbf{u}}_j, \underline{\mathbf{v}}_j\}$, l'ingresso attuale è considerato rappresentabile dal cluster attivato se e solo se risulta che l'ampiezza dell'iperrettangolo $\mathbf{R}_j \oplus \underline{\mathbf{I}} = \{\underline{\mathbf{I}} \wedge \underline{\mathbf{u}}_j, \underline{\mathbf{I}} \vee \underline{\mathbf{v}}_j\}$ è inferiore al limite $|\mathbf{R}_j \oplus \underline{\mathbf{I}}| \leq M(1 - \rho)$.

Si noti, inoltre, che durante la fase di addestramento se fosse verificato $|\mathbf{R}_j \oplus \underline{\mathbf{I}}| \leq M(1 - \rho)$, allora i pesi del cluster j^{esimo} sarebbero modificati in modo tale da includere le caratteristiche dell'ingresso attuale nel modo seguente

$$\underline{\mathbf{w}}_j^{\text{new}} = (\underline{\mathbf{u}}_j^{\text{new}}, \underline{\mathbf{I}} - \underline{\mathbf{v}}_j^{\text{new}}) \text{ dove } \underline{\mathbf{u}}_j^{\text{new}} = \underline{\mathbf{u}}_j \wedge \underline{\mathbf{I}} \text{ e } \underline{\mathbf{v}}_j^{\text{new}} = \underline{\mathbf{v}}_j \vee \underline{\mathbf{I}}$$
(4.34)

In altri termini, come risultato dell'apprendimento, l'iperrettangolo $\mathbf{R}_j = \{\underline{\mathbf{u}}_j, \underline{\mathbf{v}}_j\}$ risulta sostituito dal nuovo iperrettangolo $\mathbf{R}_j \oplus \underline{\mathbf{I}} = \{\underline{\mathbf{I}} \wedge \underline{\mathbf{u}}_j, \underline{\mathbf{I}} \vee \underline{\mathbf{v}}_j\}$.

4.4.4 Training PLP - Perfectly Learned Patterns

Si consideri l'algoritmo Fuzzy ART descritto in precedenza, e si supponga di fissare il livello del parametro di vigilanza al massimo valore consentito ($\rho = 1$) durante la fase di addestramento. La scelta del massimo livello di vigilanza durante la fase di addestramento impone che tutti i pattern utilizzati siano appresi al massimo livello di dettaglio. Da questo punto in poi tale approccio di apprendimento sarà identificato nel testo come approccio PLP (*Perfectly Learned Patterns*).

Di seguito è riportato come, la matematica dell'algoritmo Fuzzy ART si semplifica nell'ipotesi di utilizzo dell'approccio PLP. Consideriamo, dunque, l'ingresso della rete attraverso la solita notazione.

$$\underline{\mathbf{I}}^c = (\underline{\mathbf{I}}, \underline{\mathbf{I}} - \underline{\mathbf{I}}) = [I_1, I_2, \dots, I_M, 1 - I_1, 1 - I_2, \dots, 1 - I_M]$$
(4.35)

Avendo indicato con $\underline{\mathbf{I}}^c$ l'ingresso dopo il processo *Complement Coding* al livello *F1*. Indichiamo la funzione di scelta (*Choice function*) del neurone vincitore al livello *F2*, con:

$$T_j(\underline{\mathbf{I}}^c) = \frac{|\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j|}{\alpha + |\underline{\mathbf{w}}_j|}$$
(4.36)

Durante la fase di addestramento, qualunque sia la categoria vincitrice della competizione, essa sarà annullata dal sistema d'orientamento qualora risulti che

$\frac{|\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j|}{|\underline{\mathbf{I}}^c|} < 1$. Poiché dalla definizione di minimo secondo la teoria dei Fuzzy Set, e

utilizzando la definizione di ampiezza del di un vettore, è immediato constatare che vale sempre la condizione $|\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j| \leq |\underline{\mathbf{I}}^c|$, allora la categoria vincitrice è annullata sempre a meno che risulti verificata la seguente condizione:

$$|\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j| = |\underline{\mathbf{I}}^c| = M \quad (4.37)$$

Si può osservare che due condizioni sufficienti affinché l'ampiezza del vettore $|\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j|$ sia pari a M , sono le seguenti:

1. Se risulta $\underline{\mathbf{w}}_j = (1,1,\dots,1) \Rightarrow |\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j| = |\underline{\mathbf{I}}^c|$. Questo accade quando un ingresso è presentato per la prima volta alla rete, in altre parole, quando è assegnato un *template uncommitted*. In questo caso, il nuovo valore del template sarà pari a:

$$\underline{\mathbf{w}}_j^{new} = \underline{\mathbf{I}}^c \quad (4.38)$$

2. Se risulta $\underline{\mathbf{w}}_j = \underline{\mathbf{I}}^c$. Questo accade se in fase di addestramento si ripresenta un pattern già appreso in precedenza: in tal caso si ha, infatti, che $|\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j| = |\underline{\mathbf{I}}^c \wedge \underline{\mathbf{I}}^c| = |\underline{\mathbf{I}}^c|$ e vale in ogni caso la seguente relazione di addestramento.

$$\underline{\mathbf{w}}_j^{new} = \beta(\underline{\mathbf{I}}^c \wedge \underline{\mathbf{I}}^c) + (1 - \beta)\underline{\mathbf{I}}^c = \beta\underline{\mathbf{I}}^c + (1 - \beta)\underline{\mathbf{I}}^c = \underline{\mathbf{I}}^c \quad 0 < \beta \leq 1. \quad (4.39)$$

Le due condizioni (4.38 e (4.39) evidenziate in precedenza, oltre ad essere sufficienti sono anche necessarie, poiché si osserva che una volta che un template è stato assegnato ad un ingresso utilizzato nell'addestramento esso può essere riscritto come: $\underline{\mathbf{w}}_j = (\underline{\mathbf{u}}_j, \underline{\mathbf{I}} - \underline{\mathbf{u}}_j)$, quindi risulta che

$$\begin{aligned} |\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j| &= |(\underline{\mathbf{I}}, \underline{\mathbf{I}} - \underline{\mathbf{I}}) \wedge (\underline{\mathbf{u}}_j, \underline{\mathbf{I}} - \underline{\mathbf{u}}_j)| = |\underline{\mathbf{I}} \wedge \underline{\mathbf{u}}_j, (\underline{\mathbf{I}} - \underline{\mathbf{I}}) \wedge (\underline{\mathbf{I}} - \underline{\mathbf{u}}_j)| \\ |\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j| &= |\underline{\mathbf{I}} \wedge \underline{\mathbf{u}}_j, \underline{\mathbf{I}} - (\underline{\mathbf{I}} \vee \underline{\mathbf{u}}_j)| = M + |\underline{\mathbf{I}} \wedge \underline{\mathbf{u}}_j| - |\underline{\mathbf{I}} \vee \underline{\mathbf{u}}_j| \end{aligned} \quad (4.40)$$

Poiché dalla definizione di minimo e massimo fra numeri Fuzzy risulta sempre che $|\underline{\mathbf{I}} \wedge \underline{\mathbf{u}}_j| \leq |\underline{\mathbf{I}} \vee \underline{\mathbf{u}}_j|$ allora si ha che, durante la fase di addestramento PLP, una volta che un nodo è assegnato ($\underline{\mathbf{w}}_j = (\underline{\mathbf{u}}_j, \underline{\mathbf{I}} - \underline{\mathbf{u}}_j)$), risulterà sempre durante le successive presentazioni di addestramento che $|\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j| \leq M$ valendo l'uguaglianza se e solo se si ha $\underline{\mathbf{I}} = \underline{\mathbf{u}}_j$. In altri termini, impostando in fase di addestramento la massima vigilanza, allora un template formato non potrà essere modificato durante le successive fasi di apprendimento.

Il primo risultato dell'approccio PLP per l'addestramento della rete Fuzzy ART è dunque il seguente.

R1: tutti i pattern presentati durante l'addestramento sono memorizzati esattamente come template della rete. Ogni input può essere presentato una ed una sola volta per essere appreso definitivamente.

Inoltre, poiché si ha che

$$\forall \beta \in]0,1]: \underline{\mathbf{w}}_j^{new} = \underline{\mathbf{I}}^c \quad (4.41)$$

Il secondo risultato dell'approccio Perfectly Learned Patterns è che:

R2: il parametro di learning β non è influente sulle prestazioni della rete e qualunque valore maggiore di 0 può essere adottato.

Per quanto riguarda il parametro di scelta α si ricorda che esso può influenzare solo l'ordine di scelta dei template memorizzati nel livello. Si osservi che tale ordine è

del tutto ininfluyente ai fini pratici sulle prestazioni della rete. Infatti, si avrà sempre un segnale di reset emesso dal sistema di orientamento qualora il template associato non risulta identico a quello presentato in ingresso. Quindi qualunque sia l'ordine di scelta l'apprendimento proseguirà fino al riconoscimento del pattern d'ingresso come un pattern già noto e appreso in precedenza, oppure fino all'assegnazione ad un nodo uncommitted. Il terzo risultato dell'approccio Perfectly Learned Patterns è che:

R3: il parametro di scelta α non risulta influente sulle prestazioni della rete durante la fase di addestramento.

4.4.5 Testing di una rete ad addestramento PLP

Passando alla fase di testing di una rete addestrata con massima vigilanza è evidente che, per garantire una opportuna capacità di generalizzazione, sarà necessario ridurre il valore della vigilanza ad un livello $\rho < 1$, in caso contrario ($\rho = 1$), la rete si comporterà banalmente come una memoria capace di riconoscere come familiari solo ed esclusivamente i pattern presentati in addestramento.

È evidente inoltre che, tanto più basso sarà il valore del parametro di vigilanza utilizzato sulla rete durante la fase operativa, allora tanto più alta sarà la capacità della rete di generalizzare l'apprendimento. Pertanto, la scelta del valore del parametro di vigilanza durante la fase di testing, dovrà essere tale da garantire da un lato, una sufficiente capacità di generalizzazione della rete (bassi valori di vigilanza $\rho \rightarrow 0.0$), e dall'altro, un'adeguata capacità di discriminare opportunamente i dati presentati in funzione di quanto appreso in addestramento (alti valori di vigilanza $\rho \rightarrow 1.0$).

In particolare, poiché al termine della fase di training la rete avrà appreso esattamente tutti i pattern presentati, l'insieme dei pesi top-down sarà del tipo

$$\underline{w}_j = (\underline{u}_j, \underline{1} - \underline{u}_j) \quad (4.42)$$

Risulterà quindi che

$$|\underline{w}_j| = |(\underline{u}_j, \underline{1} - \underline{u}_j)| = |\underline{u}_j| + M - |\underline{u}_j| = M \quad (4.43)$$

Il corrispondente iperrettangolo R_j avrà ampiezza nulla essendo questa pari a $|R_j| = |\underline{u}_j - \underline{u}_j|$. In altri termini ogni template sarà rappresentato da un unico vettore nello spazio a M dimensioni: quello utilizzato nella fase di addestramento.

Supponiamo ora che, in fase di testing sia presentato l'ingresso indicato dalla seguente notazione.

$$\underline{I}^c = (\underline{I}, \underline{1} - \underline{I}) = [I_1, I_2, \dots, I_M, 1 - I_1, 1 - I_2, \dots, 1 - I_M] \quad (4.44)$$

Sarà scelto come template rappresentativo quello cui corrisponde il massimo valore della funzione di scelta. Ovvero il cluster di indice j tale che sia verificata la seguente relazione.

$$j = \arg \max_j T_j(\underline{I}^c) \text{ dove } T_j(\underline{I}^c) = \frac{|\underline{I}^c \wedge \underline{w}_j|}{\alpha + |\underline{w}_j|} = \frac{|\underline{I}^c \wedge \underline{w}_j|}{\alpha + M} \quad (4.45)$$

Poiché risulta come già evidenziato che

$$\begin{aligned} |\underline{I}^c \wedge \underline{w}_j| &= |(\underline{I}, \underline{1} - \underline{I}) \wedge (\underline{u}_j, \underline{1} - \underline{u}_j)| = |\underline{I} \wedge \underline{u}_j, (\underline{1} - \underline{I}) \wedge (\underline{1} - \underline{u}_j)| \\ |\underline{I}^c \wedge \underline{w}_j| &= |\underline{I} \wedge \underline{u}_j, \underline{1} - (\underline{I} \vee \underline{u}_j)| = M + |\underline{I} \wedge \underline{u}_j| - |\underline{I} \vee \underline{u}_j| \end{aligned} \quad (4.46)$$

E quindi $|\underline{\mathbf{I}}^c \wedge \underline{\mathbf{w}}_j| = M - \sum_{i=1}^M |I_i - u_{ji}|$, allora si ha che

$$j = \arg \max_j T_j(\underline{\mathbf{I}}^c) \Leftrightarrow j = \arg \max_j \left[M - \sum_{i=1}^M |I_i - u_{ji}| \right] \Leftrightarrow$$

$$j = \arg \min_j \left[\sum_{i=1}^M |I_i - u_{ji}| \right] \quad (4.47)$$

Definendo come $dis(\underline{\mathbf{I}}, \underline{\mathbf{u}}_j) = \sum_{i=1}^M |I_i - u_{ji}|$, allora il nodo rappresentativo che sarà scelto sarà quello per il quale è minima la distanza: $j = \arg \min_j dis(\underline{\mathbf{I}}, \underline{\mathbf{u}}_j)$. Questo implica che $\forall \alpha$ sarà scelto sempre il template con le maggiori similitudini con l'ingresso attuale e dunque il template con più alta possibilità di superare il test di vigilanza.

Una volta scelto il nodo j quale nodo rappresentativo, il pattern d'ingresso sarà associato al corrispondente cluster se e solo se sarà verificato che

$$|\mathbf{R}_j \oplus \underline{\mathbf{I}}| \leq M(1 - \rho) \quad (4.48)$$

dove

$$|\mathbf{R}_j \oplus \underline{\mathbf{I}}| = |(\underline{\mathbf{I}} \vee \underline{\mathbf{u}}_j) - (\underline{\mathbf{I}} \wedge \underline{\mathbf{u}}_j)| = |\underline{\mathbf{I}} \vee \underline{\mathbf{u}}_j| - |\underline{\mathbf{I}} \wedge \underline{\mathbf{u}}_j| = dis(\underline{\mathbf{I}}, \underline{\mathbf{u}}_j) = \sum_{i=1}^M |I_i - u_{ji}| \quad (4.49)$$

Ovvero il pattern d'ingresso sarà riconosciuto appartenente al nodo j se la distanza fra l'input e il vettore appreso in fase di addestramento è minore o uguale al limite $M(1 - \rho)$.

$$dis(\underline{\mathbf{I}}, \underline{\mathbf{u}}_j) \leq M(1 - \rho) \quad (4.50)$$

Risulta evidente che, il test precedente è eseguito una sola volta, poiché il primo cluster scelto fra tutti quelli memorizzate è quello che presenta la minima distanza con il vettore d'ingresso.

Infine, come indicato dalla precedente equazione (4.50), si osservi che gli unici parametri di configurazione che possono influenzare in fase di testing il funzionamento della rete Fuzzy ART addestrata con approccio PLP sono: il parametro di vigilanza ρ , e l'ampiezza della finestra di ingresso M .

Capitolo 5

Il sistema di controllo qualità Fuzzy ART

Utilizzando la descrizione dell'algoritmo ART fatta nel precedente capitolo, di seguito sarà definita una procedura neurale di monitoraggio dello stato di un processo evidenziandone tutti i parametri d'influenza. Tale procedura sarà poi analizzata in dettaglio definendo le relazioni matematiche, e statistiche, che legano i parametri dell'algoritmo neurale con le prestazioni attese nel riconoscimento di variazioni dello stato di un processo.

5.1 L'algoritmo ART per il monitoraggio della qualità

L'obiettivo del lavoro è la realizzazione di un algoritmo neurale ART, in grado di eseguire il controllo della qualità per un generico processo, cioè capace di discriminare uno stato di variabilità *naturale* delle risposte, da uno stato di variabilità *non naturale*.

La procedura utilizzata nel presente lavoro, si richiama quella proposta da Al-Ghanim (1997) secondo la quale la rete neurale ART, durante la fase di apprendimento, opera unicamente con dati naturali del processo. Attraverso un metodo auto-organizzativo poi, la rete è in grado di partizionare l'insieme di dati naturali presentati in una serie di sottoinsiemi (definiti cluster) disgiunti. La funzione eseguita da un approccio basato sul clustering, quando la rete neurale si confronta con nuovi input, è semplicemente quella di produrre una risposta indicante il cluster (ovvero il sottoinsieme) cui si conclude che il pattern appartiene (qualora tale cluster esista). Vale a dire che la funzione del sistema neurale non indicherà il tipo di pattern non naturale che si stima essere presente, ma essa fornirà unicamente l'informazione sull'appartenenza, o meno, del pattern analizzato ad uno dei sottoinsiemi che descrivono la naturale variabilità del processo e che sono stati formati durante l'addestramento. Quest'informazione fornirà un'indicazione sullo stato del processo: se esso si comporta in modo simile a quanto già appreso dalla rete, e che si suppone essere il modo naturale di variazione. Oppure, se un cambiamento nell'output si stima sia avvenuto poiché l'ingresso non è riconosciuto appartenente ai cluster naturali appresi. Lo scopo del sistema a rete neurale è dunque duplice. 1) In modo addestramento: scoprire le strutture caratteristiche (ovvero i cluster essenziali) dei dati naturali di processo; 2) in modo operativo: fornire un'indicazione del cambiamento nell'andamento del processo se esso è avvenuto.

5.2 Applicazione dell'algoritmo Fuzzy ART

Generalmente, una rete neurale può operare in due differenti fasi: training e testing. In questa sezione esamineremo le fasi operative di una rete Fuzzy ART applicata per il controllo di qualità.

Il training di una rete Fuzzy ART, per applicazioni di controllo di qualità, può essere condotto in due differenti scenari.

Nel primo scenario di training si suppone di avere una collezione di pattern d'ingresso chiamata *training list*. La *training list* è un insieme di pattern da impiegare

per l'addestramento della rete, ottenuti utilizzando dati prodotti direttamente dal processo in osservazione. Comunemente questi dati rappresentano le misure del parametro di qualità di interesse quando solo cause non assegnabili di variazione sono presenti nel processo. In altri termini, si tratta di dati prodotti dal processo in uno stato naturale di variabilità. La durata della sequenza di dati naturali utilizzati per la fase di addestramento è definita *periodo d'addestramento*: durante tale intervallo di tempo si suppone che il processo produca unicamente uscite naturali che sono apprese dalla rete neurale ART. La procedura di training della rete ha termine quando l'algoritmo d'apprendimento ha prodotto tutti i possibili cluster naturali, per l'insieme di dati d'addestramento, fissato un determinato valore del parametro di vigilanza.

L'ipotesi implicita con questo primo scenario di training è che il processo sotto ispezione produca, durante il periodo di addestramento, unicamente dati naturali. In pratica, si desidera che la rete Fuzzy ART apprenda questi dati suddividendoli autonomamente in gruppi (detti cluster) che rappresentano il naturale modo di variabilità del processo. Ovviamente, ingressi che presentano caratteristiche simili saranno raggruppati nella stessa categoria durante il training. Il processo di raggruppamento in cluster, eseguito durante la fase di apprendimento, dipenderà dal parametro di vigilanza scelto per l'algoritmo ART. Un'alta vigilanza impone un rigoroso criterio di comparazione fra due differenti ingressi per essere raggruppati nella medesima categoria. Ciò implica che l'insieme dei pattern naturali utilizzati per l'addestramento sarà suddiviso in molti raggruppamenti di pochi pattern. Ognuno di questi raggruppamenti sarà caratterizzato da un template (ovvero un prototipo rappresentativo) molto dettagliato. Viceversa, bassi parametri di vigilanza faranno sì che diversità più accentuate fra i differenti pattern naturali possano essere tollerate per considerare tali ingressi raggruppabili in una medesima categoria. Conseguentemente in questo caso, il processo di addestramento conduce alla formazione di pochi cluster che raggruppano molti pattern. Ognuno di questi cluster sarà rappresentato da un prototipo molto grezzo, in altre parole con un numero limitato di dettagli.

L'utilizzo di una procedura neurale a addestramento non supervisionato, consente quindi, di poter utilizzare un'unica classe di dati ai fini dell'addestramento: i pattern naturali. In questo caso, una volta che la rete è stata addestrata unicamente su dati naturali, un ingresso può essere accettato come un pattern che appartiene ad uno dei cluster formati, o può essere rigettato come un andamento non naturale, che non corrisponde a nessuno dei gruppi appresi. Quindi, con questo metodo, un pattern d'ingresso sarà identificato come non naturale, unicamente se non esiste alcuna classe ad esso corrispondente, cioè se la rete respinge tale ingresso come un pattern non familiare (Al-Ghanim 1997).

Nel secondo scenario di addestramento si suppone di non disporre di una predeterminata lista di pattern naturali che debbano essere appresi dalla rete neurale. Al contrario, si suppone di conoscere unicamente il target naturale del processo. Con target si intende il livello nominale atteso per le uscite prodotte dal processo. Si può considerare il target come la media nominale, vale a dire il valore che l'uscita del processo dovrebbe avere se sia le cause non assegnabili (naturali), sia quelle cause assegnabili (non naturali) di variabilità, non ne influenzano lo stato. Se il target del processo è noto e costante nel tempo, allora l'addestramento della rete può essere ottenuto utilizzando un singolo vettore di training: quello in cui tutti i componenti codificano il target nominale del processo. In questo scenario, la rete formerà un unico cluster naturale. In genere si desidera che il template prototipo di tale cluster sia il più rappresentativo possibile, in altre parole, coincidente con il pattern utilizzato nell'addestramento. Per ottenere ciò, sarà necessario fissare durante l'addestramento il

livello di vigilanza al suo massimo valore ($\rho = 1$). Se, infatti, il parametro di vigilanza assume il valore massimo durante la fase di addestramento allora la rete neurale formerà delle categorie che riproducono esattamente i pattern utilizzati durante l'addestramento (*Perfectly Learned Patterns* – PLP). Utilizzando l'approccio di apprendimento PLP, allora è sufficiente limitare il numero di presentazioni della lista di addestramento ad una sola volta.

Nella fase di testing, il processo di apprendimento è disabilitato, vale a dire che non sono più consentiti adattamenti dei pesi della rete, o la creazione di nuovi cluster. In questa fase, i pattern di una lista di test sono presentati alla rete allo scopo di valutarne le prestazioni di classificazione. In particolare, quando un pattern prelevato dalla lista di test è presentato alla rete Fuzzy ART, allora, attraverso l'algoritmo neurale, sarà scelta un'eventuale categoria fra quelle apprese in precedenza al fine di rappresentare l'ingresso. Il criterio per determinare quanto una categoria possa essere ritenuta appropriata a rappresentare l'ingresso attuale della rete dipende dal parametro di vigilanza scelto per la fase di testing (e che può essere differente da quello utilizzato per la fase di training). In particolare, alti livelli di vigilanza impongono stretti criteri di comparazione tra un ingresso e il prototipo di una categoria. Analogamente si può dire che alti livelli di vigilanza determinano ridotte capacità di generalizzazione della rete neurale. Al contrario, dei bassi livelli di vigilanza consentono di tollerare grandi differenze fra un ingresso ed il prototipo di un cluster per associare a questo ultimo l'input attuale. In altri termini, bassi livelli di vigilanza, utilizzati durante la fase di testing, implicano capacità elevate di generalizzazione della rete neurale.

Naturalmente, i risultati forniti dalla rete neurale Fuzzy ART in fase di testing sono strettamente dipendenti dal criterio implementato per decidere se un ingresso possa essere ritenuto associabile ad una categoria o meno. Da questo punto in poi, tale criterio sarà definito come il *criterio di generalizzazione* della rete neurale.

5.3 I parametri della rete Fuzzy ART

La seguente tabella riporta l'elenco di tutti i parametri che devono essere definiti nella progettazione della rete neurale di tipo ART.

<i>Parametro ART</i>	<i>Insieme di definizione</i>
Neuroni F1	$\{1,2,\dots, +\infty\}$
Parametro di Vigilanza (ρ)	$]0,1]$
Learning Rate(β)	$]0,1]$
Choice Parameter (α)	$[0,+\infty[$
Neuroni F2	$\{1,2,\dots, +\infty\}$
Numero Presentazioni	$\{1,2,\dots, +\infty\}$

Tabella 5.1 – Parametri di implementazione di una rete neurale Fuzzy ART

Dal numero di neuroni in F1 e in F2 dipendono le caratteristiche della rete ART, e la quantità di risorse computazionali richieste dall'algoritmo. In particolare, tenuto conto che ogni neurone del livello F1 è connesso con tutti i nodi del livello F2 attraverso pesi bottom-up, così come ciascuno dei neuroni del livello F2 è connesso con tutti quelli del livello F1 attraverso pesi top-down, allora è evidente che il numero di connessioni previste per la rete neurale dovrà essere pari a 2 volte il prodotto fra il numero di neuroni nel livello F1 e il numero di neuroni nel livello F2.

Tali connessioni rappresentano la memoria a lungo termine della rete (LTM). In genere, si è osservato che i tempi di addestramento crescono in modo più che lineare al crescere delle connessioni della LTM, quindi un obiettivo è quello di contenere, se possibile, al minimo la dimensione della rete neurale.

Il tasso di apprendimento (β), per definizione, deve essere inferiore o uguale a 1 e non può essere nullo. Questo è evidente perché, se il tasso di apprendimento fosse nullo, non avrebbe luogo nessun tipo di apprendimento, e la rete rimarrebbe sempre nello stato d'inizializzazione (ovvero con pesi delle connessioni nello stato iniziale) indipendentemente dal numero di pattern presentati in ingresso. D'altro canto, quanto più alto è il tasso di apprendimento tanto più "plastica" risulterà la rete, e quindi tanto più velocemente essa modificherà la propria memoria di lungo termine per tenere in conto delle nuove caratteristiche dei pattern che si susseguono nell'addestramento. Il caso particolare di tasso di apprendimento massimo, ovvero pari a 1, è definito Fast-Committ-Fast-Recode (FCFR), ed è una modalità di funzionamento dell'algoritmo ART estremamente semplificata ed efficace in molti casi. Bassi valori del tasso di apprendimento fanno sì che, affinché una determinata caratteristica sia effettivamente appresa, allora è necessario che essa risulti comune a numerosi pattern, in caso contrario la rete manterrà invariata la propria memoria di lungo termine.

Il parametro di scelta (α), da un lato influenza il valore iniziale dei pesi della rete, ma questo non riveste molta importanza ai fini pratici, dall'altro lato, esso influenza la funzione di scelta della rete, in altre parole l'ordine con cui i diversi neuroni nel livello F2 possono essere scelti per rappresentare un determinato pattern d'ingresso. Una volta che una categoria vincitrice è stata scelta, essa sarà comunque, sottoposta dal sistema di orientamento al controllo di vigilanza che, se fallito, comporterà l'annullamento della categoria vincitrice e il passaggio eventuale ad una nuova scelta.

Il parametro di vigilanza (ρ), è quello da cui prevalentemente dipendono le prestazioni della rete ART. Esso, per definizione, non può essere superiore a 1 e certamente non può essere nullo. Un parametro di vigilanza nullo, infatti, implicherebbe che qualunque fosse l'ingresso, e qualunque fosse la categoria vincitrice, allora il sistema di orientamento non emetterebbe mai un segnale di reset indicando, che ogni pattern, indipendentemente dalle sue caratteristiche, è ritenuto familiare.

Il parametro di vigilanza, durante la fase di addestramento, determina il grado di similitudine richiesto ad un ingresso per essere associato ad una predefinita classe di pattern, prima che una nuova classe sia attivata per tenere in conto delle nuove caratteristiche. Questo implica che, generalmente, più alto è il parametro di vigilanza, più alto sarà il numero di cluster formati durante l'addestramento: al limite per un approccio PLP ($\rho = 1$) si formeranno tanti cluster quanti sono i pattern (supposti differenti) utilizzati nell'addestramento della rete; inoltre, maggiore sarà il parametro di vigilanza, maggiore sarà il livello di dettaglio appreso dalla rete per ogni singolo cluster formato. In fase operativa, il parametro di vigilanza influenza il livello di conformità richiesto fra gli ingressi, e i cluster appresi durante l'addestramento, al fine di riconoscerli come pattern familiari. La regola è che: più è elevato il parametro di vigilanza, più alto sarà il livello di similitudine richiesto tra uno specifico input e almeno uno dei cluster appresi, allo scopo di riconoscere l'ingresso come un pattern familiare.

5.3.1 Scelta dei parametri della rete neurale

Supponiamo di effettuare un addestramento della rete neurale su un insieme di pattern utilizzando l'approccio a massima vigilanza PLP. Alla luce delle osservazioni

effettuate nel capitolo 4 e relative all'utilizzo di un algoritmo neurale Fuzzy ART con addestramento a massima vigilanza, si può effettuare la scelta dei parametri sia per la fase di addestramento, sia per quella operativa, come descritto nella seguente Tabella 5.2.

<i>Parametro ART Add. PLP</i>	<i>Insieme di definizione</i>	<i>Valore in fase di apprendimento</i>	<i>Valore in fase operativa</i>
Parametro di Vigilanza (ρ)	$]0,1]$	1	<1
Recode rate (β)	$]0,1]$	Non influente	Non influente
Choice Parameter (α)	$[0,+\infty]$	Non influente	Non influente
Neuroni F1	$\{1,2,\dots, +\infty\}$	2M	2M
Neuroni F2	$\{1,2,\dots, +\infty\}$	=dimensione training set	=dimensione training set
Numero Presentazioni	$\{1,2,\dots, +\infty\}$	1	1

Tabella 5.2 – Parametri di una rete Fuzzy ART ad addestramento PLP

Come già dimostrato nel capitolo 4, alcuni parametri (α e β) risultano non influenti sul comportamento della rete neurale e possono assumere qualunque valore appartenente all'insieme di definizione. Altri (il numero di neuroni in F1 e in F2) risultano dipendenti dalle specifiche di utilizzo della rete neurale (rispettivamente la dimensione della ampiezza di finestra e la dimensione dell'insieme di addestramento). Infine, la scelta del massimo parametro di vigilanza rende l'addestramento particolarmente veloce per qualunque valore del parametro M , infatti, è sufficiente che ogni pattern sia presentato una sola volta alla rete.

Si conclude, pertanto, che gli unici parametri di influenza sulle prestazioni della rete neurale Fuzzy ART addestrata in modalità PLP risultano essere: il parametro di vigilanza utilizzato in fase operativa (ρ), la window size (M).

5.4 Applicazione della rete per il controllo di qualità

Il sistema neurale per il controllo della qualità, ed il modello di simulazione del processo produttivo di riferimento (ipotizzato senza memoria), sono rappresentati dalla Figura 5.1.

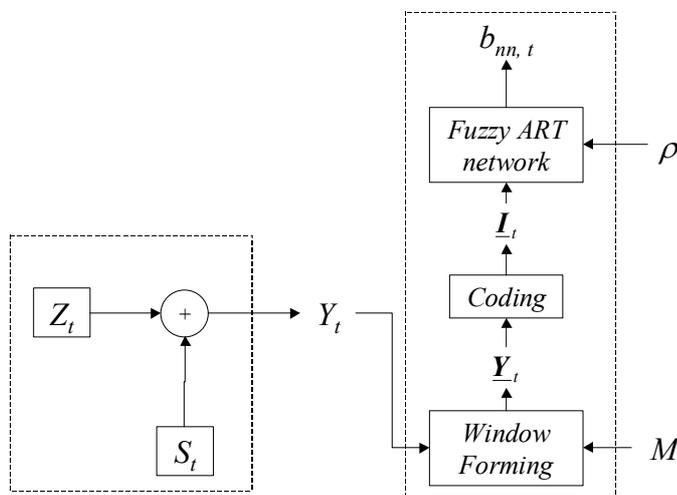


Figura 5.1 – Modello del sistema di controllo qualità

Sia $\{Y_t\}$ la serie temporale delle uscite del processo. Ad ogni istante di tempo t , il sistema di controllo riceve l'uscita del processo Y_t e produce il segnale binario $b_{m,t}$. Questo segnale rappresenta il risultato del test eseguito dalla rete Fuzzy ART sullo stato del processo. L'algoritmo è stato programmato in modo da produrre $b_{m,t} = 1$ se il processo può essere considerato in uno stato naturale, $b_{m,t} = 0$ nel caso contrario (capitolo 4).

Come evidenziato dalla Figura 5.1 precedente, prima che le uscite del processo siano elaborate dalla rete neurale Fuzzy ART è necessario realizzare alcune pre-elaborazioni. Il primo stadio di pre-elaborazione, chiamato *Window Forming*, è rivolto a trasformare la serie temporale delle uscite del processo $\{Y_t\}$ in vettori di dimensione M . Lo schema di codifica adottato per questa fase è il seguente dove \underline{Y}_t è l'uscita del processo *Window Forming* all'istante di tempo $t \geq M$.

$$\underline{Y}_t = [Y_{t-M+1}, Y_{t-M+2}, \dots, Y_{t-1}, Y_t] \quad t \geq M \quad (5.1)$$

Il parametro M rappresenta la *Window Size* (WS) del sistema di controllo neurale. Ovviamente, la necessità di dover arrangiare gli ingressi della rete in vettori di dimensione M , implica che il sistema non potrà essere in grado di fornire nessun'informazione sullo stato del processo durante i primi $M - 1$ intervalli di tempo. Tuttavia, poiché è utilizzata una finestra a passo mobile, una volta raccolte le prime $M - 1$ osservazioni, allora il sistema sarà in grado di fornire un'indicazione sullo stato del processo ad ogni singolo istante di tempo $t \geq M$.

L'obiettivo della successiva elaborazione è quello di trasformare un ingresso M -dimensionale \underline{Y}_t , in un corrispondente vettore in uscita M -dimensionale (diciamo \underline{I}_t), i cui componenti rientrano nell'intervallo $[0,1]$. Questo tipo di trasformazione è chiamata *Coding* e consiste in un re-scaling di tipo lineare degli ingressi all'interno dell'insieme di arrivo $[0,1]$. Lo schema di codifica adottato è rappresentato dalla seguente equazione (5.2). Sia $l > 0$ un limite appropriato per il valore assoluto di variazione del processo $\{Y_t\}$ dalla media nominale (il target del processo).

$$\underline{I}_t = [I_{t-M+1}, I_{t-M+2}, \dots, I_{t-1}, I_t] \quad t \geq M$$

$$\begin{cases} I_\tau = 0 & Y_\tau < -l; \\ I_\tau = \frac{1}{2} \left(1 + \frac{Y_\tau}{l} \right) & -l \leq Y_\tau \leq l; \\ I_\tau = 1 & l < Y_\tau; \end{cases} \quad (5.2)$$

Supponendo come nel caso di riferimento di assumere la sequenza delle uscite del processo in uno stato naturale rappresentabili come NID(0,1) allora un opportuno valore per l'indice l è il valore $l = 3$. In tal caso infatti, è atteso che circa il 99.74% delle osservazioni prodotte dal processo naturale ricadano all'interno dell'intervallo di variazione $[-3,3]$.

La rete neurale Fuzzy ART utilizzata in questo lavoro è stata dettagliatamente descritta nel capitolo 4. Si ricorda, in particolare, che essa è costituita da due principali sottosistemi, il sistema di attenzione e il sistema di orientamento. Il sottosistema di attenzione è, a sua volta, costituito da tre livelli di nodi $F0$, $F1$ e $F2$, mentre, il sottosistema di orientamento è caratterizzato da un singolo nodo (il nodo di reset) la cui uscita influenza i neuroni dello strato $F2$.

Nello strato di neuroni $F0$ è eseguita un'ulteriore fase di codifica sugli ingressi (*Complement Coding*) nella quale ad un ingresso M -dimensionale $\underline{I}_t = [I_{t-M+1}, \dots, I_{t-1}, I_t]$ è associata un vettore di uscita $2M$ - dimensionale \underline{I}_t^c , rispondente alla seguente relazione.

$$\underline{I}_t^c = (\underline{I}_t, \underline{I} - \underline{I}_t) = [I_{t-M+1}, I_{t-M+2}, \dots, I_t, 1 - I_{t-M+1}, 1 - I_{t-M+2}, \dots, 1 - I_t] \quad (5.3)$$

5.5 Fasi operative della rete neurale

In questo paragrafo, e in quelli successivi del presente capitolo, soffermeremo la descrizione all'uso della rete Fuzzy ART addestrata a riconoscere unicamente il target naturale del processo (ovvero la media nominale). In particolare, si suppone di utilizzare l'approccio di addestramento PLP, già descritto nel capitolo precedente, e consistente nell'impiego della massima vigilanza $\rho = 1$.

5.5.1 Training

Senza perdita di generalità consideriamo $\mu = 0$ il target di riferimento per il processo in uno stato naturale che assumiamo essere costante nel tempo, sia $\underline{Y} = [0, 0, \dots, 0]$ il vettore M -dimensionale che risulta dalla procedura di *Window Forming*. Sia \underline{I} il vettore a M componenti appartenenti all'intervallo $[0, 1]$ che deriva dalla procedura di *Coding*. È immediato verificare, utilizzando la relazione (5.2) che tutti gli M componenti del vettore \underline{I} sono uguali al valore 0.5. Sia, inoltre, $\underline{I}^c = (\underline{I}, \underline{I} - \underline{I})$ il vettore a $2M$ componenti che deriva dalla procedura (5.3) *Complement Coding* realizzata nel livello $F0$ della rete Fuzzy ART. Anche in questo caso è immediato verificare che tutti i $2M$ componenti del vettore \underline{I}^c sono uguali a 0.5.

All'avvio della fase di training, tutti i nodi di tipo top-down della rete neurale sono inizializzati al valore di 1. La presenza del vettore di addestramento \underline{I}^c produce la formazione di un cluster il cui corrispondente vettore di pesi top-down (di $2M$ componenti), che indichiamo con \underline{w}_μ^c è settato uguale a $\underline{w}_\mu^c = \underline{I}^c$. La fase di training termina quando l'unico cluster della rete è formato (\underline{w}_μ^c), in altre parole quando il target naturale \underline{I}^c è memorizzato nel vettore di pesi $\underline{w}_\mu^c = \underline{I}^c$. La procedura di training descritta, produce conseguentemente, un unico cluster nel livello $F2$.

Nel seguito, faremo riferimento al vettore \underline{w}_μ^c di $2M$ componenti nella forma $\underline{w}_\mu^c = (\underline{w}_\mu, \underline{I} - \underline{w}_\mu)$. È evidente che, in questo specifico caso di addestramento i vettori a M componenti $\underline{w}_\mu, \underline{I} - \underline{w}_\mu, \underline{I}$ risultano equivalenti ($\underline{w}_\mu = \underline{I} - \underline{w}_\mu = \underline{I}$).

5.5.2 Testing

Assumiamo che, ad un generico istante di tempo $t \geq M$ un ingresso M -dimensionale \underline{I}_t è presentato al livello $F0$ della rete neurale Fuzzy ART. La presenza del vettore in ingresso \underline{I}_t^c , produce attraverso lo strato di neuroni $F1$, l'attivazione dell'unico cluster presente nel livello $F2$, il cui vettore di pesi top-down è indicato con

\underline{w}_μ^c . La verifica della corrispondenza fra ingresso e cluster memorizzato è valutata nel sistema di orientamento comparando il valore del rapporto riportato dalla equazione (5.4) con il livello di vigilanza ρ adottato nella fase di testing della rete (Georgiopoulos *et al.* 1996, 1999, Hang *et al.* 1995).

$$\frac{|\underline{I}_t^c \wedge \underline{w}_\mu^c|}{|\underline{I}_t^c|} \quad (5.4)$$

Se il tasso riportato nella precedente relazione (5.4) è più piccolo del parametro di vigilanza ρ , allora il cluster naturale è de-attivato e l'algoritmo produce il valore di uscita $b_{nn,t} = 0$ indicante che l'ingresso attuale non può essere considerato naturale. Altrimenti, l'ingresso è considerato naturale ed il valore di uscita della rete è posto pari a $b_{nn,t} = 1$.

Considerando le relazioni dell'algoritmo Fuzzy ART, già dimostrate nel capitolo 4, ovvero che: $|\underline{I}_t^c \wedge \underline{w}_\mu^c| = M + |\underline{I}_t \wedge \underline{w}_\mu| - |\underline{I}_t \vee \underline{w}_\mu| = M - dis(\underline{I}_t, \underline{w}_\mu)$ e che $|\underline{I}_t^c| = M$, allora è possibile riformulare il criterio di valutazione con cui un generico ingresso \underline{I}_t è classificato naturale dalla rete neurale attraverso la seguente relazione.

$$dis(\underline{I}_t, \underline{w}_\mu) \leq M(1 - \rho) \quad (5.5)$$

Che risulta equivalente alla seguente

$$\sum_{r=1}^M |I_{t-M+r} - w_{\mu r}| \leq M(1 - \rho) \quad (5.6)$$

In definitiva, possiamo riformulare l'algoritmo Fuzzy ART in fase di testing attraverso i seguenti passi.

- Calcolare la distanza tra l'ingresso attuale ed il cluster memorizzato $dis(\underline{I}_t, \underline{w}_\mu) = |\underline{I}_t \vee \underline{w}_\mu| - |\underline{I}_t \wedge \underline{w}_\mu|$.
- Confrontare il valore di distanza con il limite $M(1 - \rho)$. Se tale distanza è inferiore al limite $M(1 - \rho)$ allora l'ingresso è considerato naturale e l'uscita del sistema è fissata pari a $b_{nn,t} = 1$.
- In caso contrario, l'ingresso attuale è considerato non naturale e l'output del sistema è posto pari a $b_{nn,t} = 0$, indicando in tal modo un allarme.

5.6 Modello Matematico

Dalle precedenti relazioni, è possibile ricavare con semplice algebra, un modello matematico che lega direttamente le uscite prodotte dal processo $\{Y_t\}$, con il criterio generalizzazione implementato dalla rete neurale Fuzzy ART. In particolare, la condizione necessaria e sufficiente per generare un segnale di stato naturale di variazione nell'istante di tempo $t \geq M$ può essere sintetizzata dalla seguente relazione.

$$\forall t \geq M; b_{nn,t} = 1 \Leftrightarrow \sum_{r=1}^M |I_{t-M+r} - w_{\mu r}| \leq M(1 - \rho) \quad (5.7)$$

Che, avendo assunto $\mu = 0$ il target di riferimento per il processo in uno stato naturale, può essere riscritta come segue.

$$\forall t \geq M; b_{nn,t} = 1 \Leftrightarrow \sum_{r=1}^M \left| I_{t-M+r} - \frac{1}{2} \right| \leq M(1 - \rho) \quad (5.8)$$

Ora, supponendo che $\forall \tau \geq 0; -l \leq Y_\tau \leq l$, ovvero supponendo che l'uscita prodotta dal processo possa essere considerata variabile nell'intervallo di variazione $[-l, l]$, allora sussiste una semplice relazione lineare che lega i valori di Y_τ con i corrispondenti valori I_τ di ingresso alla rete neurale, ovvero:

$$\forall \tau \geq 0; -l \leq Y_\tau \leq l \Rightarrow I_\tau = \frac{1}{2} \left(1 + \frac{Y_\tau}{l} \right) \quad (5.9)$$

Utilizzando quest'ultima relazione, è immediato riformulare il criterio di generalizzazione come indicato nelle seguenti equazioni.

$$\forall t \geq M; b_{nn,t} = 1 \Leftrightarrow \frac{1}{2} \sum_{r=1}^M \left| 1 + \frac{Y_{t-M+r}}{l} - 1 \right| \leq M(1 - \rho)$$

$$\forall t \geq M; b_{nn,t} = 1 \Leftrightarrow \frac{1}{2} \sum_{r=1}^M \left| \frac{Y_{t-M+r}}{l} \right| \leq M(1 - \rho) \quad (5.10)$$

$$\forall t \geq M; b_{nn,t} = 1 \Leftrightarrow \sum_{r=1}^M |Y_{t-M+r}| \leq 2lM(1 - \rho)$$

In altri termini, affinché una finestra di M osservazioni successive prelevate da un processo (il cui target di riferimento sia il valore nullo), sia considerabile un pattern naturale dalla rete neurale, è necessario che la grandezza ottenuta sommando i valori assoluti delle M osservazioni sia inferiore ad un limite proporzionale al prodotto della *Window Size* M e del complementare a 1 del parametro di vigilanza ρ .

Da questo punto in poi, sarà definita la grandezza *Mean Absolute Deviation (MAD)* come la media degli scarti in valore assoluto di un campione di M osservazioni da un determinato target di riferimento μ . Tale grandezza, dipendente dalla dimensione della finestra di punti, è definita in formula come segue.

$$MAD_t(M) = \frac{\sum_{r=1}^M |Y_{t-M+r} - \mu|}{M} \quad (5.11)$$

Allora la condizione necessaria e sufficiente per generare un segnale di stato naturale di variazione nell'istante di tempo $t \geq M$ può essere riformulata dalla seguente relazione supponendo di avere $\mu = 0$.

$$\forall t \geq M; b_{nn,t} = 1 \Leftrightarrow MAD_t(M) \leq 2l(1 - \rho) \quad (5.12)$$

5.6.1 Modello Statistico

Nel presente paragrafo sono riportate alcuni risultati relativi alla statistica $MAD(M)$ calcolata considerando il valore di $\mu = 0$, ovvero:

$$MAD_i(M) = \frac{\sum_{r=1}^M |Y_{t-M+r}|}{M} \quad (5.13)$$

Come indicato nella Figura 5.1, supponendo di modellare il processo ad ogni istante di tempo τ come $Y_\tau = Z_\tau + S_\tau$, essendo $Z_\tau \approx NID(0,1)$ e S_τ un generico segnale di disturbo con il quale è modellato la presenza di una causa assegnabile (non naturale) di variazione, allora si dimostra che valgono le seguenti relazioni rispettivamente per il valore atteso e per la varianza della grandezza statistica $MAD(M)$.

$$E[MAD_i(M)] = \sum_{r=1}^M E[|Y_{t-M+r}|] \quad (5.14)$$

$$VAR[MAD_i(M)] = \frac{1}{M^2} \sum_{r=1}^M VAR[|Y_{t-M+r}|] \quad (5.15)$$

Dove, assumendo $l \rightarrow +\infty$, si ha che:

$$E[|Y_\tau|] = \sqrt{\frac{2}{\pi}} e^{-\frac{S_\tau^2}{2}} + S_\tau (1 - 2\phi(-S_\tau)) \quad (5.16)$$

$$VAR[|Y_\tau|] = - \left[\sqrt{\frac{2}{\pi}} e^{-\frac{S_\tau^2}{2}} + S_\tau (1 - 2\phi(-S_\tau)) \right]^2 + 1 + S_\tau^2 \quad (5.17)$$

Con $\phi(x) = P\{Z \leq x\}$, $Z \approx N(0,1)$, ovvero la funzione di densità cumulata per una distribuzione normale di media nulla e deviazione standard 1.

Nel caso di processo in uno stato naturale ($\forall \tau: S_\tau = 0$) le precedenti relazioni si semplificano nelle seguenti.

$$E[MAD_i(M)] = \sqrt{\frac{2}{\pi}} \quad (5.18)$$

$$VAR[MAD_i(M)] = \frac{\pi - 2}{M\pi} \quad (5.19)$$

Nel seguito un valore del limite $l \in [3,6]$ sarà considerato, per semplicità, equivalente alla condizione $l \rightarrow +\infty$.

5.6.1.1 Dimostrazione media MAD

$$\forall \tau: Y_\tau = Z_\tau + S_\tau; Z_\tau \approx NID(0,1);$$

$$E[|Y_\tau|] = \int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-S_\tau)^2}{2}} dx \Rightarrow$$

$$\int_{-\infty}^{+\infty} |x| \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-S_\tau)^2}{2}} dx = \int_0^{+\infty} x \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-S_\tau)^2}{2}} dx - \int_{-\infty}^0 x \frac{1}{\sqrt{2\pi}} e^{-\frac{(x-S_\tau)^2}{2}} dx =$$

$$z = x - S_\tau \Rightarrow$$

$$\begin{aligned}
 &= \int_{-S_\tau}^{+\infty} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz + \int_{-S_\tau}^{+\infty} S_\tau \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz - \int_{-\infty}^{-S_\tau} z \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz - \int_{-\infty}^{-S_\tau} S_\tau \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} dz = \\
 &= \left(-\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \Big|_{-S_\tau}^{+\infty} \right) + S_\tau (1 - \phi(-S_\tau)) - \left(-\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \Big|_{-\infty}^{-S_\tau} \right) - S_\tau \phi(-S_\tau) = \\
 &= \frac{1}{\sqrt{2\pi}} e^{-\frac{S_\tau^2}{2}} + S_\tau (1 - \phi(-S_\tau)) + \frac{1}{\sqrt{2\pi}} e^{-\frac{S_\tau^2}{2}} - S_\tau \phi(-S_\tau) = \\
 &= \frac{1}{\sqrt{2\pi}} \left(2e^{-\frac{S_\tau^2}{2}} \right) + S_\tau (1 - 2\phi(-S_\tau))
 \end{aligned}$$

Quindi

$$\begin{aligned}
 E[MAD_t(M)] &= E \left[\frac{\sum_{t=1}^M |Y_t|}{M} \right] = \frac{\sum_{t=1}^M \left[\frac{2}{\sqrt{2\pi}} \left(e^{-\frac{(S_t)^2}{2}} \right) + S_t (1 - 2\phi(-2S_t)) \right]}{M} = \\
 E[MAD_t(M)] &= \sqrt{\frac{2}{\pi}} \frac{1}{M} \sum_{t=1}^M e^{-\frac{(S_t)^2}{2}} + \frac{1}{M} \sum_{t=1}^M S_t (1 - 2\phi(-2S_t))
 \end{aligned}$$

5.6.1.2 Dimostrazione varianza MAD

Considerando il valore asintotico della grandezza $VAR[MAD(M)]$ al limite $l \rightarrow +\infty$ risulta che:

$$Y_\tau = Z_\tau + S_\tau; Z_\tau \approx NID(0,1)$$

$$E[|Y_\tau|^2] = E[(Y_\tau)^2] = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{(x-S_\tau)^2}{2}} dx$$

$$\begin{aligned}
 z = x - S_\tau &\Rightarrow \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} x^2 e^{-\frac{(x-S_\tau)^2}{2}} dx = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} (z + S_\tau)^2 e^{-\frac{z^2}{2}} dz = \\
 &= \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} z^2 e^{-\frac{z^2}{2}} dz + \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} S_\tau^2 e^{-\frac{z^2}{2}} dz + \int_{-\infty}^{+\infty} \frac{2}{\sqrt{2\pi}} S_\tau z e^{-\frac{z^2}{2}} dz;
 \end{aligned}$$

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} z^2 e^{-\frac{z^2}{2}} dz = \int_{-\infty}^{+\infty} -\frac{1}{\sqrt{2\pi}} z d \left(e^{-\frac{z^2}{2}} \right) = -\frac{1}{\sqrt{2\pi}} \left[z e^{-\frac{z^2}{2}} \Big|_{-\infty}^{+\infty} - \int_{-\infty}^{+\infty} e^{-\frac{z^2}{2}} dz \right] = 1;$$

$$\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} S_\tau^2 e^{-\frac{z^2}{2}} dz = S_\tau^2;$$

$$\int_{-\infty}^{+\infty} \frac{2}{\sqrt{2\pi}} S_\tau z e^{-\frac{z^2}{2}} dz = 2S_\tau \left[-\frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \Big|_{-\infty}^{+\infty} \right] = 0;$$

Essendo: $VAR[Y_\tau] = -E^2[Y_\tau] + E[Y_\tau^2]$, si ha che:

$$VAR[MAD_t(M)] = VAR\left[\frac{\sum_{r=1}^M |Y_{t-M+r}|}{M}\right] = \frac{1}{M^2} \sum_{r=1}^M \{1 + E[S_{t-M+r}^2] - E^2[Y_{t-M+r}]\} \quad (5.20)$$

5.6.2 Studio di casi specifici

In questa sezione sono applicati i risultati dimostrati nei precedenti paragrafi al caso in cui la componente non naturale del processo si può rappresentare attraverso uno dei segnali speciali di disturbo (shift, variazione sistematica, mixture e ciclo). Di seguito, si richiamano brevemente le formulazioni matematiche delle serie temporali utilizzate per modellare i quattro casi speciali di disturbo esaminati.

$$Shift : S_t = \begin{cases} 0 & t < \tau \\ \varphi & t \geq \tau \end{cases} \quad (5.21)$$

$$Systematic\ Variation : S_t = \varphi(-1)^t \quad (5.22)$$

$$Mixture : S_t = \varphi m_t; m_t \in \{-1, 1\} \quad (5.23)$$

$$Cycle : S_t = \varphi \cos\left(\frac{2\pi t}{\Omega}\right) \quad (5.24)$$

In particolare, si dimostra che nel caso dei segnali shift $\tau = 1$, variazione sistematica e mixture di ampiezza φ unità di deviazione standard, la media e varianza della quantità $MAD_t(M)$ assumono rispettivamente i valori di seguito riportati.

$$E[MAD_t(M)] = E[|Y_t|] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \varphi(1 - 2\phi(-2\varphi)) \quad (5.25)$$

$$VAR[MAD_t(M)] = \frac{1 - E^2[|Y_t|] + \varphi^2}{M} \quad (5.26)$$

Dalle due precedenti relazioni appare evidente che il valore atteso e la varianza per la grandezza $MAD_t(M)$ risultano indipendenti dall'istante t che, in particolare, il valore atteso di $MAD_t(M)$ sia anche indipendente dalla ampiezza della finestra utilizzata per la rete neurale (M). Al contrario la varianza della grandezza $MAD_t(M)$ è inversamente proporzionale alla ampiezza prescelta per la finestra.

Nelle figure seguenti sono riportati rispettivamente i valori di media ($E[MAD_t(M)]$) e di deviazione standard ($\sqrt{VAR[MAD_t(M)]}$) della grandezza MAD al variare dell'ampiezza del disturbo (φ) per i tre tipi di segnali sopra riportati (shift, variazione sistematica mixture).

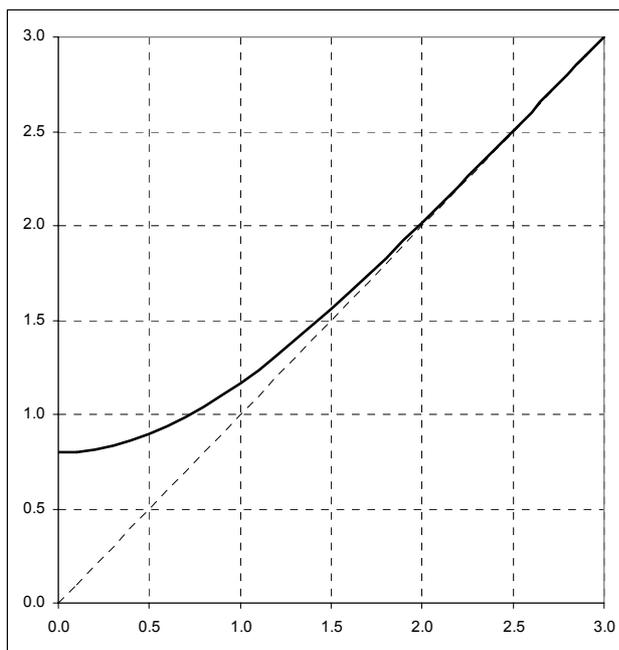


Figura 5.2 – Valore atteso del MAD al variare dell'entità di disturbo

Pattern shift, variazione sistematica e mixture.

Dal grafico precedente appare subito chiaro che piccoli disturbi sul processo non potranno essere facilmente identificabili dalla rete neurale, infatti, le variazioni del valore atteso del MAD sono sostanzialmente nulle per valori di disturbo dell'ordine di un decimo di unità di deviazione standard. Al contrario, per variazioni dell'ordine di un'unità di deviazione standard il valore atteso per la grandezza MAD appare seguire identicamente l'entità del disturbo stesso.

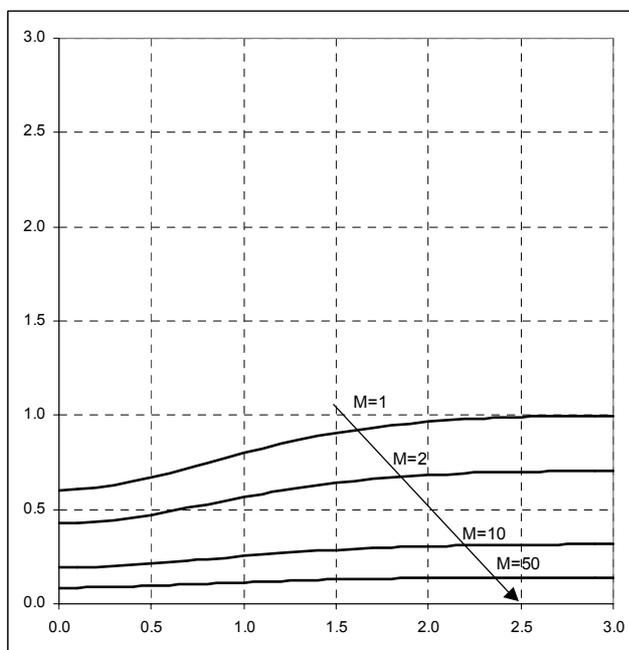


Figura 5.3 – Deviazione standard del MAD al variare dell'entità del disturbo per differenti finestre

Pattern shift, variazione sistematica e mixture.

Dal grafico della Figura 5.3 si nota che, aumentando il disturbo aumenta la dispersione della grandezza statistica MAD utilizzata nell'algoritmo neurale per decidere sullo stato

del processo. Ad ogni modo, aumentando la dimensione della finestra di punti si riduce sia la dispersione della statistica MAD, sia la dipendenza di tale dispersione dall'entità del disturbo.

Per quanto riguarda il pattern ciclo ($\Omega = 4$) è possibile dimostrare le seguenti formulazioni rispettivamente per il valore atteso e per la varianza della grandezza $MAD_i(M)$.

$$E[MAD_i(M)] = E[Y_i] = \frac{1}{2} \left[\sqrt{\frac{2}{\pi}} \left(e^{-\frac{\varphi^2}{2}} + 1 \right) + \varphi(1 - 2\phi(-2\varphi)) \right] \quad (5.27)$$

$$VAR[MAD_i(M)] = \frac{\pi - 2}{2M\pi} + \frac{1 - E^2[Y_i] + \varphi^2}{2M} \quad (5.28)$$

Nelle figure seguenti (Figura 5.4 e Figura 5.5) sono riportati rispettivamente i valori di media ($E[MAD_i(M)]$) e di deviazione standard ($\sqrt{VAR[MAD_i(M)]}$) della grandezza MAD al variare dell'entità del disturbo per un pattern di tipo ciclo di periodo 4 ($\Omega = 4$).

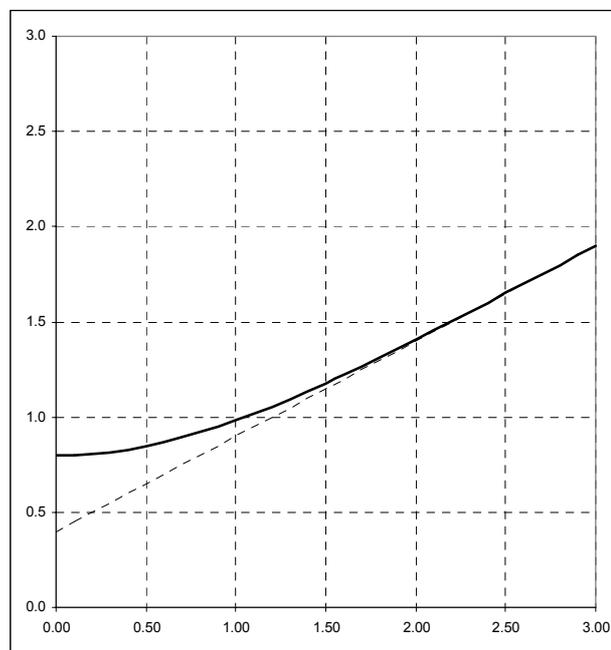


Figura 5.4 – Valore atteso del MAD al variare dell'entità di disturbo per un ciclo di periodo 4.

Il tasso di crescita del valore atteso per il MAD nel caso di ciclo di periodo 4 è, a parità di disturbo, più contenuto rispetto a quelli evidenziati nel caso di shift, variazione sistematica e mixture. Analogamente si può affermare che, a parità di entità di disturbo, mentre le capacità della rete neurale nell'identificare shift, variazioni sistematiche e mixture sono attese equivalenti, la capacità di identificare cicli appare essere più ridotta.

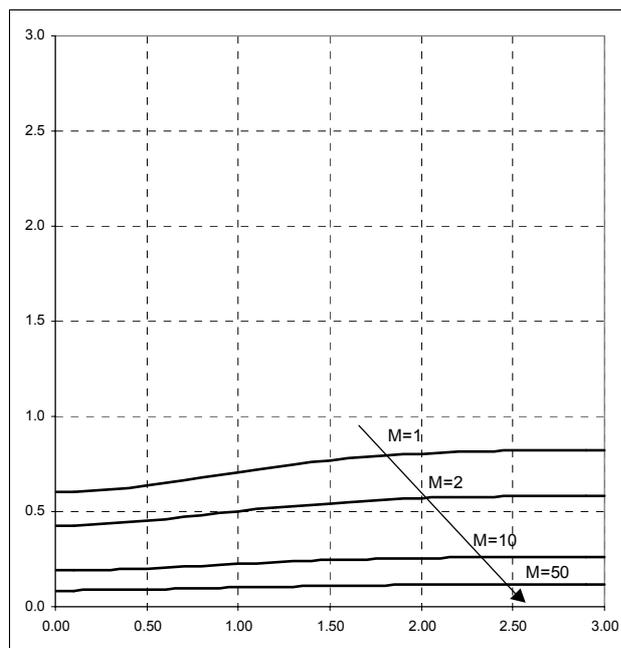


Figura 5.5 – Deviazione standard del MAD al variare dell'entità del disturbo per differenti finestre per un ciclo di periodo 4.

Nei seguenti sottoparagrafi si fornisce la dimostrazione delle relazioni riportate in precedenza.

5.6.2.1 Shift

$$E[MAD_t(M)] = E\left[\frac{\sum_{r=1}^M |Y_{t-M+r}|}{M}\right] = \sqrt{\frac{2}{\pi}} \frac{1}{M} \sum_{r=1}^M e^{-\frac{(S_{t-M+r})^2}{2}} + \frac{1}{M} \sum_{r=1}^M S_{t-M+r} (1 - 2\phi(-2S_{t-M+r}))$$

$$\text{Shift : } S_t = \begin{cases} 0 & t < \tau \\ \varphi & t \geq \tau \end{cases} \Rightarrow$$

$$E[MAD_t(M)] = \begin{cases} \sqrt{\frac{2}{\pi}} & t < \tau \\ \frac{t - \tau + 1}{M} \left[\sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \varphi(1 - 2\phi(-2\varphi)) \right] & \tau \leq t < M + \tau - 1 \\ \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \varphi(1 - 2\phi(-2\varphi)) & M + \tau - 1 \leq t \end{cases}$$

5.6.2.2 Variazione Sistemica

Nel seguito la sommatoria $\sum_{r=1}^M [\cdot]_{t-M+r}$ sarà considerata composta da due componenti:

$$\sum_{r=1}^M [\cdot]_{t-M+r} = \sum_{s=1}^{M_d} [\cdot]_s \Big|_{\forall s=1, \dots, M_p, (t-M+r)=2i+1} + \sum_{u=1}^{M_p} [\cdot]_u \Big|_{\forall u=1, \dots, M_d, (t-M+r)=2i} ; i = 1, 2, \dots$$

Vale a dire una componente di indici temporali pari ($s = 1, \dots, M_d : (t - M + r) = 2i + 1$) ed una componente di indici temporali dispari ($u = 1, \dots, M_p : (t - M + r) = 2i$) dove $M_d + M_p = M$. Pertanto, la media $E[MAD_t(M)]$ può essere espressa nel modo seguente.

$$E[MAD_t(M)] = E \left[\frac{\sum_{r=1}^M |Y_{t-M+r}|}{M} \right] = \sqrt{\frac{2}{\pi}} \frac{1}{M} \sum_{r=1}^M e^{-\frac{(S_{t-M+r})^2}{2}} + \frac{1}{M} \sum_{r=1}^M S_{t-M+r} (1 - 2\phi(-2S_{t-M+r}))$$

Systematic Variation : $S_t = \varphi(-1)^t \Rightarrow$

$$M_d + M_p = M$$

$$E[MAD_t(M)] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \frac{\varphi}{M} \left[\sum_{s=1}^{M_d} -(1 - 2\phi(2\varphi)) + \sum_{u=1}^{M_p} (1 - 2\phi(-2\varphi)) \right] =$$

$$E[MAD_t(M)] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \frac{\varphi}{M} \left[\sum_{s=1}^{M_d} [-1 + 2(1 - \phi(-2\varphi))] + \sum_{u=1}^{M_p} (1 - 2\phi(-2\varphi)) \right] =$$

$$E[MAD_t(M)] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \frac{\varphi}{M} \left[\sum_{s=1}^{M_d} (1 - 2\phi(-2\varphi)) + \sum_{u=1}^{M_p} (1 - 2\phi(-2\varphi)) \right] =$$

$$E[MAD_t(M)] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \frac{\varphi}{M} \left[\sum_{r=1}^M (1 - 2\phi(-2\varphi)) \right] =$$

$$E[MAD_t(M)] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \varphi(1 - 2\phi(-2\varphi))$$

5.6.2.3 Mixture

Consideriamo la sommatoria:

$$\sum_{r=1}^M [\cdot]_{t-M+r} = \sum_{s=1}^{M_d} [\cdot]_s \Big|_{\forall s=1, \dots, M_d : m_{t-M+r} = -1} + \sum_{u=1}^{M_p} [\cdot]_u \Big|_{\forall u=1, \dots, M_p : m_{t-M+r} = 1} ;$$

costituita da una componente di M_d indici temporali tali che $s = 1, \dots, M_d : m_{t-M+r} = -1$, ed una componente di indici M_p temporali tali che $u = 1, \dots, M_p : m_{t-M+r} = 1$, dove $M_d + M_p = M$. Pertanto, la media $E[MAD_t(M)]$ può essere espressa nel modo seguente.

$$E[MAD_t(M)] = E\left[\frac{\sum_{r=1}^M |Y_{t-M+r}|}{M}\right] = \sqrt{\frac{2}{\pi}} \frac{1}{M} \sum_{r=1}^M e^{-\frac{(S_{t-M+r})^2}{2}} + \frac{1}{M} \sum_{r=1}^M S_{t-M+r} (1 - 2\phi(-2S_{t-M+r}))$$

Mixture : $S_t = \varphi m_t, m_t \in \{-1, 1\} \Rightarrow$

$$M_d + M_p = M$$

$$E[MAD_t(M)] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \frac{\varphi}{M} \left[\sum_{s=1}^{M_d} -(1 - 2\phi(2\varphi)) + \sum_{u=1}^{M_p} (1 - 2\phi(-2\varphi)) \right] =$$

$$E[MAD_t(M)] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \frac{\varphi}{M} \left[\sum_{s=1}^{M_d} [-1 + 2(1 - \phi(-2\varphi))] + \sum_{u=1}^{M_p} (1 - 2\phi(-2\varphi)) \right] =$$

$$E[MAD_t(M)] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \frac{\varphi}{M} \left[\sum_{s=1}^{M_d} (1 - 2\phi(-2\varphi)) + \sum_{u=1}^{M_p} (1 - 2\phi(-2\varphi)) \right] =$$

$$E[MAD_t(M)] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \frac{\varphi}{M} \left[\sum_{r=1}^M (1 - 2\phi(-2\varphi)) \right] =$$

$$E[MAD_t(M)] = \sqrt{\frac{2}{\pi}} e^{-\frac{\varphi^2}{2}} + \varphi(1 - 2\phi(-2\varphi))$$

Si osservi che, il valore atteso del MAD per il pattern mixture non risulta essere dipendente dalle probabilità di transizione di stato per la catena di Markov con cui il disturbo è modellato. Questo risultato è congruente con quelli precedenti, in quanto sia uno shift che una variazione sistematica possono essere modellati come una mixture con particolari valori delle probabilità di transizione di stato.

5.6.2.4 Ciclo

Poniamo la sommatoria $\sum_{r=1}^M [\cdot]_{t-M+r}$ come somma di tre componenti.

$$\sum_{r=1}^M [\cdot]_{t-M+r} = \sum_{v=1}^{M_n} [\cdot]_v \Big|_{\forall v=1, \dots, M_n: S_{t-M+r}=0} + \sum_{s=1}^{M_d} [\cdot]_s \Big|_{\forall s=1, \dots, M_d: S_{t-M+r}=-\varphi} + \sum_{u=1}^{M_p} [\cdot]_u \Big|_{\forall u=1, \dots, M_p: S_{t-M+r}=\varphi} ;$$

vale a dire una componente di indici temporali tali che $v=1, \dots, M_n : S_{t-M+r} = 0$, una componente di indici temporali tali che $s=1, \dots, M_d : S_{t-M+r} = -\varphi$, una ulteriore componente di indici tali che $u=1, \dots, M_p : S_{t-M+r} = \varphi$, dove $M_d + M_p + M_n = M$. La media $E[MAD_t(M)]$ può essere espressa nel modo seguente.

$$E[MAD_t(M)] = E\left[\frac{\sum_{r=1}^M |Y_{t-M+r}|}{M}\right] = \sqrt{\frac{2}{\pi}} \frac{1}{M} \sum_{r=1}^M e^{-\frac{(S_{t-M+r})^2}{2}} + \frac{1}{M} \sum_{r=1}^M S_{t-M+r} (1 - 2\phi(-2S_{t-M+r}))$$

Cycle : $S_t = \varphi \cos\left(\frac{2\pi t}{4}\right) \Rightarrow$

$$M_p + M_d + M_n = M$$

$$\begin{aligned}
 E[MAD_t(M)] &= \frac{1}{M} \sqrt{\frac{2}{\pi}} \left[(M - M_n) e^{-\frac{\varphi^2}{2}} + M_n \right] + \frac{\varphi}{M} \left[\sum_{s=1}^{M_d} (1 - 2\phi(2\varphi)) + \sum_{u=1}^{M_p} (1 - 2\phi(-2\varphi)) \right] = \\
 &= \frac{1}{M} \sqrt{\frac{2}{\pi}} \left[(M - M_n) e^{-\frac{\varphi^2}{2}} + M_n \right] + \frac{\varphi}{M} \left[\sum_{s=1}^{M_d} [-1 + 2(1 - \phi(-2\varphi))] + \sum_{u=1}^{M_p} (1 - 2\phi(-2\varphi)) \right] = \\
 &= \frac{1}{M} \sqrt{\frac{2}{\pi}} \left[(M - M_n) e^{-\frac{\varphi^2}{2}} + M_n \right] + \frac{\varphi}{M} \left[\sum_{s=1}^{M_d} (1 - 2\phi(-2\varphi)) + \sum_{u=1}^{M_p} (1 - 2\phi(-2\varphi)) \right] = \\
 &= \frac{1}{M} \sqrt{\frac{2}{\pi}} \left[(M - M_n) e^{-\frac{\varphi^2}{2}} + M_n \right] + \frac{\varphi}{M} \left[\sum_{r=1}^{M-M_n} (1 - 2\phi(-2\varphi)) \right] = \\
 &= \frac{1}{M} \sqrt{\frac{2}{\pi}} \left[(M - M_n) e^{-\frac{\varphi^2}{2}} + M_n \right] + \frac{M - M_n}{M} \varphi (1 - 2\phi(-2\varphi)) \\
 M \text{ pari} \Rightarrow E[MAD_t(M)] &= \frac{1}{2} \left[\sqrt{\frac{2}{\pi}} \left(e^{-\frac{\varphi^2}{2}} + 1 \right) + \varphi (1 - 2\phi(-2\varphi)) \right] \\
 M \text{ dispari} \Rightarrow &\begin{cases} \sqrt{\frac{2}{\pi}} \left[\frac{\lfloor M - M_n \rfloor}{M} e^{-\frac{\varphi^2}{2}} + \frac{\lceil M - M_n \rceil}{M} \right] + \frac{\lfloor M - M_n \rfloor}{M} \varphi (1 - 2\phi(-2\varphi)); & t = \text{pari} \\ \sqrt{\frac{2}{\pi}} \left[\frac{\lceil M - M_n \rceil}{M} e^{-\frac{\varphi^2}{2}} + \frac{\lfloor M - M_n \rfloor}{M} \right] + \frac{\lceil M - M_n \rceil}{M} \varphi (1 - 2\phi(-2\varphi)); & t = \text{dispari} \end{cases} \\
 M \gg 1 \Rightarrow E[MAD_t(M)] &= \frac{1}{2} \left[\sqrt{\frac{2}{\pi}} \left(e^{-\frac{\varphi^2}{2}} + 1 \right) + \varphi (1 - 2\phi(-2\varphi)) \right]
 \end{aligned}$$

Una simile dimostrazione per il valore della varianza di $VAR[MAD_t(M)]$ nel caso di un ciclo di periodo $\Omega = 4$ può essere ottenuto considerando ($M \gg 1$) una media tra gli istanti di tempo in cui il disturbo ha ampiezza nulla e quelli nei quali ha ampiezza pari in valore assoluto a $u = 1, \dots, M_p : S_{t-M+r} = \varphi$, ottenendo il seguente risultato:

$$VAR[MAD_t(M)] = \frac{\pi - 2}{2M\pi} + \frac{1 - E^2[Y_t] + \varphi^2}{2M}.$$

5.7 Discussione

Il modello statistico sopra presentato, permette di fare alcune osservazioni sulle prestazioni attese dall' algoritmo neurale proposto per il monitoraggio di un processo.

La decisione sullo stato del processo è basata su un confronto a soglia della grandezza MAD con un limite dipendente dal parametro di vigilanza ρ . Il livello atteso per la grandezza MAD (che si è dimostrato essere indipendente dalla scelta dell'ampiezza di finestra) aumenta al crescere dell'entità di un disturbo φ che influisce sui dati di uscita prodotti dal processo. Ciò fornisce un'evidenza teorica all'attendibilità del risultato fornito dall' algoritmo neurale proposto nel lavoro.

L'aumento dell'ampiezza della finestra di osservazione M , pur non influenzando sul livello medio atteso per la grandezza MAD , influisce sulla dispersione dei valori osservati attorno a tale media. In particolare aumentando la finestra di osservazioni diminuisce la dispersione attorno al valore medio, e dunque, poiché la discriminazione fra stato naturale e non naturale dipende da un confronto a soglia con un limite superiore per le grandezze MAD osservate, migliora la capacità di discriminare lo stato naturale da quello non naturale riducendo sia l'errore di Tipo I, sia l'errore di Tipo II.

Gli istogrammi presentati nella Figura 5.6 esemplificano quanto sopra esposto ad un caso pratico di 2 campioni di 10000 osservazioni. E' rappresentata la distribuzione della grandezza MAD per differenti livelli del parametro M (1, 10, 25 e 75), sia per un processo in uno stato naturale di variabilità (colonne nere), sia per uno shift della media (1.5 unità di deviazione standard, colonne grigie). Negli stessi grafici, inoltre, è rappresentato il limite a soglia derivante dalla scelta di un parametro di vigilanza pari a $\rho = 0.82$. L'errore di Tipo I è pari alla percentuale di osservazioni per le quali il valore del MAD corrispondente assume valori superiori al limite stabilito (colonne nere a destra del limite). Analogamente, l'errore di Tipo II è pari alla percentuale di osservazioni per le quali il valore del MAD corrispondente assume valori inferiori al limite stabilito (colonne grigie a sinistra del limite).

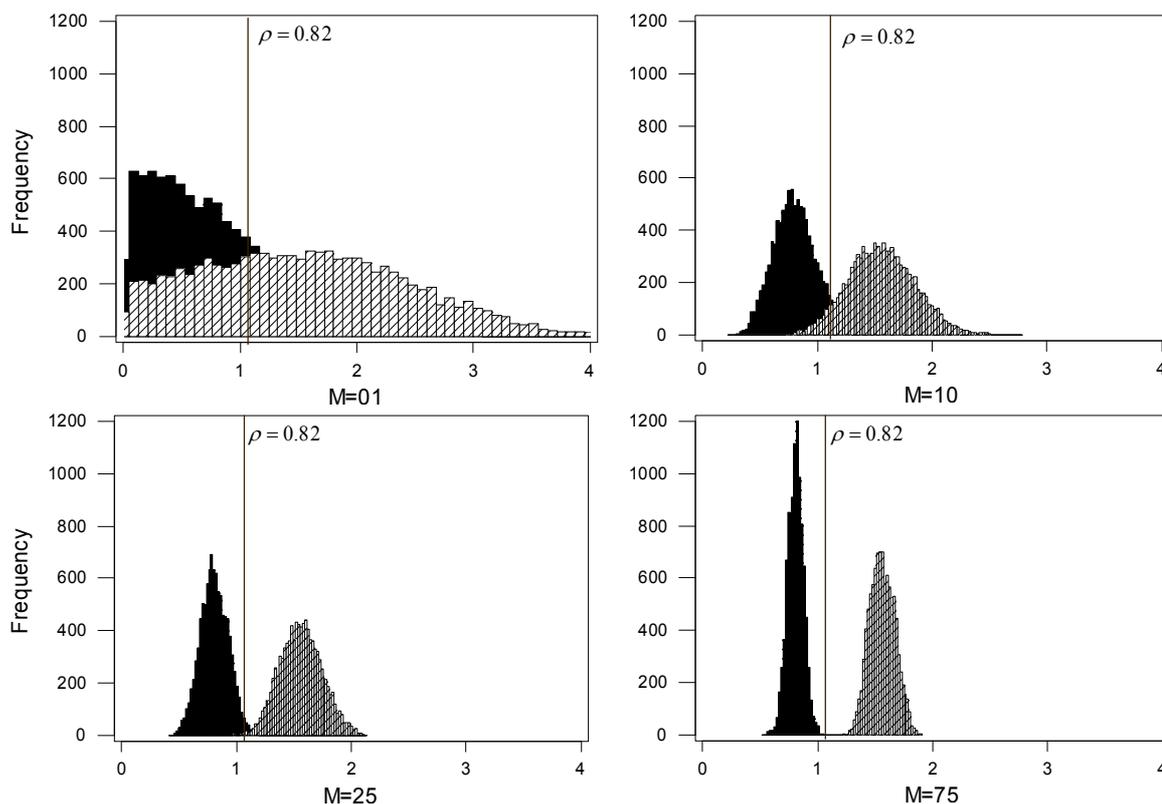


Figura 5.6 – Distribuzione di frequenza della misura MAD su 10000 dati raggruppati in finestre da 1, 10, 50 e 75 punti.

Dati normali media nulla e deviazione standard unitaria (istogramma scuro), dati normali media=1.5 e deviazione standard unitaria (istogramma chiaro)

È evidente che aumentando la finestra utilizzata per il calcolo della grandezza MAD la diminuzione della dispersione attorno al valore medio consente di limitare sia l'errore di Tipo I, sia l'errore di Tipo II.

Dai grafici sopra riportati, inoltre, appare evidente l'effetto del parametro di vigilanza. Nelle figure è stato riportato il limite dipendente dal complementare a 1 del

livello di vigilanza utilizzato (in particolare è riportato il livello $6(1 - \rho)$ calcolato per $\rho = 0.82$). Risulta che, al diminuire della vigilanza il limite superiore di variazione per la grandezza MAD per considerare la finestra di M dati generata da un processo naturale, cresce (graficamente il limite si sposta verso destra). Conseguentemente, al diminuire della vigilanza, diminuisce l'errore di Tipo I ma aumenta l'errore di Tipo II. Viceversa, all'aumentare della vigilanza diminuisce la soglia massima per la grandezza MAD che permette di considerare il processo in uno stato naturale (graficamente il limite si sposta verso sinistra), conseguentemente si riduce l'errore di Tipo II ma aumenta l'errore di Tipo I.

5.7.1 Alcune proprietà statistiche sull'errore di Tipo I

Considerando la disuguaglianza di Čebyšev (Papoulis 1973) secondo la quale $\forall \xi > 0$ risulta che, e indicando con $\alpha = P\{H_1|H_0\}$ l'errore di Tipo I, è possibile scrivere la seguente relazione.

$$P\{MAD_i(M) < E[MAD_i(M)] + \xi\} \geq P\{|MAD_i(M) - E[MAD_i(M)]| < \xi\} \geq 1 - \frac{VAR[MAD_i(M)]}{\xi^2}$$

$$E[MAD_i(M)] + \xi = 2l(1 - \rho) \Rightarrow \xi = 2l(1 - \rho) - E[MAD_i(M)] \quad (5.29)$$

$$1 - \alpha = P\{MAD_i(M) < 2l(1 - \rho)\} \geq 1 - \frac{VAR[MAD_i(M)]}{[2l(1 - \rho) - E[MAD_i(M)]]^2} \quad (5.30)$$

$$\xi > 0 \Leftrightarrow \rho < 1 - \frac{E[MAD_i(M)]}{2l} \quad (5.31)$$

$$\rho < \left(1 - \frac{E[MAD_i(M)]}{2l}\right) \Rightarrow \alpha \leq \frac{VAR[MAD_i(M)]}{[2l(1 - \rho) - E[MAD_i(M)]]^2} \quad (5.32)$$

$$(5.33)$$

In modo analogo si ha:

$$P\{MAD_i(M) > E[MAD_i(M)] - \xi\} \geq P\{|MAD_i(M) - E[MAD_i(M)]| < \xi\} \geq 1 - \frac{VAR[MAD_i(M)]}{\xi^2}$$

$$E[MAD_i(M)] - \xi = 2l(1 - \rho) \Rightarrow \xi = E[MAD_i(M)] - 2l(1 - \rho) \quad (5.34)$$

$$\xi > 0 \Leftrightarrow \rho > 1 - \frac{E[MAD_i(M)]}{2l} \quad (5.35)$$

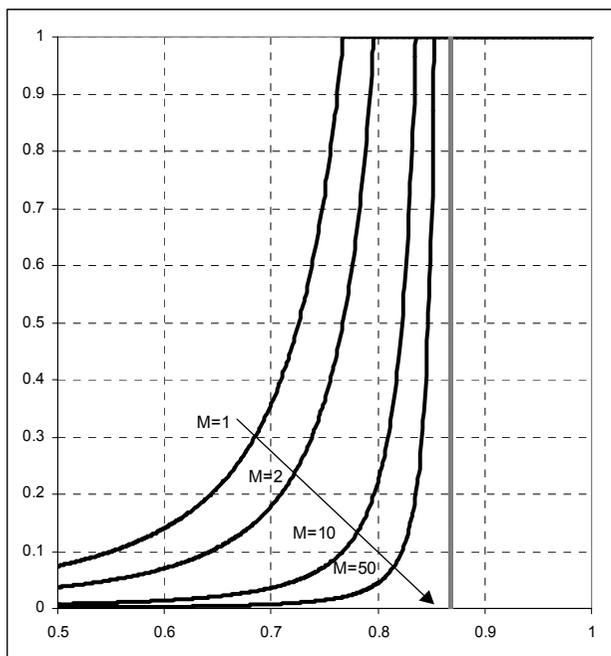
$$1 - \beta = P\{MAD_i(M) > 2l(1 - \rho)\} \geq 1 - \frac{VAR[MAD_i(M)]}{[E[MAD_i(M)] - 2l(1 - \rho)]^2} \quad (5.36)$$

$$\rho > 1 - \frac{E[MAD_i(M)]}{2l} \Rightarrow \beta \leq \frac{VAR[MAD_i(M)]}{[E[MAD_i(M)] - 2l(1 - \rho)]^2} \quad (5.37)$$

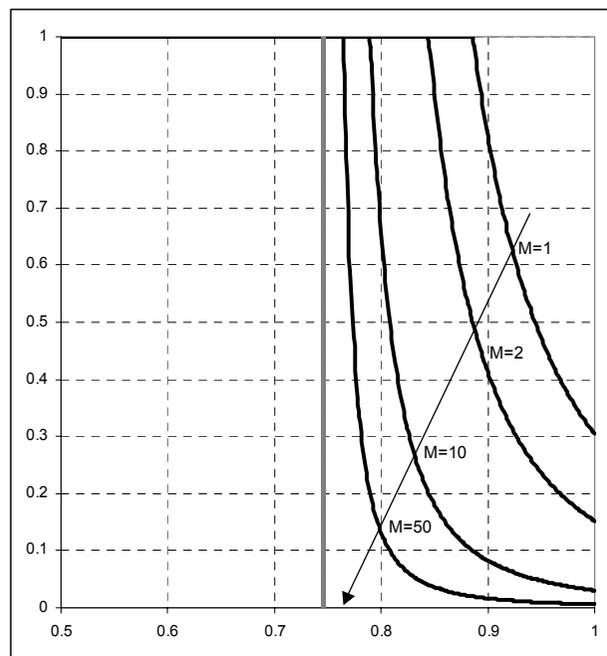
In altri termini, a parità di livello di vigilanza ρ il limite superiore dell'errore di Tipo I è inversamente proporzionale all'ampiezza della finestra M , mentre, a parità di finestra di

punti M , esso risulta direttamente dipendente dal livello di vigilanza scelto ρ . Appare, quindi, quanto già osservato empiricamente, ovvero che l'errore di Tipo I decresce all'aumentare della finestra d'osservazione e al diminuire della vigilanza adottata per la rete neurale.

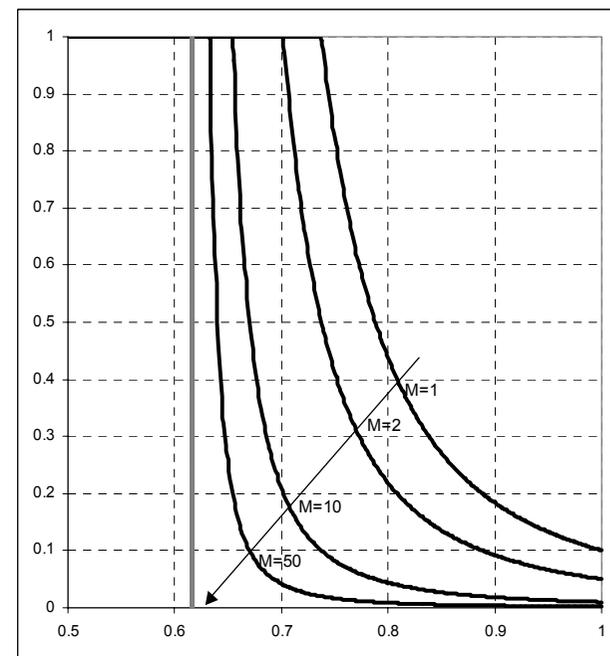
Le seguenti figure, rappresentano il valore del limite superiore ottenuto dalla disuguaglianza di Čebyšev sia per l'errore di Tipo I, che per l'errore di Tipo II. Sono stati presi in considerazione i casi: $l = 3,6$, $M = 1,2,10,50$ e $\varphi = 1.5,2.5$. In entrambi i casi $l = 3,6$ i valori di media e deviazione standard della grandezza MAD sono stati calcolati in modo esatto considerando il valore finito $l < +\infty$.



(a) Limite superiore per errore di Tipo I (ordinate) contro il parametro di vigilanza (ascissa).



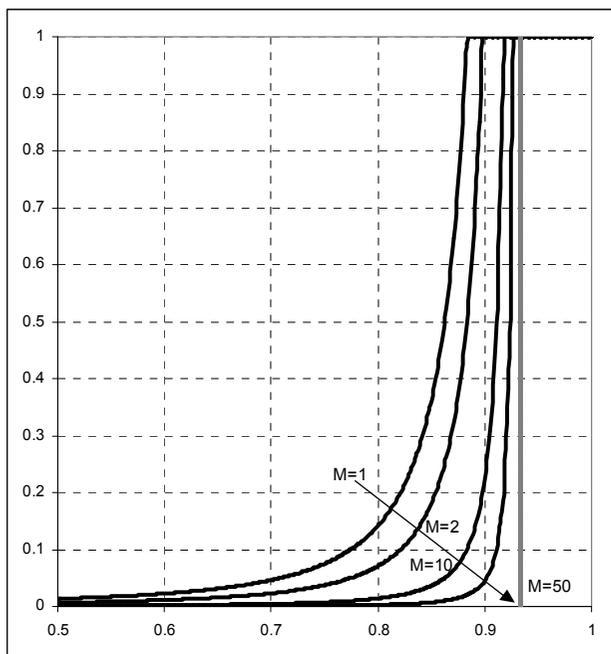
(b) Limite superiore per errore di Tipo II (ordinate) contro il parametro di vigilanza (ascissa). Grandezza del disturbo pari a 1.5 unità di deviazione standard.



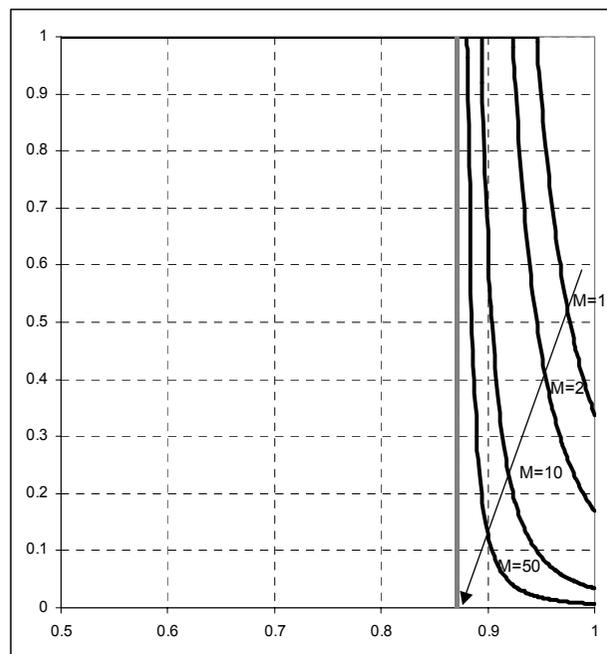
(c) Limite superiore per errore di Tipo II (ordinate) contro il parametro di vigilanza (ascissa). Grandezza del disturbo pari a 2.5 unità di deviazione standard.

Figura 5.7 – Limiti superiori per gli errori di Tipo I e di Tipo II ottenuti dalla disuguaglianza di Čebyšev. Limite $l=3$; Window Size $M=1,2,10,50$.

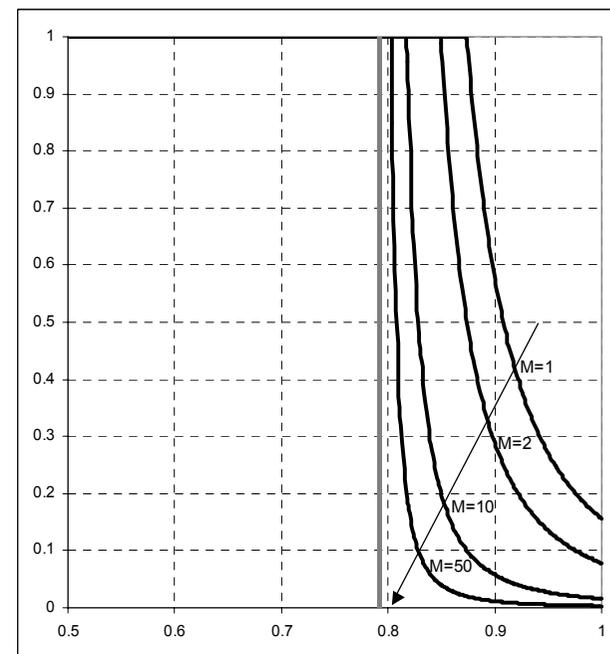
La linea verticale indica il valore superiore (a) o inferiore (b, c) per il parametro di vigilanza consentito dalla disuguaglianza di Čebyšev.



(a) Limite superiore per errore di Tipo I (ordinate) contro il parametro di vigilanza (ascissa).



(b) Limite superiore per errore di Tipo II (ordinate) contro il parametro di vigilanza (ascissa). Grandezza del disturbo pari a 1.5 unità di deviazione standard.



(c) Limite superiore per errore di Tipo II (ordinate) contro il parametro di vigilanza (ascissa). Grandezza del disturbo pari a 2.5 unità di deviazione standard.

Figura 5.8 – Limiti superiori per gli errori di Tipo I e di Tipo II ottenuti dalla disuguaglianza di Čebyšev. Limite $l=6$; Window Size $M=1,2,10,50$.

La linea verticale indica il valore superiore (a) o inferiore (b, c) per il parametro di vigilanza consentito dalla disuguaglianza di Čebyšev.

In sostanza, si può stabilire una similitudine fra una carta di controllo sulla media di un processo e la rete neurale. Si potrebbe, infatti, fare corrispondere la scelta della finestra Windows Size (M) alla dimensione del campione di osservazioni scelto nel progetto di una carta di controllo. Così come all'aumentare della dimensione del campione per una carta di controllo diminuisce la variabilità della grandezza osservata, così all'aumentare della Window Size diminuisce la variabilità della grandezza MAD valutata attraverso l'algoritmo neurale. Analogamente, il parametro di vigilanza adottato nella costruzione della rete, viene a corrispondere alla scelta dei limiti per una carta di controllo. Così come restringendo i limiti di una carta di controllo (per esempio inferiori all'ordinario valore di $\pm 3\sigma$) aumenta l'errore di Tipo I diminuendo quello di Tipo II, analogamente in un sistema a rete neurale Fuzzy ART aumentare la vigilanza implica delle condizioni più stringenti per considerare in controllo il processo e dunque un aumento dell'errore di Tipo I e una diminuzione dell'errore di Tipo II. In modo duale, una diminuzione del parametro di vigilanza equivale, agli effetti del controllo del processo, all'utilizzo di una carta di controllo con limiti di variazione più ampi (per esempio superiori a $\pm 3\sigma$).

5.8 Generalizzazione del modello

Sia data una rete Fuzzy ART a M ingressi, addestrata attraverso il metodo del Perfectly Learned Pattern (PLP) su un insieme di N . Indichiamo con

$$\underline{w}_j = (\underline{u}_j, \underline{1} - \underline{u}_j) \quad j = 1, 2, \dots, N \quad (5.38)$$

L'insieme dei N vettori di pesi top-down a $2M$ componenti. Supponiamo di aver codificato nello strato di uscita della rete neurale una funzione di mapping fra il neurone j^{esimo} del livello F2 e la classe di appartenenza (naturale, non naturale). In altri termini supponiamo di aver definito una funzione del tipo

$$B_{nn}(j): j \rightarrow b_{nn} \quad j = 1, 2, \dots, N; b_{nn} = 0, 1 \quad (5.39)$$

Si è codificato con il valore 0 un cluster da considerare non naturale, mentre un valore 1 è associato ad un cluster da considerare naturale. Assumiamo che ad un generico intervallo di tempo $t \geq M$ un ingresso M -dimensionale $\underline{I}_t = [I_{t-M+1}, \dots, I_{t-1}, I_t]$ sia presentato all'ingresso del livello F0 della rete neurale Fuzzy ART. Il cluster che risulterà conseguentemente attivato fra gli N appresi dalla rete neurale risulterà quello per cui è verificata la seguente condizione.

$$j = \arg \min_{1, 2, \dots, N} \left\{ \sum_{r=1}^M |I_{t-M+r} - u_{jr}| \right\} \quad (5.40)$$

Allora l'output del sistema neurale sarà posto uguale a $B_{nn}(j)$ se e solo se è verificata la seguente condizione

$$\sum_{r=1}^M |I_{t-M+r} - u_{jr}| \leq M(1 - \rho) \quad (5.41)$$

Se quest'ultima condizione non è verificata, il sistema non è in grado di classificare l'ingresso che risulta conseguentemente considerato non naturale ponendo il corrispondente valore di uscita all'istante t pari a 0.

5.8.1 Addestramento su target naturale e non naturale

In questa sezione si discuterà un caso particolare, del modello generale sopra presentato, che consente di ottenere un miglioramento delle prestazioni nel riconoscimento di specifici pattern non naturali.

Supponiamo di considerare l'addestramento della rete neurale basato su un insieme di due vettori: uno corrispondente al target naturale del processo, l'altro ad una alterazione non naturale del target. Impiegando il metodo di apprendimento PLP sulla coppia dei due pattern di training, il risultato dell'apprendimento consiste nella formazione di due distinte categorie: la prima rappresenta la media naturale del processo (indichiamo con \underline{w}_μ il vettore dei pesi top-down corrispondente e che l'uscita ad essa associata sia $b_{nn}=1$), e la seconda codifica il particolare disturbo non naturale (indichiamo con \underline{w}_s il vettore dei pesi top-down corrispondente e che l'uscita ad esso associata sia $b_{nn}=0$). In questo caso, quindi, la rete neurale sarà costituita da due neuroni di tipo *committed* nello strato $F2$.

In fase di testing, quando un pattern d'ingresso risulta appartenente al cluster naturale, allora l'algoritmo neurale produrrà il risultato $b_{nn,t}=1$. Al contrario, quando il pattern d'ingresso appartiene allo specifico cluster non naturale, oppure quando esso non può essere associato a nessuna delle categorie apprese dalla rete neurale, allora l'algoritmo Fuzzy ART produrrà il risultato in uscita $b_{nn,t}=0$ indicante che il vettore d'ingresso è da considerarsi non naturale.

Sia dunque $\underline{I}_t = [I_{t-M+1}, \dots, I_{t-1}, I_t]$, un ingresso presentato in fase di testing al livello $F0$ della rete neurale. Affinché esso sia riconosciuto come naturale dell'algoritmo neurale, ($b_{nn,t}=1$) è necessario che siano verificate due condizioni.

1. Il cluster naturale vinca la competizione sul cluster non naturale.
2. Il sistema di orientamento non annulli la categoria naturale.

La prima condizione si traduce nella seguente formulazione

$$\frac{|\underline{I}_t^c \wedge \underline{w}_\mu^c|}{\alpha + |\underline{w}_\mu^c|} \geq \frac{|\underline{I}_t^c \wedge \underline{w}_s^c|}{\alpha + |\underline{w}_s^c|} \quad (5.42)$$

La seconda condizione si traduce nella seguente.

$$\frac{|\underline{I}_t^c \wedge \underline{w}_\mu^c|}{|\underline{I}_t^c|} \geq \rho \quad (5.43)$$

Tenuto conto che $|\underline{I}_t^c| = |(\underline{I}_t, \underline{I} - \underline{I}_t)| = |\underline{I}_t| + M - |\underline{I}_t| = M$, così come risulta che $|\underline{w}_\mu^c| = |\underline{w}_s^c| = M$, le precedenti due condizioni possono essere riscritte rispettivamente come segue.

$$\begin{cases} |\underline{I}_t^c \wedge \underline{w}_\mu^c| \geq |\underline{I}_t^c \wedge \underline{w}_s^c| \\ |\underline{I}_t^c \wedge \underline{w}_\mu^c| \geq M\rho \end{cases} \quad (5.44)$$

O in modo equivalente come segue

$$|\underline{I}_t^c \wedge \underline{w}_\mu^c| \geq \max\{M\rho, |\underline{I}_t^c \wedge \underline{w}_s^c|\} \quad (5.45)$$

In altri termini, un dato input M -dimensionale è riconosciuto come naturale se due condizioni sono contemporaneamente verificate. 1) La sua distanza dal prototipo naturale è inferiore alla distanza presentata dal prototipo non naturale, 2) la sua distanza dal prototipo naturale è inferiore o uguale al limite $M(1 - \rho)$.

Questa estensione dell'algoritmo, come si dimostrerà nei capitoli successivi, consente di migliorare notevolmente le prestazioni dell'algoritmo neurale nel riconoscimento di specifici disturbi non naturali.

Capitolo 6

Risultati sperimentali: dati indipendenti

In questo capitolo sono riportati i risultati sperimentali (ottenuti per mezzo della simulazione Monte Carlo) dell'applicazione dell'algoritmo neurale al caso di monitoraggio di un processo manifatturiero. Il processo manifatturiero di riferimento è anch'esso riprodotto per mezzo della tecnica Monte Carlo in accordo ai modelli presentati e discussi nel capitolo 3. In particolare, i risultati presentati nel capitolo fanno riferimento al caso di un processo privo di memoria. In altre parole, si suppone di poter modellare le uscite, in uno stato naturale di variabilità, come una sequenza casuale di realizzazioni indipendenti e identicamente distribuite. I risultati sperimentali, da un lato, forniscono una validazione alle relazioni sulle prestazioni attese dalla rete e introdotte nel capitolo 5, dall'altro, confrontano le prestazioni ottenute con quelle realizzabili attraverso l'impiego di tecniche SPC.

6.1 Analisi delle prestazioni

Come riportato dal modello dell'algoritmo neurale Fuzzy ART descritto nel precedente capitolo, il *criterio di generalizzazione* dipende unicamente da due parametri: l'ampiezza di finestra M , e il parametro di vigilanza ρ . In particolare, l'ampiezza di finestra M influenza la misura MAD calcolata dall'algoritmo sulle osservazioni prodotte dal processo. Il parametro di vigilanza ρ determina il valore del limite di soglia superiore della grandezza MAD affinché il processo sia considerato in uno stato naturale.

L'ampiezza di finestra può assumere un qualunque valore intero $M \geq 1$. Il caso limite $M = 1$ implica che la rete neurale valuta lo stato del processo unicamente sulla base della più recente osservazione, mentre un valore $M > 1$, implica che l'algoritmo impiega le più recenti $M - 1$ osservazioni, oltre a quella attuale, per giudicare se il processo è in uno stato di variabilità naturale. Nel capitolo precedente è stata dimostrata una relazione di dipendenza inversa tra la varianza della misura MAD e l'ampiezza di finestra utilizzata (al crescere di M , diminuisce la varianza della grandezza MAD).

Il parametro di vigilanza può assumere valori variabili con continuità tra il limite inferiore $\rho = 0$ ed il limite superiore $\rho = 1$. Il complemento a 1 del parametro di vigilanza rappresenta (a meno di un fattore di proporzionalità che dipende dal tipo di codifica utilizzata per gli ingressi) la soglia superiore con la quale è confrontato il valore attuale della grandezza MAD . Il valore limite $\rho = 1$, pertanto, implica che la rete riconosce come naturale esclusivamente una sequenza di M punti coincidenti con il target sul quale è stata addestrata. Al contrario un valore di vigilanza pari al limite inferiore $\rho = 0$ implica che per qualunque sequenza di osservazioni il processo è considerato in uno stato naturale.

In questo paragrafo, sono esaminati in dettaglio gli effetti di questi due parametri sulle prestazioni della rete neurale misurate in termini sia di errore di Tipo I, sia di errore di Tipo II, e stimati per mezzo della simulazione Monte Carlo (Pacella *et al.* 2002a). I dettagli della metodologia di simulazione adottata sono riportati

nell'appendice A del testo. Ulteriori risultati analitici ottenuti dalle sperimentazioni sono stati raccolti nell'appendice B.

Di seguito sono valutate le prestazioni della rete neurale misurate in termini di errore di Tipo I stimato per quattro livelli dell'ampiezza di finestra $M = 10, 25, 50, 75$ ed otto livelli del parametro di vigilanza ρ variante nell'intervallo $[0.8, 0.975]$ con incrementi di 0.025. La seguente tabella riporta gli stimatori puntuali ($\hat{\alpha}$) ed intervallari (con copertura del 95% ed indicati dalla notazione $[\hat{\alpha}_-, \hat{\alpha}_+]$).

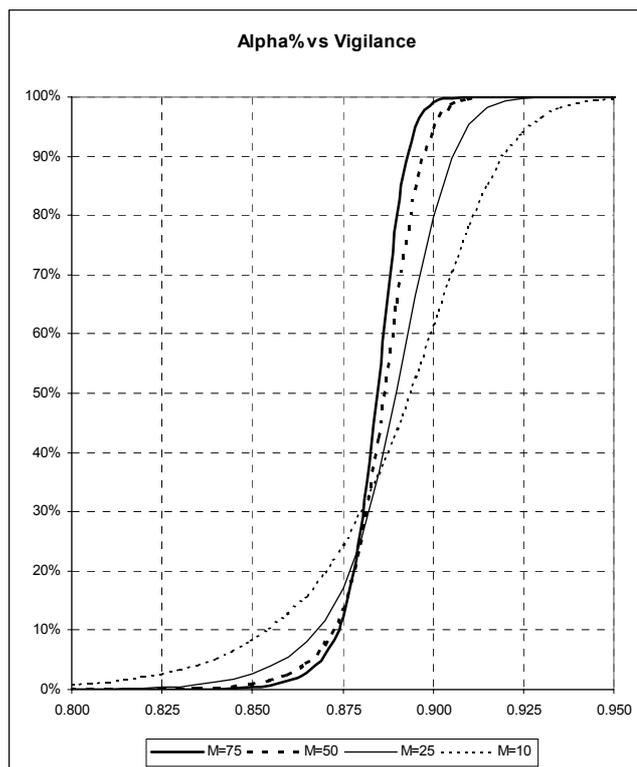
ρ	$M = 10$			$M = 25$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$
0.800	0.72%	0.77%	0.82%	0.02%	0.02%	0.03%
0.825	2.59%	2.70%	2.81%	0.32%	0.36%	0.39%
0.850	8.16%	8.38%	8.59%	2.48%	2.61%	2.73%
0.875	23.97%	24.39%	24.81%	16.67%	17.15%	17.64%
0.900	61.01%	61.62%	62.23%	79.05%	79.88%	80.70%
0.925	93.75%	94.11%	94.47%	99.64%	99.74%	99.83%
0.950	99.83%	99.86%	99.89%	100.00%	100.00%	100.00%
0.975	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

ρ	$M = 50$			$M = 75$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$
0.800	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.825	0.02%	0.03%	0.03%	0.00%	0.00%	0.01%
0.850	0.78%	0.84%	0.89%	0.34%	0.38%	0.43%
0.875	12.91%	13.48%	14.05%	11.75%	12.31%	12.88%
0.900	94.22%	94.78%	95.33%	98.83%	99.05%	99.26%
0.925	99.99%	100.00%	100.00%	100.00%	100.00%	100.00%
0.950	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
0.975	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

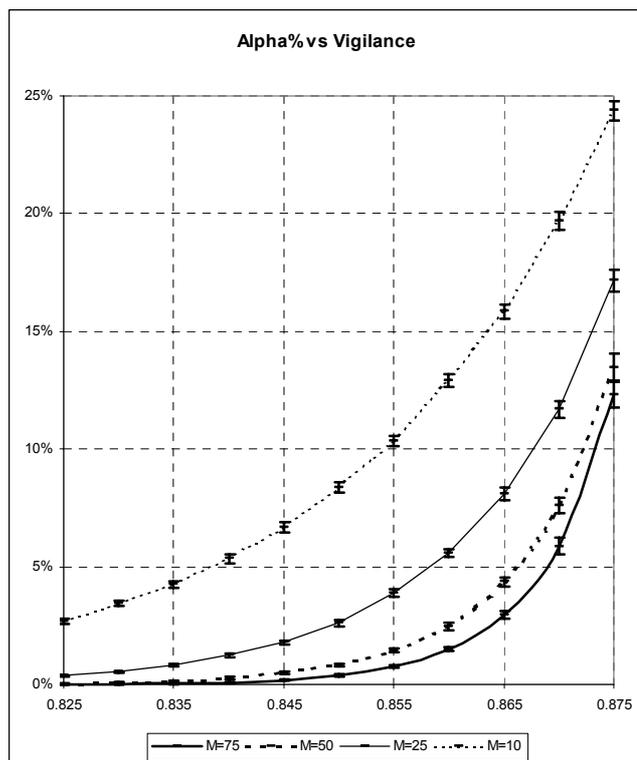
Tabella 6.1 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale (risultati di simulazione)

L'addestramento della rete Fuzzy ART è limitato sulla sola media nominale del processo.

Dai risultati della Tabella 6.1 appare evidente che, l'incremento del parametro di vigilanza causa un aumento dell'errore di Tipo I e che, se il parametro di vigilanza approssima il limite superiore $\rho=1$ allora anche l'errore di Tipo I approssima il limite superiore $\hat{\alpha}=100\%$. D'altro canto, quando il parametro di vigilanza decresce allora anche l'errore di Tipo I decresce avvicinandosi al limite inferiore $\hat{\alpha}=0\%$. Entrambe le proprietà risultano verificate a qualunque livello dell'ampiezza di finestra M , tuttavia, il tasso di crescita dell'errore di Tipo I, considerato come una funzione del parametro di vigilanza, aumenta al crescere dell'ampiezza di finestra M . In altre parole appare che gli effetti della variazione del tasso di vigilanza sull'errore di Tipo I presentato dalla rete neurale siano dipendenti dalla scelta dell'ampiezza di finestra M . La seguente Figura 6.1 riporta i grafici degli stimatori puntuali e intervallari $\hat{\alpha}$ dell'errore di Tipo I come funzione del parametro di vigilanza ρ . In particolare, le relazioni fra errore di Tipo I e parametro di vigilanza sono rappresentate per quattro differenti valori dell'ampiezza di finestra M .



a)



b)

Figura 6.1 – Stimatori puntuali degli errori di Tipo I (ordinate) e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)

a) intervallo di variazione ascissa è[0.800, 0.950], intervallo di variazione ordinate [0%, 100%]

b) intervallo di variazione ascissa è[0.825, 0.875], intervallo di variazione ordinate [0%, 25%] con rappresentazione degli stimatori intervallari a copertura 95%

Gli stimatori puntuali e intervallari dell'errore di Tipo II sono riportati nella seguente Tabella 6.2. Gli errori di Tipo II indicati si riferiscono al caso di un processo che presenta uno shift della media pari ad un'unità di deviazione standard ($\varphi = 1$), inoltre, è stato supposto che l'istante di partenza dello shift coincidesse con il primo istante di osservazione sul processo ($\tau = 1$). La notazione $[\hat{\beta}_-, \hat{\beta}_+]$ è stata usata per indicare l'intervallo di confidenza al 95% sullo stimatore $\hat{\beta}$ dell'errore di Tipo II.

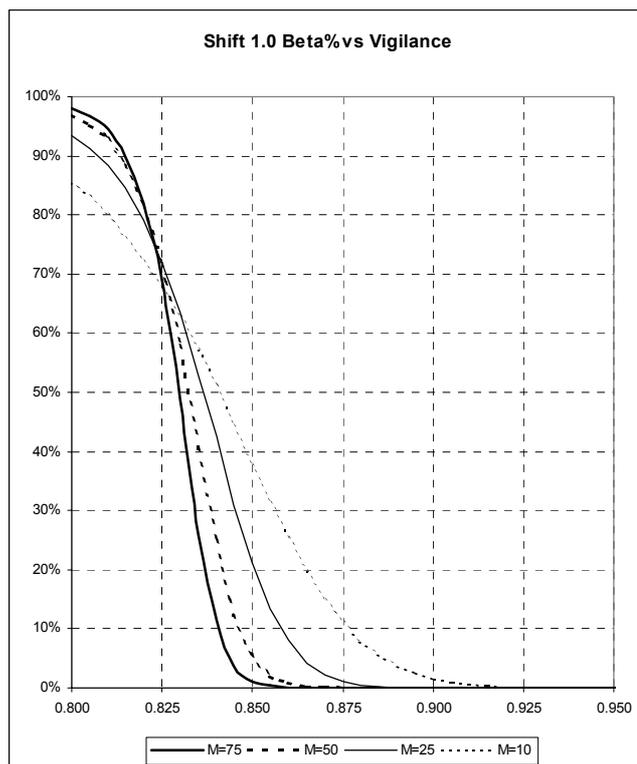
ρ	$M = 10$			$M = 25$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	85.18%	85.49%	85.80%	93.28%	93.53%	93.77%
0.825	67.35%	67.78%	68.21%	71.26%	71.94%	72.61%
0.850	37.16%	37.82%	38.48%	20.68%	21.36%	22.04%
0.875	10.84%	11.21%	11.58%	0.99%	1.18%	1.36%
0.900	1.37%	1.53%	1.68%	0.00%	0.02%	0.04%
0.925	0.06%	0.09%	0.12%	0.00%	0.00%	0.00%
0.950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

ρ	$M = 50$			$M = 75$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	96.70%	96.86%	97.02%	97.92%	98.06%	98.19%
0.825	70.86%	71.68%	72.49%	67.97%	69.04%	70.11%
0.850	4.96%	5.51%	6.05%	0.83%	1.10%	1.37%
0.875	0.00%	0.03%	0.06%	0.00%	0.00%	0.00%
0.900	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.925	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

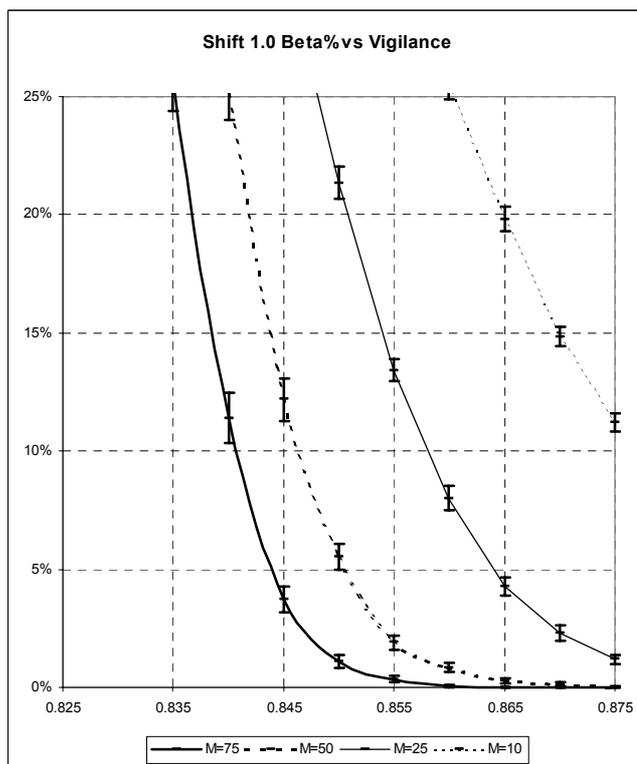
Tabella 6.2 – Stimatori puntuali e intervallari dell'errore di Tipo II generati dalla rete neurale (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard

L'addestramento della rete Fuzzy ART è limitato sulla sola media nominale del processo.

Dai risultati della Tabella 6.2 può essere facilmente dedotto che al crescere del parametro di vigilanza l'errore di Tipo II commesso dalla rete neurale diminuisce. In particolare, l'errore di Tipo II del sistema di controllo basato sulla rete neurale Fuzzy ART approssima il limite inferiore $\hat{\beta} = 0\%$ quando il parametro di vigilanza approssima il limite superiore $\rho = 1$ qualunque sia il valore dell'ampiezza scelta per la finestra di osservazione M . Tuttavia, il tasso di decrescita dell'errore di Tipo II se considerato una funzione del parametro di vigilanza ρ risulta più contenuto per bassi valori dell'ampiezza M rispetto a quello presentato per elevate ampiezze della finestra di osservazione M . Nei grafici riportati dalla Figura 6.2 è stato rappresentato l'andamento degli stimatori puntuali e intervallari $\hat{\beta}$ come funzione del parametro di vigilanza ρ per quattro livelli dell'ampiezza di finestra M .



a)



b)

Figura 6.2 – Stimatori puntuali degli errori di Tipo II (ordinate) per shift di una unità di deviazione standard, e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)

- a) intervallo di variazione ascissa è[0.800, 0.950], intervallo di variazione ordinate [0%, 100%]
- b) intervallo di variazione ascissa è[0.825, 0.875], intervallo di variazione ordinate [0%, 25%] con rappresentazione degli stimatori intervallari a copertura 95%

6.2 Comparazione delle prestazioni

In questa sezione, le prestazioni dell' algoritmo neurale saranno valutate per mezzo degli stimatori degli errori di Tipo II e saranno confrontate alle prestazioni di un metodo SPC. In particolare, poiché sono stati presi in considerazione esclusivamente cambiamenti della media di piccole e medie dimensioni (ovvero non superiori a due unità di deviazione standard), il metodo SPC appropriato per essere utilizzato come benchmark non potrà essere una tradizionale carta di controllo di Shewhart, bensì una carta di controllo CUSUM. Infatti, mentre la carta di controllo di Shewhart risulta poco sensibile per identificare piccoli cambiamenti della media, la carta di controllo a somma cumulata CUSUM rappresenta il metodo SPC più appropriato per costruire un benchmark di riferimento per il caso di studio proposto.

Come evidenziato nei capitoli precedenti, la versione tabellare della carta di controllo CUSUM si basa sulle seguenti relazioni ricorsive.

$$SH_t = \max\{0; SH_{t-1} + Y_t - k\} \text{ e } SL_t = \max\{0; SL_{t-1} - Y_t - k\} \quad (6.1)$$

Dove $\{Y_t\}$ è la serie temporale del parametro di processo monitorato, t è un indice temporale e k è un parametro arbitrario chiamato il valore di riferimento della carta CUSUM. Ad ogni istante di tempo t , le due statistiche SH_t e SL_t calcolate sui dati del processo e sono confrontate con un valore di soglia h . Un allarme sarà prodotto quando, ad un determinato istante di tempo t una, o entrambe le statistiche SH_t e SL_t , eccedono il limite superiore h . Solitamente, il valore di riferimento k è scelto pari a metà della più piccola variazione (in unità di deviazione standard) della media del processo che si ritiene importante identificare attraverso la carta CUSUM. Il corrispondente valore di h , è poi scelto in modo tale da mantenere l'errore di Tipo I circa uguale ad un predefinito valore (Montgomery 2000).

Nel caso di studio descritto in questa sezione il valore di riferimento della carta CUSUM è stato posto pari a $k = 0.5$ poiché si è supposto importante identificare variazioni della media di almeno un'unità di deviazione standard. Il corrispondente valore del limite h è stato scelto in modo da ottenere un errore medio di Tipo I uguale a circa 0.27%, ovvero uguale a quello presentato da una tradizionale carta di controllo di Shewarth quando i limiti di controllo sono posti pari a 3 volte la deviazione standard delle osservazioni. In particolare il valore del limite superiore per le statistiche della carta di controllo CUSUM è risultato pari a $h = 4.7749$.

Affinché la comparazione fra prestazioni della rete neurale con quelle della carta di controllo CUSUM sia bilanciata, tutti i risultati di seguito riportati sono stati ottenuti ignorando gli eventuali allarmi prodotti dalla carta CUSUM durante le prime $M - 1$ osservazioni. In altri termini le prestazioni dei due sistemi di controllo sono state misurate per tutti gli indici $t \geq M$. Inoltre, i parametri della rete Fuzzy ART sono stati scelti in modo tale da fornire un valore di tasso di errori di Tipo I comparabile con quello ottenibile dalla carta CUSUM di riferimento (ovvero circa 0.27%).

Poiché l' algoritmo Fuzzy ART presenta due gradi di libertà (il valore dell'ampiezza della finestra di osservazioni e il livello del parametro di vigilanza), è

possibile ottenere un determinato tasso di errori di Tipo I fissando arbitrariamente uno dei due parametri e determinando il corrispondente valore dell'altro. In questo lavoro è stato scelto di fissare l'ampiezza della finestra di osservazioni al livello $M = 75$ e di determinare conseguentemente il corrispondente livello del parametro di vigilanza in modo da ottenere il tasso di errori di Tipo I comparabile a quello fornito dalla carta di controllo CUSUM.

Successivamente, gli errori di Tipo II sono stati stimati sperimentalmente introducendo quattro segnali di disturbo (variazioni sistematiche, cicli di periodo $\Omega = 4$, shift all'istante $\tau = 1$ e mixture con probabilità di cambiamento dello stato $p_c = 0.4$). Ogni segnale di disturbo è stato inoltre valutato a cinque differenti livelli di ampiezza (0.25, 0.50, 1.00, 1.50 and 2.00). Gli stimatori puntuali ed intervallari per gli errori di Tipo II presentati sia dalla carta di controllo, sia dalla rete neurale per ognuna delle possibili combinazioni, sono poi stati misurati sperimentalmente attraverso l'approccio delle medie di batch (appendice A).

6.2.1 Confronto dei risultati

Nella seguente Tabella 6.3 sono stati riportati i confronti fra gli errori di Tipo I e di Tipo II per una carta di controllo CUSUM $k = 0.5$ e $h = 4.7749$ e per la rete neurale Fuzzy ART di parametri $M = 75$ e $\rho = 0.8475$. Accanto agli stimatori puntuali ottenuti dalla simulazione Monte Carlo per entrambi i sistemi di controllo, nella stessa Tabella 6.3 sono stati riportati gli intervalli di confidenza (con copertura del 95%) sulla differenza di prestazione ottenuta dalla rete neurale (per la quale è stato adottato il suffisso nn) e per la carta di controllo (per la quale è stato adottato il suffisso cc). Specificamente, la colonna contrassegnata con l'etichetta " $\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$ " riporta la differenza tra gli stimatori puntuali degli errori di Tipo I della rete neurale ($\hat{\alpha}_{nn}$) e della carta di controllo ($\hat{\alpha}_{cc}$). La colonna contrassegnata come " $\hat{\beta}_{nn} - \hat{\beta}_{cc}$ ", presenta l'analogia differenza con riferimento agli stimatori degli errori di Tipo II per i diversi casi di studio presi in considerazione. Inoltre, nelle due colonne adiacenti sono stati riportati gli estremi inferiori e superiori degli intervalli di confidenza sulla differenza delle prestazioni (indicati $(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$, $(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$ e $(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$, $(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$ rispettivamente).

I risultati riportati nella Tabella 6.3 dimostrano che la rete neurale di parametri $M = 75$ e $\rho = 0.8475$ risulta comparabile alla carta di controllo CUSUM ($k = 0.5$ e $h = 4.7749$) in termini di errori di Tipo I, ciò deriva dal fatto che l'intervallo di confidenza tra i rispettivi stimatori puntuali include il valore nullo indicando che non esiste evidenza statistica per rigettare l'ipotesi di uguaglianza fra le relative misure di prestazione $\alpha_{nn} = \alpha_{cc}$. Risulta, inoltre, altrettanto evidente che l'algoritmo neurale presenta prestazioni migliori rispetto alla carta CUSUM nel riconoscimento di variazioni sistematiche, cicli e mixture. In tutti questi casi, infatti, l'intervallo di confidenza sulla differenza degli stimatori β_{nn} e β_{cc} comprende unicamente valori negativi. In altri termini, per ognuno dei pattern citati è possibile concludere statisticamente che $\beta_{nn} < \beta_{cc}$, ovvero che la rete neurale presenta dei tassi di errori di Tipo II ridotti rispetto a quelli ottenibili in condizioni analoghe da una carta di controllo CUSUM.

La rete neurale, invece, presenta delle prestazioni peggiori (ovvero tassi di errore di Tipo II più elevati) rispetto a quelle della carta di controllo con riferimento al pattern shift. In particolare le prestazioni della rete neurale sono peggiori di quelle della carta di controllo se si considerano shift della media di piccola entità (vale e dire 0.25, 0.50 e

1.00 unità di deviazione standard). Al contrario, per shift della media di maggiore entità (1.50 e 2.00 unità di deviazione standard), le prestazioni della rete neurale e quelle della carta di controllo CUSUM appaiono sostanzialmente equivalenti.

	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy ART</i> $M = 75$ $\rho = 0.8475$		<i>Fuzzy ART – CUSUM</i>		
<i>Natural</i>	$\hat{\alpha}_{cc}$	$\hat{\alpha}_{nn}$		$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$	$\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$
	0.269%	0.262%		-0.053%	-0.007%	0.039%
	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy ART</i> $M = 75$ $\rho = 0.8475$		<i>Fuzzy ART – CUSUM</i>		
<i>Systematic Variation</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.709%	99.485%	-0.284%	-0.224%	-0.164%
	0.50	99.668%	97.479%	-2.323%	-2.189%	-2.055%
	1.00	99.487%	1.825%	-98.058%	-97.662%	-97.266%
	1.50	99.158%	0.000%	-99.225%	-99.158%	-99.091%
	2.00	98.424%	0.000%	-98.508%	-98.424%	-98.340%
<i>Cyclic</i> $\Omega = 4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.721%	99.630%	-0.145%	-0.091%	-0.037%
	0.50	99.689%	99.130%	-0.630%	-0.559%	-0.488%
	1.00	99.523%	84.549%	-15.667%	-14.974%	-14.281%
	1.50	99.234%	0.859%	-98.607%	-98.375%	-98.143%
	2.00	98.603%	0.000%	-98.688%	-98.603%	-98.518%
<i>Shift</i> $\tau = 1$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.034%	99.491%	0.381%	0.457%	0.533%
	0.50	92.474%	97.486%	4.639%	5.012%	5.385%
	1.00	0.019%	2.092%	1.622%	2.073%	2.524%
	1.50	0.000%	0.000%	0.000%	0.000%	0.000%
	2.00	0.000%	0.000%	0.000%	0.000%	0.000%
<i>Mixture</i> $p_c = 0.4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.618%	99.482%	-0.196%	-0.136%	-0.076%
	0.50	99.136%	97.511%	-1.799%	-1.625%	-1.451%
	1.00	95.527%	2.318%	-93.699%	-93.209%	-92.719%
	1.50	82.168%	0.000%	-82.548%	-82.168%	-81.788%
	2.00	51.802%	0.000%	-52.413%	-51.802%	-51.191%

Tabella 6.3 – Confronti fra le prestazioni della rete neurale e della carta CUSUM (risultati di simulazioni)

La rete neurale Fuzzy ART è addestrata al riconoscimento della sola media naturale del processo.

I risultati della Tabella 6.3 evidenziano che la carta di controllo CUSUM non è in grado di segnalare condizioni di variazione non naturale per i pattern di variazione sistematica, ciclo e mixture, mentre essa risulta particolarmente efficiente nel segnalare

shift costanti della media di processo. In generale ciò implica che la carta CUSUM non può essere adottata come l'unico strumento di monitoraggio per un processo, e ciò perché essa non presenta le stesse capacità di identificazione fra differenti tipi di effetti sui dati in uscita del processo.

Al contrario, dai dati riportati nella precedente Tabella 6.3 appare evidente che la rete neurale Fuzzy ART è in grado di poter riconoscere in modo più bilanciato tutti i quattro tipi di pattern non naturali utilizzati nel caso di studio. In particolare, da un rapido confronto fra i dati della tabella, risulta che fissato un determinato livello di ampiezza, le prestazioni della rete siano equivalenti nel riconoscimento dei pattern di variazione sistematica, shift e mixture, mentre esse risultano leggermente peggiori nel riconoscimento del pattern di tipo ciclo.

I confronti discussi in precedenza possono essere ragionevolmente generalizzati come segue. Mentre la carta di controllo CUSUM è in grado di identificare con successo una variazione dello stato del processo solo se questo si manifesta in una costante variazione della media, la rete neurale Fuzzy ART è in grado di estendere la capacità di identificazione su un insieme più ampio di pattern non naturali comprendendo in particolare quei pattern che presentano rapide fluttuazioni della media del processo. La rete neurale presenta capacità discriminatorie migliori se confrontate a quelle di una carta CUSUM poiché più ampio è lo spettro di pattern non naturali che possono essere riconosciuti.

6.3 Miglioramento delle prestazioni della rete neurale

I risultati riportati nelle precedenti sezioni fanno riferimento al caso di una rete neurale Fuzzy ART addestrata al riconoscimento della media naturale del processo, in altre parole del target di riferimento. In questa sezione sono stati riportati gli analoghi risultati ottenuti da una rete Fuzzy ART addestrata sia al riconoscimento del target naturale di riferimento per il processo, sia ad uno specifico segnale di disturbo. L'ipotesi sulla quale poggia l'estensione dell'algoritmo neurale discusso in questa sezione è che oltre a conoscere il target di riferimento, si conosca anche un particolare disturbo sul processo che si desidera identificare con la massima efficienza. In particolare, si richiede che la rete neurale Fuzzy ART apprenda il target naturale, ma anche una particolare deviazione della media di processo che risulta importante identificare efficientemente. L'obiettivo è quello di migliorare le prestazioni della rete neurale nel riconoscimento di un particolare pattern non naturale di variazione attraverso un ulteriore approfondimento della fase di addestramento.

Prima di proseguire, è utile evidenziare alcune differenze fra l'approccio che è esposto in questa sezione, con quelli comunemente riportati in letteratura e basati sull'uso di reti neurali a addestramento supervisionato. In primo luogo si osservi che l'addestramento della rete neurale è realizzato in modo cumulativo, vale a dire che la precedente fase è completata con una nuova fase di apprendimento senza richiedere un nuovo addestramento complessivo che sarebbe invece necessario utilizzando una rete supervisionata. In secondo luogo, l'addestramento della rete neurale su un particolare andamento non naturale, non limita le capacità di discriminazione della rete esclusivamente sul particolare andamento adottato. Infatti, e diversamente da una rete supervisionata, la rete Fuzzy ART continua a conservare la capacità di segnalare andamenti non naturali che derivano dalla mancata corrispondenza con il prototipo naturale di riferimento, ma accanto a questa proprietà, l'estensione dell'insieme di addestramento facilita l'identificazione di una particolare classe di pattern non naturali.

In altri termini ci aspettiamo che la rete neurale continui a conservare le capacità di identificazione per un ampio spettro di andamenti non naturali, presentando tuttavia una rapida capacità di riconoscimento per quella particolare classe non naturale utilizzata in fase di addestramento.

Il pattern non naturale utilizzato in fase di addestramento in questo studio è uno shift costante della media di processo. Come osservato in precedenza, in questo caso l'insieme di addestramento sarà costituito da due vettori: il primo corrisponde alla media nominale del processo, mentre il secondo codifica lo shift del target. Impiegando il metodo di apprendimento PLP, il risultato dell'apprendimento consiste nella formazione di due distinte categorie: la prima rappresenta la media naturale del processo (indichiamo con \underline{w}_μ il vettore dei pesi top-down corrispondente), e la seconda codifica il particolare disturbo non naturale che intendiamo identificare con maggiore efficienza (indichiamo con \underline{w}_s il vettore dei pesi top-down corrispondente). Evidentemente in questo caso la rete neurale sarà costituita da due neuroni di tipo *committed* nello strato *F2*. In fase di testing, quando un pattern d'ingresso risulta appartenente al cluster naturale, allora l'algoritmo neurale produrrà il risultato $b_{nn,t} = 1$. Al contrario, quando il pattern d'ingresso appartiene allo specifico cluster non naturale, oppure quando esso non può essere associato a nessuna delle categorie apprese dalla rete neurale, allora l'algoritmo Fuzzy ART produrrà il risultato in uscita $b_{nn,t} = 0$ indicante che il vettore d'ingresso è da considerarsi non naturale.

È stato dimostrato nel capitolo precedente che, quando il training della rete neurale include sia il target naturale, sia un particolare tipo di disturbo, allora il *criterio di generalizzazione* implementato dall'algoritmo Fuzzy ART per discriminare andamenti naturali da andamenti non naturali è modellabile dalla seguente relazione (6.2). Sia \underline{I}_t l'ingresso della rete neurale ad un particolare istante di tempo t , esso sarà riconosciuto come un pattern di tipo naturale (vale a dire che l'uscita della rete è posta pari a $b_{nn,t} = 1$) se e solo se risulta verificata che:

$$dis(\underline{I}_t, \underline{w}_\mu) \leq \min\{M(1 - \rho), dis(\underline{I}_t, \underline{w}_s)\} \quad (6.2)$$

In caso contrario, il pattern d'ingresso è classificato come non naturale e l'uscita prodotta dalla rete neurale è fissata pari a $b_{nn,t} = 0$. In altri termini, un dato input M -dimensionale è riconosciuto come naturale se due condizioni sono contemporaneamente verificate. 1) La sua distanza dal prototipo naturale è inferiore alla distanza presentata dal prototipo non naturale, 2) la sua distanza dal prototipo naturale è inferiore o uguale al limite $M(1 - \rho)$.

6.3.1 Analisi delle prestazioni

In questa sezione, l'analisi di sensibilità delle prestazioni della rete neurale in termini di errori di Tipo I e di Tipo II sono riconsiderate alla luce del nuovo scenario di addestramento. In particolare le tabelle di seguito riportate (Tabella 6.4 e Tabella 6.5) mostrano alcuni risultati sperimentali della rete Fuzzy ART addestrata al riconoscimento sia della media naturale del processo, sia di uno shift costante e di ampiezza pari a 0.5 unità di deviazione standard.

Un rapido confronto dei risultati riportati dalla Tabella 6.4 con quelli già presentati in precedenza nella Tabella 6.1 ci consente subito di osservare che l'estensione dell'insieme di addestramento con un pattern di tipo non naturale comporta a parità di altri fattori un aumento generalizzato dell'errore di Tipo I. In particolare, la crescita

dell'errore di Tipo I risulta particolarmente evidente per piccoli valori della finestra M (per esempio $M = 10$) piuttosto che per valori più elevati (per esempio $M = 75$).

ρ	$M = 10$			$M = 25$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$
0.800	7.63%	7.82%	8.01%	1.90%	2.01%	2.13%
0.825	9.73%	9.95%	10.18%	2.23%	2.35%	2.47%
0.850	15.88%	16.15%	16.43%	4.77%	4.95%	5.12%
0.875	32.13%	32.59%	33.05%	19.81%	20.34%	20.87%
0.900	65.39%	66.08%	66.78%	79.89%	80.71%	81.53%
0.925	94.13%	94.49%	94.84%	99.64%	99.74%	99.83%
0.950	99.83%	99.86%	99.89%	100.00%	100.00%	100.00%
0.975	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

ρ	$M = 50$			$M = 75$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$
0.800	0.50%	0.55%	0.60%	0.19%	0.22%	0.26%
0.825	0.53%	0.58%	0.62%	0.19%	0.23%	0.26%
0.850	1.34%	1.43%	1.51%	0.55%	0.60%	0.66%
0.875	13.96%	14.58%	15.20%	11.95%	12.66%	13.36%
0.900	94.53%	95.07%	95.61%	98.83%	99.05%	99.26%
0.925	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
0.950	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
0.975	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Tabella 6.4 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale (risultati di simulazione)

L'addestramento della rete Fuzzy ART comprende sia la media nominale del processo, sia uno shift di ampiezza 0.5 unità di deviazione standard.

La seguente Tabella 6.5, invece, mostra gli errori di Tipo II presentati dalla nuova rete neurale per uno shift di ampiezza $\varphi = 1$ e punto di partenza $\tau = 1$ in funzione del parametro di vigilanza ρ e della finestra M .

ρ	$M = 10$			$M = 25$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	5.32%	5.62%	5.93%	0.22%	0.30%	0.39%
0.825	4.69%	4.99%	5.28%	0.19%	0.28%	0.36%
0.850	3.21%	3.44%	3.67%	0.15%	0.23%	0.30%
0.875	1.67%	1.83%	1.99%	0.02%	0.07%	0.11%
0.900	0.45%	0.53%	0.61%	0.00%	0.01%	0.02%
0.925	0.04%	0.07%	0.09%	0.00%	0.00%	0.00%
0.950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

ρ	$M = 50$			$M = 75$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%
0.825	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%
0.850	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%
0.875	0.00%	0.00%	0.01%	0.00%	0.00%	0.00%
0.900	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.925	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

0.950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Tabella 6.5 – Stimatori puntuali e intervallari dell’errore di Tipo II generati dalla rete neurale (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard

L’addestramento della rete Fuzzy ART comprende sia la media nominale del processo, sia uno shift di ampiezza 0.5 unità di deviazione standard.

Confrontando i risultati riportati nella Tabella 6.5 di sopra con quelli precedentemente riportati (Tabella 6.2) si può subito osservare che gli errori di Tipo II commessi dalla rete neurale nel riconoscimento di uno shift della media di ampiezza pari a 1 unità di deviazione standard sono ora drasticamente ridotti. Ciò conferma quanto è stato precedentemente preventivato ovvero che l’estensione dell’insieme di addestramento su uno specifico pattern non naturale contribuisce a migliorare notevolmente le prestazioni della rete neurale al riconoscimento dello specifico disturbo.

Con l’obiettivo di valutare le prestazioni della rete neurale sulla base di un benchmark di riferimento, nella seguente Tabella 6.6 sono stati riportati i confronti fra gli errori di Tipo I e di Tipo II fra la rete neurale e la carta CUSUM di parametri $k = 0.5$ e $h = 4.7749$. Analogamente al caso mostrato in precedenza i confronti sono stati ottenuti utilizzando le stime degli errori di Tipo II per quattro pattern non naturali di variazione a cinque differenti livelli di ampiezza.

Fissando un’ampiezza della finestra $M = 75$, il parametro di vigilanza della rete neurale è stato poi scelto in modo da ottenere un tasso di errori di Tipo I comparabile con quello ottenibile dalla carta CUSUM, ovvero circa 0.27%. Il valore corrispondente del parametro di vigilanza è risultato essere $\rho = 0.8375$. Si può subito osservare che a parità di ampiezza di finestra, il nuovo addestramento della rete neurale comporta un più ridotto livello di vigilanza al fine di garantire lo stesso tasso di falsi allarmi della carta CUSUM. In altri termini, poiché l’estensione dell’insieme di addestramento comporta un aumento del tasso di falsi allarmi, al fine di compensarne l’effetto e di ottenere un livello comparabile di errori di Tipo I a quello precedentemente ottenuto, risulta necessario ridurre il livello di vigilanza.

	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy</i> <i>ART</i> $M = 75$ $\rho = 0.8375$		<i>Fuzzy</i> <i>ART –</i> <i>CUSUM</i>		
<i>Natural</i>	$\hat{\alpha}_{cc}$	$\hat{\alpha}_{nn}$		$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$	$\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$
	0.269%	0.268%		-0.047%	-0.001%	0.045%
	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy</i> <i>ART</i> $M = 75$ $\rho = 0.8375$		<i>Fuzzy</i> <i>ART –</i> <i>CUSUM</i>		
<i>Systematic</i> <i>Variation</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.709%	99.675%		-0.087%	-0.034%	0.019%
0.50	99.668%	99.014%		-0.752%	-0.654%	-0.556%
1.00	99.487%	17.819%		-82.757%	-81.668%	-80.579%
1.50	99.158%	0.000%		-99.225%	-99.158%	-99.091%
2.00	98.424%	0.000%		-98.508%	-98.424%	-98.340%

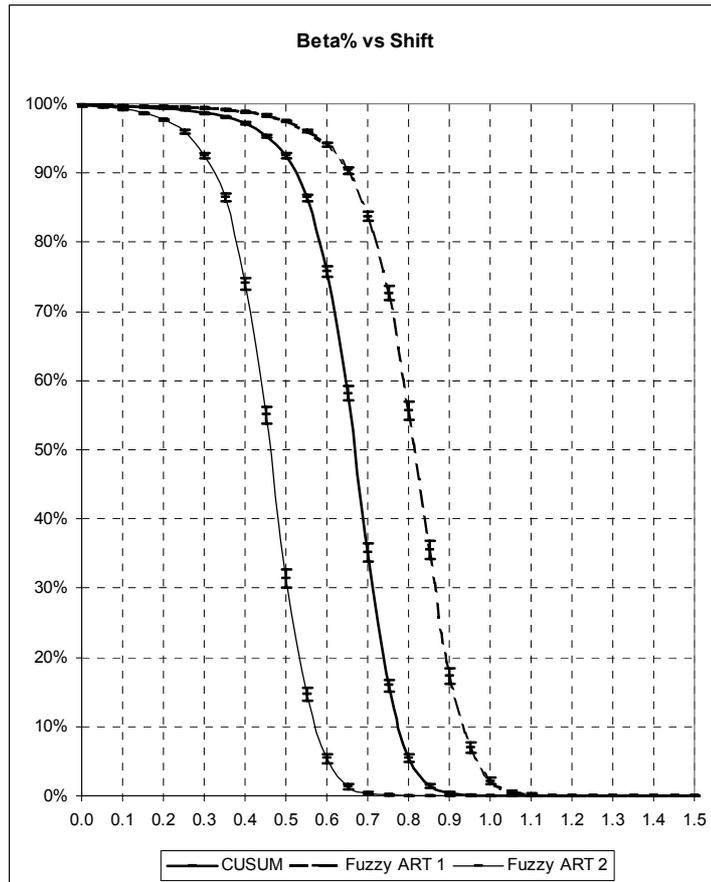
<i>Cyclic</i> $\Omega = 4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.721%	99.726%		-0.043%	0.005%	0.053%
0.50	99.689%	99.578%		-0.165%	-0.111%	-0.057%
1.00	99.523%	95.297%		-4.502%	-4.226%	-3.950%
1.50	99.234%	10.484%		-89.520%	-88.750%	-87.980%
2.00	98.603%	0.000%		-98.688%	-98.603%	-98.518%
<hr/>						
<i>Shift</i> $\tau = 1$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.034%	95.980%		-3.270%	-3.054%	-2.838%
0.50	92.474%	31.378%		-62.476%	-61.096%	-59.716%
1.00	0.019%	0.000%		-0.046%	-0.019%	0.008%
1.50	0.000%	0.000%		0.000%	0.000%	0.000%
2.00	0.000%	0.000%		0.000%	0.000%	0.000%
<hr/>						
<i>Mixture</i> $p_c = 0.4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.618%	99.662%		-0.008%	0.044%	0.096%
0.50	99.136%	98.829%		-0.407%	-0.307%	-0.207%
1.00	95.527%	16.880%		-79.688%	-78.647%	-77.606%
1.50	82.168%	0.000%		-82.548%	-82.168%	-81.788%
2.00	51.802%	0.000%		-52.413%	-51.802%	-51.191%

Tabella 6.6 – Confronti fra le prestazioni della rete neurale e della carta CUSUM (risultati di simulazioni)

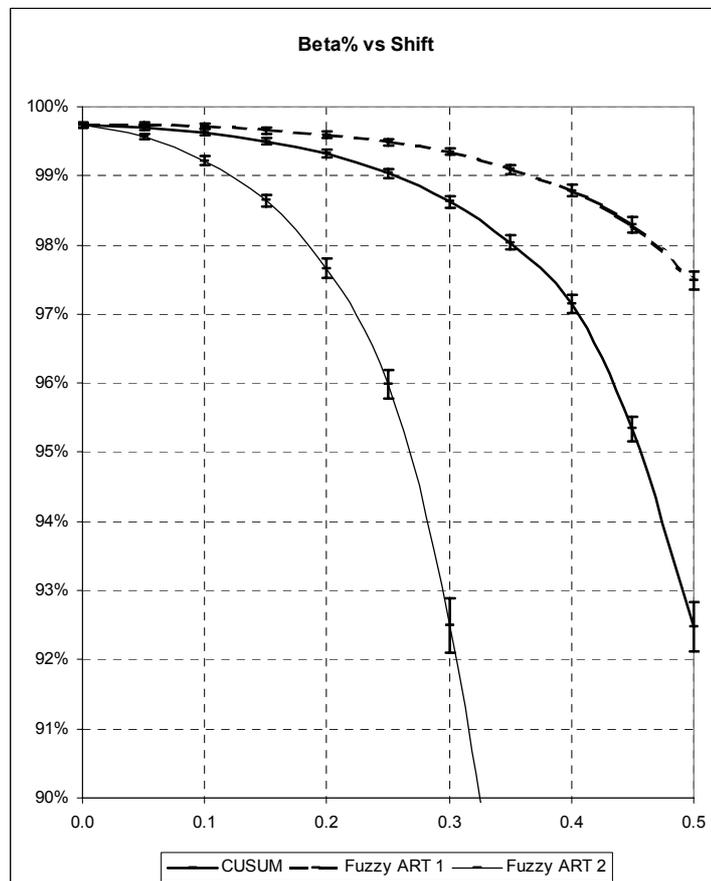
L’addestramento della rete Fuzzy ART comprende sia la media nominale del processo, sia uno shift di ampiezza 0.5 unità di deviazione standard.

Dai risultati riportati nella precedente Tabella 6.6 appare subito che la rete neurale fornisce risultati migliori rispetto alla carta CUSUM per il riconoscimento del pattern shift, e ciò è particolarmente vero per bassi livelli di ampiezza (ad esempio 0.25 o 0.50 unità di deviazione standard). Inoltre, anche se la ridotta vigilanza che è stato necessario adottare per la rete neurale ha avuto il conseguente effetto di incrementare gli errori di Tipo II per gli altri pattern, appare evidente che le prestazioni dell’algoritmo Fuzzy ART rimangono decisamente migliori, rispetto a quelle della carta CUSUM, in tutti i casi nei quali l’ampiezza del disturbo non è inferiore a 0.5 unità di deviazione standard. Viceversa, nel caso dei disturbi di ampiezza 0.25 unità di deviazione standard, le prestazioni della rete neurale e quelle della carta di controllo appaiano sostanzialmente equivalenti.

Il grafico della seguente Figura 6.3, riporta le curve beta per uno shift della media rispettivamente per i tre schemi di controllo precedentemente discussi. In particolare, la linea di spessore più ampio rappresenta le variazioni dell’errore di Tipo II per la carta CUSUM utilizzata come benchmark di riferimento (ovvero di parametri $k = 0.5$ e $h = 4.7749$). La linea a tratti rappresenta l’equivalente misura di prestazione per la rete neurale Fuzzy ART quando il training è limitato alla sola media naturale del processo (etichetta “Fuzzy ART 1”). Infine la linea sottile continua rappresenta l’andamento dell’errore di Tipo II per una rete Fuzzy ART quando l’addestramento comprende sia il target naturale del processo, sia uno shift costante della media e di ampiezza 0.5 unità di deviazione standard (etichetta “Fuzzy ART 2”).



a)



b)

Figura 6.3 – Stimatori puntuali degli errori di Tipo II (ordinate) per shift espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).

- a) intervallo di variazione ascissa è [0.0, 1.5], intervallo di variazione ordinate [0%, 100%]
- b) intervallo di variazione ascissa è [0.0, 0.5], intervallo di variazione ordinate [90%, 100%].

Carta CUSUM (linea continua in grassetto), Fuzzy ART con addestramento unico su media naturale del processo (linea tratteggiata), Fuzzy ART addestrata su media naturale e shift di ampiezza 0.5 unità di deviazione standard (linea continua sottile).

Il grafico presentato nella precedente Figura 6.3 mostra come la rete Fuzzy ART presenti già una discreta capacità di identificazione di shift della media pur avendo limitato la fase di addestramento unicamente al riconoscimento del target naturale del processo. Inoltre, le prestazioni della stessa rete neurale possono essere ulteriormente migliorate estendendo la fase di addestramento su uno specifico andamento non naturale (in questo caso proprio uno shift di ampiezza 0.5 unità di deviazione standard). In quest'ultimo caso, per esempio, le prestazioni presentate dalla rete neurale risultano notevolmente superiori a quelle del metodo CUSUM che, come ampiamente dimostrato in letteratura, è fra i metodi SPC quello che presenta le migliori prestazioni per il riconoscimento di piccoli shift della media di un processo.

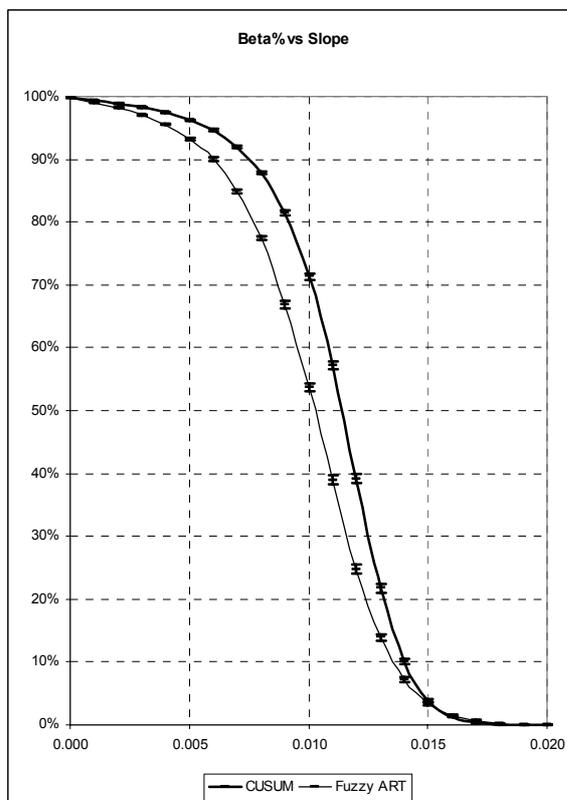
6.4 Il pattern di tipo trend

In questa sezione è esaminato il comportamento della rete nel riconoscimento di un ulteriore tipo di pattern: il trend (Pacella *et al.* 2002c). Questo tipo di pattern è rappresentato da una costante variazione della media del processo secondo un tasso (definito slope) positivo (ed in tal caso il trend è di crescita), o negativo (in questo caso il trend è di decrescita).

L'analisi per questo tipo di disturbo è stata separata da quella dei pattern esaminati nelle precedenti sezioni perché la caratterizzazione è eseguita per mezzo di un parametro di diverso ordine di grandezza. In particolare, un trend è caratterizzato da un parametro (definito *slope* e indicato con s) che esprime, in termini di unità di deviazione standard, il tasso di variazione della media del processo per ogni istante di osservazione successivo a quello di variazione dello stato del processo. In pratica, un tasso di crescita $0 < s \leq 1$ implica che la media del processo è variata di una unità di deviazione standard (in aumento o in diminuzione a seconda del segno di s) dopo ogni gruppo di circa $1/|s|$ osservazioni. I valori assunti dal parametro s risultano tipicamente di un ordine di grandezza inferiori rispetto a quelli assunti dal parametro di ampiezza φ che caratterizza, invece, i pattern esaminati nelle precedenti sezioni.

Il grafico riportato dalla seguente Figura 6.4 riporta le curve beta per uno trend di crescita della media del processo ottenute per mezzo della simulazione Monte Carlo su 50 batch di 2000 osservazioni. In particolare, la linea di spessore più ampio rappresenta le variazioni dell'errore di Tipo II nella identificazione del trend di associato al tasso indicato in ascissa per la carta CUSUM utilizzata come benchmark di riferimento (ovvero di parametri $k = 0.5$ e $h = 4.7749$). La linea sottile continua rappresenta l'andamento dell'errore di Tipo II per una rete Fuzzy ART considerando un addestramento che comprende sia il target naturale del processo, sia uno shift costante della media e di ampiezza 0.5 unità di deviazione standard. In particolare i risultati della rete neurale sono stati ottenuti impostando l'ampiezza di finestra $M = 75$, e scegliendo

il parametro di vigilanza in modo da ottenere un tasso di falsi allarmi statisticamente equivalente a quello che caratterizza il benchmark di riferimento (ovvero di circa 0.27%). Ciò ha comportato l'impostazione del parametro di vigilanza al livello $\rho = 0.8375$.



a)

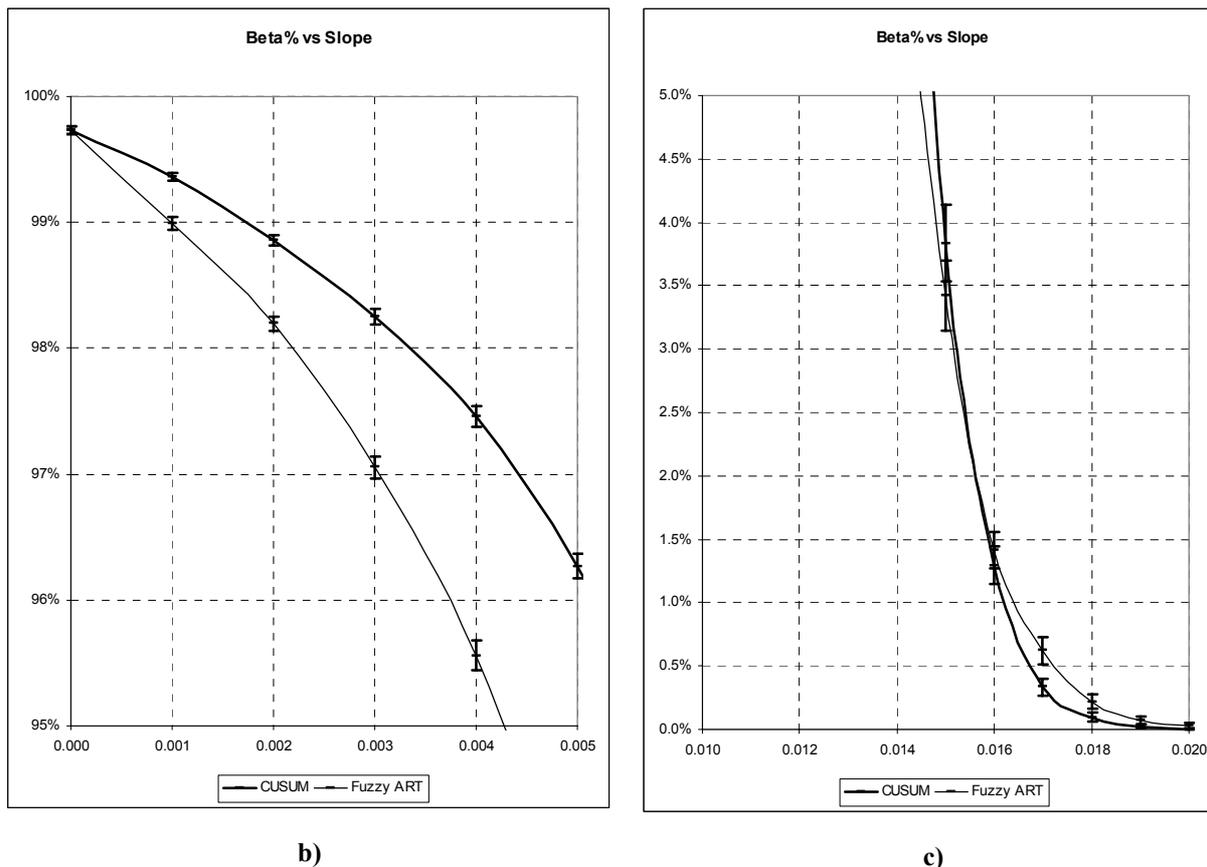


Figura 6.4 – Stimatori puntuali degli errori di Tipo II (ordinate) per trend con tassi (slope) espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).

- a) intervallo di variazione ascissa è [0.000, 0.020], intervallo di variazione ordinate [0%, 100%]
- b) intervallo di variazione ascissa è [0.000, 0.005], intervallo di variazione ordinate [90%, 100%].
- c) intervallo di variazione ascissa è [0.010, 0.020], intervallo di variazione ordinate [0%, 5%].

Carta CUSUM (linea continua in grassetto), Fuzzy ART addestrata su media naturale e shift di ampiezza 0.5 unità di deviazione standard (linea continua sottile).

Dai risultati schematizzati dalla Figura 6.4 si osserva che la rete neurale Fuzzy ART addestrata al riconoscimento di uno shift costante della media è in grado di generalizzare in modo soddisfacente nel riconoscimento di un ulteriore tipo di pattern (il trend), migliorando le prestazioni ottenibili attraverso un metodo SPC quale la carta CUSUM.

Il miglioramento delle prestazioni è particolarmente evidente per bassi livelli di trend di crescita (inferiori a 0.005 unità di deviazione standard), mentre al contrario, la carta CUSUM presenta prestazioni migliori per *slope* di più elevato valore (maggiori di 0.016 unità di deviazione standard). Per valori di *slope* superiori a 0.020 unità di deviazione standard, i due metodi di controllo posti a confronto, presentano prestazioni sostanzialmente equivalenti con errori di Tipo II prossimi al limite 0%.

Capitolo 7

Risultati sperimentali: dati dipendenti

In questo capitolo è rivista l'applicazione della procedura neurale al caso di un processo le cui uscite non sono modellabili come realizzazioni indipendenti. In particolare, attraverso l'utilizzo della simulazione Monte Carlo, saranno analizzate le misure di prestazione per il monitoraggio di un processo manifatturiero nel caso (tipicamente tra i più comuni) in cui le uscite possano essere modellare secondo uno schema Auto-Regressivo (AR) del primo ordine. Anche in questo caso, le prestazioni dell' algoritmo neurale saranno confrontate con quelle ottenibili dai metodi SPC per il monitoraggio di sequenze di dati auto-correlati.

7.1 Analisi caso AR -0.8

In questa sezione, saranno esaminate le prestazioni dell' algoritmo neurale Fuzzy ART nel caso di dati caratterizzati da una dipendenza temporale. In particolare, si farà riferimento a dati che presentano una correlazione modellabile secondo uno schema Auto-Regressivo (AR) del primo ordine e di fattore pari a -0.8 .

Nella Figura 7.1 seguente, è stato riportato l'istogramma sulla distribuzione di densità della misura *MAD* ottenuta su 25000 successive osservazioni raggruppate in finestre di 75 dati. In particolare, l'istogramma rappresentato con linea sottile si riferisce alla misura *MAD* ottenuta da dati normali e indipendentemente distribuiti, mentre l'istogramma in linea più spessa, fa riferimento a dati auto-correlati secondo uno schema AR del primo ordine e di fattore -0.8 .

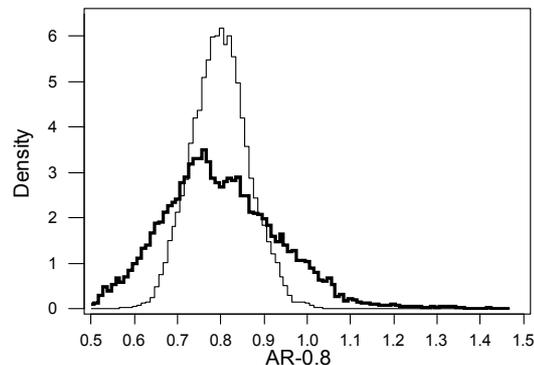


Figura 7.1 – Distribuzione densità statistica MAD per finestre di 75 osservazioni

Dati naturali per un processo AR di primo ordine a fattore -0.8 (linea spessa), dati naturali e indipendenti (linea sottile).

Poiché il funzionamento dell' algoritmo Fuzzy ART, ad addestramento di tipo PLP sul target naturale del processo, consiste in un decisore a soglia superiore (proporzionale a $(1-\rho)$), è immediato notare che fissata una ampiezza di finestra M e all' aumentare del parametro di vigilanza, aumenti necessariamente anche l' errore medio di Tipo I commesso dalla rete neurale essendo quest'ultimo dipendente dalla densità della statistica *MAD* inferiore del limite di soglia.

Dall'istogramma sopra riportato appare che la presenza di una forte correlazione negativa comporti un aumento della dispersione della misura MAD attorno ad un valore medio che, al contrario, sembra essere invariato. Ciò lascia presupporre che, considerando la variazione dell'errore di Tipo I come funzione del parametro di vigilanza, la presenza di un'auto-correlazione negativa comporti una diminuzione del tasso di crescita dei falsi allarmi al crescere di ρ (ovvero al diminuire del limite di soglia superiore con cui la grandezza MAD è confrontata).

Per valutare l'influenza dei parametri ρ e M , sui falsi allarmi prodotti dalla rete neurale alla presenza di una forte auto-correlazione negativa, sono state calcolate, attraverso le medesime procedure sperimentali, le misure di prestazione della rete che sono state descritte nel precedente capitolo con riferimento a dati indipendenti. In particolare, nella seguente tabella sono stati riportati gli stimatori puntuali ($\hat{\alpha}$) ed intervallari (con copertura del 95% ed indicati dalla notazione [$\hat{\alpha}_-, \hat{\alpha}_+$]) relativi all'errore di Tipo I ottenuti attraverso la simulazione Monte Carlo su un processo AR del primo ordine di fattore -0.8 .

ρ	$M=10$			$M=25$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$
0.800	1.94%	2.04%	2.14%	0.45%	0.50%	0.55%
0.825	3.24%	3.36%	3.48%	1.10%	1.18%	1.25%
0.850	5.68%	5.86%	6.04%	2.62%	2.75%	2.87%
0.875	10.64%	10.91%	11.18%	6.98%	7.22%	7.46%
0.900	22.99%	23.47%	23.95%	24.44%	25.12%	25.80%
0.925	54.23%	54.91%	55.59%	78.80%	79.53%	80.25%
0.950	93.35%	93.68%	94.00%	99.76%	99.81%	99.86%
0.975	99.96%	99.98%	99.99%	100.00%	100.00%	100.00%

ρ	$M=50$			$M=75$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$
0.800	0.10%	0.12%	0.15%	0.03%	0.04%	0.05%
0.825	0.34%	0.38%	0.43%	0.16%	0.19%	0.22%
0.850	1.27%	1.36%	1.45%	0.71%	0.79%	0.87%
0.875	5.13%	5.38%	5.62%	4.40%	4.66%	4.93%
0.900	31.73%	32.71%	33.69%	40.95%	42.06%	43.18%
0.925	95.70%	96.19%	96.68%	99.10%	99.33%	99.56%
0.950	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
0.975	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Tabella 7.1 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale su dati AR(-0.8) (risultati di simulazione)

L'addestramento della rete Fuzzy ART è limitato sulla sola media nominale del processo.

Dai dati riportati nella Tabella 7.1 precedente si può subito osservare che le considerazioni relative all'influenza dei fattori ρ ed M sull'errore di Tipo I, fatte nel caso di dati indipendenti, valgono anche nel caso in cui i dati del processo presentano una correlazione temporale negativa. Inoltre, confrontando i risultati della Tabella 7.1 con quelli omologhi ottenuti su dati indipendenti è facile dedurre che la presenza di una forte correlazione auto-regressiva negativa provochi una riduzione dei tassi di crescita della funzione errore di Tipo I dal parametro di vigilanza. Ciò, com'è stato anticipato, dipende principalmente dall'aumento della dispersione della misura MAD attorno al valore medio rispetto al caso presentato da dati indipendenti.

La seguente Figura 7.2 riporta il grafico degli stimatori puntuali $\hat{\alpha}$ dell'errore di Tipo I come funzione del parametro di vigilanza ρ per quattro differenti valori della ampiezza di finestra M .

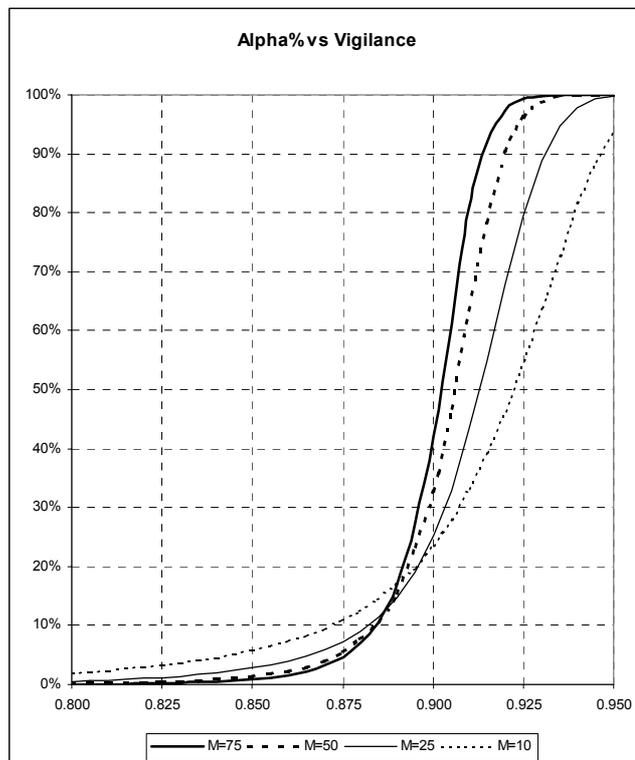


Figura 7.2 – Stimatori puntuali degli errori di Tipo I (ordinate) su dati AR(-0.8) e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)

Appare, inoltre, che l'influenza dell'ampiezza di finestra M sui tassi di errore di Tipo I risulti invariata nel caso di dati che presentano un'auto-correlazione. All'aumentare della finestra, infatti, cresce il tasso di variazione dell'errore di Tipo I in funzione del parametro di vigilanza. Questo risulta chiaramente dovuto all'effetto che un aumento della finestra M ha sulla dispersione della grandezza MAD . In particolare, essa comporta una diminuzione della dispersione della grandezza MAD attorno al valore medio e dunque ad una più rapida variazione del tasso di errore in funzione del parametro di vigilanza.

Per valutare il modo in cui la presenza di un'auto-correlazione fra i dati modifichi il funzionamento della rete neurale nel riconoscimento di variazioni della media del processo, nel seguente istogramma è stato riportata la distribuzione di densità relativa a dati che presentano uno shift costante di un'unità di deviazione standard, sia nel caso di indipendenza temporale, sia nel caso di presenza di auto-correlazione del primo ordine di fattore -0.8 . Come nel caso precedente, l'istogramma rappresentato con linea spessa si riferisce a dati dipendenti, mentre quello rappresentato con linea sottile fa riferimento a dati indipendenti.

Le considerazioni che possono essere derivate dall'istogramma sono opposte a quelle fatte in precedenza, infatti, la presenza di auto-correlazione sembra provocare una diminuzione della dispersione della statistica MAD attorno al valore medio. Ciò lascia presupporre che la presenza di un'auto-correlazione negativa dei dati implichi un incremento del tasso di diminuzione degli errori di Tipo II all'aumentare del parametro di vigilanza.

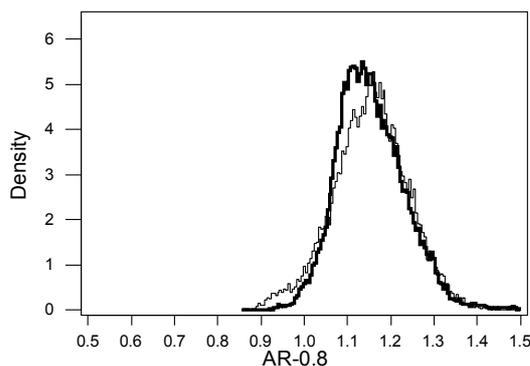


Figura 7.3 – Distribuzione densità statistica MAD per finestre di 75 osservazioni

Dati non naturali (shift 1 unità di deviazione standard) per un processo AR di primo ordine a fattore -0.8 (linea spessa), dati non naturali (shift 1 unità di deviazione standard) indipendenti (linea sottile).

Gli stimatori puntuali e intervallari dell'errore di Tipo II sono riportati nella seguente Tabella 7.2. Gli errori di Tipo II si riferiscono al caso di un processo che presenta uno shift della media pari ad un'unità di deviazione standard ($\varphi = 1$), inoltre, è stato supposto che l'istante di partenza dello shift coincidesse con il primo istante di osservazione sul processo ($\tau = 1$). La notazione $[\hat{\beta}_-, \hat{\beta}_+]$ è stata usata per indicare l'intervallo di confidenza al 95% sullo stimatore $\hat{\beta}$ dell'errore di Tipo II.

ρ	$M = 10$			$M = 25$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	90.78%	91.00%	91.22%	95.96%	96.08%	96.20%
0.825	65.11%	65.61%	66.11%	67.52%	68.19%	68.85%
0.850	13.22%	13.62%	14.03%	1.84%	2.08%	2.32%
0.875	0.30%	0.35%	0.41%	0.00%	0.00%	0.00%
0.900	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.925	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

ρ	$M = 50$			$M = 75$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	97.91%	98.01%	98.11%	98.69%	98.78%	98.87%
0.825	61.71%	62.85%	63.99%	53.39%	54.72%	56.05%
0.850	0.01%	0.04%	0.07%	0.00%	0.00%	0.00%
0.875	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.900	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.925	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Tabella 7.2 – Stimatori puntuali e intervallari dell'errore di Tipo II generati dalla rete neurale su dati AR(-0.8) (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard

L'addestramento della rete Fuzzy ART è limitato sulla sola media nominale del processo.

Come anticipato si può notare che in questo caso la presenza di una forte correlazione di tipo auto-regressivo negativo determina la diminuzione del tasso di allarmi di Tipo II a parità di fattori ρ e M (se confrontato con quello ottenuto da dati indipendenti). Nel grafico riportato dalla seguente Figura 7.4 è stato rappresentato

l'andamento degli stimatori puntuali $\hat{\beta}$ come funzione del parametro di vigilanza ρ per quattro livelli dell'ampiezza di finestra M .

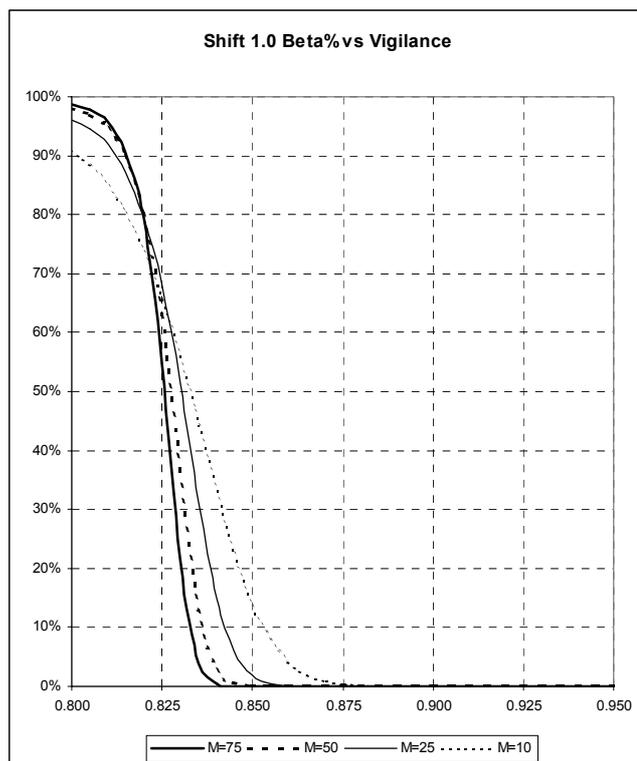


Figura 7.4 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(-0.8) per shift di una unità di deviazione standard, e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)

7.1.1 Comparazione delle prestazioni caso AR -0.8

Nella seguente Tabella 7.3 sono stati riportati i confronti fra gli errori di Tipo I e di Tipo II per una carta di controllo CUSUM $k = 0.5$ e $h = 4.7749$ e per la rete neurale Fuzzy ART di parametri $M = 75$ e $\rho = 0.8300$. Si evidenzia che, tutti i risultati riportati nel presente capitolo e inerenti l'uso della carta di controllo CUSUM fanno riferimento ai residui tra i dati simulati del processo e quelli di previsione ottenuti dal corrispondente modello AR. In altri termini la costruzione della carta di controllo CUSUM poggia sull'ipotesi implicita di conoscenza del modello di auto-correlazione del processo.

	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy ART</i> $M = 75$ $\rho = 0.8300$		<i>Fuzzy ART - CUSUM</i>		
<i>Natural</i>	$\hat{\alpha}_{cc}$ 0.269%	$\hat{\alpha}_{nn}$ 0.261%		$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$ -0.056%	$\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$ -0.008%	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$ 0.040%
	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy ART</i> $M = 75$ $\rho = 0.8300$		<i>Fuzzy ART - CUSUM</i>		

<i>Systematic Variation</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.730%	99.643%		-0.140%	-0.087%	-0.034%
0.50	99.714%	99.187%		-0.601%	-0.527%	-0.453%
1.00	99.675%	92.503%		-7.572%	-7.172%	-6.772%
1.50	99.569%	30.416%		-70.609%	-69.153%	-67.697%
2.00	99.416%	0.371%		-99.219%	-99.045%	-98.871%
<hr/>						
<i>Cyclic $\Omega = 4$</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.681%	99.711%		-0.020%	0.030%	0.080%
0.50	99.463%	99.646%		0.121%	0.183%	0.245%
1.00	97.299%	98.506%		1.068%	1.207%	1.346%
1.50	84.196%	34.402%		-51.156%	-49.794%	-48.432%
2.00	53.524%	0.000%		-53.955%	-53.524%	-53.093%
<hr/>						
<i>Shift $\tau = 1$</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	15.897%	99.696%		82.918%	83.799%	84.680%
0.50	0.000%	99.466%		99.415%	99.466%	99.517%
1.00	0.000%	21.412%		20.146%	21.412%	22.678%
1.50	0.000%	0.000%		0.000%	0.000%	0.000%
2.00	0.000%	0.000%		0.000%	0.000%	0.000%
<hr/>						
<i>Mixture $p_c = 0.4$</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	98.248%	99.696%		1.341%	1.448%	1.555%
0.50	87.414%	99.440%		11.722%	12.026%	12.330%
1.00	16.628%	58.796%		40.995%	42.168%	43.341%
1.50	0.555%	0.000%		-0.618%	-0.555%	-0.492%
2.00	0.023%	0.000%		-0.037%	-0.023%	-0.009%

Tabella 7.3 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(-0.8) (risultati di simulazioni)

La rete neurale Fuzzy ART è addestrata al riconoscimento della sola media naturale del processo.

In questo caso si può notare che, la rete neurale presenta prestazioni migliori rispetto a quelle della carta di controllo CUSUM sui residui solo nel caso del pattern di variazione sistematica. Negli altri casi, le prestazioni risultano peggiori o equivalenti per basse ampiezze del disturbo, oppure migliori come nel caso del ciclo di ampiezze 1.50 e 2.00 unità di deviazione standard. Complessivamente, si può osservare che la presenza di un'auto-correlazione negativa comporti un degrado delle prestazioni della rete neurale confrontate ad una carta di controllo CUSUM. Ciò è principalmente dovuto ad una diminuzione del parametro di vigilanza, che è stato necessario adottare per garantire il mantenimento di un tasso di errori di Tipo I di circa 0.27%, questo a sua volta ha comportato un aumento del limite di soglia superiore per la misura *MAD* affinché si segnali una condizione di allarme con un conseguente generalizzato aumento degli errori di Tipo II. D'altro canto, anche le prestazioni della carta CUSUM appaiono migliorate per effetto della maggiore influenza che un disturbo ha sui residui provenienti da uno schema auto-regressivo di fattore negativo (vedi capitolo 3). Tuttavia, è necessario sottolineare che le migliori prestazioni della carta CUSUM sono dipendenti dall'ipotesi di conoscere il modello di auto-regressione del processo, mentre l'approccio neurale, ne risulta del tutto indipendente (approccio *Model Free*).

7.1.2 Miglioramento delle prestazioni della rete neurale

In questo paragrafo sono riesaminate le prestazioni della rete neurale, nel monitoraggio di dati che presentano un'auto-regressione del primo ordine di fattore – 0.8, supponendo di estendere la fase di addestramento al riconoscimento sia del target naturale del processo, sia di uno shift di 0.5 unità di deviazione standard.

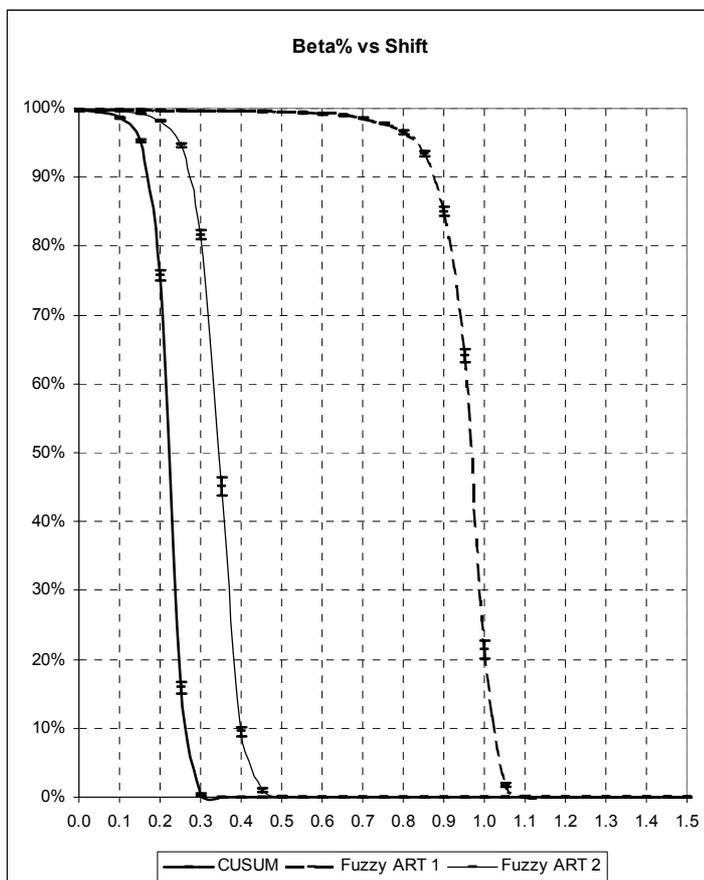
	<i>CUSUM</i> $k=0.5$ $h=4.7749$	<i>Fuzzy ART</i> $M=75$ $\rho=0.8300$		<i>Fuzzy ART – CUSUM</i>		
<i>Natural</i>	$\hat{\alpha}_{cc}$	$\hat{\alpha}_{nn}$		$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$	$\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$
	0.269%	0.263%		-0.054%	-0.006%	0.042%
	<i>CUSUM</i> $k=0.5$ $h=4.7749$	<i>Fuzzy ART</i> $M=75$ $\rho=0.8300$		<i>Fuzzy ART – CUSUM</i>		
<i>Systematic Variation</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.730%	99.634%	-0.147%	-0.096%	-0.045%
	0.50	99.714%	99.186%	-0.604%	-0.528%	-0.452%
	1.00	99.675%	92.481%	-7.630%	-7.194%	-6.758%
	1.50	99.569%	30.416%	-70.609%	-69.153%	-67.697%
	2.00	99.416%	0.371%	-99.219%	-99.045%	-98.871%
<i>Cyclic</i> $\Omega = 4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.681%	99.705%	-0.025%	0.024%	0.073%
	0.50	99.463%	99.627%	0.106%	0.164%	0.222%
	1.00	97.299%	98.388%	0.944%	1.089%	1.234%
	1.50	84.196%	33.906%	-51.568%	-50.290%	-49.012%
	2.00	53.524%	0.000%	-53.955%	-53.524%	-53.093%
<i>Shift</i> $\tau = 1$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	15.897%	94.689%	77.864%	78.792%	79.720%
	0.50	0.000%	0.080%	0.020%	0.080%	0.140%
	1.00	0.000%	0.000%	0.000%	0.000%	0.000%
	1.50	0.000%	0.000%	0.000%	0.000%	0.000%
	2.00	0.000%	0.000%	0.000%	0.000%	0.000%
<i>Mixture</i> $p_c = 0.4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	98.248%	99.680%	1.326%	1.432%	1.538%
	0.50	87.414%	99.299%	11.579%	11.885%	12.191%
	1.00	16.628%	53.404%	35.371%	36.776%	38.181%
	1.50	0.555%	0.000%	-0.618%	-0.555%	-0.492%
	2.00	0.023%	0.000%	-0.037%	-0.023%	-0.009%

Tabella 7.4 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(-0.8) (risultati di simulazioni)

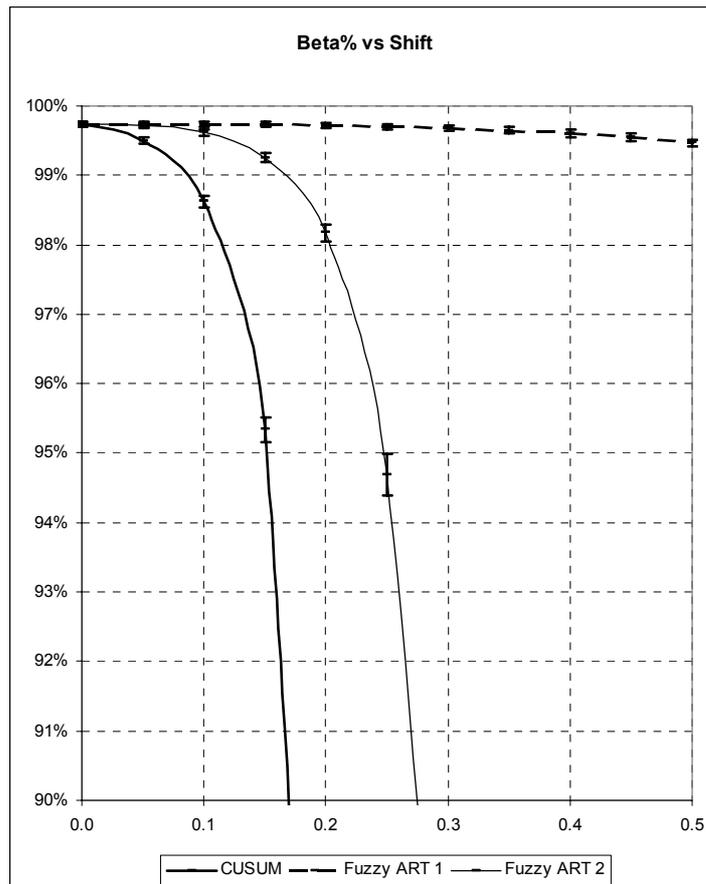
L'addestramento della rete Fuzzy ART comprende sia la media nominale del processo, sia uno shift di ampiezza 0.5 unità di deviazione standard.

Si può notare che, poiché il parametro di vigilanza è rimasto invariato per garantire il medesimo tasso di falsi allarmi, le prestazioni della rete sono rimaste sostanzialmente invariate per i pattern di variazione sistematica, ciclo e mixture. Invece, le prestazioni nel riconoscimento di shift sono notevolmente migliorate, e possono essere ritenute equivalenti a quelle presentate da una carta di controllo CUSUM costruita sui residui già per variazioni della media dell'ordine di 0.5 unità di deviazione standard.

I grafici riportati dalla seguente Figura 7.5 riportano le curve beta per uno shift della media rispettivamente per i tre schemi di controllo precedentemente discussi. In particolare, la linea di spessore più ampio rappresenta le variazioni dell'errore di Tipo II per la carta CUSUM utilizzata come benchmark di riferimento (ovvero di parametri $k = 0.5$ e $h = 4.7749$). La linea a tratti rappresenta l'equivalente misura di prestazione per la rete neurale Fuzzy ART quando il training è limitato alla sola media naturale del processo (etichetta "Fuzzy ART 1"). Infine la linea sottile continua rappresenta l'andamento dell'errore di Tipo II per una rete Fuzzy ART quando l'addestramento comprende sia il target naturale del processo, sia uno shift costante della media e di ampiezza 0.5 unità di deviazione standard (etichetta "Fuzzy ART 2").



a)



b)

Figura 7.5 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(-0.8) per shift espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).

- c) intervallo di variazione ascissa è [0.0, 1.5], intervallo di variazione ordinate [0%, 100%]
- d) intervallo di variazione ascissa è [0.0, 0.5], intervallo di variazione ordinate [90%, 100%].

Carta CUSUM (linea continua in grassetto), Fuzzy ART con addestramento unico su media naturale del processo (linea tratteggiata), Fuzzy ART addestrata su media naturale e shift di ampiezza 0.5 unità di deviazione standard (linea continua sottile).

7.2 Analisi caso AR -0.2

L'analisi sulle prestazioni della rete neurale (sia in termini di errori di Tipo I sia in termini di errore di Tipo II) eseguita nel precedente paragrafo viene di seguito riconsiderata per un processo che presenta una dipendenza temporale delle osservazioni di tipo AR del primo ordine di fattore -0.2.

La Figura 7.6 seguente riporta l'istogramma delle osservazioni MAD ottenute su 25000 dati raggruppati in finestre di 75 punti.

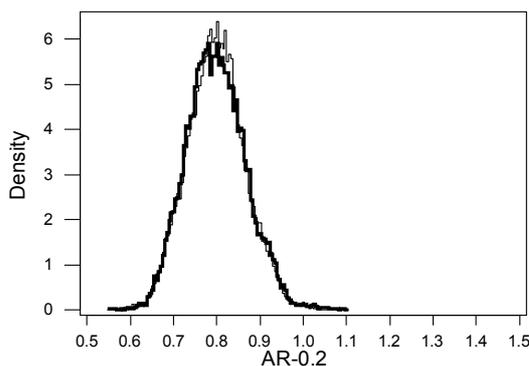


Figura 7.6 – Distribuzione densità statistica MAD per finestre di 75 osservazioni

Dati naturali per un processo AR di primo ordine a fattore -0.2 (linea spessa), dati naturali e indipendenti (linea sottile).

Si può intuire che, la presenza di un'auto-correlazione AR di fattore ridotto non comporti notevoli variazioni alla distribuzione di densità della grandezza MAD . La seguente Tabella 7.5 riporta gli stimatori puntuali ($\hat{\alpha}$) ed intervallari (con copertura del 95% ed indicati dalla notazione $[\hat{\alpha}_-, \hat{\alpha}_+]$) relativi all'errore di Tipo I presentato dalla rete neurale nel caso di un processo che presenti auto-correlazione AR del primo ordine con fattore -0.2 .

ρ	$M = 10$			$M = 25$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$
0.800	0.79%	0.85%	0.91%	0.02%	0.03%	0.04%
0.825	2.63%	2.72%	2.81%	0.36%	0.39%	0.43%
0.850	7.91%	8.16%	8.40%	2.53%	2.65%	2.77%
0.875	23.20%	23.60%	24.01%	15.95%	16.41%	16.86%
0.900	59.02%	59.70%	60.38%	76.23%	77.16%	78.08%
0.925	93.35%	93.66%	93.96%	99.62%	99.71%	99.79%
0.950	99.77%	99.82%	99.86%	100.00%	100.00%	100.00%
0.975	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

ρ	$M = 50$			$M = 75$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$
0.800	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.825	0.03%	0.04%	0.05%	0.00%	0.01%	0.01%
0.850	0.79%	0.84%	0.90%	0.36%	0.40%	0.44%
0.875	12.34%	12.90%	13.46%	11.02%	11.60%	12.18%
0.900	92.89%	93.46%	94.04%	98.42%	98.69%	98.96%
0.925	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
0.950	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
0.975	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Tabella 7.5 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale su dati AR(-0.2) (risultati di simulazione)

L'addestramento della rete Fuzzy ART è limitato sulla sola media nominale del processo.

Come era preventivabile osservando l'andamento delle distribuzioni di densità riportate nel precedente istogramma, le variazioni rispetto al caso di dati indipendenti risultano molto contenute.

La seguente Figura 7.7 riporta il grafico degli stimatori puntuali $\hat{\alpha}$ dell'errore di Tipo I come funzione del parametro di vigilanza ρ . In particolare le relazioni fra errore

di Tipo I e parametro di vigilanza sono rappresentate per quattro differenti valori della ampiezza di finestra M .

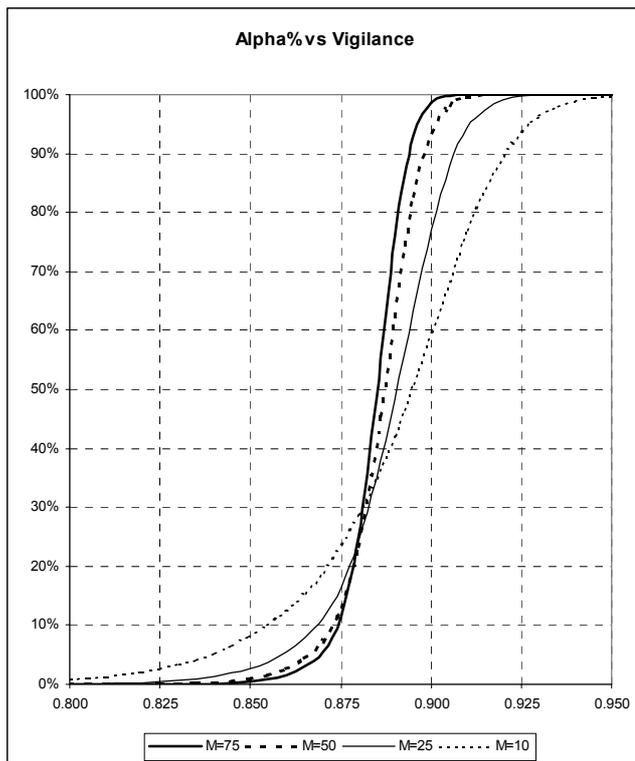


Figura 7.7 – Stimatori puntuali degli errori di Tipo I (ordinate) su dati AR(-0.2) e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)

Per valutare il modo in cui la presenza di un’auto-correlazione fra i dati modifichi il funzionamento della rete neurale nel riconoscimento di variazioni della media del processo, nel seguente istogramma è stato riportata la distribuzione di densità relativa a dati che presentano uno shift costante di un’unità di deviazione standard, sia nel caso di indipendenza temporale, sia nel caso di presenza di auto-correlazione del primo ordine di fattore -0.2 .

Anche in questo caso, come in quello descritto nel precedente paragrafo, appare che la presenza di auto-correlazione sembra provocare una diminuzione della dispersione della statistica MAD attorno al valore medio. Ciò lascia presupporre che la presenza di un’auto-correlazione negativa dei dati implichi un incremento del tasso di diminuzione degli errori di Tipo II all’aumentare del parametro di vigilanza.

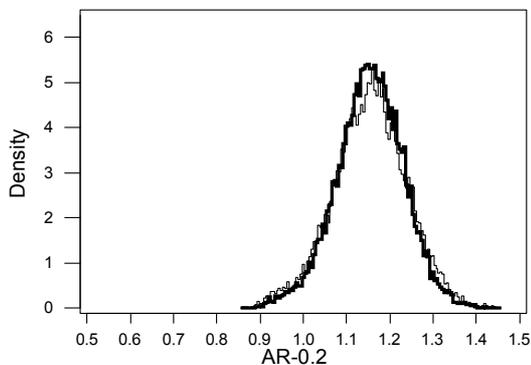


Figura 7.8– Distribuzione densità statistica MAD per finestre di 75 osservazioni

Dati non naturali (shift 1 unità di deviazione standard) per un processo AR di primo ordine a fattore -0.2 (linea spessa), dati non naturali (shift 1 unità di deviazione standard) indipendenti (linea sottile).

Gli stimatori puntuali e intervallari dell'errore di Tipo II sono riportati nella seguente tabella. La notazione $[\hat{\beta}_-, \hat{\beta}_+]$ è stata usata per indicare l'intervallo di confidenza al 95% sullo stimatore $\hat{\beta}$ dell'errore di Tipo II.

ρ	$M=10$			$M=25$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	84.64%	84.98%	85.33%	93.28%	93.51%	93.73%
0.825	63.57%	64.00%	64.43%	66.28%	66.98%	67.67%
0.850	29.17%	29.76%	30.36%	11.84%	12.56%	13.28%
0.875	6.10%	6.37%	6.64%	0.24%	0.33%	0.43%
0.900	0.60%	0.70%	0.81%	0.00%	0.00%	0.00%
0.925	0.00%	0.02%	0.03%	0.00%	0.00%	0.00%
0.950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

ρ	$M=50$			$M=75$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	96.82%	96.98%	97.14%	98.07%	98.19%	98.31%
0.825	63.80%	64.61%	65.43%	58.06%	59.34%	60.62%
0.850	1.60%	1.98%	2.36%	0.18%	0.34%	0.49%
0.875	0.00%	0.01%	0.01%	0.00%	0.00%	0.00%
0.900	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.925	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Tabella 7.6 – Stimatori puntuali e intervallari dell'errore di Tipo II generati dalla rete neurale su dati AR(-0.2) (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard

L'addestramento della rete Fuzzy ART è limitato sulla sola media nominale del processo.

Da un confronto fra i dati relativi all'errore di Tipo II ottenuti per un processo che presenta indipendenza fra le osservazioni, e quelli riportati nella precedente Tabella 7.6 si può osservare un aumento del tasso di decrescita dell'errore di Tipo II (come funzione del parametro di vigilanza) che risulta essere tanto più marcato quanto maggiore è l'ampiezza di finestra presa in considerazione.

Nel grafico riportato dalla seguente Figura 7.9 è stato rappresentato l'andamento degli stimatori puntuali $\hat{\beta}$ come funzione del parametro di vigilanza ρ per quattro livelli dell'ampiezza di finestra M .

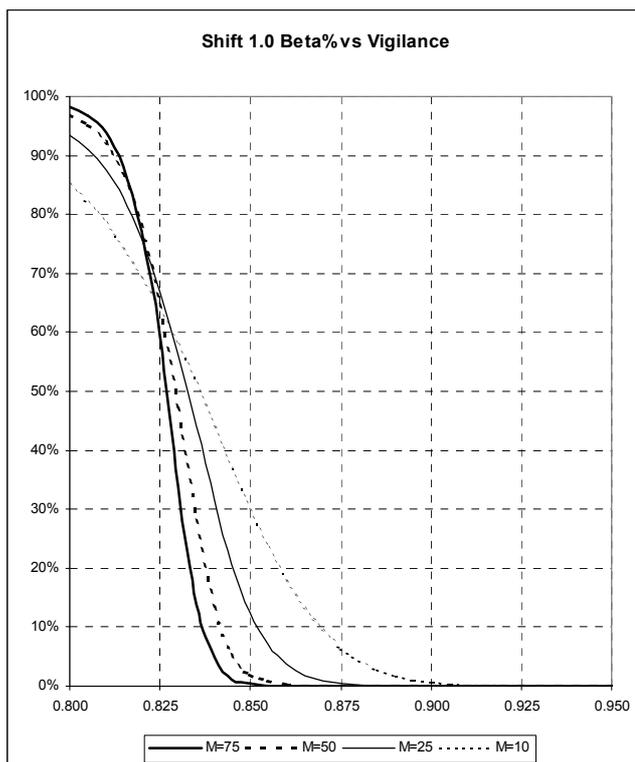


Figura 7.9 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(-0.2) per shift di una unità di deviazione standard, e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)

7.2.1 Comparazione delle prestazioni caso AR -0.2

Nella seguente tabella sono stati riportati i confronti fra gli errori di Tipo I e di Tipo II per una carta di controllo CUSUM $k = 0.5$ e $h = 4.7749$ e per la rete neurale Fuzzy ART di parametri $M = 75$ e $\rho = 0.8470$.

	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy ART</i> $M = 75$ $\rho = 0.8470$		<i>Fuzzy ART - CUSUM</i>		
<i>Natural</i>	$\hat{\alpha}_{cc}$	$\hat{\alpha}_{nn}$		$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$	$\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$
	0.269%	0.255%		-0.060%	-0.014%	0.032%
	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy ART</i> $M = 75$ $\rho = 0.8470$		<i>Fuzzy ART - CUSUM</i>		
<i>Systematic Variation</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.713%	99.464%	-0.307%	-0.249%	-0.191%
	0.50	99.695%	97.626%	-2.198%	-2.069%	-1.940%
	1.00	99.577%	8.413%	-91.943%	-91.164%	-90.385%
	1.50	99.358%	0.000%	-99.407%	-99.358%	-99.309%
	2.00	98.920%	0.000%	-98.994%	-98.920%	-98.846%

<i>Cyclic</i> $\Omega = 4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.719%	99.630%		-0.138%	-0.089%	-0.040%
0.50	99.695%	99.121%		-0.656%	-0.574%	-0.492%
1.00	99.494%	85.768%		-14.268%	-13.726%	-13.184%
1.50	99.115%	0.944%		-98.570%	-98.171%	-97.772%
2.00	98.283%	0.000%		-98.714%	-98.283%	-97.852%
<hr/>						
<i>Shift</i> $\tau = 1$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	98.555%	99.518%		0.082%	0.963%	1.844%
0.50	72.387%	97.644%		25.119%	25.257%	25.395%
1.00	0.000%	0.720%		0.447%	0.720%	0.993%
1.50	0.000%	0.000%		0.000%	0.000%	0.000%
2.00	0.000%	0.000%		0.000%	0.000%	0.000%
<hr/>						
<i>Mixture</i> $p_c = 0.4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.560%	99.509%		-0.159%	-0.051%	0.057%
0.50	98.839%	97.630%		-1.532%	-1.209%	-0.886%
1.00	92.809%	1.884%		-91.562%	-90.925%	-90.288%
1.50	70.165%	0.000%		-70.228%	-70.165%	-70.102%
2.00	31.051%	0.000%		-31.065%	-31.051%	-31.037%

Tabella 7.7 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(-0.2) (risultati di simulazioni)

La rete neurale Fuzzy ART è addestrata al riconoscimento della sola media naturale del processo.

Le prestazioni ottenute in questo caso sono simili a quelle riscontrate nel caso di dati indipendenti. Risulta, infatti, evidente che l’algoritmo neurale presenta prestazioni migliori rispetto alla carta CUSUM nel riconoscimento di variazioni sistematiche, cicli e mixture. La rete neurale, invece, presenta delle prestazioni peggiori (ovvero tassi di errore di Tipo II più elevati) rispetto a quelle della carta di controllo con riferimento al pattern shift. In particolare le prestazioni della rete neurale sono peggiori di quelle della carta di controllo se si considerano shift della media di piccola entità (vale e dire 0.25, 0.50 e 1.00 unità di deviazione standard). Al contrario, per shift della media di maggiore entità (1.50 e 2.00 unità di deviazione standard), le prestazioni della rete neurale e quelle della carta di controllo CUSUM appaiono sostanzialmente equivalenti.

7.2.2 Miglioramento delle prestazioni della rete neurale

In questo paragrafo sono riesaminate le prestazioni della rete neurale, nel monitoraggio di dati che presentano un’auto-regressione del primo ordine di fattore – 0.2, supponendo di estendere la fase di addestramento al riconoscimento sia del target naturale del processo, sia di uno shift di 0.5 unità di deviazione standard.

	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy</i> <i>ART</i> $M = 75$ $\rho = 0.8440$		<i>Fuzzy</i> <i>ART</i> – <i>CUSUM</i>		
<i>Natural</i>	$\hat{\alpha}_{cc}$	$\hat{\alpha}_{nn}$		$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$	$\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$
	0.269%	0.292%		-0.022%	0.023%	0.068%

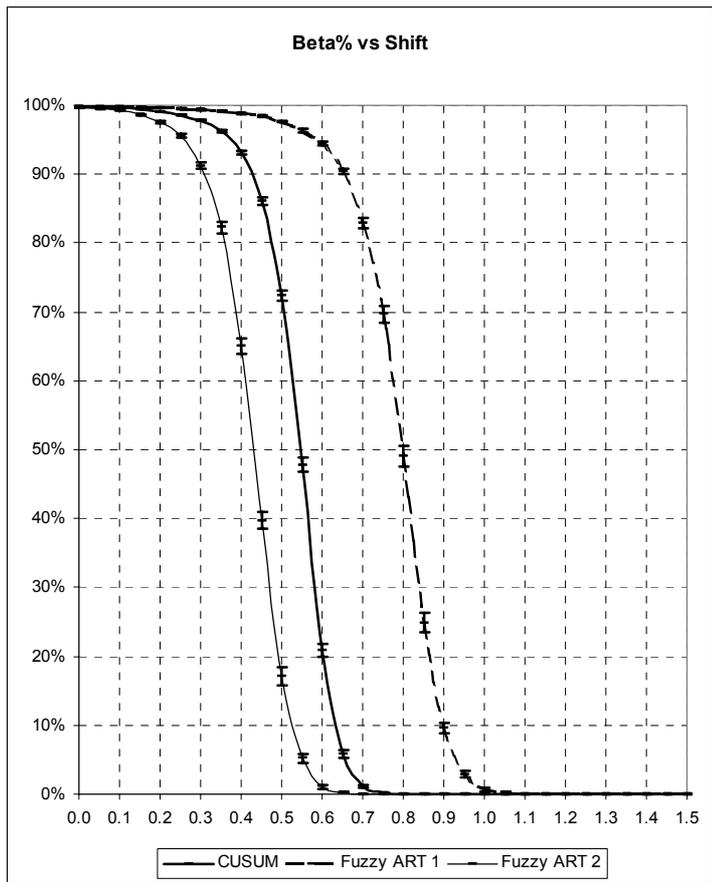
	<i>CUSUM</i> $k=0.5$ $h=4.7749$	<i>Fuzzy ART</i> $M=75$ $\rho=0.8440$		<i>Fuzzy ART – CUSUM</i>		
<i>Systematic Variation</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.713%	99.523%		-0.242%	-0.190%	-0.138%
0.50	99.695%	98.108%		-1.703%	-1.587%	-1.471%
1.00	99.577%	14.456%		-86.000%	-85.121%	-84.242%
1.50	99.358%	0.000%		-99.407%	-99.358%	-99.309%
2.00	98.920%	0.000%		-98.994%	-98.920%	-98.846%
<i>Cyclic</i> $\Omega = 4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.719%	99.638%		-0.125%	-0.081%	-0.037%
0.50	99.695%	99.276%		-0.488%	-0.419%	-0.350%
1.00	99.494%	90.151%		-9.842%	-9.343%	-8.844%
1.50	99.115%	2.318%		-97.304%	-96.797%	-96.290%
2.00	98.283%	0.000%		-98.714%	-98.283%	-97.852%
<i>Shift</i> $\tau = 1$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	98.555%	95.551%		-3.919%	-3.004%	-2.089%
0.50	72.387%	17.025%		-56.661%	-55.362%	-54.063%
1.00	0.000%	0.000%		0.000%	0.000%	0.000%
1.50	0.000%	0.000%		0.000%	0.000%	0.000%
2.00	0.000%	0.000%		0.000%	0.000%	0.000%
<i>Mixture</i> $p_c = 0.4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.560%	99.500%		-0.171%	-0.060%	0.051%
0.50	98.839%	97.994%		-1.167%	-0.845%	-0.523%
1.00	92.809%	3.749%		-89.807%	-89.060%	-88.313%
1.50	70.165%	0.000%		-70.228%	-70.165%	-70.102%
2.00	31.051%	0.000%		-31.065%	-31.051%	-31.037%

Tabella 7.8 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(-0.2) (risultati di simulazioni)

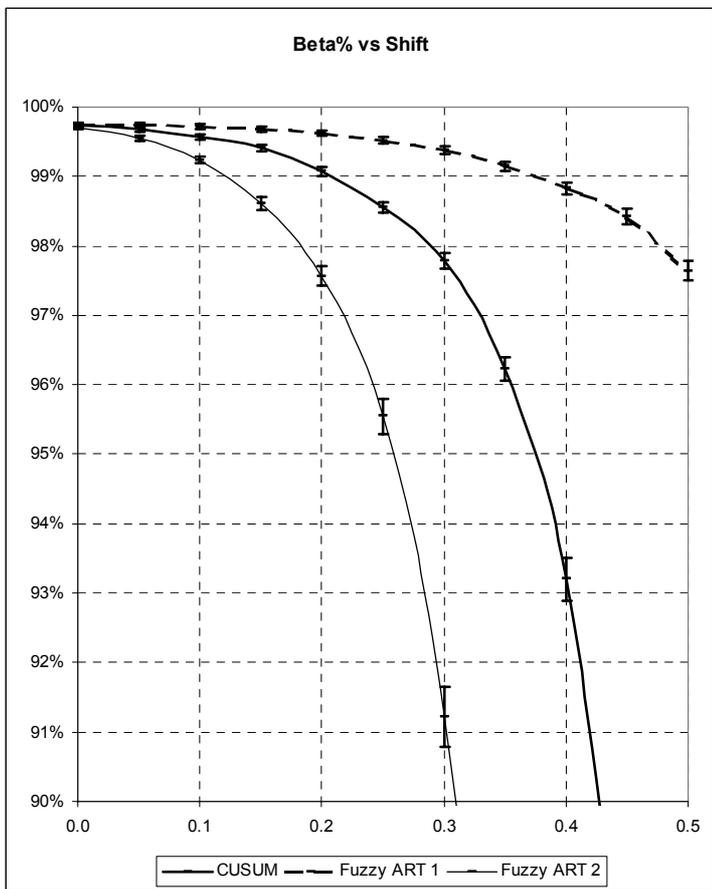
L’addestramento della rete Fuzzy ART comprende sia la media nominale del processo, sia uno shift di ampiezza 0.5 unità di deviazione standard.

Dai risultati riportati nella precedente Tabella 7.8 appare subito che la rete neurale fornisce risultati migliori rispetto alla carta CUSUM per il riconoscimento del pattern shift, e ciò è particolarmente vero per bassi livelli di ampiezza (ad esempio 0.25 o 0.50 unità di deviazione standard). In aggiunta, la rete neurale conserva prestazioni migliori rispetto a quelle presentate dalla carta CUSUM sui residui per tutti gli altri tipi di pattern e per qualunque livello di ampiezza sia considerata.

I grafici riportati dalla seguente Figura 7.10, schematizzano le curve beta per uno shift della media rispettivamente per i tre schemi di controllo precedentemente discussi.



a)



b)

Figura 7.10 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(-0.2) per shift espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).

- a) intervallo di variazione ascissa è [0.0, 1.5], intervallo di variazione ordinate [0%, 100%]
- b) intervallo di variazione ascissa è [0.0, 0.5], intervallo di variazione ordinate [90%, 100%].

Carta CUSUM (linea continua in grassetto), Fuzzy ART con addestramento unico su media naturale del processo (linea tratteggiata), Fuzzy ART addestrata su media naturale e shift di ampiezza 0.5 unità di deviazione standard (linea continua sottile).

7.3 Analisi caso AR 0.2

L'analisi sulle prestazioni della rete neurale eseguita nel precedente paragrafo viene di seguito riconsiderata per un processo che presenta una dipendenza temporale delle osservazioni di tipo AR del primo ordine di fattore 0.2. La Figura 7.11 seguente riporta l'istogramma delle osservazioni MAD ottenute su 25000 dati raggruppate in finestre di 75 punti.

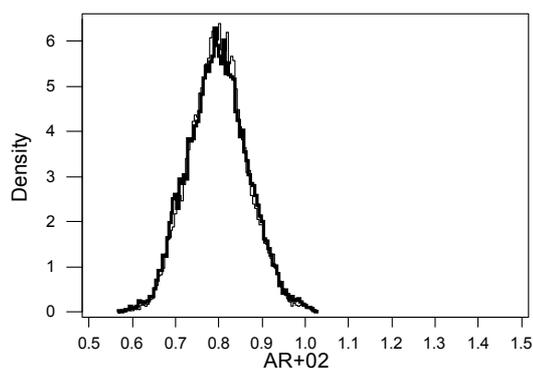


Figura 7.11 – Distribuzione densità statistica MAD per finestre di 75 osservazioni

Dati naturali per un processo AR di primo ordine a fattore 0.2 (linea spessa), dati naturali e indipendenti (linea sottile).

La seguente tabella riporta gli stimatori puntuali ($\hat{\alpha}$) ed intervallari relativi all'errore di Tipo I presentato dalla rete neurale nel caso di un processo che presenti auto-correlazione AR del primo ordine di fattore 0.2.

ρ	$M = 10$			$M = 25$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$
0.800	0.81%	0.87%	0.92%	0.02%	0.04%	0.05%
0.825	2.78%	2.87%	2.96%	0.36%	0.40%	0.44%
0.850	7.98%	8.22%	8.46%	2.59%	2.71%	2.84%
0.875	22.91%	23.39%	23.86%	16.02%	16.53%	17.05%
0.900	58.94%	59.55%	60.15%	76.19%	76.97%	77.76%
0.925	93.13%	93.43%	93.74%	99.65%	99.74%	99.83%
0.950	99.78%	99.83%	99.88%	100.00%	100.00%	100.00%
0.975	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

ρ	$M = 50$			$M = 75$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$

0.800	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.825	0.02%	0.04%	0.05%	0.00%	0.00%	0.01%
0.850	0.81%	0.87%	0.93%	0.38%	0.42%	0.46%
0.875	12.25%	12.79%	13.34%	10.95%	11.55%	12.14%
0.900	92.88%	93.51%	94.15%	98.55%	98.81%	99.06%
0.925	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
0.950	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
0.975	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Tabella 7.9 – Stimatori puntuali e intervallari dell'errore di Tipo I generati dalla rete neurale su dati AR(0.2) (risultati di simulazione)

L'addestramento della rete Fuzzy ART è limitato sulla sola media nominale del processo.

La seguente figura riporta i grafici degli stimatori puntuali $\hat{\alpha}$ dell'errore di Tipo I come funzione del parametro di vigilanza ρ .

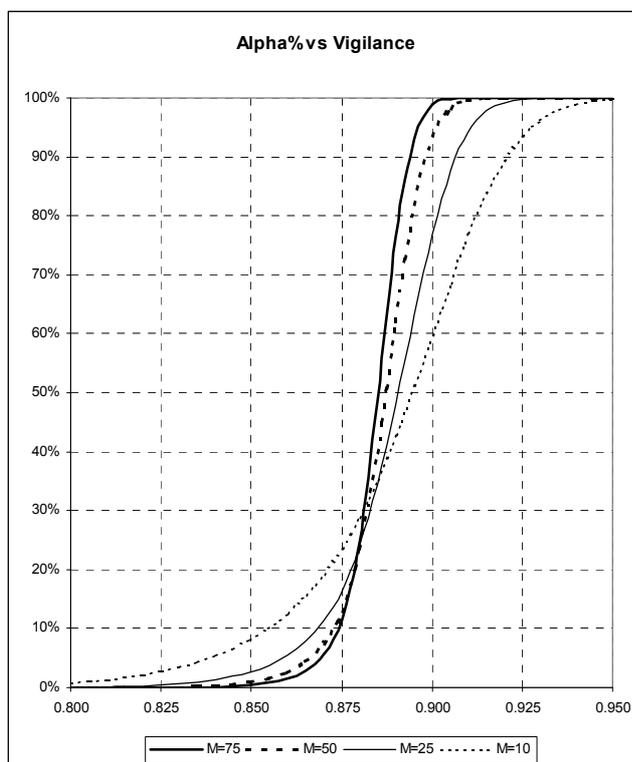


Figura 7.12 – Stimatori puntuali degli errori di Tipo I (ordinate) su dati AR(0.2) e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)

Analogamente a quanto fatto nei precedenti paragrafi, per valutare il modo in cui la presenza di un'auto-correlazione fra i dati modifichi il funzionamento della rete neurale nel riconoscimento di variazioni della media del processo, nel seguente istogramma è stato riportata la distribuzione di densità relativa a dati che presentano uno shift costante di un'unità di deviazione standard, sia nel caso di indipendenza temporale, sia nel caso di presenza di auto-correlazione del primo ordine di fattore 0.2.

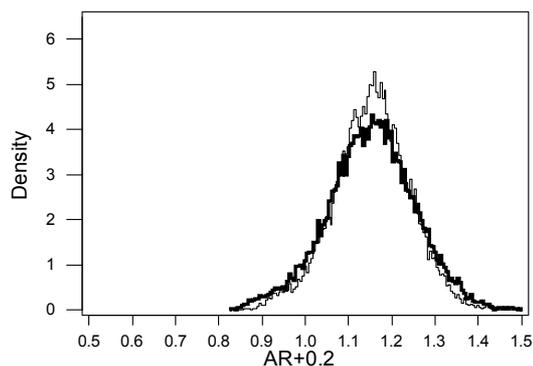


Figura 7.13– Distribuzione densità statistica MAD per finestre di 75 osservazioni

Dati non naturali (shift 1 unità di deviazione standard) per un processo AR di primo ordine a fattore 0.2 (linea spessa), dati non naturali (shift 1 unità di deviazione standard) indipendenti (linea sottile).

In questo caso, a differenza di quanto è stato osservato nei paragrafi precedenti, appare che la presenza di auto-correlazione di termine positivo tra i dati sembra provocare un aumento della dispersione della statistica *MAD* attorno al valore medio. Ciò lascia presupporre che la presenza di un'auto-correlazione positiva dei dati implichi una diminuzione del tasso di decrescita degli errori di Tipo II all'aumentare del parametro di vigilanza.

Gli stimatori puntuali e intervallari dell'errore di Tipo II sono riportati nella seguente tabella.

ρ	$M = 10$			$M = 25$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	86.07%	86.38%	86.69%	93.51%	93.76%	94.00%
0.825	71.86%	72.27%	72.69%	76.66%	77.31%	77.97%
0.850	46.86%	47.46%	48.06%	32.98%	33.88%	34.78%
0.875	18.32%	18.81%	19.30%	3.01%	3.32%	3.63%
0.900	3.53%	3.75%	3.97%	0.04%	0.10%	0.15%
0.925	0.29%	0.36%	0.43%	0.00%	0.00%	0.00%
0.950	0.00%	0.01%	0.02%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

ρ	$M = 50$			$M = 75$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	96.65%	96.81%	96.98%	97.84%	97.98%	98.12%
0.825	77.59%	78.31%	79.03%	76.00%	76.86%	77.73%
0.850	13.13%	14.16%	15.20%	4.43%	5.15%	5.87%
0.875	0.08%	0.14%	0.21%	0.00%	0.00%	0.00%
0.900	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.925	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Tabella 7.10 – Stimatori puntuali e intervallari dell'errore di Tipo II generati dalla rete neurale su dati AR(0.2) (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard

L'addestramento della rete Fuzzy ART è limitato sulla sola media nominale del processo.

Nel grafico riportato dalle seguente Figura 7.14 è stato rappresentato l'andamento degli stimatori puntuali $\hat{\beta}$ come funzione del parametro di vigilanza ρ per quattro livelli dell'ampiezza di finestra M .

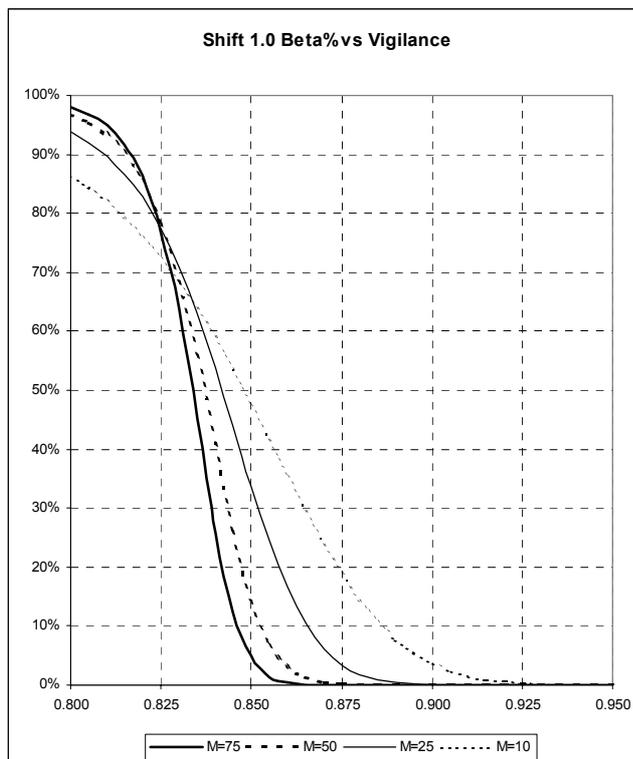


Figura 7.14 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(0.2) per shift di una unità di deviazione standard, e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)

7.3.1 Comparazione delle prestazioni caso AR 0.2

Nella seguente tabella sono stati riportati i confronti fra gli errori di Tipo I e di Tipo II per una carta di controllo CUSUM $k = 0.5$ e $h = 4.7749$ e per la rete neurale Fuzzy ART di parametri $M = 75$ e $\rho = 0.8475$.

	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy ART</i> $M = 75$ $\rho = 0.8470$		<i>Fuzzy ART – CUSUM</i>		
<i>Natural</i>	$\hat{\alpha}_{cc}$	$\hat{\alpha}_{nn}$		$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$	$\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$
	0.269%	0.264%		-0.053%	-0.005%	0.043%
	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy ART</i> $M = 75$ $\rho = 0.8470$		<i>Fuzzy ART – CUSUM</i>		
<i>Systematic Variation</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.709%	99.487%	-0.276%	-0.222%	-0.168%
	0.50	99.638%	97.655%	-2.119%	-1.983%	-1.847%

1.00	99.403%	0.591%		-99.006%	-98.812%	-98.618%
1.50	98.762%	0.000%		-98.811%	-98.762%	-98.713%
2.00	97.405%	0.000%		-97.479%	-97.405%	-97.331%
Cyclic						
$\Omega = 4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.721%	99.628%		-0.150%	-0.093%	-0.036%
0.50	99.695%	99.171%		-0.602%	-0.524%	-0.446%
1.00	99.500%	85.623%		-14.513%	-13.877%	-13.241%
1.50	99.155%	0.851%		-98.672%	-98.304%	-97.936%
2.00	98.392%	0.000%		-98.823%	-98.392%	-97.961%
Shift						
$\tau = 1$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.316%	99.476%		-0.721%	0.160%	1.041%
0.50	96.874%	97.648%		0.651%	0.774%	0.897%
1.00	3.616%	9.037%		4.449%	5.421%	6.393%
1.50	0.000%	0.000%		0.000%	0.000%	0.000%
2.00	0.000%	0.000%		0.000%	0.000%	0.000%
Mixture						
$p_c = 0.4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.659%	99.475%		-0.293%	-0.184%	-0.075%
0.50	99.311%	97.591%		-2.057%	-1.720%	-1.383%
1.00	96.967%	3.296%		-94.316%	-93.671%	-93.026%
1.50	88.545%	0.000%		-88.608%	-88.545%	-88.482%
2.00	65.668%	0.000%		-65.682%	-65.668%	-65.654%

Tabella 7.11 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(0.2) (risultati di simulazioni). La rete neurale Fuzzy ART è addestrata al riconoscimento della sola media naturale del processo.

I risultati riportati nella Tabella 7.11 dimostrano che la rete neurale risulta comparabile alla carta di controllo CUSUM in termini di errori di Tipo I. Risulta, inoltre, evidente che l’algoritmo neurale presenta prestazioni migliori rispetto alla carta CUSUM nel riconoscimento di variazioni sistematiche, cicli e mixture. In tutti questi casi, infatti, l’intervallo di confidenza sulla differenza degli stimatori β_{nn} e β_{cc} comprende unicamente valori negativi. In altri termini, per ognuno dei pattern citati è possibile concludere statisticamente che $\beta_{nn} < \beta_{cc}$, ovvero che la rete neurale presenta dei tassi di errori di Tipo II ridotti rispetto a quelli ottenibili in condizioni analoghe da una carta di controllo CUSUM.

La rete neurale, invece, presenta delle prestazioni leggermente peggiori (ovvero tassi di errore di Tipo II più elevati) rispetto a quelle della carta di controllo con riferimento al pattern shift.

7.3.2 Miglioramento delle prestazioni della rete neurale

In questo paragrafo sono riesaminate le prestazioni della rete neurale, nel monitoraggio di dati che presentano un’auto-regressione del primo ordine di fattore 0.2, supponendo di estendere la fase di addestramento al riconoscimento sia del target naturale del processo, sia di uno shift di 0.5 unità di deviazione standard.

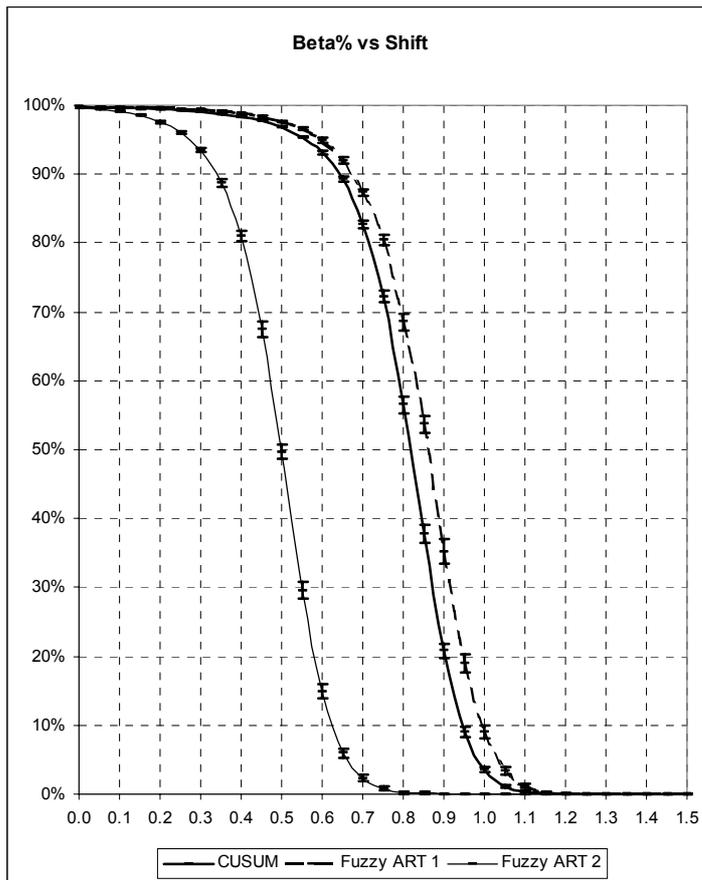
	<i>CUSUM</i> $k = 0.5$	<i>Fuzzy</i> <i>ART</i>		<i>Fuzzy</i> <i>ART</i> –
--	---------------------------	----------------------------	--	------------------------------

	$h = 4.7749$	$M = 75$ $\rho = 0.8000$	CUSUM			
Natural	$\hat{\alpha}_{cc}$	$\hat{\alpha}_{nn}$	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$	$\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$	
	0.269%	0.326%	0.006%	0.057%	0.108%	
	CUSUM $k = 0.5$ $h = 4.7749$	Fuzzy ART $M = 75$ $\rho = 0.8000$	Fuzzy ART – CUSUM			
Systematic Variation	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$	
	0.25	99.709%	99.674%	-0.084%	-0.035%	0.014%
	0.50	99.638%	99.663%	-0.027%	0.025%	0.077%
	1.00	99.403%	97.474%	-2.069%	-1.929%	-1.789%
	1.50	98.762%	0.000%	-98.811%	-98.762%	-98.713%
	2.00	97.405%	0.000%	-97.479%	-97.405%	-97.331%
Cyclic $\Omega = 4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$	
	0.25	99.721%	99.676%	-0.088%	-0.045%	-0.002%
	0.50	99.695%	99.678%	-0.072%	-0.017%	0.038%
	1.00	99.500%	99.606%	-0.012%	0.106%	0.224%
	1.50	99.155%	97.169%	-2.335%	-1.986%	-1.637%
	2.00	98.392%	17.526%	-82.351%	-80.866%	-79.381%
Shift $\tau = 1$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$	
	0.25	99.316%	96.088%	-4.134%	-3.228%	-2.322%
	0.50	96.874%	49.709%	-48.259%	-47.165%	-46.071%
	1.00	3.616%	0.000%	-3.616%	-3.616%	-3.616%
	1.50	0.000%	0.000%	0.000%	0.000%	0.000%
	2.00	0.000%	0.000%	0.000%	0.000%	0.000%
Mixture $p_c = 0.4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$	
	0.25	99.659%	99.649%	-0.114%	-0.010%	0.094%
	0.50	99.311%	99.562%	-0.049%	0.251%	0.551%
	1.00	96.967%	96.622%	-0.853%	-0.345%	0.163%
	1.50	88.545%	0.326%	-88.397%	-88.219%	-88.041%
	2.00	65.668%	0.000%	-65.682%	-65.668%	-65.654%

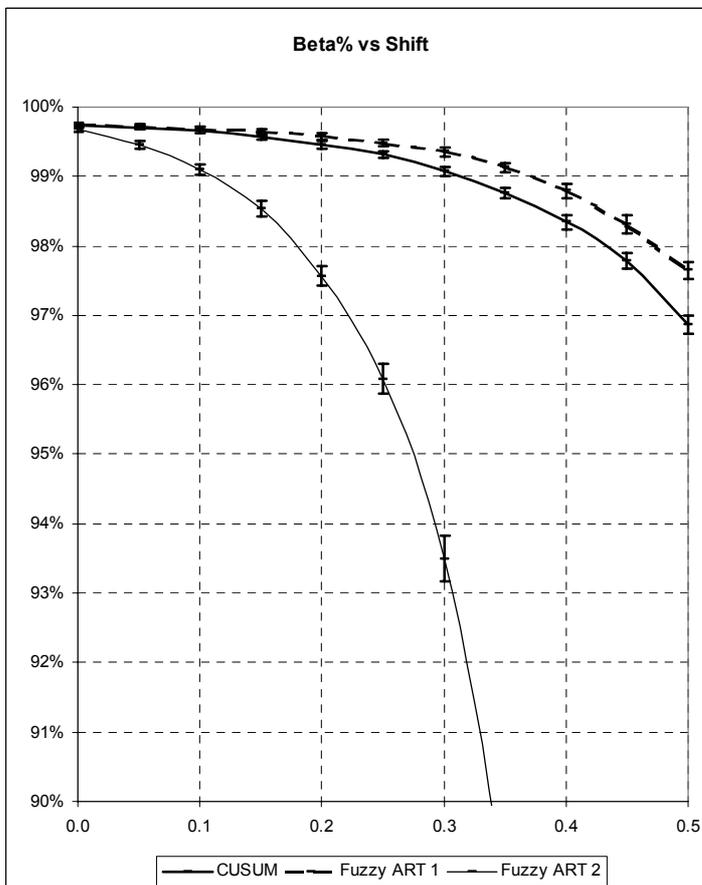
Tabella 7.12 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(0.2) (risultati di simulazioni)

L'addestramento della rete Fuzzy ART comprende sia la media nominale del processo, sia uno shift di ampiezza 0.5 unità di deviazione standard.

In questo caso è necessario notare che, per la dimensione della finestra prescelta, non è stato possibile individuare un livello del parametro di vigilanza che garantisca un tasso di falsi allarmi comparabile con quello ottenuto dalla carta di controllo CUSUM. In effetti, l'intervallo di confidenza sulla differenza degli errori di Tipo I comprende unicamente valori positivi variando da un massimo di 0.108% ad un minimo di 0.006%. I grafici riportati dalla seguente Figura 7.15 riportano le curve beta per uno shift della media rispettivamente per i tre schemi di controllo precedentemente discussi.



a)



b)

Figura 7.15 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(0.2) per shift espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).

- a) intervallo di variazione ascissa è [0.0, 1.5], intervallo di variazione ordinate [0%, 100%]
- b) intervallo di variazione ascissa è [0.0, 0.5], intervallo di variazione ordinate [90%, 100%].

Carta CUSUM (linea continua in grassetto), Fuzzy ART con addestramento unico su media naturale del processo (linea tratteggiata), Fuzzy ART addestrata su media naturale e shift di ampiezza 0.5 unità di deviazione standard (linea continua sottile).

7.4 Analisi caso AR 0.8

L'analisi sulle prestazioni della rete neurale eseguita nel precedente paragrafo viene di seguito riconsiderata per un processo che presenta una dipendenza temporale delle osservazioni di tipo AR del primo ordine di fattore 0.8. La Figura 7.16 seguente riporta l'istogramma delle osservazioni *MAD* ottenute su 25000 dati raggruppate in finestre di 75 punti.

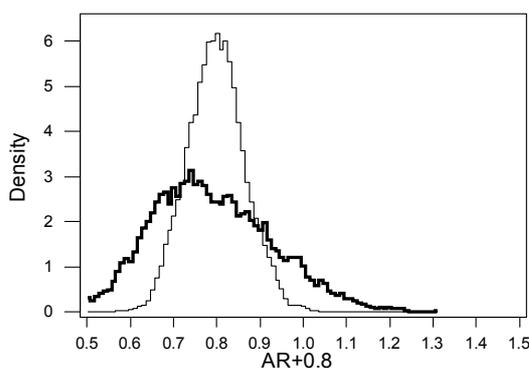


Figura 7.16 – Distribuzione densità statistica MAD per finestre di 75 osservazioni

Dati naturali per un processo AR di primo ordine a fattore 0.8 (linea spessa), dati naturali e indipendenti (linea sottile).

In questo caso, le considerazioni che possono essere fatte sono simili a quelle già dedotte per il caso di un processo con auto-correlazione AR di fattore negativo -0.8 . In particolare, dall'istogramma sopra riportato appare che la presenza di una forte correlazione positiva fra le osservazioni del processo comporti un aumento della dispersione della misura *MAD* attorno ad un valore medio che, al contrario, sembra essere invariato. Ciò lascia presupporre che, considerando la variazione dell'errore di Tipo I come funzione del parametro di vigilanza, la presenza di un'auto-correlazione negativa comporti una diminuzione del tasso di crescita dei falsi allarmi al crescere di ρ (ovvero al diminuire del limite di soglia superiore con cui la grandezza *MAD* è confrontata).

La seguente tabella riporta gli stimatori puntuali ($\hat{\alpha}$) ed intervallari relativi all'errore di Tipo I presentato dalla rete neurale nel caso di un processo che presenti auto-correlazione AR del primo ordine di fattore positivo 0.8.

ρ	$M=10$			$M=25$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$

0.800	1.96%	2.05%	2.14%	0.52%	0.57%	0.63%
0.825	3.32%	3.43%	3.54%	1.15%	1.23%	1.30%
0.850	5.69%	5.88%	6.07%	2.70%	2.82%	2.94%
0.875	10.64%	10.93%	11.22%	7.23%	7.47%	7.72%
0.900	22.78%	23.24%	23.71%	24.88%	25.52%	26.15%
0.925	54.10%	54.76%	55.42%	78.46%	79.16%	79.85%
0.950	93.73%	94.07%	94.40%	99.81%	99.86%	99.91%
0.975	99.98%	99.99%	100.00%	100.00%	100.00%	100.00%

ρ	$M = 50$			$M = 75$		
	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$	$\hat{\alpha}_-$	$\hat{\alpha}$	$\hat{\alpha}_+$
0.800	0.11%	0.14%	0.16%	0.03%	0.05%	0.06%
0.825	0.39%	0.45%	0.50%	0.16%	0.19%	0.22%
0.850	1.28%	1.37%	1.46%	0.80%	0.87%	0.95%
0.875	5.38%	5.61%	5.85%	4.45%	4.68%	4.91%
0.900	31.90%	32.89%	33.87%	41.28%	42.77%	44.26%
0.925	95.89%	96.30%	96.71%	99.23%	99.42%	99.62%
0.950	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%
0.975	100.00%	100.00%	100.00%	100.00%	100.00%	100.00%

Tabella 7.13 – Stimatori puntuali e intervallari dell’errore di Tipo I generati dalla rete neurale su dati AR(0.8) (risultati di simulazione)

L’addestramento della rete Fuzzy ART è limitato sulla sola media nominale del processo.

La seguente Figura 7.17 riporta il grafico degli stimatori puntuali $\hat{\alpha}$ dell’errore di Tipo I come funzione del parametro di vigilanza ρ . In particolare, le relazioni fra errore di Tipo I e parametro di vigilanza sono rappresentate per quattro differenti valori dell’ampiezza di finestra M .

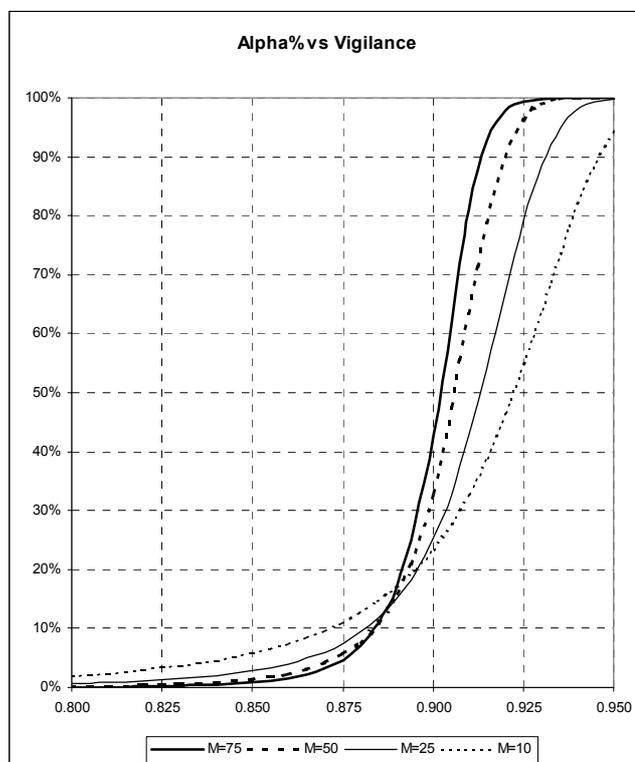


Figura 7.17 – Stimatori puntuali degli errori di Tipo I (ordinate) su dati AR(0.8) e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)

Nel seguente istogramma è stato riportata la distribuzione di densità relativa a dati che presentano uno shift costante di un'unità di deviazione standard, sia nel caso di indipendenza temporale, sia nel caso di presenza di auto-correlazione del primo ordine di fattore 0.8. Come per i casi precedenti, l'istogramma di rappresentato con linea spessa si riferisce a dati dipendenti, mentre quello rappresentato con linea più sottile, fa riferimento a dati indipendenti.

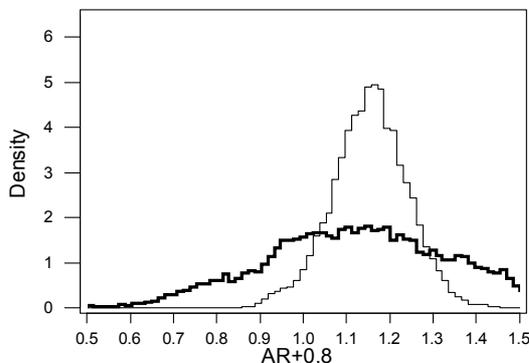


Figura 7.18– Distribuzione densità statistica MAD per finestre di 75 osservazioni

Dati non naturali (shift 1 unità di deviazione standard) per un processo AR di primo ordine a fattore 0.8 (linea spessa), dati non naturali (shift 1 unità di deviazione standard) indipendenti (linea sottile).

Le considerazioni che possono essere derivate dall'istogramma sono analoghe a quelle già fatte nel caso naturali. In particolare, la presenza di auto-correlazione positiva sembra provocare un aumento della dispersione della statistica MAD attorno al valore medio. Ciò lascia presupporre che la presenza di un'auto-correlazione positiva fra i dati implichi una diminuzione del tasso di decrescita degli errori di Tipo II all'aumentare del parametro di vigilanza.

Gli stimatori puntuali e intervallari dell'errore di Tipo II sono riportati nella seguente Tabella 7.14. Gli errori di Tipo II riportati nella tabella, si riferiscono al caso di un processo che presenta uno shift della media pari ad un'unità di deviazione standard ($\varphi = 1$), inoltre, è stato supposto che l'istante di partenza dello shift coincidesse con il primo istante di osservazione sul processo ($\tau = 1$). Come per tutti i casi precedenti, la notazione $[\hat{\beta}_-, \hat{\beta}_+]$ è stata usata per indicare l'intervallo di confidenza al 95% sullo stimatore $\hat{\beta}$ dell'errore di Tipo II.

ρ	$M = 10$			$M = 25$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	91.58%	91.81%	92.04%	95.26%	95.45%	95.63%
0.825	87.72%	88.05%	88.39%	91.39%	91.69%	91.99%
0.850	81.26%	81.68%	82.10%	83.21%	83.77%	84.33%
0.875	68.81%	69.42%	70.03%	64.60%	65.52%	66.43%
0.900	46.44%	47.13%	47.82%	28.67%	29.67%	30.68%
0.925	16.52%	17.06%	17.61%	2.79%	3.12%	3.45%
0.950	1.26%	1.42%	1.57%	0.00%	0.02%	0.04%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

ρ	$M = 50$			$M = 75$		
	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$	$\hat{\beta}_-$	$\hat{\beta}$	$\hat{\beta}_+$
0.800	97.05%	97.20%	97.35%	97.87%	98.00%	98.13%

0.825	93.18%	93.48%	93.79%	93.97%	94.33%	94.68%
0.850	81.71%	82.46%	83.22%	79.77%	80.57%	81.37%
0.875	50.77%	51.96%	53.14%	34.04%	35.51%	36.98%
0.900	8.05%	8.78%	9.52%	1.86%	2.23%	2.60%
0.925	0.05%	0.12%	0.19%	0.00%	0.00%	0.00%
0.950	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
0.975	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

Tabella 7.14 – Stimatori puntuali e intervallari dell’errore di Tipo II generati dalla rete neurale su dati AR(0.8) (risultati di simulazione) per uno shift della media di 1 unità di deviazione standard

L’addestramento della rete Fuzzy ART è limitato sulla sola media nominale del processo.

Nel grafico riportato dalle seguente Figura 7.19 è stato schematizzato l’andamento degli stimatori puntuali $\hat{\beta}$ come funzione del parametro di vigilanza ρ per quattro livelli dell’ampiezza di finestra M .

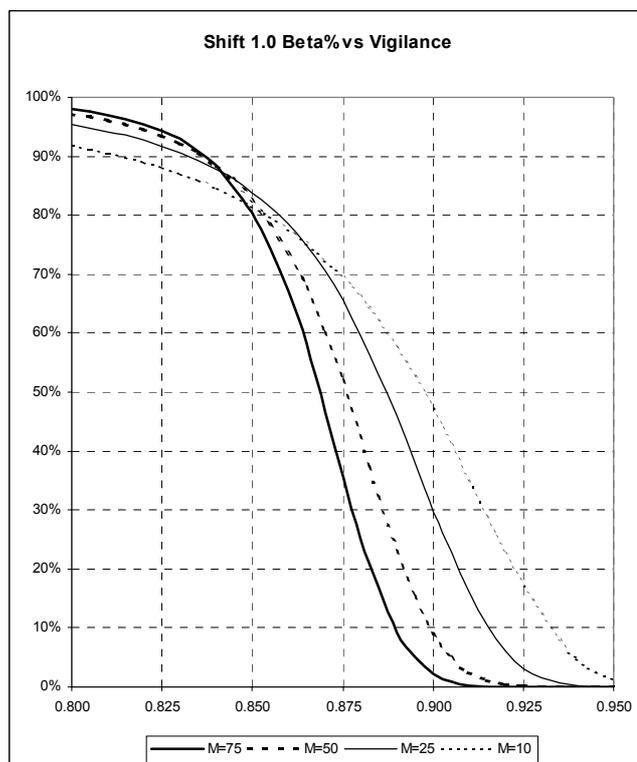


Figura 7.19 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(0.8) per shift di una unità di deviazione standard, e parametro di vigilanza della rete Fuzzy ART (ascissa) a quattro differenti livelli della ampiezza di finestra M (risultati di simulazione)

7.4.1 Comparazione delle prestazioni caso AR 0.8

Nella seguente tabella sono stati riportati i confronti fra gli errori di Tipo I e di Tipo II per una carta di controllo CUSUM $k = 0.5$ e $h = 4.7749$ e per la rete neurale Fuzzy ART di parametri $M = 75$ e $\rho = 0.8300$.

	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy ART</i> $M = 75$ $\rho = 0.8300$		<i>Fuzzy ART – CUSUM</i>		
<i>Natural</i>	$\hat{\alpha}_{cc}$	$\hat{\alpha}_{nn}$		$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$	$\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$

	0.269%	0.273%		-0.049%	0.004%	0.057%
	<i>CUSUM</i> $k = 0.5$ $h = 4.7749$	<i>Fuzzy ART</i> $M = 75$ $\rho = 0.8300$		<i>Fuzzy ART – CUSUM</i>		
<i>Systematic Variation</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.599%	99.680%		0.025%	0.081%	0.137%
0.50	99.198%	99.422%		0.153%	0.224%	0.295%
1.00	95.507%	21.925%		-74.775%	-73.582%	-72.389%
1.50	68.459%	0.000%		-68.508%	-68.459%	-68.410%
2.00	18.271%	0.000%		-18.345%	-18.271%	-18.197%
<i>Cyclic</i> $\Omega = 4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.669%	99.707%		-0.011%	0.038%	0.087%
0.50	99.458%	99.603%		0.083%	0.145%	0.207%
1.00	97.709%	98.483%		0.623%	0.774%	0.925%
1.50	89.312%	34.394%		-56.242%	-54.918%	-53.594%
2.00	70.740%	0.000%		-71.171%	-70.740%	-70.309%
<i>Shift</i> $\tau = 1$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.652%	99.616%		-0.917%	-0.036%	0.845%
0.50	99.451%	99.128%		-0.395%	-0.323%	-0.251%
1.00	98.252%	92.873%		-5.834%	-5.379%	-4.924%
1.50	92.169%	34.922%		-58.766%	-57.247%	-55.728%
2.00	45.869%	0.342%		-45.672%	-45.527%	-45.382%
<i>Mixture</i> $p_c = 0.4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
0.25	99.629%	99.666%		-0.071%	0.037%	0.145%
0.50	99.202%	99.338%		-0.167%	0.136%	0.439%
1.00	95.582%	73.601%		-23.144%	-21.981%	-20.818%
1.50	75.334%	0.023%		-75.388%	-75.311%	-75.234%
2.00	28.990%	0.000%		-29.004%	-28.990%	-28.976%

Tabella 7.15 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(0.8) (risultati di simulazioni)

La rete neurale Fuzzy ART è addestrata al riconoscimento della sola media naturale del processo.

I risultati riportati nella Tabella 7.15 dimostrano che la rete neurale risulta comparabile alla carta di controllo CUSUM in termini di errori di Tipo I. Risulta, inoltre, altrettanto evidente che l’algoritmo neurale presenta prestazioni migliori rispetto alla carta CUSUM nel riconoscimento di variazioni sistematiche, cicli e mixture ma solo per ampiezze maggiori o uguali a un’unità di deviazione standard.

La rete neurale, invece, presenta delle prestazioni migliori rispetto a quelle della carta di controllo con riferimento al pattern shift per qualunque livello di ampiezza sia preso in considerazione.

7.4.2 Miglioramento delle prestazioni della rete neurale

Anche se la rete neurale presenta già prestazioni migliori se confrontate a quelle della carta CUSUM nel riconoscimento di shift della media del processo, in questo

paragrafo sono riesaminate le prestazioni della rete neurale, nel monitoraggio di dati che presentano un'auto-regressione del primo ordine di fattore 0.8, supponendo di estendere la fase di addestramento al riconoscimento sia del target naturale del processo, sia di uno shift di 0.5 unità di deviazione standard.

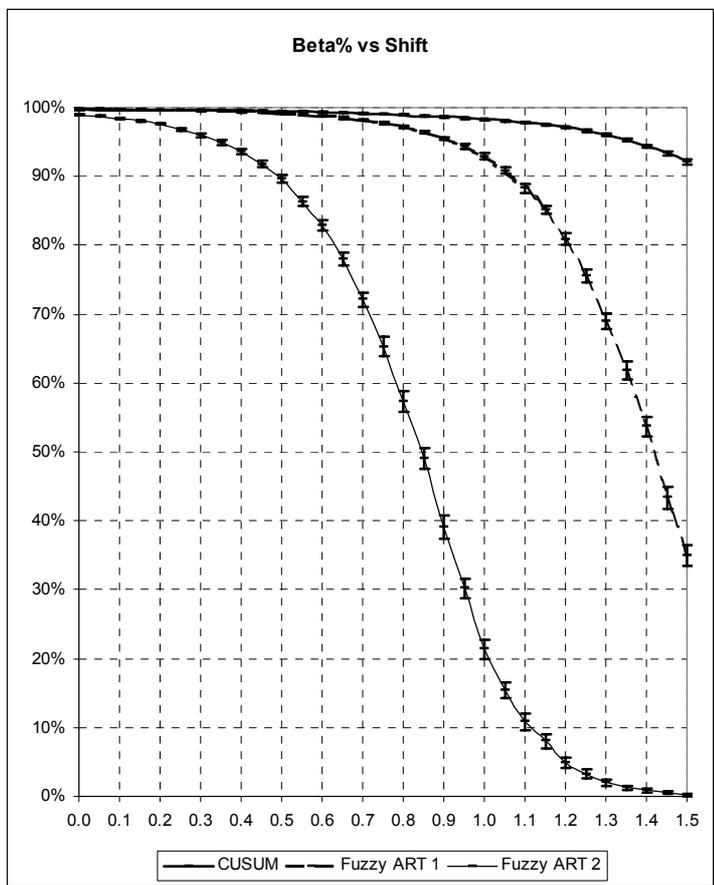
	<i>CUSUM</i> $k=0.5$ $h=4.7749$	<i>Fuzzy ART</i> $M=75$ $\rho=0.8000$		<i>Fuzzy ART - CUSUM</i>		
<i>Natural</i>	$\hat{\alpha}_{cc}$	$\hat{\alpha}_{nn}$		$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_-$	$\hat{\alpha}_{nn} - \hat{\alpha}_{cc}$	$(\hat{\alpha}_{nn} - \hat{\alpha}_{cc})_+$
	0.269%	1.034%		0.686%	0.765%	0.844%
	<i>CUSUM</i> $k=0.5$ $h=4.7749$	<i>Fuzzy ART</i> $M=75$ $\rho=0.8000$		<i>Fuzzy ART - CUSUM</i>		
<i>Systematic Variation</i>	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.599%	98.933%	-0.746%	-0.666%	-0.586%
	0.50	99.198%	98.890%	-0.393%	-0.308%	-0.223%
	1.00	95.507%	97.303%	1.646%	1.796%	1.946%
	1.50	68.459%	0.000%	-68.508%	-68.459%	-68.410%
	2.00	18.271%	0.000%	-18.345%	-18.271%	-18.197%
<i>Cyclic</i> $\Omega = 4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.669%	98.947%	-0.799%	-0.722%	-0.645%
	0.50	99.458%	98.955%	-0.589%	-0.503%	-0.417%
	1.00	97.709%	98.831%	0.980%	1.122%	1.264%
	1.50	89.312%	96.712%	7.055%	7.400%	7.745%
	2.00	70.740%	3.571%	-67.854%	-67.169%	-66.484%
<i>Shift</i> $\tau = 1$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.652%	96.886%	-3.664%	-2.766%	-1.868%
	0.50	99.451%	89.641%	-10.345%	-9.810%	-9.275%
	1.00	98.252%	21.352%	-78.316%	-76.900%	-75.484%
	1.50	92.169%	0.192%	-92.098%	-91.977%	-91.856%
	2.00	45.869%	0.000%	-45.869%	-45.869%	-45.869%
<i>Mixture</i> $p_c = 0.4$	$\hat{\beta}_{cc}$	$\hat{\beta}_{nn}$		$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_-$	$\hat{\beta}_{nn} - \hat{\beta}_{cc}$	$(\hat{\beta}_{nn} - \hat{\beta}_{cc})_+$
	0.25	99.629%	98.931%	-0.823%	-0.698%	-0.573%
	0.50	99.202%	98.870%	-0.639%	-0.332%	-0.025%
	1.00	95.582%	95.953%	-0.155%	0.371%	0.897%
	1.50	75.334%	1.160%	-74.420%	-74.174%	-73.928%
	2.00	28.990%	0.000%	-29.004%	-28.990%	-28.976%

Tabella 7.16 – Confronti fra le prestazioni della rete neurale e della carta CUSUM su dati AR(0.8) (risultati di simulazioni)

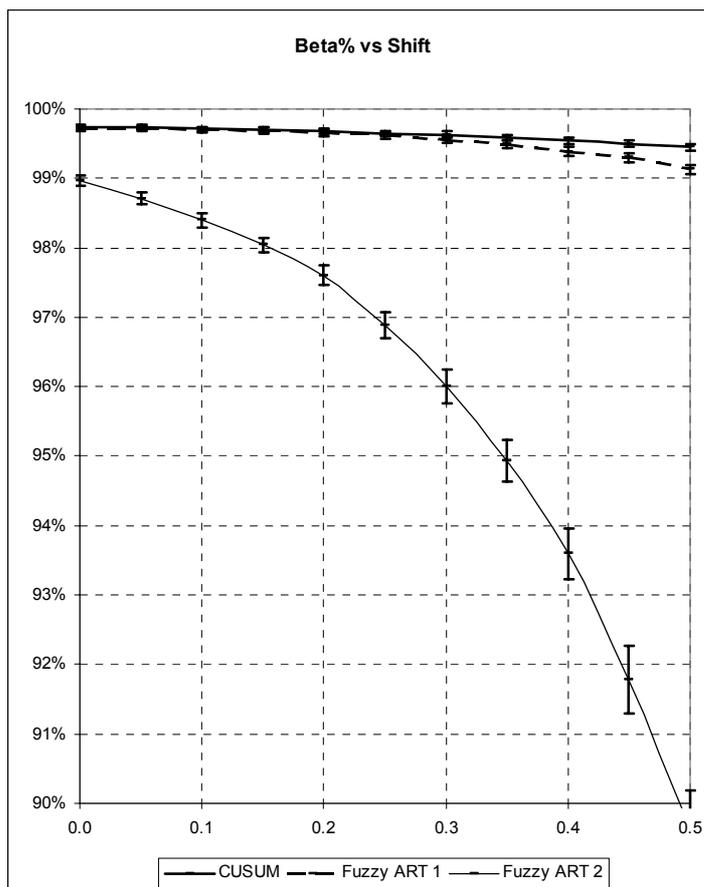
L'addestramento della rete Fuzzy ART comprende sia la media nominale del processo, sia uno shift di ampiezza 0.5 unità di deviazione standard.

Anche in questo caso, non è stato possibile tarare l'algoritmo neurale addestrato a riconoscere sia il target naturale del processo, sia una deviazione di mezza unità di deviazione standard, a fornire un tasso di falsi allarmi inferiore a circa 1%.

I grafici riportati dalla seguente Figura 7.20 riportano le curve beta per uno shift della media rispettivamente per i tre schemi di controllo precedentemente discussi.



a)



b)

Figura 7.20 – Stimatori puntuali degli errori di Tipo II (ordinate) su dati AR(0.8) per shift espressi in termini di unità di deviazione standard (ascissa) con rappresentazione degli stimatori intervallari a copertura del 95% (risultati di simulazione).

- a) intervallo di variazione ascissa è [0.0, 1.5], intervallo di variazione ordinate [0%, 100%]
- b) intervallo di variazione ascissa è [0.0, 0.5], intervallo di variazione ordinate [90%, 100%].

Carta CUSUM (linea continua in grassetto), Fuzzy ART con addestramento unico su media naturale del processo (linea tratteggiata), Fuzzy ART addestrata su media naturale e shift di ampiezza 0.5 unità di deviazione standard (linea continua sottile).

Conclusioni e sviluppi futuri

Le conclusioni dell'approccio proposto nel presente testo, e dei relativi risultati in precedenza discussi, possono essere tracciate sulla base delle osservazioni con cui è stato presentato il lavoro nel capitolo introduttivo.

Partendo dai metodi SPC, è possibile fare le seguenti considerazioni sulle prestazioni che possono fornire nelle applicazioni pratiche.

- La carta di controllo di Shewhart, essendo limitata all'utilizzo della osservazione più recente per la segnalazione di un fuori controllo, non risulta adeguata per l'identificazione di variazioni della media di bassa entità (per esempio di valori inferiori a 2 unità di deviazione standard).
- I metodi SPC, complementari alla carta di controllo di Shewhart, e proposti per migliorare le prestazioni del monitoraggio nel caso di piccole variazioni della media (ad esempio la carta CUSUM), si dimostrano adeguati all'identificazione di alterazioni costanti della media di processo (il che coincide con l'identificazione di un pattern di tipo shift) ma si rivelano inutili per identificare diversi tipi di alterazione, in particolare quelle che influenzano il processo con rapide fluttuazioni della media.
- I metodi SPC di riconoscimento pattern sulle carte di controllo (per esempio le run rules Western Electric) non possono essere ritenuti adeguati per implementare un sistema automatico di supporto per il miglioramento delle prestazioni di monitoraggio di un processo. Per questo motivo, infatti, a partire dai primi anni '90 diversi ricercatori hanno affrontato il problema di progettare nuovi metodi automatici (utilizzando algoritmi neurali o tecniche euristiche *ad hoc*) di supporto al riconoscimento automatico di pattern sulle carte di controllo.

Per quanto riguarda l'analisi della letteratura riguardante l'utilizzo di algoritmi neurali per il riconoscimento di pattern, alcuni limiti sono già stati discussi e si sintetizzano brevemente di seguito.

- Si tratta di algoritmi che richiedono lunghe fasi di addestramento per il raggiungimento di livelli di prestazioni superiori a quelli dei corrispondenti metodi SPC. Inoltre, per adottare in un'applicazione pratica un simile sistema, è necessario avere un insieme di dati di ampiezza sufficientemente ampio da raggiungere i livelli di prestazioni desiderati.
- Inoltre, gli addestramenti sono non cumulativi. Una volta addestrata una rete neurale di questo tipo si comporta come un sistema rigido di definite prestazioni e non più modificabile. Se si intende migliorarne le prestazioni, ad esempio aggiungendo nuovi dati di addestramento, è necessario sostituire completamente il sistema di monitoraggio con una nuova rete neurale per la quale sia riformulata nuovamente tutta la fase di addestramento.
- L'addestramento di un sistema neurale di tipo supervisionato, richiede che sia ipotizzabile a priori ogni condizione operativa nella quale il sistema di monitoraggio si troverà a funzionare. Questo, nel campo del controllo di qualità, implica preventivare non solo il possibile andamento dei dati osservati in condizioni naturali di variazione, ma anche tutte le possibili variazioni non naturali che possono essere osservate sul processo in fase operativa.

Si nota quindi che, per rendere operativa una simile rete neurale per il controllo di qualità, è necessario avere a disposizione un numero sufficiente di dati per garantire un'adeguata fase di addestramento (oppure disporre di dettagliati modelli matematici per simulare tali dati con tecnica Monte Carlo), ed è necessario che lo stato di funzionamento del processo sia noto a priori sia nella condizione naturale, sia in ogni possibile stato non naturale che si può manifestare.

È evidente che queste ipotesi limitino notevolmente il campo di applicabilità di molti dei sistemi neurali per il controllo di qualità proposti in letteratura. Per esempio, non è possibile la loro applicazione per monitorare un nuovo processo per il quale non sono disponibili dati di addestramento, oppure per il quale non è possibile formulare nessun modello matematico o statistico né per gli andamenti naturali né per quelli non naturali.

Tuttavia, in quei campi di applicazione nei quali il funzionamento del processo è ben noto, se si dispone di un numero sufficiente di dati storici che possono essere utilizzati per l'addestramento o per derivare modelli molto precisi sugli stati operativi sia in condizioni naturali di variabilità che in condizioni non naturali, le reti neurali supervisionate rappresentano indubbiamente la tecnica più efficace per il monitoraggio della qualità. Se, infatti, il funzionamento del processo è ben noto in ogni suo stato, naturale o non naturale, è evidente che è possibile definire un ambiente controllato di addestramento nel quale sono definiti esattamente i modi operativi di un sistema di controllo. In tal caso, la rete neurale supervisionata fornisce, grazie alla sua notevole capacità di generalizzazione, una grande efficienza del controllo che permette di migliorare notevolmente le prestazioni nel monitoraggio di un sistema rispetto ai metodi tradizionali SPC (come è già stato documentato ampiamente dalla letteratura esistente).

Al contrario, in tutti quei casi nei quali il funzionamento del processo non è noto, perché è un nuovo processo o ha subito delle modifiche importanti come un resettaggio, non è possibile applicare i metodi di controllo con algoritmo neurale supervisionato. Si ritiene che questa categoria di processi siano tra i più comuni a differenza di quelli per i quali è possibile stabilire modelli esatti e immutabili.

Queste osservazioni hanno motivato il lavoro che è stato presentato in questo testo. In particolare, il lavoro è stato mirato allo sviluppo di un sistema di monitoraggio, per il controllo qualità di un processo, che presenti caratteristiche di reale flessibilità ed efficienza.

In particolare, alcune caratteristiche innovative del metodo proposto possono essere elencate nei seguenti punti.

- La procedura neurale proposta non richiede una lunga fase di addestramento in un ambiente controllato ottenuto sulla base di una notevole mole di dati e di modelli. Al contrario, l'addestramento si può adattare alle sole informazioni disponibili. Queste potrebbero essere ridotte alla conoscenza del solo *target* di riferimento per il processo (questa è, ad esempio, l'unica informazione plausibile di cui si può disporre per un processo nuovo o per il quale non si ha nessun'informazione a priori). Oppure si può pensare di adottare un insieme di dati naturali di funzionamento (come discusso in dettaglio nell'appendice C). In ogni caso non è necessario disporre di alcun'informazione sui modi con cui eventuali stati non naturali di variazione si manifesteranno nel processo per rendere operativo il metodo di controllo proposto.
- La possibilità di un addestramento adattativo, implica la capacità di estendere la procedura di training non appena nuovi dati e informazioni sono disponibili. Per esempio nei risultati sperimentali è stato dimostrato

che le prestazioni della rete neurale proposta addestrata unicamente a riconoscere il target naturale del processo, possono essere notevolmente migliorate se si addestra successivamente la rete al riconoscimento anche di un particolare pattern non naturale.

- Il metodo neurale proposto, inoltre, si è dimostrato sufficientemente robusto anche nel caso di un processo con uscite dipendenti nel tempo. Ad esempio sono state valutate le prestazioni della rete neurale nel caso di un processo le cui uscite possano essere modellate come un processo Auto-Regressivo del primo ordine con diversi valori del fattore (sia positivi che negativi).

L'analisi sviluppata nel testo, inoltre, non è stata unicamente limitata a presentare dei risultati sperimentali ottenuti da una rete neurale tarata in modo vagamente empirico (come nella maggior parte dei casi presentati in letteratura), ma si è voluto, in particolare approfondire gli aspetti matematici dell'algoritmo neurale e quindi alcune implicazioni teoriche sulle prestazioni attese dal metodo nel caso in cui si adotti un modello statistico delle uscite del processo monitorato. Questo ha consentito da un lato di dimostrare teoricamente la validità dell'approccio, dall'altro di chiarire esattamente l'influenza di tutti i parametri dell'algoritmo sulle prestazioni attese.

L'analisi matematica dell'algoritmo Fuzzy ART, inoltre, ha consentito la definizione di un particolare metodo di addestramento (che è stato definito approccio PLP – Perfectly Learned Patterns) che si traduce nella semplificazione della fase di addestramento sia in termini di complessità computazionale, sia, conseguentemente nei tempi richiesti.

In conclusione, si ritiene che l'algoritmo neurale proposto possa efficacemente ed efficientemente trovare applicazione in quei casi pratici di controllo della qualità nei quali, da un lato è indispensabile garantire prestazioni migliori rispetto ai metodi tradizionali SPC, ma dall'altro non è possibile definire con esattezza un ambiente controllato di addestramento che permetta di implementare gli algoritmi neurali proposti nell'ultimo decennio nella letteratura di riferimento.

Alcuni degli sviluppi futuri che possono essere definiti sulla base di quanto sviluppato nel lavoro di ricerca sviluppato nel presente testo, possono essere brevemente tracciati dai seguenti punti.

- Approfondire il modello matematico e statistico del sistema in modo da prevedere con maggiore esattezza le prestazioni attese dall'uso della rete in determinati contesti. Per esempio, può essere interessante poter derivare la distribuzione probabilistica della grandezza utilizzata nell'algoritmo (MAD) per il controllo di qualità.
- Approfondire il comportamento della rete neurale, nel caso di sequenze di dati che presentano una dipendenza temporale, schematizzabile secondo una determinata classe di modelli ARIMA. Per esempio caratterizzare la grandezza MAD (con media e varianza) nel caso di alcune famiglie (tra la più comuni) di schemi di correlazione.
- Approfondire lo studio relativo alla combinazione dei metodi di monitoraggio automatico su rete neurale con i metodi di regolazione EPC.

In generale, si ritiene che la direzione verso la quale dovrà essere orientata la ricerca di metodi neurali per il monitoraggio della qualità, non può prescindere da una più approfondita indagine sull'apporto che essi possono fornire nel controllo automatico di processi che presentano una correlazione temporale delle osservazioni. Sarà inoltre importante in quest'ambito, valutare i modi con cui i metodi neurali possono combinarsi con le tecniche di regolazione EPC utilizzate per regolare i processi che presentano sia

dipendenza delle osservazioni, sia non stazionarietà di funzionamento, per migliorare ulteriormente le prestazioni di controllo.

Bibliografia

- [1] AL-GHANIM A., KAMAT S. J., 1995, Unnatural pattern recognition on control charts using correlation analysis techniques. *Computers and Industrial Engineering*, **29**(1), 43-47.
- [2] AL-GHANIM A., 1997, An unsupervised learning neural algorithm for identifying process behavior on control charts and a comparison with supervised learning approaches. *Computers and Industrial Engineering*, **32**(3), 627-639.
- [3] AL-GHANIM A., LUDEMAN L. C., 1997, Automated unnatural pattern recognition on control charts using correlation analysis techniques. *Computers and Industrial Engineering*, **32**(3), 679-690.
- [4] ANGLANI A., PACELLA M., SEMERARO Q., 2001, A Special-Purpose Neural Network Recogniser to Detect Non-Random Pattern on Control Charts. *Proc. of the 5th International Conference on Artificial Neural Networks and Genetic Algorithms (ICANNGA 2001), Prague, Czech Republic, April 22-25, pagg. 449-452, (ed. Springer-Verlag Wien).*
- [5] BISHOP C. M., 1995, *Neural Networks for Pattern Recognition*, (Oxford University Press Inc.).
- [6] BLUME M., ESENER S. C., 1997, An efficient mapping of Fuzzy ART onto a Neural Architecture. *Neural Networks Letters*, **10**(3), 409-411.
- [7] BOX G. E. P., JENKINS G. M., REINSEL G. C., 1994, *Time Series Analysis. Forecasting and Control*. 3rd ed. (Prentice Hall).
- [8] BURKE L., IGNIZIO J. P., 1997, A practical overview of neural networks, *Journal of Intelligent Manufacturing*, **8**, 157-165.
- [9] BURWICK T., JOUBLIN F., 1998, Optimal Algorithmic Complexity of Fuzzy ART *Neural Processing Letters*, **7**, 37-41.
- [10] CHANG S. I., AW C. A., 1996, A neural fuzzy control chart for detecting and classifying process mean shifts. *International Journal of Production Research*, **34**(8), 2265-2278.
- [11] CHANG S. I., HO E. S., 1999, A two-stage neural network approach for process variance change detection and classification. *International Journal of Production Research*, **37**(7), 1581-1599.
- [12] CHENG C. -S., 1995, A multi-layer neural network model for detecting changes in the process mean. *Computers and Industrial Engineering*, **28**(1), 51-61.
- [13] CHENG C.-S., HUBELE N. F., 1996, A pattern recognition algorithm for an X control chart. *IIE Transactions*, **28**, 215-224.
- [14] CHENG C. -S. 1997, A neural network approach for the analysis of control chart patterns. *International Journal of Production Research*, **35**(3), 667-697.
- [15] CHENG C. -S., CHENG S. S., 2001, A neural network-based procedure for the monitoring of the exponential mean. *Computers and Industrial Engineering*, **40**, 309-321.
- [16] COOK D. F., CHIU C. C., 1998, Using radial basis function neural networks to recognize shifts in correlated manufacturing process parameters. *IIE Transactions*, **30**, 227-234.
- [17] COOK D. F., ZOBEL C. W., NOTTINGHAM Q. J., 2001, Utilization of neural networks for the recognition of variance shifts in correlated manufacturing process parameters, *International Journal of Production Research*, **39**(17), 3881-3887.
- [18] DEDEAKAYOĞULLARI I., BURNAK N., 1999, The determination of mean and/or variance shifts with artificial neural networks. *International Journal of Production Research*, **37**(10), 2191-2200.
- [19] GEORGIPOULOS M., FERNLUND H., BEBIS G., HEILEMAN G. L., 1996, Order of search in Fuzzy ART and Fuzzy ARTMAP: effect of the choice parameter. *Neural Networks*, **9**(9), 1541-1559.
- [20] GEORGIPOULOS M., DAGHER I., HEILEMAN G. L., BEBIS G., 1999, Properties of learning of a Fuzzy ART Variant, *Neural Networks*, **12**, 837-850.

- [21] GUH R. -S., HSIEH Y. C., 1999, A neural network based model for abnormal pattern recognition of control charts. *Computers and Industrial Engineering*, **36**, 97-108.
- [22] GUH R. -S., TANNOCK J. D. T., 1999a, Recognition of control chart concurrent patterns using a neural network approach. *International Journal of Production Research*, **37**(8), 1743-1765.
- [23] GUH R. -S., TANNOCK J. D. T., 1999b, A neural network approach to characterize pattern parameters in process control charts. *Journal of Intelligent Manufacturing*, **10**, 449-462.
- [24] GUH R. S., ZORRIASSATINE F., TANNOCK J. D. T., O'BRIEN C., 1999, On-line control chart pattern detection and discrimination – a neural network approach. *Artificial Intelligence in Engineering*, **13**, 413-25.
- [25] GUO Y., DOOLEY K. J., 1992, Identification of changes structure in statistical process control. *International Journal of Production Research*, **30**(7), 1655-1669.
- [26] HAYKIN S., 1999, *Neural Networks: a comprehensive foundation*, 2nd ed., (Prentice-Hall).
- [27] HAGAN M. T., DEMUTH H. B., BEALE M., 1996, *Neural Network Design*, (PWS Publishing Company).
- [28] HUANG J., GEORGIPOULOS M., HEILEMAN J. L., 1995, Fuzzy ART Proprieties. *Neural Networks*, **8**(2), 203-213.
- [29] HWARNG H. B., HUBELE N. F., 1993a, X-bar control chart pattern identification through efficient off-line neural network training. *IIE Transactions*, **25**(3), 27-40.
- [30] HWARNG H. B., HUBELE N. F., 1993b, Back-propagation pattern recognizers for X-bar control charts: methodology and performance. *Computers and Industrial Engineering*, **24**(2), 219-235.
- [31] HWARNG H. B., 1995a, Proper and effective training of a pattern recognizer for cyclic data, *IIE Transactions* **27**, 746-756.
- [32] HWARNG H. B., 1995b, Multilayer perceptrons for detecting cyclic data on control chart, *International Journal of Production Research*, **33**(11), 3101-3117.
- [33] HWARNG H. B., CHONG C. W., 1995, Detecting process non-randomness through a fast and cumulative learning ART-based pattern recognizer, *International Journal of Production Research*, **33**(7), 1817-1833.
- [34] KANG B.-S., PARK S.-C., 2000, Integrated machine learning approaches for complementing statistical process control procedures. *Decision Support Systems*, **29**, 59-72.
- [35] KIM M.-H., JANG D.-S., YANG, Y.-K. 2001, A robust-invariant pattern recognition model using Fuzzy ART. *Pattern Recognition*, **34**, 1685-1696.
- [36] KOUFAKOU A., GEORGIPOULOS M., ANAGNOSTOPOULOS G., KASPARIS T., 2001, Cross-validation in Fuzzy ARTMAP for large databases, *Neural Networks*, **14**, 1279-1291.
- [37] LEGER R. P., GARLAND W. J., POEHLMAN W. F. S., 1998, Fault detection and diagnosis using statistical control charts and artificial neural networks. *Artificial Intelligence in Engineering*, **12**, 35-47.
- [38] MARRIOT S., HARRISON R. F., 1995, A Modified Fuzzy ARTMAP Architecture for Approximation of Noisy Mappings. *Neural Networks*, **8**(4), 619-641.
- [39] MONTGOMERY D. C., 2000, *Introduction to Statistical Quality Control*, 4th ed. (J. Wiley).
- [40] PACELLA M., SEMERARO Q., ANGLANI A., 2002a, On the use of adaptive resonance theory based neural algorithms for manufacturing process quality control. Submitted to *International Journal of Production Research*, ref. IJPR 20229 (September 2002).
- [41] PACELLA M., SEMERARO Q., ANGLANI A., 2002b, Manufacturing quality control by means of a Fuzzy ART network trained on natural process data. Submitted to *Engineering Applications of Artificial Intelligence*, ref. 20021025/1 (October 2002).
- [42] PACELLA M., SEMERARO Q., ANGLANI A., 2002c, Manufacturing process quality control by means of a Fuzzy ART neural algorithm. Submitted to *6th International Conference on Artificial Neural Networks and Genetic Algorithms (ICANNGA 2003)*, Roanne, France.
- [43] PAO Y. H., 1989, *Adaptive pattern recognition and neural networks*. (Addison-Wesley).

- [44] PAPOULIS A., 1973, *Probabilità, Variabili Aleatorie e Processi Stocastici*, (Boringhieri).
- [45] PERRY M. B., SPOERRE J. K., VELASCO T., 2001, Control chart pattern recognition using back propagation artificial neural networks. *International Journal of Production Research*, **39**(15), 3399-3418.
- [46] PHAM D. T., OZTEMELE E., 1994, Control chart pattern recognition using learning vector quantization networks, *International Journal of Production Research* **23**(3), 721-729.
- [47] PHAM D.T., SAGIROGLU S., 2001, Training multilayered perceptrons for pattern recognition: a comparative study of four training algorithms. *International Journal of Machine Tools & Manufacture*, **41**, 419-430.
- [48] PHAM D. T., WANI M. A., 1997, Feature-based control chart pattern recognition. *International Journal of Production Research*, **35**(7), 1875-1890.
- [49] PUGH G. A., 1989, Synthetic Neural Networks for Process Control, *Computers and Industrial Engineering*, **17**, 24-26.
- [50] PUGH G. A., 1991, A Comparison of Neural Networks to SPC Charts, *Computers and Industrial Engineering*, **21**, 253-255.
- [51] REDDY D. C., GHOSH K., VARDHAN V. A., 1998, Identification and Interpretation of Manufacturing Process Patterns through Neural Networks, *Mathematical Computing Modelling*, **27**(5), 15-36.
- [52] SMITH A. E., 1994, X-bar and R control chart interpretation using neural computing. *International Journal of Production Research*, **32**(2), 309-320.
- [53] STUART M., MULLINS E., DREW E., 1996, Statistical quality control and improvement. *European Journal of Operational Research* **88**, 203-214.
- [54] SWIFT J. L., MIZE, J. H., 1995, Out-of-control pattern recognition and analysis for quality control charts using LISP-based systems. *Computers and Industrial Engineering*, **28**(1), 81-91.
- [55] WANG T.-Y., CHEN, L.-H., 2002, Mean shifts detection and classification in multivariate process: a neural-fuzzy approach. *Journal of Intelligent Manufacturing*, **13**, 211-221.
- [56] WEST D. A., MANGIAMELI P. M., CHEN S. K., 1999, Control of complex manufacturing processes: a comparison of SPC methods with radial basis function neural network. *Omega, International Journal Management Science*, **27**, 349-362.
- [57] ZORRIASSANTINE F., TANNOCK J. D. T., 1998, A review of neural networks for statistical process control. *Journal of Intelligent Manufacturing*, **9**, 209-224.

Appendice A

La metodologia di sperimentazione

A.1 Stima delle prestazioni

Un sistema di monitoraggio della qualità, come ad esempio una carta di controllo oppure un algoritmo neurale, è progettato allo scopo di eseguire una distinzione fra classi di dati naturali e non naturali. Pertanto, per misurarne le prestazioni, due caratteristiche devono essere valutate analiticamente, o stimate per mezzo della simulazione Monte Carlo.

La prima di queste caratteristiche è la capacità del sistema di controllo di modellare le cause naturali (non assegnabili) senza creare errori di Tipo I. Gli errori di Tipo I rappresentano falsi allarmi, generati erroneamente dal sistema di controllo, per un processo in uno stato naturale. Questa proprietà può essere stimata riportando la media degli errori di Tipo I, (ovvero la media campionaria dei segnali di errore in un run di simulazione, diciamo $\hat{\alpha}$), che si verificano nel controllo di un processo che presenta unicamente cause naturali (ovvero non assegnabili) di variazione. Questo valore rappresenta uno stimatore puntuale consistente del parametro $\alpha = P\{H_1|H_0\}$, ovvero l'effettivo tasso atteso di segnali di allarme generati dal sistema di controllo quando il processo è in uno stato naturale di variazione.

Inoltre, il sistema di controllo deve essere in grado di identificare velocemente le condizioni non naturali di variazione nei dati di uscita di un processo. Questa proprietà è misurata sperimentalmente valutando il tasso di errori di Tipo II (ovvero la media del numero di segnali di non-allarme generati dal sistema di controllo in un run di simulazione, diciamo $\hat{\beta}$) prodotti dal sistema di controllo su un processo in uno stato non naturale di variazione (ovvero soggetto a cause assegnabili di variazione). Questo valore è uno stimatore puntuale consistente del parametro $\beta = P\{H_0|H_1\}$, ovvero del tasso atteso di segnali di non allarme, emessi dal sistema di controllo quando il processo monitorato presenta delle cause speciali di variazione. Ovviamente, quest'ultima misura delle prestazioni dipenderà dai parametri del segnale utilizzato per simulare gli effetti sull'uscita del processo di una causa assegnabile di variazione.

A.2 Analisi dei risultati della simulazione

Al fine di valutare correttamente i risultati ottenuti dalla simulazione, in questo lavoro, è stato scelto il metodo *batch mean*. Questo metodo consiste nel dividere l'output di un lungo run di simulazione in un numero continuo di batch e di usare la media campionaria ottenuta considerando le misure di prestazione valutate in ogni singolo batch (ovvero le medie di batch) al fine di produrre stimatori puntuali e intervallari delle misure di prestazione per il sistema di controllo.

Il principale problema con l'applicazione del metodo del *batch mean* in pratica, è la scelta dell'ampiezza dei batch. Infatti, se l'ampiezza dei batch è eccessivamente ridotta le medie delle misure di prestazione calcolate in ognuno di essi potrebbero risultare correlate (ovvero non indipendenti statisticamente) e ciò potrebbe risultare in intervalli di confidenza con coperture inferiori rispetto a quelle nominali previste (ad esempio coperture inferiori rispetto al 95% con la quale lo stimatore intervallare è stato calcolato). Nel caso contrario, ipotizzando una lunghezza limitata del run di simulazione, un'ampiezza molto ampia dei batch può comportare un numero limitato di batch con potenziali problemi nell'applicazione del teorema del limite centrale per il calcolo degli stimatori intervallari.

In tutti gli esperimenti, riportati nel presente lavoro, è stato scelto un numero di 50 batch di 2000 dati come dimensione dell'esperimento di simulazione al fine di stimare sia l'errore di Tipo I che l'errore di Tipo II. Questa dimensione è stata scelta con l'obiettivo di soddisfare i seguenti due criteri: 1) indipendenza della sequenza delle medie di batch verifica attraverso un test sulla correlazione al lag 1; 2) ampiezza massima degli stimatori intervallari (con copertura del 95%) inferiore a 3% sia sull'errore di Tipo I sia sull'errore di Tipo II.

A.3 Relazioni di stima della prestazioni

Come accennato in precedenza, il metodo del batch mean divide le uscite di un lungo run di simulazione in un numero di batch continui. Indichiamo con $\{b_t\}$ la sequenza di segnali generati da un sistema di controllo (rete neurale o carta di controllo), dove $t = 1, 2, \dots$ rappresenta un indice temporale intero, corrispondente alla generazione di un output nel processo sottoposto a controllo. Assumiamo che tali segnali risultino binari ed in particolare che risultino le seguenti associazioni.

$$\{b_t\}; b_t = 1 \rightarrow H_0; b_t = 0 \rightarrow H_1 \quad \text{[A 1]}$$

In altri termini, ad un determinato istante di tempo t , si assume che un valore di uscita pari a $b_t = 1$ indichi una situazione di non allarme (processo in uno stato naturale). Al contrario, un segnale di uscita pari a $b_t = 0$ è utilizzato per segnalare un allarme (processo in uno stato non naturale). Allora, le seguenti relazioni possono essere utilizzate per rappresentare rispettivamente l'errore di Tipo I e l'errore di Tipo II.

$$\alpha = \lim_{t \rightarrow +\infty} \frac{1}{t} \sum_{\tau=1}^t (1 - b_\tau) \Big|_{H_0}; \beta = \lim_{t \rightarrow +\infty} \frac{1}{t} \sum_{\tau=1}^t b_\tau \Big|_{H_1} \quad \text{[A 2]}$$

Ovviamente di tali misure possiamo calcolare solo ed esclusivamente degli stimatori affetti da errori poiché limitiamo la simulazione ad un istante finito di tempo t . Pertanto, saranno utilizzati i seguenti stimatori puntuali consistenti rispettivamente per l'errore di Tipo I e per l'errore di Tipo II.

$$\hat{\alpha} = \frac{1}{t} \sum_{\tau=1}^t (1 - b_\tau) \Big|_{H_0}; \hat{\beta} = \frac{1}{t} \sum_{\tau=1}^t b_\tau \Big|_{H_1} \quad \text{[A 3]}$$

Al fine di valutare l'errore che si commette nel misurare gli errori di Tipo I e di Tipo II con gli stimatori puntuali riportati nell'equazione, si adotta il metodo delle medie di batch. Assumiamo che il processo $\{b_t\}$ sia debolmente stazionario con $\lim_{t \rightarrow +\infty} t \text{Var}[\hat{\alpha}] < \infty$

e $\lim_{t \rightarrow +\infty} tVar[\hat{\beta}] < \infty$, e supponiamo di suddividere i dati in k batch ognuno consistente di m osservazioni. Assumendo per semplicità che $t = km$, allora l' s^{esimo} batch consisterà delle seguenti osservazioni.

$$b_{(s-1)m+1}, b_{(s-1)m+2}, \dots, b_{sm} \quad [\text{A } 4]$$

Le medie del batch s^{esimo} sono date dalle seguenti relazioni rispettivamente per l'errore di Tipo I e per l'errore di Tipo II.

k batches of size m ; $s = 1, 2, \dots, k$

$$\hat{\alpha}_s(m) = \frac{1}{m} \sum_{\tau=1}^m (1 - b_{(s-1)m+\tau}) \Big|_{H_0} ; \hat{\beta}_s(m) = \frac{1}{m} \sum_{\tau=1}^m b_{(s-1)m+\tau} \Big|_{H_1} \quad [\text{A } 5]$$

È evidente che le medie di batch sono dipendenti dall'ampiezza scelta, e che per semplicità è stata supposta uguale per tutti i k batch della simulazione. È ovvio che gli stimatori puntuali rispettivamente per gli errori di Tipo I e di Tipo II possano essere riscritti come segue.

$$\hat{\alpha} = \frac{1}{k} \sum_{s=1}^k \hat{\alpha}_s(m) \Big|_{H_1} ; \hat{\beta} = \frac{1}{k} \sum_{s=1}^k \hat{\beta}_s(m) \Big|_{H_1} \quad [\text{A } 6]$$

Vale a dire come grande media delle medie di batch. Supponendo, quindi, che la lunghezza del run di simulazione sia fissata e pari a t . Indichiamo con

$$\sigma_{\hat{\alpha}}^2 = Var[\hat{\alpha}], \quad \sigma_{\hat{\beta}}^2 = Var[\hat{\beta}] \quad [\text{A } 7]$$

la varianza dei due stimatori che otteniamo dalla simulazione, allora, poiché il processo è debolmente stazionario per ipotesi, e dunque anche la serie degli stimatori risulta debolmente stazionario, si può scrivere che:

$$\begin{aligned} \sigma_{\hat{\alpha}}^2 &= Var[\hat{\alpha}] = Var\left[\frac{1}{k} \sum_s \hat{\alpha}_s(m)\right] = \frac{1}{k^2} \sum_{s,r} Cov[\hat{\alpha}_s(m), \hat{\alpha}_r(m)] \\ \sigma_{\hat{\alpha}}^2 &= \frac{\sigma_{am}^2}{k} + \frac{1}{k^2} \sum_{s \neq r} Cov[\hat{\alpha}_s(m), \hat{\alpha}_r(m)] = \frac{\sigma_{am}^2}{k} + \left(\frac{t\sigma_{\hat{\alpha}}^2 - m\sigma_{am}^2}{t}\right) \end{aligned} \quad [\text{A } 8]$$

Avendo indicato con σ_{am}^2 la varianza fra i batch dello stimatore dell'errore di Tipo I (ovvero $\sigma_{am}^2 = Var[\hat{\alpha}_s(m)]$), che risulterà essere dipendente dalla ampiezza m scelta per i batch.

Poiché $t = km$, $\lim_{t \rightarrow +\infty} tVar[\hat{\alpha}] < \infty$ e $\lim_{t \rightarrow +\infty} tVar[\hat{\alpha}_s(m)] < \infty$ risulta quindi che

$$\left(\frac{t\sigma_{\hat{\alpha}}^2 - m\sigma_{am}^2}{t}\right) \xrightarrow{t \rightarrow \infty} 0 \quad \text{allora si ha che } \frac{\sigma_{am}^2}{k} \text{ approssima } \sigma_{\hat{\alpha}}^2 \text{ con un errore che}$$

diminuisce all'aumentare di t e di m . Equivalentemente, si può dire che: la correlazione tra le medie di batch diminuiscono come t e m approssimano $+\infty$.

Ovviamente il risultato precedente vale anche per lo stimatore dell'errore di Tipo II ovvero:

$$\sigma_{\hat{\beta}}^2 = Var[\hat{\beta}] = Var\left[\frac{1}{k} \sum_s \hat{\beta}_s(m)\right] = \frac{1}{k^2} \sum_s Cov[\hat{\beta}_s(m), \hat{\beta}_r(m)]$$

$$\sigma_{\hat{\beta}}^2 = \frac{\sigma_{\beta m}^2}{k} + \frac{1}{k^2} \sum_{s \neq r} Cov[\hat{\beta}_s(m), \hat{\beta}_r(m)] = \frac{\sigma_{\beta m}^2}{k} + \left(\frac{t\sigma_{\hat{\beta}}^2 - m\sigma_{\beta m}^2}{t}\right) \quad \text{[A 9]}$$

Avendo posto $\sigma_{\beta m}^2 = Var[\hat{\beta}_s(m)]$. In definitiva, utilizzando gli stimatori puntuali degli errori di Tipo I e di Tipo II di seguito riportati

$$\hat{\alpha} = \frac{1}{t} \sum_{\tau=1}^t (1 - b_{\tau}) \Big|_{H_0} ; \hat{\beta} = \frac{1}{t} \sum_{\tau=1}^t b_{\tau} \Big|_{H_1} \quad \text{[A 10]}$$

e stimando le varianze tra i batch (sopra indicata con $\sigma_{\alpha m}^2$ e $\sigma_{\beta m}^2$ rispettivamente) dei diversi stimatori calcolati attraverso le seguenti relazioni

$$\hat{\sigma}_{\alpha m}^2 = \frac{1}{k-1} \sum_{s=1}^k (\hat{\alpha}_s(m) - \hat{\alpha}); \hat{\sigma}_{\beta m}^2 = \frac{1}{k-1} \sum_{s=1}^k (\hat{\beta}_s(m) - \hat{\beta}) \quad \text{[A 11]}$$

è possibile ottenere un intervallo di confidenza al 95% (utilizzando la statistica t di Student) attraverso le seguenti relazioni.

$$\hat{\alpha} \pm t_{k-1, 1-\frac{0.05}{2}} \sqrt{\frac{\hat{\sigma}_{\alpha m}^2}{k}}; \hat{\beta} \pm t_{k-1, 1-\frac{0.05}{2}} \sqrt{\frac{\hat{\sigma}_{\beta m}^2}{k}} \quad \text{[A 12]}$$

A.3.1 Alcuni vantaggi del metodo batch mean

- Le medie dei batch tendono ad essere distribuite in modo normale poiché esse rappresentano delle medie. Questa proprietà può essere molto utile quando si utilizzano delle statistiche di calcolo degli stimatori basate sull'ipotesi di normalità.
- Le medie batch tendono ad essere distribuite in modo indipendente. In altri termini, considerare le medie batch tende a ridurre la eventuale correlazione temporale tra le osservazioni. Questa proprietà risulta particolarmente utile quando si utilizzano statistiche per il calcolo degli stimatori basate sull'ipotesi di indipendenza fra le osservazioni.
- L'uso di differenti ampiezze dei batch permette di poter ottenere differenti varianze sugli stimatori. Questo può essere particolarmente utile quando si utilizza il metodo dei batch per confrontare le medie di due differenti sistemi utilizzando statistiche di confronto basate sull'ipotesi di uguaglianza delle varianze fra le due popolazioni.

A.3.2 Confronto

Supponiamo di confrontare le prestazioni fra due differenti sistemi di monitoraggio di un processo. Per esempio confrontiamo un metodo neurale (per il quale utilizzeremo il suffisso nn) ed una carta di controllo SPC (per la quale utilizzeremo il suffisso cc). Allora, utilizzando il medesimo esperimento (stesso numero di batch e stessa ampiezza), è possibile calcolare gli intervalli di confidenza sulla differenza degli stimatori per mezzo delle seguenti relazioni.

$$\hat{\alpha}_{nn} - \hat{\alpha}_{cc} \pm t_{2k-2,1-\frac{0.05}{2}} \sqrt{\frac{\hat{\sigma}_{m\alpha_{nn}}^2 + \hat{\sigma}_{m\alpha_{cc}}^2}{k}}; \hat{\beta}_{nn} - \hat{\beta}_{cc} \pm t_{2k-2,1-\frac{0.05}{2}} \sqrt{\frac{\hat{\sigma}_{m\beta_{nn}}^2 + \hat{\sigma}_{m\beta_{cc}}^2}{k}} \quad \text{[A 13]}$$

Appendice B

Risultati sperimentali analitici

B.1 Descrizione

In questa sezione sono stati riportati i risultati analitici ottenuti dalla simulazione Monte Carlo e inerenti l'utilizzo dell'algoritmo neurale Fuzzy ART descritto nei capitoli precedenti. I risultati fanno riferimento sia al caso di dati di processo indipendenti, sia correlati secondo uno schema auto-regressivo del primo ordine con fattore $\phi \in \{-0.8, -0.2, 0.2, 0.8\}$.

Per ognuno dei casi presi in considerazione sono stati riportati gli errori di Tipo I e di Tipo II, questi ultimi con riferimento ad uno shift costante della media di processo pari ad un'unità di deviazione standard ($\phi=1$) a partire dal primo istante di osservazione ($\tau=1$). Per entrambe le misure (errore di Tipo I e di Tipo II), sono indicati sia gli stimatori puntuali, (indicati con $\hat{\alpha}$ e $\hat{\beta}$ rispettivamente), sia le semiampiezze dei corrispondenti stimatori intervallari con copertura del 95%, di punto medio pari allo stimatore puntuale, (indicati con $\pm\Delta\hat{\alpha}$ e $\pm\Delta\hat{\beta}$ rispettivamente per l'errore di Tipo I e di Tipo II). In formula, valgono relazioni riportate nella seguente equazione.

$$\Delta\hat{\alpha} = t_{k-1,1-\frac{0.05}{2}} \sqrt{\frac{\hat{\sigma}_{\alpha m}^2}{k}}; \Delta\hat{\beta} = t_{k-1,1-\frac{0.05}{2}} \sqrt{\frac{\hat{\sigma}_{\beta m}^2}{k}} \quad \text{[B 1]}$$

Avendo indicato con k il numero di batch di simulazione ($k=50$), con m la lunghezza dei batch ($m=2000$), con $t_{k-1,1-\frac{0.05}{2}}$ il quantile corrispondente al 95% della distribuzione cumulata t di Student a $k-1$ gradi di libertà ($t_{k-1,1-\frac{0.05}{2}} = 2.0096$) e con

$\hat{\sigma}_{\alpha m}^2$ ($\hat{\sigma}_{\beta m}^2$) la varianza campionaria stimata fra i k batch di lunghezza m nella stima puntuale dell'errore di Tipo I (o di Tipo II rispettivamente).

Entrambe le misure di prestazione sono state calcolate per 4 livelli di ampiezza della finestra $M \in \{10, 25, 50, 75\}$ e 32 livelli del parametro di vigilanza ρ variabile nell'intervallo $[0.800, 0.950]$ con incrementi di 0.005.

Nelle stesse sezioni, inoltre, sono stati riportati i dati di confronto fra le prestazioni dell'algoritmo Fuzzy ART e della carta di controllo CUSUM di parametri $k=0.5$ e $h=4.7749$. I dati riportano lo stimatore puntuale dell'errore di Tipo II (indicato con $\hat{\beta}$), e la semi ampiezza del corrispondente intervallo di confidenza al 95%, relativi ad uno shift della media del processo di ampiezza variabile nell'intervallo $[0.0, 1.5]$ con incrementi di 0.05 unità di deviazione standard. Si osservi che uno shift di ampiezza nulla implica il processo nel suo stato naturale di variazione, dunque la misura indicata

deve essere considerata il complemento a 1 dell'errore di Tipo I presentato dal sistema di controllo.

In particolare, accanto ai dati relativi la carta di controllo CUSUM, sono stati riportati quelli ottenuti dalla rete neurale considerando due distinte strategie di addestramento. Con l'etichetta "Fuzzy ART 1" sono stati indicati i risultati della rete neurale il cui training è limitato alla sola media naturale del processo. Con l'etichetta "Fuzzy ART 2" sono stati indicati i risultati ottenuti dalla rete neurale addestrata al riconoscimento sia del target naturale del processo, sia di uno shift costante e di ampiezza 0.5 unità di deviazione standard. In entrambi i casi i risultati sono stati ottenuti da una rete Fuzzy ART di parametro $M = 75$ e ρ determinato in modo tale da ottenere un tasso di errori di Tipo I comparabile con quello della carta di controllo assunta come riferimento (ovvero di circa 0.27%). Per garantire che la comparazione fra prestazioni della rete neurale con quelle della carta di controllo CUSUM sia bilanciata, tutti i risultati di seguito riportati sono stati ottenuti ignorando gli eventuali allarmi prodotti dalla carta CUSUM durante le prime $M - 1$ osservazioni. In altri termini le prestazioni dei due sistemi di controllo sono state misurate per tutti gli indici $t \geq M$.

Tutti i risultati sono stati ottenuti attraverso simulazione Monte Carlo. In particolare, gli stimatori puntuali e intervallari sono stati calcolati attraverso il metodo *batch mean* sulla base di 50 run di 2000 osservazioni temporali. Le stime dei tassi di allarme, inoltre, sono stati ottenuti ri-generando la successione di osservazioni ad ogni allarme segnalato dal sistema di controllo.

Il generatore di numeri pseudo-casuali per la simulazione Monte Carlo è quello interno all'ambiente MATLAB 5 (identico a quello delle versioni 5.3 o 6.0), il seme del generatore è pari a 931316785, ovvero quello di default dell'ambiente di simulazione impostato all'avvio dell'applicazione. Lo stato iniziale del generatore corrisponde alla coppia di valori [362436069,521288629] vale a dire lo stato default (indice nullo) del generatore all'avvio dell'applicazione.

B.2 Dati indipendenti

B.2.1 Errori di Tipo I

ρ	$M=10$		$M=25$		$M=50$		$M=75$	
	$\hat{\alpha}$	$\pm \Delta \hat{\alpha}$						
0.800	0.00770	0.00051	0.00024	0.00008	0.00000	0.00000	0.00000	0.00000
0.805	0.00989	0.00066	0.00036	0.00010	0.00000	0.00000	0.00000	0.00000
0.810	0.01266	0.00066	0.00076	0.00020	0.00000	0.00000	0.00000	0.00000
0.815	0.01667	0.00076	0.00135	0.00022	0.00000	0.00000	0.00000	0.00000
0.820	0.02160	0.00088	0.00231	0.00029	0.00000	0.00000	0.00000	0.00000
0.825	0.02699	0.00110	0.00356	0.00038	0.00025	0.00009	0.00003	0.00003
0.830	0.03423	0.00107	0.00524	0.00049	0.00066	0.00017	0.00010	0.00006
0.835	0.04256	0.00139	0.00820	0.00060	0.00134	0.00026	0.00032	0.00011
0.840	0.05343	0.00179	0.01241	0.00077	0.00273	0.00036	0.00068	0.00015
0.845	0.06680	0.00199	0.01789	0.00082	0.00482	0.00047	0.00168	0.00029
0.850	0.08375	0.00217	0.02609	0.00125	0.00836	0.00058	0.00384	0.00043
0.855	0.10342	0.00242	0.03893	0.00164	0.01438	0.00085	0.00752	0.00064
0.860	0.12905	0.00291	0.05577	0.00186	0.02462	0.00141	0.01492	0.00085
0.865	0.15844	0.00313	0.08092	0.00252	0.04367	0.00194	0.02947	0.00178
0.870	0.19680	0.00390	0.11695	0.00361	0.07595	0.00319	0.05870	0.00351
0.875	0.24389	0.00419	0.17153	0.00484	0.13483	0.00570	0.12314	0.00567

0.880	0.30004	0.00510	0.24885	0.00691	0.24575	0.00944	0.26510	0.01154
0.885	0.36384	0.00572	0.36410	0.00777	0.42483	0.01150	0.51574	0.01329
0.890	0.43775	0.00673	0.50613	0.00930	0.65350	0.01200	0.79994	0.01169
0.895	0.52534	0.00601	0.66534	0.00910	0.84499	0.00870	0.94942	0.00586
0.900	0.61617	0.00612	0.79875	0.00825	0.94778	0.00553	0.99047	0.00215
0.905	0.70535	0.00566	0.89603	0.00584	0.98695	0.00245	0.99849	0.00082
0.910	0.78596	0.00707	0.95305	0.00375	0.99642	0.00125	1.00000	0.00000
0.915	0.85378	0.00526	0.98278	0.00231	0.99936	0.00041	1.00000	0.00000
0.920	0.90704	0.00391	0.99327	0.00126	1.00000	0.00000	1.00000	0.00000
0.925	0.94111	0.00361	0.99739	0.00094	1.00000	0.00000	1.00000	0.00000
0.930	0.96650	0.00230	0.99958	0.00024	1.00000	0.00000	1.00000	0.00000
0.935	0.98252	0.00207	0.99986	0.00015	1.00000	0.00000	1.00000	0.00000
0.940	0.99162	0.00119	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000
0.945	0.99592	0.00074	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000
0.950	0.99860	0.00034	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000

Tabella B 1– Errori di Tipo I 50 run 2000 punti. Osservazioni indipendenti

B.2.2 Errori di Tipo II

ρ	$M = 10$		$M = 25$		$M = 50$		$M = 75$	
	$\hat{\beta}$	$\pm \Delta\hat{\beta}$						
0.800	0.85488	0.00311	0.93527	0.00246	0.96862	0.00161	0.98055	0.00131
0.805	0.83042	0.00341	0.91289	0.00315	0.95143	0.00242	0.96696	0.00185
0.810	0.80056	0.00362	0.88343	0.00376	0.92760	0.00300	0.94450	0.00321
0.815	0.76139	0.00409	0.84673	0.00459	0.88376	0.00477	0.89871	0.00471
0.820	0.72218	0.00447	0.79103	0.00706	0.81665	0.00539	0.82103	0.00675
0.825	0.67779	0.00430	0.71937	0.00676	0.71675	0.00812	0.69041	0.01070
0.830	0.62938	0.00553	0.63572	0.00780	0.57850	0.01016	0.49013	0.01290
0.835	0.57322	0.00554	0.52889	0.00756	0.40666	0.01064	0.25903	0.01492
0.840	0.51103	0.00576	0.42460	0.00997	0.25095	0.01074	0.11388	0.01061
0.845	0.44298	0.00660	0.30924	0.00942	0.12186	0.00903	0.03706	0.00545
0.850	0.37817	0.00658	0.21360	0.00680	0.05505	0.00542	0.01098	0.00270
0.855	0.31287	0.00619	0.13429	0.00486	0.01897	0.00307	0.00344	0.00136
0.860	0.25435	0.00527	0.08012	0.00543	0.00836	0.00192	0.00058	0.00035
0.865	0.19823	0.00509	0.04261	0.00403	0.00280	0.00114	0.00003	0.00004
0.870	0.14835	0.00417	0.02290	0.00316	0.00128	0.00070	0.00000	0.00000
0.875	0.11212	0.00372	0.01178	0.00186	0.00027	0.00028	0.00000	0.00000
0.880	0.07879	0.00371	0.00457	0.00111	0.00000	0.00000	0.00000	0.00000
0.885	0.05626	0.00257	0.00224	0.00069	0.00000	0.00000	0.00000	0.00000
0.890	0.03828	0.00219	0.00099	0.00053	0.00000	0.00000	0.00000	0.00000
0.895	0.02549	0.00206	0.00048	0.00037	0.00000	0.00000	0.00000	0.00000
0.900	0.01525	0.00156	0.00022	0.00022	0.00000	0.00000	0.00000	0.00000
0.905	0.01042	0.00116	0.00016	0.00018	0.00000	0.00000	0.00000	0.00000
0.910	0.00631	0.00089	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.915	0.00338	0.00057	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.920	0.00218	0.00049	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.925	0.00088	0.00029	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.930	0.00064	0.00028	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.935	0.00015	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.940	0.00009	0.00008	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.945	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.950	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Tabella B 2 – Errori di Tipo II 50 run 2000 punti. Osservazioni indipendenti

B.2.3 Confronto carta CUSUM

φ	Fuzzy ART 2		Fuzzy ART 1		CUSUM	
	$\hat{\beta}$	$\pm \Delta \hat{\beta}$	$\hat{\beta}$	$\pm \Delta \hat{\beta}$	$\hat{\beta}$	$\pm \Delta \hat{\beta}$
0.00	0.99732	0.00032	0.99738	0.00033	0.99731	0.00033
0.05	0.99568	0.00043	0.99738	0.00034	0.99698	0.00038
0.10	0.99217	0.00068	0.99716	0.00039	0.99631	0.00042
0.15	0.98644	0.00083	0.99653	0.00042	0.99497	0.00050
0.20	0.97657	0.00143	0.99595	0.00043	0.99322	0.00055
0.25	0.95980	0.00210	0.99491	0.00049	0.99034	0.00059
0.30	0.92503	0.00393	0.99351	0.00048	0.98629	0.00086
0.35	0.86419	0.00584	0.99096	0.00065	0.98039	0.00100
0.40	0.74030	0.00868	0.98789	0.00086	0.97145	0.00131
0.45	0.55037	0.01241	0.98295	0.00113	0.95340	0.00178
0.50	0.31378	0.01353	0.97486	0.00137	0.92474	0.00352
0.55	0.14669	0.00925	0.96121	0.00200	0.86430	0.00493
0.60	0.05372	0.00618	0.94105	0.00314	0.75758	0.00748
0.65	0.01389	0.00375	0.90344	0.00484	0.58153	0.01084
0.70	0.00458	0.00192	0.83716	0.00636	0.35161	0.01292
0.75	0.00096	0.00065	0.72628	0.01090	0.15897	0.00862
0.80	0.00019	0.00033	0.55630	0.01351	0.05410	0.00565
0.85	0.00000	0.00000	0.35448	0.01303	0.01477	0.00256
0.90	0.00000	0.00000	0.17241	0.01122	0.00394	0.00125
0.95	0.00000	0.00000	0.06978	0.00733	0.00063	0.00047
1.00	0.00000	0.00000	0.02092	0.00456	0.00019	0.00027
1.05	0.00000	0.00000	0.00546	0.00167	0.00000	0.00000
1.10	0.00000	0.00000	0.00190	0.00110	0.00000	0.00000
1.15	0.00000	0.00000	0.00028	0.00025	0.00000	0.00000
1.20	0.00000	0.00000	0.00003	0.00004	0.00000	0.00000
1.25	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.30	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.35	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.40	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.45	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.50	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Tabella B 3 – Confronto NN/SPC 50 run 2000 punti. Osservazioni indipendenti

B.2.4 Confronto carta CUSUM, pattern di tipo trend

In questa sezione sono riportati i risultati sperimentali relativi al pattern di tipo trend. Il parametro che caratterizza il pattern (s) varia nell'intervallo di valori $[0.00,0.02]$ con incrementi di 0.001 unità di deviazione standard.

La rete Fuzzy ART è stata addestrata al riconoscimento sia del target naturale del processo, sia di uno shift costante e di ampiezza 0.5 unità di deviazione standard (analogamente alle tabelle precedenti si è adottata l'etichetta "Fuzzy ART 2"). I parametri utilizzati per la rete sono $M = 75$ e $\rho = 0.8375$, con i quali è stato possibile ottenere dal metodo di controllo neurale un tasso di falsi allarmi comparabile con quello della carta di controllo CUSUM (ovvero di circa 0.27%). Anche in questo caso, per garantire che la comparazione fra prestazioni della rete neurale con quelle della carta di controllo CUSUM sia bilanciata, sono stati ignorati gli eventuali allarmi prodotti dal metodo SPC durante le prime $M-1$ osservazioni. In altri termini le prestazioni dei due sistemi di controllo sono state misurate per tutti gli indici $t \geq M$.

s	Fuzzy ART 2		CUSUM	
	$\hat{\beta}$	$\pm \Delta\hat{\beta}$	$\hat{\beta}$	$\pm \Delta\hat{\beta}$
0.000	0.99732	0.00032	0.99731	0.00033
0.001	0.98991	0.00050	0.99358	0.00030
0.002	0.98196	0.00058	0.98860	0.00043
0.003	0.97054	0.00085	0.98255	0.00063
0.004	0.95560	0.00122	0.97459	0.00078
0.005	0.93261	0.00176	0.96269	0.00098
0.006	0.90030	0.00246	0.94635	0.00123
0.007	0.84931	0.00333	0.92002	0.00195
0.008	0.77471	0.00396	0.87831	0.00273
0.009	0.66932	0.00583	0.81436	0.00449
0.010	0.53706	0.00646	0.71347	0.00516
0.011	0.38954	0.00711	0.57216	0.00640
0.012	0.24756	0.00760	0.39172	0.00724
0.013	0.13946	0.00549	0.21739	0.00724
0.014	0.07153	0.00357	0.10106	0.00461
0.015	0.03419	0.00277	0.03835	0.00305
0.016	0.01416	0.00141	0.01294	0.00148
0.017	0.00623	0.00106	0.00336	0.00068
0.018	0.00219	0.00057	0.00095	0.00035
0.019	0.00068	0.00031	0.00023	0.00014
0.020	0.00030	0.00021	0.00000	0.00000

Tabella B 4 – Confronto NN/SPC 50 run 2000 punti (pattern trend)

B.3 Dati dipendenti AR 1° ordine fattore-0.8

B.3.1 Errori di Tipo I

ρ	$M=10$		$M=25$		$M=50$		$M=75$	
	$\hat{\alpha}$	$\pm \Delta\hat{\alpha}$						
0.800	0.02040	0.00103	0.00501	0.00053	0.00124	0.00024	0.00038	0.00012
0.805	0.02272	0.00098	0.00605	0.00055	0.00156	0.00025	0.00060	0.00016
0.810	0.02509	0.00101	0.00709	0.00061	0.00197	0.00035	0.00082	0.00017
0.815	0.02759	0.00117	0.00831	0.00066	0.00260	0.00033	0.00107	0.00021
0.820	0.03008	0.00104	0.00990	0.00067	0.00311	0.00042	0.00142	0.00023
0.825	0.03359	0.00123	0.01179	0.00075	0.00384	0.00044	0.00191	0.00030
0.830	0.03710	0.00119	0.01411	0.00087	0.00510	0.00056	0.00261	0.00036
0.835	0.04140	0.00149	0.01671	0.00088	0.00641	0.00056	0.00339	0.00045
0.840	0.04679	0.00148	0.01973	0.00108	0.00822	0.00065	0.00441	0.00060
0.845	0.05216	0.00177	0.02323	0.00112	0.01055	0.00070	0.00605	0.00056
0.850	0.05859	0.00183	0.02745	0.00129	0.01360	0.00087	0.00788	0.00080
0.855	0.06566	0.00207	0.03256	0.00135	0.01742	0.00108	0.01115	0.00096
0.860	0.07394	0.00218	0.03985	0.00166	0.02260	0.00122	0.01576	0.00104
0.865	0.08421	0.00240	0.04824	0.00196	0.02942	0.00146	0.02219	0.00128
0.870	0.09497	0.00247	0.05915	0.00193	0.04009	0.00185	0.03196	0.00210
0.875	0.10909	0.00268	0.07223	0.00238	0.05376	0.00245	0.04664	0.00269
0.880	0.12577	0.00317	0.09099	0.00330	0.07571	0.00319	0.06951	0.00325
0.885	0.14507	0.00359	0.11496	0.00342	0.10653	0.00457	0.10592	0.00446
0.890	0.17030	0.00317	0.14673	0.00438	0.15455	0.00537	0.16951	0.00678
0.895	0.19662	0.00445	0.19143	0.00557	0.23269	0.00731	0.27374	0.00756
0.900	0.23472	0.00477	0.25121	0.00677	0.32711	0.00982	0.42063	0.01113

0.905	0.27854	0.00548	0.32798	0.00759	0.46602	0.00961	0.61031	0.01318
0.910	0.33206	0.00583	0.43269	0.00883	0.63321	0.01216	0.80765	0.01110
0.915	0.39250	0.00553	0.54735	0.00848	0.78319	0.01101	0.91983	0.00933
0.920	0.46325	0.00607	0.68049	0.00788	0.90370	0.00728	0.97444	0.00396
0.925	0.54914	0.00681	0.79525	0.00725	0.96192	0.00492	0.99331	0.00232
0.930	0.63775	0.00627	0.88797	0.00687	0.98756	0.00266	0.99828	0.00107
0.935	0.72636	0.00555	0.94755	0.00418	0.99721	0.00104	0.99987	0.00026
0.940	0.81630	0.00446	0.97885	0.00270	0.99912	0.00075	1.00000	0.00000
0.945	0.88617	0.00423	0.99312	0.00124	1.00000	0.00000	1.00000	0.00000
0.950	0.93675	0.00326	0.99809	0.00051	1.00000	0.00000	1.00000	0.00000

Tabella B 5 – Errori di Tipo I 50 run 2000 punti. Osservazioni AR(-0.8)

B.3.2 Errori di Tipo II

ρ	$M = 10$		$M = 25$		$M = 50$		$M = 75$	
	$\hat{\beta}$	$\pm \Delta\hat{\beta}$						
0.800	0.91001	0.00223	0.96081	0.00122	0.98011	0.00097	0.98781	0.00087
0.805	0.88431	0.00272	0.94471	0.00206	0.96870	0.00150	0.97786	0.00142
0.810	0.84898	0.00346	0.91789	0.00228	0.94490	0.00258	0.95657	0.00245
0.815	0.80287	0.00409	0.87324	0.00380	0.89896	0.00419	0.90601	0.00443
0.820	0.73646	0.00421	0.79701	0.00551	0.80358	0.00682	0.79512	0.00761
0.825	0.65609	0.00499	0.68186	0.00666	0.62846	0.01139	0.54722	0.01331
0.830	0.56022	0.00443	0.51587	0.00740	0.36739	0.01264	0.21412	0.01282
0.835	0.44401	0.00552	0.31433	0.00759	0.13041	0.00919	0.03985	0.00570
0.840	0.33029	0.00563	0.14875	0.00762	0.02720	0.00342	0.00539	0.00185
0.845	0.22533	0.00556	0.06105	0.00423	0.00655	0.00137	0.00021	0.00020
0.850	0.13623	0.00406	0.02079	0.00238	0.00040	0.00026	0.00000	0.00000
0.855	0.07858	0.00315	0.00462	0.00098	0.00000	0.00000	0.00000	0.00000
0.860	0.04177	0.00244	0.00097	0.00038	0.00000	0.00000	0.00000	0.00000
0.865	0.02031	0.00138	0.00023	0.00018	0.00000	0.00000	0.00000	0.00000
0.870	0.00905	0.00097	0.00007	0.00012	0.00000	0.00000	0.00000	0.00000
0.875	0.00353	0.00056	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.880	0.00144	0.00035	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.885	0.00069	0.00024	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.890	0.00017	0.00009	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.895	0.00006	0.00009	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.900	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.905	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.910	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.915	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.920	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.925	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.930	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.935	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.940	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.945	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.950	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Tabella B 6 – Errori di Tipo II 50 run 2000 punti. Osservazioni AR(-0.8)

B.3.3 Confronto carta CUSUM

φ	Fuzzy ART 2		Fuzzy ART 1		CUSUM	
	$\hat{\beta}$	$\pm \Delta \hat{\beta}$	$\hat{\beta}$	$\pm \Delta \hat{\beta}$	$\hat{\beta}$	$\pm \Delta \hat{\beta}$
0.00	0.99737	0.00035	0.99739	0.00036	0.99731	0.00033
0.05	0.99717	0.00040	0.99739	0.00036	0.99497	0.00050
0.10	0.99617	0.00046	0.99736	0.00035	0.98629	0.00086
0.15	0.99253	0.00066	0.99730	0.00039	0.95340	0.00180
0.20	0.98171	0.00125	0.99718	0.00039	0.75769	0.00725
0.25	0.94689	0.00297	0.99696	0.00041	0.15897	0.00891
0.30	0.81664	0.00735	0.99676	0.00040	0.00394	0.00125
0.35	0.45094	0.01249	0.99649	0.00045	0.00001	0.00002
0.40	0.09506	0.00699	0.99605	0.00048	0.00000	0.00000
0.45	0.01054	0.00283	0.99550	0.00052	0.00000	0.00000
0.50	0.00080	0.00060	0.99466	0.00052	0.00000	0.00000
0.55	0.00006	0.00010	0.99346	0.00065	0.00000	0.00000
0.60	0.00000	0.00000	0.99211	0.00070	0.00000	0.00000
0.65	0.00000	0.00000	0.98971	0.00075	0.00000	0.00000
0.70	0.00000	0.00000	0.98540	0.00087	0.00000	0.00000
0.75	0.00000	0.00000	0.97821	0.00127	0.00000	0.00000
0.80	0.00000	0.00000	0.96564	0.00231	0.00000	0.00000
0.85	0.00000	0.00000	0.93451	0.00335	0.00000	0.00000
0.90	0.00000	0.00000	0.84986	0.00643	0.00000	0.00000
0.95	0.00000	0.00000	0.64118	0.01006	0.00000	0.00000
1.00	0.00000	0.00000	0.21412	0.01282	0.00000	0.00000
1.05	0.00000	0.00000	0.01749	0.00320	0.00000	0.00000
1.10	0.00000	0.00000	0.00059	0.00060	0.00000	0.00000
1.15	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.20	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.25	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.30	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.35	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.40	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.45	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.50	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Tabella B 7 – Confronto NN/SPC 50 run 2000 punti. Osservazioni AR(-0.8)

B.4 Dati dipendenti AR 1° ordine fattore-0.2

B.4.1 Errori di Tipo I

ρ	$M=10$		$M=25$		$M=50$		$M=75$	
	$\hat{\alpha}$	$\pm \Delta \hat{\alpha}$						
0.800	0.00849	0.00056	0.00031	0.00009	0.00000	0.00000	0.00000	0.00000
0.805	0.01073	0.00068	0.00056	0.00014	0.00000	0.00000	0.00000	0.00000
0.810	0.01380	0.00074	0.00094	0.00020	0.00002	0.00003	0.00000	0.00000
0.815	0.01755	0.00072	0.00168	0.00026	0.00006	0.00005	0.00000	0.00000
0.820	0.02207	0.00085	0.00249	0.00030	0.00016	0.00007	0.00001	0.00002
0.825	0.02721	0.00089	0.00393	0.00034	0.00041	0.00013	0.00007	0.00005
0.830	0.03422	0.00112	0.00608	0.00047	0.00078	0.00015	0.00014	0.00007
0.835	0.04273	0.00145	0.00885	0.00065	0.00160	0.00025	0.00040	0.00011
0.840	0.05296	0.00179	0.01264	0.00076	0.00294	0.00038	0.00086	0.00017

0.845	0.06494	0.00207	0.01874	0.00092	0.00496	0.00047	0.00178	0.00028
0.850	0.08158	0.00244	0.02649	0.00123	0.00843	0.00057	0.00402	0.00040
0.855	0.10056	0.00234	0.03778	0.00146	0.01505	0.00089	0.00803	0.00060
0.860	0.12449	0.00237	0.05499	0.00210	0.02576	0.00145	0.01542	0.00092
0.865	0.15395	0.00332	0.07840	0.00247	0.04330	0.00191	0.02974	0.00163
0.870	0.18957	0.00327	0.11270	0.00361	0.07336	0.00296	0.05589	0.00303
0.875	0.23602	0.00403	0.16407	0.00454	0.12900	0.00558	0.11599	0.00578
0.880	0.28837	0.00457	0.24021	0.00731	0.23215	0.00895	0.24780	0.01079
0.885	0.34955	0.00596	0.34785	0.00778	0.40044	0.01024	0.48428	0.01403
0.890	0.42197	0.00668	0.48315	0.01000	0.62233	0.01286	0.75888	0.01408
0.895	0.50771	0.00609	0.63269	0.00982	0.82618	0.00836	0.92874	0.00780
0.900	0.59700	0.00683	0.77156	0.00929	0.93464	0.00573	0.98690	0.00273
0.905	0.68215	0.00654	0.87933	0.00646	0.98151	0.00322	0.99829	0.00082
0.910	0.76970	0.00578	0.94254	0.00429	0.99468	0.00150	0.99987	0.00015
0.915	0.83985	0.00613	0.97434	0.00328	0.99936	0.00047	1.00000	0.00000
0.920	0.89402	0.00452	0.99150	0.00130	1.00000	0.00000	1.00000	0.00000
0.925	0.93655	0.00307	0.99705	0.00087	1.00000	0.00000	1.00000	0.00000
0.930	0.96388	0.00245	0.99941	0.00031	1.00000	0.00000	1.00000	0.00000
0.935	0.98012	0.00182	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000
0.940	0.99001	0.00130	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000
0.945	0.99557	0.00068	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000
0.950	0.99817	0.00044	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000

Tabella B 8 – Errori di Tipo I 50 run 2000 punti. Osservazioni AR(-0.2)

B.4.2 Errori di Tipo II

ρ	$M=10$		$M=25$		$M=50$		$M=75$	
	$\hat{\beta}$	$\pm \Delta\hat{\beta}$						
0.800	0.84982	0.00344	0.93507	0.00223	0.96982	0.00162	0.98191	0.00119
0.805	0.82044	0.00395	0.90903	0.00324	0.95166	0.00241	0.96652	0.00187
0.810	0.78520	0.00427	0.87507	0.00393	0.92048	0.00371	0.93868	0.00306
0.815	0.74201	0.00483	0.82881	0.00517	0.86797	0.00522	0.88184	0.00444
0.820	0.69145	0.00447	0.75898	0.00560	0.78249	0.00656	0.77272	0.01023
0.825	0.64001	0.00428	0.66975	0.00697	0.64613	0.00815	0.59338	0.01282
0.830	0.57972	0.00587	0.56033	0.00782	0.47154	0.01373	0.33856	0.01342
0.835	0.51341	0.00623	0.44111	0.00929	0.27971	0.01089	0.13870	0.01014
0.840	0.43794	0.00710	0.31313	0.00870	0.13462	0.00868	0.04746	0.00614
0.845	0.36811	0.00680	0.20622	0.00748	0.05149	0.00506	0.01140	0.00293
0.850	0.29761	0.00594	0.12560	0.00722	0.01981	0.00377	0.00339	0.00156
0.855	0.23467	0.00566	0.06999	0.00488	0.00783	0.00214	0.00044	0.00034
0.860	0.17854	0.00427	0.03713	0.00381	0.00241	0.00127	0.00018	0.00027
0.865	0.13189	0.00421	0.01679	0.00231	0.00054	0.00035	0.00000	0.00000
0.870	0.09477	0.00406	0.00774	0.00181	0.00018	0.00014	0.00000	0.00000
0.875	0.06371	0.00274	0.00332	0.00097	0.00005	0.00006	0.00000	0.00000
0.880	0.04304	0.00230	0.00224	0.00082	0.00000	0.00000	0.00000	0.00000
0.885	0.02757	0.00186	0.00099	0.00055	0.00000	0.00000	0.00000	0.00000
0.890	0.01712	0.00157	0.00048	0.00037	0.00000	0.00000	0.00000	0.00000
0.895	0.01102	0.00130	0.00017	0.00021	0.00000	0.00000	0.00000	0.00000
0.900	0.00702	0.00104	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.905	0.00405	0.00067	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.910	0.00205	0.00051	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.915	0.00119	0.00038	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.920	0.00044	0.00014	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.925	0.00015	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.930	0.00013	0.00011	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

0.935	0.00004	0.00005	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.940	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.945	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.950	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Tabella B 9 – Errori di Tipo II 50 run 2000 punti. Osservazioni AR(-0.2)

B.4.3 Confronto carta CUSUM

φ	Fuzzy ART 2		Fuzzy ART 1		CUSUM	
	$\hat{\beta}$	$\pm \Delta \hat{\beta}$	$\hat{\beta}$	$\pm \Delta \hat{\beta}$	$\hat{\beta}$	$\pm \Delta \hat{\beta}$
0.00	0.99708	0.00031	0.99745	0.00033	0.99764	0.00033
0.05	0.99554	0.00040	0.99735	0.00033	0.99727	0.00037
0.10	0.99240	0.00048	0.99721	0.00036	0.99615	0.00043
0.15	0.98618	0.00092	0.99685	0.00039	0.99461	0.00050
0.20	0.97568	0.00139	0.99621	0.00040	0.99137	0.00062
0.25	0.95551	0.00256	0.99518	0.00042	0.98631	0.00076
0.30	0.91216	0.00437	0.99378	0.00060	0.97895	0.00107
0.35	0.82246	0.00840	0.99150	0.00061	0.96393	0.00164
0.40	0.65027	0.01172	0.98833	0.00084	0.93510	0.00306
0.45	0.39717	0.01191	0.98423	0.00114	0.86628	0.00541
0.50	0.17025	0.01316	0.97644	0.00140	0.73151	0.00764
0.55	0.05182	0.00651	0.96355	0.00227	0.48779	0.01012
0.60	0.01081	0.00266	0.94393	0.00281	0.21867	0.00997
0.65	0.00272	0.00114	0.90343	0.00399	0.06342	0.00564
0.70	0.00044	0.00033	0.82923	0.00808	0.01406	0.00233
0.75	0.00020	0.00022	0.69668	0.01175	0.00273	0.00083
0.80	0.00000	0.00000	0.49063	0.01426	0.00051	0.00027
0.85	0.00000	0.00000	0.24864	0.01373	0.00000	0.00000
0.90	0.00000	0.00000	0.09582	0.00788	0.00000	0.00000
0.95	0.00000	0.00000	0.02841	0.00473	0.00000	0.00000
1.00	0.00000	0.00000	0.00720	0.00276	0.00000	0.00000
1.05	0.00000	0.00000	0.00184	0.00125	0.00000	0.00000
1.10	0.00000	0.00000	0.00031	0.00027	0.00000	0.00000
1.15	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.20	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.25	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.30	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.35	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.40	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.45	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.50	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Tabella B 10 – Confronto NN/SPC 50 run 2000 punti. Osservazioni AR(-0.2)

B.5 Dati dipendenti AR 1° ordine fattore+0.2

B.5.1 Errori di Tipo I

ρ	$M=10$		$M=25$		$M=50$		$M=75$	
	$\hat{\alpha}$	$\pm \Delta \hat{\alpha}$						
0.800	0.00867	0.00054	0.00036	0.00012	0.00000	0.00000	0.00000	0.00000
0.805	0.01119	0.00065	0.00061	0.00016	0.00000	0.00000	0.00000	0.00000
0.810	0.01392	0.00079	0.00108	0.00020	0.00002	0.00003	0.00000	0.00000
0.815	0.01796	0.00092	0.00168	0.00025	0.00008	0.00005	0.00000	0.00000
0.820	0.02277	0.00096	0.00266	0.00030	0.00021	0.00009	0.00002	0.00003
0.825	0.02868	0.00091	0.00401	0.00040	0.00036	0.00011	0.00004	0.00004
0.830	0.03517	0.00117	0.00582	0.00051	0.00079	0.00018	0.00014	0.00007
0.835	0.04325	0.00148	0.00872	0.00061	0.00166	0.00025	0.00037	0.00011
0.840	0.05374	0.00183	0.01286	0.00081	0.00306	0.00036	0.00092	0.00018
0.845	0.06610	0.00203	0.01867	0.00088	0.00524	0.00044	0.00201	0.00031
0.850	0.08219	0.00237	0.02713	0.00126	0.00871	0.00064	0.00420	0.00044
0.855	0.10105	0.00273	0.03808	0.00156	0.01486	0.00104	0.00801	0.00073
0.860	0.12550	0.00291	0.05387	0.00183	0.02469	0.00149	0.01552	0.00111
0.865	0.15500	0.00349	0.07793	0.00249	0.04290	0.00217	0.02855	0.00162
0.870	0.18974	0.00370	0.11474	0.00352	0.07357	0.00365	0.05620	0.00328
0.875	0.23385	0.00471	0.16534	0.00515	0.12794	0.00547	0.11545	0.00593
0.880	0.28639	0.00491	0.23945	0.00647	0.23485	0.00874	0.24390	0.00943
0.885	0.35010	0.00487	0.34383	0.00787	0.39704	0.01194	0.48591	0.01467
0.890	0.42754	0.00542	0.47952	0.00855	0.62054	0.01208	0.76459	0.01272
0.895	0.51029	0.00671	0.63122	0.00800	0.81885	0.00893	0.93010	0.00642
0.900	0.59545	0.00610	0.76973	0.00786	0.93512	0.00635	0.98805	0.00258
0.905	0.68446	0.00568	0.87759	0.00693	0.98322	0.00267	0.99829	0.00086
0.910	0.76865	0.00539	0.94034	0.00405	0.99591	0.00129	0.99987	0.00022
0.915	0.83681	0.00447	0.97751	0.00224	0.99934	0.00043	1.00000	0.00000
0.920	0.89248	0.00421	0.99079	0.00147	1.00000	0.00000	1.00000	0.00000
0.925	0.93434	0.00307	0.99740	0.00086	1.00000	0.00000	1.00000	0.00000
0.930	0.96296	0.00225	0.99925	0.00034	1.00000	0.00000	1.00000	0.00000
0.935	0.98096	0.00170	0.99976	0.00020	1.00000	0.00000	1.00000	0.00000
0.940	0.99013	0.00122	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000
0.945	0.99532	0.00082	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000
0.950	0.99830	0.00047	1.00000	0.00000	1.00000	0.00000	1.00000	0.00000

Tabella B 11 – Errori di Tipo I 50 run 2000 punti. Osservazioni AR(+0.2)

B.5.2 Errori di Tipo II

ρ	$M=10$		$M=25$		$M=50$		$M=75$	
	$\hat{\beta}$	$\pm \Delta \hat{\beta}$						
0.800	0.86382	0.00308	0.93757	0.00245	0.96814	0.00163	0.97982	0.00137
0.805	0.84426	0.00377	0.91942	0.00293	0.95432	0.00228	0.96732	0.00185
0.810	0.82054	0.00411	0.89586	0.00365	0.93326	0.00340	0.95014	0.00294
0.815	0.79223	0.00355	0.86518	0.00446	0.90400	0.00451	0.91730	0.00391
0.820	0.75850	0.00446	0.82828	0.00501	0.85418	0.00492	0.86376	0.00577
0.825	0.72272	0.00413	0.77314	0.00656	0.78307	0.00720	0.76862	0.00865
0.830	0.68228	0.00520	0.70850	0.00636	0.68340	0.00992	0.63865	0.01380
0.835	0.63772	0.00557	0.62850	0.00764	0.55822	0.01109	0.45115	0.01635
0.840	0.59049	0.00595	0.53817	0.00921	0.41045	0.01145	0.25594	0.01792
0.845	0.53500	0.00597	0.44023	0.01150	0.26030	0.01277	0.12494	0.01078

0.850	0.47462	0.00602	0.33878	0.00899	0.14164	0.01039	0.05147	0.00720
0.855	0.41420	0.00703	0.24525	0.00832	0.06721	0.00659	0.01421	0.00287
0.860	0.35332	0.00646	0.16615	0.00836	0.03031	0.00419	0.00423	0.00142
0.865	0.29365	0.00572	0.10555	0.00566	0.01289	0.00254	0.00109	0.00053
0.870	0.23710	0.00552	0.06131	0.00443	0.00536	0.00160	0.00016	0.00020
0.875	0.18811	0.00493	0.03318	0.00313	0.00144	0.00065	0.00000	0.00000
0.880	0.14312	0.00447	0.01862	0.00220	0.00078	0.00048	0.00000	0.00000
0.885	0.10777	0.00407	0.00942	0.00167	0.00009	0.00011	0.00000	0.00000
0.890	0.07571	0.00311	0.00385	0.00102	0.00000	0.00000	0.00000	0.00000
0.895	0.05395	0.00240	0.00183	0.00077	0.00000	0.00000	0.00000	0.00000
0.900	0.03749	0.00217	0.00099	0.00054	0.00000	0.00000	0.00000	0.00000
0.905	0.02430	0.00194	0.00048	0.00039	0.00000	0.00000	0.00000	0.00000
0.910	0.01572	0.00145	0.00023	0.00019	0.00000	0.00000	0.00000	0.00000
0.915	0.00984	0.00106	0.00016	0.00017	0.00000	0.00000	0.00000	0.00000
0.920	0.00591	0.00088	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.925	0.00361	0.00066	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.930	0.00183	0.00047	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.935	0.00088	0.00030	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.940	0.00057	0.00025	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.945	0.00022	0.00013	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
0.950	0.00010	0.00010	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Tabella B 12 – Errori di Tipo II 50 run 2000 punti. Osservazioni AR(+0.2)

B.5.3 Confronto carta CUSUM

φ	Fuzzy ART 2		Fuzzy ART 1		CUSUM	
	$\hat{\beta}$	$\pm \Delta\hat{\beta}$	$\hat{\beta}$	$\pm \Delta\hat{\beta}$	$\hat{\beta}$	$\pm \Delta\hat{\beta}$
0.00	0.99674	0.00039	0.99736	0.00036	0.99731	0.00033
0.05	0.99449	0.00055	0.99717	0.00038	0.99706	0.00034
0.10	0.99095	0.00076	0.99690	0.00033	0.99653	0.00036
0.15	0.98534	0.00107	0.99639	0.00043	0.99572	0.00043
0.20	0.97568	0.00147	0.99582	0.00048	0.99458	0.00050
0.25	0.96088	0.00220	0.99476	0.00047	0.99316	0.00053
0.30	0.93499	0.00332	0.99353	0.00061	0.99075	0.00062
0.35	0.88720	0.00526	0.99128	0.00075	0.98757	0.00073
0.40	0.81091	0.00753	0.98792	0.00100	0.98345	0.00104
0.45	0.67467	0.01056	0.98315	0.00131	0.97788	0.00107
0.50	0.49709	0.01108	0.97648	0.00125	0.96874	0.00130
0.55	0.29539	0.01209	0.96555	0.00207	0.95386	0.00177
0.60	0.14914	0.00986	0.94931	0.00301	0.93181	0.00316
0.65	0.05956	0.00677	0.91969	0.00463	0.89304	0.00451
0.70	0.02291	0.00476	0.87335	0.00514	0.82672	0.00586
0.75	0.00789	0.00265	0.80425	0.00701	0.72227	0.00859
0.80	0.00184	0.00106	0.68522	0.01159	0.56522	0.01174
0.85	0.00184	0.00110	0.53674	0.01265	0.37842	0.01306
0.90	0.00000	0.00000	0.35190	0.01819	0.20821	0.01056
0.95	0.00000	0.00000	0.18993	0.01360	0.09026	0.00680
1.00	0.00000	0.00000	0.09037	0.00984	0.03616	0.00416
1.05	0.00000	0.00000	0.03332	0.00537	0.01173	0.00232
1.10	0.00000	0.00000	0.01191	0.00254	0.00394	0.00123
1.15	0.00000	0.00000	0.00269	0.00115	0.00099	0.00066
1.20	0.00000	0.00000	0.00093	0.00045	0.00024	0.00027
1.25	0.00000	0.00000	0.00028	0.00028	0.00000	0.00000
1.30	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.35	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

1.40	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.45	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000
1.50	0.00000	0.00000	0.00000	0.00000	0.00000	0.00000

Tabella B 13 – Confronto NN/SPC 50 run 2000 punti. Osservazioni AR(+0.2)

N.B. Si osservi che in quest'ultimo caso non è stato possibile ottenere dalla rete Fuzzy ART 2 un tasso di allarmi statisticamente equivalente a quello del benchmark SPC. Nella tabella sono stati riportati i risultati ottenuti dalla rete Fuzzy ART 2 con parametro di vigilanza tarato in modo da ottenere il minimo valore possibile di falsi allarmi.

B.6 Dati dipendenti AR 1° ordine fattore+0.8

B.6.1 Errori di Tipo I

ρ	$M = 10$		$M = 25$		$M = 50$		$M = 75$	
	$\hat{\alpha}$	$\pm \Delta \hat{\alpha}$						
0.800	0.02047	0.00091	0.00574	0.00054	0.00135	0.00021	0.00049	0.00014
0.805	0.02267	0.00102	0.00676	0.00053	0.00177	0.00025	0.00066	0.00016
0.810	0.02489	0.00097	0.00774	0.00063	0.00221	0.00029	0.00082	0.00021
0.815	0.02733	0.00116	0.00898	0.00070	0.00272	0.00038	0.00109	0.00022
0.820	0.03047	0.00126	0.01044	0.00069	0.00358	0.00047	0.00144	0.00029
0.825	0.03429	0.00111	0.01227	0.00078	0.00445	0.00053	0.00190	0.00033
0.830	0.03802	0.00128	0.01432	0.00088	0.00550	0.00057	0.00273	0.00042
0.835	0.04215	0.00151	0.01663	0.00086	0.00691	0.00056	0.00359	0.00049
0.840	0.04705	0.00167	0.01953	0.00105	0.00853	0.00080	0.00476	0.00055
0.845	0.05252	0.00166	0.02386	0.00109	0.01090	0.00079	0.00638	0.00070
0.850	0.05878	0.00191	0.02820	0.00122	0.01371	0.00093	0.00873	0.00076
0.855	0.06607	0.00177	0.03337	0.00140	0.01728	0.00126	0.01157	0.00089
0.860	0.07443	0.00228	0.03993	0.00150	0.02289	0.00118	0.01548	0.00103
0.865	0.08466	0.00223	0.04937	0.00183	0.03050	0.00139	0.02231	0.00124
0.870	0.09629	0.00240	0.06023	0.00204	0.04086	0.00194	0.03219	0.00201
0.875	0.10930	0.00285	0.07473	0.00244	0.05613	0.00234	0.04681	0.00228
0.880	0.12702	0.00282	0.09403	0.00294	0.07712	0.00303	0.07112	0.00404
0.885	0.14685	0.00321	0.11826	0.00348	0.10999	0.00395	0.10920	0.00466
0.890	0.17063	0.00387	0.15093	0.00443	0.15797	0.00544	0.17163	0.00730
0.895	0.19924	0.00362	0.19400	0.00558	0.22735	0.00788	0.27992	0.01064
0.900	0.23244	0.00467	0.25516	0.00636	0.32887	0.00983	0.42770	0.01491
0.905	0.27553	0.00477	0.32676	0.00624	0.47533	0.00944	0.61856	0.01595
0.910	0.32706	0.00536	0.42692	0.00597	0.63310	0.01153	0.80817	0.01262
0.915	0.38897	0.00522	0.54109	0.00723	0.78686	0.01076	0.92824	0.00796
0.920	0.46445	0.00602	0.66790	0.00831	0.89982	0.00763	0.97755	0.00460
0.925	0.54756	0.00660	0.79157	0.00693	0.96301	0.00406	0.99422	0.00194
0.930	0.64035	0.00511	0.88564	0.00604	0.98929	0.00214	0.99878	0.00099
0.935	0.73410	0.00660	0.94698	0.00499	0.99840	0.00068	0.99996	0.00006
0.940	0.81955	0.00466	0.97986	0.00267	0.99988	0.00014	1.00000	0.00000
0.945	0.89109	0.00395	0.99382	0.00146	1.00000	0.00000	1.00000	0.00000
0.950	0.94066	0.00338	0.99862	0.00053	1.00000	0.00000	1.00000	0.00000

Tabella B 14 – Errori di Tipo I 50 run 2000 punti. Osservazioni AR(+0.8)

B.6.2 Errori di Tipo II

ρ	$M = 10$		$M = 25$		$M = 50$		$M = 75$	
	$\hat{\beta}$	$\pm \Delta\hat{\beta}$						
0.800	0.91810	0.00228	0.95446	0.00184	0.97202	0.00150	0.97997	0.00129
0.805	0.91191	0.00241	0.94840	0.00192	0.96768	0.00163	0.97633	0.00146
0.810	0.90503	0.00243	0.94186	0.00260	0.96132	0.00220	0.97004	0.00173
0.815	0.89834	0.00277	0.93547	0.00264	0.95387	0.00248	0.96373	0.00226
0.820	0.89014	0.00297	0.92742	0.00274	0.94497	0.00256	0.95440	0.00266
0.825	0.88053	0.00335	0.91687	0.00300	0.93482	0.00304	0.94325	0.00358
0.830	0.87143	0.00299	0.90549	0.00304	0.92208	0.00390	0.92873	0.00461
0.835	0.85903	0.00320	0.89228	0.00340	0.90604	0.00438	0.90891	0.00524
0.840	0.84660	0.00369	0.87750	0.00411	0.88376	0.00474	0.88382	0.00575
0.845	0.83246	0.00454	0.86085	0.00502	0.85867	0.00540	0.84644	0.00770
0.850	0.81680	0.00420	0.83771	0.00558	0.82462	0.00754	0.80570	0.00805
0.855	0.79803	0.00435	0.81410	0.00509	0.78378	0.00832	0.74417	0.01004
0.860	0.77643	0.00535	0.78400	0.00680	0.73528	0.00910	0.67132	0.01221
0.865	0.75176	0.00534	0.74816	0.00760	0.67784	0.00908	0.58212	0.01314
0.870	0.72276	0.00623	0.70532	0.00730	0.60161	0.01268	0.46722	0.01446
0.875	0.69420	0.00605	0.65516	0.00918	0.51957	0.01182	0.35508	0.01470
0.880	0.66203	0.00692	0.59634	0.00945	0.42607	0.01144	0.25129	0.01350
0.885	0.62112	0.00699	0.52886	0.00895	0.32303	0.01309	0.16857	0.01235
0.890	0.57525	0.00728	0.46010	0.00947	0.23032	0.01114	0.09258	0.00869
0.895	0.52625	0.00687	0.37858	0.01015	0.15378	0.01165	0.05064	0.00698
0.900	0.47128	0.00691	0.29672	0.01005	0.08783	0.00734	0.02230	0.00366
0.905	0.41081	0.00738	0.22714	0.00938	0.04755	0.00533	0.00962	0.00292
0.910	0.34961	0.00704	0.16204	0.00714	0.02451	0.00360	0.00256	0.00113
0.915	0.28830	0.00656	0.10495	0.00774	0.01014	0.00203	0.00074	0.00057
0.920	0.22564	0.00720	0.06218	0.00516	0.00416	0.00129	0.00016	0.00023
0.925	0.17064	0.00542	0.03121	0.00331	0.00117	0.00071	0.00000	0.00000
0.930	0.12213	0.00509	0.01541	0.00209	0.00043	0.00047	0.00000	0.00000
0.935	0.07803	0.00330	0.00717	0.00129	0.00000	0.00000	0.00000	0.00000
0.940	0.04809	0.00258	0.00240	0.00101	0.00000	0.00000	0.00000	0.00000
0.945	0.02640	0.00202	0.00095	0.00052	0.00000	0.00000	0.00000	0.00000
0.950	0.01417	0.00152	0.00018	0.00019	0.00000	0.00000	0.00000	0.00000

Tabella B 15 – Errori di Tipo II 50 run 2000 punti. Osservazioni AR(+0.8)

B.6.3 Confronto carta CUSUM

φ	Fuzzy ART 2		Fuzzy ART 1		CUSUM	
	$\hat{\beta}$	$\pm \Delta\hat{\beta}$	$\hat{\beta}$	$\pm \Delta\hat{\beta}$	$\hat{\beta}$	$\pm \Delta\hat{\beta}$
0.00	0.98966	0.00072	0.99727	0.00042	0.99731	0.00033
0.05	0.98714	0.00077	0.99717	0.00041	0.99731	0.00035
0.10	0.98402	0.00100	0.99704	0.00036	0.99711	0.00030
0.15	0.98045	0.00106	0.99684	0.00049	0.99698	0.00038
0.20	0.97606	0.00143	0.99658	0.00047	0.99678	0.00035
0.25	0.96886	0.00180	0.99616	0.00048	0.99652	0.00037
0.30	0.96000	0.00238	0.99559	0.00054	0.99631	0.00042
0.35	0.94932	0.00297	0.99490	0.00057	0.99596	0.00037
0.40	0.93599	0.00367	0.99389	0.00059	0.99545	0.00044
0.45	0.91780	0.00488	0.99299	0.00063	0.99497	0.00050
0.50	0.89641	0.00542	0.99128	0.00073	0.99451	0.00050
0.55	0.86293	0.00654	0.98964	0.00085	0.99411	0.00050
0.60	0.82902	0.00691	0.98739	0.00099	0.99322	0.00055

0.65	0.78074	0.00920	0.98494	0.00118	0.99244	0.00049
0.70	0.72127	0.01068	0.98159	0.00114	0.99151	0.00051
0.75	0.65301	0.01345	0.97793	0.00145	0.99034	0.00059
0.80	0.57308	0.01441	0.97201	0.00154	0.98902	0.00068
0.85	0.49037	0.01508	0.96430	0.00184	0.98766	0.00070
0.90	0.39133	0.01658	0.95485	0.00270	0.98629	0.00087
0.95	0.30190	0.01469	0.94344	0.00356	0.98403	0.00103
1.00	0.21352	0.01434	0.92873	0.00461	0.98252	0.00099
1.05	0.15401	0.01175	0.90875	0.00520	0.98037	0.00101
1.10	0.10809	0.01226	0.88279	0.00606	0.97792	0.00105
1.15	0.08031	0.01059	0.85189	0.00527	0.97514	0.00129
1.20	0.04846	0.00793	0.80878	0.00847	0.97109	0.00139
1.25	0.03221	0.00663	0.75559	0.00941	0.96609	0.00165
1.30	0.02019	0.00482	0.68983	0.01201	0.95988	0.00159
1.35	0.01231	0.00343	0.61820	0.01297	0.95266	0.00177
1.40	0.00855	0.00303	0.53706	0.01419	0.94402	0.00214
1.45	0.00478	0.00191	0.43334	0.01522	0.93358	0.00308
1.50	0.00192	0.00123	0.34922	0.01538	0.92169	0.00375

Tabella B 16 – Confronto NN/SPC 50 run 2000 punti. Osservazioni AR(+0.8)

N.B. Si osservi che in questo caso non è possibile ottenere dalla rete Fuzzy ART 2 un tasso di allarmi statisticamente equivalente a quello del benchmark SPC. Nella tabella sono stati riportati i risultati ottenuti dalla rete Fuzzy ART 2 con parametro di vigilanza tarato in modo da ottenere il minimo valore possibile di falsi allarmi.

Appendice C

Addestramento sull'uscita del processo

C.1 Lo scenario della fase di addestramento

In questa sezione è presentato uno studio, basato sulla simulazione, per analizzare il comportamento della rete neurale Fuzzy ART nell'ipotesi che la fase di addestramento sia realizzata utilizzando una serie di dati direttamente prodotti dal processo. In altri termini, lo scenario di addestramento adottato è quello proposto da Al-Ghanim (1997) che prevede un training realizzato direttamente sui dati prodotti dal processo in una condizione naturale di variabilità. Durante tale intervallo si suppone che il processo produca unicamente uscite naturali che sono apprese dalla rete neurale ART. La procedura di addestramento della rete ha termine quando l'algoritmo d'apprendimento ha prodotto tutti i possibili cluster naturali, per l'insieme di dati d'addestramento, fissato un determinato valore del parametro di vigilanza (Pacella *et al.* 2002b).

A differenza, tuttavia, di quanto indicato nel lavoro di Al-Ghanim (1997), anche per questo scenario di addestramento si è deciso di adottare in fase di training il massimo parametro di vigilanza ($\rho = 1$). La scelta del massimo parametro di vigilanza durante la fase di addestramento (*Perfectly Learned Patterns – PLP*) implica che la rete neurale apprenderà un numero di categorie corrispondente al numero di differenti pattern presentati durante il training. Ognuna di tali categorie memorizza uno specifico tipo di pattern utilizzato nell'addestramento.

La prima conseguenza dell'approccio di addestramento PLP è che gli unici parametri della rete neurale che possono influenzare le prestazioni del sistema sono: il parametro di vigilanza adottato in fase operativa (ρ), e la dimensione della ampiezza di finestra adottata (M). Il numero di neuroni nello stato $F2$, invece, dovrà essere maggiore o al più uguale al numero di differenti pattern utilizzati durante la fase di addestramento.

Tuttavia, a differenza del caso presentato nei precedenti capitoli, oltre ai parametri ρ e M , altri fattori possono influenzare le prestazioni della rete neurale nello scenario di addestramento descritto. Le prestazioni, infatti, dipenderanno anche dall'insieme di addestramento, e dalla sua dimensione (la lunghezza del periodo di training prescelta).

L'obiettivo dello studio presentato in questa sezione è quello di porre in evidenza l'influenza dei differenti fattori sulle prestazioni della rete neurale misurate in termini di errore di Tipo I, e di Tipo II per uno specifico pattern non naturale.

C.2 Il modello di riferimento

Consideriamo una rete neurale di tipo Fuzzy ART addestrata con approccio PLP su un insieme di addestramento che indichiamo con L (Learning). Supponiamo di

verificare le prestazioni della rete su un insieme di prova che indichiamo con T (Testing).

Consideriamo una famiglia d'insiemi di addestramento, che indichiamo attraverso l'indice i , e una famiglia di insiemi di prova che indichiamo con l'indice j . In formula, abbiamo che:

$$\begin{aligned} L_i &= \{l_t\}; l_t = NID(0,1); t = 1,2,\dots |L_i|; i \in \{1,2,\dots + \infty\}; \\ T_j &= \{Y_t\}; t = 1,2,\dots |T_j|; j \in \{1,2,\dots + \infty\}; \end{aligned} \quad \text{[C 1]}$$

Ovviamente, per ogni fissata coppia di indici i e j è possibile definire i seguenti insiemi numerabili di learning e di testing a valle della codifica di *Window Forming* a parametro M .

$$\begin{aligned} \bar{L}_i^M &= \{[l_1, l_2, \dots, l_M], [l_2, l_3, \dots, l_{1+M}], \dots, [l_d, l_{d+1}, \dots, l_{d+M}]\} \\ \bar{T}_j^M &= \{[Y_1, Y_2, \dots, Y_M], [Y_2, Y_3, \dots, Y_{1+M}], \dots, [Y_p, Y_{p+1}, \dots, Y_{p+M}]\} \end{aligned} \quad \text{[C 2]}$$

Dove supponiamo che la dimensione (o cardinalità) sia costante (d) al variare dell'indice i per ogni insieme di addestramento e pari a $|L_i| = M + d$, così come si suppone costante la dimensione degli insiemi di testing al variare dell'indice j e pari a $|T_j| = M + p$.

I parametri $d \geq 1$ e $p \geq 1$, pertanto, determinano la dimensione dell'insieme di learning e di testing della rete neurale. Nelle prove descritte di seguito, mentre di considererò d un fattore dell'esperimento, il parametro p sarà, invece, considerato fissato ($p = 10000$).

Con riferimento al criterio di generalizzazione implementato dall'algorithmo neurale è immediato constatare che indicato con

$$\underline{Y}_t = [Y_{t-M+1}, Y_{t-M+2}, \dots, Y_t] \quad \forall t \geq M \quad \text{[C 3]}$$

un vettore di M successive osservazioni prodotte dal processo monitorato, allora il sistema sarà considerato all'istante di tempo t in uno stato naturale se e solo se risulta verificata la seguente condizione

$$b_{nn,t} = 1 \Leftrightarrow \exists \underline{L}_j = [l_{1j}, l_{2j}, \dots, l_{Mj}] \in \bar{L}^M : \sum_{r=1}^M |Y_{t-M+r} - l_{rj}| \leq 2lM(1 - \rho) \quad \text{[C 4]}$$

Ovvero se e solo se esiste almeno un pattern dell'insieme di addestramento tale che la media degli scarti in valore assoluto fra le corrispondenti componenti con il pattern presentato in ingresso ($MAD = \frac{1}{M} \sum_{r=1}^M |Y_{t-M+r} - l_{rj}|$) sia minore o uguale al limite $2l(1 - \rho)$, dove ρ è il parametro di vigilanza scelto per la rete, e l è il limite di variazione in valore assoluto scelto per la codifica di *Coding* (ad esempio $l = 3$).

Si osservi che la precedente relazione vale, a rigore, solo supponendo che tutte le osservazioni rientrino nell'insieme limitato di variabilità $[-l, l]$.

C.2.1 Costruzione dell'esperimento

Le prestazioni di una rete Fuzzy ART (che misuriamo in termini degli errori di Tipo I e di Tipo II) sono influenzate da due famiglie di parametri:

1. i parametri di configurazione dell'algorithmo ART;
2. i parametri di configurazione dell'esperimento.

I parametri di configurazione della rete Fuzzy ART a addestramento PLP sono riportati nella seguente Tabella C 1, mentre i parametri dell'esperimento sono riportati nella Tabella C 2.

<i>Parametro ART</i>	<i>Insieme di definizione</i>
Neuroni F1 (M)	$\{1,2,\dots, +\infty\}$
Parametro di Vigilanza (ρ)	$]0,1]$

Tabella C 1 – Parametri rete neurale

<i>Parametro Esperimento</i>	<i>Insieme di definizione</i>
Indice Insieme Learning (i)	$\{1,2,\dots, +\infty\}$
Indice Insieme Testing (j)	$\{1,2,\dots, +\infty\}$
Dimensione Insieme Learning (d)	$\{1,2,\dots, +\infty\}$

Tabella C 2 – Parametri esperimento

La probabilità di un errore di Tipo I, si può esprimere come funzione di cinque parametri.

$$\alpha = P\{H_1|H_0\} = F(\rho, M, i, j, d) \quad [C 5]$$

Tenuto conto che i parametri relativi sia all'insieme di learning (indicato con indice i) sia all'insieme di testing (indicato con indice j) sono scelti in modo casuale, allora l'espressione [C 5] si può modellare come segue:

$$\begin{aligned} \alpha_{ij}(\rho, M, d) &= f(\rho, M, d) + e_i(\rho, M, d) + e_j(\rho, M, d) + e_{ij}(\rho, M, d) \\ e_i(\rho, M, d) &\approx NID(0, \sigma_i^2(\rho, M, d)) \\ e_j(\rho, M, d) &\approx NID(0, \sigma_j^2(\rho, M, d)) \\ e_{ij}(\rho, M, d) &\approx NID(0, \sigma_{ij}^2(\rho, M, d)) \end{aligned} \quad [C 6]$$

Le precedenti considerazioni relative all'errore di Tipo I, possono essere estese anche al caso dell'errore di Tipo II che può essere espresso dalla seguente relazione generale.

$$\beta = P\{H_0|H_1\} = G(\rho, M, i, j, d, s, \underline{\Phi}) \quad [C 7]$$

Dove, in particolare, è stato indicato con l'indice s il tipo di pattern utilizzato per modellare le uscite del processo influenzate da cause assegnabili (non naturali), mentre con $\underline{\Phi}$ si è indicato il vettore dei parametri che caratterizzano l'andamento non naturale s . Considerando nel seguito unicamente un pattern di tipo shift di parametri $\tau = 1$ e $\varphi = 1.5$ (capitolo 3), è possibile riscrivere la relazione come segue.

$$\beta = P\{H_0|H_1\} = G(\rho, M, i, j, d) \quad [C 8]$$

Tenuto conto ancora una volta che i parametri relativi sia all'insieme di addestramento (indicato con indice i) sia all'insieme di testing (indicato con indice j) sono scelti in modo casuale, allora l'espressione si può modellare come segue:

$$\beta_{ij}(\rho, M, d) = g(\rho, M, d) + e_i(\rho, M, d) + e_j(\rho, M, d) + e_{ij}(\rho, M, d)$$

$$e_i(\rho, M, d) \approx NID(0, \sigma_i^2(\rho, M, d))$$

$$e_j(\rho, M, d) \approx NID(0, \sigma_j^2(\rho, M, d))$$

$$e_{ij}(\rho, M, d) \approx NID(0, \sigma_{ij}^2(\rho, M, d))$$

[C 9]

C.3 Risultati sperimentali

Per valutare l'effetto di ogni singolo fattore è stato costruito il piano sperimentale riportato nella seguente Tabella C 3. L'esperimento consiste di tre fattori fissi binari e 2 fattori casuali per un totale di 1800 combinazioni. Per i fattori a effetti fissi binari si utilizzerà la notazione riportata nella Tabella C 4.

<i>Fattore ART</i>	<i>Livelli</i>	<i>Valori</i>	<i>Tipo</i>
Neuroni F1 (<i>M</i>)	2	{10,25}	Fisso
Parametro di Vigilanza (ρ)	2	{0.85,0.90}	Fisso
<i>Fattore Esperimento</i>	<i>Livelli</i>	<i>Valori</i>	
Indice Insieme Learning (<i>i</i>)	15	{1,2,..., 15}	Casuale
Indice Insieme Testing (<i>j</i>)	15	{1,2,..., 15}	Casuale
Dimensione Insieme Learning (<i>d</i>)	2	{100, 1000}	Fisso

Tabella C 3– Fattori dell'esperimento di simulazione

	- (<i>low</i>)	+ (<i>high</i>)
<i>M</i>	10	25
ρ	0.85	0.90
<i>d</i>	100	1000

Tabella C 4– Notazione esperimento

La procedura per la valutazione delle prestazioni della rete neurale è la seguente:

1. Analisi del piano fattoriale completo attraverso la tecnica ANOVA. In tal modo si intende porre in evidenza l'effetto di ogni singolo fattore nonché di tutte le interazioni (doppie e triple) fra i fattori binari ρ , *M* e *d*.
2. Per ognuno delle otto combinazioni dei parametri binari ρ , *M* e *d* si considerano le variabilità sulla prestazione della rete neurale (rispettivamente errore di Tipo I e di Tipo II) derivanti dalla variazione dell'insieme di addestramento e dalla variazione dell'insieme di testing. In tal modo si intende porre in evidenza l'influenza della scelta casuale dell'insieme di addestramento e dell'insieme di testing sulla variabilità della misura di prestazione ottenuta dalla rete neurale.
3. Schematizzazione degli effetti principali e di interazione doppia dei parametri binari ρ , *M* e *d* sulle prestazioni della rete neurale.

I risultati delle fasi sopra esposte sono riportati di seguito sia per l'errore di Tipo I, sia per l'errore di Tipo II per il riconoscimento di uno shift costante della media ($\tau = 1$ e $\varphi = 1.5$).

C.3.1 Errore di Tipo I

I risultati del piano fattoriale, analizzati per mezzo della analisi della varianza ANOVA, relativi all'errore di Tipo I sono riportati nella seguente tabella.

ANOVA: Alpha versus M; d; rho; i; j										
Factor	Type	Levels	Values							
M	fixed	2	10	25						
d	fixed	2	100	1000						
rho	fixed	2	0.85	0.90						
i	random	15	1	2	3	4	5	6	7	
			8	9	10	11	12	13	14	
j	random	15	1	2	3	4	5	6	7	
			8	9	10	11	12	13	14	
			15							

Analysis of Variance for Alpha						
Source	DF	SS	MS	F	P	
M	1	54.2380	54.2380	2.3E+05	0.000	
d	1	9.0982	9.0982	3.9E+04	0.000	
rho	1	139.5610	139.5610	5.9E+05	0.000	
M*d	1	0.5912	0.5912	2505.75	0.000	
M*rho	1	44.0424	44.0424	1.9E+05	0.000	
d*rho	1	4.5210	4.5210	1.9E+04	0.000	
M*d*rho	1	2.0490	2.0490	8684.26	0.000	
i	14	0.1533	0.0109	46.40	0.000	
j	14	0.0416	0.0030	12.60	0.000	
Error	1764	0.4162	0.0002			
Total	1799	254.7119				

Tabella C 5– Tabella ANOVA errore di Tipo I

Per valutare analiticamente l'influenza dei fattori casuali i e j sulla risposta del sistema, nella Tabella C 6 sono stati riportati (per ognuna delle 8 combinazioni dei fattori ad effetti fissi), i livelli medi dell'errore di Tipo I (indicati nella colonna $\hat{f}(\rho, M, d)$) e le relative deviazioni standard (riportata nella colonna indicata con l'etichetta "dev.st"). Livello medio e deviazione standard per ognuno degli otto campioni sono ottenuti da una popolazione di 225 osservazioni ottenute combinando i 15 differenti insiemi di addestramento con i 15 insiemi di testing.

La variabilità per ognuno degli otto campioni corrispondenti alle diverse combinazioni dei parametri ρ , M e d , è stata poi ulteriormente suddivisa nelle diverse componenti:

- la componente legata alla scelta dell'insieme di addestramento ($\hat{\sigma}_I(\rho, M, d)$);
- la componente legata alla scelta dell'insieme testing ($\hat{\sigma}_J(\rho, M, d)$);
- quella legata all'errore, ovvero all'interazione tra insieme di addestramento e di testing ($\hat{\sigma}_{IJ}(\rho, M, d)$).

ρ	M	d	$\hat{f}(\rho, M, d)$	$\hat{\sigma}_I(\rho, M, d)$	$\hat{\sigma}_J(\rho, M, d)$	$\hat{\sigma}_{IJ}(\rho, M, d)$	dev.st.
-	-	-	1.100%	0.197%	0.187%	0.136%	0.304%
-	-	+	0.027%	0.004%	0.017%	0.018%	0.025%
-	+	-	7.656%	2.293%	0.898%	0.509%	2.515%
-	+	+	0.337%	0.031%	0.134%	0.084%	0.161%
+	-	-	42.276%	3.773%	1.191%	0.671%	4.013%
+	-	+	7.661%	0.189%	0.662%	0.292%	0.748%

+	+	-	97.906%	1.090%	0.187%	0.319%	1.151%
+	+	+	84.036%	1.781%	1.018%	0.604%	2.138%

Tabella C 6– Errore di Tipo I e variabilità

Gli effetti principali e dei tre parametri binari ρ , M e d sono schematizzati nei grafici di Figura C 1. In particolare, i *Countor Plot* evidenziano per una fissata dimensione dell'insieme di addestramento ($d = 100, 1000$) i luoghi di punti nel piano $M - \rho$ nei quali la rete neurale presenta i medesimi livelli di errore di Tipo I. Il diagramma di *Pareto* inoltre, evidenzia il livello di influenza che ogni singolo fattore, o l'interazione fra i fattori, ha sulla risposta del sistema.

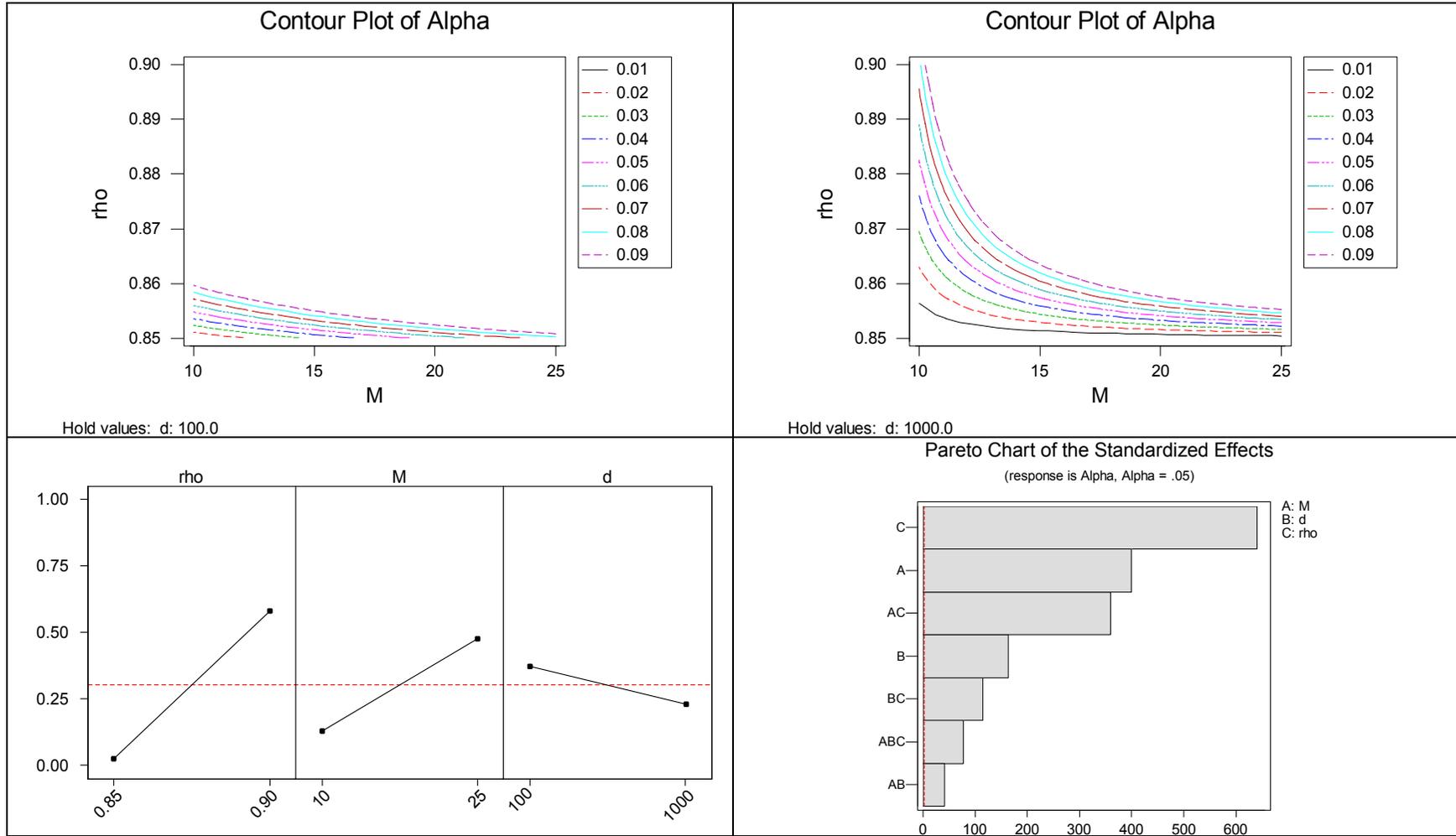


Figura C 1 – Schematizzazione risultati ANOVA per errore di Tipo I

C.3.2 Errore di Tipo II

Il piano fattoriale relativo all'errore di Tipo II è riportato nella seguente tabella.

ANOVA: Beta versus M; d; rho; i; j									
Factor	Type	Levels	Values						
M	fixed	2	10	25					
d	fixed	2	100	1000					
rho	fixed	2	0.85	0.90					
i	random	15	1	2	3	4	5	6	7
			8	9	10	11	12	13	14
			15						
j	random	15	1	2	3	4	5	6	7
			8	9	10	11	12	13	14
			15						
Analysis of Variance for Beta									
Source	DF	SS	MS	F	P				
M	1	22.7089	22.7089	4849.77	0.000				
d	1	4.5100	4.5100	963.16	0.000				
rho	1	17.9776	17.9776	3839.34	0.000				
M*d	1	3.1847	3.1847	680.14	0.000				
M*rho	1	13.4501	13.4501	2872.43	0.000				
d*rho	1	2.2510	2.2510	480.72	0.000				
M*d*rho	1	1.3488	1.3488	288.06	0.000				
i	14	4.2303	0.3022	64.53	0.000				
j	14	0.0123	0.0009	0.19	1.000				
Error	1764	8.2599	0.0047						
Total	1799	77.9336							

Tabella C 7– Tabella ANOVA errore di Tipo II

Come nel caso dell'errore di Tipo I, nella seguente Tabella C 8 sono stati riportati (per ognuna delle 8 combinazioni dei fattori ad effetti fissi), i livelli medi dell'errore di Tipo II (indicati nella colonna $\hat{g}(\rho, M, d)$) e le relative deviazioni standard (riportata nella colonna indicata con l'etichetta "dev.st"). La configurazione dell'esperimento è identica a quella già descritta nel caso dell'errore di Tipo I. La componente di variabilità è stata suddivisa in:

- la componente legata alla scelta dell'insieme di addestramento: $\hat{\sigma}_I(\rho, M, d)$;
- la componente legata alla scelta dell'insieme testing: $\hat{\sigma}_J(\rho, M, d)$;
- quella legata all'errore o interazione: $\hat{\sigma}_{IJ}(\rho, M, d)$.

ρ	M	d	$\hat{g}(\rho, M, d)$	$\hat{\sigma}_I(\rho, M, d)$	$\hat{\sigma}_J(\rho, M, d)$	$\hat{\sigma}_{IJ}(\rho, M, d)$	dev.st.
-	-	-	26.970%	20.174%	0.829%	0.547%	20.198%
-	-	+	57.938%	11.517%	1.042%	0.632%	11.581%
-	+	-	1.101%	3.533%	0.089%	0.301%	3.548%
-	+	+	4.298%	3.998%	0.531%	0.387%	4.052%
+	-	-	2.238%	3.486%	0.139%	0.237%	3.497%
+	-	+	8.114%	3.405%	0.477%	0.406%	3.461%
+	+	-	0.000%	0.000%	0.000%	0.001%	0.001%
+	+	+	0.001%	0.001%	0.000%	0.004%	0.004%

Tabella C 8– Errore di Tipo I e variabilità

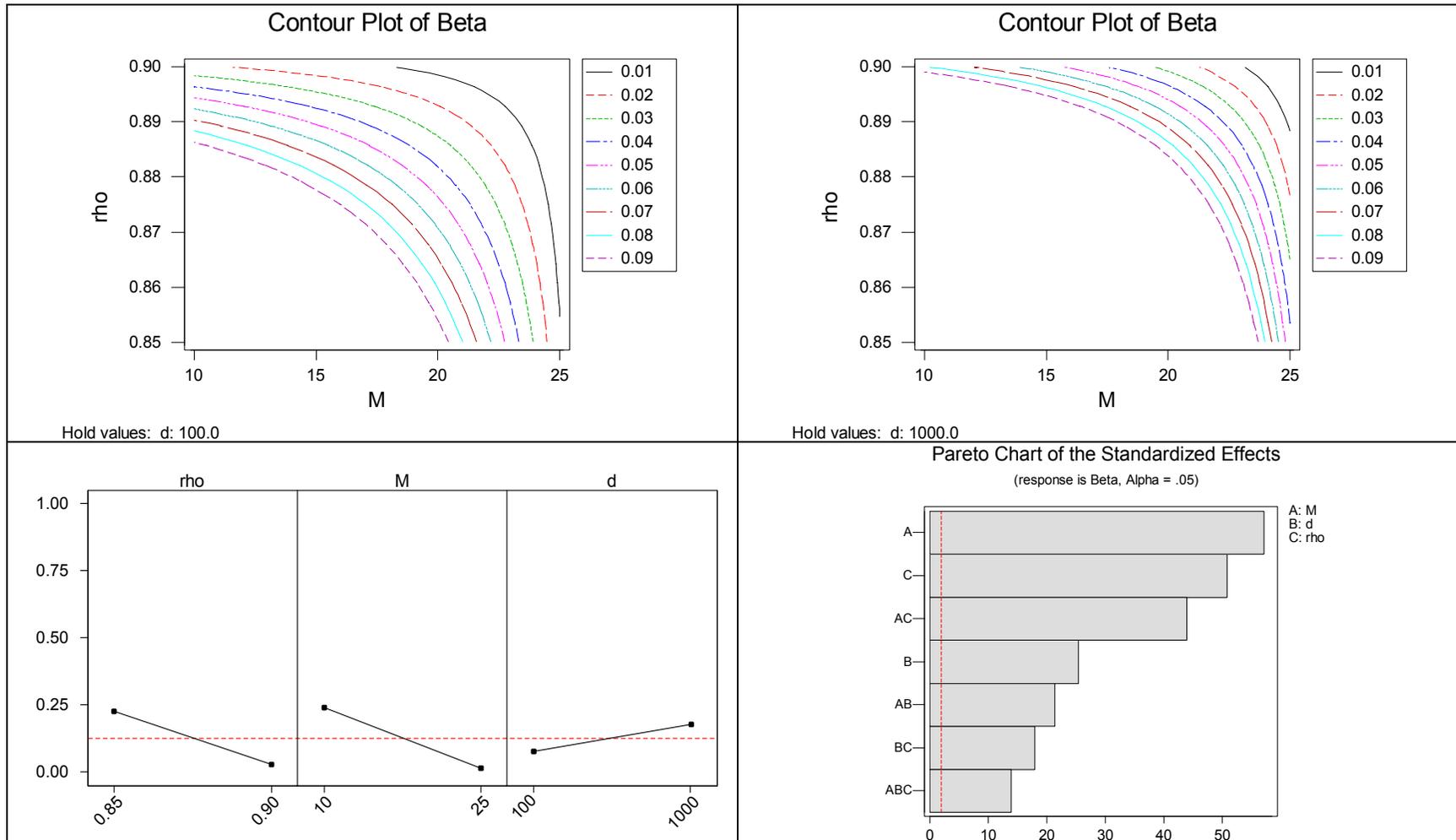


Figura C 2 – Schematizzazione risultati ANOVA per errore di Tipo II

C.3.3 Considerazioni

Dai risultati riportati in precedenza è possibile dedurre le seguenti considerazioni.

- L'influenza dei parametri M e ρ sulle prestazioni della rete neurale risulta identica a quella osservata nel caso di addestramento limitato esclusivamente sul target del processo.
- Il funzionamento della rete misurato sia in termini di errore di Tipo I sia in termini di errore di Tipo II è scarsamente influenzato dall'insieme di testing utilizzato per valutarne le prestazioni.
- Il funzionamento della rete neurale misurato in termini di errore di Tipo I è "debolmente" influenzato dalla scelta dell'insieme di addestramento.
- Il funzionamento della rete neurale misurato in termini di errore di Tipo II è "fortemente" influenzato dalla scelta dell'insieme di addestramento.
- Aumentando la dimensione della finestra M si stabilizza il funzionamento della rete neurale (sia in termini di errore di Tipo I che in termini di errore di Tipo II) al variare dell'insieme di addestramento.

C.4 Miglioramento delle prestazioni

In questa sezione sono presentati i risultati delle prove sopra esposte nel caso $M = 75$ e $d = 7500$. Si può osservare che l'aumento della dimensione per la finestra di dati M , associato inoltre ad un aumento della dimensione dell'insieme di addestramento d , consente di ridurre l'influenza della scelta dell'insieme di addestramento sulle prestazioni della rete neurale.

ρ	$\hat{f}(\rho, M, d)$	$\hat{\sigma}_I(\rho, M, d)$	$\hat{\sigma}_J(\rho, M, d)$	$\hat{\sigma}_{II}(\rho, M, d)$	<i>dev.st.</i>
-	3.472%	0.267%	0.974%	0.405%	1.088%
+	100.00%	0.000%	0.000%	0.000%	0.000%

Tabella C 9– Errore di Tipo I e variabilità (caso M=75, d=7500)

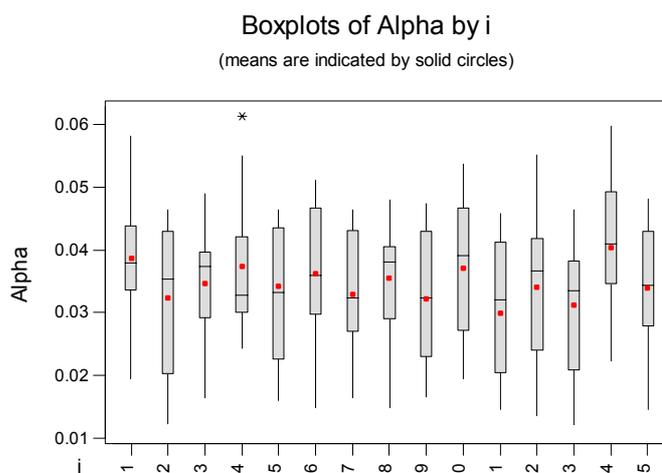


Figura C 3 – Schematizzazione errore di Tipo I per 15 insiemi di addestramento (caso M=75, d=7500, $\rho=0.85$)

ρ	$\hat{g}(\rho, M, d)$	$\hat{\sigma}_I(\rho, M, d)$	$\hat{\sigma}_J(\rho, M, d)$	$\hat{\sigma}_{IJ}(\rho, M, d)$	dev.st.
-	0.000%	0.000%	0.000%	0.000%	0.000%
+	0.000%	0.000%	0.000%	0.000%	0.000%

Tabella C 10– Errore di Tipo II e variabilità (caso M=75, d=7500)

C.5 Funzione errore di Tipo I e di Tipo II

In questa sezione sono rappresentate le funzioni errore di Tipo I ($\alpha = f(\rho, M, d)$) e di Tipo II ($\beta = g(\rho, M, d)$) in relazione al livello del parametro di vigilanza ρ , e per diverse combinazioni dei parametri d (dimensione dell'insieme di addestramento), M (ampiezza della finestra della rete neurale). Accanto alla rappresentazione delle funzioni $\alpha = f(\rho, M, d)$ ed $\beta = g(\rho, M, d)$, sono stati riportate le funzioni derivata parziale rispetto al parametro di vigilanza: ovvero $\left. \frac{\partial f}{\partial \rho}(\rho) \right|_{M,d}$ e $\left. \frac{\partial g}{\partial \rho}(\rho) \right|_{M,d}$ rispettivamente.

Le funzioni sono state stimate attraverso la simulazione su un insieme di 26000 osservazioni. La rete è stata addestrata su un insieme di 10000 osservazioni naturali del processo. Per due prefissati valori del parametro di vigilanza (0.85 e 0.90) sono, inoltre, stati rappresentati i valori ottenibili modificando l'insieme di addestramento.

La funzione $\alpha = f(\rho, M, d)$ per ogni fissata coppia di valori M e d si presenta monotona crescente rispetto al parametro ρ , ovvero $\forall \rho \in [0,1]: \left. \frac{\partial f}{\partial \rho}(\rho) \right|_{M,d} \geq 0$. Al contrario la funzione $\beta = g(\rho, M, d)$ per ogni coppia di valori M e d si presenta monotona decrescente rispetto al parametro ρ , ovvero $\forall \rho \in [0,1]: \left. \frac{\partial g}{\partial \rho}(\rho) \right|_{M,d} \leq 0$.

Alcune importanti conclusioni che possono essere dedotte sono le seguenti.

- Aumentando l'ampiezza della finestra M si stabilizza il funzionamento della rete poiché decresce l'influenza sulle prestazioni che può essere causata da un cambio dell'insieme di addestramento.
- Aumentando la dimensione dell'insieme di addestramento d si stabilizza il funzionamento della rete poiché decresce l'influenza sulle prestazioni che può essere causata da un cambio dell'insieme di addestramento.
- Aumentando l'ampiezza della finestra M aumenta il tasso di crescita (decrescita) dell'errore di Tipo I (Tipo II) presentato dalla funzione $\alpha = f(\rho, M, d)$ ($\beta = g(\rho, M, d)$) in relazione all'aumento del parametro di vigilanza.
- Aumentando la dimensione dell'insieme di addestramento d aumenta il tasso di crescita (decrescita) dell'errore di Tipo I (Tipo II) presentato dalla funzione $\alpha = f(\rho, M, d)$ ($\beta = g(\rho, M, d)$) in relazione all'aumento del parametro di vigilanza.

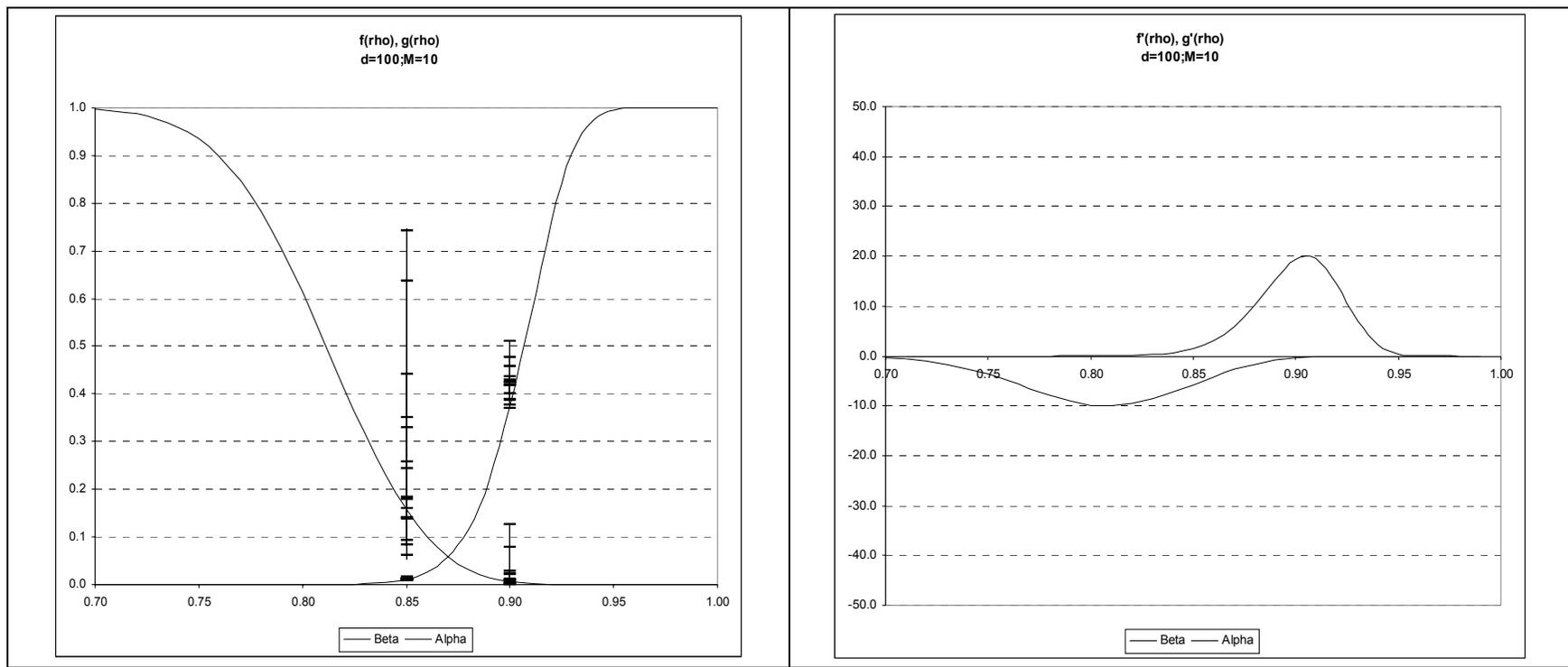


Figura C 4 – Funzioni errore di Tipo I e di Tipo II (sinistra) e derivate parziali rispetto al paramero di vigilanza ($d=100, M=10$).

I tratti verticali sul diagramma delle funzioni errore rappresentano l'intervallo di variabilità osservato variando la scelta dell'insieme di addestramento per due valori del parametro di vigilanza (0.85 e 0.90).

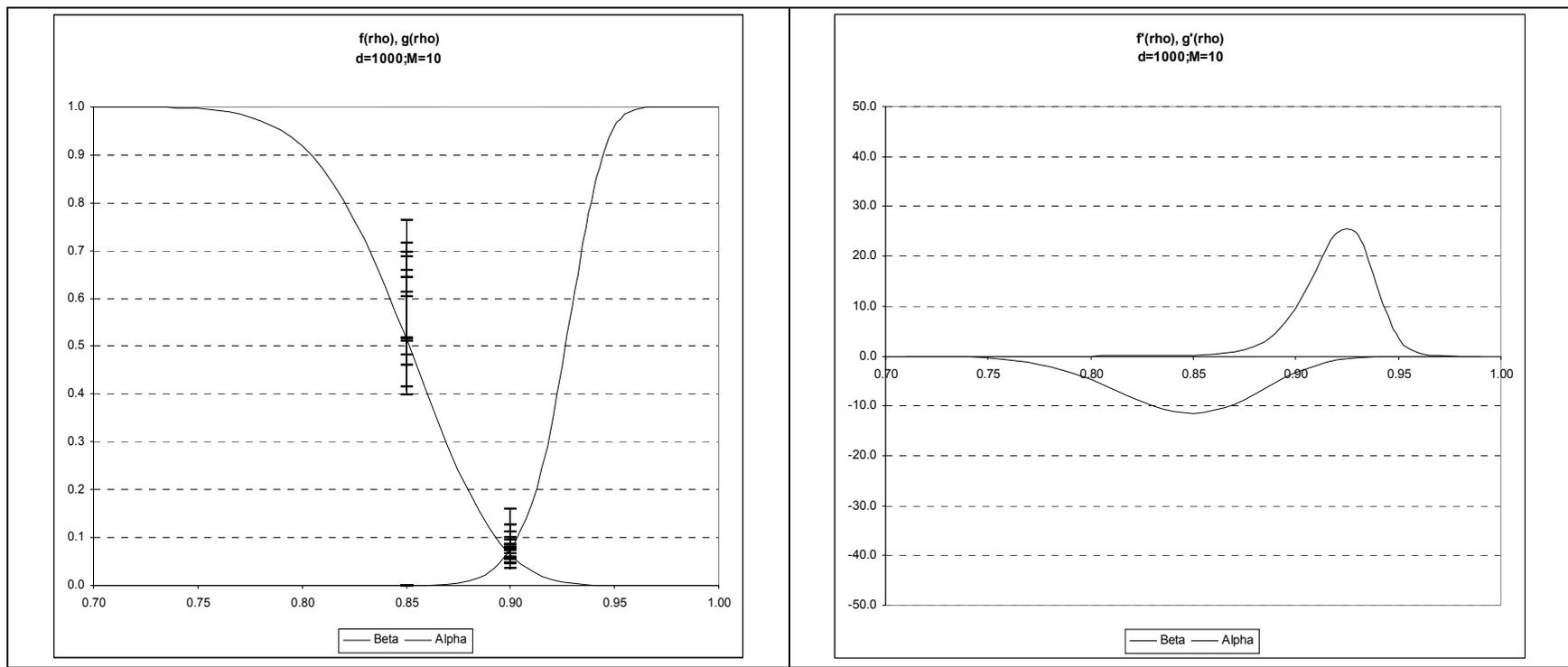


Figura C 5 – Funzioni errore di Tipo I e di Tipo II (sinistra) e derivate parziali rispetto al paramero di vigilanza ($d=1000$, $M=10$).

I tratti verticali sul diagramma delle funzioni errore rappresentano l'intervallo di variabilità osservato variando la scelta dell'insieme di addestramento per due valori del parametro di vigilanza (0.85 e 0.90).

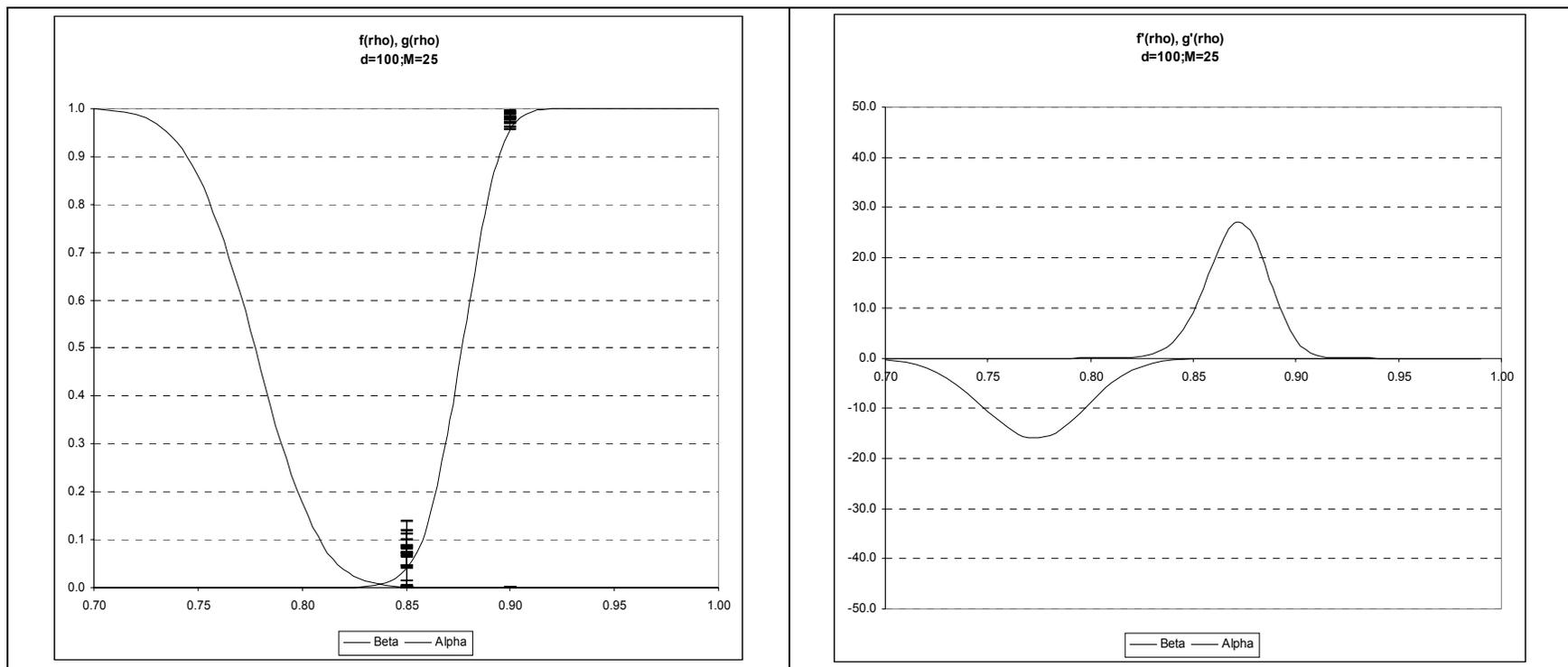


Figura C 6 – Funzioni errore di Tipo I e di Tipo II (sinistra) e derivate parziali rispetto al parametro di vigilanza ($d=100, M=25$).

I tratti verticali sul diagramma delle funzioni errore rappresentano l'intervallo di variabilità osservato variando la scelta dell'insieme di addestramento per due valori del parametro di vigilanza (0.85 e 0.90).

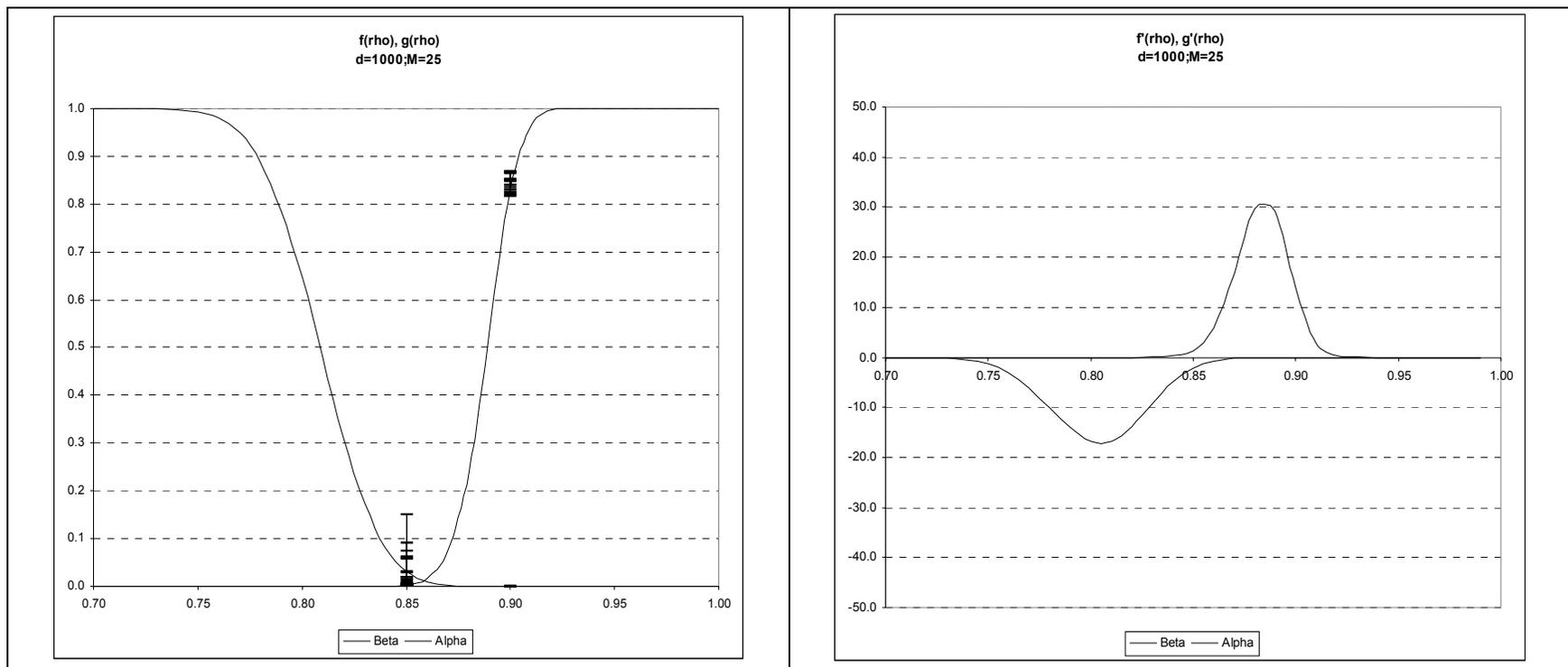


Figura C 7 – Funzioni errore di Tipo I e di Tipo II (sinistra) e derivate parziali rispetto al paramero di vigilanza (d=1000, M=25).

I tratti verticali sul diagramma delle funzioni errore rappresentano l'intervallo di variabilità osservato variando la scelta dell'insieme di addestramento per due valori del parametro di vigilanza (0.85 e 0.90).

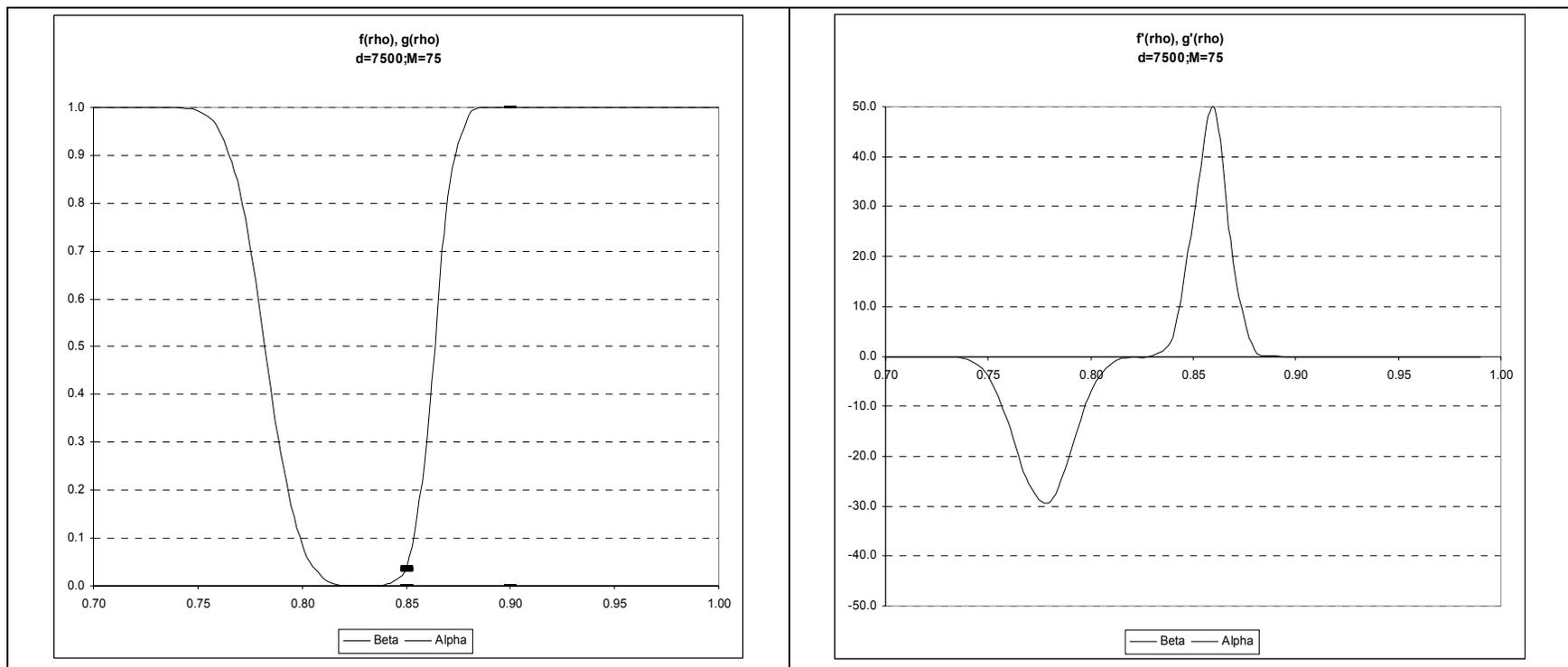


Figura C 8 – Funzioni errore di Tipo I e di Tipo II (sinistra) e derivate parziali rispetto al paramero di vigilanza ($d=7500$, $M=75$).

I tratti verticali sul diagramma delle funzioni errore (in tal caso notevolmente ridotti) rappresentano l'intervallo di variabilità osservato variando la scelta dell'insieme di addestramento per due valori del parametro di vigilanza (0.85 e 0.90).

