

**UNIVERSITA' DEGLI STUDI DI BARI**

FACOLTA' DI ECONOMIA

CORSO DI LAUREA IN SCIENZE STATISTICHE ED ECONOMICHE

---

TESI DI LAUREA

IN

STATISTICA ECONOMICA

**ANALISI MODERNA DELLE SERIE STORICHE  
RELATIVE AI CONSUMI IRRIGUI  
NELL'AGRO DI FOGGIA**

**Relatori:**

**Chiar.mo Prof. Antonio Mastrodonato**

**Dott. Marcello Mastrorilli**

**Laureanda:**

**SIMONA PISANI**

---

**ANNO ACCADEMICO 2004-2005**

*Si ringrazia per la guida qualificata e per la grande disponibilità dimostrata il Prof. Antonio Mastrodonato (direttore del Dipartimento del Corso di Laurea in Scienza Statistiche ed Economiche della Facoltà di Economia dell'Università di Bari).*

*Si ringrazia, inoltre, per il valido supporto tecnico fornito il Dott. Marcello Mastrorilli (direttore della Sezione Irrigazione dell'Istituto Sperimentale Agronomico di Bari).*

# ***INDICE***

<b>INTRODUZIONE.....</b>	<b>6</b>
<b>CAPITOLO 1: AGRICOLTURA E IRRIGAZIONE IN PUGLIA.....</b>	<b>11</b>
1.1 Struttura e caratteristiche dell'agricoltura.....	11
1.2 Distribuzione territoriale delle colture .....	16
1.3 I Consorzi di Bonifica .....	19
1.3.1 Consorzio di Bonifica del Gargano.....	21
1.3.2 Consorzio di Bonifica della Capitanata .....	23
1.3.3 Consorzio di Bonifica Terre d'Apulia .....	25
1.3.4 Consorzio di Bonifica Stornara e Tara.....	27
1.3.5 Consorzio di Bonifica Arneo .....	28
1.3.6 Consorzio di Bonifica Ugento Li Foggi.....	30
1.4 Acque sotterranee e sorgive.....	33
1.5 Corsi d'acqua superficiali.....	36
1.6 Schemi idrici ad uso plurimo esistenti .....	37
1.7 Caratteristiche e struttura dell'irrigazione.....	39
1.8 Conclusioni .....	45
<b>CAPITOLO 2: IL PIANO DELL'INDAGINE.....</b>	<b>47</b>
2.1 I dati .....	47
2.2 Calcolo dell'evapotraspirazione delle colture.....	48
2.3 Il bilancio idrico, ovvero quando e quanto irrigare .....	53
2.4 Obiettivo dell'indagine .....	56
2.5 Software utilizzati per l'analisi .....	57

### **CAPITOLO 3: PROCESSI STOCASTICI E MODELLI LINEARI .. 60**

3.1	Introduzione.....	60
3.2	Lo studio dei fenomeni che variano nel tempo.....	61
3.3	Diversi approcci allo studio delle serie storiche.....	63
3.3.1	Approccio classico .....	64
3.3.2	Approccio moderno o statistico .....	67
3.4	I processi stocastici .....	68
3.5	Processi stazionari.....	74
3.6	Stima dei momenti di un processo stocastico stazionario.....	81
3.7	Alcuni processi stocastici stazionari.....	85
3.7.1	Processo puramente casuale (white-noise) .....	85
3.7.2	Processo AutoRegressivo - AR(p) .....	87
3.7.3	Processo a Media Mobile - MA(q) .....	92
3.7.4	Processo Autoregressivo a Media – ARMA (p,q) .....	97
3.8	Processi non stazionari.....	101
3.8.1	Processo passeggiata casuale ( <i>random walk</i> ) .....	103
3.8.2	Processo AutoRegressivo Integrato a Media Mobile - ARIMA (p,d,q).....	105

### **CAPITOLO 4: ANALISI TEMPORALE DEI DATI MEDIANTE LA PROCEDURA DI BOX-JENKINS ..... 108**

4.1	La procedura di Box e Jenkins.....	108
4.1.1	Identificazione.....	110
4.1.2	Stima dei parametri .....	112
4.1.3	Controllo diagnostico .....	115
4.2	Specificazione di modelli stocastici applicati ai consumi irrigui nell'agro di Foggia in relazione alle due principali colture irrigue praticate in quest'area.....	119
4.2.1	Olivo.....	120
4.2.2	Pomodoro .....	133

### **CONSIDERAZIONI CONCLUSIVE..... 146**

<b>APPENDICE</b> .....	153
TAVOLA A.1 Spiegazione dei comandi inseriti in "R 2.1.0" per l'analisi temporale dei dati.....	154
TAVOLA A.2 Comandi inseriti in "R 2.1.0" e relativi risultati per l'analisi temporale dei consumi irrigui dell'olivo nell'agro di Foggia.....	156
TAVOLA A.3 Comandi inseriti in "R 2.1.0" e relativi risultati per l'analisi temporale dei consumi irrigui del pomodoro nell'agro di Foggia.....	160
<b>BIBLIOGRAFIA</b> .....	167

# INTRODUZIONE

Per il settore agricolo l'acqua costituisce un elemento determinante in termini quantitativi e qualitativi, soprattutto nelle regioni meridionali dove la maggior parte delle colture attualmente praticate risulta possibile solo grazie al supporto dell'irrigazione. Inoltre, come risulta da numerose esperienze condotte in materia, laddove si sono insediati sistemi agricoli in irriguo, sono migliorate le tecniche produttive e consolidati i legami con l'industria alimentare, determinando una produttività dell'intero settore paragonabile a quella conseguita nello stesso settore industriale.

L'indagine conoscitiva sull'approvvigionamento idrico e sugli usi agricoli delle acque, presentata dalla nona Commissione Agricoltura e Produzione Agroalimentare del Senato nel febbraio del 2003 ha evidenziato come l'irrigazione si stia trasformando da un mero mezzo

per raggiungere elevate rese produttive ad un fattore in grado di indirizzare le produzioni agricole verso la qualità e di contribuire alla salvaguardia del territorio e alla tutela delle risorse naturali nel quadro dei nuovi indirizzi di politica agraria comunitaria. Viene peraltro sottolineato che questa tendenza può essere assecondata solo se si è in grado di assicurare adeguate disponibilità idriche da destinare all'irrigazione.

In Italia, considerato un Paese potenzialmente ricco di risorse idriche, la natura del territorio, la disomogenea distribuzione delle precipitazioni tra contesti geografici, la conseguente irregolarità dei deflussi superficiali e lo stato infrastrutturale delle reti di distribuzione non consentono di utilizzare appieno le risorse potenzialmente disponibili. Ulteriori difficoltà per gli approvvigionamenti idrici derivano dai sempre più frequenti periodi siccitosi che nel corso degli ultimi anni hanno interessato vaste aree del nostro Paese, in particolare del Mezzogiorno, dove si sono registrate forti riduzioni idriche negli invasi, e con segnali preoccupanti anche nel Centro e nel Nord, aree tradizionalmente non soggette a questo tipo di problematiche.

Alla diminuzione delle disponibilità idriche si aggiunge la tendenza all'aumento della domanda di acqua da parte di tutti i settori produttivi. Gli insediamenti abitativi richiedono maggiori disponibilità, le

installazioni industriali e artigianali richiedono quantitativi sempre maggiori e aumenta la domanda di acqua per il settore turistico. Si determina dunque l'inasprimento della competizione per l'allocazione settoriale. In particolare, il settore turistico rappresenta il maggiore antagonista per il settore primario in prossimità delle fasce costiere, poiché aumenta la richiesta di acqua proprio nel periodo di maggior stress idrico, cioè quello estivo.

Lo stesso settore agricolo registra un aumento della domanda di acqua, sia perché sono irrigate colture tradizionalmente non irrigue, sia perché le modifiche intervenute negli anni sulle condizioni meteorologiche, che ormai caratterizzano le stagioni irrigue, hanno determinato un aumento, non solo nel periodo estivo ma anche in quello autunnale, del fabbisogno di acqua da parte delle colture. L'andamento estremamente irregolare delle risorse superficiali, che impone l'esistenza di sistemi di accumulo per usufruire di sufficienti disponibilità idriche sia nel tempo (elevata richiesta nei mesi estivi) che nello spazio (trasferimenti interregionali) e l'utilizzazione di molte fonti di approvvigionamento non solo per scopi irrigui, ha indirizzato le aziende agricole delle regioni meridionali a ricorrere significativamente alle risorse sotterranee (complessivamente in Puglia, Sicilia e Campania si concentra l'80% delle aziende irrigue meridionali che attuano questa



forma di approvvigionamento) che in molti casi, soprattutto lungo le fasce costiere delle regioni meridionali, ha determinato un abbassamento del livello delle falde e un depauperamento delle risorse idriche sotterranee a causa della crescente intrusione di acqua di mare.

Inoltre, negli ultimi quaranta anni la rapida diffusione, nelle regioni meridionali, delle specie orticole e frutticole caratterizzate da elevate necessità idriche, se da un lato ha contribuito a migliorare il reddito e ad aumentare l'occupazione di manodopera nel settore agricolo, dall'altro ha avuto e continua ad avere un impatto ambientale negativo, soprattutto in relazione ai fenomeni di inquinamento delle acque e di degrado del territorio, conseguenti, appunto, al tipo di gestione delle risorse naturali in particolare dell'acqua.

Acqua-agricoltura-ambiente si configurano ormai come un trionomio inscindibile, che deve concretizzarsi in una efficace politica di tutela ambientale, alla quale l'agricoltura ed in particolare l'impiego irriguo delle acque, può e deve attivamente contribuire.

I cambiamenti descritti sono stati, in qualche modo, percepiti dalla collettività ma non sono stati statisticamente dimostrati e quantificati.

Nel presente lavoro, pertanto, è stato analizzato l'andamento temporale dei consumi irrigui, che rappresentano una variabile strettamente collegata ai cambiamenti in atto, nell'Agro di Foggia (zona

della Puglia particolarmente interessata alle problematiche ambientali succitate) in relazione alle due principali colture presenti in quest'area (olivo, pomodoro); tale studio è stato condotto mediante l'approccio moderno o statistico all'analisi delle serie storiche.

Nel primo capitolo di questa tesi, si procede ad una disamina preliminare dell'agricoltura e dello stato dell'irrigazione in Puglia, allo scopo di comprendere il contesto nel quale effettuare l'analisi oggetto di studio. Esposto il piano dell'indagine nel secondo capitolo, nel terzo si illustra il supporto teorico all'analisi moderna delle serie storiche, che viene applicata nel capitolo successivo (mediante la procedura di Box e Jenkins) ai dati in esame.

# **CAPITOLO 1**

## **AGRICOLTURA E IRRIGAZIONE**

### **IN PUGLIA**

#### **1.1 Struttura e caratteristiche dell'agricoltura**

Il territorio regionale si estende per circa 2 milioni di ettari, di cui la parte occupata dai boschi ammonta a soli 120 mila ettari. Questo dato evidenzia come la Puglia sia, tra le regioni italiane, quella a più bassa intensità boschiva, a fronte di una media nazionale pari al 22%; infatti i boschi in Puglia coprono appena il 6% del territorio regionale.

La superficie agricola utilizzata (SAU) costituita dall'insieme dei seminativi, prati permanenti e pascoli, coltivazioni legnose agrarie, orti

familiari e castagneti da frutto, è pari a 1.445.561 ettari e costituisce circa il 76% dell'intera superficie territoriale regionale.

Quasi la metà del suolo regionale è occupata dalle colture arboree da frutto; la Puglia, infatti, pur essendo povera di boschi, si caratterizza per l'elevata presenza delle colture arboree, tra le quali predomina l'olivo. Vigneti, mandorleti, ciliegeti ed agrumeti risultano comunque diffusi in maniera significativa.

Nel complesso, la superficie agricola pugliese, che rappresenta il 21% di quella dell'intero Mezzogiorno ed il 9,7% di quella nazionale, è prevalentemente interessata dalla cerealicoltura e dall'olivicoltura. I cereali, con circa 482 mila ettari investiti (coltivati prevalentemente su superfici agricole che variano tra i 20 e 50 ettari, *Tabella 1*), e l'olivo, con circa 353 mila ettari (coltivato, invece, su superfici agricole più piccole rispetto a quelle dei cereali, che variano tra i 2 e i 5 ettari, *Tabella 2*), occupano circa il 57% e quindi ben oltre la metà della superficie agricola regionale (*Figura 1*). Tra le altre colture più praticate in ordine di superficie investita, un posto di rilievo è occupato dalla vite con circa 153 mila ettari (circa l'11% della SAU, *Figura 1*), anch'essa come l'olivo coltivata principalmente su superfici agricole di piccole dimensioni, che variano fra i 2 e i 5 ettari (*Tabella 2*) e i fruttiferi (fragole, fichi, ciliegie, etc.) con circa 43 mila ettari investiti (circa il 3%

di SAU, *Figura 1*), coltivati in maniera uniforme su tutte le classi di Superficie Agricola Utilizzata, ma prevalentemente su quelle comprese tra i 2 e i 5 ettari (*Tabella 2*); seguono gli ortaggi con circa 37 mila ettari di superficie agricola utilizzata (circa il 3% di SAU, *Figura 1*), anch'essi come i cereali coltivati su superfici agricole comprese tra i 20 e i 50 ettari (*Tabella 1*).

**Tabella 1 - Aziende con seminativi e relativa superficie investita per classe di superficie agricola autorizzata (SAU) e principali coltivazioni praticate per classe di superficie (*superficie in ettari*)**

COLTIVAZIONI	CLASSI DI SUPERFICIE AGRICOLA UTILIZZATA								
	Meno di 1	da 1 a 2	da 2 a 5	da 5 a 10	da 10 a 20	da 20 a 50	da 50 a 100	100 ed oltre	TOTALE
SEMINATIVI	7.950,38	17.411,35	49.475,66	75.130,72	103.450,88	195.943,24	88.493,63	97.337,64	635.193,50
Cereali	4.505,89	11.500,16	39.085,82	59.232,54	73.517,04	150.144,61	68.802,36	75.341,88	482.130,30
Culture proteiche	446,08	523,62	691,55	309,06	1.375,98	1.894,44	1.458,85	142,48	6.797,06
Patata	265,20	392,09	383,30	503,85	54,69	326,68	54,36	103,97	2.048,14
Barbabietola da zucchero	-	176,94	281,48	1.611,58	1.717,61	2.719,67	3.049,62	814,37	10.371,27
Piante sarchiate da foraggio	-	-	-	-	-	-	133,74	34,23	167,97
Piante industriali	77,42	17,74	336,47	757,48	638,07	4.573,26	384,87	208,65	6.993,96
Ortive	1.502,43	2.260,66	4.181,99	4.656,50	8.700,26	9.878,91	2.780,03	3.931,78	37.892,56
Fiori e piante ornamentali	37,30	99,41	152,76	16,06	102,99	-	-	-	408,52
Piantine	14,40	10,46	5,98	0,57	586,40	2,00	14,54	-	634,35
Foraggere avicendate	366,17	845,81	1.690,54	4.043,32	12.132,75	18.952,65	6.945,28	9.376,10	54.352,62
Sementi	10,60	6,54	-	-	-	26,75	-	103,14	147,03
Terreni a riposo	724,89	1.577,92	2.665,77	3.999,76	4.625,09	7.469,27	4.869,98	7.281,04	33.213,72

Fonte: Istituto Nazionale di Statistica (ISTAT), *Struttura e produzioni delle aziende agricole, Anno 2003*

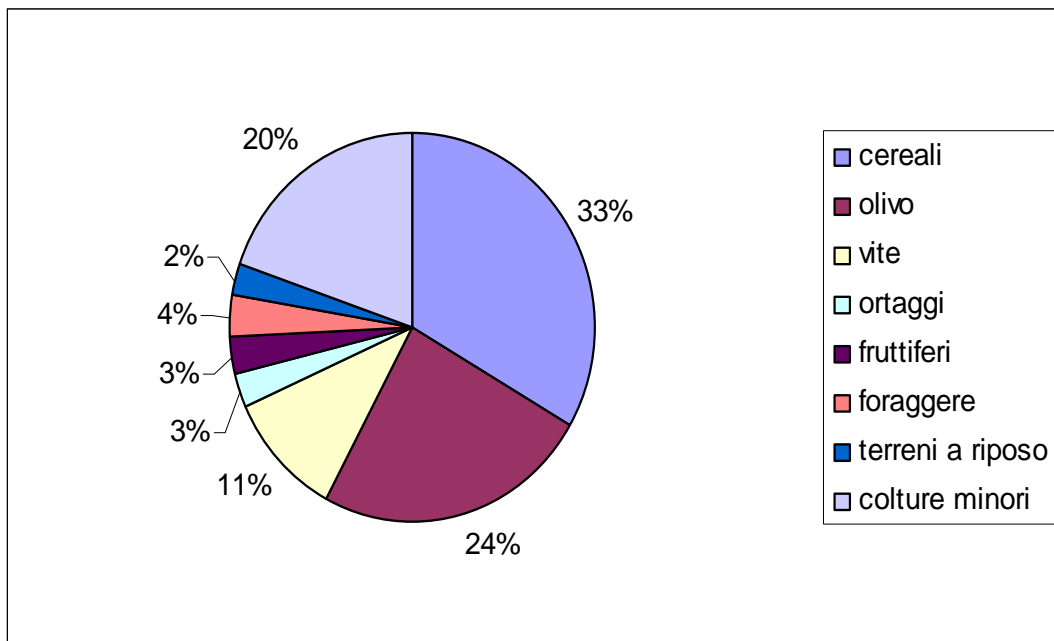
Notevole importanza in termini di superficie agricola utilizzata assumono anche i foraggi (prati ed erbai) con circa 54 mila ettari investiti (circa il 4% di SAU, *Figura 1*) che ricoprono in prevalenza superfici agricole comprese tra i 20 e 50 ettari (*Tabella 1*) ed i terreni a riposo che rivestono principalmente superfici agricole di grandi dimensioni (*Tabella 1*) con circa 33 mila ettari investiti, pari a circa il 2% di SAU (*Figura 1*).

**Tabella 2 - Aziende con seminativi e relativa superficie investita per classe di superficie agricola autorizzata (SAU) e principali coltivazioni praticate per classe di superficie (*superficie in ettari*)**

COLTIVAZIONI	CLASSI DI SUPERFICIE AGRICOLA UTILIZZATA								TOTALE
	Meno di 1	da 1 a 2	da 2 a 5	da 5 a 10	da 10 a 20	da 20 a 50	da 50 a 100	100 ed oltre	
COLTIVAZIONI LEGNOSE AGRARIE	48.815,74	73.241,01	122.590,98	86.670,47	75.722,31	92.860,63	39.070,63	25.280,00	564.251,77
Vite	11.362,04	15.012,49	33.875,60	25.617,77	21.128,63	26.588,13	7.894,57	12.242,10	153.721,33
Olivo	33.607,11	51.862,33	74.494,64	51.928,58	45.868,88	57.509,13	26.298,48	12.102,63	353.671,78
Agrumi	374,15	1.151,42	2.020,02	2.311,38	1.595,49	4.520,34	85,91	557,74	12.616,45
Fruttiferi	3.299,57	5.105,07	12.158,16	6.714,50	7.128,84	4.089,53	4.791,67	337,01	43.624,35
Frutta fresca di origine temperata	1.829,64	2.359,74	5.465,49	3.428,37	1.716,38	1.006,63	1.470,59	234,82	17.511,66
Frutta fresca di origine subtropicale	7,98	15,18	-	-	49,49	72,84	-	-	145,49
Frutta in guscio	1.461,95	2.730,15	6.692,67	3.286,13	5.362,97	3.010,06	3.321,08	102,19	25.967,20
Vivai	99,27	109,70	19,76	98,24	-	125,40	-	-	425,37
Coltivazioni legnose agrarie in serra	-	-	-	-	-	0,50	-	-	0,50
Altre coltivazioni legnose agrarie	73,60	-	22,80	-	0,47	27,60	-	40,52	164,99

Fonte: Istituto Nazionale di Statistica (ISTAT), *Struttura e produzioni delle aziende agricole, Anno 2003*

**Figura 1 – Suddivisione della Superficie Agricola Utilizzata (SAU)  
in Puglia (2003)**



Dalla *Tabella 3* si evince, inoltre, che le aziende agricole pugliesi sono poco meno di 286.000 (il 23% delle aziende del Mezzogiorno) e che continua ad essere evidente il processo di senilizzazione dell'imprenditoria agricola, con una spiccata prevalenza di conduttori con oltre 54 anni (circa il 66% di tutti gli addetti) ed una ancora limitata presenza di giovani imprenditori (appena il 7,5% ha meno di 40 anni). Inoltre, dalla *Tabella 3*, si evidenzia che in Puglia, la percentuale di occupati nel settore agricolo è pari all'11,42% di quelli totali, rappresentando un valore superiore, non solo, a quella degli occupati

agricoli in Italia (5,48% degli occupati totali) ma, anche, alla percentuale occupati agricoli nel Mezzogiorno (9,65% degli occupati totali).

**Tabella 3 - Indicatori strutturali delle aziende per orientamento produttivo dominante (1999)**

DESCRIZIONE	UNITÀ DI MISURA	PUGLIA	MEZZOGIORNO	ITALIA
<b>Aziende</b>	<i>numero</i>	285.852	1.244.952	2.269.574
<i>per classi di età del conduttore</i>				
< 40 anni	%	7,50	7,99	8,89
tra 40 e 54 anni	%	26,18	25,88	25,74
> 54 anni	%	66,32	66,13	65,37
<b>Occupati agricoli/occupati totali</b>	%	11,42	9,65	5,48

*Fonte: Istituto Nazionale di Economia Agraria (INEA); Osservatorio Puglia sul mondo rurale e sul sistema agroindustriale 2000-2002 ;elaborazioni su dati ISTAT*

## 1.2 Distribuzione territoriale delle colture

La Puglia presenta aspetti produttivi e paesaggistici del territorio rurale estremamente diversificati.

L'uomo nel corso dell'attività agricola e a fini produttivi è intervenuto sul paesaggio naturale, trasformandolo e rimodellandolo in funzione delle mutevoli esigenze produttive.

L'attuale territorio rurale pugliese è caratterizzato da vari sottotipi di paesaggio agrario, dai più antichi (come gli oliveti a terrazzamento del Gargano) ai più recenti (come i tendoni dell'uva da tavola del Sud Barese e dell'Arco Ionico), dal più incontaminato (pascoli dell'Alta



Murgia, Fossa Bradanica e certe aree del foggiano), a quello delle piantagioni di olivo che, sterminato e diversificato caratterizza gran parte del territorio rurale pugliese. Considerando questa estrema variabilità del paesaggio pugliese, risulta difficile rappresentarlo o descriverlo sinteticamente.

Parallelamente al paesaggio agrario, anche la produttività del suolo si presenta estremamente variabile in Puglia. Le aree meno fertili (Murgia e Gargano, Appennino Dauno e parte del Salento) sono caratterizzate da un'agricoltura estensiva e da fenomeni di abbandono; i terreni fertili, particolarmente vocati all'attività agricola (il Tavoliere e Valle dell'Ofanto, arco Ionico-Tarantino, ecc.) presentano un'agricoltura di tipo intensivo, molto spesso irrigua.

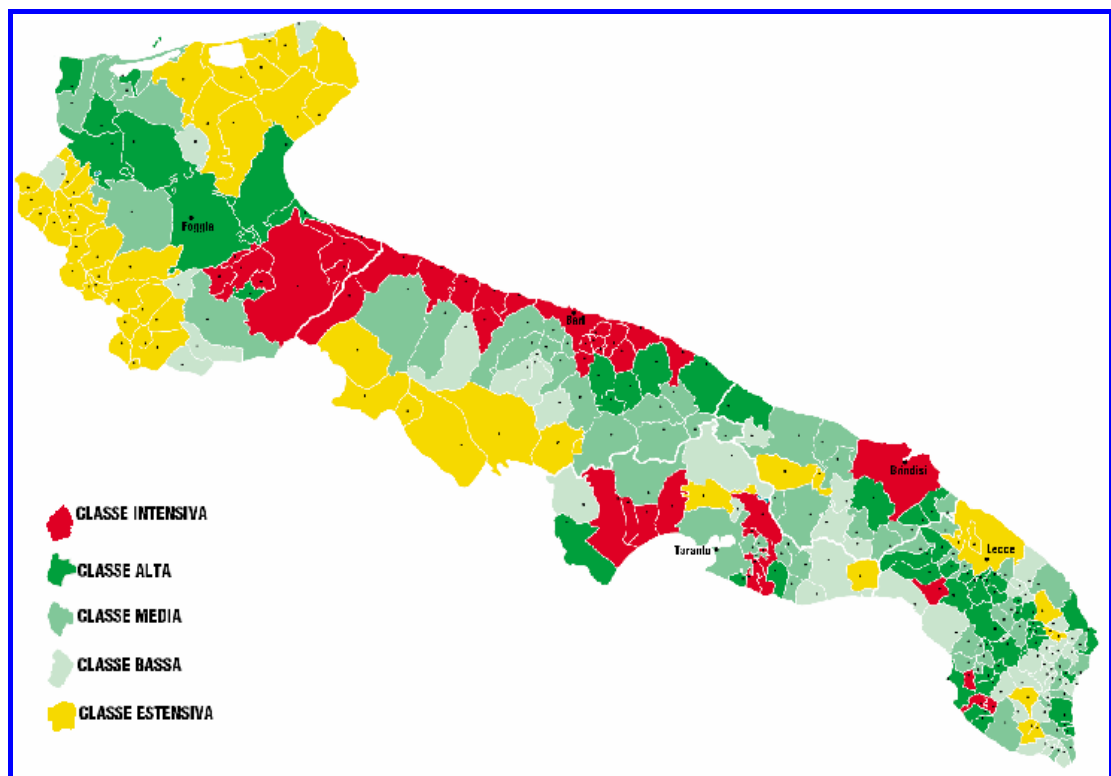
Lo studio INEA (2001), relativo alla produttività di tutti i comuni pugliesi (*Figura 2*), evidenzia quanto segue:

- i territori comunali con più elevata produttività agricola (classe di produttività intensiva ed alta) sono quelli del Fortore, del Tavoliere, della Valle dell'Ofanto, del litorale barese e brindisino, dell'arco Ionico-Tarantino ed alcuni comuni del leccese; sono gli stessi comuni le cui superfici sono interessate all'irrigazione;
- i territori comunali a bassa produttività agricola (classe di produttività bassa ed estensiva) sono quelli della Murgia, del

Gargano, dell' Appennino Dauno e delle Serre Salentine;

- le aree a produttività media sono quelle della fascia centrale interna della provincia di Bari, Brindisi e Taranto.

**Figura 2 - I comuni della Puglia per classi di produttività agricola (2001)**



*Fonte: Istituto Nazionale di Economia Agraria (INEA)*

Nella Figura 2 vengono indicate le seguenti classi:

- classe intensiva, quando la produttività comunale supera quella media regionale di oltre il 35%;
- classe alta, quando la produttività comunale supera quella media regionale di oltre il 10% fino al 35% compreso;

- classe media, quando la produttività comunale è compresa tra il +10% e il -10% di quella media regionale;
- classe bassa, quando la produttività comunale è inferiore a quella media regionale di oltre il 10% fino al 35% compreso;
- classe estensiva, quando la produttività comunale è più bassa di quella media regionale di oltre il 35%.
- lo scarto di produttività media tra i comuni rientranti nelle classi di produttività estreme estensiva ed intensiva è del rapporto di 1 a 7.

### **1.3 I Consorzi di Bonifica**

La descrizione delle caratteristiche e dell'importanza dell'agricoltura irrigua pugliese non può prescindere dall'analisi delle diverse realtà territoriali, produttive e socio economiche presenti in Puglia, anche alla luce della considerazione che ormai l'agricoltura non può essere più vista come un settore a sé stante, ma va collocata in un ambito più ampio di intersettorialità, che tenga conto delle molteplici e multiformi realtà economico-culturali che caratterizzano lo scenario di riferimento.

Sulla base di queste considerazioni è possibile individuare alcuni contesti territoriali ben definiti, identificabili in massima parte con i

territori delimitati e serviti dai Consorzi di Bonifica della regione, ciascuno dei quali presenta caratteristiche agro-climatiche distinte e dinamiche socio-economiche e produttive tali da permetterne un'agevole classificazione. I comprensori consortili sono, infatti, molto diversi fra loro per estensione, per caratteristiche ed utilizzazione della superficie agricola.

Da un punto di vista strutturale, la superficie territoriale dei sei Consorzi di Bonifica (del Gargano, della Capitanata, Terre d'Apulia, Stornara e Tara, Arneo e Ugento Li Foggi) della Puglia è pari a 1.743.591 ettari, praticamente quasi tutto il territorio regionale. Restano fuori dai confini amministrativi dei Consorzi parte dell'Appennino Dauno e alcuni comuni della provincia di Brindisi.

Il Consorzio con la maggiore superficie amministrativa è quello di Terre d'Apulia, con poco meno di 570 mila ettari; il più piccolo è Stornara e Tara con all'incirca 143 mila ettari. Il Gargano si caratterizza, tra i Consorzi pugliesi, per la maggiore incidenza della superficie forestale, il 23%. Il Consorzio della Capitanata presenta, invece, alta intensità irrigua, elevata produttività dei terreni ed aziende agricole dall'ampiezza medio-alta, spesso superiore ai 20 ettari. Al contrario, nel territorio consortile di Ugento è presente una frammentazione fondiaria

molto accentuata: oltre il 90% delle ditte presenta infatti una superficie inferiore ai 2 ettari.

Con le risorse idriche gestite dai Consorzi di Bonifica si irrigano 75.517 ettari di superficie. Tra le diverse colture irrigue, quelle con maggiore superficie irrigata (con acqua gestita dai Consorzi) sono prevalentemente la vite, le colture orticole e l'olivo. Questo fenomeno riguarda sia la superficie irrigata complessivamente sia quella ricadente nell'ambito del singolo Consorzio, con la sola eccezione del Consorzio di Stornara e Tara, dove l'agrumeto è la seconda coltura irrigua più estesa. Vite, orticole e olivo, insieme, costituiscono i 4/5 di tutta la superficie irrigata con la rete irrigua consortile.

La produzione delle colture che si praticano solo con l'irrigazione (pomodoro, uva da tavola, carciofo, agrumi, ecc.) rappresenta il 43% del valore di tutta la produzione agricola vegetale della regione.

Di seguito si riportano le caratteristiche dell'agricoltura nei territori dei singoli Consorzi di Bonifica.

### **1.3.1 Consorzio di Bonifica del Gargano**

Il Promontorio del Gargano, estrema propaggine nord-orientale della Puglia, rientra nell'omonimo Consorzio di Bonifica, che

comprende 13 comuni, per una superficie territoriale pari a 150.337 ettari (*Tabella 4*).

Come si è già detto, è il Consorzio con la più elevata incidenza di superficie forestale (all'incirca 34.000 ettari), pari al 23% della superficie territoriale.

La SAU (superficie agricola utilizzata) è occupata in prevalenza da prati e pascoli.

In questo comprensorio l'agricoltura irrigua riveste importanza marginale, in quanto le attività agricole sono principalmente imperniate sull'allevamento bovino ed ovino che utilizzano, per lo più i pascoli demaniali, e su un'olivicoltura poco specializzata ed a bassa produttività.

Le aree interne sono caratterizzate da una bassa densità di popolazione e da una elevata incidenza degli attivi agricoli sul totale, in virtù della limitata presenza di attività alternative a quella agricola.

Le aree interessate all'irrigazione sono modeste e circoscritte a pochi comuni.

Tra le colture irrigue figurano gli agrumeti, gli ortaggi, e gli oliveti nei comuni del Versante Adriatico; la barbabietola da zucchero ed il pomodoro nei terreni adiacenti al Tavoliere.

Complessivamente la superficie irrigata non supera i 500 ettari (*Tabella 4*).

### **1.3.2 Consorzio di Bonifica della Capitanata**

Nel Consorzio ricade un territorio comprendente la Piana del Tavoliere e le propaggini estreme dei comuni dell'Appennino Dauno e del Gargano, al confine con il Tavoliere. Esso è identificato con l'area di pianura della provincia di Foggia, tradizionalmente destinata alla coltivazione di frumento duro, che un tempo costituiva la coltura principale e che oggi si coltiva in rotazione alle colture industriali e a quelle orticole.

Le aziende agricole presentano una dimensione media superiore a 20 ettari, non sono rare inoltre aziende con più di 100 ettari. La tipologia aziendale, le colture ed anche la pratica irrigua non sono distribuite uniformemente su tutto il Tavoliere, ma cambiano man mano che ci si sposta dal Nord al Sud, dall'Alto al Medio al Basso Tavoliere.

Nell'Alto Tavoliere prevalgono le aziende dall'ampiezza media di 20 ettari.

Il frumento duro occupa più della metà di tutta la superficie irrigua della zona, mentre tra le colture irrigue prevalgono la barbabietola da zucchero, il pomodoro e la vite per uva da vino.

Nella zona di Lesina sono più diffuse, invece, le orticole. La fonte di approvvigionamento prevalente è quella pubblica gestita dal Consorzio.

Anche nel Tavoliere Centrale, che comprende il territorio che ha per estremi il comune di San Severo e il Torrente Carapelle, prevalgono aziende ad ampiezza elevata. La densità e l'estensione della coltura, infatti, sono tali da farne il più grande centro produttivo di pomodoro dell'Italia e dell'Unione Europea.

Quest'area si caratterizza per la scarsa presenza dell'irrigazione con fonte idrica pubblica e la straordinaria diffusione dell'irrigazione basata sullo sfruttamento dei pozzi privati.

Il Basso Tavoliere comprende la parte più meridionale del Tavoliere, che dal Torrente Carapelle si estende fino alla sinistra dell'Ofanto. In questa parte del Tavoliere prevalgono le colture arboree (olivo da tavola e da olio, vite da vino e da tavola, pescheti) e gli ortaggi, soprattutto carciofi e asparagi, finocchi e broccoletti.

La maglia aziendale risulta decisamente ridotta (5 ettari mediamente).

Per l'irrigazione, molto diffusa nell'area, si utilizza prevalentemente la risorsa idrica fornita dal Consorzio della Capitanata.

In tutto il Tavoliere, si irrigano all'incirca 54.000 ettari con acque gestite dal Consorzio.

In Capitanata si concentra ben il 70% di tutta la superficie regionale irrigata con rete idrica pubblica.



### 1.3.3 Consorzio di Bonifica Terre d'Apulia

Si tratta del Consorzio più esteso della Puglia, comprendendo tutta la provincia di Bari e 6 comuni della provincia di Taranto, per una superficie complessiva di circa 570 mila ettari (*Tabella 4*). Nel vasto territorio consortile sono delimitabili tre aree omogenee, che grosso modo si equivalgono per estensione:

- il litorale barese ed il comprensorio irriguo di Minervino-Loconia, dove l'irrigazione viene praticata diffusamente (area irrigua);
- la Murgia (Alta e Bassa) che da Minervino (Bari) si estende fino a Mottola (Taranto), dove la suscettività alla trasformazione irrigua è bassa o nulla (terreni non idonei all'irrigazione);
- due fasce adiacenti alla dorsale dell'Alta Murgia, ovvero la fascia orientale verso il mare e la fascia occidentale (Fossa Bradanica) a ridosso del confine con la Basilicata (area suscettibile all'irrigazione).

Le caratteristiche agricole delle tre aree vengono descritte nel seguente modo:

- *area irrigua* - comprende il litorale barese, che si estende (per circa 150 mila ettari) lungo l'Adriatico dalla foce dell'Ofanto ai confini della provincia di Brindisi, ed il comprensorio irriguo di Minervino-Loconia, esteso per 12 mila ettari. In tutta l'area è diffusa l'irrigazione, che si pratica esclusivamente con l'acqua dei pozzi

privati situati lungo il litorale e con l'acqua derivata dalla diga del Locone, nel comprensorio irriguo di Minervino-Loconia.

In tutto il territorio irriguo la tipologia produttiva più diffusa è quella arboricola, sia specializzata che mista, quella orticola specializzata (carciofi, patate, insalate) ed anche l'olivicola-orticola.

In tutta l'area si rileva una frammentazione fondiaria diffusa con una maglia aziendale media tra i 2 e i 5 ettari.

- area non suscettibile alla trasformazione irrigua – il vasto territorio della Murgia, che da Minervino si estende per tutta la lunghezza della provincia di Bari e continua oltre il confine della provincia di Taranto fino a Mottola per una superficie di circa 200 mila ettari, non è idonea alla trasformazione irrigua. Esso è destinato prevalentemente a pascolo di scarsa produttività ed ai cereali nell'Alta Murgia, mentre nella Murgia Sud-Orientale prevalgono le foraggere coltivate in appezzamenti adiacenti ai boschi di quercia o ai pascoli arborati. Nella Bassa Murgia barese e tarantina è concentrato il grosso dell'allevamento bovino della Puglia, mentre nell'Alta Murgia sono presenti prevalentemente gli allevamenti ovini, ormai in fase di contrazione.
- area suscettibile alla trasformazione irrigua – comprende le due fasce di territorio adiacenti alla dorsale dell'Alta Murgia: quella

occidentale di 80 mila ettari e quella orientale avente una estensione di 120 mila ettari. Le due sub-aree sono molto diverse per colture praticate e per tipologia aziendale. Nella fascia interna, denominata Fossa Bradanica, prevale la coltivazione del frumento duro in aziende cerealicole specializzate, dall'ampiezza medio-grande (che varia tra i 20 e i 50 ettari). Nella fascia orientale, collocata tra il litorale e l'Alta Murgia, prevalgono le colture arboree (olivo, mandorlo, ciliegio e vite da vino). Le aziende sono arboricole miste e olivicole specializzate. Anche in quest'aria, come nel litorale barese, la proprietà fondiaria è assai frazionata. L'irrigazione è sporadica e, là dove è presente, l'acqua viene prelevata da pozzi privati trivellati a qualche centinaio di metro di profondità.

#### **1.3.4 Consorzio di Bonifica Stornara e Tara**

Si tratta del più piccolo Consorzio della Puglia, con circa 143 mila ettari di superficie territoriale (*Tabella 4*). Fatta eccezione per la parte più alta dei comuni della Murgia Sud-Orientale, tutto il territorio del Consorzio è interessato all'irrigazione.

Tra le colture praticate prevalgono le arboree e quelle erbacee. I prati e i pascoli interessano l'8% della SAU totale.

Quest'area si caratterizza in particolare per:

- la presenza degli agrumeti (la maggior parte di quelli coltivati in Puglia);
- la più elevata densità vinicola della regione (il 30% di tutta la SAU è investita a vigneto sia da vino che da tavola).

La frammentazione fondiaria non è molto accentuata. La dimensione media più diffusa è compresa tra i 5 e i 10 ettari.

In tutto il territorio si irrigano all'incirca 13 mila ettari utilizzando acque gestite dal Consorzio di Bonifica.

La coltura maggiormente irrigata è la vite, seguita da agrumi (sia quelli specializzati che quelli consociati con l'olivo), olivo e orticole.

### **1.3.5 Consorzio di Bonifica Arneo**

Il territorio consortile è esteso per circa 249 mila ettari e comprende i comuni di tre province: Lecce, Brindisi e Taranto (*Tabella 4*).

Le caratteristiche più rilevanti dell'utilizzazione del territorio sono:

- la scarsa presenza di boschi;
- la prevalente utilizzazione arbicola della superficie agricola;
- la presenza di aree di rilevanza naturalistica.

I boschi occupano appena 1 ettaro ogni 100 di superficie territoriale, mentre le colture arboree, oliveti in prevalenza, occupano all'incirca 67 ettari ogni 100 di SAU consortile.

All'interno del territorio consortile sono individuabili tre aree omogenee in funzione della concentrazione delle colture praticate:

- area settentrionale adriatica: occupa la parte nord-occidentale del Consorzio e comprende alcuni comuni della provincia di Brindisi. L'aspetto colturale più rilevante di quest'area è dato dalla diffusione delle colture orticole, che occupano il 50% della coltura di superficie a seminativi. Quest'area, infatti, è nota per la produzione del carciofo e del pomodoro da industria;
- area centrale: include la parte centrale del territorio consortile, a cavallo dei confini delle tre province comprendenti alcuni comuni della provincia di Brindisi, altri della provincia di Taranto ed altri ancora della provincia di Lecce. In quest'area i 3/4 della SAU sono occupati dalla vite e dall'olivo le cui superfici si equivalgono. La vite è esclusivamente dedicata alla produzione di vino;
- area meridionale jonica: comprende i comuni della provincia di Lecce che si affacciano sul Mar Ionio (Porto Cesario, Nardò, Galatone, Sannicola) e quelli interni a loro confinanti. La superficie agricola viene utilizzata per metà dalle colture arboree e per l'altra metà dalle colture erbacee. Tra le specie erbacee, risultano più coltivate le patate, gli ortaggi estivi e l'anguria. È, inoltre, diffusa la floricoltura, in pieno campo ed in serra (Leverano e comuni limitrofi).

In definitiva il territorio di questo Consorzio presenta tre specificità in termini di utilizzazione agricola della superficie:

- la floricoltura di Leverano;
- l'elevata concentrazione della vite da vino nell'area centrale;
- l'elevata concentrazione orticola (pomodoro e carciofo) nei comuni brindisini dell'area adriatica.

Infine in questo Consorzio è molto diffuso il prelievo delle acque attraverso pozzi privati, per cui la superficie irrigua complessiva è di gran lunga superiore ai 1.127 ettari rilevati come irrigati con acqua gestita da pubblici consorzi.

### **1.3.6 Consorzio di Bonifica Ugento Li Foggi**

Il territorio consortile si estende per circa 189 mila ettari e comprende 78 comuni della provincia di Lecce (*Tabella 4*).

Le caratteristiche territoriali di questo Consorzio sono molto simili a quelle del Consorzio dell'Arneo, per le seguenti motivazioni:

- assenza di boschi;
- presenza di aree di interesse naturalistico;
- prevalenza delle colture arboree nell'ambito della superficie agricola utilizzata.

Tra le colture arboree prevale l'olivicoltura tradizionale, costituita

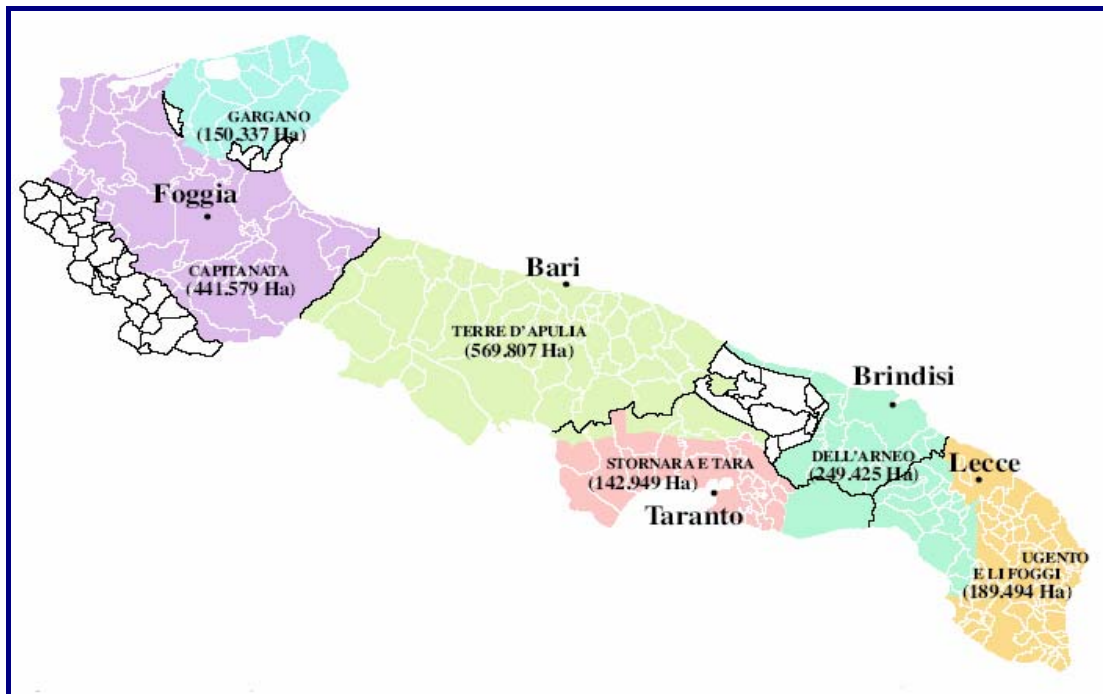
da piante ultra-secolari. La viticoltura, esclusivamente destinata alla produzione del vino, è meno diffusa di un tempo ed è concentrata in alcuni comuni come Alezio, Matino e Ugento. Tra le colture erbacee prevalgono i cereali, le orticole ed il tabacco. Persiste, infatti, la tabacchicoltura, condotta prevalentemente con manodopera familiare. Per quanto riguarda la dimensione aziendale, l'area si caratterizza per una spinta frammentazione fondiaria: su circa 198 mila ditte iscritte al catasto consortile, il 90% presenta estensione inferiore a due ettari ed occupa circa la metà della superficie consortile. La fonte di approvvigionamento idrica più diffusa è quella dei pozzi privati. Con la rete irrigua consortile si irrigano soltanto 2.120 ettari (*Tabella 4*).

**Tabella 4 – Consorzi di Bonifica in Puglia (2004)**

Consorzi di Bonifica	Province di riferimento del consorzio	Superficie Amministrativa (Ha)	Superficie Attrezzata (Ha)	Superficie Irrigata (Ha)	Comuni
GARGANO	Foggia	150.337	570	428	13
CAPITANATA	Foggia	441.579	140.378	53.667	39
TERRE D'APULIA	Bari, Taranto	569.807	17.645	4.972	54
STORNARA E TARA	Taranto	142.949	43.705	13.203	23
ARNEO	Brindisi, Lecce, Taranto	249.425	18.552	1.127	48
UGENTO E LI FOGGI	Lecce	189.494	14.685	2.120	78
	Totale	1.743.591	235.535	75.517	255

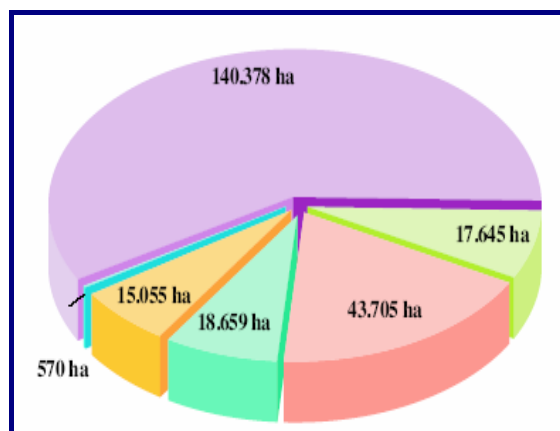
Fonti: Istituto Nazionale di Economia Agraria (INEA) e Associazione Nazionale dei Biotecnologi Italiani (ANBI)

**Figura 3 - Consorzi di Bonifica in Puglia: confini amministrativi e superficie territoriale (2004)**



Fonte: Istituto Nazionale di Economia Agraria (INEA)

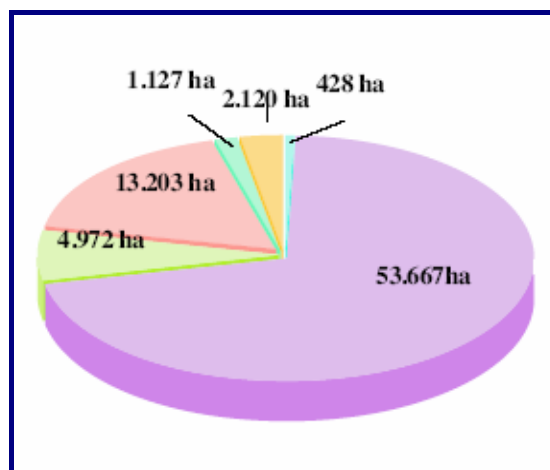
**Figura 4 – Superficie attrezzata per singolo Consorzio (2004)**



Fonte: Istituto Nazionale di Economia Agraria (INEA)



**Figura 5 – Superficie irrigata per singolo Consorzio (2004)**



*Fonte: Istituto Nazionale di Economia Agraria (INEA)*

## **1.4 Acque sotterranee e sorgive**

Il territorio, eccezione fatta per il Tavoliere, ha natura prevalentemente carsica e ciò rende la regione estremamente povera di acque superficiali. Nel contempo sono presenti rilevanti risorse idriche sotterranee che hanno incentivato in vaste aree la vocazione prevalentemente agricola.

In effetti, nelle zone pianeggianti prossime alla costa, la falda di acqua dolce si trova a poca profondità, per cui nel passato è stato possibile raggiungerla anche mediante pozzi a scavo. Ciò ha consentito che, in prossimità dei litorali del barese, a Nord (fino a Trani-Barletta) e

a Sud (Polignano-Monopoli) e nei territori pianeggianti del Salento, si siano diffuse coltivazioni di tipo ortivo, notoriamente idroesigenti.

Il sollevamento dell'acqua, inizialmente praticato con le "norie", azionate da asini o muli che, con gli occhi bendati, giravano torno-torno per intere giornate, nella seconda metà del secolo scorso ha cominciato ad avvalersi di pompe di aspirazione che, posizionate a livelli adeguati, hanno consentito di catturare falde sempre più profonde; successivamente le tecniche di estrazione si sono notevolmente perfezionate fino alle attuali possibilità di perforare fino a profondità di alcune centinaia di metri, pozzi di piccolo diametro che, attrezzati con pompe sommerse, ormai consentono in tutta la regione di emungere dalla falda più profonda.

In tal modo però è stata favorito un indiscriminato sfruttamento e, talvolta, la dissipazione della preziosa risorsa con conseguenze notevoli sul naturale assetto della circolazione sotterranea, provocando, tra l'altro, l'inaridimento di alcune sorgenti millenarie.

Si evidenziano, inoltre, i pericoli a cui sono esposte le acque di falda per effetto dell'inquinamento derivante dall'intensa attività antropica presente nelle aree di alimentazione e per effetto della progressiva salinizzazione conseguente all'intrusione della sottostante acqua marina, causata dall'incondizionato emuginamento. Ai prelievi per

usi potabili durante tutto l'anno, si sommano nei mesi più caldi i prelievi per l'irrigazione, che danno luogo a veri e propri *shock*.

Fra le principali manifestazioni sorgive si possono ricordare quelle che bordano il Gargano di cui, alcune alimentano i laghi di Lesina e Varano, altre confluiscono direttamente in mare; altre sorgenti si rinvengono nel Subappennino. Tutte queste hanno portate assai modeste dell'ordine di pochi litri al secondo (l/s).

Le Murge, come il Gargano, hanno emergenze carsiche sottocosta: fra tutte va ricordata quella di Torre Canne che alimenta le antiche e famose terme.

Il Salento, infine, è ricco di sorgenti: nel tarantino alcune contornano il Mar Piccolo e, pur avendo buone portate (la sorgente Galese arriva anche oltre 500 l/s) non sono adeguatamente sfruttate perché spesso sono sottomarine o salmastre. Importanti sono le sorgenti salmastre dell'Idume (minimo 700 l/s) e del Chidro presso Manduria (anche 3000 l/s come portata massima) e quelle termali di Santa Cesarea Terme a sud di Otranto. Alcune di queste determinano la formazione di piccoli laghi e di brevi corsi d'acqua<sup>1</sup>.

---

<sup>1</sup> AA.VV., *Atlante Geografico De Agostini*, Istituto Geografico De Agostini S.p.A., Novara, 2003.

## 1.5 Corsi d'acqua superficiali

Il reticolo idrografico superficiale del territorio pugliese è scarsamente sviluppato a causa della natura fondamentale calcarea dei terreni; solo nella zona pedegarganica e del Tavoliere la minore permeabilità del suolo ha consentito la formazione di alcuni corsi d'acqua, che invece nelle altre aree sono praticamente assenti.

Fra i fiumi, il più importante è l'Ofanto che nasce presso Nusco, in Irpinia, e dopo circa 165 Km si versa nell'Adriatico, a nord di Barletta. Esso attraversa la regione per un tratto della lunghezza di circa 50 Km, inferiore a quella dei corsi d'acqua che solcano il Tavoliere che sono il Candelaro (70 Km), il Sosola (60 Km), il Cervaro (80 Km), il Carapelle (85 Km) e il Celone (59 Km). Altri corsi d'acqua di interesse regionale sono il Fortore (86 Km, di cui 25 Km in Puglia), il Lato ed il Galese nel tarantino, il Canale Reale presso Brindisi. Non viene considerato il Bardano che scorre quasi per intero in Basilicata.

Le portate medie di questi corsi sono assai esigue; il regime è fortemente irregolare e torrentizio caratterizzato da magre estive e da piene autunnali-invernali, che in passato hanno anche dato luogo a rovinose esondazioni.

Il bacino idrografico più esteso risulta quello dell'Ofanto, il cui corso segna grosso modo il confine fra le due unità morfologico-

strutturali dell'Altopiano della Murgia e del Bassopiano del Tavoliere di Foggia. Tale bacino interessa il territorio di tre regioni, Campania, Basilicata e Puglia ed ha una forma pressoché trapezoidale, una superficie di 2.790 Km<sup>2</sup> e un'altitudine media di 450 m.

I corsi d'acqua affluenti del fiume Ofanto si sviluppano in un ambiente geologico e morfostrutturale chiaramente appenninico, con rare eccezioni (per esempio il Torrente Locone)<sup>2</sup>.

## **1.6 Schemi idrici ad uso plurimo esistenti**

Per schema idrico si intende l'insieme di grandi opere idrauliche mediante le quali è possibile realizzare un collegamento tra fonti di approvvigionamento (sorgenti, laghi naturali o artificiali, corsi d'acqua, pozzi, etc.) e utilizzatori finali delle risorse idriche (per uso potabile, irriguo e industriale).

I principali schemi idrici interregionali ad uso plurimo sono:

- Schema del Fortore (Puglia e Molise): interessa i territori ubicati nei comprensori irrigui del Fortore (Consorzio della Capitanata in Puglia e Consorzio Larinese in Molise).

---

<sup>2</sup> AA.VV., *Atlante Geografico...ecc.*, op. cit..

Le fonti di approvvigionamento sono costituite dall'invaso di Occhito, sul fiume Fortore, e dalla diga del Celone sull'omonimo torrente, attualmente in fase di invaso sperimentale;

- Schema dell'Ofanto (Campania, Basilicata e Puglia): presenta cinque invasi: Conza e Osento in Campania, Rendina in Basilicata, Marana Capacciotti e Locone in Puglia.

Le risorse idriche rese disponibili dallo schema soddisfano i fabbisogni potabile, irriguo ed industriali dei territori lucani e pugliesi del medio e basso Ofanto;

- Schema Jonico-Sinni (Basilicata, Puglia e Calabria): interessa un territorio molto vasto, comprendente l'arco Jonico Lucano e Pugliese, il Salento ed in parte, la zona jonica calabrese. L'utilizzazione della risorsa idrica resa disponibile dallo schema interessa gli usi potabile, agricolo ed industriale.

Questi schemi sono tutti di carattere interregionale, interessando oltre alla Puglia, anche le regioni limitrofe, in particolare la Basilicata.

Ciò trova giustificazione nelle caratteristiche idromorfologiche della regione che, come si è precedentemente detto, impediscono la formazione di un adeguato reticolo di acque superficiali, capace di soddisfare i fabbisogni potabili, irrigui ed industriali.

Questa situazione ha reso indispensabile la messa a punto di grandi schemi idrici interregionali, che hanno il compito di rendere disponibili risorse sufficienti a soddisfare i fabbisogni idrici della Puglia e delle regioni limitrofe (Basilicata e Molise, soprattutto). Ad incidere fortemente sul deficit di disponibilità idrica pugliese è in particolare il settore primario: infatti la presenza di territori fortemente suscettibili all'irrigazione e l'elevata capacità da parte degli operatori agricoli di adottare ordinamenti produttivi e tecniche innovative e remunerative, ha portato l'agricoltura pugliese, soprattutto grazie alle colture irrigue (vite, ortaggi, etc.) a conquistare un posto di rilievo.

Vi sono anche schemi idrici minori di interesse regionale, alcuni dei quali sono ancora in fase di progettazione, quali il Carapelle e il Gravina-Pentecchia<sup>3</sup>.

## **1.7 Caratteristiche e struttura dell'irrigazione**

In base a quanto detto, si può affermare che la pratica irrigua è molto diffusa nell'agricoltura pugliese. Fatta eccezione per alcune aree (Gargano, Alta Murgia e Appennino Dauno), tutto il territorio regionale

---

<sup>3</sup> Istituto Nazionale di Economia Agraria (INEA), *Lo stato dell'irrigazione in Puglia*, volume 1, 2001

con un diverso grado di intensità ne è interessato.

Il territorio agricolo pugliese, in un contesto di carenza generalizzata di precipitazioni e di corsi di acqua superficiali, presenta una buona (o moderatamente buona) attitudine dei suoli all'irrigazione. Infatti è stato stimato che i suoi agricoli adatti all'irrigazione sono estesi per 735.292 ettari, in buona parte non ancora utilizzati per la pratica irrigua. Inoltre il territorio, povero di acque superficiali, presenta una grande ricchezza di acque sotterranee di falda, prelevabili con una relativa facilità e, in qualche area, anche da pochi metri di profondità.

Sul territorio agricolo regionale c'è un'ampia diffusione di colture intensive ad alto reddito, quali il pomodoro, il carciofo, l'uva da tavola, l'oliveto intensivo, gli agrumi, il pescheto, ecc., per la cui coltivazione si ricorre all'irrigazione.

Si è pertanto determinata una forte domanda di acqua irrigua (consumi irrigui), che presenta un andamento temporale crescente, in quanto i coltivatori che hanno acquistato una straordinaria capacità tecnica, sono portati ad ampliare le aree irrigue, sia aumentando la superficie delle colture irrigate ad alto reddito, sia estendendo l'irrigazione alle colture come l'oliveto, tradizionalmente in asciutto.

Attualmente il quadro di riferimento dell'irrigazione regionale è costituito dai seguenti dati caratteristici e strutturali:



- superficie irrigata complessivamente nella regione: 362.130 ettari (di cui 352.459 ettari nei territori amministrati dai Consorzi);
- superficie irrigata con rete irrigua consortile: 75.517 ettari;
- fabbisogno idrico irriguo complessivo: 812.043.526 m<sup>3</sup>;
- disponibilità di risorsa idrica pubblica per l'irrigazione: 274.100.000 m<sup>3</sup>;
- superficie complessivamente attrezzata con rete di distribuzione pubblica della risorsa idrica: 235.535 ettari (di cui in esercizio 187.129 ettari)<sup>4</sup>.

Si noti che per fabbisogno irriguo si intende il volume di acqua necessario per il soddisfacimento idrico di un dato sistema colturale.

La risorsa idrica per l'irrigazione è gestita sia dagli enti pubblici che dai privati. Gli enti pubblici, ossia i Consorzi di Bonifica, gestiscono tutta la risorsa idrica prelevata dalle sorgenti, dagli invasi e da numerosi pozzi che prelevano acqua di falda. I privati, quasi sempre agricoltori (singoli o associati), gestiscono e prelevano l'acqua esclusivamente dai pozzi che essi stessi hanno trivellato. In Puglia, pertanto, la risorsa idrica è gestita dai Consorzi di Bonifica (irrigazione pubblica). Accanto a

---

<sup>4</sup> Studio condotto dall'Istituto Nazionale di Economia Agraria (INEA) e dall'Associazione Nazionale dei Biotecnologi Italiani (ANBI) nel 2004.

questa coesiste un'irrigazione privata, tramite pozzi gestiti dagli agricoltori. Il fenomeno di irrigazione da pozzi privati assume una dimensione rilevante. Esso costituisce la parte più ampia dell'irrigazione regionale. I pozzi irrigui privati sono distribuiti su tutto il territorio regionale, con una densità variabile in funzione dei fattori incentivanti del fenomeno: l'assenza di acqua pubblica e il costo e la facilità di trivellazione. Essi sono più diffusi dove manca la rete pubblica irrigua, ma sono anche presenti all'interno dei comprensori irrigui consortili, dove la rete irrigua è disattivata o non pienamente funzionante.

Dai dati riportati emergono con chiarezza quali sono gli aspetti caratterizzanti l'irrigazione regionale, che sono sintetizzabili nella seguente maniera:

- ▶ ***penuria di risorsa idrica per l'irrigazione:*** la disponibilità complessiva è 274,1 Mm<sup>3</sup>. Il confronto di questi volumi disponibili con i fabbisogni idrici complessivi (812 Mm<sup>3</sup>) richiesti dalle colture irrigue che si praticano nella regione evidenzia l'entità della penuria di acqua pubblica per l'irrigazione. In altri termini, si può ritenere che circa i 2/3 degli attuali fabbisogni irrigui delle colture vengono soddisfatti con l'acqua da pozzi privati. Dalla *Figura 6*, inoltre, si evince che la risorsa idrica pubblica disponibile per l'irrigazione proviene principalmente (circa 80%) dagli invasi regionali ed

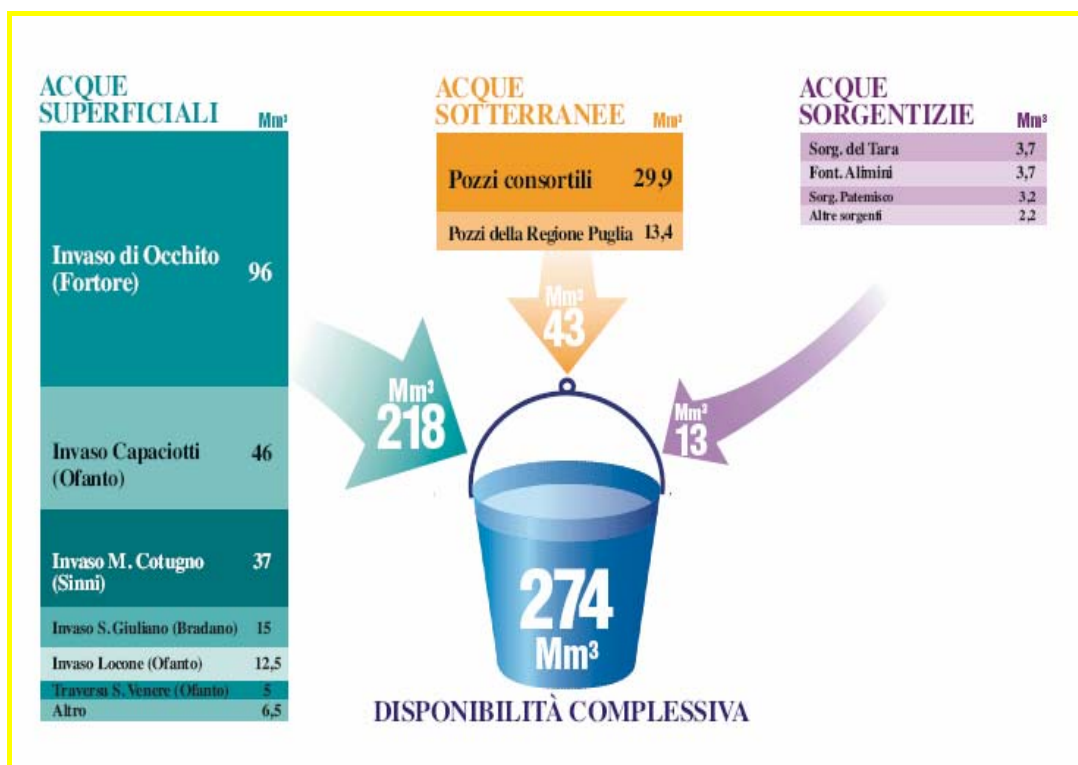
extraregionali ed in misura minore dalle sorgenti (5%) e dai pozzi (15%) situati nel territorio regionale. L'acqua pubblica per l'irrigazione, disponibile negli invasi ubicati nelle regioni limitrofe è pari a 55,5 Mm<sup>3</sup>, che rappresenta circa il 20% di tutta l'acqua disponibile per l'irrigazione pubblica;

► ***prevalenza della superficie irrigata con pozzi privati:*** la superficie irrigata, servita da rete irrigua pubblica pari a 75.517 ettari (Tabella 4), rappresenta solamente 1/5 della superficie (362.130 ettari) complessivamente irrigata nella regione. Di quest'ultima, una piccola parte sfugge al controllo diretto (evasione) da parte degli enti gestori, mentre la parte più rilevante è costituita dalle superficie irrigata con pozzi privati.

► ***sottoutilizzazione della superficie attrezzata con rete pubblica:*** con la rete irrigua gestita dai Consorzi di Bonifica è stata attrezzata complessivamente una superficie di 235.535 ettari (Tabella 4), di questa soltanto una parte (187.129 ettari) è in esercizio; di quest'ultima la superficie irrigata è pari a 75.517 ettari. La presenza di circa 50 mila ettari di superficie attrezzata con rete consortile, ma non in esercizio, è dovuta sia a carenze strutturali degli schemi idrici, sia a deficienze di interventi di manutenzione e di ammodernamento della rete irrigua. Ad esempio nel Consorzio di Bonifica dell'Arneo

è stata ultimata la rete di distribuzione, ma non sono mai state completate le opere di accumulo e/o adduzione della risorsa idrica. Un caso analogo si è verificato per il Consorzio di Bonifica di Ugento Li Foggi, in cui è stata costruita la rete di distribuzione in previsione dell'utilizzo di acque reflue, ma queste non sono mai state utilizzate a causa dell'elevato costo di gestione dell'impianto di depurazione. In altri casi la rete di distribuzione è inattiva perché degradata (Stornara) per mancanza di manutenzione e di interventi di ammodernamento.

**Figura 6 - La disponibilità di risorsa idrica per l'irrigazione: tipologia e fonti di approvvigionamento (2004)**



Fonte: Istituto Nazionale di Economia Agraria (INEA)

## 1.8 Considerazioni conclusive

In questo capitolo, sono state descritte le principali caratteristiche dell'agricoltura e dello stato dell'irrigazione in Puglia, allo scopo di comprendere l'ambito nel quale si effettua l'analisi dei dati. Dalla descrizione emerge che l'agricoltura, tale descrizione consente di dire che l'agricoltura svolge un ruolo ancora determinante nell'economia pugliese, essendo la superficie agricola attrezzata (SAU) della Puglia il 21% di quella dell'intero mezzogiorno e il 9,7% di quella nazionale (cfr 1, § 1.1). Le colture maggiormente praticate in Puglia sono quelle dei cereali, seguite da quelle dell'olivo e dalla vite, che occupano il 68% della SAU del territorio pugliese (*Figura 1*). È risultato, inoltre, evidente nella regione (cfr 1, § 1.1), un processo di senilizzazione dell'imprenditoria agricola (66% di conduttori con oltre 54 anni) ed una presenza ancora rilevante di occupati nel settore agricolo (11,42% di quelli totali). Occorre, inoltre, notare che il territorio rurale della Puglia presenta aspetti produttivi e paesaggistici molto diversificati. I territori comunali con più elevata produttività agricola sono quelli del Fortore, del Tavoliere, della Valle dell'Ofanto, del litorale barese e brindisino ed alcuni comuni del leccese dove è anche maggiormente diffusa l'irrigazione; mentre i territori comunali a bassa produttività agricola

sono quelli della Murgia, del Gargano, dell'Appennino Dauno e della Serre Salentine (cfr 1, § 1.2).

Il territorio pugliese, inoltre, è caratterizzato da un reticolo idrografico superficiale scarsamente sviluppato, ma in compenso presenta rilevanti risorse idriche sotterranee che hanno incentivato la vocazione all'agricoltura ed hanno determinato l'importanza economica dei sistemi colturali irrigui.

In questo capitolo è stata esaminata, anche la gestione delle risorse irrigue in Puglia, limitatamente ai Consorzi irrigui pubblici. Tale disamina consente di affermare che gli aspetti che caratterizzano l'irrigazione in Puglia sono:

- la penuria di risorsa idrica per l'irrigazione;
- la prevalenza della superficie irrigata con pozzi privati;
- la sottoutilizzazione della superficie attrezzata con rete pubblica.

Una volta esaminato il contesto nel quale realizzare l'analisi, si passa, nel capitolo successivo, alla descrizione del piano di indagine del presente lavoro.

# **CAPITOLO 2**

## **IL PIANO DELL'INDAGINE**

### **2.1 I dati**

Le serie storiche dei dati sono state fornite dall'Istituto Sperimentale Agronomico (C.R.A.-I.S.A.) di Bari. Esse si riferiscono ai consumi irrigui annui dell'Agro di Foggia dal 1951 al 2002 in relazione alle colture dell'olivo e del pomodoro. Tali valori annui sono stati ottenuti cumulando, rispettivamente, i consumi irrigui giornalieri dell'olivo e del pomodoro nei singoli anni considerati. Si noti che il calcolo dei consumi irrigui di una determinata coltura, si attua, generalmente, attraverso due fasi:

- a) calcolo dell'evapotraspirazione della coltura;
- b) calcolo del bilancio idrico.

## 2.2 Calcolo dell'evapotraspirazione della coltura

L'evapotraspirazione della coltura ( $ET_c$ ) è stata calcolata moltiplicando l'evapotraspirazione di riferimento ( $ET_0$ ) per il coefficiente colturale ( $Kc$ ); in simboli si ha:

$$ET_c = ET_0 * Kc \quad (2.1)$$

L'evapotraspirazione di riferimento, che permette di determinare la quantità di acqua dispersa nell'atmosfera da un ipotetico prato di ampia estensione i cui processi di crescita avvengono in condizioni ottimali, è stata stimata mediante il modello di *Penman-Monteith*.

L'impostazione del modello deriva dal lavoro teorico di *Penman* del 1948<sup>5</sup>, che ha fornito il fondamento per il calcolo dell'evaporazione da una superficie di acqua libera. Nella formula elaborata da *Penman* sono considerati i due termini che agiscono sull'evaporazione: radiativo e aerodinamico. Se il passaggio dell'acqua dallo stato liquido a quello di vapore fosse paragonato al movimento di un'automobile, l'energia disponibile (radiazione netta,  $Rn$ , al netto dell'energia che si accumula nel terreno sottoforma di calore,  $G$ ) sarebbe il "motore" ed il vento ( $u_2$ ),

---

<sup>5</sup> PENMAN H.L., *Natural evaporation from open water, bare soil and grass*, Proc. Royal Soc. A., 1948.



congiuntamente al deficit di saturazione del vapore nell'aria ( $e_s - e_a$ ), rappresenterebbe "l'acceleratore", dove  $e_s$  rappresenta la pressione di vapore saturo alla temperatura dell'aria e  $e_a$  indica la pressione di vapore saturo effettiva.

*Monteith* (1965)<sup>6</sup>, nell'applicare la formula di *Penman* alle colture, ha introdotto il termine di resistenza che funge da freno all'evaporazione<sup>7</sup>. La resistenza superficiale,  $r_s$  (resistenza del terreno, più quella cuticolare e stomatica), agendo in serie con la resistenza aerodinamica ( $r_a$ ), si oppone alla perdita di acqua della superficie verso l'atmosfera.

Con lo spirito di semplificare le procedure di calcolo, senza tradire l'impostazione teorica, la FAO<sup>8</sup> propone la seguente formulazione del modello di *Penman-Monteith* per la coltura di riferimento:

---

<sup>6</sup> MONTEITH J.L., *Evaporation and environment*, In: Fogg G.E. (Ed), *The State and Movement of Water in Living Organism*, Soc. Exp. Biol. Symp. 19, 1965.

<sup>7</sup> RANA G. - GIGOTTI R. - MASTRORILLI M. - JOVANOVIC N. - INCARNATO D., *Evapotranspiration measurement under drought condition*, , international conference on "Supplementary irrigation and drought water management", Valenzano (Bari), 27 settembre-2 ottobre 1992.

<sup>8</sup> ALLEN R.G – PEREIRA L:S. – RAES D., SMITH M., *Crop evapotranspiration: guidelines for computing crop water requirements*, Irrigation and Drainage Paper No.56, F.A.O., ROME, 1998.

$$ET_0 = \frac{0,408 \Delta (Rn - G) + \gamma \frac{900}{T + 273} u_2 (e_s - e_a)}{\Delta + \gamma (1 + 0,34 u_2)} \quad (2.2)$$

Con l'adozione degli opportuni fattori di conversione,  $ET_0$  è espresso in  $mm d^{-1}$ .

Come dati di ingresso al modello, rappresentati dall'energia disponibile ( $Rn - G$ ), dalla temperatura dell'aria ( $T$ ), dalla velocità del vento ( $u_2$ ) e dal deficit di saturazione del vapore acqueo ( $e_s - e_a$ ), sono stati utilizzati i rilievi agro-meteorologici eseguiti a Foggia, presso la stazione dell'Istituto Sperimentale Agronomico, realizzata secondo gli standard del WMO (1981)<sup>9</sup>.

Inoltre, nella formula (2.2),  $\Delta$ , che indica la pendenza della curva che correla pressione di vapore saturo e temperatura, può essere calcolata in funzione della temperatura media dell'aria ( $T$ , in  $^{\circ}C$ ), nel seguente modo:

$$\Delta = \frac{4098 \left[ 0,61082, 7183^{\frac{17,27 T_{med}}{T_{med} + 237,3}} \right]}{(T_{med} + 237,3)^2} = \frac{4098 \left[ 0,61082, 7183^{\frac{17,27 \cdot 26,24}{26,24 + 237,3}} \right]}{(26,24 + 237,3)^2} = 0,2007 \text{ KPa } ^{\circ}C^{-1}$$

---

<sup>9</sup>W.M.O., *Agricultural meteorological practices*, W.M.O.-No134, Geneva.

Mentre la costante  $\gamma$  si può calcolare come una semplice funzione della pressione atmosferica (P), ottenendo l'espressione:

$$\gamma = 0,665 \cdot 10^{-3} P = 0,655 \cdot 10^{-3} \cdot 100,24 = 0,0667 \text{ kPa}^\circ \text{C}^{-1}$$

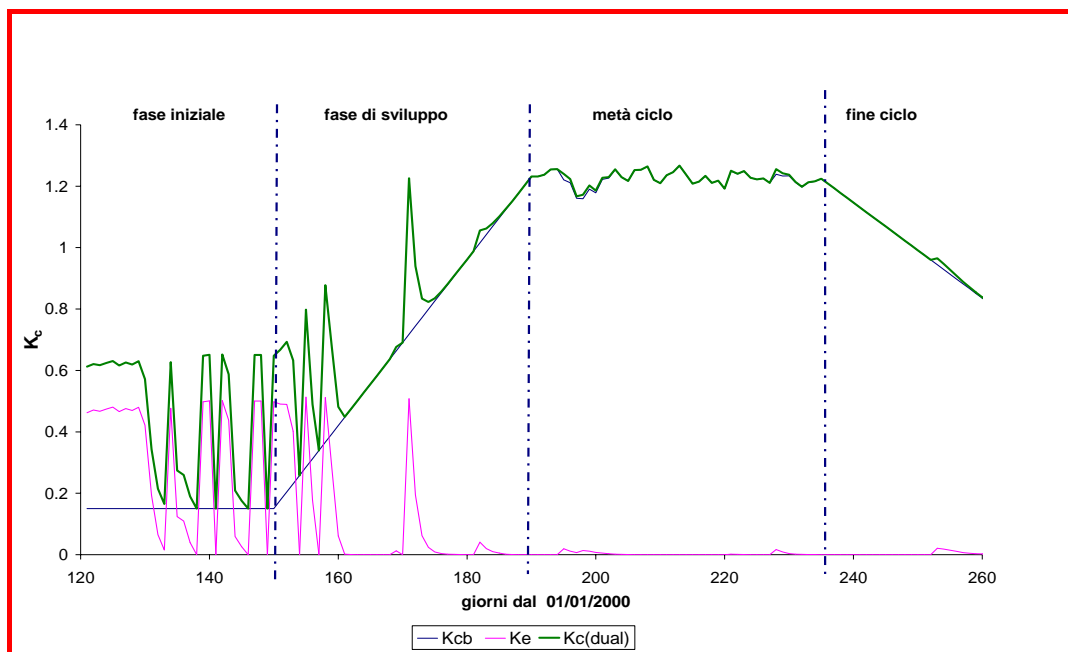
Il coefficiente colturale  $Kc$ , che compare nella formula (2.1), può essere rappresentato da un'unica funzione (*single Kc*) o diviso in due fattori (*dual Kc*) che descrivono separatamente l'evaporazione del terreno ( $Ke$ ) e la traspirazione della coltura ( $Kcb$ ). Il valore di  $Kc$  varia in funzione del grado di copertura del terreno da parte della coltura, della sua altezza e dell'area fogliare.

Dal punto di vista pratico è assai utile suddividere il ciclo colturale in quattro stadi di crescita: (1) iniziale ( $Kc_{ini}$ , dalla semina o trapianto, o dall'inizio dell'emissione di nuove foglie nel caso delle colture arboree, fino a quando la vegetazione copre circa il 10% del terreno), (2) di sviluppo ( $Kc_{svi}$ , fino alla completa copertura del terreno da parte delle foglie o, a seconda della tecnica di coltivazione, al pieno sviluppo della coltura), (3) metà ciclo ( $Kc_{med}$ , fino all'inizio della maturazione o ingiallimento o senescenza delle foglie) e (4) fine ciclo ( $Kc_{fin}$ , fino alla senescenza completa o raccolta).

La durata di ogni stadio dipende dalla specie, dalla varietà coltivata,

dal clima e dalle agro-tecniche impiegate. Per questo, l'intervallo tra una fenofase e quella successiva, non può essere definito a priori, ma occorre osservare periodicamente la coltura per rilevare la lunghezza temporale delle fenofasi principali.

**Figura 7-Evoluzione del *dual Kc* durante il ciclo colturale del pomodoro**



Fonte: Istituto Sperimentale Agronomico (C.R.A.)

Ad esempio, in Figura 7, dove viene riportata, l'evoluzione del *Kc dual* durante il ciclo colturale del pomodoro, la durata in giorni delle quattro fenofasi considerate deriva da osservazioni eseguite in campo direttamente sulla coltura in esame; inoltre, si osserva che il coefficiente *Kc dual* (suddiviso nelle sue due componenti *Ke* e *Kcb*) presenta un valore costante nella fenofase iniziale per poi salire nella fase di sviluppo

della coltura e raggiungere il massimo a metà ciclo. L'ultima fase, chiaramente, mostra un decrescita del coefficiente in esame.

## 2.3 Il bilancio idrico, ovvero quando e quanto irrigare

Si irriga se la pioggia è insufficiente a compensare le perdite per evapotraspirazione. Ma quando bisogna irrigare? E quanto? Queste due risposte vengono fornite dal bilancio irriguo:

$$U_{r,i} = U_{r,i-1} - (P - RO)_i - I_i - CR_i + ET_{c,i} + DP_i$$

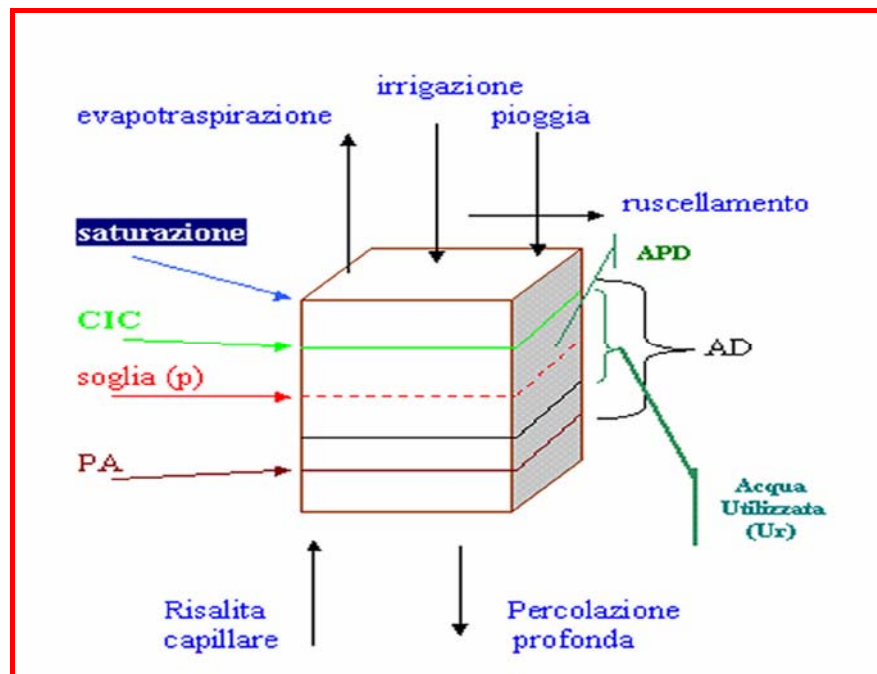
Con questa formula si stima la quantità di acqua utilizzata dalla coltura fino al giorno  $i$ -esimo ( $U_{r,i}$ , in  $mm$ ). Questo dato deriva dal volume di acqua utilizzato dalla coltura fino al giorno precedente ( $U_{r,i-1}$ ). Per il giorno  $i$ -esimo, si devono contabilizzare gli apporti di acqua dovuti a (vedi Figura 8):

- piogge utili  $P$  (espresse in  $mm$ ), ovvero le precipitazioni superiori a  $0,2 \cdot ET_0$ , alle quali vengono sottratti eventuali deflussi superficiali ( $RO$ );
- acqua irrigua ( $I$ );
- contributo di acqua dovuto alla risalita capillare ( $CR$ );

e le perdite per:

- evapotraspirazione della coltura ( $ET_c$ );
- percolazione ( $DP$ ), che si stima come la quantità di acqua che supera la capacità idrica di campo a seguito di un evento molto piovoso o di un'irrigazione abbondante.

**Figura 8 – Schematizzazione del bilancio idrico**



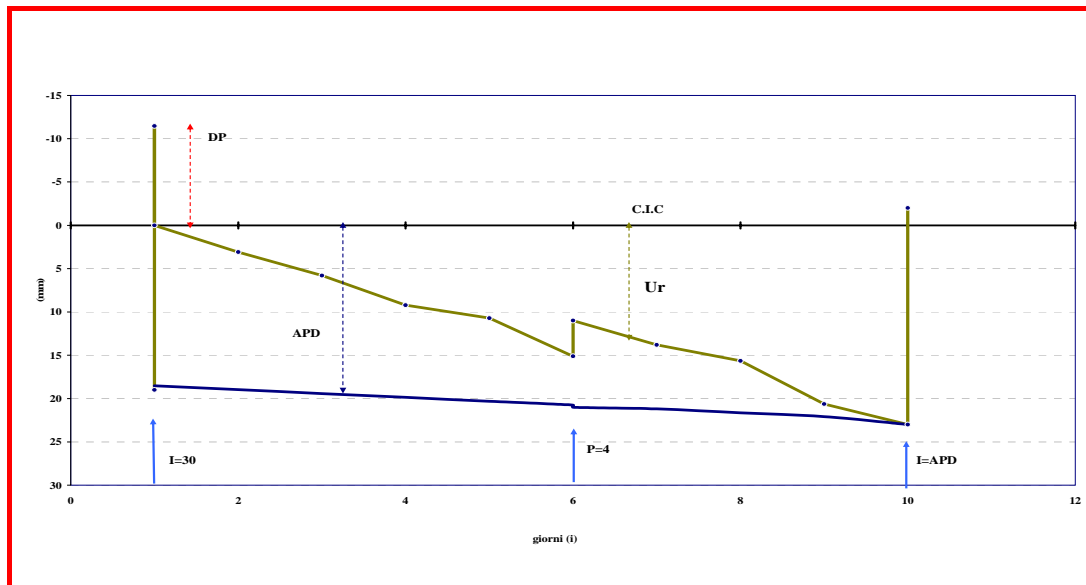
Fonte: Istituto Sperimentale Agronomico (CRA)

L'applicazione in scala giornaliera di tale modello permette di dimensionare accuratamente due variabili irrigue: momento e volume di adacquamento. Teoricamente le colture assorbono l'acqua dal suolo nell'intervallo compreso tra la capacità idrica di campo (CIC) e il punto

di appassimento (PA). In realtà, il lavoro che compie l'apparato radicale per assorbire l'acqua dal terreno aumenta progressivamente con la diminuzione dell'acqua disponibile; in altre parole, più essa si avvicina al PA e meno disponibile diventa l'acqua per le piante.

Per evitare conseguenze sulla produttività, l'irrigazione deve essere eseguita prima del completo esaurimento dell'acqua prontamente disponibile (APD) nello strato del terreno colonizzato dalle radici e cioè quando  $U_{r,i} \leq APD$  (vedi Figura 9).

**Figura 9 - Evoluzione di  $U_r$  e APD durante un periodo di 10 giorni**



Fonte: Istituto Sperimentale Agronomico (C.R.A.)

L'APD si ottiene dalla seguente formula:

$$APD = AD * p$$

dove:

- AD è l'acqua disponibile totale e corrisponde al contenuto idrico del terreno compreso tra PA e CIC.
- $p$  è la frazione di AD che, generalmente, varia tra 0,3 (nel caso delle colture a modesto sviluppo radicale) e 0,7 (nel caso di colture con radici profonde). I valori tabellati di  $p$  (indicati con  $p_{tab}$ ), comunque, dovrebbero essere corretti in base al pedo-clima. Per quanto riguarda la correzione relativa al clima una buona approssimazione potrebbe essere:

$$p = p_{tab} + 0,04(5 - ET_c).$$

Invece, per tener conto della pedologia,  $p_{tab}$  dovrebbe essere ridotto o aumentato del 5-10%, a seconda che la coltura si trovi in terreno tendenzialmente argilloso o tendenzialmente sabbioso.

## 2.4 Obiettivo dell'indagine

Come si è già accennato nell'introduzione, oggetto di studio del presente lavoro è l'analisi temporale dei dati in esame (cfr. capitolo 2, § 2.1), mediante l'approccio moderno o statistico all'analisi delle serie storiche. Tale studio, se da un lato permette di descrivere il fenomeno in



esame, dall'altro è in grado di identificare il processo stocastico sottostante alla serie storica considerata e quindi spiegare il meccanismo generatore della serie stessa. L'approccio moderno all'analisi delle serie storiche permette, inoltre, il conseguimento di un altro importante obiettivo: la previsione dei valori futuri della serie storica considerata. Occorre, tuttavia, tener presente che il fenomeno osservato (rappresentato dai consumi irrigui nell'Agro di Foggia) potrebbe rappresentare solo uno degli aspetti di un fenomeno più complesso, la cui descrizione richiederebbe l'analisi contemporanea di più serie storiche. Per cui se utilizzati a fini previsionali i modelli individuati dovrebbero essere integrati con altre variabili ambientali, poiché essi non sono in grado di fornire previsioni attendibili del fenomeno oggetto di studio. Nella seguente trattazione si procederà alla definizione di modelli che si ritiene generino le serie storiche considerate, riconoscendo, tuttavia, che tali modelli possano essere solo una parte o un'approssimazione di un sistema strutturale più ampio.

## **2.5 Software utilizzati per l'analisi**

Per l'analisi temporale dei dati è stato utilizzato il software "R" per Windows nella versione "2.1.0".

“R” può essere definito come un sistema di analisi statistica, che è contemporaneamente un linguaggio ed un software; esso è disponibile gratuitamente sotto i vincoli della GPL (*General Public Licence*) e scaricabile dal sito <http://www.r-project.org>.

Le sue principali caratteristiche possono essere così riassunte:

- semplicità nella manipolazione e gestione dei dati;
- disponibilità di una *suite* di strumenti per calcoli su vettori, matrici ed altre operazioni complesse;
- accesso ad un vasto insieme di strumenti integrati per l’analisi statistica;
- produzione di numerose potenzialità grafiche particolarmente flessibili;
- possibilità di adoperare un vero e proprio linguaggio di programmazione che consente l’uso di funzioni create dall’utente.

“R” è basato sul linguaggio “S” a cui è strettamente legato un altro ambiente commerciale probabilmente più conosciuto, S-Plus. Trattandosi di un linguaggio, è bene ribadire che, per un determinato problema possono esistere (ed in generale esistono) diverse soluzioni, tutte ugualmente valide<sup>10</sup>.

---

<sup>10</sup>IACUS S. – MASAROTTO G., *Laboratorio di statistica con R*, McGraw-Hill, Milano, 2003.

Il cuore di “R” è rappresentato dal modulo base (che offre gli strumenti fondamentali per l’analisi statistica) e attorno a questo “modulo” ruotano una serie di altre librerie (*packages*) addizionali, alcune delle quali sono già comprese nel programma “R” al momento in cui lo si installa, mentre altre librerie ancora, in relazione alle esigenze e necessità, possono essere aggiunte e installate dall’utente dal sito <http://cran.at.r-project.org/bin/windows/contrib>.

In particolare, si noti che per lo studio condotto sui dati a disposizione, è stato necessario scaricare il pacchetto “*ast*”, dedicato, appunto, all’analisi temporale moderna delle serie storiche.

Per la rappresentazione grafica di alcune serie di dati , ci si è avvalsi, inoltre, del programma *Excel 2003* appartenente al *package Office* disponibile per il sistema operativo Windows.

# **CAPITOLO 3**

## **PROCESSI STOCASTICI E**

### **MODELLI LINEARI**

#### **3.1 Introduzione**

Nel presente capitolo sono stati spiegati e definiti alcuni concetti teorici relativi allo studio moderno delle serie storiche, che sono fondamentali per comprendere l'analisi dei dati, condotta nel capitolo successivo mediante la procedura di Box-Jenkins. Esso è stato sviluppato nella seguente maniera: dopo aver chiarito il concetto di serie storica, si procede alla distinzione fra approccio classico e moderno all'analisi delle stessa e si spiega il motivo per cui si è preferito scegliere l'approccio moderno; tale analisi si basa sul concetto di processo stocastico, di cui

viene data una definizione formale. Infine, dopo aver spiegato cosa si intende per processo stazionario e non stazionario, vengono descritti quelli più comunemente impiegati nell'analisi moderna delle serie storiche, i quali rientrano nella cosiddetta classe dei modelli ARIMA.

### **3.2 Lo studio dei fenomeni che variano nel tempo**

Per serie storica si intende un insieme di dati ordinati secondo un criterio qualitativo. Quando il criterio ordinatore dei dati è il tempo, inteso come progressione cronologica, si ha una serie storica o temporale. Possiamo pertanto definire una serie storica come “una successione di dati numerici, in cui ogni dato è associato ad un particolare istante o intervallo di tempo”<sup>11</sup>. Se in ciascun punto o intervallo di tempo viene osservato un solo fenomeno, la serie che ne deriva viene detta univariata. Se invece le variabili sono più di una, si ha una serie storica di più variabili (o multivariata), detta anche serie storica multipla. Tuttavia, ai fini dello studio che si andrà a condurre nel presente lavoro, si considereranno solo serie storiche univariate.

---

<sup>11</sup> VIANELLI S., *L'analisi delle serie temporali nello sviluppo storico e metodologico della statistica*, Atto del Convegno nazionale di studio sull'“Analisi moderna delle serie storiche”, Napoli, 19-22 maggio 1981.

Il parametro temporale  $t$ , che definisce l'ordinamento dei dati, appartiene ad un insieme parametrico  $T$ , che può essere continuo e discreto. Nel primo caso la serie storica viene detta a parametro continuo (o, più semplicemente, continuo) e nel secondo caso a parametro discreto (o, più semplicemente, discreto). Le serie storiche che si andranno ad analizzare, riguardano sequenze di osservazioni rilevate ad intervalli di tempo equidistanti (anni), con variabili nell'insieme dei numeri interi relativi. Si andranno a considerare, pertanto, serie storiche equispaziate a parametro discreto.

Assumendo l'intervallo temporale come unità di riferimento, i valori di una serie storica ottenuta osservando il fenomeno  $Y$  ai tempi  $i_1, \dots, i_t, \dots, i_n$ , dati da:

$$y(i_1), \dots, y(i_t), \dots, y(i_n)$$

possono essere più semplicemente scritti come:

$$y_1, \dots, y_t, \dots, y_n.$$

Un'altra importante distinzione è tra serie deterministiche e serie storiche stocastiche. Una serie storica  $\{y_t\}_{t=1}^n$  è detta deterministica se esiste una funzione:

$$\Psi_t = \Psi(t, y_{t-1}, y_{t-2}, \dots)$$

tale per cui:

$$E(y_t - \Psi_t)^2 = 0 \quad t \in T.$$

Una serie deterministica può dunque essere prevista esattamente sulla base della propria storia passata.

Una serie storica stocastica, invece, rappresenta un fenomeno che evolve nel tempo seguendo determinate leggi probabilistiche<sup>12</sup>.

La maggioranza delle serie storiche, comprese quelle che si andranno ad analizzare, è tuttavia di tipo stocastico, nel senso che il futuro viene determinato dal passato solo in misura parziale, rendendo quindi impossibile l'elaborazione di previsioni del tutto prive di errori.

### **3.3 Diversi approcci allo studio delle serie storiche**

Le serie storiche possono essere analizzate seguendo due approcci: il primo, cosiddetto classico, consiste nella descrizione delle caratteristiche della serie storica con strumenti propri della statistica, quali medie mobili, regressioni ed interpolazioni.

---

<sup>12</sup> DI FONZO T. - LISI F., *Complementi di statistica economica*, Cleup, Padova, 2001.

Il secondo approccio, più moderno, detto anche stocastico, consiste nella determinazione di un modello teorico di riferimento in grado di generare una serie “simile” a quella oggetto di analisi e nel fare inferenza su di essa. Di seguito si analizzeranno i fondamenti teorici di entrambi i metodi focalizzando l’attenzione sull’approccio statistico, oggetto di questa tesi.

### **3.3.1 Approccio classico**

Nell’approccio classico l’analisi della serie storica si realizza attraverso la scomposizione della serie nelle componenti di tendenza, ciclo e stagionalità e mediante la successiva determinazione delle componenti della serie. Perciò questa metodologia viene spesso denominata anche analisi delle componenti.

Nell’approccio classico la previsione si realizza attraverso la determinazione di una funzione matematica (lineare, parabolica, iperbolica, etc.), spesso stimata con il metodo dei minimi quadrati. I valori futuri della serie vengono poi estrapolati dalla funzione matematica ed aggiustati dalle altre componenti di ciclo e stagionalità.

Il modello può essere sintetizzato nel seguente modo:

$$x_t = \varphi (T_t, C_t, S_t, A_t)$$



dove  $x_t$  è la variabile osservata al tempo  $t$ ,  $T_t$ ,  $C_t$ ,  $S_t$  e  $A_t$ , sono le componenti sistematiche osservate nello stesso istante di tempo. Più in particolare:

- $T_t$  rappresenta la componente di trend, cioè la tendenza generale della serie storica e l'andamento di lungo periodo che la variabile oggetto di studio ha avuto nell'arco di tempo considerato;
- $C_t$  è la componente ciclica, cioè l'alternanza registrata dalla serie nel corso del periodo in esame;
- $S_t$  è la componente stagionale legata, cioè al diverso andamento che la serie registra nel corso di un anno solare, dovuta quindi a cause direttamente connesse al succedersi delle stagioni naturali;
- $A_t$  rappresenta la componente accidentale o di errore.

Mentre le cause della variabilità nel tempo delle componenti di tendenza, di ciclo e di stagionalità sono direttamente connesse al fenomeno oggetto di osservazione, la componente accidentale è spesso dovuta a *shock* esterni al fenomeno, o ad eventi di natura puramente aleatoria, difficili da prevedere in quanto privi di una certa regolarità nel tempo.

Del modello classico sono state proposte varie specificazioni a seconda di come interferiscono le diverse componenti sulla variabile

osservata. In particolare, dalla teoria sono stati proposti i seguenti modelli funzionali:

1) Modello additivo

$$x_t = T_t + C_t + S_t + A_t$$

2) Modello moltiplicativo:

$$x_t = T_t \cdot C_t \cdot S_t \cdot A_t$$

che può essere ricondotto attraverso una trasformazione logaritmica al seguente modello:

$$\text{Log } x_t = \text{Log } T_t + \text{Log } C_t + \text{Log } S_t + \text{Log } A_t$$

3) Modello misto

$$x_t = T_t \cdot C_t \cdot S_t + A_t \quad \text{oppure} \quad x_t = T_t \cdot C_t + S_t + A_t$$

quando il trend e il ciclo vengono attribuiti ad un'unica componente.

Da notare che nel modello additivo le componenti sono espresse nella stessa unità di misura della variabile osservata e ciascuna componente è stimata in maniera autonoma e indipendente dalle altre. Il modello moltiplicativo, di solito più usato per la rappresentazione dei fenomeni economici, ha solo la componente di tendenza espressa nell'unità di misura della variabile osservata. Le altre componenti sono rappresentate da coefficienti che amplificano o riducono l'effetto della

componente di tendenza a seconda di come vanno ad influire in quel dato istante le componenti cicliche, stagionali ed accidentali<sup>13</sup>.

### **3.3.2 Approccio moderno o statistico**

Come detto in precedenza nell'approccio moderno, l'analisi delle serie temporali avviene attraverso la specificazione di un modello teorico di riferimento che è in grado di generare, in base a dei coefficienti stimati, una serie temporale quanto più possibile vicina a quella oggetto di analisi.

Per tale ragione, l'analisi moderna delle serie storiche è in grado di giungere a previsioni più precise, quindi con un minore margine di errore, rispetto a quelle ottenute mediante la metodologia di scomposizione della serie nelle varie componenti di trend, ciclo e stagionalità<sup>14</sup>. Poiché uno degli obiettivi del presente lavoro è quello di poter utilizzare il modello identificato per effettuare previsioni sul fenomeno in esame, lo studio temporale delle serie storiche considerate sarà svolto mediante l'approccio moderno. Occorre, tuttavia, notare che

---

<sup>13</sup> ALVARO G., *Contabilità nazionale e statistica economica*, Cacucci, Bari, 1999.

<sup>14</sup> DI FONZO T. - LISI F., *Complementi di statistica economica...ecc., op. cit.*

il modello individuato mediante tale approccio rappresenta solo la base di un modello strutturale più complesso, perciò per poter effettuare previsioni attendibili, è necessario integrare il modello con altre variabili collegate al fenomeno oggetto di studio.

L'analisi moderna delle serie storiche si basa sul concetto di processo stocastico, che verrà definito nel paragrafo successivo.

### **3.4 I processi stocastici**

Alla base dell'approccio moderno all'analisi delle serie storiche c'è la teoria dei processi stocastici, già definiti (cfr. capitolo 3, § 3.2) come un "fenomeno che evolve seguendo delle leggi probabilistiche".

Per una definizione più formale bisogna invece considerare uno spazio di probabilità  $(\Omega, F, P)$ , dove  $\Omega$  è lo spazio degli eventi elementari,  $F$  è la classe degli eventi che si richiede sia chiusa rispetto alle operazioni logiche di unione e negazione su  $\Omega$  (sia, cioè, una  $\sigma$ -algebra su  $\Omega$ ) e  $P$  una misura di probabilità.

Dato uno spazio parametrico  $T$ , si definisce processo stocastico una funzione finita e a valori reali di  $\omega \in \Omega$  e  $t \in T$ , tale che per ogni  $t$ ,  $Y_t(\omega)$  è una funzione misurabile di  $\omega$ . Nel seguito il parametro  $t$

indicherà sempre il tempo, inoltre supporremo che l'insieme  $T$  sia discreto.

Un processo stocastico è dunque una funzione di due variabili di natura diversa:

- $t$ , di natura matematica, che ha lo scopo di creare un ordine nella famiglia di variabili casuali;
- $\omega$ , evento che specifica su  $\Omega$  quale risultato si verifica per  $t$  fissato.

Si ha inoltre:

- per ogni  $\omega = \omega_0$  fissato,  $Y_t(\omega_0)$  è una funzione di  $t$  detta realizzazione o traiettoria del processo stocastico;
- per ogni  $t = t_0$  fissato,  $Y_{t_0}(\omega)$  è una funzione misurabile di  $\omega \in \Omega$ , e quindi una variabile casuale;
- per  $t = t_0$  e  $\omega = \omega_0$ ,  $Y_{t_0}(\omega_0)$  è un numero reale.

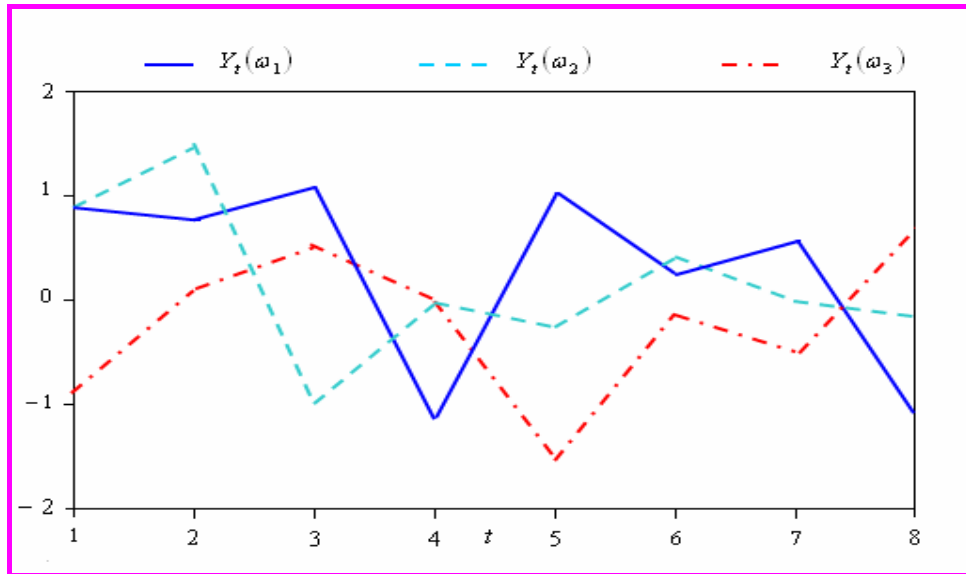
A titolo di esempio, nella Figura 10 vengono rappresentate tre realizzazioni finite di uno stesso processo stocastico<sup>15</sup>.

Le tre serie, pur generate dallo stesso processo, differiscono per l'evento elementare  $\omega$  su cui opera il meccanismo casuale di generazione dei valori osservati.

---

<sup>15</sup> DI FONZO T. - LISI F., Complementi di statistica economica...ecc., op. cit..

**Figura 10 - Realizzazioni di uno stesso processo stocastico**



Semplificando, possiamo dunque dire che un processo stocastico è una collezione di variabili casuali indicizzate nel tempo<sup>16</sup>:

$$\{Y_t, t = t_1, t_2, \dots\}. \quad (3.1)$$

Per la descrizione completa di un processo stocastico occorrerebbe specificare la distribuzione di probabilità congiunta di  $(Y_{t_1}, \dots, Y_{t_n})$  per ogni insieme di istanti  $t_1, \dots, t_n$  ed ogni valore di  $n$ . Ciò è piuttosto complicato ed in genere si preferisce descrivere il processo stocastico sulla base dei momenti, in particolare dei primi due, delle variabili casuali  $Y_t$ . Al variare di  $t$ , media, varianza e covarianza definiscono

---

<sup>16</sup> Per comodità di notazione, nel seguito si assume  $Y_t \equiv Y_t(\omega)$ .

funzioni e valori reali e precisamente:

- Funzione Media:

$$\mu_t = E[Y_t].$$

- Funzione Varianza:

$$\sigma_t^2 = \text{Var}[Y_t] = E[Y_t - \mu_t]^2.$$

- Funzione Autocovarianza:

$$\gamma_{t_1, t_2} = E\{[Y_{t_1} - \mu_{t_1}][Y_{t_2} - \mu_{t_2}]\}.$$

Si noti che, per  $t_1 = t_2 = t$ ,  $\gamma_{t_1, t_2} = \sigma_t^2$ . Inoltre, varianza e autocovarianza sono entrambe espresse nel quadrato dell'unità di misura di  $Y_t$ .

Il significato della media e della varianza di una variabile casuale sono ben noti, per cui risulta facile l'estensione delle funzioni “media” e “varianza” di un processo stocastico.

Ci si sofferma invece sul senso della funzione “autocovarianza”: essa non è altro che la covarianza tra variabili casuali del medesimo

processo stocastico spaziate fra loro di uno sfasamento temporale<sup>17</sup>, pari a  $k = |t_2 - t_1|$ .

La funzione varianza si può dunque interpretare come l'autocovarianza in corrispondenza di  $k = 0$ .

Per facilitare l'interpretazione della autocovarianza, è opportuno effettuare la normalizzazione. Definiamo, pertanto, la funzione di autocorrelazione (generalmente indicata con *ACF*, da *AutoCorrelation Function*):

$$\rho_{t_1, t_2} = \frac{\gamma_{t_1, t_2}}{\sigma_{t_1, t_2}}. \quad (3.2)$$

La (3.2) è una misura della correlazione tra  $Y_{t_1}$  e  $Y_{t_2}$ . Essa ha lo stesso contenuto informativo della funzione di autocovarianza, ma ha il vantaggio di esser un numero puro, indipendente dall'unità di misura di  $Y_t$ .

Una proprietà di cui godono i processi stocastici, è l'ergodicità, che assicura che le medie nel tempo forniscano stime consistenti dei

---

<sup>17</sup> Si noti che nel secondo paragrafo di questo capitolo, si è detto che ci si riferirà a serie storiche osservate ad istanti di tempo equispaziate.



momenti (media, varianza ed autocovarianza) del processo stocastico. In sostanza l'ergodicit  richiede che la memoria del processo sia limitata, cos  che osservazioni molto distanti nel tempo siano fra loro non correlate.

Tale condizione viene sintetizzata nella seguente propriet  dei processi stocastici che deriva dal teorema ergodico di *Slutsky*:

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n \text{Cov}(x_t, x_{t-k}) = 0.$$

Di conseguenza, se un processo   ergodico,   possibile usare le informazioni contenute nella sua evoluzione nel tempo e fare inferenza sulle caratteristiche.

Il "teorema ergodico" dice appunto che l'osservazione di una serie abbastanza lunga   equivalente, ai fini inferenziali, all'osservazione di un gran numero di realizzazioni.

Ad esempio, se un processo ergodico ha valore atteso  $\mu$ , allora la sua media aritmetica nel tempo   uno stimatore consistente di  $\mu$ , quindi  $\mu$  pu  essere stimato in modo consistente come se si disponesse di molte realizzazioni del processo anzich  di una sola<sup>18</sup>.

---

<sup>18</sup> JOHONSTON J., *Econometrica*, Franco angeli, Milano, 1993.

### 3.5 Processi stazionari

Da un punto di vista intuitivo un processo stocastico è stazionario se la sua media e la sua variabilità non presentano cambiamenti di natura sistematica e se la sua dinamica non presenta variazioni strettamente periodiche.

Più formalmente, un processo stocastico si dice stazionario in senso stretto se le distribuzioni congiunte, rispettivamente di  $(Y_{t_1}, \dots, Y_{t_n})$  e di  $(Y_{t_1+\tau}, \dots, Y_{t_n+\tau})$  per ogni insieme di istanti  $t_1, \dots, t_n$  e per ogni  $\tau$  sono uguali. Per  $n=1$ , la stazionarietà in senso stretto implica che la distribuzione di  $Y_t$  sia la stessa per ogni  $t$ ; dunque, purché i primi due momenti siano finiti, ne consegue che la media e la varianza sono entrambe costanti e non dipendono dall'istante  $t$ , ovvero:

$$\mu_t = \mu, \quad \sigma_t^2 = \sigma^2.$$

Per quanto riguarda invece la funzione di autocovarianza si ha:

$$\gamma_{t_1, t_2} = E \{ [Y_{t_1} - \mu][Y_{t_2} - \mu] \} = \gamma_{|t_2 - t_1|} \quad (3.3)$$

L'autocovarianza di un processo stazionario in senso stretto dipende, cioè solo dallo sfasamento temporale tra le due variabili e non

dagli istanti  $t_1$  e  $t_2$ . Ponendo, pertanto,  $t_1 = t$  e  $t_2 = t + k$ , la (3.3) può essere scritta nel modo seguente:

$$\gamma_{t,t+k} = E \{ [Y_t - \mu][Y_{t+k} - \mu] \} \quad k = 0, \pm 1, \pm 2, \dots \quad (3.4)$$

Nei processi stocastici stazionari l'autocovarianza assume un ruolo fondamentale. Essa è un indice delle relazioni lineari esistenti tra coppie di variabili casuali, componenti il processo stocastico, prese con uno sfasamento pari a  $k$ . Inoltre, l'autocovarianza di un processo stocastico stazionario è una funzione pari di  $k$ , nel senso che  $\gamma_k = \gamma_{-k}$ ; ciò consente di limitare il calcolo della funzione di autocovarianza ai soli valori positivi di  $k$ .

Una proprietà analoga alla (3.4) vale per la funzione di autocorrelazione (globale), che è pari a:

$$\rho_{t,t+k} = \rho_k = \frac{\gamma_k}{\gamma_0}, \quad k = 0, \pm 1, \pm 2, \dots \quad (3.5)$$

La funzione di autocorrelazione (3.5) soddisfa le seguenti proprietà<sup>19</sup>:

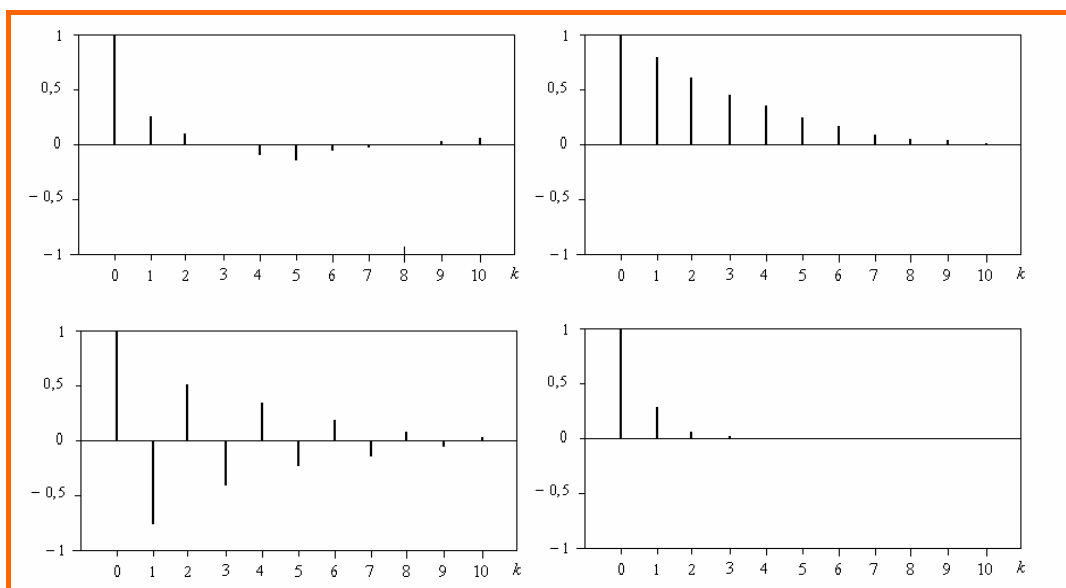
---

<sup>19</sup> WEI W.W.S., *Time series analysis. Univariate and multivariate methods*, Addison-Wesley, Redwood City, 1990.

- 1)  $\rho_0 = 1$
- 2)  $\rho_k = \rho_{-k}$
- 3)  $|\rho_k| \leq 1$

Il grafico dei valori  $\rho_k$ , per  $k = 0, 1, 2, \dots$  viene chiamato correlogramma. Alcuni esempi di correlogramma sono riportati nella seguente figura (*Figura 11*).

**Figura 11 - Esempi di correlogrammi**



In generale, la correlazione tra due variabili può essere dovuta al fatto che esiste effettivamente un legame lineare diretto tra le variabili o al fatto che queste ultime sono correlate con una terza variabile. Nel contesto delle serie storiche, buona parte della correlazione tra  $Y_t$  e  $Y_{t+k}$

può essere dovuta alla correlazione che tali variabili hanno con  $Y_{t+1}, Y_{t+2}, \dots, Y_{t+k-1}$ . Un possibile modo per tenere conto di ciò è considerare la funzione di autocorrelazione parziale (*PACF*, da *Partial AutoCorrelation Function*) che misura l'autocorrelazione tra  $Y_t$  e  $Y_{t+k}$  al netto delle variabili intermedie.

La funzione di autocorrelazione parziale è dunque pari alla correlazione condizionata:

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

che può anche essere espressa come<sup>20</sup>:

$$P_k = \frac{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_1 \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_2 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & \rho_k \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \cdots & \rho_{k-2} & \rho_{k-1} \\ \rho_1 & 1 & \cdots & \rho_{k-3} & \rho_{k-2} \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \cdots & \rho_1 & 1 \end{vmatrix}}, \quad (3.6)$$

dove  $|\cdot|$  rappresenta il determinante di una matrice.

---

<sup>20</sup> WEI W.W.S., *Time series analysis...ecc., op. cit.*

La condizione di stazionarietà in senso stretto è piuttosto forte, perché impone dei “vincoli” sull’intera distribuzione del processo e sui momenti di qualsiasi ordine.

Una forma di stazionarietà meno restrittiva è quella cosiddetta debole o del secondo ordine, così chiamata perché impone condizioni solamente sui primi due momenti.

Un processo è definito stazionario del secondo ordine, o debolmente stazionario, se la sua media è costante e la sua funzione di autocovarianza dipende solo da  $k$ , cosicché:

$$\begin{aligned} E[Y_t] &= \mu & \forall t, \\ \text{Cov}[Y_t, Y_{t+k}] &= \gamma_k & \forall t, \forall k \end{aligned} \tag{3.7}$$

Dalla (3.7), per  $k = 0$ , consegue che anche la varianza  $\gamma_0$  di un processo stocastico debolmente stazionario è costante. Si noti, inoltre, che non viene fatta nessuna assunzione sui momenti di ordine superiore al secondo.

Un importante risultato, che rende particolarmente interessanti i processi debolmente stazionari, è il teorema di decomposizione di *Wold*. Esso afferma che per ogni processo stocastico debolmente stazionario  $Y_t$ , si ha:

$$Y_t = f(t) + \sum_{i=1}^{\infty} \psi_i \varepsilon_{t-i}$$

in cui  $f(t)$  è una funzione deterministica,  $\varepsilon_t$  è un processo *white-noise*,

$$\psi_0 = 1 \text{ e } \sum_{i=1}^{\infty} \psi_i^2 < +\infty.$$

Pertanto, in base al teorema di *Wold*, ogni processo stocastico debolmente stazionario può essere espresso come la somma di una componente deterministica, corrispondente alla media o ad eventuali trend e/o stagionalità della serie, più una componente stocastica, data dalla somma ponderata di un processo *white-noise*.

La nozione di stazionarietà debole è quella più usata in pratica, in parte per comodità, in parte perché, per alcuni processi, la stazionarietà del secondo ordine è già soddisfacente.

Ciò è vero, in particolare, per i processi gaussiani, cioè quelli per cui  $(Y_{t_1}, Y_{t_2}, \dots, Y_{t_k})$  si distribuisce come una variabile casuale normale  $k$ -variata per ogni  $k$ -upla  $t_1, t_2, \dots, t_k$  e per ogni  $k \geq 1$ . Un processo gaussiano, infatti, è completamente caratterizzato dai suoi primi momenti, dunque, in questo caso, stazionarietà forte e debole coincidono.

Poiché nell'analisi delle serie storiche, che si effettuerà nel capitolo

successivo, si considererà la stazionarietà in senso debole di un processo stocastico, nel seguito si useranno i termini “stazionario” e “stazionarietà” come sinonimi di “debolmente stazionario” e di “stazionarietà in senso debole”.

Dato un processo stazionario  $\{Y_t\}$ , è dunque possibile calcolare in maniera univoca la sua funzione di autocovarianza  $\gamma_k$ , quindi  $\rho_k$ . Ci si può chiedere se sia vero il contrario, cioè se data una funzione di autocovarianza è unico il processo stazionario  $\{Y_t\}$  che possiede quella funzione di autocovarianza.

La risposta, in generale, è negativa, perché la funzione di autocovarianza non caratterizza pienamente un processo e si può dimostrare che esistono più processi con la stessa funzione di autocovarianza. La risposta è invece affermativa se, oltre alla condizione di stazionarietà, si introduce anche la condizione di invertibilità. Infatti, tra i vari processi con la stessa funzione di autocovarianza, uno solo è anche invertibile.

La condizione di invertibilità riguarda la possibilità di esprimere un processo  $\{Y_t\}$  come una funzione delle variabili casuali del “passato”<sup>21</sup>.

---

<sup>21</sup> DI FONZO T. - LISI F., Complementi di statistica economica...ecc., op. cit..



Un processo  $\{Y_t\}$  è detto invertibile se esiste una funzione lineare  $h(\cdot)$  e un processo white noise  $\{\varepsilon_t\}$  tale che, per ogni  $t$ , si ha:

$$Y_t = h(Y_{t-1}, Y_{t-2}, \dots) + \varepsilon_t$$

### **3.6 Stima dei momenti di un processo stocastico stazionario**

Quanto detto fin qui, riguarda la definizione dei momenti “teorici” di un processo stocastico. Quando però si analizza la serie storica di un fenomeno reale, il processo generatore è ignoto e si pone il problema di ricercare, con metodi inferenziali, le caratteristiche del processo stocastico, partendo dalla serie storica dei dati osservati  $\{y_t\}_{t=1}^n$  che costituisce l’unica realizzazione conosciuta del processo generatore dei dati.

Se il processo stocastico è stazionario e sotto la condizione di ergodicità, i momenti possono essere stimati consistentemente in base ai dati osservati della serie storica.

L’espressione:

$$\hat{\mu} = \frac{1}{n} \sum_{t=1}^n y_t \quad (3.8)$$

fornisce uno stimatore corretto e consistente della media  $\mu$  del processo stocastico.

Si può infatti mostrare che:

$$E[\hat{\mu}] = \mu,$$

$$\text{Var}[\hat{\mu}] = \frac{\sigma^2}{n} \left[ 1 + 2 \sum_{r=1}^{n-1} \left( 1 - \frac{r}{n} \right) \rho_r \right].$$

Poiché:

$$\lim_{n \rightarrow \infty} \left( 1 + 2 \sum_{r=1}^{n-1} \left( 1 - \frac{r}{n} \right) \rho_r \right) = 1 + 2 \sum_{r=1}^{\infty} \rho_r,$$

affinché sia  $\lim_{n \rightarrow \infty} \text{Var}[\hat{\mu}] = 0$  è sufficiente che  $\sum_{r=1}^{\infty} \rho_r < \infty$ . Pertanto, se:

$$\lim_{r \rightarrow \infty} \rho_r = 0 \quad (3.9)$$

lo stimatore (3.8) è consistente. La condizione (3.9) è detta di ergodicità rispetto alla media.

Per l'autocovarianza sono stati proposti vari stimatori. Quello solitamente preferito è:

$$\hat{\gamma}_k = \frac{1}{n} \sum_{t=1}^{n-k} (y_t - \hat{\mu})(y_{t+k} - \hat{\mu}). \quad (3.10)$$

Sotto condizioni di ergodicità, analoghe ma lievemente più stringenti della (3.9),  $\hat{\gamma}_k$  è uno stimatore consistente. In particolare si dimostra che<sup>22</sup>:

$$E[\hat{\gamma}_k] = \gamma_k - \frac{k}{n}\gamma_k - \frac{n-k}{n}\text{Var}[\hat{\mu}].$$

Lo stimatore  $\hat{\gamma}_k$  è quindi distorto, ma la sua distorsione tende a zero al crescere di  $n$ . Affinché le stime di  $\hat{\gamma}_k$  possano ritenersi soddisfacenti, è necessario che  $n$  sia sufficientemente elevato (solitamente si richiede  $n > 50$ ) e che  $k$  sia molto minore di  $n$  (in genere è sufficiente  $k \leq n/4$ ).

Lo stimatore della varianza si ottiene ponendo  $k = 0$  nella (3.10):

$$\hat{\sigma}^2 = \hat{\gamma}_0 = \frac{1}{n} \sum_{t=1}^n (y_t - \hat{\mu})^2.$$

Uno stimatore ragionevole dell'autocorrelazione è dunque:

$$\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}. \quad (3.11)$$

Lo stimatore (3.11) è distorto, ma si può dimostrare che è

---

<sup>22</sup> WEI W.W.S., *Time series analysis...ecc.*, op. cit..

consistente. In particolare sotto opportune ipotesi,  $Var[\hat{\rho}_k]$  può essere approssimata da:

$$Var[\hat{\rho}_k] \cong \frac{1}{n} \sum_{i=-\infty}^{+\infty} (\rho_i^2 + \rho_{i-k} \rho_{i+k} + 2\rho_i^2 \rho_k^2 - 4\rho_i \rho_k \rho_{i-k}). \quad (3.12)$$

Se  $\rho_k \cong 0$  per  $k > q$  la (3.12), nota come approssimazione di *Bartlett*<sup>23</sup>, può essere ulteriormente approssimata da:

$$Var[\hat{\rho}_k] \cong \frac{1}{n} \sum_{i=-q}^q \rho_i^2.$$

Uno stimatore  $\hat{P}_k$  di  $P_k$  può essere ottenuto sostituendo  $\hat{\rho}_k$  a  $\rho_k$  nella (3.6). Per  $Var[\hat{P}_k]$  vale inoltre l'approssimazione di *Quenouille*<sup>24</sup>:

$$Var[\hat{P}_k] \cong \frac{1}{n}.$$

Mediante la stima dei momenti sino al secondo ordine è dunque possibile descrivere le principali caratteristiche di un processo

<sup>23</sup> BARTLETT M.S., "On the theoretical specification of sampling properties of autocorrelated time series", *Journal of the Royal Statistical Society*, B, 8, 1946.

<sup>24</sup> BOX G.E.P. - JENKINS G.M., *Time Series Analysis, Forecasting and Control*, Holden - Day, San Francisco, 1976.

stazionario, usando solo la serie dei valori osservati.

### 3.7 Alcuni processi stocastici stazionari

In questo paragrafo vengono descritti alcuni dei processi stocastici più comunemente utilizzati nell'analisi delle serie storiche.

#### 3.7.1 Processo puramente casuale (*white-noise*)

Il processo *white noise* è uno dei più importanti processi stocastici. Si tratta di un processo puramente casuale,  $\{\varepsilon_t\}$ , che consiste in una sequenza di variabili casuali indipendenti ed identicamente distribuite, di media nulla e varianza costante.

Per tale processo, sinteticamente indicato con la notazione  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ , si ha:

- $E[\varepsilon_t] = 0 \quad \forall t,$
  - $Var[\varepsilon_t] = \sigma_\varepsilon^2 \quad \forall t,$
  - $E[\varepsilon_r, \varepsilon_s] = 0 \quad \forall r, s, \quad r \neq s$
- (3.13)

Dalla (3.13) è immediato dedurre che tale processo è debolmente stazionario; in questo caso, poi, il processo è anche stazionario in senso stretto.

Inoltre, se le variabili casuali  $\varepsilon_t$  hanno distribuzione normale si

parla di *white noise gaussiano*.

La funzione di autocorrelazione di un processo *white-noise* è:

$$\rho_k = \begin{cases} 1 & k = 0 \\ 0 & k = \pm 1, \pm 2, \dots \end{cases}$$

La funzione di autocorrelazione parziale  $P_k$  è identicamente uguale a  $\rho_k$ , dal momento che le componenti sono serialmente incorrelate.

Si noti che, utilizzando l'approssimazione di *Bartlett* (cfr. 3, § 3.6), si ha:

$$\text{Var}[\hat{\rho}_k] \cong \frac{1}{n} \quad \forall k \neq 0 \quad (3.14)$$

La (3.14) implica che, per l'autocorrelazione empirica di un processo *white noise*, asintoticamente vale il risultato  $\hat{\rho}_k \sim N(0, 1/n)$ . Questo risultato può essere utilizzato per verificare se i residui di un modello stimato sono incorrelati. Il rifiuto di tale ipotesi, ad un opportuno livello di significatività, comporta solitamente l'abbandono dell'intero modello, in quanto i residui non possono ritenersi imputabili solamente al caso<sup>25</sup>.

---

<sup>25</sup> DI FONZO T. - LISI F., *Complementi di statistica economica...ecc.*, op. cit..

### 3.7.2 Processo AutoRegressivo - AR(p)

Sia  $\{\varepsilon_t\}$  un processo *white noise* di media nulla e varianza  $\sigma_\varepsilon^2$ . Si dice che  $\{Y_t\}$  è un processo autoregressivo di ordine  $p$ , e lo si indica con  $AR(p)$  (dall'inglese *AutoRegressive*), se:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \dots + \phi_p Y_{t-p} + \varepsilon_t, \quad (3.15)$$

con  $\phi_i$ ,  $i = 0, \dots, p$ , parametri costanti.

Nella (3.15) la variabile  $Y_t$  è ottenuta come il risultato di una somma pesata di valori passati e di uno *shock* casuale contemporaneo.

Utilizzando l'operatore ritardo  $B$ , tale che  $B^h Y_t = Y_{t-h}$ ,  $h = 0, 1, 2, \dots$ , e indicando con:

$$\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$$

l'operatore polinomiale autoregressivo, un processo  $AR(p)$  può essere scritto come:

$$\phi(B)Y_t = \phi_0 + \varepsilon_t.$$

La media di un processo  $AR(p)$  è pari a:

$$E[Y_t] = \frac{\phi_0}{1 - \phi_1 - \dots - \phi_p}.$$

Perché il processo abbia media nulla è dunque sufficiente che  $\phi_0 = 0$ .

La funzione di autocovarianza del processo è data da:

$$\gamma_k = \begin{cases} \phi_1 \gamma_1 + \dots + \phi_p \gamma_p + \sigma_\varepsilon^2 & k = 0 \\ \phi_1 \gamma_{k-1} + \dots + \phi_p \gamma_{k-p} & k > 0 \end{cases} \quad (3.16)$$

Dividendo ambo i membri della (3.16) per  $\gamma_0$ , si ricava che la varianza del processo può essere espressa nel modo seguente:

$$\gamma_0 = \frac{\sigma_\varepsilon^2}{1 - \phi_1 \rho_1 - \dots - \phi_p \rho_p}.$$

Inoltre, per la funzione di autocorrelazione, vale la seguente equazione omogenea alle differenze finite,

$$\rho_k = \phi_1 \rho_{k-1} + \dots + \phi_p \rho_{k-p} \quad k > 0$$

che, scritta per esteso per  $k = 1, 2, \dots, p$ , e tenuto conto che  $\rho_0 = 1$  e

$\rho_k = \rho_{-k}$ , porta al seguente sistema di equazioni lineari:



$$\begin{aligned}
\rho_1 &= \phi_1 + \phi_2 \rho_1 + \phi_3 \rho_2 + \dots + \phi_p \rho_{p-1} \\
\rho_2 &= \phi_1 \rho_1 + \phi_2 + \phi_3 \rho_1 + \dots + \phi_p \rho_{p-2} \\
&\vdots \\
\rho_p &= \phi_1 \rho_{p-1} + \phi_2 \rho_{p-2} + \phi_3 \rho_{p-3} + \dots + \phi_p.
\end{aligned} \tag{3.17}$$

Le (3.17) sono note come equazioni di *Yule-Walker* e consentono di pervenire, dalla funzione di autocorrelazione, ai parametri del processo  $AR(p)$ .

Si può mostrare che in un processo  $AR(p)$  stazionario l'ACF tende ad annullarsi al divergere di  $k$ . In particolare a seconda del valore dei parametri  $\phi_i$ ,  $\rho_k$  tende a zero con un comportamento misto tra l'esponenziale e lo pseudo-periodico. Relativamente alla funzione di autocorrelazione parziale  $P_k$ , si può invece mostrare che, al divergere di  $k$ , essa è diversa da zero per  $k \leq p$  e si annulla per  $k > p$ . Il processo  $AR(p)$  è sempre invertibile: nella (3.15), infatti, la variabile  $Y_t$  è già espressa in funzione dei suoi valori passati.

Quanto alla stazionarietà, si dimostra<sup>26</sup> che un processo  $AR(p)$  è

---

<sup>26</sup> WEI W.W.S., *Time series analysis...ecc.*, op. cit..

stazionario, se tutte le  $p$  radici dell'equazione caratteristica  $\phi(B)=0$  sono, in modulo, maggiori di uno.

### **ESEMPIO: AR(1)**

Un processo  $AR(1)$  senza costante assume la forma:

$$Y_t = \phi_1 Y_{t-1} + \varepsilon_t. \quad (3.18)$$

L'equazione caratteristica  $1 - \phi_1 B = 0$  ha soluzione  $\frac{1}{\phi_1}$ , per cui un processo  $AR(1)$ , che è sempre invertibile, è stazionario se  $|\phi_1| < 1$ .

Quanto ai momenti si ha:

$$E[Y_t] = E[\phi_1 Y_{t-1} + \varepsilon_t] = \phi_1 E[Y_{t-1}] + E[\varepsilon_t].$$

In ipotesi di stazionarietà si ha  $E[Y_t] = E[Y_{t-1}]$  e dunque:

$$E[Y_t] = \phi_1 E[Y_{t-1}]$$

da cui si ricava

$$E[Y_t] = 0.$$

L'autocovarianza è data da:

$$\gamma_k = \begin{cases} \frac{\sigma_\varepsilon^2}{1-\phi_1^2} & k=0 \\ \phi_1 \gamma_{k-1} = \sigma_1^k \gamma_0 & k>0 \end{cases} . \quad (3.19)$$

La (3.19) si ricava moltiplicando ambo i membri della (3.18) per  $Y_{t-k}$  e calcolando il valore atteso:

$$\begin{aligned} \gamma_k &= E[Y_{t-k} Y_t] = E[Y_{t-k} (\phi_1 Y_{t-1} + \varepsilon_t)] \\ &= \phi_1 E[Y_{t-k} Y_{t-1}] + E[Y_{t-k} \varepsilon_t] \\ &= \phi_1 \gamma_{k-1} + E[Y_{t-k} \varepsilon_t] \end{aligned}$$

Poiché:

$$E[Y_{t-k} \varepsilon_t] = \begin{cases} \sigma_\varepsilon^2 & k=0 \\ 0 & k>0 \end{cases}$$

risulta essere  $\gamma_k = \phi_1 \gamma_{k-1}$ . Operando sostituzioni successive, si ricava

$\gamma_k = \phi_1^k \gamma_0$ . Dalla (3.19) si ottiene:

$$\rho_k = \frac{\gamma_k}{\gamma_0} = \phi_1^k \quad k \geq 0.$$

Si noti che  $\rho_1 = \phi_1$  e che la velocità con cui  $\rho_k$  tende a zero

dipende dal valore di  $\phi_1$ . Per valori prossimi a 1 la funzione di autocorrelazione decresce molto lentamente, mentre per valori prossimi a zero  $\rho_k$  si annulla molto velocemente<sup>27</sup>.

La funzione di autocorrelazione parziale è:

$$P_k = \begin{cases} \phi_1 & k = 1 \\ 0 & k > 1 \end{cases}$$

### 3.7.3 Processo a Media Mobile - MA(q)

Sia  $\{\varepsilon_t\}$  un processo *white-noise* di media zero e varianza  $\sigma_\varepsilon^2$ . Si dice che  $\{Y_t\}$  è un processo a media mobile di ordine  $q$  e lo si indica  $MA(q)$  (dall'inglese *Moving Average*), se:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \quad (3.20)$$

con  $\theta_j, j = 0, \dots, q$ , parametri costanti.

Nella (3.20) la variabile  $Y_t$  è pensata come il risultato di una somma pesata di impulsi casuali presenti e passati.

In modo analogo a quanto si è visto per il processo  $AR(p)$ ,

---

<sup>27</sup> DI FONZO T. - LISI F., *Complementi di statistica economica...ecc.*, op. cit..

indicando con:

$$\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$$

l'operatore polinomiale a media mobile, la (3.20) può essere scritta anche come:

$$Y_t = \theta(B)\varepsilon_t.$$

Media e autocovarianza di un processo  $MA(q)$  sono pari rispettivamente a:

$$E[Y_t] = E[\varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}] = 0,$$

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 \left( 1 + \sum_{i=1}^q \theta_i^2 \right) & k = 0 \\ \sigma_\varepsilon^2 \left( -\theta_k + \theta_1 \theta_{k+1} + \theta_2 \theta_{k+2} + \dots + \theta_{q-k} \theta_q \right) & k = 1, \dots, q. \\ 0 & k > q \end{cases}$$

La funzione di autocorrelazione è quindi:

$$\rho_k = \begin{cases} 1 & k = 0 \\ \frac{-\theta_k + \theta_1\theta_{k+1} + \theta_2\theta_{k+2} + \dots + \theta_{q-k}\theta_q}{1 + \theta_1^2 + \dots + \theta_q^2} & k = 1, \dots, q.. \\ 0 & k > q \end{cases}$$

La funzione di autocorrelazione parziale  $P_k$  al divergere di  $k$  tende ad annullarsi con un comportamento misto tra l'esponenziale e lo pseudo-periodico, a seconda del valore dei parametri del processo.

Poiché media, varianza e covarianza sono costanti e non dipendono dal tempo, un processo  $MA(q)$  è sempre stazionario.

La condizione di invertibilità si può ottenere considerando l'equazione caratteristica  $\theta(B)=0$ . Si dimostra<sup>28</sup> che un processo  $MA(q)$  è invertibile se tutte le radici dell'equazione caratteristica sono, in modulo, maggiori di uno.

Indicate tali radici con  $B_i$ , il processo  $MA(q)$  è dunque invertibile se:

$$|B_i| > 1, \quad i = 1, \dots, q.$$

---

<sup>28</sup> HAMILTON J.D., *Econometria delle serie storiche*, Monduzzi, Bologna, 1995

### ESEMPIO: MA(1)

Un processo  $MA(1)$  assume la forma:

$$Y_t = \varepsilon_t - \theta_1 \varepsilon_{t-1} = (1 - \theta_1 B) \varepsilon_t. \quad (3.21)$$

Poiché l'unica radice dell'equazione caratteristica  $1 - \theta_1 B = 0$  è  $B_1 = \frac{1}{\theta_1}$ ,

il processo è invertibile se  $|\theta_1| < 1$ .

Il processo è sempre stazionario, in quanto, media, varianza e covarianza sono costanti e non dipendono dal tempo.

Infatti:

$$E[Y_t] = E[\varepsilon_t - \theta_1 \varepsilon_{t-1}] = E[\varepsilon_t] - \theta_1 E[\varepsilon_{t-1}] = 0;$$

$$\gamma_k = \begin{cases} \sigma_\varepsilon^2 (1 + \theta_1^2) & k = 0 \\ -\sigma_\varepsilon^2 \theta_1 & k = 1. \\ 0 & k > 1 \end{cases} \quad (3.22)$$

La (3.22) si ottiene moltiplicando ambo i membri della (3.21) per  $Y_{t-k}$  e calcolando il valore atteso:

$$\begin{aligned} \gamma_k &= E[Y_{t-k} Y_t] = E[Y_{t-k} (\varepsilon_t - \theta_1 \varepsilon_{t-1})] \\ &= E[Y_{t-k} \varepsilon_t] - \theta_1 E[Y_{t-k} \varepsilon_{t-1}] \end{aligned}$$

Ora, se  $k = 0$ ,

$$E[Y_t \varepsilon_t] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1}) \varepsilon_t] = \sigma_\varepsilon^2,$$

$$E[Y_t \varepsilon_{t-1}] = E[(\varepsilon_t - \theta_1 \varepsilon_{t-1}) \varepsilon_{t-1}] = -\theta_1 \sigma_\varepsilon^2,$$

e dunque

$$\gamma_0 = \sigma_\varepsilon^2 + \theta_1^2 \sigma_\varepsilon^2 = \sigma_\varepsilon^2 (1 + \theta_1^2).$$

Se  $k = 1$ ,

$$E[Y_{t-1} \varepsilon_t] = E[(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2}) \varepsilon_t] = 0,$$

$$E[Y_{t-1} \varepsilon_{t-1}] = E[(\varepsilon_{t-1} - \theta_1 \varepsilon_{t-2}) \varepsilon_{t-1}] = \sigma_\varepsilon^2,$$

e dunque

$$\gamma_1 = -\sigma_\varepsilon^2 \theta_1.$$

Per  $k = 2$  si ha:

$$E[Y_{t-2} \varepsilon_t] = E[(\varepsilon_{t-2} - \theta_1 \varepsilon_{t-3}) \varepsilon_t] = 0,$$

$$E[Y_{t-2} \varepsilon_{t-1}] = E[(\varepsilon_{t-2} - \theta_1 \varepsilon_{t-3}) \varepsilon_{t-1}] = 0.$$

Per  $k > 2$  i valori attesi continuano ad essere nulli e dunque la (3.22) è verificata.

La funzione di autocorrelazione è pari a:



$$\gamma_k = \begin{cases} -\frac{\theta_1}{1 + \theta_1^2} & k = 1 \\ 0 & k > 1 \end{cases} .$$

Infine, per la funzione di autocorrelazione parziale si ha:

$$P_k = \frac{-\theta_1^k (1 - \theta_1^2)}{1 - \theta_1^{2(k+1)}}, \quad k \geq 1,$$

da cui si ricava facilmente che  $\lim_{k \rightarrow \infty} P_k = 0$ <sup>29</sup>.

### 3.7.4 Processo Autoregressivo a Media – ARMA (p,q)

Nelle forme pure dei processi *AR* e *MA* si possono avere un gran numero di parametri.

Una rappresentazione più parsimoniosa può essere ottenuta considerando un modello misto.

Si parte come sempre da un processo *white noise* di media nulla e varianza  $\sigma_\varepsilon^2$ . Si dice che  $\{Y_t\}$  è un processo autoregressivo a media

---

<sup>29</sup> DI FONZO T. - LISI F., Complementi di statistica economica...ecc., op. cit..

mobile di ordine  $(p, q)$ , e lo si indica con  $ARMA(p, q)$  (dall' inglese *AutoRegressive Moving Average*), se si può scrivere:

$$Y_t - \sum_{i=1}^p \phi_i Y_{t-i} = \phi_0 + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}. \quad (3.23)$$

Ovviamente, quando  $p = 0$  si torna ad avere un  $MA(q)$  e quando  $q = 0$  si ha un modello  $AR(p)$ .

Modelli di questo tipo sono stati proposti da Box e Jenkins<sup>30</sup> come una classe standard per il trattamento di serie storiche.

Anche in questo caso la (3.23) può essere espressa in una forma più compatta, utilizzando gli operatori polinomiali  $\theta(B)$  e  $\phi(B)$ :

$$\phi(B)Y_t = \phi_0 + \theta(B)\varepsilon_t.$$

Un processo  $ARMA(p, q)$  è stazionario se tutte le  $p$  radici dell'equazione caratteristica  $\phi(B) = 0$  sono in modulo, maggiori di uno, mentre è invertibile se le  $q$  radici dell'equazione  $\theta(B) = 0$  sono tutte, in modulo, maggiori di uno.

Si ha, inoltre, che:

---

<sup>30</sup> BOX G.E.P. - JENKINS G.M., *Time Series Analysis...ecc.*, op. cit..



### ESEMPIO: ARMA(1,1)

Un processo ARMA(1,1) assume la forma:

$$Y_t = \phi_0 - \phi_1 Y_{t-1} + \varepsilon_t - \theta_1 \varepsilon_{t-1}.$$

Considerando le condizioni di stazionarietà e di invertibilità, è facile vedere che un processo è stazionario se  $|\phi_1| < 1$  ed è invertibile se  $|\theta_1| < 1$ . Relativamente ai momenti del processo si ha:

$$E[Y_t] = \frac{\phi_0}{1 - \phi_1}.$$

Supponiamo ora che  $\phi_0 = 0$ , cosicché  $E[Y_t] = 0$ . In questo caso si ha:

$$\text{Var}[Y_t] = \gamma_0 = \frac{(1 + \theta_1^2 - 2\theta_1\phi_1)}{1 - \phi_1^2} \sigma_\varepsilon^2.$$

La funzione di autocovarianza è:

$$\gamma_k = \begin{cases} \frac{(\phi_1 - \theta_1)(1 - \phi_1\theta_1)}{1 - \phi_1^2} \sigma_\varepsilon^2 & k = 1 \\ \phi_1 \gamma_{k-1} = \phi_1^{k-1} \gamma_1 & k \geq 2 \end{cases}$$

Come si vede, per  $k > 1$  la funzione di autocovarianza è quella di un processo  $AR(1)$ . La funzione di autocorrelazione risulta essere:

$$\rho_k = \begin{cases} 1 & k = 0 \\ \frac{(\phi_1 - \theta_1)(1 - \phi_1 \theta_1)}{1 + \theta_1^2 - 2\theta_1 \phi_1} & k = 1 \\ \phi_1^{k-1} \rho_1 & k \geq 2 \end{cases}$$

Per  $k = 1$ , pertanto, la funzione di autocorrelazione di un processo  $ARMA(1,1)$  combina insieme le caratteristiche, tanto di un processo  $AR(1)$ , quanto di un processo  $MA(1)$ . Per  $k > 1$ , invece, la funzione di autocovarianza e quella di autocorrelazione tendono ad annullarsi, come per un processo  $AR(1)$ .

Infine, per la funzione di autocorrelazione parziale di un  $ARMA(1,1)$  si ha  $P_1 = \rho_1$ ; per  $k > 1$  l'autocorrelazione parziale decade a zero esponenzialmente al crescere di  $k$ <sup>32</sup>.

### 3.8 Processi non stazionari

I modelli di serie storiche fin qui considerati sono tutti stazionari.

---

<sup>32</sup> DI FONZO T. - LISI F., Complementi di statistica economica...ecc., op. cit..

Tuttavia è piuttosto frequente il caso di serie per le quali è ragionevole ipotizzare un processo non stazionario. Limitandosi ai primi due momenti, la non stazionarietà può dipendere dal fatto che la media non è costante e/o che non sono costanti i momenti secondi. Nel primo caso si parla di non stazionarietà in media, e tipicamente si verifica in presenza di un trend, nel secondo si parla di non stazionarietà in varianza (o in covarianza). Al fine di poterle esaminare mediante l'approccio moderno all'analisi delle serie storiche, è necessario renderle stazionarie tramite alcuni accorgimenti. La stabilizzazione di una serie avviene attraverso la trasformazione funzionale della serie osservata e quindi attraverso l'isolamento della componente deterministica dalla componente di errore. La trasformazione può essere di diverso tipo:

- trasformazione logaritmica;
- utilizzo della media mobile della serie;
- impiego delle variazioni relative fra un periodo ed il precedente al posto del valore assoluto del fenomeno osservato.

La trasformazione logaritmica viene utilizzata per stabilizzare una serie in quanto l'operazione produce l'effetto di diminuire la variabilità, poiché essa cresce in termini assoluti al crescere del trend. Ad esempio, una variabilità del 5% su valori elevati è maggiore in termini assoluti della stessa percentuale applicata però a valori bassi della serie. In altri

termini, mentre la serie originaria può essere pensata come il risultato moltiplicativo di più componenti, la trasformazione logaritmica rende additiva tale relazione al trascorrere del tempo.

Anche l'utilizzo delle medie mobili ha un effetto stabilizzante. Infatti l'utilizzo di un valore centrale degli  $n$  termini precedenti produce l'effetto di eliminare la variabilità all'interno degli  $n$  termini scelti per il calcolo della media. Dal punto di vista grafico il risultato dell'impiego delle medie mobili si traduce in una "smussatura" del tracciato del grafico e in uno spostamento della curva verso il centro o verso destra, a seconda che si utilizzino rispettivamente medie mobili centrate o medie mobili degli  $n$  termini precedenti.

Infine, spesso può essere utile l'utilizzo delle variazioni che il fenomeno osservato subisce nel tempo tra un periodo e il precedente (la cosiddetta differenziazione), piuttosto che considerare il valore assoluto del fenomeno. Tale trasformazione è alla base di tutti i processi integrati illustrati nel sottoparagrafo 3.6.2.

Consideriamo di seguito alcuni processi non stazionari.

### **3.8.1 Processo passeggiata casuale (*random walk*)**

Sia  $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$ . Si dice che  $\{Y_t\}$  è una passeggiata casuale (*random walk*) se:

$$Y_t = Y_{t-1} + \varepsilon_t, \quad Y_0 = \mu \quad (3.24)$$

ovvero se si sta considerando un processo  $AR(1)$  di parametro  $\phi_1 = 1$ .

Essendo  $Y_t = \mu + \sum_{i=1}^t \varepsilon_i$ , si ha:

- $E[Y_t] = \mu$ ,
- $Var[Y_t] = t\sigma_\varepsilon^2$ .

Poiché la varianza cambia nel tempo e, in particolare, cresce con esso, il processo è non stazionario. La non stazionarietà del processo può anche essere desunta dal fatto che la (3.24) definisce un processo  $AR(1)$  con  $\phi_1 = 1$ ; esso, evidentemente, non rispetta la condizione di stazionarietà  $|\phi_1| < 1$ .

Consideriamo, ora, la funzione di autocorrelazione. Essa è pari a:

$$\rho_k = \sqrt{\frac{t}{t+k}}, \quad k = 0, 1, 2, \dots$$

Per un processo *random walk*, dunque, per ogni prefissato valore di  $k$ ,  $\rho_k \rightarrow 1$  al crescere di  $t$ . Per questo si dice anche che il processo ha memoria infinita.

Analogamente la funzione di autocorrelazione parziale è pari a:



$$P_k = \begin{cases} 1 & k = 1 \\ 0 & k > 1 \end{cases} .$$

Infine, indicato con  $\Delta_B = 1 - B$  l'operatore differenza prima, tale per cui  $\Delta_B Y_t = Y_t - Y_{t-1}$ , si osservi che  $\Delta_B Y_t$  è invece un processo *white noise* che, come sappiamo, è stazionario. La (3.24) costituisce, pertanto, un esempio di processo non stazionario che può essere reso stazionario tramite una opportuna trasformazione<sup>33</sup>.

### 3.8.2 Processo AutoRegressivo Integrato a Media Mobile - ARIMA (p,d,q)

Box e Jenkins<sup>34</sup> hanno proposto un'estensione dei modelli ARMA che riguarda certi speciali tipi di processi non stazionari. Essi, cioè, considerano processi omogenei non stazionari di grado  $d$  (ovvero processi che diventano stazionari dopo  $d$  differenziazioni successive).

Sia ancora  $\{\varepsilon_t\}$  un processo *white-noise* di media zero e varianza  $\sigma_\varepsilon^2$ . Si indichi con  $X_t$  la  $d$ -esima differenza di  $Y_t$ ,

---

<sup>33</sup> DI FONZO T. - LISI F., *Complementi di statistica economica...ecc.*, op. cit..

<sup>34</sup> BOX G.E.P. - JENKINS G.M., *Time Series Analysis...ecc.*, op. cit..

$X_t = (1-B)^d Y_t = \Delta_B^d Y_t$ . Si dice che  $\{Y_t\}$  è un processo autoregressivo integrato a media mobile di ordine  $(p, d, q)$  e lo si indica con  $ARIMA(p, d, q)$  (da *AutoRegressive Integrated Moving Average*), se  $\{X_t\}$  è un processo ARMA  $(p, q)$ . In sintesi valgono le seguenti relazioni<sup>35</sup>:

$$\begin{aligned} X_t &= \Delta_B^d Y_t, \\ X_t &= \sum_{i=1}^p \phi_i X_{t-i} + \varepsilon_t - \sum_{j=1}^q \theta_j \varepsilon_{t-j}. \end{aligned} \quad (3.25)$$

Utilizzando l'operatore differenza ed i polinomi, rispettivamente, autoregressivo e a media mobile, la (3.25) può anche essere scritta in termini di  $Y_t$ :

$$\phi(B) \Delta_B^d Y_t = \theta(B) \varepsilon_t.$$

È evidente, che per  $d = q = 0$  la classe di processi  $ARIMA(p, d, q)$  si riduce a quella dei processi  $AR(p)$ ; per  $p = d = 0$  si ottiene la classe dei processi  $MA(q)$  e per  $d = 0$ ,  $Y_t \sim ARMA(p, q)$ . Quando poi si considera il caso molto particolare  $p = d = q = 0$ , si ottiene il processo

---

<sup>35</sup> DI FONZO T. - LISI F., *Complementi di statistica economica...ecc.*, op. cit..

*white-noise*. È immediato riconoscere, infine, che un modello *random walk* non è altro che un ARIMA (0,1,0), per questo esso viene anche indicato con la sigla I (1).

# CAPITOLO 4

## Analisi temporale dei dati mediante la procedura di Box e Jenkins

### 4.1 La procedura di Box e Jenkins

*Box e Jenkins*<sup>36</sup> hanno proposto un metodo per costruire, a partire dai dati, un modello ARIMA che rappresenti adeguatamente il processo generatore di una serie storica.

Esso consta di tre fasi fondamentali:

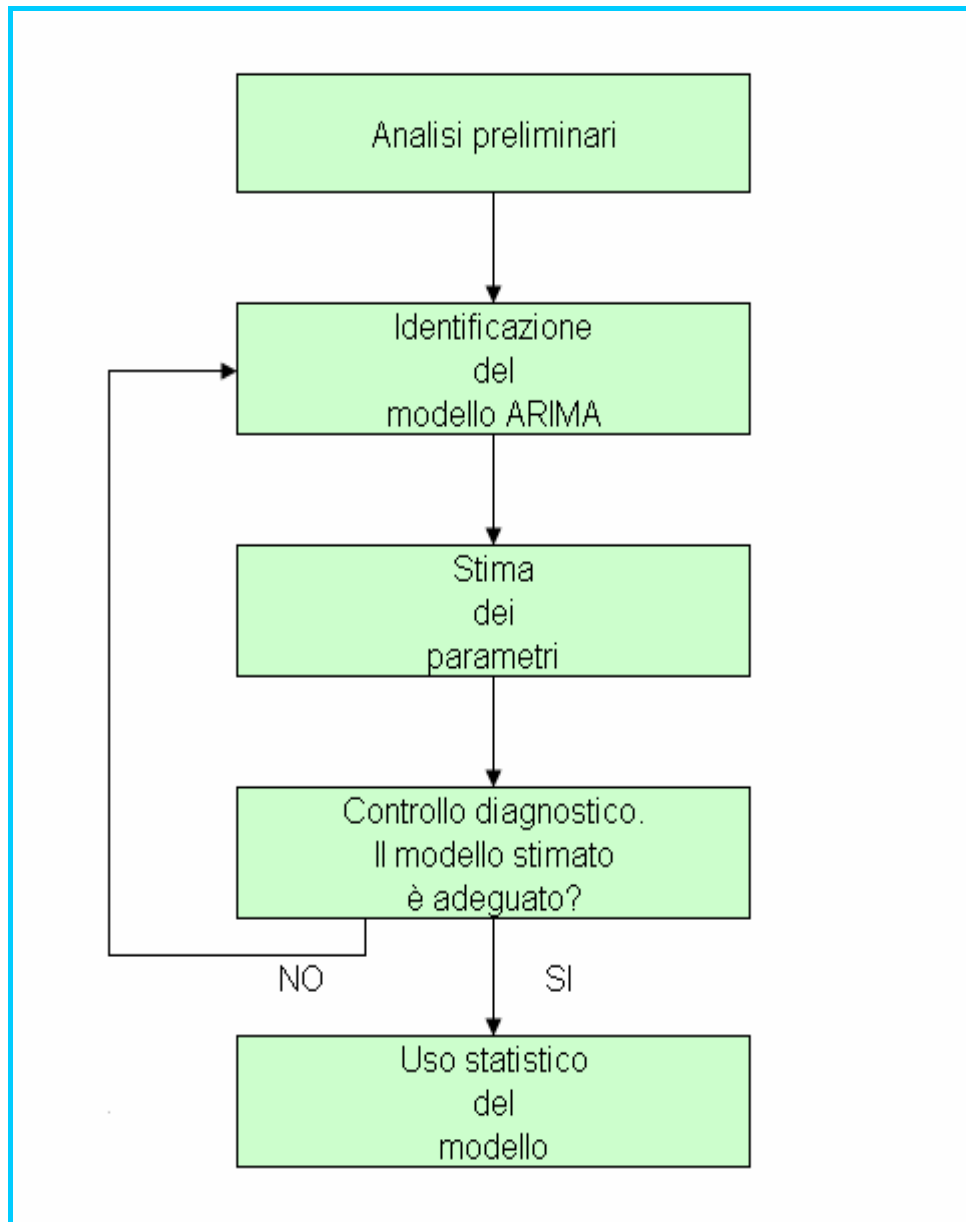
- a) Identificazione;
- b) stima dei parametri;
- c) controllo diagnostico;

---

<sup>36</sup> *BOX G.E.P. - JENKINS G.M., Time Series Analysis...ecc., op. cit..*

In alcuni casi esse possono essere ripetute più volte in maniera iterativa secondo lo schema riportato nella *Figura 12*.

**Figura 12: Schema logico della procedura di Box e Jenkins**



### 4.1.1 Identificazione

Il primo passo consiste nella specificazione dell'ordine del modello, ovvero nell'identificazione dei parametri  $p, d, q$ .

I principali strumenti da usare sono la funzione di autocorrelazione e la funzione di autocorrelazione parziale (entrambe stimate sul campione).

L'idea di base è quella di riconoscere nella struttura della funzione di autocorrelazione empirica la struttura di una funzione di autocorrelazione teorica. Con questo fine, le relazioni viste nei capitoli precedenti possono essere riassunte nel modo seguente.

- La costruzione di un modello ARMA poggia sull'assunto che il processo sia stazionario. Se il processo generatore della serie è non stazionario è spesso opportuna una delle trasformazioni descritte nel capitolo precedente; in particolare, nel caso di un trend la non stazionarietà può essere rimossa differenziando  $d$  volte i dati fino ad ottenere una serie stazionaria. Una volta che la serie appaia ragionevolmente stazionaria si può cercare di identificare uno specifico modello.
- Per un processo  $AR(p)$ , l'ACF  $\rho_k$  tende a zero con un comportamento che dipende dal valore dei parametri  $\phi_i$ , mentre la

PACF  $P_k$  è diversa da zero per  $k \leq p$  e si annulla per  $k \geq p + 1$ .

- Per un processo MA ( $q$ ), l'ACF  $\rho_k$  è diversa da zero per  $k \leq q$  e si annulla per  $k \geq q + 1$ .

L'autocorrelazione parziale, invece, tende a zero. Il modo con cui tende a zero dipende dal valore dei parametri  $\theta_j$ .

- Quando non c'è un chiaro segno di trovarsi nelle due condizioni precedenti è probabilmente più appropriato un modello ARMA. L'utilizzo di un modello misto ha anche il vantaggio di rendere il modello più parsimonioso in termini di numero di parametri.

In quest'ultimo caso, ed in generale per evitare il rischio di sovrapparametrizzazione (*overfitting*), è utile fare ricorso a criteri diversi per l'identificazione del modello. Infatti, se da una parte la varianza residua diminuisce al crescere di  $p$  e  $q$ , dall'altra un buon adattamento non conduce necessariamente ad una buona capacità previsiva del modello stesso. Per ovviare a questo inconveniente sono stati introdotti dei criteri che tentano di evitare la sovrapparametrizzazione, assegnando un "costo" all'introduzione di ogni nuovo parametro addizionale. Uno di questi è il cosiddetto AIC (*Asymptotic Information Criterion, Akaike*)<sup>37</sup>,

---

<sup>37</sup> AKAIKE H., *A new look at the statistical model identification*, IEEE Transactions on Automatic Control, 1974.

dato da:

$$AIC(k) = -\frac{2}{n}(\log L(\hat{\delta}) - k),$$

in cui  $k = p + q + 2$  è il numero dei parametri del modello,

$$\hat{\delta} = (\hat{\phi}_0, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q, \hat{\sigma}_\varepsilon^2)'$$

è il vettore ( $k \times 1$ ) contenente i parametri stimati.

La quantità  $\frac{2k}{n}$  rappresenta una sorta di “penalità”, legata al

numero di parametri del modello, che diminuisce al crescere di  $n$ .

I valori prescelti di  $p$  e  $q$  sono quelli che minimizzano AIC<sup>38</sup>.

#### 4.1.2 Stima dei parametri

Identificati gli ordini  $p$  e  $q$  del modello si può passare alla fase di stima dei parametri che lo caratterizzano, ossia alla stima del vettore:

$$\delta = (\phi_0, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_\varepsilon^2)'$$

---

<sup>38</sup> AKAIKE H., *A new look...ecc., op. cit.*



Il metodo di stima più usato è il metodo della massima verosimiglianza<sup>39</sup>.

Esso consiste nel massimizzare una funzione di verosimiglianza, definita in base alla probabilità di osservare una data realizzazione campionaria, condizionatamente ai valori assunti dai parametri oggetti di stima.

Si consideri, a tal proposito, una distribuzione di probabilità  $D$ , con funzione di probabilità  $L_D$  caratterizzata da un parametro  $\delta$  ed un campione di dati  $\{x_i\}_{i=1}^n$  di dimensione  $n$ ; la probabilità associata ai dati osservati è data da:

$$\Pr(\{x_i\}_{i=1}^n) = L_D(x_1, \dots, x_n | \delta)$$

Il metodo della massima verosimiglianza ricerca il valore più verosimile di  $\delta$ , ossia massimizza la verosimiglianza dei dati osservati, sullo spazio di tutti i possibili valori di  $\delta$ , dato da  $\Theta$ <sup>40</sup>.

Da un punto di vista matematico,  $L_D(x_1, \dots, x_n | \delta)$  è detta funzione di verosimiglianza, e lo stimatore di massima verosimiglianza è

---

<sup>39</sup> PARPINEL F. – PROVASI C., *Probabilità e statistica per le scienze economiche*, Giappichelli, Torino, 1999.

<sup>40</sup> AZZALINI A., *Inferenza statistica Una presentazione basata sul concetto di verosimiglianza*, Springer, Milano, 2001.

è ottenuto come:

$$\hat{\delta} = \max_{\delta \in \Theta} L_D(x_1, \dots, x_n | \delta).$$

Se, come spesso accade, la funzione di verosimiglianza assume una forma particolarmente complicata, è possibile ricorrere ad opportune approssimazioni che rendono i calcoli più semplici al prezzo di una contenuta perdita di precisione.

Inoltre, il carattere ricorsivo delle relazioni tra  $Y_t$  e  $Y_{t-k}$  e fra  $\varepsilon_t$  e  $\varepsilon_{t-k}$ ,  $k = 1, 2, 3, \dots$ , implica la perdita delle osservazioni iniziali, per cui è necessaria una ricostruzione di tali dati, denominata *back-forecasting* da *Box e Jenkins*<sup>41</sup>.

Una possibile alternativa consiste nel porre pari a zero i valori iniziali mancanti: in questo caso le stime che si ottengono vengono dette condizionate.

Attualmente, però, quasi tutti i programmi di calcolo (compreso il software “R 2.1.0” che andremo ad utilizzare per l’analisi) sono in grado di produrre stime dei parametri massimizzando la verosimiglianza esatta di un processo ARMA.

---

<sup>41</sup> *BOX G.E.P. - JENKINS G.M., Time Series Analysis...ecc., op. cit..*

Questa può essere ottenuta in vari modi, tutti essenzialmente volti a trasformare la funzione di verosimiglianza delle osservazioni originarie in quella di osservazioni gaussiane indipendenti.

Fissato come livello di significatività  $\alpha = 0,05$ , una regola operativa di pratica utilità consiste nel ritenere un parametro significativo, quindi accettabile, quando il rapporto tra la sua stima ed il suo errore standard sia non minore di 1,96 in valore assoluto (dove 1,96 rappresenta il valore soglia della variabile casuale normale standardizzata al livello di significatività prefissato)<sup>42</sup>.

L'eventuale significatività di un parametro non basta ovviamente a formulare un giudizio sulla validità del modello nel suo complesso.

Per fare ciò, è necessario procedere alla verifica del modello secondo quanto specificato nel seguito.

### **4.1.3 Controllo diagnostico**

Il passo finale nella costruzione del modello è quello di controllare l'adeguatezza del modello stesso, facendo uso di test diagnostici.

Alla base di tali test c'è la considerazione che, se il modello è stato

---

<sup>42</sup> DI FONZO T. - LISI F., *Complementi di statistica economica...ecc.*, op. cit..

correttamente identificato e stimato, allora sui residui  $e_t = (y_t - \hat{y}_t)$  devono potersi riscontrare le ipotesi fatte a priori sui disturbi  $\varepsilon_t$ . In particolare, i residui devono risultare incorrelati.

Alcuni tipi di analisi diagnostiche sono le seguenti:

- Analisi grafiche: se il modello ARIMA prescelto è un'adeguata rappresentazione del processo generatore dei dati, la serie dei residui non dovrebbe mostrare alcun tipo di regolarità né valori particolarmente diversi gli uni dagli altri. Al riguardo, utili indicazioni si possono ottenere dal diagramma di dispersione dei punti  $(e_t, e_{t+1})$ , che dovrebbe dar luogo ad una nuvola di punti senza alcuna struttura;
- Autocorrelazioni dei residui: la serie dei residui stimati  $e_t$  può essere trattata come una serie storica a sé stante per la quale è possibile calcolare la funzione di autocorrelazione empirica.

Per ogni valore di  $k$  è possibile utilizzare  $\hat{\rho}_k$  per verificare se  $\rho_k$  è significativamente diversa dalla autocorrelazione di un processo *white noise* che, come è noto (cfr. 3, § 3.7.1), è nulla.

Poiché per un *white noise* la varianza di  $\hat{\rho}_k$  è approssimativamente  $1/n$  per ogni  $k$  (cfr. 3, § 3.7.1), sotto l'ipotesi di normalità la regione

di accettazione, al livello di significatività del 5%, è data da  $[-1,96/\sqrt{n}, 1,96/\sqrt{n}]$ .

Un valore di  $\hat{\rho}_k$  fuori da questo intervallo porta a ritenere  $\rho_k$  significativamente diverso da zero.

Questo risultato è utile perché, permette di studiare l'ipotesi di incorrelazione seriale, tracciando le due rette parallele all'asse delle ascisse in  $-1.96/\sqrt{n}, 1,96/\sqrt{n}$  e osservando il comportamento della funzione di autocorrelazione empirica al variare di  $k$ .

- Test di tipo portmanteau: per una verifica complessiva di assenza di autocorrelazione si può utilizzare anche la seguente statistica sviluppata da *Ljung e Box*<sup>43</sup>:

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

dove  $m$  è il numero di autocorrelazioni prese in esame. Se il modello ARMA è correttamente specificato, allora  $Q(m)$  è ben approssimata ad una variabile casuale  $\chi^2_{m-p-q}$ ;

- Test di normalità dei residui: infine, può essere utile verificare se i

---

<sup>43</sup> LJUNG G.M. – BOX G.E.P, “On a measure of lack of fit in time series models”, *Biometrika*, 65, N. 2, 1978.

residui sono distribuiti normalmente. Tale aspetto è molto importante perché, nel caso di gaussianità l'incorrelazione dei residui implica anche la loro indipendenza. In ultima analisi, ciò significa che il modello lineare è in grado di spiegare l'intera struttura di dipendenza seriale della serie in esame. Un controllo preliminare sulla normalità dei residui può avvenire attraverso la rappresentazione del *normal plot*, che è il grafico dei punti  $(q_s, q_n)$ , dove  $q_s$  è il quantile di ordine  $q$  dei residui, mentre  $q_n$  è il quantile di ordine  $q$  di una normale standardizzata  $N(0,1)$ . Più i punti si concentrano sulla bisettrice, più i dati sono vicini ad una gaussiana standardizzata. Un'ulteriore conferma sulla normalità dei residui del modello si ha considerando il test di *Shapiro-Wilk*, che è ritenuto uno dei test più potenti per tale verifica, soprattutto per piccoli campioni. Il controllo della normalità avviene confrontando due stimatori alternativi della varianza  $\sigma^2$ : uno stimatore non parametrico basato sulla combinazione lineare ottimale della statistica d'ordine di una variabile aleatoria normale, al numeratore e l'usuale stimatore parametrico, ovvero la varianza campionaria, al denominatore<sup>44</sup>.

Supponendo che  $X \sim N(\mu, \sigma^2)$ , la statistica test di *Shapiro-Wilk*,

indicata con  $W$ , si calcola come segue:

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

dove gli  $a_i$ ,  $i = 1, \dots, n$  sono i pesi riportati in apposite tavole.

La regione di rifiuto dell'ipotesi nulla di normalità della serie in favore dell'ipotesi di non normalità della stessa, è data da:

$$R = \{x: W < w_\alpha\},$$

per opportuni valori critici di  $w_\alpha$ <sup>44</sup>.

## **4.2 Specificazione di modelli stocastici applicati ai consumi irrigui nell'agro di Foggia in relazione alle due principali colture irrigue praticate in quest'area.**

Enunciata la metodologia che si andrà ad utilizzare nell'analisi dei dati, si procede allo studio temporale dei consumi irrigui, quindi alla

---

<sup>44</sup> DI FONZO T. - LISI F., Complementi di statistica economica...ecc., op. cit..

specificazione di un modello stocastico relativo alla coltura irrigua considerata.

Di seguito si condurrà l'analisi, considerando le due principali colture irrigue praticate nell'Agro di Foggia.

È importante osservare che, quando si lavora con dati reali, è molto improbabile avere delle funzioni di autocorrelazione, totale o parziale, perfettamente nulle dopo un certo numero di ritardi. Si avranno piuttosto dei valori molto bassi, coerenti con l'ipotesi che  $\rho_k$  sia significativamente diversa da zero.

La nullità o meno della funzione di autocorrelazione va dunque sempre riferita alle fasce di confidenza tipiche di un processo *white noise*.

#### **4.2.1 Olivo**

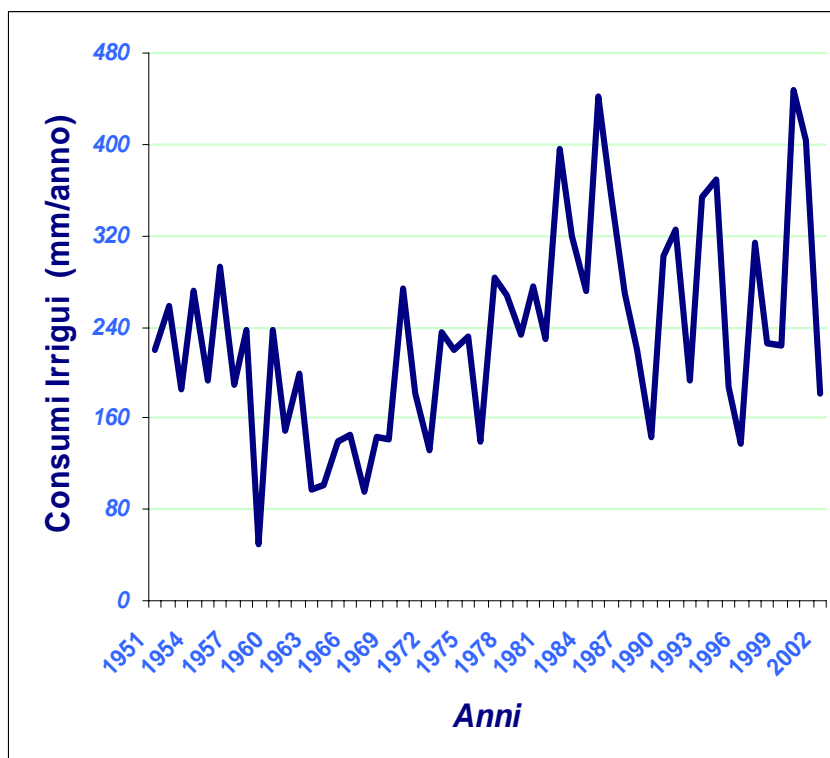
Si comincia, innanzitutto, col rappresentare graficamente la serie storica  $\{y_t\}_{t=1}^{52}$  relativa ai consumi irrigui della coltura dell'olivo nell'Agro di Foggia dal 1951 al 2002, che indichiamo con  $Y_1$ .

Osservando la *Figura 13*, si nota che la serie presenta una lieve decrescita fino all'anno 1968 e una graduale risalita da quest'anno in poi.



Tale crescita si interrompe nel 1989, anno particolarmente piovoso, per poi mantenersi a livelli piuttosto alti negli anni successivi.

**Figura 13 - Andamento dei consumi irrigui (mm/anno) relativi alla coltura dell'olivo dal 1951 al 2002 nell'Agro di Foggia**



Le motivazioni di questo andamento sono dovute al decorso meteorologico: infatti, tra il 1951 e il 1968 il regime evapotraspirativo è stato diverso da quello che si è registrato successivamente. A partire dagli anni settanta, l'evapotraspirazione tende ad aumentare e di conseguenza anche i consumi irrigui. Tale andamento è stato osservato anche per variabili indipendenti (come, ad esempio, la temperatura) e a

diverse latitudini. Alcuni ricercatori ipotizzano che questo fenomeno sia legato all'effetto serra che si è accentuato nell'ultimo trentennio dello scorso secolo, in corrispondenza di un maggiore uso di combustibili. Altre ipotesi correlano l'aumento del regime termico alla diversa tecnica di rilievo dei dati agrometeorologici (introduzione di strumentazione elettronica in sostituzione di quella meccanica tradizionale ed espansione delle aree urbane intorno a parchi agrometeorologici).

Ciò premesso, le principali caratteristiche della serie possono essere così riassunte:

**Tabella 5: Principali caratteristiche della serie storica relativa ai consumi irrigui dell'olivo (mm/anno)**

Numero oss.	Minimo	Media	Massimo	Varianza
52	49	233,3	448	8210,5

In particolare, si noti che il minimo valore della variabile oggetto di studio si è determinato in corrispondenza dell'anno 1959, mentre il massimo valore della stessa si è riscontrato nell'anno 2000.

Ricordando quanto detto nel secondo capitolo (cfr. 2, § 2.3), ovvero, che si irriga quando la pioggia non è in grado di compensare le perdite determinate dall'evapotraspirazione, il minimo valore della variabile in esame per il periodo considerato, verificatosi nel 1959, è

attribuibile ad un valore molto basso, (51,5 mm), della differenza fra il valore di evapotraspirazione (505 mm) della coltura in esame e quello di pioggia (435,5mm), ottenuto quest'ultimo cumulando i dati giornalieri di pioggia per la stagione irrigua dell'olivo (aprile-settembre). Osservando tale differenza, si nota che, ad un valore di evapotraspirazione dell'olivo rientrante nella media annua (pari a 505 mm), corrisponde un valore di pioggia molto alto (pari a 435,5 mm), che ha permesso un minore consumo irriguo per la coltura considerata nell'anno 1959.

Considerando le stesse variabili, nel 2000, (anno in cui si è verificato il massimo valore dei consumi irrigui per l'olivo nell'Agro di Foggia), la differenza fra il valore di evapotraspirazione dell'olivo e quello di pioggia è risultata molto elevata e pari a  $(681-118,4)=562,6$  mm.

Poiché nell'anno 2000 ad un'evapotraspirazione molto elevata dell'olivo è corrisposta una stagione irrigua scarsamente piovosa, i consumi irrigui sono risultati molto alti per l'olivo.

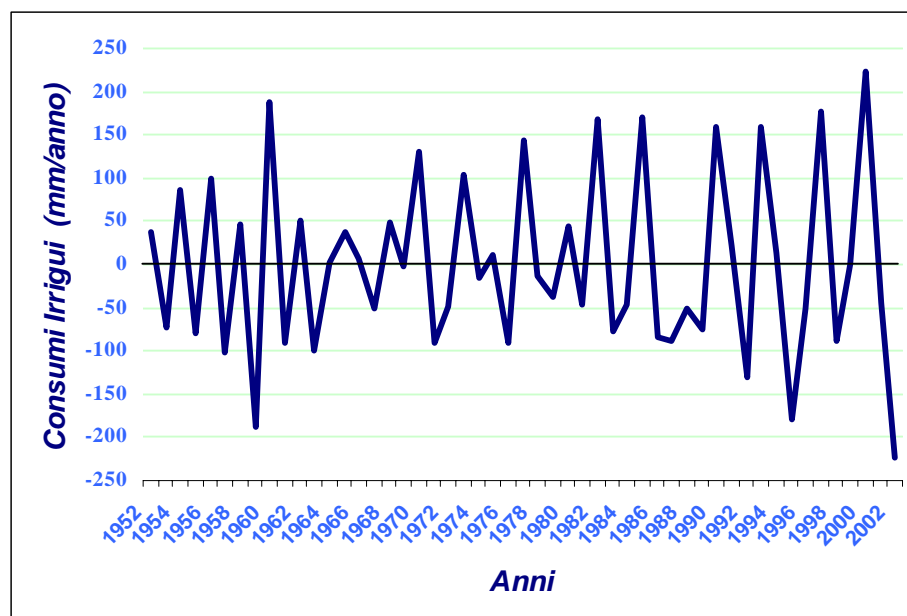
L'andamento assunto dalla serie storica considerata, permette di dire che la serie  $Y_1$  non è stazionaria.

Come è noto, la costruzione di un modello ARMA poggia sull'assunto che il processo sia stazionario, per cui, prima di procedere

all'identificazione del processo generatore della serie  $Y_1$ , è necessario rendere la serie stazionaria mediante differenziazione.

Osservando il grafico della serie differenziata (*Figura 14*), si può notare che la non stazionarietà è stata eliminata; tale grafico, infatti, assume un andamento altalenante intorno allo zero.

**Figura 14 - Serie storica differenziata dei consumi irrigui (mm/anno) relativi alla coltura dell'olivo dal 1952 al 2002 nell'Agro di Foggia**



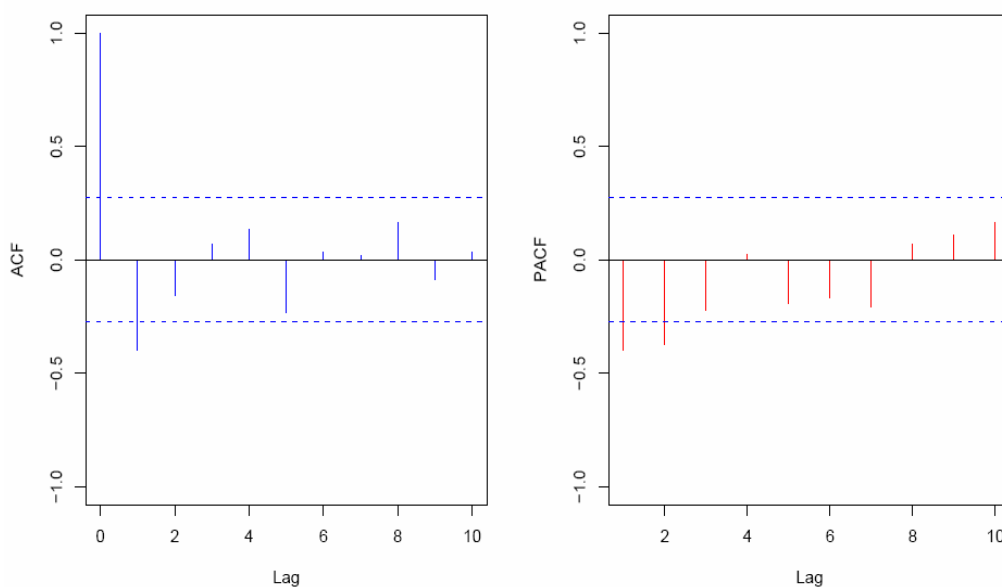
Dal momento che la dinamica della serie differenziata appare coerente con un processo generatore stazionario, possiamo procedere alla specificazione di un modello della classe ARMA.

I correlogrammi stimati per la serie differenziata considerata sono riportati in *Figura 15*; poiché  $n = 52$ , gli estremi dell'intervallo di

confidenza per l'ACF e il PACF di un processo *white noise* vengono posti pari a  $\pm 1,96/\sqrt{52} \cong \pm 0,27$ .

Per quanto riguarda i *lags*, si è già detto (cfr. 3, § 3.6), che è necessario che il numero di ritardi sia minore o uguale di  $\frac{n}{4} = \frac{52}{4} = 13$ , si è assunto, pertanto,  $k = 10$ .

**Figura 15 - Correlogrammi stimati per la serie differenziata dei consumi irrigui relativi alla coltura dell'olivo**



In Figura 15, il grafico della stima della funzione di autocorrelazione  $\hat{\rho}_k$  mette in evidenza che quest'ultima è statisticamente nulla dopo il primo ritardo (si osservi che il primo coefficiente di correlazione è disegnato al ritardo 0 e vale 1, poiché rappresenta la

correlazione della serie con se stessa), mentre il grafico della funzione di autocorrelazione parziale  $\hat{P}_k$  tende a diminuire all'aumentare di  $k$ . Si osservi che quest'ultimo "inizia" al ritardo 1, non al ritardo 0 come quello prodotto da  $\hat{\rho}_k$ .

Tale risultato induce a pensare che la serie differenziata in questione sia generata da un modello MA(1); ciò implica che la serie iniziale  $Y_1$  dei consumi irrigui riferiti alla coltura dell'olivo sia identificata da un processo ARIMA(0,1,1), ovvero da un processo autoregressivo integrato a media mobile di ordine (0,1,1).

Una volta individuato il modello ARIMA generatore della serie storica  $Y_1$ , è possibile passare alla fase successiva della procedura di *Box-Jenkins*, rappresentata dalla stima dei parametri del modello considerato.

**Tabella 6 - Stima dei parametri del modello**

**ARIMA(0,1,1)**

Coefficiente
$\hat{\theta}_1$
0,7668
Deviazione standard
$\hat{\sigma}_{\theta_1}$
0,0922

I risultati della stima, ottenuti con il metodo della massima verosimiglianza, sono riportati nella Tabella 6.

Poiché la stima del parametro  $\theta_1$ , è data da  $\hat{\theta}_1 = 0,7668$ , il modello stimato è pertanto:

$$(1 - B)Y_t = (1 - 0,7668 B)\varepsilon_t$$

con  $\hat{\sigma}_\varepsilon^2 = 6821$ . B, come si è già detto nel capitolo precedente, rappresenta l'operatore ritardo, tale che  $B^h Y_t = Y_{t-h}$ ,  $h = 0,1,2,\dots$

Nella Tabella 6 compare anche la stima della deviazione standard del parametro stimato,  $\hat{\sigma}_{\hat{\theta}_1}$ , necessaria al calcolo della statistica test  $t$  per la verifica di ipotesi  $H_0 : \theta_1 = 0$ , ossia:

$$t = \frac{\hat{\theta}_1}{\hat{\sigma}_{\hat{\theta}_1}} = \frac{0,7668}{0,0922} = 8,32$$

Fissato  $\alpha = 0,05$ , poiché  $|t| > 1,96$ , si può affermare che il parametro è significativo.

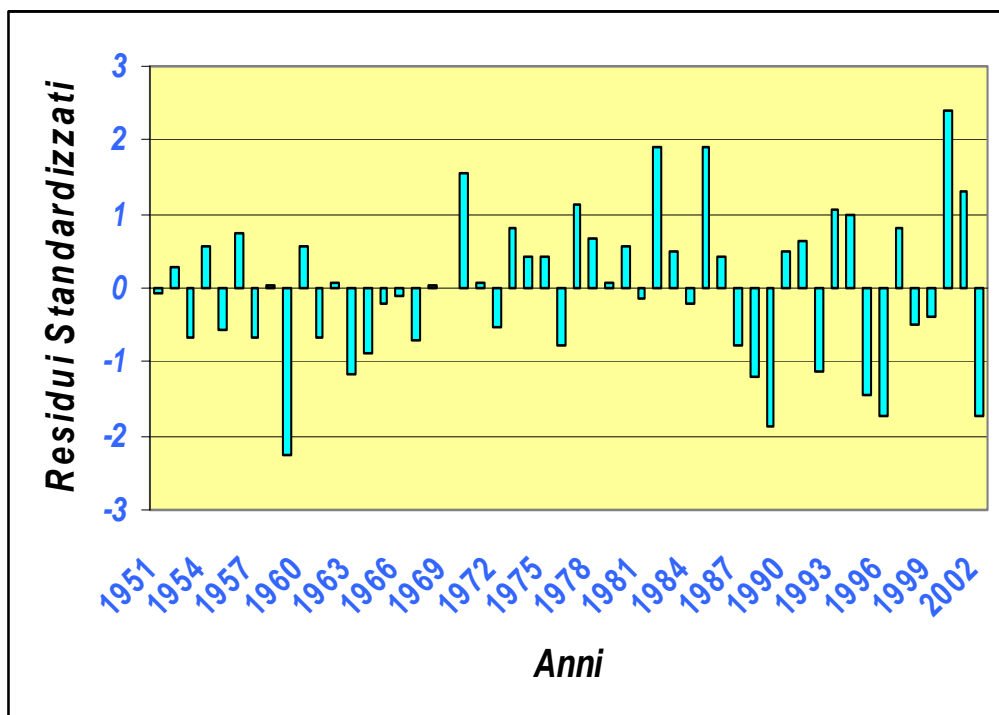
Il passo seguente consiste nella valutazione dell'adeguatezza del modello stimato tramite lo studio dei residui definiti come:

$$e_t = (y_t - \hat{y}_t)$$

dove  $\hat{y}_t$  è il valore ottenuto dal modello stimato una volta sostituite le stime dei parametri.

Affinché il modello stimato risulti adeguato, è necessario che i residui risultino incorrelati, quindi che il processo generatore degli  $e_t$  sia un *white noise*; tale verifica può essere fatta considerando opportune misure diagnostiche. Generalmente, è preferibile operare sui residui standardizzati allo scopo di lavorare con numeri puri.

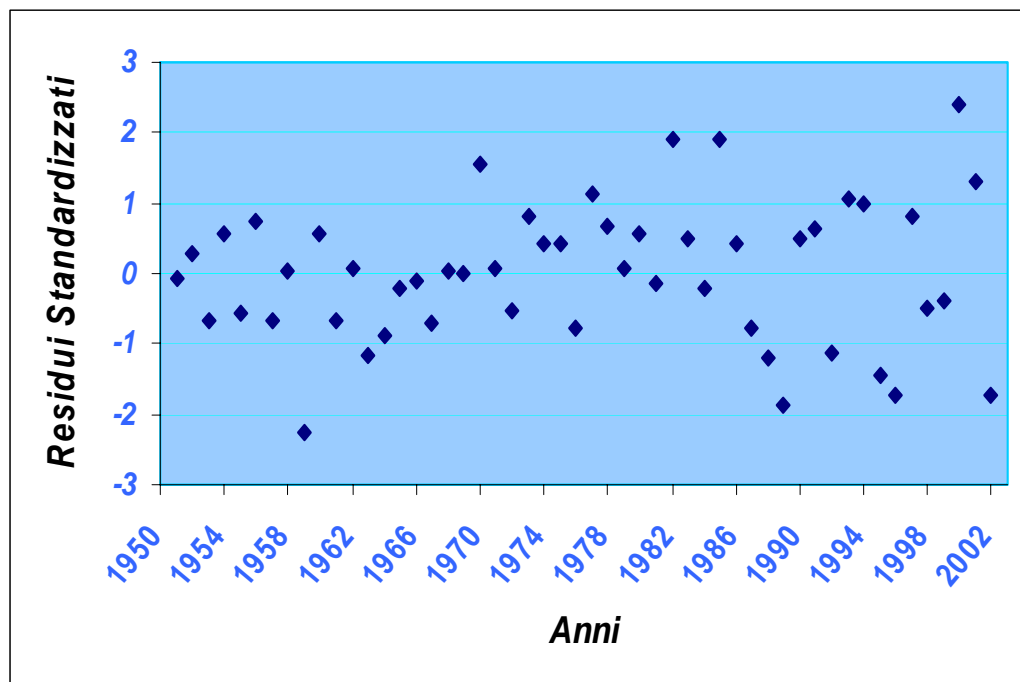
**Figura 16 - Serie storica dei residui standardizzati del modello stimato per la serie  $Y_1$**





Rappresentando graficamente la serie storica dei residui standardizzati (*Figura 16*), si nota che essa ha un andamento del tutto casuale, privo di ogni tipo di regolarità; anche il diagramma di dispersione degli stessi, rappresentato in *Figura 17*, dà luogo ad una nuvola di punti senza alcuna struttura.

**Figura 17 - Grafico di dispersione dei residui standardizzati del modello stimato per la serie  $Y_1$**

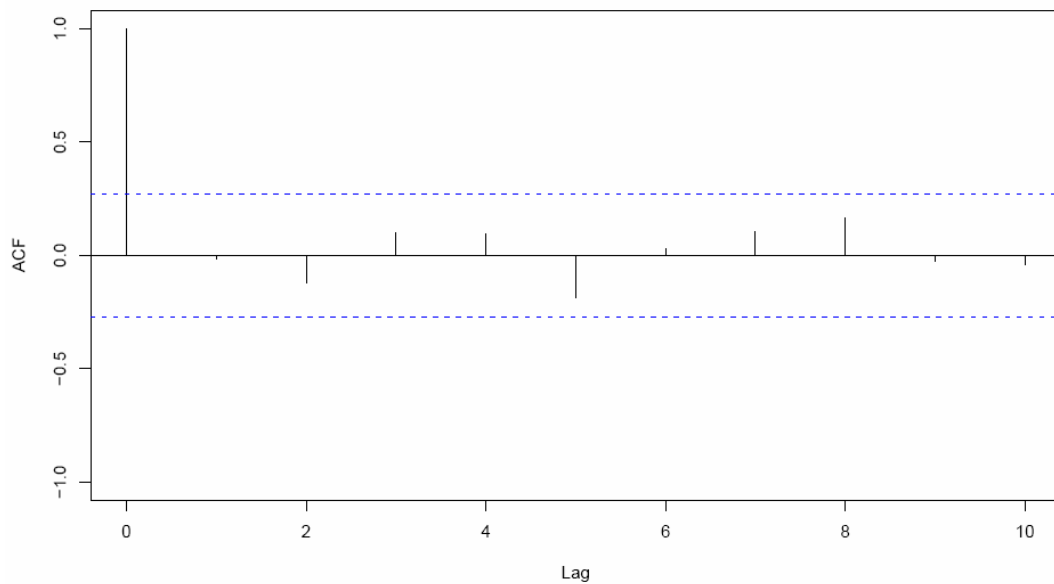


Questo primo risultato porta a ritenere che il processo sottostante alla serie dei residui standardizzati sia un processo *white noise*.

Un'ulteriore conferma si ha considerando il grafico della funzione di autocorrelazione parziale dei valori standardizzati di  $e_t$  ( $\cdot$ ), dove si può

notare che nessun valore di  $\hat{\rho}_k$  esce al di fuori della regione di accettazione  $\pm 1,96/\sqrt{52} \cong \pm 0,27$ .

**Figura 18 - ACF dei residui standardizzati del modello stimato per la serie  $Y_1$**



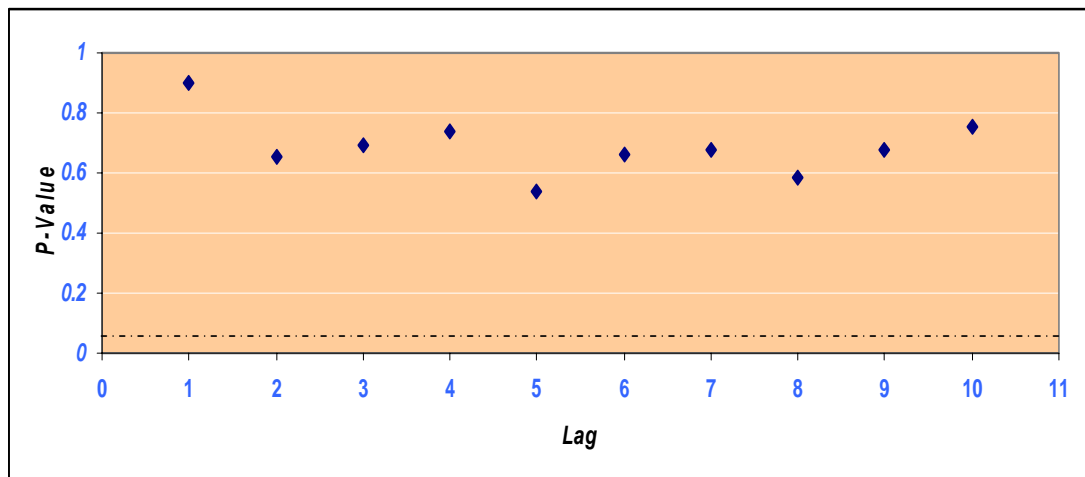
Ad un risultato coerente a quanto visto finora, si giunge considerando il test di *Ljung-Box*, utilizzato per una verifica complessiva di assenza di autocorrelazione dei residui.

I *p-values* del test di *Ljung-Box* ottenuti considerando i primi 10 *lags*, rappresentati in *Figura 19*, risultano tutti maggiori del livello di confidenza considerato (pari a 0,05); si può, quindi, accettare l'ipotesi di base  $H_0$  di assenza di autocorrelazione dei residui.

Poiché la serie dei residui soddisfa tutti i test diagnostici effettuati si

può affermare che i residui standardizzati sono incorrelati, quindi che il modello stimato ben si adatta ai dati.

**Figura 19 - *P-values* della statistica di Ljung-Box relativi ai residui standardizzati del modello stimato per la serie  $Y_1$**



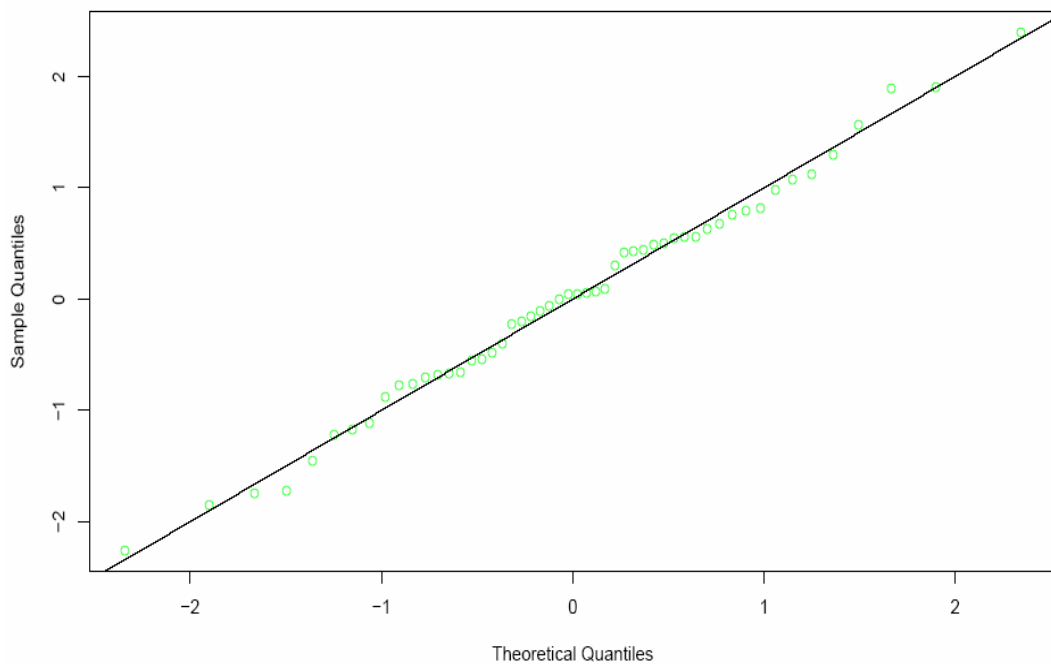
Infine, può essere utile verificare se il processo stocastico sottostante alla serie dei residui considerati sia un *white noise* gaussiano, cioè può essere utile stabilire se i valori standardizzati degli  $e_t$  si distribuiscono normalmente.

A tale scopo si rappresenta il *normal plot* dei residui standardizzati del modello identificato (*Figura 20*).

Poiché i punti si concentrano sulla bisettrice, si può dire che i residui si distribuiscono come una normale standardizzata.

Un'ulteriore conferma a quanto detto, si ottiene applicando il test di *Shapiro-Wilk* sulla normalità ai residui standardizzati.

**Figura 20: Normal plot dei residui standardizzati del modello stimato per la serie  $Y_1$**



Poiché il *p-value* relativo a tale test (pari a 0,9655) è maggiore di 0,05 (livello di confidenza utilizzato), si accetterà l'ipotesi di base di normalità dei residui.

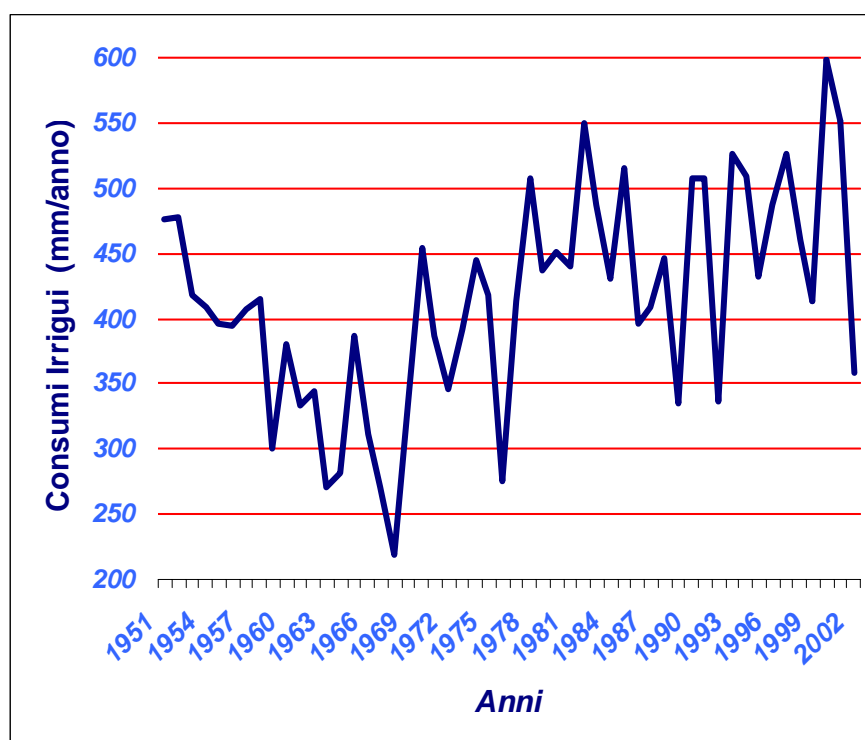
Si può concludere dicendo che, i residui  $e_t$ , non solo sono incorrelati, ma si distribuiscono anche come una normale standardizzata; ciò significa che il modello lineare stimato è in grado di spiegare l'intera

struttura di dipendenza seriale della serie dei consumi irrigui relativi alla coltura dell'olivo nella Agro di Foggia.

#### 4.2.2 Pomodoro

Si raffigura, ora, la serie storica  $\{y_t\}_{t=1}^{52}$  dei consumi irrigui riferiti alla coltura del pomodoro nell'Agro di Foggia, che denotiamo con  $Y_2$ ; essa riguarda l'intervallo di tempo 1951-2002.

**Figura 21 - Andamento dei consumi irrigui (mm/anno) relativi alla coltura del pomodoro dal 1951 al 2002 nell'Agro di Foggia**



Come si può notare dalla *Figura 21*, la serie considerata denota un andamento simile alla serie storica dell'olivo studiata precedentemente, anche se con periodi di crescita e decrescita più accentuati; infatti il suo tracciato è decrescente fino all'anno 1968 e cresce da quest'anno in poi, fino a stabilizzarsi intorno a valori piuttosto alti negli ultimi anni del periodo considerato.

Le motivazioni di questo andamento sono le stesse di quelle esposte nel caso della coltura dell'olivo e sono determinate, appunto, da un diverso andamento del regime evapotraspirativo nel periodo considerato.

Nella *Tabella 7*, in cui vengono sintetizzate le principali caratteristiche della serie  $Y_2$ , si può osservare che il minimo valore della variabile in esame, pari a 219, si ha in corrispondenza dell'anno 1968, mentre il massimo valore della stessa, pari a 599, è relativo all'anno 2000.

**Tabella 7 : Principali caratteristiche della serie storica riferita ai consumi irrigui del pomodoro (mm/anno)**

Numero oss.	Minimo	Media	Massimo	Varianza
52	219	414,9	599	6888,095

Considerando quanto detto nel paragrafo precedente, ossia che si irriga solo nel caso in cui la pioggia non è in grado di compensare le

perdite dovute all'evapotraspirazione stagionale, nell'anno 1968 la differenza tra il valore dell'evapotraspirazione del pomodoro e quello della pioggia (calcolato quest'ultimo cumulando i dati giornalieri di pioggia nella stagione irrigua della coltura in esame che va da aprile ad agosto) ha prodotto un valore piuttosto basso, pari a  $(376-232,6)=143,4$  mm. Tale differenza riscontrata nel 1968, indica che ad un valore evapotraspirativo del pomodoro molto basso (pari a 376 mm), è corrisposto un valore della pioggia rientrante nella media, pari a 232,6 mm; tale situazione, nel complesso, ha condotto ad una notevole riduzione dei consumi irrigui per la coltura del pomodoro nell'anno considerato. Viceversa, nell'anno 2000, la differenza in esame è risultata molto alta e pari a  $(661-117,6)=543,4$  mm, che risulta da un valore molto alto di evapotraspirazione del pomodoro, pari a 661 mm, e da una stagione irrigua poco piovosa (il cui valore è risultato di gran lunga inferiore alla media). Questa situazione ha inevitabilmente portato ad un rilevante aumento dei consumi irrigui per la coltura in esame nell'anno 2000.

Inoltre risulta evidente, dalla *Tabella 7*, che il valore medio della serie storica dei consumi irrigui relativi alla coltura del pomodoro, pari a 414,9 mm, è quasi il doppio di quello riferito ai consumi irrigui della coltura dell'olivo, pari a 233,3 mm (*Tabella 5*); questo si verifica perché

il pomodoro è una coltura che, generalmente, necessita di una maggiore quantità di acqua rispetto all'olivo, infatti il valore  $p$  del pomodoro (ossia il rapporto tra acqua prontamente utilizzabile dalla coltura e acqua disponibile totale) è minimo (0,5), mentre per l'olivo è prossimo al punto di appassimento. Per cui, la presenza di periodi di crescita e decrescita più accentuati, nell'andamento dei consumi irrigui del pomodoro (denotati precedentemente) rispetto a quello dei consumi irrigui dell'olivo, è legata proprio a questa maggiore necessità idrica della coltura del pomodoro.

Per quanto riguarda la varianza, invece, si può osservare che essa è minore per la serie storica dei consumi irrigui del pomodoro rispetto a quella dei consumi irrigui dell'olivo; ciò si determina perché questa specie è necessariamente in irriguo, quindi ha bisogno in modo costante di una maggiore quantità di acqua rispetto all'olivo, che può essere, invece, coltivato in arido.

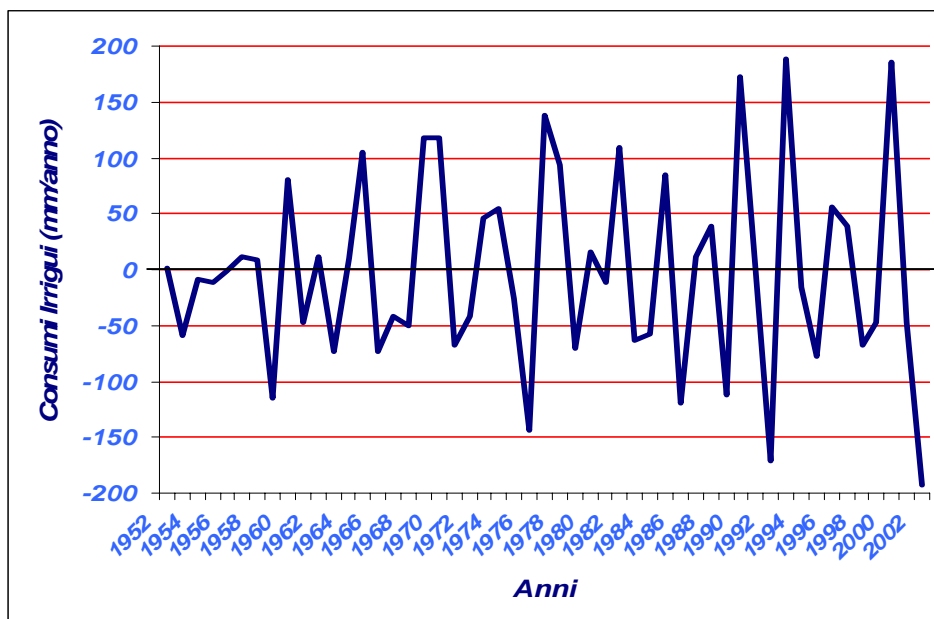
La descrizione dell'andamento della serie  $Y_2$ , fatta precedentemente, fa supporre che il processo generatore di quest'ultima sia non stazionario; poiché, come è noto, per poter applicare un modello di tipo ARMA, è necessario che il processo sia stazionario, è opportuno tramite differenziazione rendere stazionaria la serie considerata.

La serie differenziata, rappresentata in Figura 22, appare



compatibile con un processo generatore stazionario, poiché il suo tracciato segue un'andatura irregolare, fluttuando intorno al valore zero.

**Figura 22 - Serie storica differenziata dei consumi irrigui (mm/anno) relativi alla coltura del pomodoro dal 1952 al 2002 nell'Agro di Foggia**



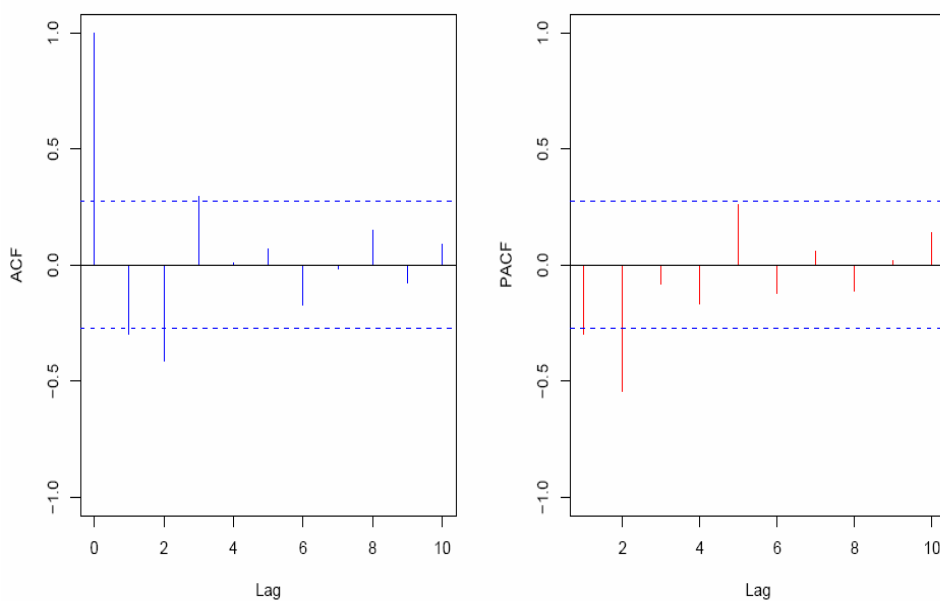
Essendo la serie differenziata stazionaria, è possibile considerare un modello della classe ARMA.

Le stime delle funzioni di autocorrelazione totale  $\hat{\rho}_k$  e parziale  $\hat{P}_k$ , riportate in Figura 23, tendono ad annullarsi velocemente all'aumentare di  $k$ , suggerendo un modello misto del tipo ARMA ( $p, q$ ).

Si noti che, anche in questo caso, poiché  $n = 52$ , gli estremi dell'intervallo di confidenza per l'ACF e il PACF di un processo *white*

*noise* vengono posti uguali a  $\pm 1,96/\sqrt{52} \cong \pm 0,27$ ; si considera, anche in questo caso, un numero di ritardi pari a 10.

**Figura 23 - Correlogrammi stimati per la serie differenziata dei consumi irrigui relativi alla coltura del pomodoro**



In questa circostanza per evitare il rischio di sovrapparametrizzazione (*overfitting*) è possibile considerare il criterio AIC (*Asymptotic Information Criterion*) che, come è noto, assegna un “costo” all’introduzione di ogni nuovo parametro aggiuntivo, consentendo, in questo modo, di ottenere un modello più parsimonioso in termini di numero di parametri.

I risultati del test AIC, applicato ai dati della serie storica differenziata riferita ai consumi irrigui della coltura del pomodoro, sono

riportati in *Tabella 8*, dove si può notare che il minimo valore del test si ha in corrispondenza dei parametri  $p = q = 2$ ; il modello identificato per la serie storica differenziata è, dunque, un ARMA(2,2), il che vuol dire che il processo sottostante alla serie storica iniziale  $Y_2$  è un processo autoregressivo integrato a media mobile di ordine (2,1,2), ovvero un ARIMA (2,1,2).

**Tabella 8 - AIC (  $p, q$  ) per la serie differenziata di  $Y_2$**

p	q		
	0	1	2
0	604,44	587,49	588,02
1	601,38	589,48	581,45
2	580,13	580,37	<b>579,6</b>

Specificato il modello per la serie storica  $Y_2$ , si passa alla fase di stima dei parametri del modello identificato.

Le stime dei parametri del modello e le relative deviazioni standard, ottenute mediante il metodo della massima verosimiglianza, sono riportate nella *Tabella 9*. Poiché le stime dei parametri  $\phi_1, \phi_2, \theta_1, \theta_2$ , sono

date rispettivamente da  $\hat{\phi}_1 = -0,8058$ ,  $\hat{\phi}_2 = -0,3577$ ,

$\hat{\theta}_1 = -0,3698$ ,  $\hat{\theta}_2 = 0,6302$ , il modello stimato è:

$$(1 + 0,8033 B + 0,3571 B^2)(1 - B)Y_t = (1 + 0,3638 B - 0,6362 B^2)$$

con  $\hat{\sigma}_\varepsilon^2 = 3694$ .

**Tabella 9 - Stima dei parametri del modello  
ARIMA (2,1,2)**

Coefficients			
$\hat{\phi}_1$	$\hat{\phi}_2$	$\hat{\theta}_1$	$\hat{\theta}_2$
-0,8058	-0,3577	-0,3698	0,6302
Deviazione Standard			
$\hat{\sigma}_{\hat{\phi}_1}$	$\hat{\sigma}_{\hat{\phi}_2}$	$\hat{\sigma}_{\hat{\theta}_1}$	$\hat{\sigma}_{\hat{\theta}_2}$
0,1643	0,1627	0,1480	0,1334

Inoltre, fissato  $\alpha$ , poiché risulta che:

$$|t_{\hat{\phi}_1}| = \left| \frac{\hat{\phi}_1}{\hat{\sigma}_{\hat{\phi}_1}} \right| = \left| \frac{-0,8058}{0,1643} \right| = 4,90 > 1,96,$$

$$|t_{\hat{\phi}_2}| = \left| \frac{\hat{\phi}_2}{\hat{\sigma}_{\hat{\phi}_2}} \right| = \left| \frac{-0,3571}{0,1633} \right| = 2,19 > 1,96,$$

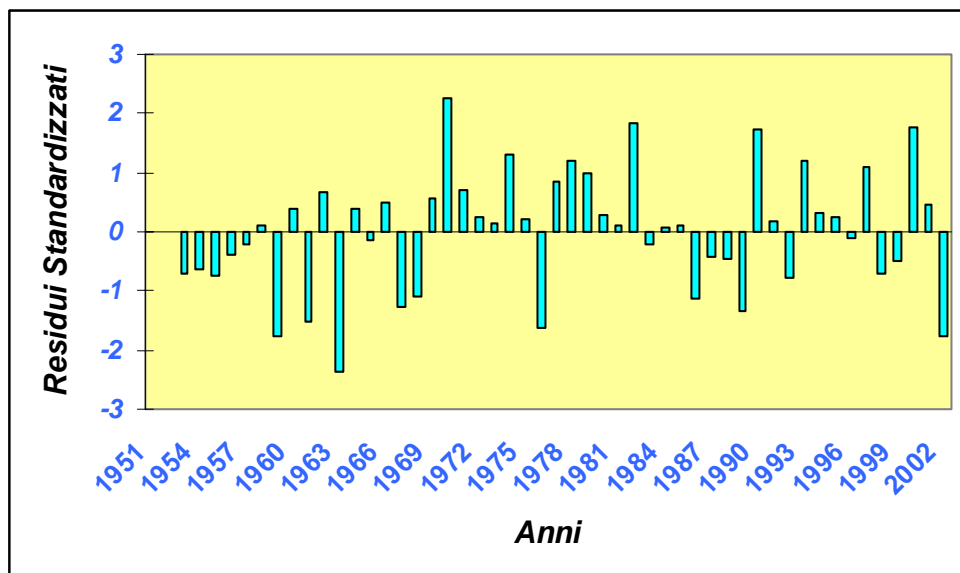
$$|t_{\hat{\theta}_1}| = \left| \frac{\hat{\theta}_1}{\hat{\sigma}_{\hat{\theta}_1}} \right| = \left| \frac{-0,3698}{0,1480} \right| = 2,50 > 1,96,$$

$$|t_{\hat{\theta}_2}| = \left| \frac{\hat{\theta}_2}{\hat{\sigma}_{\hat{\theta}_2}} \right| = \left| \frac{0,6302}{0,1334} \right| = 4,72 > 1,96.$$

si può affermare che tutti i parametri del modello sono significativi.

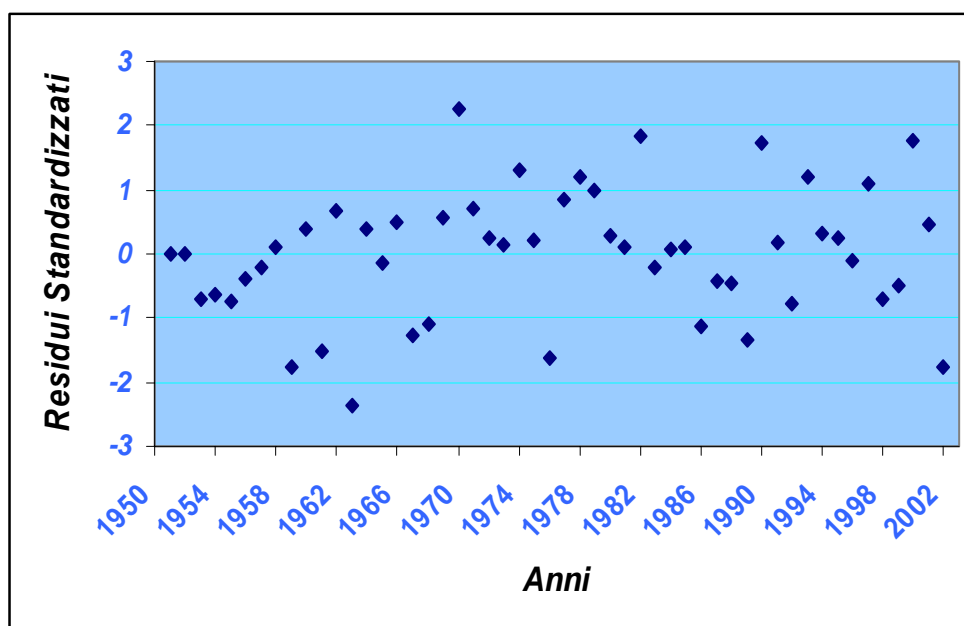
L'ultimo passo della procedura di *Box-Jenkins* consta di alcuni test diagnostici, volti a valutare la bontà di adattamento dei dati al modello stimato; tali misure diagnostiche si basano, come già detto nell'analisi precedente, sullo studio dei residui standardizzati del modello stimato, i quali devono risultare incorrelati; ciò vuol dire che il processo generatore della serie storica dei residui standardizzati deve essere un *white noise*.

**Figura 24 - Serie storica dei residui standardizzati del modello stimato per la serie  $Y_2$**



La rappresentazione grafica della serie dei residui standardizzati (*Figura 24*) e del suo diagramma di dispersione (*Figura 25*) porta a pensare che il processo sottostante alla serie dei residui standardizzati sia un *white noise*, dal momento che la serie presenta un tracciato del tutto irregolare, mentre il grafico di dispersione assume la forma di un insieme di punti privi di struttura.

**Figura 25 - Grafico di dispersione dei residui standardizzati del modello stimato per la serie  $Y_2$**

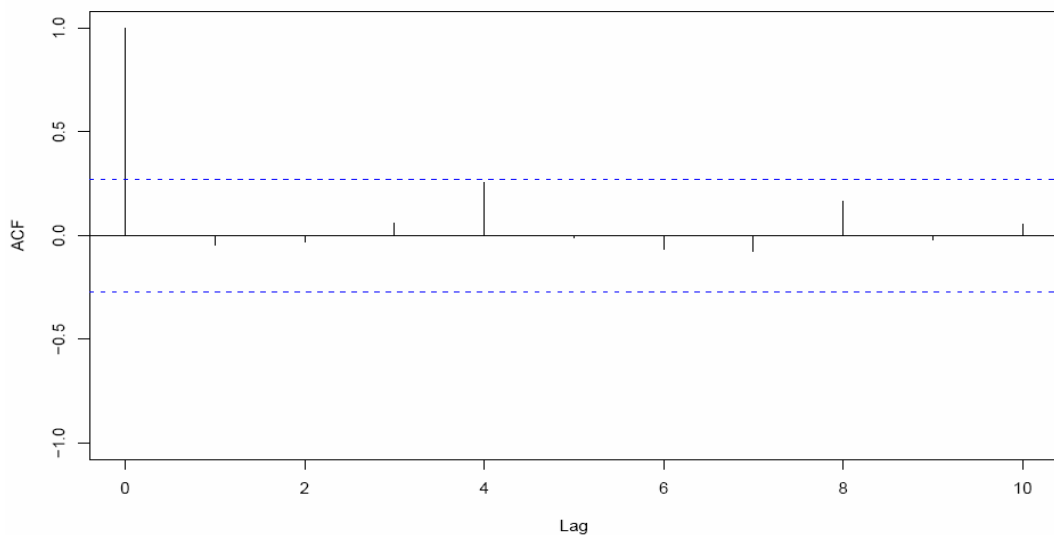


Anche dal grafico rappresentante la stima della funzione di autocorrelazione totale (*Figura 26*) si giunge alla conclusione che la serie dei residui standardizzati sia un *white noise*, dal momento che tutti i

valori di  $\hat{\rho}_k$  risultano compresi nell'intervallo di confidenza  $\pm 1,96/\sqrt{52} \cong \pm 0,27$ .

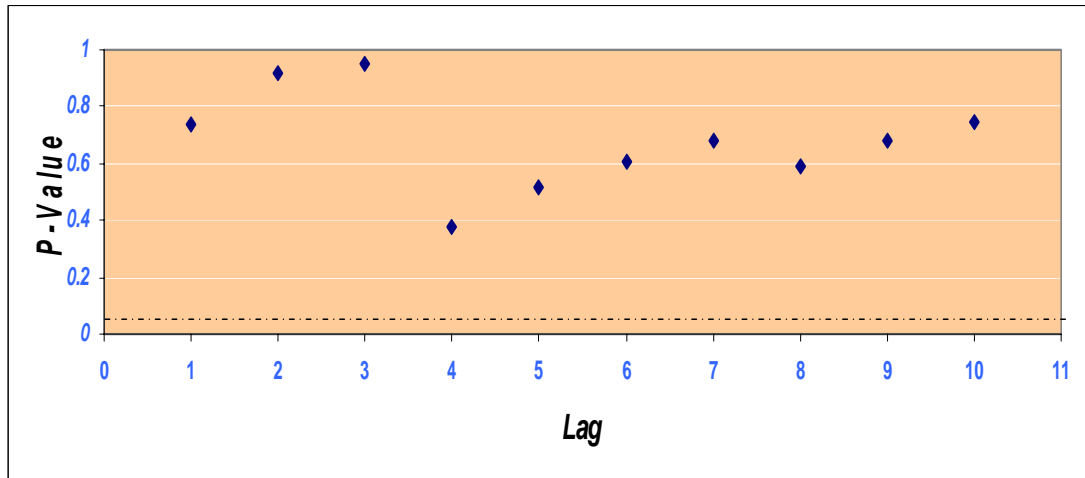
Un'ulteriore conferma a quanto detto, si ha applicando il test di *Ljung-Box* ai primi 10 lag della serie dei residui standardizzati del modello.

**Figura 26 - ACF dei residui standardizzati del modello stimato per la serie  $Y_2$**



Osservando la Figura 27, dove il risultato del test di *Ljung-Box* è rappresentato graficamente, si ha la conferma che i *p-values* calcolati per i primi 10 lags sono tutti superiori al livello di confidenza assunto (pari a 0,05); quindi si accetta l'ipotesi di base  $H_0$  che i residui del modello ARIMA sono realizzazioni di un processo *white noise*.

**Figura 27 - P-values della statistica di Ljung-Box relativi ai residui standardizzati del modello stimato per la serie  $Y_2$**



La significatività di tutti i parametri del modello, l'analisi grafica ed infine il test di *Ljung-Box* permettono di ritenere adeguato il modello stimato.

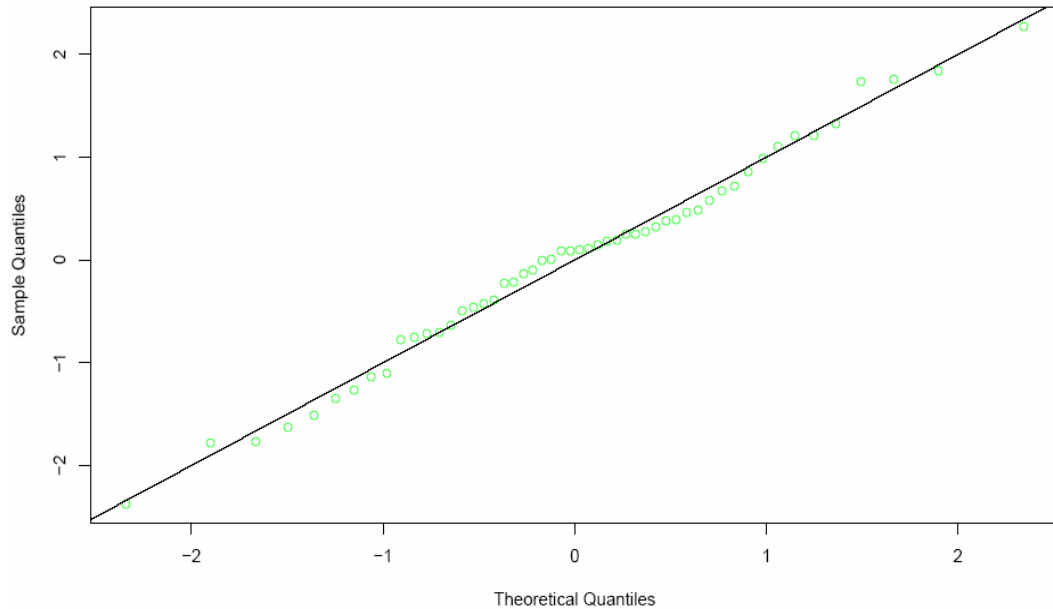
Infine, è anche possibile verificare che i residui standardizzati del modello stimato, oltre ad essere incorrelati, si distribuiscono come una normale standardizzata  $N(0,1)$ . Osservando la *Figura 28*, in cui viene riportato il *normal plot* di quest'ultimi, si può notare che i punti si concentrano sulla bisettrice del grafico, il che ci permette di avere una prima conferma sulla normalità dei residui standardizzati del modello.

Anche il test sulla normalità di *Shapiro-Wilk* ci porta a ritenere che i residui standardizzati si distribuiscono normalmente; infatti, il *p-value*



del test pari a 0,9008 supera di gran lunga il valore 0,05 (livello di confidenza adottato).

**Figura 28: Normal plot dei residui standardizzati del modello stimato per la serie  $Y_2$**



Si può, dunque, affermare che il processo generatore dei residui standardizzati del modello è un *white noise* gaussiano. Il modello stimato è, pertanto, in grado di spiegare completamente la struttura di dipendenza della serie storica dei consumi irrigui relativi alla coltura del pomodoro nell'Agro di Foggia.

## CONSIDERAZIONI CONCLUSIVE

Tra i diversi sistemi colturali, quelli irrigui stanno assumendo sempre più rilevanza negli scenari di sviluppo del Mezzogiorno, poiché più degli altri assicurano reddito e occupazione. In una regione scarsamente dotata di acqua, l'agricoltura irrigua deve saper fare un uso razionale di una risorsa naturale sempre più limitata. L'agricoltura irrigua, infatti, più degli altri comparti produttivi utilizza l'acqua, quindi deve concorrere al risparmio della risorsa idrica, migliorando l'efficienza di impiego delle acque tradizionali e riutilizzando le acque reflue.

Oltre che a produrre e ad usare le risorse, altrettanto importante è il ruolo che l'agricoltura svolge a tutela dell'ambiente, soprattutto in relazione ai fenomeni di erosione superficiale del suolo e di degrado del paesaggio. Una buona pratica agricola, infatti, può concorrere in maniera determinante alla tutela dell'assetto idrogeologico e alla riduzione dei

fenomeni di desertificazione che cominciano a manifestarsi in alcune fasce costiere del territorio meridionale dell'Italia.

Negli ambienti a clima mediterraneo, gli ultimi decenni sono stati caratterizzati da una forte espansione delle aree irrigue. La crescita della pratica irrigua è stata, però, frenata, oltre che dalla cronica penuria di acqua di buona qualità, anche dagli irregolari andamenti meteorologici che sono stati registrati nelle ultime annate, soprattutto per quanto riguarda le precipitazioni.

Per diverse annate successive, all'aumento dei valori di evapotraspirazione potenziale (anche durante la stagione invernale) è corrisposta una minore piovosità. Questa situazione ha provocato variazioni significative nelle falde, sia da un punto di vista quantitativo che qualitativo.

L'espansione delle superfici irrigabili e la maggiore durata della stagione irrigua hanno determinato un più intensivo prelievo dalle riserve idriche; di conseguenza l'abbassamento dei livelli delle falde ha aumentato i rischi di intrusione marina nelle zone costiere o, nei casi meno drammatici, ha provocato fenomeni stagionali di salinizzazione.

Il ripetersi di annate caratterizzate da scarsa pluviometria con l'aumento della salinità del terreno e la diminuzione delle risorse idriche rinnovabili innescano i processi di desertificazione.

I cambiamenti climatici sono percepiti in modo sempre crescente dalla popolazione e sono quantificabili attraverso alcuni indicatori.

L'analisi di tali indicatori, tuttavia, non ha mai fornito interpretazioni univoche, perciò ancora adesso si dibatte se i cambiamenti climatici siano una "percezione" o un fenomeno effettivamente quantificabile. Per quanto riguarda l'agricoltura, i consumi irrigui rappresentano uno di questi indicatori.

All'origine della controversia descritta potrebbe esserci un approccio all'analisi statistica delle serie storiche disponibili non scientificamente aggiornato.

Questa situazione piuttosto critica ha ispirato il contenuto del presente lavoro, nel quale è stato analizzato l'andamento temporale dei consumi irrigui dal 1951 al 2002 di due colture tipiche dell'agricoltura locale (olivo e pomodoro), in un'area agricola (Foggia) in fase di intensificazione colturale e particolarmente esposta ai rischi dovuti alla scarsità delle risorse idriche.

Tale studio, condotto mediante l'approccio moderno o statistico all'analisi delle serie storiche, ha permesso il conseguimento dei seguenti risultati:

➤ descrizione: l'analisi ha permesso di descrivere sinteticamente l'andamento dei volumi annui di irrigazione. Lo strumento utilizzato

allo scopo è stato il grafico della serie rispetto al tempo, mediante il quale è stato possibile trarre le prime, utili informazioni sulle caratteristiche delle serie studiate. L'analisi grafica della due serie storiche ha permesso di avere una prima conferma sulla crescita dei consumi irrigui "percepiti" negli ultimi anni dagli agricoltori. Nel periodo considerato (1951-2002), infatti, si è manifestata la presenza di un trend a partire dagli anni settanta, sia nel caso della serie storica relativa all'olivo che di quella relativa al pomodoro. Accanto all'ispezione grafica, altre indicazioni sono state fornite dalla media e dalla variabilità. I valori medi delle due serie storiche hanno evidenziato che i fabbisogni irrigui del pomodoro sono quasi il doppio di quelli dell'olivo. Viceversa, la variabilità delle serie storiche considerate è risultata maggiore per l'olivo che per il pomodoro, essendo quest'ultima una specie necessariamente irrigua (che richiede acqua in maniera costante). L'olivo invece è una specie che può essere coltivata anche in aridocoltura (quindi resistente anche ai periodi di carenza d'acqua).

- spiegazione: un altro risultato conseguito è rappresentato dalla spiegazione del fenomeno mediante l'identificazione del modello generatore della serie considerata. L'analisi, effettuata analizzando i grafici delle funzioni di autocorrelazione totale e parziale delle due

serie storiche differenziate, ha condotto alla specificazione di un modello Autoregressivo Integrato a Media Mobile di ordine (0,1,1) (ARIMA (0,1,1)) nel caso dei consumi irrigui relativi alla coltura dell'olivo, mentre per i consumi irrigui riguardanti la coltura del pomodoro, essendo ambigua l'identificazione del modello tramite le funzioni di autocorrelazione totale e parziale, si è fatto ricorso al criterio AIC (*Asymptotic Information Criterion*) che ha permesso di individuare un modello Autoregressivo Integrato a Media Mobile di ordine (2,1,2) (ARIMA (2,1,2)). Poiché entrambi i modelli identificati risultano integrati, si può affermare che le due serie storiche non sono stazionarie; questo deriva dalla presenza di un trend e quindi di una crescita nel periodo considerato. La bontà dei modelli specificati è stata verificata attraverso alcuni test diagnostici (analisi grafica della serie dei residui standardizzati dei modelli e delle loro funzioni di autocorrelazione e test di Ljung-Box calcolato per i primi 10 ritardi delle serie storiche dei residui standardizzati), basati sul controllo dell'incorrelazione dei residui del modello. Tali test hanno confermato che entrambi i modelli identificati sono corretti. Inoltre, poiché i residui, oltre ad essere incorrelati sono risultati anche indipendenti, i modelli specificati spiegano interamente la struttura di dipendenza delle serie storiche

considerate. La verifica sulla normalità dei residui è stata effettuata mediante analisi grafica (*normal plot*) e tramite il test di *Shapiro-Wilk*.

Si noti che l'individuazione di un modello sottostante alla serie storica osservata consente il conseguimento di un altro importante obiettivo:

- la previsione che permette di fare inferenza sui valori futuri del fenomeno di interesse in base alla sua storia passata (scenari climatici).

Tuttavia, occorre notare che, se utilizzati a fini previsionali, i modelli individuati, non forniscono buone stime, in quanto dovrebbero essere integrati con altre variabili di carattere ambientale (pioggia, tecniche colturali, metodi irrigui, etc.).

L'approccio *time series* potrebbe essere impiegato in tutti quei campi di analisi (compreso quello dei consumi irrigui) in cui è difficile modellare correttamente il funzionamento del fenomeno e si prescinde dalle cause che influenzano la variabile osservata.

Se effettuato con finalità di analisi e di ricerca dei fattori che influiscono sull'andamento della serie, questo approccio rappresenta una prima fase di studio del fenomeno. Infatti, esso permette di descrivere il

fenomeno considerato e consente di passare alla fase successiva di costruzione di un modello strutturale più complesso<sup>45</sup>.

---

<sup>45</sup> DI FONZO T. - LISI F., *Complementi di statistica economica...ecc.*, op. cit..



## **APPENDICE**

## TAVOLA A.1

### Spiegazione dei comandi inseriti in “R 2.1.0” per l’analisi temporale dei dati

I comandi utilizzati in "R 2.1.0" per l'analisi temporale dei consumi irrigui nella Agro di Foggia, in relazione alle principali colture praticate in quest'area, assumono il seguente significato:

**read.table()**: serve per leggere i dati da un file esterno ad "R" e rappresenta il modo più facile per costruire in "R" un nuovo *data frame*;

**summary()**: permette di ottenere il valore minimo e massimo, il primo e il terzo quantile, la mediana e la media aritmetica della variabile di un *data frame*;

**var()**: restituisce la varianza della variabile di un *data frame*;

**ts()**: consente al programma di identificare i dati caricati, come serie temporale;

**diff()**: serve per differenziare la serie di dati della variabile considerata;

**par(mfrow)**: specifica in modo permanente in quante parti deve essere divisa l'area di lavoro nella quale vengono rappresentati i grafici;

**arima()**: specifica il modello arima individuato per la serie storica di dati della variabile oggetto di studio;

**acf()**: permette di rappresentare graficamente la funzione di autocorrelazione relativa ai dati della variabile in esame;

**pacf()**: consente di rappresentare graficamente la funzione di autocorrelazione parziale relativa ai dati della variabile in esame;

**box.test()**: applica ai dati della variabile in esame il test di Ljung-Box o di Box-Pierce, a seconda dell'opzione "type" specificata e restituisce il valore della statistica utilizzata, il relativo valore (X-squared), i corrispondenti gradi di libertà (df) e il livello di significatività osservato;

**tsdiag()**: restituisce il grafico dei residui standardizzati, della funzione di autocorrelazione dei residui e i *p-value* della statistica di Ljung-Box;

**plot()**: è una funzione generica e produce grafici diversi a seconda del tipo di argomento a cui è riferita;

**qqnorm()**: permette di rappresentare il normal plot riferito ai dati della variabile in esame;

**shapiro.test()**: applica ai dati della variabile in esame il test di *Wilk-Shapiro* e restituisce come risultato il valore della statistica *W* e il relativo *p-value*.

## TAVOLA A.2

**Comandi inseriti in “R 2.1.0” e relativi risultati per l’analisi temporale dei consumi irrigui dell’olivo nell’Agro di Foggia**

```
> x<-read.table("irrigazione.txt",header=T)

> olivo<-ts(x$oli,start=1951, end=2002)

> olivo
Time Series:
Start = 1951
End = 2002
Frequency = 1

> summary(olivo)
  Min. 1st Qu.  Median    Mean 3rd Qu.   Max.
 49.0  172.8  227.0  233.3  277.8  448.0

> var(olivo)
[1] 8210.466

> diffolivo<-diff(olivo)

> diffolivo
Time Series:
Start = 1952
End = 2002
Frequency = 1

> old.p<-par(mfrow=c(1,2))

> acf(diffolivo, col="blue", xlab="Lag", ylab="ACF", ylim=c(-1,1))

> pacf(diffolivo, col="red", xlab="Lag", ylab="PACF", ylim=c(-1,1))

> par(old.p)

> m<-arima(olivo, c(0,1,1))

> m
Call:
arima(x = olivo, order = c(0, 1, 1))
```

Coefficients:

mal  
-0.7668  
s.e. 0.0922

sigma^2 estimated as 6821: log likelihood = -297.92, aic = 599.83

**> residuals(m)**

Time Series:

Start = 1951

End = 2002

Frequency = 1

[1] 0.2199998 29.3619337 -49.0565815 50.3328876 -41.1519079  
67.2533225 -50.7672423 8.3708029 -181.0744165  
[10] 50.4799146 -51.3078222 10.6918054 -91.7743071 -67.3322207 -  
13.6168258 -4.4395555 -53.4016855 8.0540489  
[19] 4.1754898 134.2004995 11.9016431 -39.8739576 72.4253155  
39.5343208 41.3141658 -59.3210227 97.5136794  
[28] 60.7717506 9.5987027 51.3601136 -7.6179701 162.1586724  
46.3404144 -11.4669858 161.2073166 40.6109333  
[37] -57.8602509 -95.3662222 -148.1251410 45.4202483 56.8274472 -  
87.4257134 92.9635013 85.2827770 -114.6066672  
[46] -138.8783758 70.5105051 -34.9337854 -27.7866119 202.6937149  
111.4219740 -137.5635909

**> stand<-function(x) {m=mean(x)**

**+ s=(var(x)^0.5)**

**+ z=(x-m)/s**

**+ return(z) }**

**> res.stand<-stand(residuals(m))**

**> res.stand**

Time Series:

Start = 1951

End = 2002

Frequency = 1

[1] -0.057575706 0.295932832 -0.655329180 0.550322649 -0.559440910  
0.755577325 -0.676080486 0.041298256 -2.256781987  
[10] 0.552106172 -0.682638033 0.069453361 -1.173519935 -0.877023269 -  
0.225424416 -0.114098811 -0.708037809 0.037455847  
[19] -0.009593322 1.567685359 0.084129392 -0.543938634 0.818316595  
0.419329778 0.440920330 -0.779842797 1.122652982  
[28] 0.676952104 0.056193392 0.562783496 -0.152654821 1.906834191  
0.501891637 -0.199345612 1.895293694 0.432389713  
[37] -0.762122789 -1.217091945 -1.857088190 0.490729489 0.629105418 -  
1.120768975 1.067456685 0.974285051 -1.450489688  
[46] -1.744919572 0.795088837 -0.484011483 -0.397312129 2.398547735  
1.291368652 -1.728970470

```
> acf(res.stand, 10, xlab="Lag", ylab="ACF", ylim=c(-1,1))
```

```
> Box.test(res.stand, lag=1, type="Ljung-Box")
```

```
Box-Ljung test
```

```
data: res.stand  
X-squared = 0.0153, df = 1, p-value = 0.9016
```

```
> Box.test(res.stand, lag=2, type="Ljung-Box")
```

```
Box-Ljung test
```

```
data: res.stand  
X-squared = 0.8546, df = 2, p-value = 0.6523
```

```
> Box.test(res.stand, lag=3, type="Ljung-Box")
```

```
Box-Ljung test
```

```
data: res.stand  
X-squared = 1.4499, df = 3, p-value = 0.6939
```

```
> Box.test(res.stand, lag=4, type="Ljung-Box")
```

```
Box-Ljung test
```

```
data: res.stand  
X-squared = 1.967, df = 4, p-value = 0.7418
```

```
> Box.test(res.stand, lag=5, type="Ljung-Box")
```

```
Box-Ljung test
```

```
data: res.stand  
X-squared = 4.068, df = 5, p-value = 0.5397
```

```
> Box.test(res.stand, lag=6, type="Ljung-Box")
```

```
Box-Ljung test
```

```
data: res.stand  
X-squared = 4.1201, df = 6, p-value = 0.6604
```

```
> Box.test(res.stand, lag=7, type="Ljung-Box")
```

```
Box-Ljung test
```

```
data: res.stand
X-squared = 4.8311, df = 7, p-value = 0.6806

> Box.test(res.stand, lag=8, type="Ljung-Box")

    Box-Ljung test

data: res.stand
X-squared = 6.5731, df = 8, p-value = 0.5833

> Box.test(res.stand, lag=9, type="Ljung-Box")

    Box-Ljung test

data: res.stand
X-squared = 6.6103, df = 9, p-value = 0.6776

> Box.test(res.stand, lag=10, type="Ljung-Box")

    Box-Ljung test

data: res.stand
X-squared = 6.718, df = 10, p-value = 0.7518

> qqnorm(res.stand, col="green")

> abline(0,1)

> shapiro.test(res.stand)

    Shapiro-Wilk normality test

data: res.stand
W = 0.9912, p-value = 0.9655
```

## TAVOLA A.3

**Comandi inseriti in “R 2.1.0” e relativi risultati per l’analisi temporale dei consumi irrigui del pomodoro nell’Agro di Foggia**

```
> z<-read.table("irrigazione.txt", header=T)

> z

> pomodoro<-ts(z$pom, start=1951, end=2002)

> pomodoro

Time Series:
Start = 1951
End = 2002
Frequency = 1

> summary(pomodoro)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 219.0   354.8   414.0   414.9   476.3   599.0

> var(pomodoro)
[1] 6888.095

> diffpomodoro<-diff(pomodoro)

> diffpomodoro

Time Series:
Start = 1952
End = 2002
Frequency = 1

> old.p<-par(mfrow=c(1,2))

> acf(diffpomodoro, col="blue", xlab="Lag", ylab="ACF",
ylim=c(-1,1))

> pacf(diffpomodoro, col="red", xlab="Lag", ylab="PACF",
ylim=c(-1,1))

> par(old.p)
```



```

> m<-arima(diffpomodoro, c(0,0,0))

> m

Call:
arima(x = diffpomodoro, order = c(0, 0, 0))

Coefficients:
      intercept
      -2.3137
s.e.      12.2051

sigma^2 estimated as 7597:  log likelihood = -300.22,
aic = 604.44

> m<-arima(diffpomodoro, c(1,0,0))

> m

Call:
arima(x = diffpomodoro, order = c(1, 0, 0))

Coefficients:
      ar1  intercept
      -0.3196  -1.4153
s.e.   0.1382   8.8386

sigma^2 estimated as 6865:  log likelihood = -297.69,
aic = 601.38

> m<-arima(diffpomodoro, c(0,0,1))

> m

Call:
arima(x = diffpomodoro, order = c(0, 0, 1))

Coefficients:
      ma1  intercept
      -0.7243  0.7701
s.e.   0.0961  2.9735

sigma^2 estimated as 5163:  log likelihood = -290.74,
aic = 587.49

> m<-arima(diffpomodoro, c(1,0,1))

> m

```

```

Call:
arima(x = diffpomodoro, order = c(1, 0, 1))

Coefficients:
          ar1          ma1  intercept
      0.0198  -0.7324      0.7881
s.e.  0.1927   0.1245      2.9584

sigma^2 estimated as 5162:  log likelihood = -290.74,
aic = 589.48

> m<-arima(diffpomodoro, c(2,0,0))

> m

Call:
arima(x = diffpomodoro, order = c(2, 0, 0))

Coefficients:
          ar1          ar2  intercept
     -0.5488  -0.6352      0.3488
s.e.   0.1171   0.1132      4.2613

sigma^2 estimated as 4263:  log likelihood = -286.07,
aic = 580.13

> m<-arima(diffpomodoro, c(0,0,2))

> m

Call:
arima(x = diffpomodoro, order = c(0, 0, 2))

Coefficients:
          ma1          ma2  intercept
     -0.3514  -0.6481      2.2884
s.e.   0.2256   0.2183      1.0266

sigma^2 estimated as 4753:  log likelihood = -290.01,
aic = 588.02

> m<-arima(diffpomodoro, c(1,0,2))

> m

Call:
arima(x = diffpomodoro, order = c(1, 0, 2))

Coefficients:

```

	ar1	ma1	ma2	intercept
	-0.5438	0.2072	-0.7928	1.1429
s.e.	0.1352	0.1188	0.1125	2.6585

sigma^2 estimated as 4031: log likelihood = -285.73,  
aic = 581.45

**> m<-arima(diffpomodoro, c(2,0,1))**

**> m**

Call:

arima(x = diffpomodoro, order = c(2, 0, 1))

Coefficients:

	ar1	ar2	ma1	intercept
	-0.3614	-0.5761	-0.3051	0.6026
s.e.	0.1776	0.1341	0.1947	3.3085

sigma^2 estimated as 4110: log likelihood = -285.18,  
aic = 580.37

**> m<-arima(diffpomodoro, c(2,0,2))**

**> m**

Call:

arima(x = diffpomodoro, order = c(2, 0, 2))

Coefficients:

	ar1	ar2	ma1	ma2	intercept
	-0.8033	-0.3571	0.3638	-0.6362	0.8896
s.e.	0.1652	0.1633	0.1496	0.1356	2.9980

sigma^2 estimated as 3686: log likelihood = -283.8, aic  
= 579.6

**> m<-arima(pomodoro, c(2,1,2))**

**> m**

Call:

arima(x = pomodoro, order = c(2, 1, 2))

Coefficients:

	ar1	ar2	ma1	ma2
	-0.8058	-0.3577	0.3698	-0.6302
s.e.	0.1643	0.1627	0.1480	0.1334

```
sigma^2 estimated as 3694: log likelihood = -283.84,  
aic = 577.68
```

```
> residuals(m)
```

```
Time Series:
```

```
Start = 1951
```

```
End = 2002
```

```
Frequency = 1
```

```
[1] 0.4759995 0.6807975 -43.0116567 -38.2919401  
-45.2988593 -23.5110911 -12.4313566 6.1370038 -  
107.3425952  
[10] 24.3589015 -91.0816478 41.5844755 -143.5612516  
23.9844990 -7.5414597 30.2598071 -76.1805703 -  
66.3479063  
[19] 35.9212773 138.9242326 44.4300485 15.7653581  
9.7257691 80.8388950 12.5012994 -98.0749423  
52.6991680  
[28] 73.7914009 60.5032294 17.1252835 7.0002450  
112.5249492 -12.9212748 5.8032752 6.4141205 -  
68.5765802  
[37] -25.3235217 -27.3583186 -81.4585170 105.8195869  
11.6855458 -46.3397305 74.2436307 19.9654911  
15.7786864  
[46] -5.6207107 67.3268707 -42.3542693 -29.7931406  
107.2062923 28.2788963 -107.2287116
```

```
> stand<-function(x){m=mean(x)
```

```
+ s=(var(x)^0.5)
```

```
+ z=(x-m)/s
```

```
+ return(z) }
```

```
> res.stand<-stand(residuals(m))
```

```
res.stand
```

```
Time Series:
```

```
Start = 1951
```

```
End = 2002
```

```
Frequency = 1
```

```
[1] -0.002326822 0.001043156 -0.717921993 -0.640258428  
-0.755558216 -0.397037591 -0.214719084 0.090825758 -  
1.776496148  
[10] 0.390669509 -1.508920094 0.674118622 -2.372478959  
0.384508665 -0.134255181 0.487769696 -1.263720881 -  
1.101923086  
[19] 0.580929942 2.275857279 0.720942904 0.249261609  
0.149879368 1.320055320 0.195551088 -1.623995686  
0.857012366
```

```

[28] 1.204087866 0.985429234 0.271639362 0.105030505
1.841453557 -0.222780754 0.085334209 0.095385749 -
1.138596212
[37] -0.426861378 -0.460344230 -1.350570196 1.731115930
0.182127754 -0.772685890 1.211529368 0.318375356
0.249480928
[46] -0.102649001 1.097713160 -0.707104596 -0.500409551
1.753934351 0.455173546 -1.774622178
> acf(res.stand, 10, xlab="Lag", ylab="ACF", ylim=c(-
1,1))

```

```
> Box.test(res.stand, lag=1, type="Ljung-Box")
```

Box-Ljung test

```

data: res.stand
X-squared = 0.1139, df = 1, p-value = 0.7358

```

```
> Box.test(res.stand, lag=2, type="Ljung-Box")
```

Box-Ljung test

```

data: res.stand
X-squared = 0.1643, df = 2, p-value = 0.9211

```

```
> Box.test(res.stand, lag=3, type="Ljung-Box")
```

Box-Ljung test

```

data: res.stand
X-squared = 0.3567, df = 3, p-value = 0.949

```

```
> Box.test(res.stand, lag=4, type="Ljung-Box")
```

Box-Ljung test

```

data: res.stand
X-squared = 4.2377, df = 4, p-value = 0.3748

```

```
> Box.test(res.stand, lag=5, type="Ljung-Box")
```

Box-Ljung test

```

data: res.stand
X-squared = 4.2432, df = 5, p-value = 0.515

```

```
> Box.test(res.stand, lag=6, type="Ljung-Box")
```

Box-Ljung test

```

data: res.stand
X-squared = 4.491, df = 6, p-value = 0.6105

> Box.test(res.stand, lag=7, type="Ljung-Box")

      Box-Ljung test

data: res.stand
X-squared = 4.8309, df = 7, p-value = 0.6806

> Box.test(res.stand, lag=8, type="Ljung-Box")

      Box-Ljung test

data: res.stand
X-squared = 6.5234, df = 8, p-value = 0.5888

> Box.test(res.stand, lag=9, type="Ljung-Box")

      Box-Ljung test

data: res.stand
X-squared = 6.5492, df = 9, p-value = 0.684

> Box.test(res.stand, lag=10, type="Ljung-Box")

      Box-Ljung test

data: res.stand
X-squared = 6.7522, df = 10, p-value = 0.7486

> qqnorm(res.stand, col="green")

> abline(0,1)

> shapiro.test(res.stand)

      Shapiro-Wilk normality test

data: res.stand
W = 0.9887, p-value = 0.9008

```

## BIBLIOGRAFIA

- 1) AA.VV., *Atlante Geografico De Agostini*, Istituto Geografico De Agostini S.p.A., Novara, 2003.
- 2) AA.VV., *Irrigazioni e acquedotti rurali in Puglia, Storia e ruolo dei Consorzi di Bonifica nello sviluppo della regione*, Unione Regionale delle Bonifiche e delle Irrigazioni della Puglia, Promedit, Casamassima (Bari), 1995.
- 3) AA.VV. (a cura di SCANDELLA P. – MECCELLA G.), *Irrigazione sostenibile, La buona pratica irrigua*, Ministero delle Politiche Agrarie e Forestali, Progetto Editoriale PANDA, Volume No. 5, L'Informatore Agrario, Verona, 2004.
- 4) AA.VV., *Agricultural meteorological practices*, Technical note n°134, Secretariat of the World Meteorological Organization, Geneva 1982.
- 5) AKAIKE H., *A new look at the statistical model identification*, IEEE Transactions on Automatic Control, 1974.

- 6) ALLEN R.G – PEREIRA L.S – RAES D. – SMITH M., *Crop evapotranspiration: guidelines for computing crop water requirements*, Irrigation and Drainage paper No. 56, F.A.O., Rome, 1998.
- 7) ALVARO G., *Contabilità nazionale e statistica economica*, Cacucci, Bari, 1999.
- 8) AZZALINI A., *Inferenza statistica, Una presentazione basata sul concetto di verosimiglianza*, Sprinter, Milano, 2001.
- 9) BARTLETT M.S., “On the theoretical specification of sampling properties of autocorrelated time series”, *Journal of the Royal Statistical Society*, B, 8, 1946.
- 10) BORTOT P. – VENTURA L. – SALVAN A., *Inferenza Statistica, Applicazioni con S-PLUS e R*, Cedam, Padova, 2002.
- 11) BOX G.E.P. – JENKINS G.M., *Time Series Analysis, Forecasting and Control*, Holden-Day, San Francisco, 1976.
- 12) DI FONZO T. - LISI F., *Complementi di statistica economica, Analisi delle serie storiche univariate*, Cleup, Padova, 2001.
- 13) DOBB J.L., *Stochastic processes*, Wiley, New York, (1953).
- 14) GIUSTI F – VITALI O., *Statistica Economica*, Cacucci, Bari, 1990.
- 15) HAMILTON J.D., *Econometria delle serie storiche*, a cura di Sitzia B., Monduzzi, Bologna, 1995.



- 16) IACUS S. – MASAROTTO G., *Laboratorio di statistica con R*, McGraw-Hill, Milano, 2003.
- 17) IHAKA R. – GENTLEMAN R., *R: A language for data analysis and graphics*, Journal of Computational and Graphical Statistics, 1996.
- 18) JOHNSTON J., *Econometrica*, Franco Angeli, Milano, 1993.
- 19) LJUNG G.M. – BOX G.E.P., “On a measure of lack of fit in time series models”, *Biometrika*, 65, N.2, 1978.
- 20) MASTRORILLI M., *L'acqua e l'ambiente: il contributo della ricerca agronomica*, 2° Workshop progetto CLIMAGRI, Cambiamenti climatici e agricoltura, Istituto Sperimentale Agronomico, Bari, 2003.
- 21) MONTEITH J.L., *Evaporation and environment*, In: Fogg G.E.(Ed), The State and Movement of Water in Living Organism. Soc. Exp. Biol. Symp. 19, 1965.
- 22) PARPINEL F. – PROVASI C., *Probabilità e statistica per le scienze economiche*, Giappichelli, Torino, 1999.
- 23) PENMAN H.L., *Natural evaporation from open water, bare soil and grass*, Proc. Royal Soc. A. 193, 1948.
- 24) PICCOLO D., *Introduzione all'analisi delle serie storiche*, La Nuova Italia Scientifica, 1990.

- 25) POLLICE A., *Esercitazioni con R*, Materiale didattico, Dipartimento di Scienze Statistiche, Università degli Studi di Bari, 2004.
- 26) RANA G. – GIGOTTI R. – MASTRORILLI M. – JOVANOVIĆ N. – INCARNATO D., *Evapotranspiration measurement under drought conditions*, International conference on “Supplementary irrigation and drought water management”, Valenzano (Bari), 27 Settembre-2 Ottobre, 1992, Vol.3.
- 27) SNEYRES R., *On the statistical analysis of series observation*, Technical note n°143, Secretariat of the World Meteorological Organization, Geneva, 1990.
- 28) VIANELLI S., *L'analisi delle serie temporali nello sviluppo storico e metodologico della statistica*, Atto del Convegno nazionale di studio sull' “Analisi Moderna delle serie storiche”, Napoli, 19-22 maggio 1981.
- 29) WEI W.W.S., *Time series analysis. Univariate and multivariate methods*, Addison-Wesley, Redwood City, 1990.

### **Siti internet:**

<http://www.cran.r-project.org/doc/contrib/Ricci-ts-italian.pdf>

<http://www.dip.statistica.uniba.it/html/docenti/pollice/materiale.htm>

<http://www.inea.it/pdf/filespdf.cfm>

[http://www.istat.it/dati/dataset/20050421\\_00](http://www.istat.it/dati/dataset/20050421_00)

<http://www.sirio.stat.unipd.it/files/ts04-05>

<http://www.statistica.unipd.it/servizi/matdid.asp?idins=7#appunti>