

UNIVERSITA' DEGLI STUDI DI BARI

FACOLTA' DI ECONOMIA

CORSO DI LAUREA IN SCIENZE STATISTICHE ED ECONOMICHE

TESI DI LAUREA

IN

STATISTICA ECONOMICA

ANALISI DI DATI TERRITORIALI E AMBIENTALI:

ANALISI STATISTICA SULLA QUALITA' DELLE ACQUE
DI FALDA IN PUGLIA

Relatori:

Chiar.mo Prof. Antonio Mastrodonato

Dott. Marcello Mastrorilli

Laureando:

DOMENICO VITALE

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 Scienze Statistiche ed Economiche della Facoltà di Economia dell'Università degli Studi di Bari).

Si ringrazia, inoltre, per il valido supporto tecnico fornito, il Dott. Marcello Mastrorilli (Direttore della Sezione Irrigazione dell'Istituto Sperimentale Agronomico del Consiglio per la Ricerca e sperimentazione in Agricoltura di Bari).

INDICE

<i>INTRODUZIONE</i>		4
<i>CAPITOLO 1</i>	<i>CONTESTO TERRITORIALE.....</i>	10
1.1	Aspetti geomorfologici generali.....	10
1.2	Il clima	13
1.3	Aspetti socio-economici.....	21
<i>CAPITOLO 2</i>	<i>PIANO DELL'INDAGINE.....</i>	36
2.1	Obiettivo dell'indagine	36
2.2	Le fonti dei dati	39
2.3	Il software utilizzato.....	45
<i>CAPITOLO 3</i>	<i>ANALISI STATISTICA SULLA QUALITÀ DELLE ACQUE DI FALDA.....</i>	48
3.1	Premessa.....	48
3.2	Idrostruttura delle Murge	52
3.3	Idrostruttura della Capitanata.....	62
3.4	Idrostruttura del Salento.....	72
3.5	Idrostruttura Arco Jonico	82
3.6	Considerazioni sulla variabilità dei parametri chimico-fisici nelle idrostrutture pugliesi.	84
<i>CAPITOLO 4</i>	<i>L'ANALISI CLASSICA DELLE SERIE STORICHE: ASPETTI TEORICI.....</i>	88
4.1	Premessa.....	88
4.2	Il coefficiente di correlazione lineare di Bravais-Pearson	92
4.3	Funzione di correlazione incrociata	93
4.4	L'analisi delle serie storiche	95
4.5	Approccio classico o moderno	97

4.6	Richiami all'analisi di regressione	102
4.7	La determinazione della componente tendenziale	110
4.8	La determinazione della componente stagionale	115
4.9	La determinazione della componente ciclica	119
4.10	Il metodo delle medie mobili	120
4.11	Regressione non parametrica	129
4.12	L'analisi dei residui.....	138
<i>CAPITOLO 5</i>	<i>ANALISI SULLA DIPENDENZA TRA PIOGGIA E CONDUCIBILITA' ELETTRICA</i>	<i>143</i>
5.1	Una misura della dipendenza	143
5.2	Analisi temporale della pioggia	148
5.3	Analisi temporale della ECw	151
5.4	Scelta del modello di aggregazione	153
5.5	Scelta di uno stimatore del trend.....	154
5.6	Scelta di uno stimatore per la componente stagionale	157
5.7	Stima del modello	161
5.8	I risultati dell'analisi sulla serie della ECw	164
5.9	Le relazioni di dipendenza tra la pioggia e la ECw	165
<i>CAPITOLO 6</i>	<i>CONSIDERAZIONI CONCLUSIVE</i>	<i>172</i>
<i>APPENDICE</i>		<i>178</i>
<i>BIBLIOGRAFIA</i>		<i>185</i>

INTRODUZIONE

La disponibilità di risorse idriche, in particolare per gli usi potabili e per quelli agricoli, rappresenta un problema di grande importanza e gravità per il futuro della popolazione mondiale. Dinamiche legate ai cambiamenti climatici, unitamente a quelle concernenti la situazione demografica e socio-economica, il progressivo inquinamento e il maggior bisogno di risorse idriche, in futuro, aggraveranno situazioni già oggi molto critiche.

Alla scarsa disponibilità delle risorse idriche si accompagna molto spesso un degrado qualitativo delle stesse, causato sia da fenomeni d'inquinamento antropico generalizzato nel territorio (es. scarichi nei corsi d'acqua di reflui civili ed industriali o spandimento sulla superficie del suolo) che da intrusione salina lungo le coste, spesso causata da eccessivo emungimento.

La consistenza delle risorse idriche in Italia non è drammatica.

Dai dati riportati nella Tabella 1, le previsioni circa le disponibilità idriche in Italia sono in controtendenza rispetto a quelle di molti altri Paesi del Bacino Mediterraneo. Infatti, mentre nei diversi Paesi considerati si prevede una riduzione della risorsa idrica per abitante, le previsioni al 2025 per l'Italia ne indicano un aumento pari a circa 250 m³/anno rispetto al 1990. Più che all'aumento delle risorse naturali, la maggiore disponibilità sarà da attribuire al calo demografico.

Tabella 1 - Le risorse idriche nel bacino del Mediterraneo

	Risorse naturali rinnovabili [km ³ anno ⁻¹]			Disponibilità per abitante [m ³ anno ⁻¹]	
	teoriche totali	interne al Paese	esterne al Paese	1990	2025
Italia	187	179	8	3277	3531
Francia	185	170	15	3295	3064
Spagna	113	112	1	2885	2672
Malta	0.07	0.06	0	198	180
Albania	50	44	6	15408	9973
Israele	2	1	0.5	370	246
Gaza	0.06	0.03	0.035	84	0
Turchia	236	227	9	4222	2690
Algeria	14	14	0.2	565	271
Tunisia	4	4	0.6	511	313

Fonte: MASTRORILLI M., L'acqua e l'ambiente: il contributo della ricerca agronomica, I.S.A., Bari, 2°Workshop progetto CLIMAGRI, Cambiamenti Climatici e Agricoltura, 2003.

Si può essere ottimisti guardando al futuro, tuttavia la distribuzione delle risorse idriche all'interno del nostro Paese risulta disomogenea.

Infatti, come si rileva dalla Tabella 2, considerando la disponibilità idrica per grandi compartimenti idrografici (Nord, Centro, Sud-Isole), risulta che sono le regioni settentrionali a beneficiare del 65 per cento dell'intero ammontare della risorsa nazionale; il restante 35 per cento viene ripartito tra le regioni centrali (15 per cento) e meridionali-insulari (20 per cento), in quantità senza alcun dubbio esigue se confrontate alle dimensioni e alle esigenze del territorio, caratterizzato da condizioni climatiche che, in genere, determinano una richiesta evaporativa dell'ambiente maggiore rispetto alle regioni settentrionali.

Tabella 2 - Disponibilità idrica in Italia per compartimenti geografici

Compartimenti idrografici	Disponibilità (milioni di m³)	Utilizzazioni %
Nord	33.925	65
Centro	7.825	15
Sud-Isole	10.058	20
Italia	51.808	100

Fonte: MASTRORILLI M., L'acqua e l'ambiente:.....ecc..op.cit.

Concorrono a questa disuguaglianza distributiva vari fattori, come la natura del territorio, la difforme distribuzione delle precipitazioni tra contesti geografici, la conseguente irregolarità dei deflussi superficiali,

lo stato infrastrutturale delle reti di distribuzione che non consentono di utilizzare efficacemente le risorse potenzialmente disponibili.

Le difficoltà legate all'approvvigionamento idrico riguardano, pertanto, il Sud dell'Italia, dove, negli ultimi anni, si sono registrati periodi siccitosi più prolungati e intensi rispetto al resto del Paese. A ciò, si aggiunge la tendenza all'aumento della domanda di acqua da parte di tutti i settori produttivi. Gli insediamenti urbani richiedono maggiori disponibilità, le installazioni industriali e artigianali utilizzano quantitativi sempre maggiori e cresce la domanda di acqua per il settore turistico.

Anche il settore agricolo registra un aumento della domanda di acqua, sia perché sono state riconvertite in irriguo colture tradizionalmente "in asciutto" sia perché le modifiche delle condizioni meteorologiche, registrate negli ultimi anni, hanno allungato la stagione irrigua e aumentato i consumi idrici. Infine, c'è da considerare che i sistemi colturali irrigui sono ritenuti i più remunerativi e i prodotti ottenuti in irriguo sono quantitativamente e qualitativamente stabili (si riduce la variabilità interannuale).

Per far fronte a tali esigenze, molte aziende agricole delle regioni meridionali fanno ricorso alle risorse sotterranee (complessivamente in

Puglia, Sicilia e Campania si concentra l'80% delle aziende irrigue meridionali che attuano questa forma di approvvigionamento).

Lungo le fasce costiere, l'eccessivo emungimento si accompagna, spesso, all'abbassamento del livello delle falde e al progressivo peggioramento qualitativo delle risorse idriche sotterranee, provocato da una fase di ricarica della falda meno efficiente. L'intrusione dell'acqua di mare impedisce definitivamente lo sfruttamento delle falde ai fini irrigui.

In questo lavoro s'intendono applicare alcuni strumenti statistici per interpretare la situazione in cui riversano alcune falde pugliesi, dalle quali si attinge l'acqua per l'irrigazione.

Nel primo capitolo viene analizzato il territorio pugliese nei suoi aspetti geomorfologici e climatici. Saranno inoltre considerati alcuni aspetti socio-economici della regione Puglia, del Mezzogiorno e dell'Italia.

Chiarito l'obiettivo dell'indagine, nel secondo capitolo, si presentano il database a disposizione e il software utilizzato per l'elaborazione dei dati.

Nel terzo capitolo, dopo un'attenta analisi esplorativa che evidenzia la struttura dei dati, si procede ad una classificazione qualitativa (a fini irrigui) delle acque di falda monitorate.

Infine, dopo aver esposto, nel quarto capitolo, la metodologia statistica di cui si farà uso, si conduce uno studio sulle relazioni di dipendenza tra la pioggia, evento meteorico che alimenta le falde, e la Conducibilità Elettrica, uno dei parametri chimico-fisici utilizzato come indicatore di salinità delle acque sotterranee.

CAPITOLO 1

CONTESTO TERRITORIALE

1.1 Aspetti geomorfologici generali

La Puglia si estende per una superficie complessiva di 1.935.000 ettari. Dal punto di vista morfologico, gran parte del territorio regionale è pianeggiante; larga estensione presentano le forme collinari, mentre marginale risulta la presenza di rilievi montuosi (Tabella 3).

Nel complesso, sotto il profilo orografico, la Puglia è distinta in cinque zone: Gargano, sub-Appennino Dauno, Tavoliere, Murgia e Salento. Le aree montuose sono rappresentate dal massiccio del Gargano e dal sub-Appennino Dauno, ambedue ubicate a nord, in provincia di Foggia.

Tabella 3: Classificazione della superficie territoriale per zone altimetriche

Zone altimetriche	Superficie	
	ha (migliaia)	%
Pianura	1.030	53,2
Collina	876	45,3
Montagna	29	1,5
Totale Puglia	1.935	100,0

Fonte: ISTAT

Il Gargano, promontorio indipendente e peninsulare, situato a nord-ovest della regione, è bagnato sui tre versanti nord, est e sud dal mare Adriatico.

Il sub-Appennino Dauno si eleva nella parte occidentale della regione tra i corsi dei fiumi Fortore ed Ofanto, collegandosi alla dorsale dell'Appennino Campano.

Tra le predette aree montuose si sviluppa la pianura del Tavoliere, un bassopiano a pendio lievissimo su cui si sono accumulati, in modo particolare lungo i bacini fluviali, i materiali di alluvione trascinati dai corsi d'acqua torrentizi, provenienti dall'Appennino. Il Tavoliere si affaccia sul mare Adriatico in direzione sud-est, delimitando il golfo di Manfredonia.

Il rilievo delle Murge caratterizza l'area centrale del territorio regionale ed interessa gran parte della provincia di Bari e parte del territorio provinciale tarantino e brindisino. Costituisce un altopiano di natura calcarea, variamente inciso da lame e gravine. La roccia calcarea, spesso affiorante, limita fortemente il franco di coltivazione (profondità del terreno); tuttavia, il processo di antropizzazione ha esteso la superficie coltivabile tramite rottura meccanica del friabile substrato roccioso (spietramento). Il massiccio delle Murge degrada a pendio, più o meno dolce, verso le coste adriatiche e ioniche.

Sull'Adriatico si sviluppa, lungo la linea di costa, un'area pianeggiante, denominata "Litorale Barese", mentre sul versante ionico si estende la piana indicata comunemente come "Arco Jonico Tarantino".

Il nord del Salento è costituito da una zona completamente priva di rilievi denominata "Tavoliere di Lecce"; verso sud, invece, si riscontrano tre serie allungate di basse elevazioni, le "Serre Salentine", la cui altitudine massima si aggira intorno ai 200 m s.l.m..

La regione manca di una vera e propria idrografia superficiale e presenta, quindi, corsi d'acqua fluviali a regime prevalentemente torrentizio. L'area che è direttamente interessata da tali corsi è il Tavoliere, attraversato dai fiumi Fortore, Candelaro (con i suoi affluenti Sasola, Celone), Cervaro, Carapelle, Ofanto, i quali sfociano tutti

nell'Adriatico, il primo a nord-ovest del promontorio del Gargano, gli altri nel Golfo di Manfredonia.

L'Ofanto è il più importante dei fiumi italiani del versante adriatico, a sud del Reno; esso nasce presso Nusco in Irpinia e, dopo 165 Km, si versa nell'Adriatico a nord di Barletta; ha una portata media annua di 15,20 m³ al secondo e attraversa la Puglia per circa 50 km, molto meno degli altri corsi d'acqua che solcano il Tavoliere: il Candelaro (70 Km), il Sasola (60 Km), il Cervaro (80 Km), il Carapelle (85 Km), il Celone (59 Km). Altri corsi d'acqua di interesse regionale sono il Fortore (86 Km di cui 25 in Puglia), il Lato e il Galese nel tarantino, il Canale Reale presso Brindisi.

Le portate medie di questi torrenti sono esigue; il regime è molto irregolare ed è caratterizzato da magre estive e piene autunnali-invernali che, in passato, hanno dato luogo a rovinose inondazioni. I fiumi pugliesi presentano un tracciato irregolare.

1.2 Il clima

La Puglia ha un clima tipicamente temperato, la sua latitudine la pone al centro dell'omonima zona climatica. Inoltre, l'ampia area a

contatto con il mare e la scarsa altitudine fanno sì che il clima, in generale, si mantenga temperato.

La caratteristica climatica, che maggiormente condiziona lo sviluppo della vegetazione nel territorio regionale, è la coincidenza del periodo delle più elevate temperature con quello della quasi assoluta mancanza di precipitazioni. Gli apporti idrici meteorici vengono a mancare proprio nei mesi in cui la vegetazione è attiva e, più elevata è la traspirazione delle piante, causata dall'alta richiesta evaporativa dell'atmosfera.

Il clima ha condizionato significativamente lo sviluppo dell'agricoltura pugliese fino agli anni '60, ossia, prima dello sviluppo dell'irrigazione pubblica, quando le colture agrarie si limitavano a quelle erbacee autunno-invernali e a colture arboree tradizionali resistenti alla siccità, quali l'olivo, la vite ad alberello, il fico e il mandorlo.

1.2.1 Le Temperature

Per quanto concerne le temperature, gran parte del territorio regionale presenta medie annuali comprese tra i 16° e i 17° C, con valori più alti (17° - 18° C) nei paesi litoranei del canale d'Otranto e del golfo di Taranto, ed inferiori ai 16° C (fino a 11° - 12° C) nella zona

settentrionale delle Murge, in parte del Tavoliere, del Gargano e nel sub-Appennino. Il mese più freddo è ovunque gennaio, con temperature medie mensili generalmente comprese tra i 6 ° e i 10 ° C, con punte inferiori nelle aree montane e superiori nel Salento. Il mese più caldo è agosto, con temperature medie comprese tra i 24 ° e i 26 ° C, con valori inferiori nelle aree montane e in parte delle Murge.

L'escursione termica annuale si aggira tra i 16 ° e i 20 ° C e aumenta col crescere della latitudine e spostandosi dalla costa verso le aree interne. L'analisi delle sequenze agrometeorologiche evidenzia diversi estremi climatici: per quanto riguarda i valori minimi di temperatura dell'aria, non è raro registrare valori inferiori agli 0 ° C, fino a minimi di -6 ° C nell'alto Gargano e nel Tavoliere, e, con frequenza minore, nelle Murge e nel Salento. Temperature rigide si registrano solitamente tra novembre e febbraio, più frequentemente a gennaio; occasionalmente si verificano anche in marzo ed aprile, con grave danno per le colture agrarie.

1.2.2 Pioggia e siccità

Le piogge sono concentrate nel periodo autunno-invernale, mentre nella stagione estiva è evidente l'esiguo numero di giorni piovosi, con un minimo assoluto nel mese di agosto. Non di rado si registrano periodi di persistente deficienza della piovosità di due o tre mesi o anche di più; ciò rende la Puglia una regione ad elevato rischio di siccità.

Le condizioni meteorologiche verificatesi negli ultimi anni, caratterizzate da una diminuzione delle precipitazioni, soprattutto nel periodo autunnale ed invernale, hanno determinato nelle regioni dell'Italia meridionale una crisi idrica, sia per l'agricoltura sia per gli altri settori produttivi. Tale situazione, infatti, non permettendo l'alimentazione regolare degli invasi, che assicurano gli approvvigionamenti idrici, limita le disponibilità di acqua. La gravità del fenomeno siccitoso è, di seguito, analizzata tramite i trend di precipitazione in nove stazioni di rilevazioni meteorologiche della Puglia.

Le precipitazioni variano tra 414 mm, a Taranto, e 659 mm, a Monte Sant'Angelo. Queste quantità annuali evidenziano la bassa piovosità rispetto ad altre regioni.

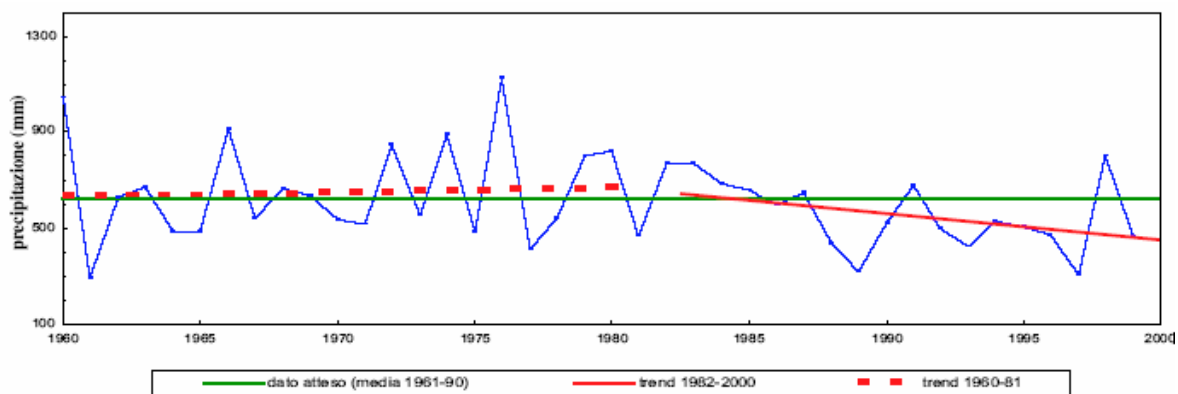
Per quanto riguarda l'andamento temporale delle precipitazioni, sono state rappresentate graficamente (Figure 1 - 9) le piogge annue del quarantennio 1960-1999, al fine di individuare l'eventuale presenza di un trend all'aumento o alla diminuzione della pluviometrica.

Le distribuzioni dei dati lasciano ipotizzare una tendenza, piuttosto generale, alla diminuzione dell'apporto pluviometrico, con l'eccezione delle stazioni di Brindisi (Figura 3) e Taranto (Figura 9), che mostrano un andamento di sostanziale mantenimento del dato medio. Le stazioni di Bari (Figura 2) e Santa Maria di Leuca (Figura 7) mostrano, dagli anni '80, valori poco oscillanti, ma di circa 150-200 mm al di sotto del dato medio. Nelle altre stazioni (Figura 1, Figura 4 e Figura 5), il trend è sempre negativo, con tendenze alla diminuzione più accentuate a Grottaglie (Figura 8) e a Monte Sant'Angelo (Figura 6).

In conclusione, l'analisi qui riportata rende evidente, non solo la bassa piovosità in Puglia, ma soprattutto una tendenza alla riduzione delle precipitazioni nel medio-lungo periodo. L'analisi dei dati pluviometrici evidenzia che ci sono differenze tra le stazioni, ma tutte indicano un aumento della siccità.

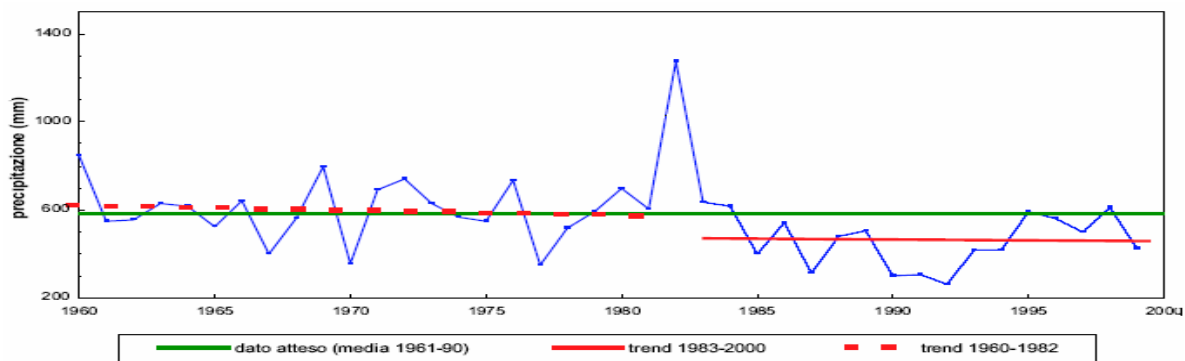
Trend di precipitazione delle stazioni agrometeorologiche della Puglia.

Figura 1 - Stazione meteorologica di Lecce



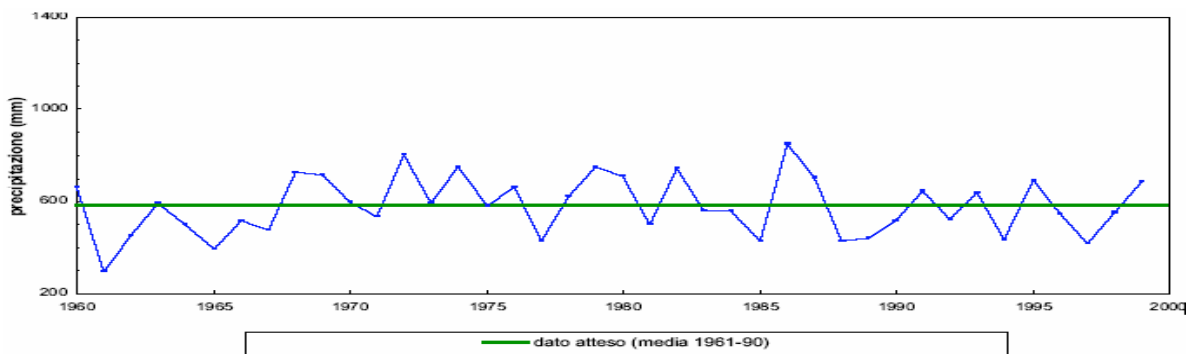
Fonte: INEA, Lo stato dell'irrigazione in Puglia, vol. 1, 2001.

Figura 2 - Stazione meteorologica di Bari-Palese



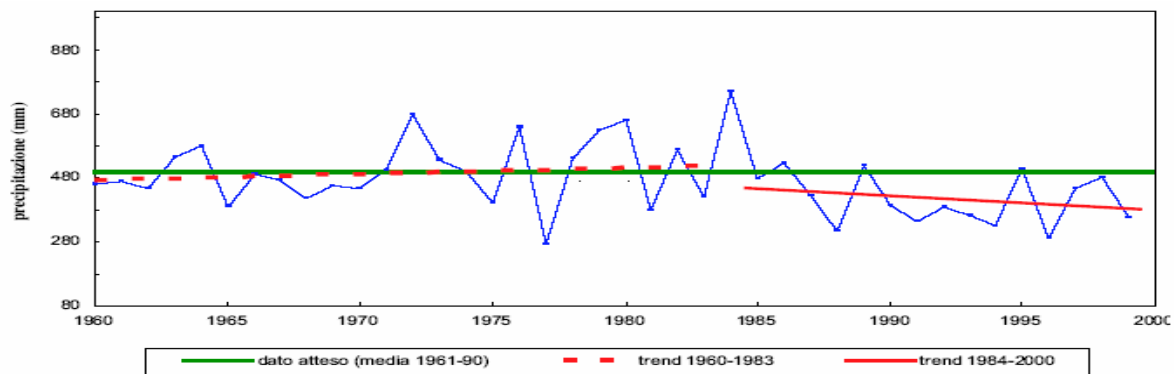
Fonte: INEA, Lo stato dell'irrigazione in Puglia, vol. 1, 2001.

Figura 3 - Stazione meteorologica di Brindisi



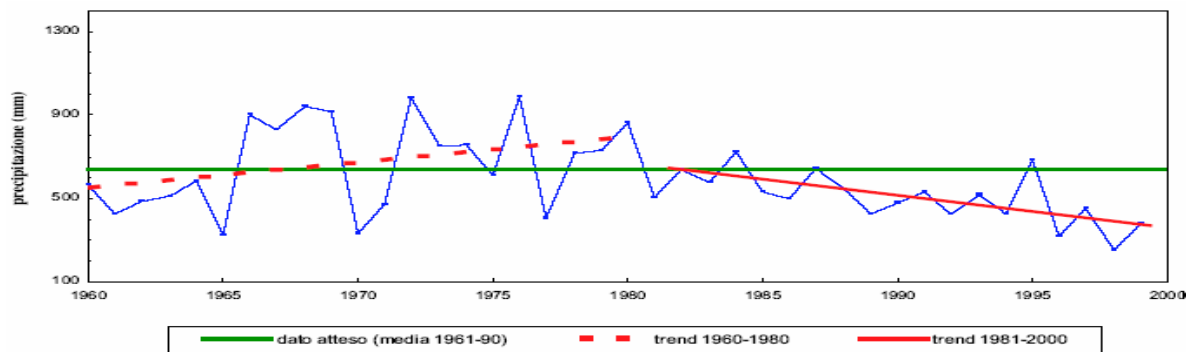
Fonte: INEA, Lo stato dell'irrigazione in Puglia, vol. 1, 2001.

Figura 4 - Stazione meteorologica di Foggia



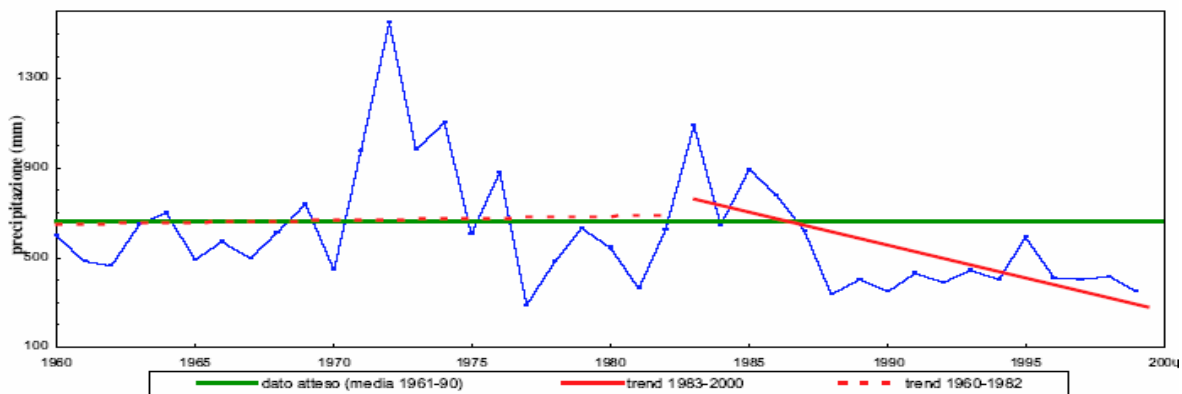
Fonte: INEA, Lo stato dell'irrigazione in Puglia, vol. 1, 2001.

Figura 5 - Stazione meteorologica di Gioia del Colle



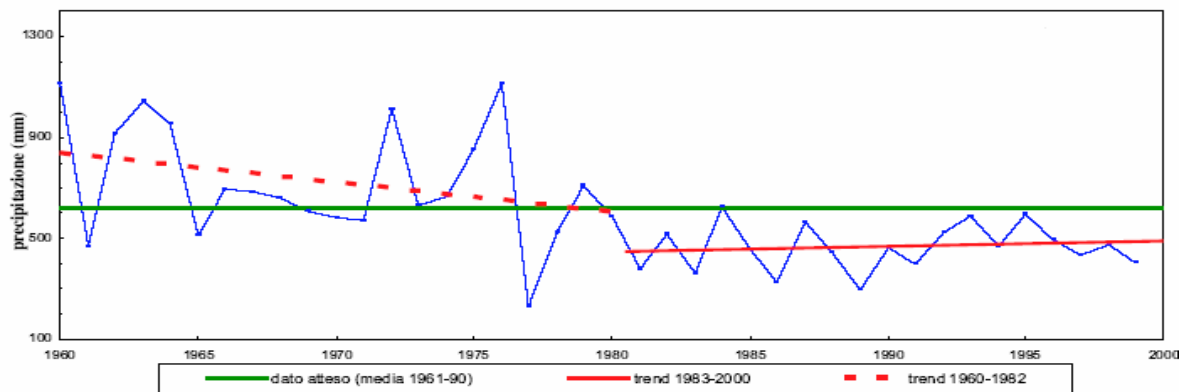
Fonte: INEA, Lo stato dell'irrigazione in Puglia, vol. 1, 2001.

Figura 6 - Stazione meteorologica di Monte Sant'Angelo



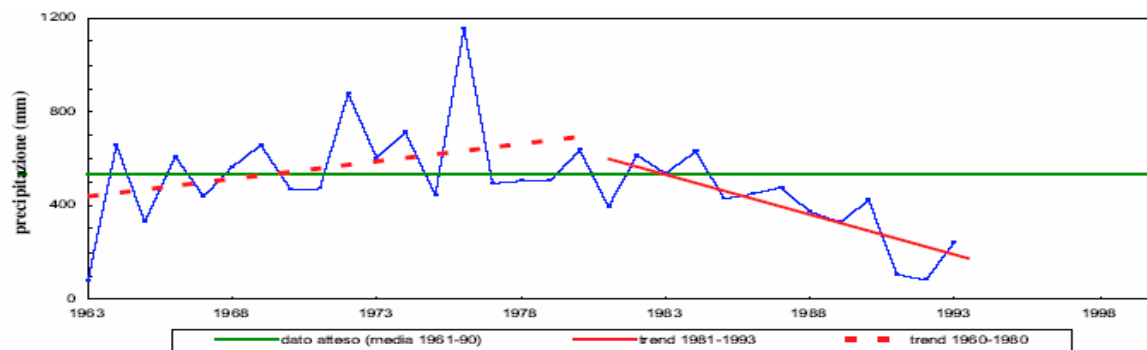
Fonte: INEA, Lo stato dell'irrigazione in Puglia, vol. 1, 2001.

Figura 7 - Stazione meteorologica di Santa Maria di Leuca



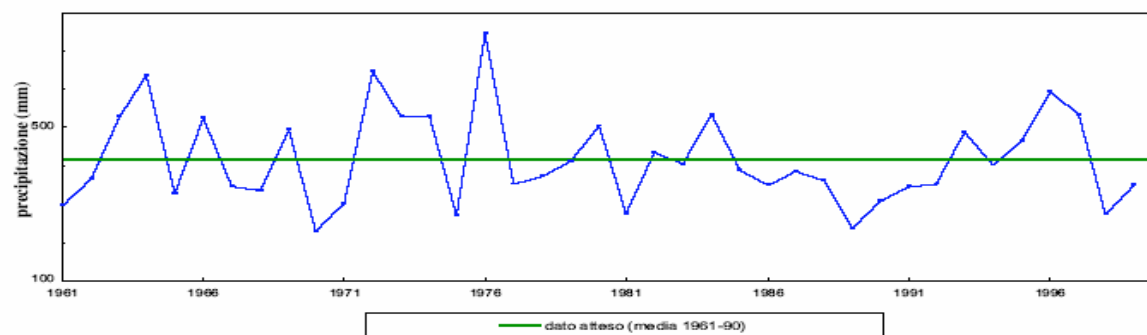
Fonte: INEA, Lo stato dell'irrigazione in Puglia, vol. 1, 2001.

Figura 8 - Stazione meteorologica di Grottaglie



Fonte: INEA, Lo stato dell'irrigazione in Puglia, vol. 1, 2001.

Figura 9 - Stazione meteorologica di Taranto

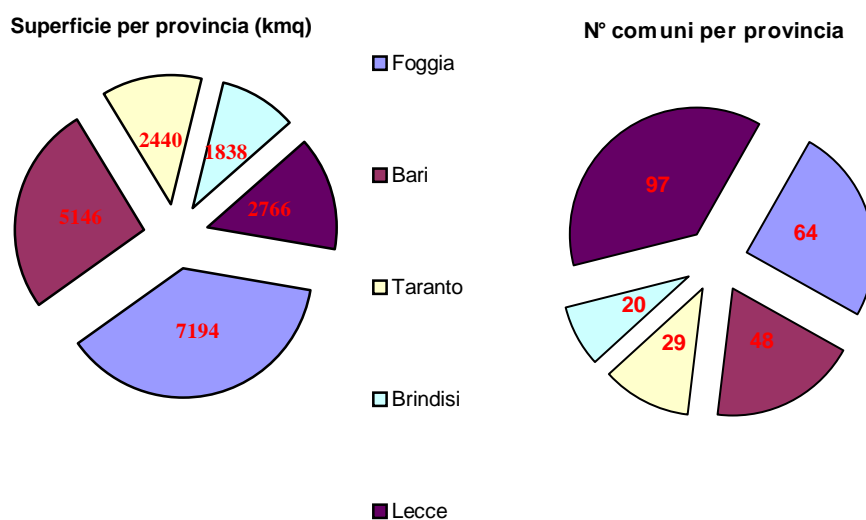


Fonte: INEA, Lo stato dell'irrigazione in Puglia, vol. 1, 2001.

1.3 Aspetti socio-economici

1.3.1 Aspetti demografici

La popolazione residente in Puglia al 31 dicembre 2004, ammontava a 4.068.167 unità, per una densità di popolazione di 210 abitanti/Km², con una significativa variabilità per provincia.

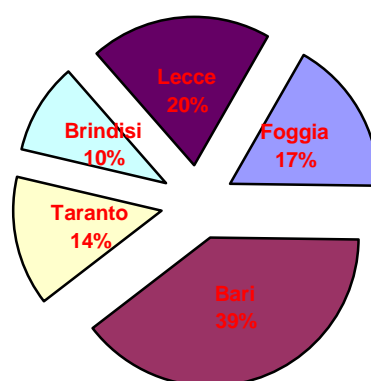


Fonte: dati ISTAT

La popolazione residente per provincia nei comuni della provincia di Foggia è di 686.856 abitanti (il 16,9 per cento del totale), nei comuni della provincia di Bari di 1.594.109 abitanti (il 39,2 per cento), in provincia di Taranto di 580.588 abitanti (il 14,3 per cento), in provincia

di Brindisi di 401.217 abitanti (il 9,9 per cento) e in provincia di Lecce di 805.397 abitanti (il 19,8 per cento). La diversa ripartizione territoriale della popolazione dipende dalla natura geografica del territorio, ovvero, dalla presenza dei rilievi, dalla natura dei terreni e dalla loro produttività che dipende, a sua volta, dalla facilità di reperire acque sotterranee e dalle condizioni climatiche.

Distribuzione della popolazione per provincia

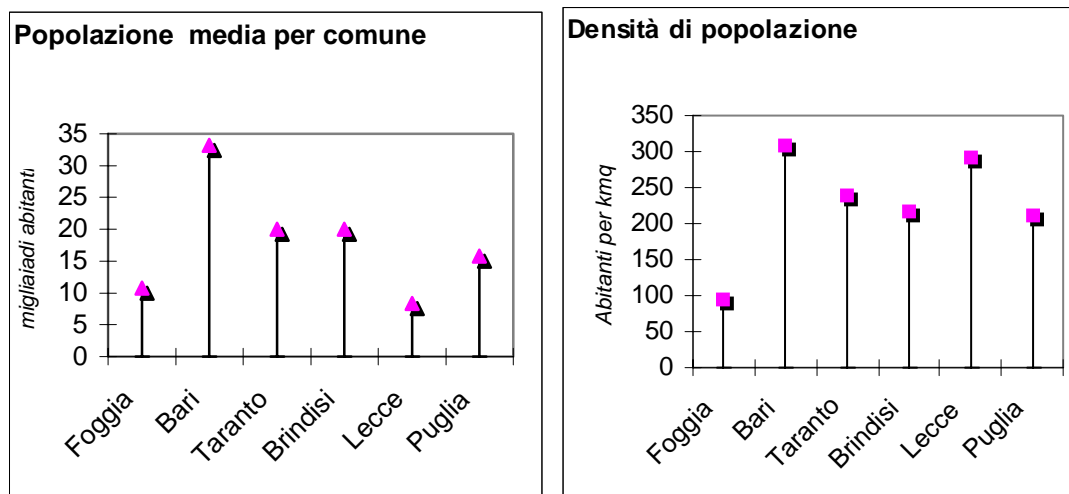


Fonte: dati ISTAT

Un altro fattore che ha assunto notevole rilevanza nella distribuzione della popolazione, è stata la presenza di aree paludose e malariche lungo le coste e la pianura del Tavoliere, che sono state bonificate solo nel periodo 1950-1970.

Un aspetto di rilevante interesse nella dinamica della popolazione regionale è il processo di “concentrazione urbana” che, in Puglia, raggiunge valori massimi nei comuni con più di 20.000 abitanti, in cui si

concentra il 62,7% della popolazione, contro il 55,1% del Mezzogiorno e il 52,7% del valore nazionale.



Fonte: dati ISTAT

L'analisi dei dati *censuari* (vedi figura 10) mostra che, a fronte di una diminuzione delle variazioni dall'8 al 4 per cento circa nei periodi intercensuari del 1971-81 e del 1981-91, la Puglia sia stata interessata da un rallentamento della crescita demografica, anche se di entità modesta, se confrontata alla media nazionale. La causa del rallentamento è legata soprattutto alla notevole flessione della fecondità. Nell'ultimo intervallo, 1991-2001, la popolazione pugliese registra un indice di variazione negativo prossimo allo zero, pari a -0,3 per cento.

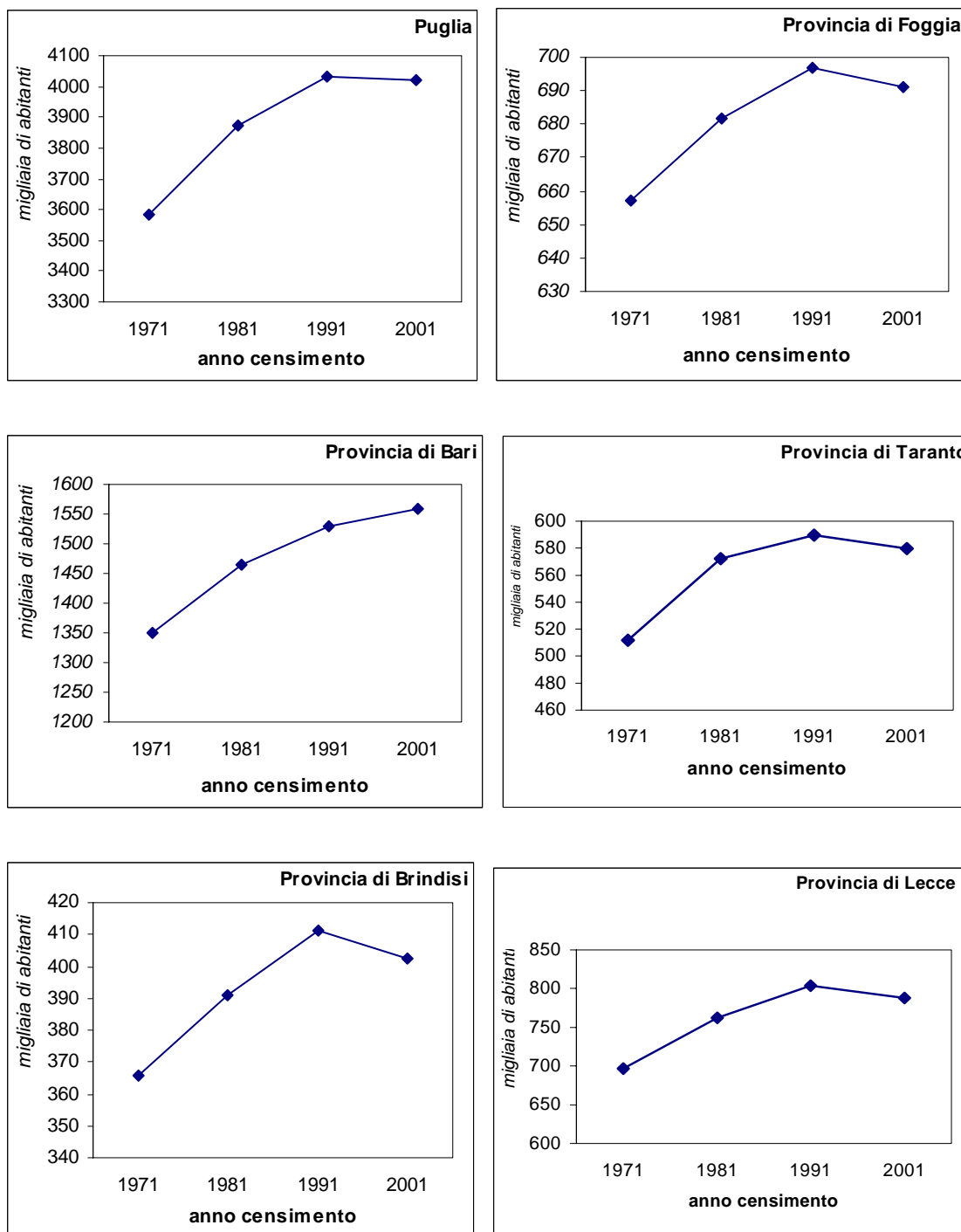
A livello provinciale, soltanto per quella di Bari si assiste a variazioni intercensuarie della popolazione sempre positive.

Nel 2004 si è registrato un incremento della popolazione residente, in larga parte dovuto alle iscrizioni anagrafiche successive alla regolarizzazione degli stranieri presenti in Puglia. L'incremento demografico della regione Puglia è dovuto, in misura consistente, alle immigrazioni che sono largamente superiori alle emigrazioni.

Il tasso di natalità nel 2004, è stato di 10,6 nati per mille abitanti in provincia di Foggia, di 10,3 in provincia di Bari, di 9,6 nella provincia di Taranto, di 9,2 nella provincia di Brindisi, di 9,4 nella provincia di Lecce. Nel complesso la media in Puglia è di 10 nati per mille abitanti ed è in sintonia con la media nazionale, pari a 9,7.

La mortalità presenta quozienti più elevati nelle province a più forte tasso di invecchiamento; i tassi di mortalità vanno da 8,2 morti ogni mille abitanti della provincia di Lecce a 7,2 per mille abitanti della provincia di Bari; seguono le province di Taranto (7,3 per mille abitanti), Brindisi (7,9 per mille abitanti) e Foggia (8,0 per mille abitanti). Il tasso di mortalità in Puglia è di 7,7 morti ogni mille abitanti, in tendenza con i valori del Sud e delle Isole ed è inferiore alla media nazionale (9,4 morti per mille abitanti).

Figura 10 – Andamento demografico in Puglia e nelle province



Fonte: dati ISTAT, “14° Censimento Generale della popolazione”.

1.3.2 Le forze di lavoro

L'Istituto Nazionale di Statistica ha condotto, con riferimento al periodo che va dal 3 gennaio al 3 aprile 2005, la nuova rilevazione continua sulle forze di lavoro. Per la Puglia (vedi Tabella 4), il tasso di attività (rapporto tra le persone appartenenti alle forze di lavoro e la corrispondente popolazione di riferimento) è risultato del 52,3 %; il tasso di occupazione (rapporto tra gli occupati e la corrispondente popolazione di riferimento) del 44,1 %; il tasso di disoccupazione (rapporto tra le persone in cerca di occupazione e le corrispondenti forze di lavoro) del 15,5 %. I tassi sono in linea con i dati riferiti al Mezzogiorno, ma non con quelli nazionali, rispettivamente pari a 62,3, 57,1 e 8,2 per cento.

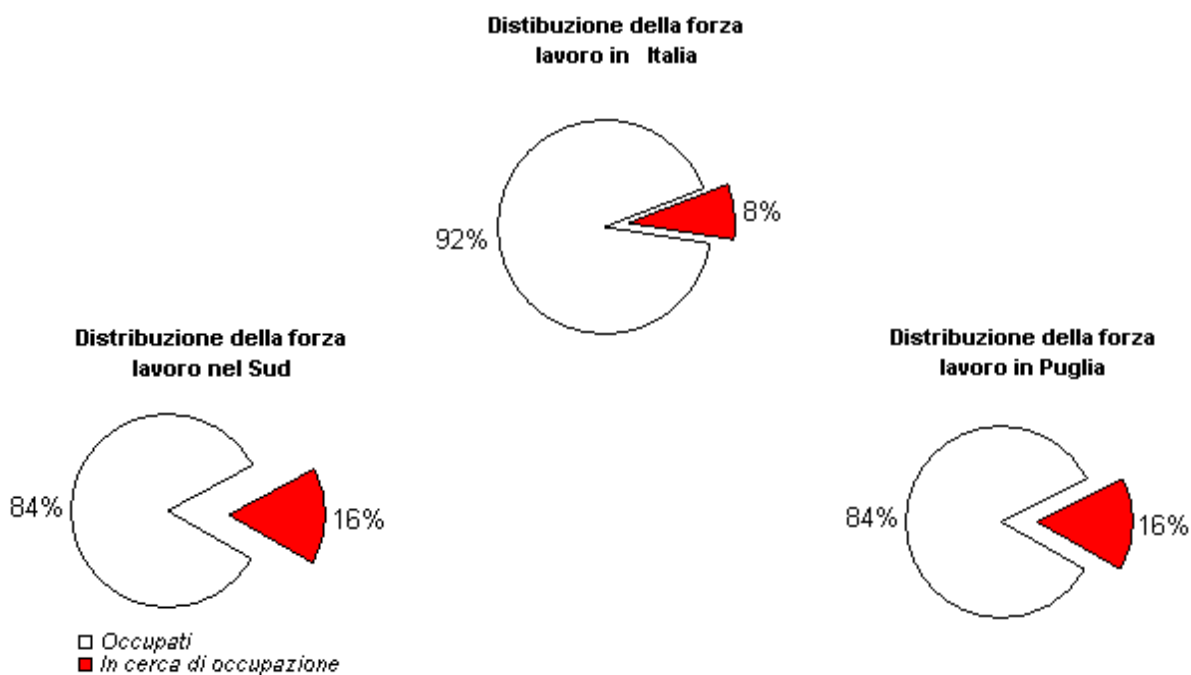
Tabella 4-Forze di lavoro e principali indicatori del mercato del lavoro

	ITALIA	MEZZOGIORNO	PUGLIA
<i>Forze di lavoro</i> ¹	24383	7488	1431
<i>Occupati</i> ¹	22373	6321	1209
<i>In cerca di occupazione</i> ¹	2011	1167	222
<i>Tasso di attività</i> ²	62,3	53,7	52,3
<i>Tasso di occupazione</i> ²	57,1	45,3	44,1
<i>Tasso di disoccupazione</i> ²	8,2	15,6	15,5

Fonte: ISTAT, Rilevazione trimestrale delle forze di lavoro, I trimestre 2005

¹ dati espressi in migliaia di unità

² valori percentuali



Fonte: ISTAT, Rilevazione trimestrale delle forze di lavoro, I trimestre 2005

Con riferimento al periodo 1995-2003, sono state considerate le occupazioni dipendenti, indipendenti e totali nelle tre aree geografiche (Italia, Mezzogiorno, Puglia), classificandole nei principali aggregati economici: primario (agricoltura, silvicoltura e pesca), industria, terziario. Come riportato nella Tabella 5, in Italia, nel periodo di riferimento, l'occupazione alle dipendenze è aumentata del 13,26 per cento (2.131 mila unità); anche le posizioni lavorative indipendenti (Tabella 6) negli anni sono aumentate del 2,72 per cento. Nel complesso, l'occupazione registra un incremento del 10,43 per cento, pari a 2.293 mila unità (Tabella 7).

L'agricoltura registra un forte calo della domanda di lavoro, pari al 18,65 per cento rispetto al 1995. La riduzione degli occupati nel settore

primario ha riguardato prevalentemente la componente autonoma e tutte le aree territoriali.

In compenso, l'industria e, in particolar modo, il terziario presentano incrementi rilevanti soprattutto nel lavoro subordinato, dove i servizi registrano un aumento di occupati del 20 per cento, 1.998 mila unità circa.

Il Mezzogiorno (si vedano le Tabella 8, Tabella 9, Tabella 10) segue le tendenze osservate in Italia, distinguendosi, però, per un maggiore incremento dell'occupazione dipendente nell'industria e, allo stesso tempo, per una più accentuata flessione della componente autonoma del settore primario.

In Puglia la tendenza negativa, riscontrata nel settore primario in Italia e nel Mezzogiorno, assume un calo più lieve con una riduzione di occupati del 15 per cento circa (si veda Tabella 13). Il dato caratterizzante è senz'altro quello riguardante la componente autonoma (vedi Tabella 12) che, nel complesso, registra un incremento di oltre 5 punti percentuali. I settori dell'industria e, in particolar modo, dei servizi denotano una dinamica positiva in quanto registrano, rispettivamente, incrementi del 4 e del 12 per cento circa. Dalla Tabella 11, relativa all'occupazione dipendente, si evince una variazione complessiva del lavoro subordinato meno marcata rispetto alle altre unità territoriali.

Tabella 5 – Occupati dipendenti in Italia per settore di attività economica (migliaia di unità)

Aggregati	Occupati dipendenti		
	1995	2003	var. %
PRIMARIO	638	545	-14.59
INDUSTRIA	5353	5581	4.25
SERVIZI	10080	12078	19.82
TOTALE	16072	18203	13.26

Fonte: dati ISTAT, "Conti regionali"

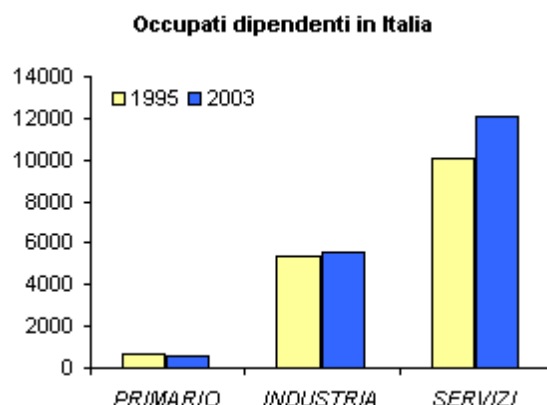


Tabella 6 – Occupati indipendenti in Italia per settore di attività economica (migliaia di unità)

Aggregati	Occupati indipendenti		
	1995	2003	var. %
PRIMARIO	689	535	-22.40
INDUSTRIA	1455	1474	1.34
SERVIZI	3777	4073	7.84
TOTALE	5921	6082	2.72

Fonte: dati ISTAT, "Conti regionali"

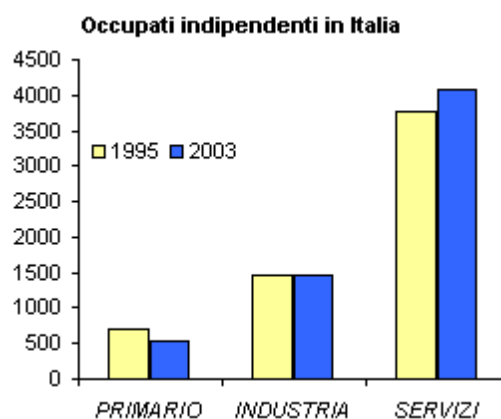


Tabella 7 – Occupati totali in Italia per settore di attività economica (migliaia di unità)

Aggregati	Occupati totali		
	1995	2003	var. %
PRIMARIO	1327	1080	-18.65
INDUSTRIA	6808	7055	3.63
SERVIZI	13857	16151	16.55
TOTALE	21993	24286	10.43

Fonte: dati ISTAT, "Conti regionali"

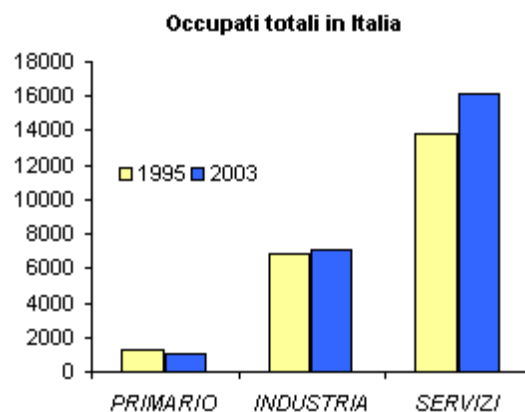


Tabella 8 – Occupati dipendenti nel Mezzogiorno per settore di attività economica
(migliaia di unità)

<i>Aggregati</i>	<i>Occupati dipendenti</i>		
	1995	2003	var. %
PRIMARIO	450	372	-17.24
INDUSTRIA	1058	1150	8.64
SERVIZI	2978	3517	18.11
TOTALE	4486	5039	12.33

Fonte: dati ISTAT, "Conti regionali"

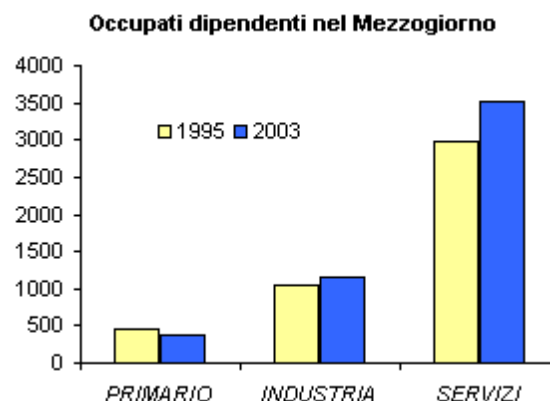


Tabella 9 - Occupati indipendenti nel Mezzogiorno per settore di attività economica
(migliaia di unità)

<i>Aggregati</i>	<i>Occupati indipendenti</i>		
	1995	2003	var. %
PRIMARIO	281	197	-29.75
INDUSTRIA	326	332	1.90
SERVIZI	1117	1196	7.02
TOTALE	1724	1725	0.06

Fonte: dati ISTAT, "Conti regionali"

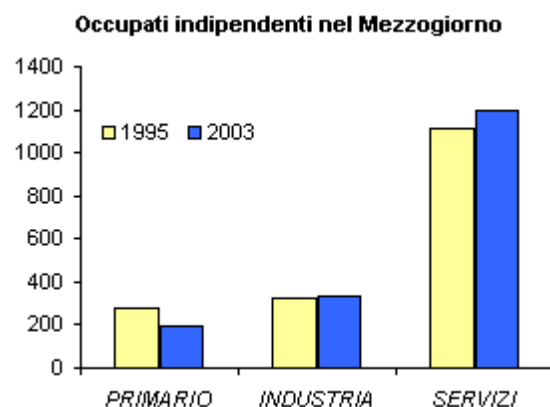


Tabella 10 - Occupati totali nel Mezzogiorno per settore di attività economica
(migliaia di unità)

<i>Aggregati</i>	<i>Occupati totali</i>		
	1995	2003	var. %
PRIMARIO	731	570	-22.05
INDUSTRIA	1384	1482	7.05
SERVIZI	4095	4713	15.08
TOTALE	6210	6764	8.92

Fonte: dati ISTAT, "Conti regionali"

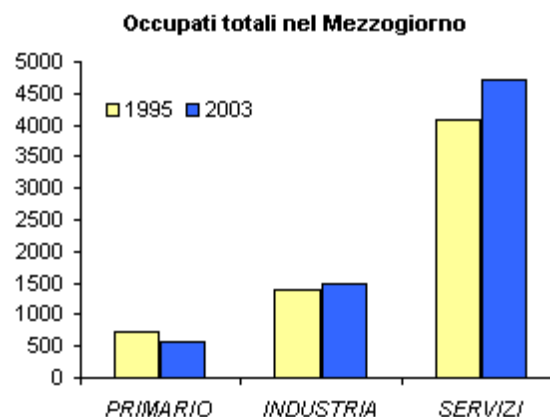


Tabella 11 – Occupati dipendenti in Puglia per settore di attività economica (migliaia di unità)

Aggregati	Occupati dipendenti		
	1995	2003	var. %
PRIMARIO	114	100	-12.62
INDUSTRIA	237	251	5.79
SERVIZI	561	642	14.59
TOTALE	912	993	8.90

Fonte: dati ISTAT, "Conti regionali"

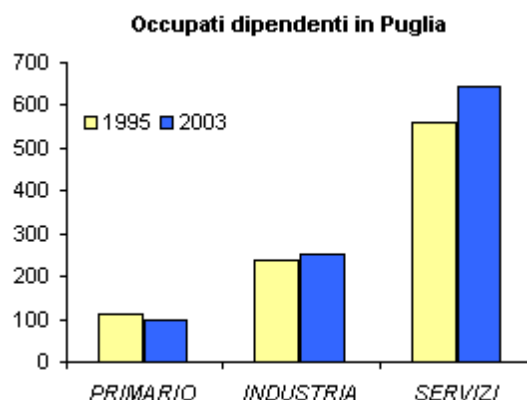


Tabella 12 – Occupati indipendenti in Puglia per settore di attività economica (migliaia di unità)

Aggregati	Occupati indipendenti		
	1995	2003	var. %
PRIMARIO	52	41	-21.31
INDUSTRIA	66	69	3.95
SERVIZI	222	248	11.80
TOTALE	340	358	5.20

Fonte: dati ISTAT, "Conti regionali"

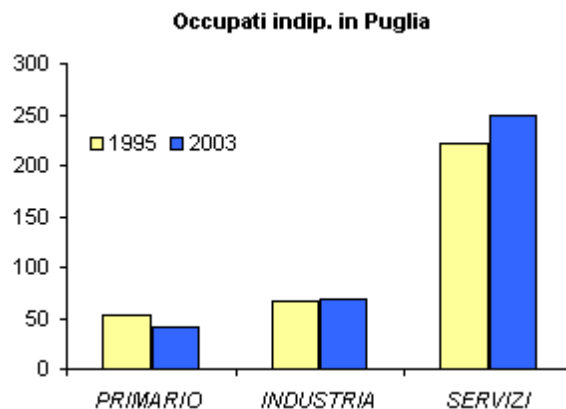
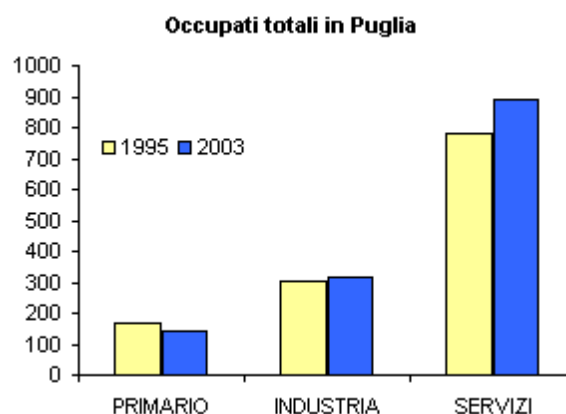


Tabella 13 – Occupati totali in Puglia per settore di attività economica (migliaia di unità)

Aggregati	Occupati totali		
	1995	2003	var. %
PRIMARIO	166	141	-15.34
INDUSTRIA	303	319	5.38
SERVIZI	783	891	13.80
TOTALE	1252	1350	7.89

Fonte: dati ISTAT, "Conti regionali"



Per rilevare e misurare le differenziazioni territoriali delle attività economiche, si farà riferimento ad alcuni strumenti e indicatori che la Statistica Economica ha sviluppato per rispondere all'esigenza informativa connessa con l'esame e l'approfondimento dei problemi riguardanti le interrelazioni fra economia e territorio. Tra questi, assumono particolare importanza, per la semplicità di calcolo, gli indici di localizzazione (o specializzazione), gli indici di dotazione, gli indici di vocazione territoriale¹.

Indicando con:

L_{ir} : gli occupati nell'attività i . ma del territorio r . mo;

L_i : il totale degli occupati nel paese nell'attività i . ma;

L_r : il totale degli occupati nel territorio r . mo;

L : il totale degli occupati nel paese;

P_r : la popolazione presente nel territorio;

P : la popolazione presente nel paese;

si hanno, con riferimento all'unità territoriale r . ma e all'attività economica i . ma i seguenti:

¹ ALVARO G., Contabilità nazionale e statistica economica, Cacucci, Bari, 1999.

a) indice di localizzazione (o specializzazione)

$$I_{Li}^{(r)} = L_{ir} : L_r / L_i : L$$

che pone in relazione l'incidenza degli occupati nell'attività i. ma del territorio r. mo sul complesso degli occupati dell'unità territoriale r. ma rispetto all'incidenza del totale dell'occupazione nella stessa attività produttiva sul complesso del paese. Calcolato con riferimento ai settori dell'agricoltura, dell'industria e del terziario, si hanno, rispettivamente, gli indici di localizzazione agricola, industriale e terziaria;

b) indice di dotazione

$$I_{Di}^{(r)} = L_{ir} : P_r$$

che fornisce l'incidenza degli occupati nell'attività i. ma del territorio r. mo rispetto alla popolazione presente in questo territorio;

c) indice di vocazione territoriale

$$I_{Vi}^{(r)} = L_{ir} : P_r / L_i : P$$

che misura il rapporto fra l'incidenza degli occupati nell'attività i. ma rispetto alla popolazione del territorio r. mo e quella omologa riferita al paese.

Sapendo che la popolazione al 31 dicembre 2003 in Italia, nel Mezzogiorno e in Puglia era rispettivamente di 57.321.070, 20.557.362,

4.023.957 abitanti e considerando i dati sull'occupazione per l'anno 2003 come riportati nelle Tabella 7, Tabella 10, Tabella 13, sono stati calcolati i suddetti indici.

Tabella 14 – Indici di distribuzione locale delle imprese

INDICI DI LOCALIZZAZIONE			
SETTORE	MEZZOG. ¹	PUGLIA ¹	PUGLIA ²
PRIMARIO	1.8937	2.3431	1.2373
INDUSTRIA	0.7541	0.8132	1.0783
SERVIZI	1.0476	0.9918	0.9467

INDICI DI DOTAZIONE			
SETTORE	ITALIA	MEZZOG.	PUGLIA
PRIMARIO	0.0188	0.0277	0.0350
INDUSTRIA	0.1231	0.0721	0.0793
SERVIZI	0.2818	0.2293	0.2213

INDICI DI VOCAZIONE TERRITORIALE			
SETTORE	MEZZOG. ¹	PUGLIA ¹	PUGLIA ²
PRIMARIO	1.4707	1.8560	1.2619
INDUSTRIA	0.5857	0.6441	1.0997
SERVIZI	0.8136	0.7856	0.9655

Fonte: dati ISTAT, "Conti regionali"

¹ gli indici sono calcolati con riferimento all'Italia

² gli indici sono calcolati con riferimento al Mezzogiorno

Dai risultati si evince che l'incidenza degli occupati nel settore dei servizi risulta pressoché omogenea fra le unità territoriali. Le maggiori differenziazioni territoriali delle attività economiche riguardano i settori dell'agricoltura e dell'industria.

Il peso del settore agricolo, infatti, influisce particolarmente sull'economia del Mezzogiorno e ancor più su quella pugliese. L'agricoltura rappresenta un settore importante per l'economia pugliese.

Nel Mezzogiorno, è meno rilevante l'incidenza degli occupati nel settore dell'industria, anche se, per la Puglia, si denota, pur lievemente, una maggiore concentrazione territoriale delle attività economiche.

Nel caso del sistema economico italiano, infatti, la dicotomia nord-sud appare come la conseguenza naturale di una economia di mercato che, non adeguatamente indirizzata e governata, ha prodotto una concentrazione delle attività produttive in prossimità dei centri localizzativi del nord, i quali offrivano (ed offrono) larga disponibilità di forza-lavoro professionalizzata ed una integrazione con i mercati delle materie prime e dei prodotti finiti a livello sia nazionale che europeo.

CAPITOLO 2

PIANO DELL'INDAGINE

2.1 Obiettivo dell'indagine

La Puglia presenta rischi ambientali riferibili al particolare assetto idrogeologico del territorio e all'uso non sempre sostenibile delle risorse naturali, in particolare suolo e acqua. Le problematiche più evidenti sono legate allo sviluppo intenso dei centri urbani, in particolare sulla costa, ed all'inquinamento di alcune zone marine e costiere. Con riferimento alle attività agricole, si segnalano l'uso di acque di falda per l'irrigazione, il deterioramento della loro qualità per l'eccessivo emungimento e il rischio di desertificazione.

Infatti, la scarsità di corpi idrici superficiali e la concomitante ricchezza della circolazione idrica sotterranea stanno determinando, in Puglia, una condizione generalizzata di eccessivo prelievo da falda, spesso in forme illecite e abusive, per l'uso potabile ed irriguo.

Il contenuto salino della falda dipende, in condizioni naturali, dalla concentrazione salina dell'acqua che la ravvena e dall'influenza esercitata dall'acqua marina. La distribuzione del contenuto salino è molto influenzata dall'azione esercitata dagli attingimenti: quando questi superano determinati limiti di prelievo, si esercita un richiamo d'acqua salata dal basso (dalla zona di transizione fra acque dolci e acque salate), o direttamente dal mare, con conseguente riduzione del volume di "acque dolci" in seno all'acquifero.

L'uso di queste acque in agricoltura determina una progressiva salinizzazione dei suoli irrigati, con conseguenze dirette di fitotossicità di alcuni ioni, di alterazioni nella fisiologia delle colture e, nel medio-lungo periodo, di produttività dei suoli. A tal riguardo, è opportuno sottolineare che il fenomeno della progressiva salinizzazione è da ritenersi in molti casi irreversibile, in considerazione dei lunghi tempi necessari per il ripristino della fertilità, una volta eliminate le cause del degrado sulle falde e sui suoli.

Le aree del territorio pugliese in cui si ravvisano tali fenomeni sono il Salento, l'Arco Jonico Tarantino ed il Litorale Adriatico; qui, negli ultimi dieci anni, si è riscontrata una sensibile salinizzazione di alcune falde. Non a caso, proprio su tali aree, gravano i maggiori rischi di degrado e desertificazione.

La diffusione di sistemi colturali intensivi e l'adozione di strategie commerciali, influenzate dalle politiche nazionali ed europee, hanno favorito le scelte produttive non compatibili, dal punto di vista ecologico, con l'ambiente e, se da una parte hanno contribuito ad aumentare le produzioni e di conseguenza i profitti, dall'altra hanno generato situazioni di eccessivo sfruttamento del territorio.

Infine, le variazioni climatiche tendono ad esporre a rischio di aridità superfici sempre più ampie, ad amplificare gli eventi siccitosi ed a moltiplicare gli effetti erosivi della pioggia.

Concludendo, in Puglia, la scarsità di corpi idrici superficiali rende le acque di falda spesso l'unica fonte di approvvigionamento disponibile, dato che i consorzi operano solo su una modesta frazione del territorio regionale.

La qualità delle acque sotterranee regionali risulta particolarmente vulnerabile a causa:

- del delicato assetto idrogeologico;
- dei cambiamenti climatici in atto;
- dei prelievi indiscriminati;
- della contaminazione da inquinanti di varia natura (reflui domestici e zootecnici, scarichi abusivi);
- dell'intrusione marina.

Non esiste un patrimonio informativo sufficientemente esteso sulle caratteristiche idrogeologiche dei corpi idrici sotterranei, sui prelievi e sulla loro qualità chimico-fisica, perciò è difficile quantificare il danno ambientale prodotto; sicuramente le dimensioni del problema destano preoccupazione. La complessità dei problemi ambientali del territorio rende necessario un quadro conoscitivo, organico ed esaustivo indispensabile prima che venga intrapresa qualsiasi azione sull'utilizzo delle risorse. Per contribuire alla conoscenza dello stato attuale delle risorse ambientali si valuterà l'evoluzione temporale della qualità delle acque di alcune falde della regione Puglia, prese a campione.

Inoltre, vista l'importanza dei fenomeni climatici, in particolare della pioggia, si cercheranno di individuare le possibili relazioni di dipendenza tra i principali indicatori di qualità delle acque di falda e il fenomeno meteorico.

2.2 Le fonti dei dati

La fonte dei dati è costituita dalle rilevazioni effettuate dall'Istituto Sperimentale Agronomico di Bari, presso il quale è possibile eseguire una consultazione.

Il dataset comprende due tipologie di dati. La prima serie riguarda le principali caratteristiche chimico-fisiche registrate, con frequenza quindicinale, nel periodo 2001-2004, in una rete di falde monitorate a scala regionale. La seconda considera l'altezza di pioggia giornaliera registrata nello stesso periodo di riferimento, presso le stazioni agrometeorologiche dell'Istituto Sperimentale Agronomico.

Di seguito, nella Tabella 15, si riporta una classificazione dei siti monitorati per idrostruttura d'appartenenza, profondità della falda, periodo di rilevazione dei dati.

Tabella 15 - Classificazione delle falde monitorate

Codice sito	Sito falda	Idrostruttura di riferimento	Profondità metri	Periodo rilevazione dati Dal - - al
BA1	Az.Scorbeto-Molfetta	MURGE	30	02/04/2001-24/09/2004
BA2	Az.Cooperativa-Molfetta	MURGE	100	02/04/2001-27/09/2004
BA3	Az.Le Coppe-Molfetta	MURGE	50	27/04/2001-25/09/2004
BA4	Az.Agostinelli-Rutigliano	MURGE	200	09/07/2002-15/09/2004
BA5	Az.La Noira(CNR)-Mola	MURGE	20	03/12/2001-03/09/2004
BR1	Az.Laghezza -Brindisi	SALENTO	80	27/03/2001-01/10/2004
BR2	Az.Annichiarico-Carovigno	SALENTO	110	03/05/2001-01/10/2004
BR3	Az.Greco-Carovigno	SALENTO	110	15/07/2002-01/10/2004
FG1	AZ.FiorentinoGuerriero-Lesina	CAPITANATA	20	13/06/2001-14/19/2004
FG2	Az.LiberoV.- Sannicandro	CAPITANATA	20	13/06/2001-30/05/2003
FG3	Az.Chiarella-Lucera	CAPITANATA	35	25/03/2001-03/08/2004
FG4	I.S.A. - Foggia	CAPITANATA	30	24/03/2001-18/11/2003
FG5	Az.Salcuni-S.Giovanni Rotondo	CAPITANATA	120	21/01/2002-15/09/2004
LE1	Ist.Tab. -Lecce	SALENTO	120	29/03/2001-01/10/2004
LE2	Az.Pasanisi-Ruffano	SALENTO	60	13/07/2001-30/10/2004
TA1	Az.Perniola-Taranto	ARCO JONICO	100	20/06/2001-05/09/2004

Per ogni sito monitorato sono state rilevate e calcolate le variabili indicate nella Tabella 16.

Tabella 16 - Variabili rilevate nei vari siti monitorati

DETERMINAZIONI CHIMICO FISICHE		Unità di misura
pH		
Conducibilità Elettrica	ECw	dSm ⁻¹
Durezza		°F
S.A.R.		Adimensionale
GAS DISCIOLTI		Unità di misura
Ione Cloruro	Cl ⁻	g/l
Ione Sodio	Na ⁺	meq/l
Ione Magnesio	Mg ⁺⁺	meq/l
Ione Calcio	Ca ⁺⁺	meq/l
Ione Potassio	K ⁺	meq/l
Ione Nitrato	NO ₃ ⁻	ppm
Ione Solfato	SO ₄ ⁻⁻	meq/l
Carbonati	CaCO ₃	meq/l
Bicarbonati	HCO ₃	meq/l
Estratto Saturo		g/l

Di seguito, sono presi in considerazione i principali parametri che caratterizzano la qualità delle acque e i limiti d'accettabilità per l'uso irriguo.

2.2.1 Reazione in pH

La reazione in pH è un parametro che regola tutte le funzioni biologiche. I valori normali per le acque destinate all'uso irriguo sono quelli compresi nell'intervallo tra 5,5 e 8,5; valori ottimi sono compresi

tra 6,5 e 7,5. Una reazione molto lontana dalla neutralità è sempre indice di qualche anomalia, come il contenuto di sostanze tossiche o l'eccesso di certi sali².

2.2.2 Salinità e sodicità

L'acqua usata per l'irrigazione contiene sempre una certa quantità di sali disciolti, la cui concentrazione varia entro limiti piuttosto ampi in relazione alla loro origine. Le acque provenienti da corsi superficiali naturali (es. fiumi ad una sufficiente distanza dalla foce marina) e da invasi (naturali e artificiali) generalmente hanno un basso contenuto in sali disciolti; invece le acque provenienti da falde più o meno profonde possono presentare un contenuto salino elevato.

L'eccessivo emungimento delle acque di falda lungo zone costiere caratterizzate da formazioni geologiche permeabili (formazioni sabbiose o rocce fessurate) consente all'acqua del mare di invadere il sottosuolo dell'entroterra fino a distanze dalla costa anche notevoli e con pendenza variabile in relazione alla conducibilità idrica del mezzo poroso attraversato ed al carico di acqua dolce stratificatosi su di essa.

² AA.VV. (a cura di P.SCANDELLA, G. MECELLA), Irrigazione sostenibile: la buona pratica irrigua, Ministero delle Politiche Agrarie e Forestali, Progetto Editoriale PANDA, Volume No 5, cap. 2, Edizioni L'Informatore Agrario, Verona, 2004.

Le acque di falda (generalmente dolce), per la minore densità rispetto a quella dell'acqua del mare, si stratifica al disopra di quest'ultima e tende a scorrere verso la costa con pendenza che dipende dal carico idraulico e dalla conducibilità idrica della zona acquifera.

Il contenuto salino di un'acqua viene comunemente valutato in termini di **Conducibilità Elettrica (ECw)** e la misura viene espressa in dS m^{-1} . Un'acqua si definisce salmastra³ se il valore di ECw supera i $3,0 \text{ dS m}^{-1}$.

Un altro aspetto da considerare è la qualità dei sali disciolti nell'acqua, in quanto questa può influenzare le proprietà chimiche e fisiche del terreno. L'azione dell'acqua nei riguardi delle proprietà fisiche del terreno dipendono molto anche dai cationi presenti in esso e, principalmente, dal **Ca** sottoforma di calcare attivo.

Per tenere conto di questi fatti sono stati proposti vari indici di qualità dell'acqua irrigua. Tra questi il più comune è il rapporto di assorbimento del sodio (**S.A.R. "Sodium Absorbation Ratio"**) :

$$\text{SAR} = \frac{\text{Na}^+}{\sqrt{\frac{\text{Ca}^{++} + \text{Mg}^{++}}{2}}}$$

dove la concentrazione del **Na**, **Ca**, **Mg**, è espressa in meq L^{-1} .

³ AA.VV. (a cura di P.SCANDELLA, G. MECELLA), Irrigazione sostenibile: la buona pratica irrigua,.....ecc., op.cit..

In conformità a tali indici si riporta (Tabella 17) la classificazione dell'acqua irrigua proposta dalla F.A.O. :

Tabella 17 - Classificazione dei parametri ECw e SAR nella valutazione della qualità delle acque di falda

Problemi potenziali per l'irrigazione	Parametro	Limitazioni d'uso		
		nessuna	lieve	severa
<i>Salinità</i> ¹	ECw	< 0,7	0,7 - 0,3	> 0,3
<i>Infiltrazione</i> ²	SAR = 0 - 3 con ECw	> 0,7	0,7 - 0,2	< 0,2
	SAR = 3 - 6 con ECw	> 1,2	1,2 - 0,3	< 0,3
	SAR = 6 - 12 con ECw	> 1,9	1,9 - 0,5	< 0,5
	SAR = 12 - 20 con ECw	> 2,9	2,9 - 1,3	< 1,3
	SAR = 20 - 40 con ECw	> 5,0	5,0 - 2,9	< 2,9

Fonte: F.A.O.

¹ : influenza la qualità di acqua per la coltura

² : influenza la velocità di infiltrazione dell'acqua nel terreno tenendo conto, contemporaneamente, della ECw e del SAR

2.2.3 Durezza

Il grado di durezza o grado idrotimetrico rappresenta il complesso di sali di calcio e di magnesio contenuti nell'acqua e calcolati tutti come carbonato o come ossido di calcio. Chiamasi poi durezza totale quella dovuta a tutti i sali di calcio e di magnesio disciolti nell'acqua; durezza

permanente quella dovuta ai detti sali che rimangono sciolti anche dopo l'ebollizione dell'acqua; durezza temporanea quella dovuta ai sali che precipitano dall'acqua con l'ebollizione, che cioè erano prima sciolti allo stato di bicarbonato⁴.

2.2.4 Carbonati e bicarbonati

Un elevato contenuto in **carbonati** e **bicarbonati** può provocare intasamento degli erogatori dell'acqua irrigua, specialmente con impianti di micro-irrigazione⁵.

2.3 Il software utilizzato

Per l'elaborazione e la rappresentazione grafica dei dati ci si è avvalsi prevalentemente del software R per Windows nella versione 2.1.1. Questo, più che un software statistico può essere definito come un ambiente, costituito da una varietà di strumenti, orientato alla gestione, all'analisi dei dati e alla produzione di grafici, basato sul linguaggio S creato da AT&T Bell Laboratories, ambiente dal quale è nato un altro

⁴ VILAVECCHIA V., Trattato di chimica analitica applicata, Ulrico Hoepli, vol. 1, Milano 1967.

⁵ AA.VV. (a cura di P.SCANDELLA, G. MECELLA), Irrigazione sostenibile: la buona pratica irrigua,.....ecc., op.cit..

software commerciale più noto, S-Plus.

R, a differenza di quest'ultimo, è disponibile gratuitamente sotto i vincoli della GPL (General Public License) ed è disponibile per diverse architetture hardware e sistemi informativi. Sul sito <http://www.r-project.org> è possibile scaricare, oltre al programma base, anche una serie di moduli aggiuntivi e un'ampia modulistica sull'ambiente, che va dall'installazione del software al suo utilizzo nell'analisi dei dati.

La versione iniziale di R fu sviluppata nel 1996 dai ricercatori del dipartimento di Statistica dell'Università di Auckland in Nuova Zelanda. In seguito, altri ricercatori iniziarono ad aggiungersi e a fornire il loro contributo alla scrittura del codice sorgente e al miglioramento dell'applicazione.

Oggi R è uno dei software maggiormente utilizzato a livello mondiale dai ricercatori in campo statistico.

L'ambiente R è basato sul concetto di "packages" tradotto di solito in italiano con il termine pacchetto. Un package è un insieme di strumenti che svolgono determinate funzioni, ma può anche contenere solo dati oppure sola documentazione.

Nell'analisi statistica dei dati si è utilizzato il package "base" che contiene gli strumenti per le più importanti e diffuse analisi statistiche

esplorative, descrittive ed inferenziali e il package “ast” che mette a disposizione validi strumenti per l’analisi delle serie storiche⁶.

⁶ Il package ‘base’ fa parte delle biblioteche standard di R e, quindi, viene automaticamente installata con l’installazione del software; ‘ast’, invece, è una biblioteca di funzioni aggiuntive e può essere scaricata dal sito: <http://sirio.stat.unipd.it>

CAPITOLO 3

ANALISI STATISTICA SULLA QUALITA' DELLE ACQUE DI FALDA

3.1 Premessa

La Statistica fornisce metodi che consentono di fare valide induzioni dall'insieme dei dati di osservazione, permettendo di penetrare più a fondo sul meccanismo o sui fattori che hanno determinato il presentarsi del fenomeno con varie modalità.

Si suole oggi distinguere una statistica descrittiva, tendente ad evidenziare le regolarità presenti nei dati, da una statistica inferenziale, tendente a giustificare le osservazioni in termini di modelli teorici esplicativi dei fenomeni collettivi.

Negli ultimi anni, accanto alla sempre maggiore rilevazione di grandi masse di dati ed alla disponibilità di strumenti atti ad elaborarle rapidamente, la statistica descrittiva ha assunto una maggiore importanza, assumendo la nuova etichetta di ‘analisi dei dati’⁷.

Le analisi che si affrontano in questo capitolo seguono questo approccio, perseguendo l’obiettivo di sintetizzare le cifre grezze in un unico valore che sappia cogliere il sottofondo costante della molteplicità dei valori campionati e possa quindi essere comparato con il valore limite dettato dalla F.A.O..

Non si dimentichi, però, che tutte le volte che si sintetizzano più dati con un solo valore si perdono delle informazioni. Il metodo statistico rende più obiettiva questa sintesi.

I valori caratteristici che saranno considerati nell’analisi sono forniti dall’output del comando *summary*. Trattasi:

- dei valori minimo e massimo della distribuzione;
- della media aritmetica μ ;
- della mediana;
- del primo e del terzo quartile.

⁷ GIRONE G., SALVEMINI T., Lezioni di statistica, CACUCCI, Bari, 2002.

La media aritmetica, per le sue note proprietà, sarà presa come valore sintetico. Gli altri risultati, invece saranno utilizzati per trarre valide informazioni circa la forma distributiva della grandezza esaminata.

Per avere una misura della dispersione dei valori rilevati intorno al valore effettivo, in altre parole, per misurare di quanto le quantità rilevate differiscono in media dalla grandezza assunta a rappresentare l'intensità del carattere si farà uso dello scarto quadratico medio σ .

Tuttavia, poiché le modalità delle distribuzioni a confronto sono espresse con unità di misura diverse, tra le quali o non intercede alcuna relazione (ad esempio, ° F e g/l, dsm^{-1} e ppm) oppure intercede un rapporto costante (ad esempio, g/l e meq/l), si è ricorso ad un indice di variabilità relativo al massimo, poiché espresso in termini della variabilità massima, dato dalla formula:

$$V_r = \frac{\sigma}{\max \sigma} \quad \text{con } 0 \leq V_r \leq 1,$$

dove
$$\max \sigma = \sqrt{(\mu - x_{(\min)}) (x_{(\max)} - \mu)}$$

Infine, per descrivere in forma visiva e sintetica le osservazioni, riguardanti i diversi parametri, i dati empirici sono stati raccolti in una

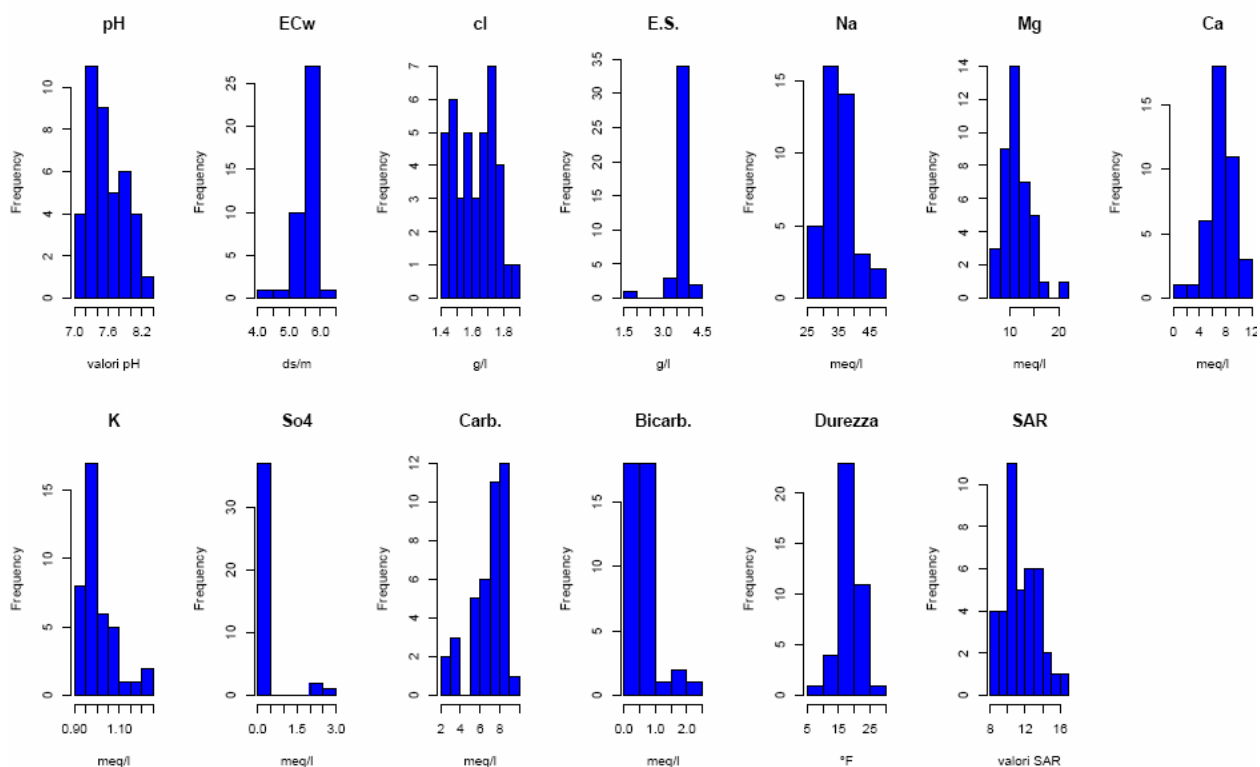
distribuzione di frequenze e rappresentati graficamente tramite istogrammi. La rappresentazione grafica, infatti, facilitando l'intuizione, consente di:

- selezionare l'intervallo di valori in cui ricadono più osservazioni;
- di verificare l'omogeneità delle osservazioni;
- di identificare eventuali valori anomali.

3.2 Idrostruttura delle Murge

3.2.1 Sito BA1

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Durezza	SAR
Min. :7.095	Min. :4.200	Min. : 8.835	Min. : 8.465
1st Qu.:7.287	1st Qu.:5.487	1st Qu.:16.984	1st Qu.:10.262
Median :7.543	Median :5.600	Median :18.416	Median :11.389
Mean :7.569	Mean :5.562	Mean :18.514	Mean :11.548
3rd Qu.:7.801	3rd Qu.:5.753	3rd Qu.:20.399	3rd Qu.:12.808
Max. :8.291	Max. :6.210	Max. :29.284	Max. :16.084
Cloruri	Estr. Saturato	Na	Mg
Min. :1.407	Min. :1.826	Min. :27.17	Min. : 7.667
1st Qu.:1.489	1st Qu.:3.567	1st Qu.:30.77	1st Qu.: 9.439
Median :1.631	Median :3.704	Median :34.15	Median :10.571
Mean :1.611	Mean :3.660	Mean :34.66	Mean :11.363
3rd Qu.:1.706	3rd Qu.:3.789	3rd Qu.:37.59	3rd Qu.:12.956
Max. :1.879	Max. :4.180	Max. :48.48	Max. :20.072

Ca		K		NO3	SO4	
Min.	: 0.362	Min.	:0.900	tracce:40	Min.	:0.01480
1st Qu.:	6.266	1st Qu.:	0.961		1st Qu.:	0.07838
Median :	6.958	Median :	0.995		Median :	0.10480
Mean :	7.151	Mean :	1.012		Mean :	0.27990
3rd Qu.:	8.172	3rd Qu.:	1.028		3rd Qu.:	0.17520
Max.	:11.352	Max.	:1.238		Max.	:2.55700
Carbonati		Bicarbonati				
Min.	:2.440	Min.	:0.00			
1st Qu.:	6.180	1st Qu.:	0.00			
Median :	7.560	Median :	0.56			
Mean :	6.964	Mean :	0.53			
3rd Qu.:	8.120	3rd Qu.:	0.69			
Max.	:9.200	Max.	:2.08			

Valori dello scarto quadratico medio calcolato per ciascun parametro:

σ_{pH}	= 0.3160062	σ_{ECw}	= 0.3466358
σ_{Cl}	= 0.1313459	$\sigma_{E.S.}$	= 0.3483482
σ_{Na}	= 4.9150280	σ_{Mg}	= 2.6763780
σ_{Ca}	= 2.0133280	σ_K	= 0.0758324
σ_{SO4}	= 0.5980578	σ_{Carb}	= 1.7780840
$\sigma_{Bic.}$	= 0.5346458	σ_{Dur}	= 3.6023760
σ_{SAR}	= 1.8733840		

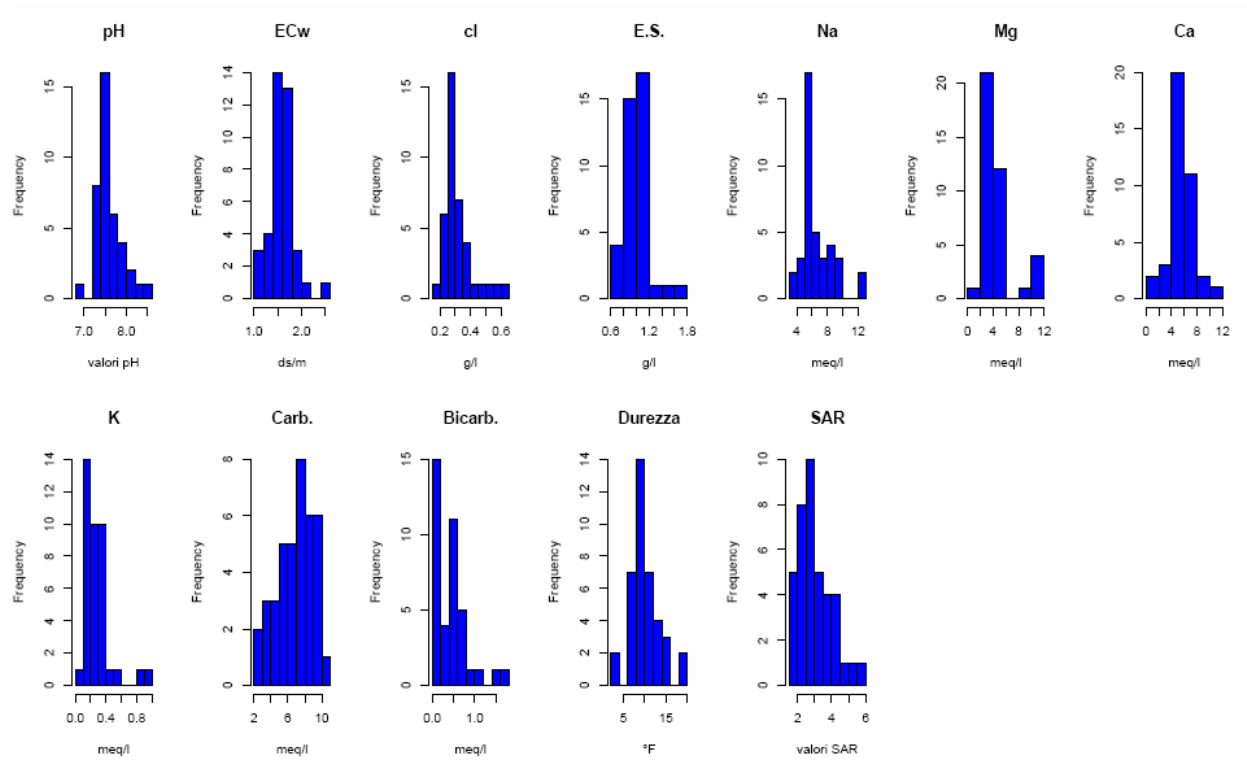
Valori dell'indice di variabilità relativo:

V_r^{pH}	= 0.5401939	V_r^{ECw}	= 0.3690126
V_r^{Cl}	= 0.5618789	$V_r^{E.S.}$	= 0.3566829
V_r^{Na}	= 0.4832479	V_r^{Mg}	= 0.4717318
V_r^{Ca}	= 0.3770000	V_r^K	= 0.4766787
V_r^{SO4}	= 0.7697425	V_r^{Carb}	= 0.5590550
$V_r^{Bic.}$	= 0.5898783	V_r^{Dur}	= 0.3528297
V_r^{SAR}	= 0.5009604		

3.2.2

Sito BA2

Distribuzioni di frequenza per ogni singola caratteristica rilevata



Risultati del comando “summary”:

pH	ECw	Durezza	SAR
Min. :6.865	Min. :1.076	Min. : 3.195	Min. :1.529
1st Qu.:7.423	1st Qu.:1.513	1st Qu.: 8.111	1st Qu.:2.405
Median :7.538	Median :1.582	Median : 9.656	Median :2.776
Mean :7.596	Mean :1.596	Mean :10.066	Mean :3.026
3rd Qu.:7.739	3rd Qu.:1.673	3rd Qu.:11.224	3rd Qu.:3.646
Max. :8.502	Max. :2.590	Max. :18.854	Max. :5.524

Cl	E.S	Na	Mg
Min. :0.1910	Min. :0.648	Min. : 3.260	Min. : 0.484
1st Qu.:0.2655	1st Qu.:0.910	1st Qu.: 5.218	1st Qu.: 3.303
Median :0.2830	Median :1.006	Median : 5.783	Median : 3.965
Mean :0.3171	Mean :1.009	Mean : 6.555	Mean : 4.522
3rd Qu.:0.3345	3rd Qu.:1.089	3rd Qu.: 7.826	3rd Qu.: 4.603
Max. :0.6030	Max. :1.640	Max. :12.087	Max. :10.448

Ca	K	NO3	SO4
Min. : 0.250	Min. :0.0860	0.0280: 1	tracce :21
1st Qu.: 4.601	1st Qu.:0.1850	tracce:38	0.0203 : 1
Median : 5.519	Median :0.2180		0.0242 : 1
Mean : 5.544	Mean :0.2853		0.0273 : 1
3rd Qu.: 6.455	3rd Qu.:0.3285		0.0280 : 1
Max. :10.204	Max. :0.9560		0.0337 : 1
			(Other):13

Carb.		Bic.	
Min.	: 2.440	Min.	:0.0000
1st Qu.	: 5.360	1st Qu.	:0.0000
Median	: 7.360	Median	:0.4400
Mean	: 6.845	Mean	:0.4036
3rd Qu.	: 8.320	3rd Qu.	:0.6000
Max.	:10.120	Max.	:1.6400

Valori dello scarto quadratico medio calcolato per ciascun parametro:

$$\sigma_{pH} = 0.2973092$$

$$\sigma_{ECw} = 0.2621881$$

$$\sigma_{Cl} = 0.0931166$$

$$\sigma_{E.S.} = 0.1961325$$

$$\sigma_{Na} = 1.9985640$$

$$\sigma_{Mg} = 2.3172840$$

$$\sigma_{Ca} = 1.8581110$$

$$\sigma_K = 0.1718566$$

$$\sigma_{Carb.} = 2.0658460$$

$$\sigma_{Bic.} = 0.4239134$$

$$\sigma_{Dur} = 3.2831580$$

$$\sigma_{SAR} = 0.9669778$$

Valori dell'indice di variabilità relativo:

$$V_r^{pH} = 0.3653105$$

$$V_r^{ECw} = 0.3646552$$

$$V_r^{Cl} = 0.4904388$$

$$V_r^{E.S.} = 0.4109490$$

$$V_r^{Na} = 0.4680978$$

$$V_r^{Mg} = 0.4737362$$

$$V_r^{Ca} = 0.3741013$$

$$V_r^K = 0.4700781$$

$$V_r^{Carb} = 0.5439031$$

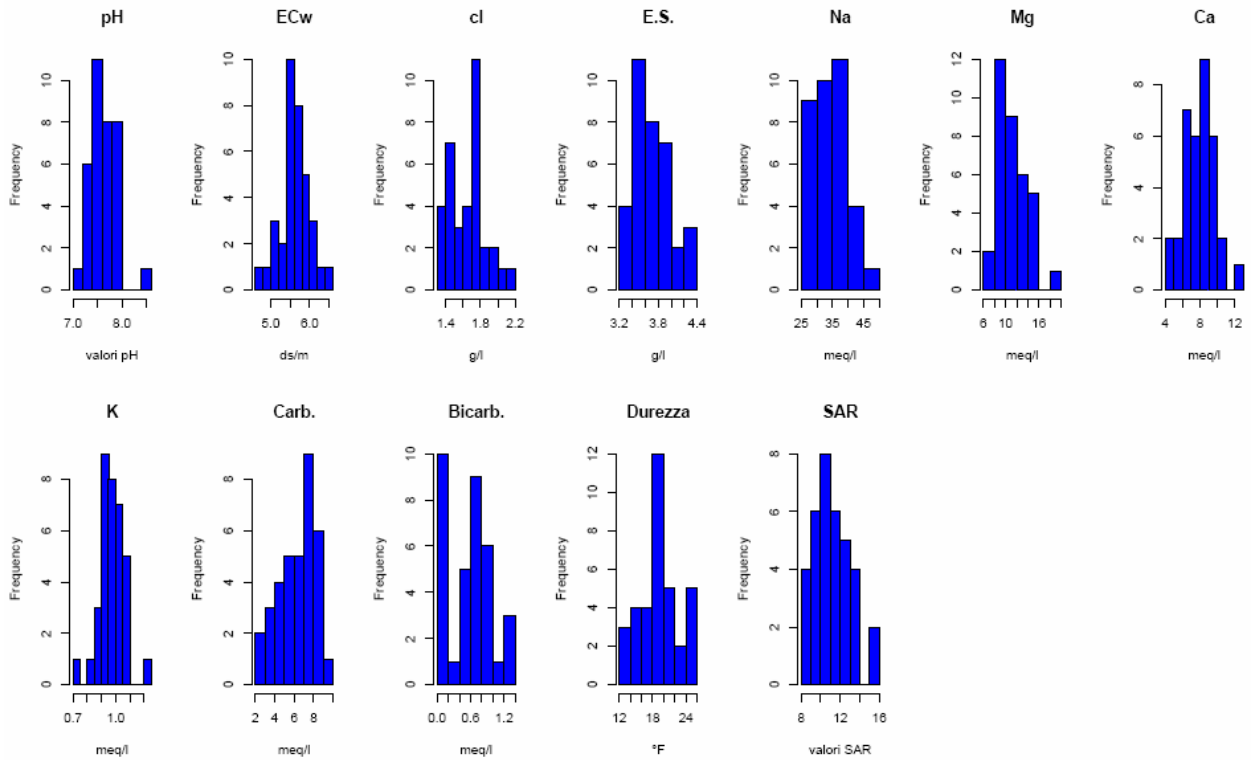
$$V_r^{Bic.} = 0.6001030$$

$$V_r^{Dur} = 0.4225079$$

$$V_r^{SAR} = 0.5000352$$

3.2.3 Sito BA3

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH		ECw		Durezza		SAR	
Min.	:7.157	Min.	:4.640	Min.	:13.17	Min.	: 8.337
1st Qu.	:7.427	1st Qu.	:5.480	1st Qu.	:16.97	1st Qu.	: 9.917
Median	:7.594	Median	:5.610	Median	:18.69	Median	:10.964
Mean	:7.624	Mean	:5.612	Mean	:19.13	Mean	:11.259
3rd Qu.	:7.809	3rd Qu.	:5.830	3rd Qu.	:21.16	3rd Qu.	:12.316
Max.	:8.493	Max.	:6.420	Max.	:25.48	Max.	:15.468
Cl		E.S		Na		Mg	
Min.	:1.312	Min.	:3.350	Min.	:27.09	Min.	: 7.157
1st Qu.	:1.489	1st Qu.	:3.525	1st Qu.	:30.02	1st Qu.	: 8.896
Median	:1.684	Median	:3.670	Median	:34.39	Median	:10.694
Mean	:1.648	Mean	:3.723	Mean	:34.47	Mean	:11.250
3rd Qu.	:1.737	3rd Qu.	:3.873	3rd Qu.	:38.59	3rd Qu.	:13.245
Max.	:2.128	Max.	:4.380	Max.	:47.13	Max.	:18.504
Ca		K		NO3		SO4	
Min.	: 4.635	Min.	:0.7470	tracce:35		0.0904	: 2
1st Qu.	: 6.711	1st Qu.	:0.9223			0.1062	: 2
Median	: 8.154	Median	:0.9620			0.1071	: 2
Mean	: 7.883	Mean	:0.9737			0.0804	: 1
3rd Qu.	: 8.987	3rd Qu.	:1.0165			0.0868	: 1
Max.	:12.110	Max.	:1.2280			0.0877	: 1
						(Other)	:26

Carb.		Bic.	
Min.	:2.280	Min.	:0.0000
1st Qu.	:5.040	1st Qu.	:0.0000
Median	:6.560	Median	:0.6400
Mean	:6.383	Mean	:0.5669
3rd Qu.	:7.980	3rd Qu.	:0.9200
Max.	:9.320	Max.	:1.3600

Valori dello scarto quadratico medio calcolato per ciascun parametro:

$$\sigma_{pH} = 0.2799079$$

$$\sigma_{ECw} = 0.3689049$$

$$\sigma_{Cl} = 0.1943621$$

$$\sigma_{E.S.} = 0.2682197$$

$$\sigma_{Na} = 5.1046600$$

$$\sigma_{Mg} = 2.7765560$$

$$\sigma_{Ca} = 1.6614990$$

$$\sigma_K = 0.0835236$$

$$\sigma_{SO4} = 9.2315680$$

$$\sigma_{Carb.} = 1.9218930$$

$$\sigma_{Bic.} = 0.4257020$$

$$\sigma_{Dur} = 3.5595000$$

$$\sigma_{SAR} = 1.8076060$$

Valori dell'indice di variabilità relativo:

$$V_r^{pH} = 0.4393120$$

$$V_r^{ECw} = 0.4162804$$

$$V_r^{Cl} = 0.4838994$$

$$V_r^{E.S.} = 0.5417018$$

$$V_r^{Na} = 0.5280519$$

$$V_r^{Mg} = 0.5095598$$

$$V_r^{Ca} = 0.4484107$$

$$V_r^K = 0.3478655$$

$$V_r^{Carb} = 0.5536350$$

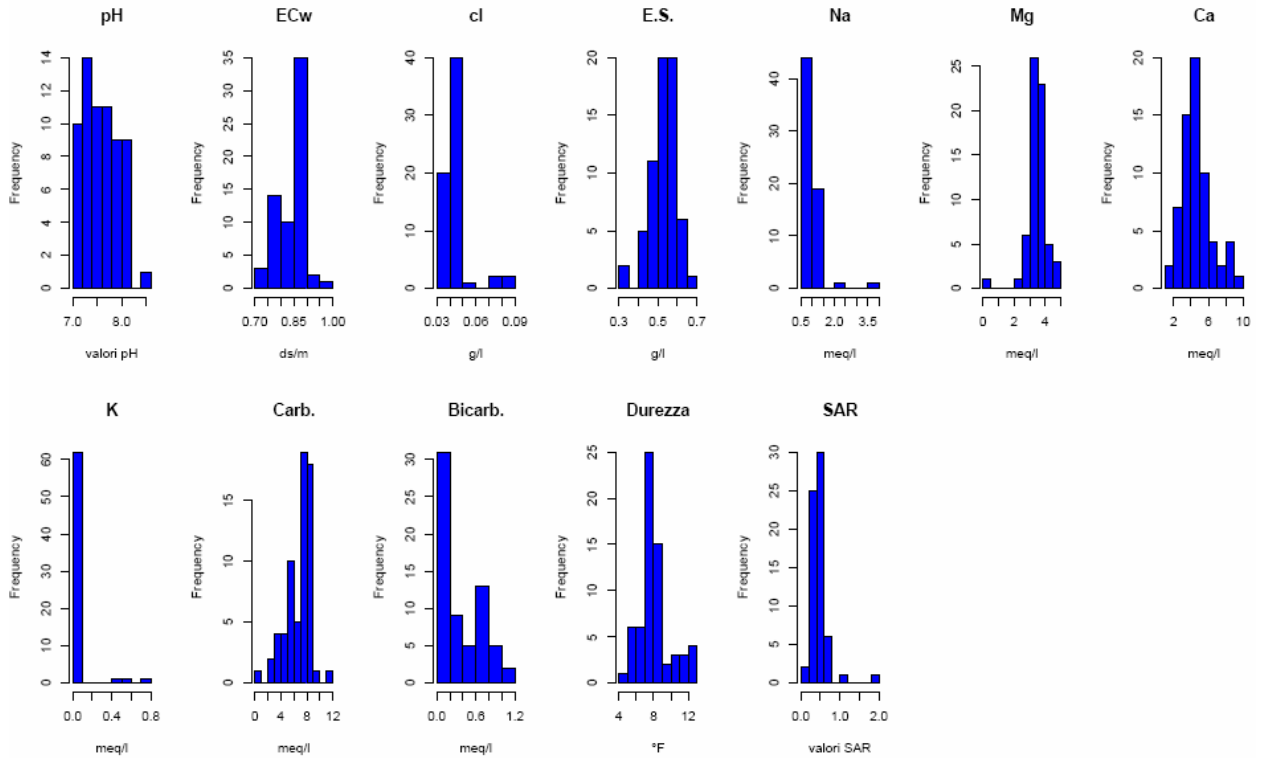
$$V_r^{Bic.} = 0.6348822$$

$$V_r^{Dur} = 0.5786256$$

$$V_r^{SAR} = 0.5154391$$

3.2.4 Sito BA4

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Durezza	SAR
Min. :7.012	Min. :0.7050	Min. : 4.516	Min. :0.0062
1st Qu.:7.333	1st Qu.:0.8000	1st Qu.: 7.214	1st Qu.:0.3662
Median :7.577	Median :0.8600	Median : 7.715	Median :0.4194
Mean :7.578	Mean :0.8456	Mean : 8.107	Mean :0.4624
3rd Qu.:7.854	3rd Qu.:0.8810	3rd Qu.: 8.753	3rd Qu.:0.5308
Max. :8.425	Max. :0.9600	Max. :12.828	Max. :1.8392
Cl	E.S	Na	Mg
Min. :0.03100	Min. :0.3200	Min. :0.521	Min. :0.177
1st Qu.:0.03900	1st Qu.:0.4980	1st Qu.:0.730	1st Qu.:3.184
Median :0.04200	Median :0.5420	Median :0.878	Median :3.479
Mean :0.04427	Mean :0.5325	Mean :0.945	Mean :3.442
3rd Qu.:0.04600	3rd Qu.:0.5680	3rd Qu.:1.052	3rd Qu.:3.718
Max. :0.08800	Max. :0.6580	Max. :3.608	Max. :4.862
Ca	K	NO3	SO4
Min. :1.689	Min. :0.00650	tracce:65	tracce :31
1st Qu.:3.583	1st Qu.:0.06000		0.0186 : 3
Median :4.416	Median :0.06400		0.011 : 2
Mean :4.665	Mean :0.08856		0.0137 : 2
3rd Qu.:5.289	3rd Qu.:0.06900		0.0183 : 2
Max. :9.316	Max. :0.71000		0.0003 : 1
			(Other):24

Carb.		Bic.	
Min.	: 0.360	Min.	:0.0000
1st Qu.	: 5.640	1st Qu.	:0.0000
Median	: 7.480	Median	:0.2800
Mean	: 6.767	Mean	:0.3323
3rd Qu.	: 8.080	3rd Qu.	:0.6800
Max.	:11.360	Max.	:1.1600

Valori dello scarto quadratico medio calcolato per ciascun parametro:

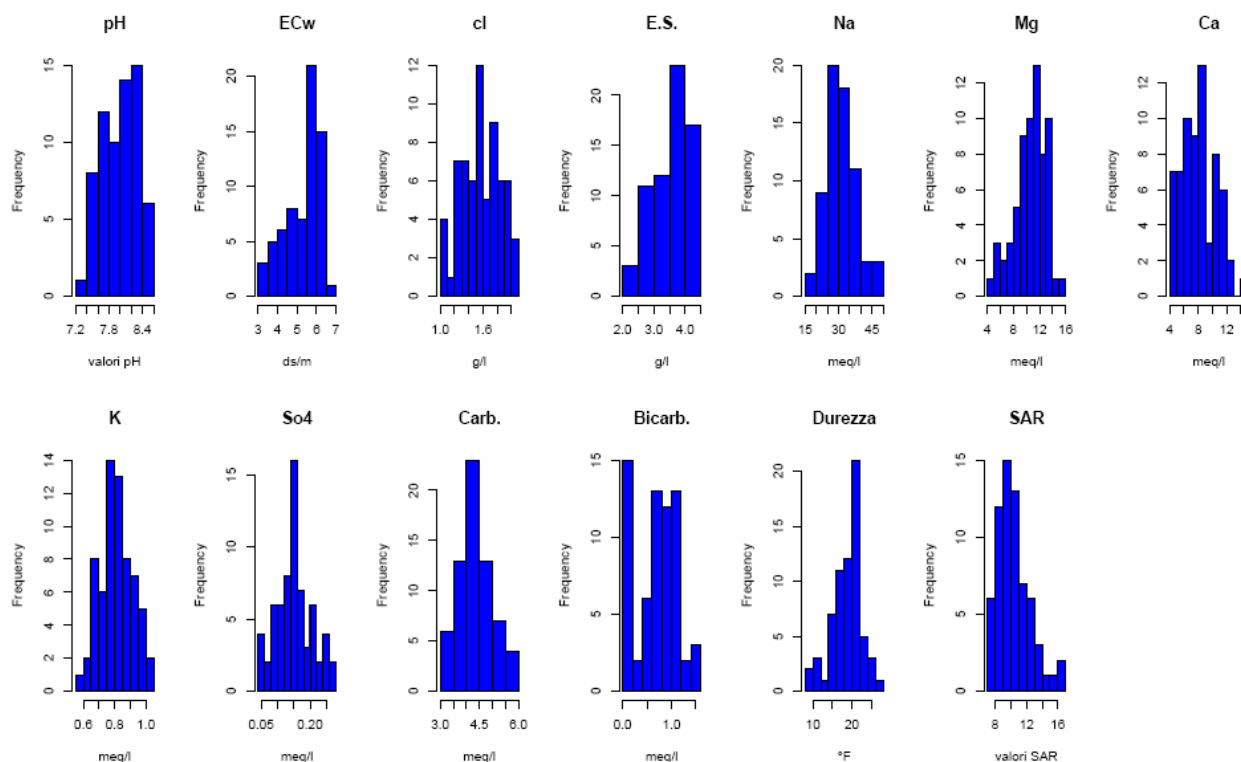
σ_{pH}	= 0.3342574	σ_{ECw}	= 0.0499802
σ_{Cl}	= 0.0102869	$\sigma_{E.S.}$	= 0.0658144
σ_{Na}	= 0.4144249	σ_{Mg}	= 0.6264680
σ_{Ca}	= 1.6396610	σ_K	= 0.1141311
$\sigma_{Carb.}$	= 1.9513060	$\sigma_{Bic.}$	= 0.3544660
σ_{Dur}	= 1.8060170	σ_{SAR}	= 0.2339828

Valori dell'indice di variabilità relativo:

V_r^{pH}	= 0.4827228	V_r^{ECw}	= 0.3940922
V_r^{Cl}	= 0.4270408	$V_r^{E.S.}$	= 0.4030286
V_r^{Na}	= 0.3899817	V_r^{Mg}	= 0.2909573
V_r^{Ca}	= 0.4407199	V_r^K	= 0.5054122
$V_r^{Carb.}$	= 0.3597120	$V_r^{Bic.}$	= 0.6758805
V_r^{Dur}	= 0.4386249	V_r^{SAR}	= 0.2952347

3.2.5 Sito BA5

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Durezza	SAR
Min. :7.260	Min. :3.350	Min. : 8.404	Min. : 7.046
1st Qu.:7.754	1st Qu.:4.645	1st Qu.:16.510	1st Qu.: 8.913
Median :8.040	Median :5.700	Median :19.382	Median :10.049
Mean :7.995	Mean :5.333	Mean :18.697	Mean :10.423
3rd Qu.:8.262	3rd Qu.:5.992	3rd Qu.:20.920	3rd Qu.:11.496
Max. :8.539	Max. :6.580	Max. :26.289	Max. :16.533
Cl	E.S	Na	Mg
Min. :1.028	Min. :2.054	Min. :19.17	Min. : 4.253
1st Qu.:1.347	1st Qu.:3.062	1st Qu.:27.35	1st Qu.: 9.294
Median :1.595	Median :3.661	Median :30.76	Median :11.007
Mean :1.574	Mean :3.540	Mean :31.53	Mean :10.661
3rd Qu.:1.791	3rd Qu.:4.001	3rd Qu.:35.19	3rd Qu.:12.195
Max. :2.092	Max. :4.490	Max. :48.56	Max. :15.054
Ca	K	NO3	SO4
Min. : 4.151	Min. :0.5880	tracce:66	Min. :0.0536
1st Qu.: 6.208	1st Qu.:0.7515		1st Qu.:0.1137
Median : 7.992	Median :0.8160		Median :0.1460
Mean : 8.036	Mean :0.8134		Mean :0.1520
3rd Qu.: 9.909	3rd Qu.:0.8838		3rd Qu.:0.1793
Max. :14.361	Max. :1.0340		Max. :0.2724

Carb.		Bic.	
Min.	:3.160	Min.	:0.0000
1st Qu.	:3.900	1st Qu.	:0.4100
Median	:4.400	Median	:0.7800
Mean	:4.365	Mean	:0.7073
3rd Qu.	:4.720	3rd Qu.	:1.0400
Max.	:5.720	Max.	:1.4800

Valori dello scarto quadratico medio calcolato per ciascun parametro:

$$\sigma_{pH} = 0.3160040$$

$$\sigma_{ECw} = 0.8975364$$

$$\sigma_{Cl} = 0.2747657$$

$$\sigma_{E.S.} = 0.5757998$$

$$\sigma_{Na} = 6.3845900$$

$$\sigma_{Mg} = 2.3669640$$

$$\sigma_{Ca} = 2.3702590$$

$$\sigma_K = 0.1036990$$

$$\sigma_{SO4} = 0.0545253$$

$$\sigma_{Carb.} = 0.6324756$$

$$\sigma_{Bic.} = 0.4400928$$

$$\sigma_{Dur} = 3.7552110$$

$$\sigma_{SAR} = 2.1194690$$

Valori dell'indice di variabilità relativo:

$$V_r^{pH} = 0.4997461$$

$$V_r^{ECw} = 0.5707549$$

$$V_r^{Cl} = 0.5166590$$

$$V_r^{E.S.} = 0.4845826$$

$$V_r^{Na} = 0.4400481$$

$$V_r^{Mg} = 0.4461152$$

$$V_r^{Ca} = 0.4781554$$

$$V_r^K = 0.4650445$$

$$V_r^{SO4} = 0.5009591$$

$$V_r^{Carb} = 0.4949754$$

$$V_r^{Bic.} = 0.5953025$$

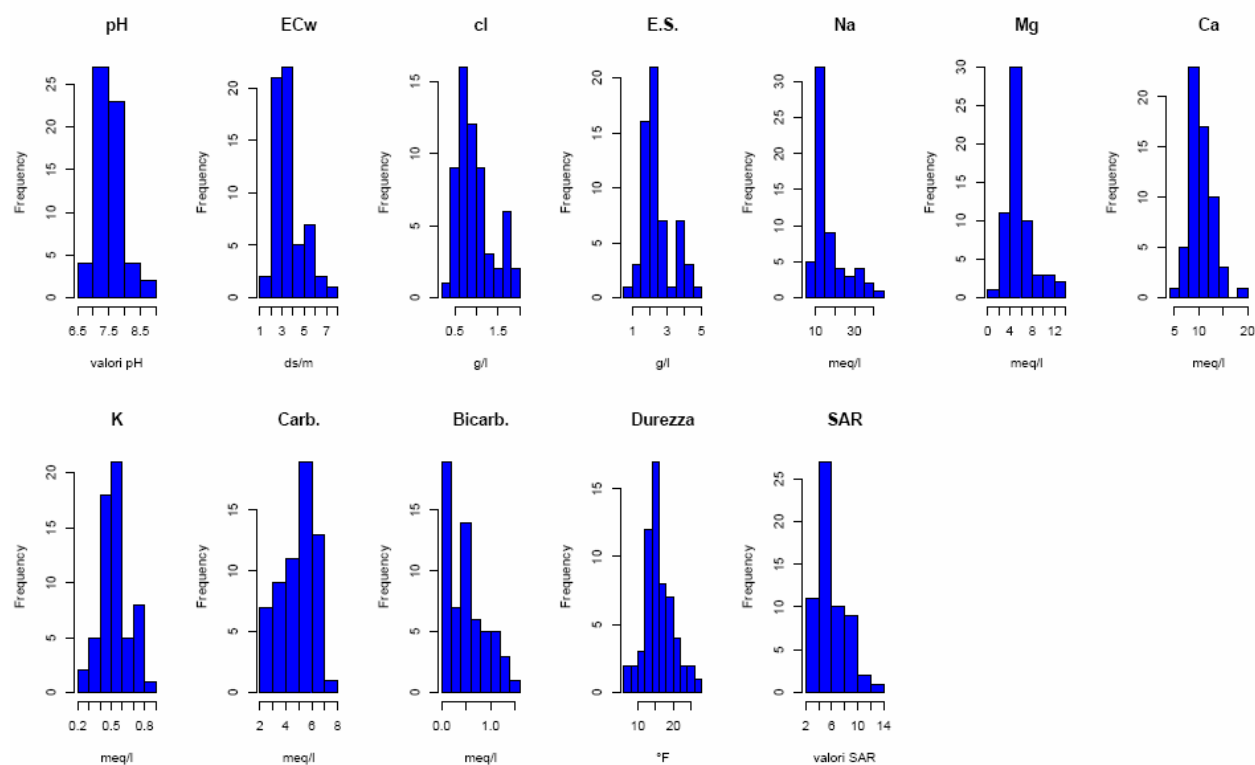
$$V_r^{Dur} = 0.4247997$$

$$V_r^{SAR} = 0.4665820$$

3.3 Idrostruttura della Capitanata

3.3.1 Sito FG1

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Dur	SAR
Min. :6.731	Min. :1.449	Min. : 7.876	Min. : 3.045
1st Qu.:7.326	1st Qu.:2.600	1st Qu.:13.735	1st Qu.: 4.151
Median :7.490	Median :3.180	Median :15.480	Median : 4.837
Mean :7.583	Mean :3.510	Mean :15.956	Mean : 5.776
3rd Qu.:7.814	3rd Qu.:4.018	3rd Qu.:18.547	3rd Qu.: 6.807
Max. :8.716	Max. :7.110	Max. :26.430	Max. :13.695
Cl	E.S	Na	Mg
Min. :0.3190	Min. :0.884	Min. : 6.087	Min. : 1.933
1st Qu.:0.6643	1st Qu.:1.826	1st Qu.:11.384	1st Qu.: 4.436
Median :0.8685	Median :2.173	Median :13.456	Median : 4.959
Mean :0.9616	Mean :2.431	Mean :16.587	Mean : 5.584
3rd Qu.:1.0728	3rd Qu.:2.864	3rd Qu.:17.532	3rd Qu.: 6.433
Max. :1.9140	Max. :4.740	Max. :44.782	Max. :12.504

Ca		K		NO3	SO4
Min.	: 5.359	Min.	:0.2240	tracce:60	tracce : 5
1st Qu.:	8.932	1st Qu.:	0.4545		0.0687 : 3
Median	:10.369	Median	:0.5230		0.0768 : 2
Mean	:10.372	Mean	:0.5304		0.0868 : 2
3rd Qu.:	11.087	3rd Qu.:	0.5880		0.0877 : 2
Max.	:18.912	Max.	:0.8490		0.1053 : 2
					(Other):44

Carb.		Bic.	
Min.	:2.160	Min.	:0.00
1st Qu.:	3.990	1st Qu.:	0.00
Median	:5.220	Median	:0.50
Mean	:4.974	Mean	:0.50
3rd Qu.:	5.940	3rd Qu.:	0.77
Max.	:7.120	Max.	:1.44

Valori dello scarto quadratico medio calcolato per ciascun parametro:

$$\sigma_{pH} = 0.39311$$

$$\sigma_{ECw} = 1.198095$$

$$\sigma_{Cl} = 0.39614$$

$$\sigma_{E.S.} = 0.857316$$

$$\sigma_{Na} = 8.14892$$

$$\sigma_{Mg} = 2.373696$$

$$\sigma_{Ca} = 2.30841$$

$$\sigma_K = 0.124126$$

$$\sigma_{SO4} = 14.95812$$

$$\sigma_{Carb.} = 1.361929$$

$$\sigma_{Bic.} = 0.43235$$

$$\sigma_{Dur} = 3.953243$$

$$\sigma_{SAR} = 2.24799$$

Valori dell'indice di variabilità relativo:

$$V_r^{pH} = 0.4001273$$

$$V_r^{ECw} = 0.4398340$$

$$V_r^{Cl} = 0.5063767$$

$$V_r^{E.S.} = 0.4536195$$

$$V_r^{Na} = 0.4736085$$

$$V_r^{Mg} = 0.4722574$$

$$V_r^{Ca} = 0.3528023$$

$$V_r^K = 0.3972792$$

$$V_r^{Carb} = 0.5542142$$

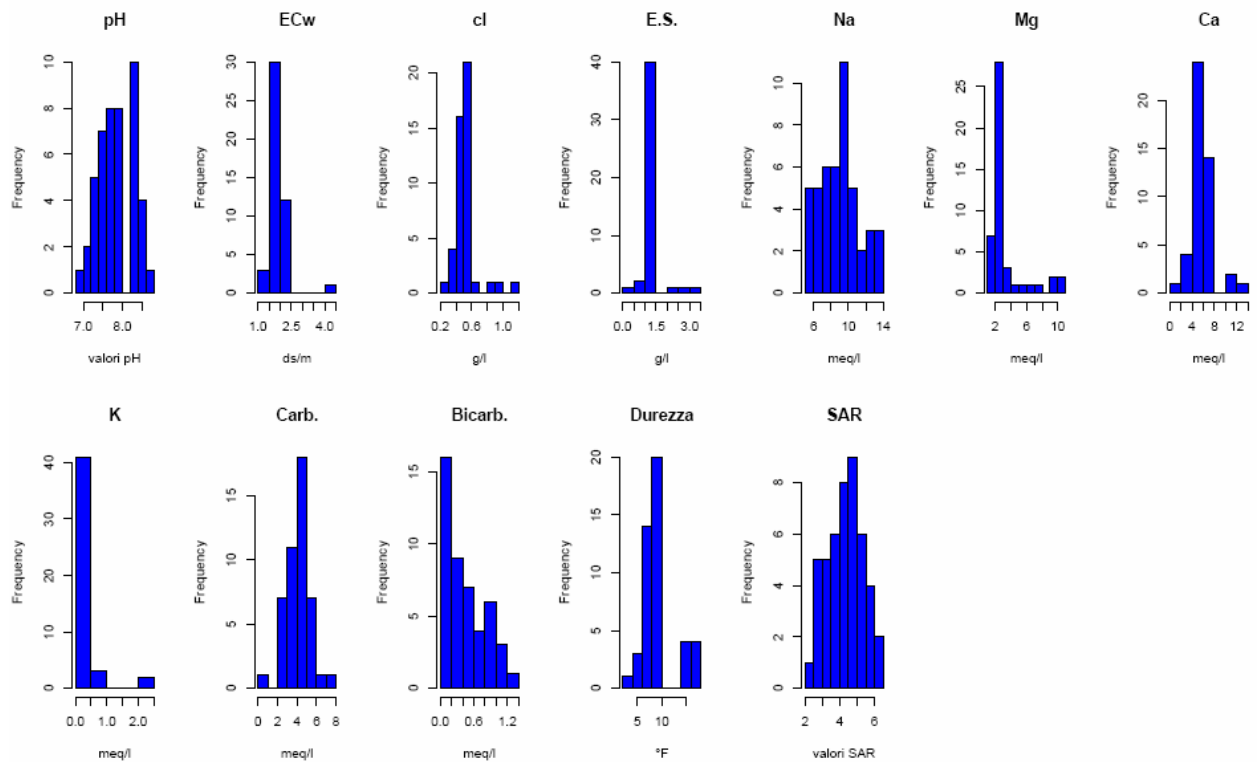
$$V_r^{Bic.} = 0.6306512$$

$$V_r^{Dur} = 0.4297273$$

$$V_r^{SAR} = 0.4834445$$

3.3.2 Sito FG2

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Dur	SAR
Min. :6.985	Min. :1.195	Min. : 3.480	Min. :2.421
1st Qu.:7.539	1st Qu.:1.805	1st Qu.: 7.595	1st Qu.:3.584
Median :7.790	Median :1.918	Median : 8.400	Median :4.335
Mean :7.842	Mean :1.935	Mean : 9.246	Mean :4.305
3rd Qu.:8.292	3rd Qu.:2.027	3rd Qu.: 9.434	3rd Qu.:5.050
Max. :8.671	Max. :4.290	Max. :17.417	Max. :6.472
Cl	E.S	Na	Mg
Min. :0.2130	Min. :0.119	Min. : 5.043	Min. : 1.431
1st Qu.:0.4610	1st Qu.:1.195	1st Qu.: 7.359	1st Qu.: 2.375
Median :0.5050	Median :1.242	Median : 9.065	Median : 2.550
Mean :0.5216	Mean :1.296	Mean : 9.024	Mean : 3.446
3rd Qu.:0.5320	3rd Qu.:1.292	3rd Qu.:10.108	3rd Qu.: 2.825
Max. :1.1700	Max. :3.018	Max. :13.740	Max. :10.612
Ca	K	NO3	SO4
Min. : 0.938	Min. :0.1300	tracce:46	tracce :26
1st Qu.: 4.975	1st Qu.:0.2143		0.0519 : 3
Median : 5.664	Median :0.2575		0.0571 : 2
Mean : 5.801	Mean :0.3734		0.0334 : 1
3rd Qu.: 6.427	3rd Qu.:0.3583		0.0384 : 1
Max. :12.210	Max. :2.2580		0.0406 : 1
			(Other):12

Carb.		Bic.	
Min.	:0.360	Min.	:0.0000
1st Qu.	:3.620	1st Qu.	:0.0000
Median	:4.240	Median	:0.4000
Mean	:4.118	Mean	:0.4235
3rd Qu.	:4.800	3rd Qu.	:0.7300
Max.	:7.120	Max.	:1.2400

Valori dello scarto quadratico medio calcolato per ciascun parametro:

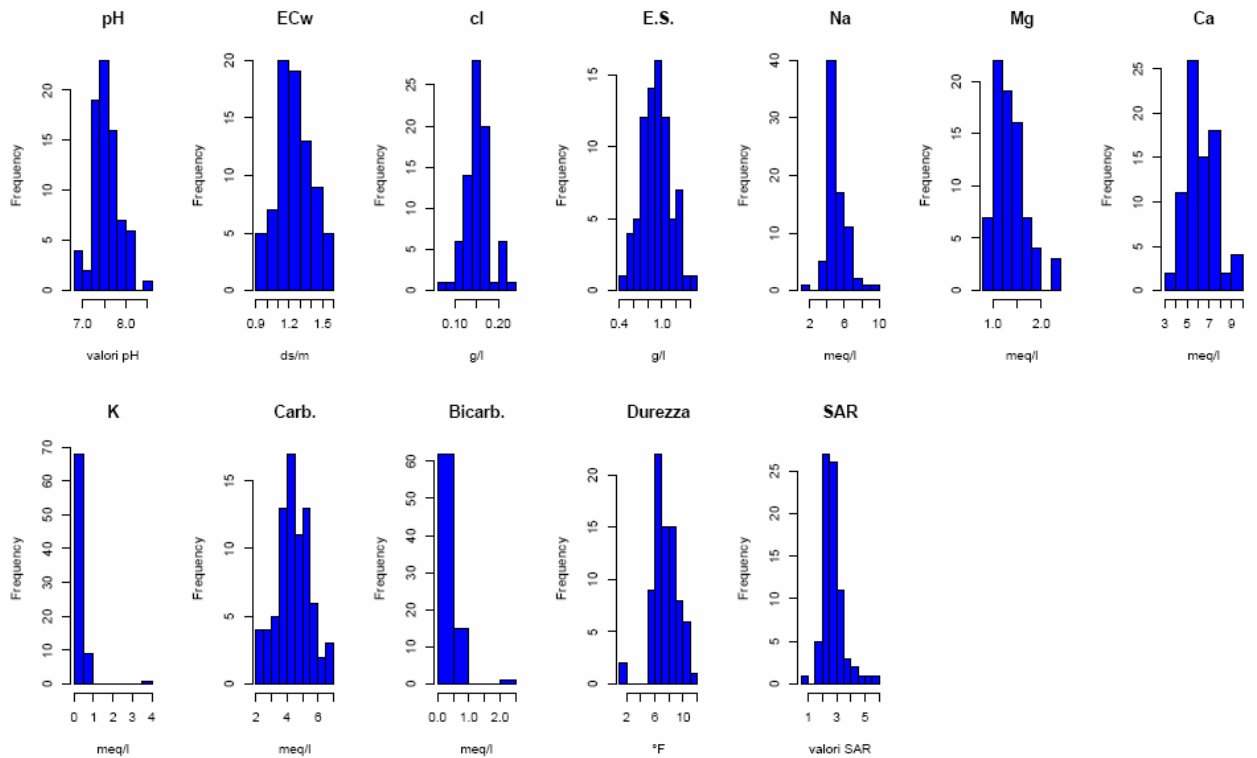
σ_{pH}	= 0.444637	σ_{ECw}	= 0.4070416
σ_{Cl}	= 0.154816	$\sigma_{E.S.}$	= 0.4454824
σ_{Na}	= 2.262398	σ_{Mg}	= 2.3852740
σ_{Ca}	= 1.908844	σ_K	= 0.4114257
$\sigma_{Carb.}$	= 1.196212	$\sigma_{Bic.}$	= 0.3891480
σ_{Dur}	= 3.402078	σ_{SAR}	= 1.0011280

Valori dell'indice di variabilità relativo:

V_r^{pH}	= 0.5275189	V_r^{ECw}	= 0.3083912
V_r^{Cl}	= 0.3460864	$V_r^{E.S.}$	= 0.3129118
V_r^{Na}	= 0.5221350	V_r^{Mg}	= 0.6277221
V_r^{Ca}	= 0.3419229	V_r^K	= 0.6074748
$V_r^{Carb.}$	= 0.3561372	$V_r^{Bic.}$	= 0.6617834
V_r^{Dur}	= 0.4956374	V_r^{SAR}	= 0.4954814

3.3.3 Sito FG3

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Dur	SAR
Min. :6.817	Min. :0.903	Min. : 1.143	Min. :0.8974
1st Qu.:7.353	1st Qu.:1.134	1st Qu.: 6.513	1st Qu.:2.2706
Median :7.484	Median :1.224	Median : 7.314	Median :2.5936
Mean :7.539	Mean :1.243	Mean : 7.490	Mean :2.7194
3rd Qu.:7.753	3rd Qu.:1.330	3rd Qu.: 8.490	3rd Qu.:2.9811
Max. :8.508	Max. :1.593	Max. :11.132	Max. :5.6130
Cl	E.S	Na	Mg
Min. :0.0740	Min. :0.4440	Min. :1.869	Min. :0.872
1st Qu.:0.1370	1st Qu.:0.7935	1st Qu.:4.485	1st Qu.:1.114
Median :0.1475	Median :0.9120	Median :4.795	Median :1.316
Mean :0.1525	Mean :0.9226	Mean :5.110	Mean :1.352
3rd Qu.:0.1683	3rd Qu.:1.0100	3rd Qu.:5.771	3rd Qu.:1.491
Max. :0.2270	Max. :1.4020	Max. :9.174	Max. :2.303
Ca	K	NO3	SO4
Min. :3.832	Min. :0.0190	tracce:78	tracce :19
1st Qu.:5.339	1st Qu.:0.3635		0.0607 : 5
Median :6.003	Median :0.4020		0.0554 : 2
Mean :6.280	Mean :0.4358		0.0571 : 2
3rd Qu.:7.225	3rd Qu.:0.4567		0.0781 : 2
Max. :9.760	Max. :3.6830		0.0849 : 2
			(Other):46

Carb.		Bic.	
Min.	:2.080	Min.	:0.0000
1st Qu.	:3.880	1st Qu.	:0.0000
Median	:4.440	Median	:0.0800
Mean	:4.442	Mean	:0.2564
3rd Qu.	:5.180	3rd Qu.	:0.4400
Max.	:6.920	Max.	:2.4000

Valori dello scarto quadratico medio calcolato per ciascun parametro:

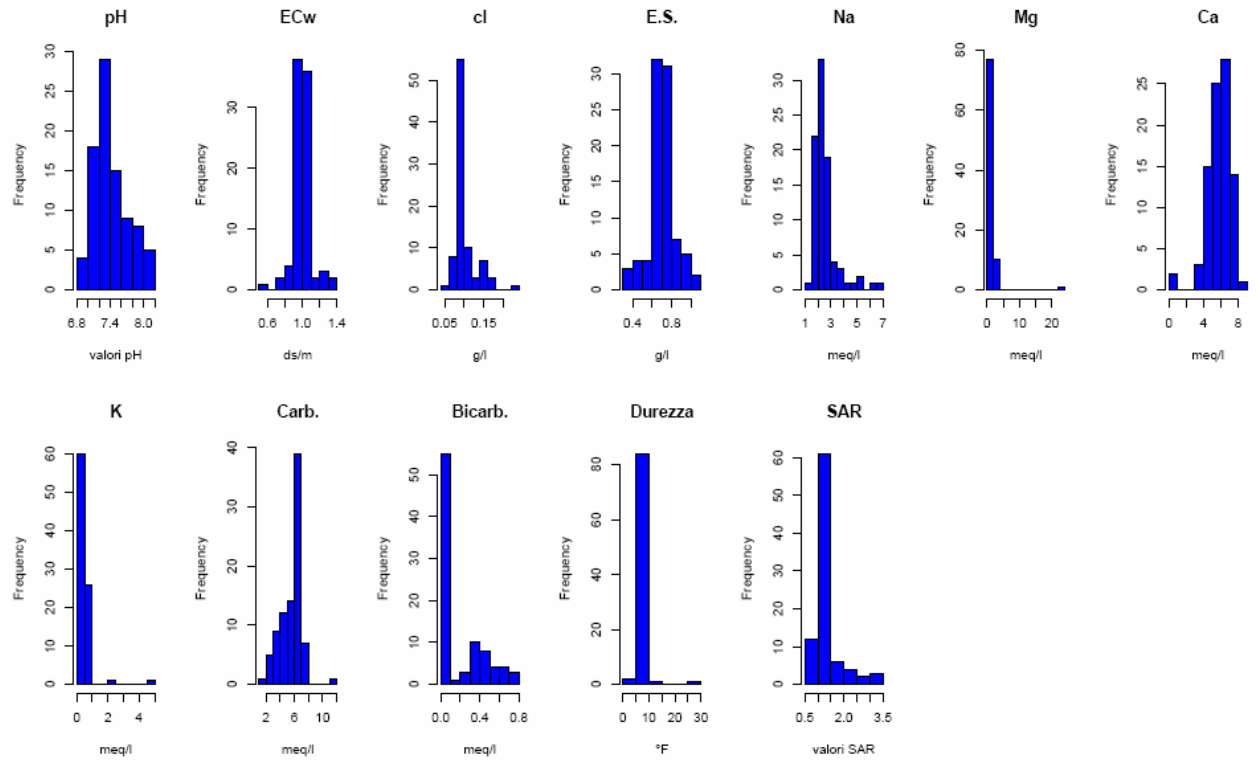
σ_{pH}	= 0.313190	σ_{ECw}	= 0.151174
σ_{Cl}	= 0.028897	$\sigma_{E.S.}$	= 0.203445
σ_{Na}	= 1.111938	σ_{Mg}	= 0.324689
σ_{Ca}	= 1.343662	σ_K	= 0.386142
σ_{SO4}	= 16.58228	$\sigma_{Carb.}$	= 1.067634
$\sigma_{Bic.}$	= 0.363843	σ_{Dur}	= 1.746234
σ_{SAR}	= 0.754929		

Valori dell'indice di variabilità relativo:

V_r^{pH}	= 0.3744406	V_r^{ECw}	= 0.4382298
V_r^{Cl}	= 0.3778615	$V_r^{E.S.}$	= 0.4247287
V_r^{Na}	= 0.3063837	V_r^{Mg}	= 0.4806526
V_r^{Ca}	= 0.4603696	V_r^K	= 0.3319006
$V_r^{Carb.}$	= 0.4412997	$V_r^{Bic.}$	= 0.4907672
V_r^{Dur}	= 0.3631961	V_r^{SAR}	= 0.3287849

3.3.4 Sito FG4

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Dur	SAR
Min. :6.815	Min. :0.5540	Min. : 1.482	Min. :0.745
1st Qu.:7.203	1st Qu.:0.9527	1st Qu.: 6.461	1st Qu.:1.088
Median :7.330	Median :0.9960	Median : 7.406	Median :1.221
Mean :7.418	Mean :1.0116	Mean : 7.589	Mean :1.354
3rd Qu.:7.600	3rd Qu.:1.0690	3rd Qu.: 8.465	3rd Qu.:1.383
Max. :8.200	Max. :1.3400	Max. :28.465	Max. :3.082
Cl	E.S	Na	Mg
Min. :0.04600	Min. :0.3100	Min. :1.404	Min. : 1.045
1st Qu.:0.08500	1st Qu.:0.6450	1st Qu.:2.000	1st Qu.: 1.301
Median :0.09200	Median :0.7050	Median :2.387	Median : 1.406
Mean :0.09954	Mean :0.7031	Mean :2.570	Mean : 1.775
3rd Qu.:0.10600	3rd Qu.:0.7585	3rd Qu.:2.682	3rd Qu.: 1.842
Max. :0.22700	Max. :1.0380	Max. :6.739	Max. :22.293
Ca	K	NO3	SO4
Min. :0.265	Min. :0.2080	tracce:88	tracce :34
1st Qu.:5.107	1st Qu.:0.4365		0.0313 : 2
Median :5.955	Median :0.4780		0.0327 : 2
Mean :5.815	Mean :0.5449		0.0397 : 2
3rd Qu.:6.677	3rd Qu.:0.5122		0.0827 : 2
Max. :8.253	Max. :4.5020		0.0155 : 1
			(Other):45

Carb.		Bic.	
Min.	: 1.760	Min.	:0.0000
1st Qu.	: 4.430	1st Qu.	:0.0000
Median	: 6.460	Median	:0.0000
Mean	: 5.652	Mean	:0.1773
3rd Qu.	: 6.760	3rd Qu.	:0.4000
Max.	:11.040	Max.	:0.8000

Valori dello scarto quadratico medio calcolato per ciascun parametro:

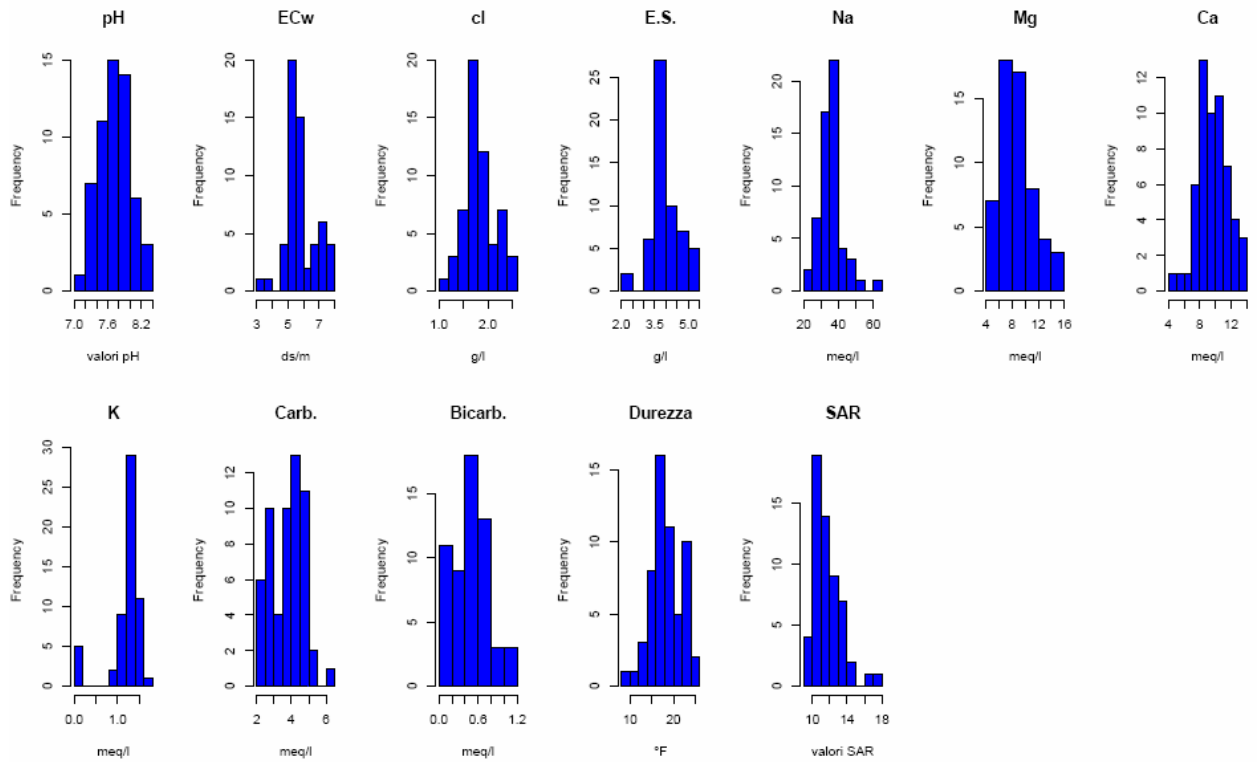
σ_{pH}	= 0.302763	σ_{ECw}	= 0.109842
σ_{Cl}	= 0.027051	$\sigma_{E.S.}$	= 0.133517
σ_{Na}	= 0.922272	σ_{Mg}	= 2.239844
σ_{Ca}	= 1.374227	σ_K	= 0.473337
$\sigma_{Bic.}$	= 0.248419	$\sigma_{Carb.}$	= 1.572117
σ_{SAR}	= 0.487284	σ_{Dur}	= 2.700081

Valori dell'indice di variabilità relativo:

V_r^{pH}	= 0.4408979	V_r^{ECw}	= 0.2833541
V_r^{Cl}	= 0.3274725	$V_r^{E.S.}$	= 0.3679824
V_r^{Na}	= 0.4183041	V_r^{Mg}	= 0.5787995
V_r^{Ca}	= 0.3735742	V_r^K	= 0.4099242
$V_r^{Carb.}$	= 0.3433053	$V_r^{Bic.}$	= 0.7476786
V_r^{Dur}	= 0.2391288	V_r^{SAR}	= 0.4750744

3.3.5 Sito FG5

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Dur	SAR
Min. :7.116	Min. :3.290	Min. : 9.468	Min. : 9.426
1st Qu.:7.538	1st Qu.:5.110	1st Qu.:16.239	1st Qu.:10.528
Median :7.711	Median :5.550	Median :17.943	Median :11.488
Mean :7.733	Mean :5.789	Mean :18.381	Mean :11.741
3rd Qu.:7.937	3rd Qu.:6.400	3rd Qu.:20.928	3rd Qu.:12.506
Max. :8.237	Max. :7.850	Max. :25.169	Max. :17.526
Cl	E.S	Na	Mg
Min. :1.170	Min. :2.200	Min. :22.09	Min. : 4.648
1st Qu.:1.634	1st Qu.:3.610	1st Qu.:31.22	1st Qu.: 6.812
Median :1.741	Median :3.904	Median :35.09	Median : 8.168
Mean :1.832	Mean :3.991	Mean :35.53	Mean : 8.692
3rd Qu.:1.915	3rd Qu.:4.268	3rd Qu.:37.30	3rd Qu.:10.036
Max. :2.517	Max. :5.282	Max. :62.17	Max. :14.478
Ca	K	NO3	SO4
Min. : 4.820	Min. :0.065	tracce:57	0.0510 : 2
1st Qu.: 8.383	1st Qu.:1.166		0.0571 : 2
Median : 9.720	Median :1.269		0.0616 : 2
Mean : 9.689	Mean :1.188		0.0705 : 2
3rd Qu.:10.943	3rd Qu.:1.386		0.0969 : 2
Max. :13.688	Max. :1.645		0.0268 : 1
			(Other):46

	Carb.	Bic.
Min.	:2.200	Min. :0.0000
1st Qu.	:3.000	1st Qu.:0.2800
Median	:4.000	Median :0.4800
Mean	:3.828	Mean :0.4932
3rd Qu.	:4.480	3rd Qu.:0.6800
Max.	:6.360	Max. :1.2000

Valori dello scarto quadratico medio calcolato per ciascun parametro:

$$\sigma_{pH} = 0.268394$$

$$\sigma_{ECw} = 0.9712771$$

$$\sigma_{Cl} = 0.306902$$

$$\sigma_{E.S.} = 0.6139806$$

$$\sigma_{Na} = 6.717796$$

$$\sigma_{Mg} = 2.5322030$$

$$\sigma_{Ca} = 1.914986$$

$$\sigma_K = 0.3798162$$

$$\sigma_{SO4} = 15.121690$$

$$\sigma_{Carb.} = 0.9383235$$

$$\sigma_{Bic.} = 0.293118$$

$$\sigma_{Dur} = 3.4408480$$

$$\sigma_{SAR} = 1.606973$$

Valori dell'indice di variabilità relativo:

$$V_r^{pH} = 0.4812728$$

$$V_r^{ECw} = 0.4279757$$

$$V_r^{Cl} = 0.4557445$$

$$V_r^{E.S.} = 0.4037767$$

$$V_r^{Na} = 0.3549675$$

$$V_r^{Mg} = 0.5234796$$

$$V_r^{Ca} = 0.4339774$$

$$V_r^K = 0.5303094$$

$$V_r^{Carb} = 0.4621844$$

$$V_r^{Bic.} = 0.4964653$$

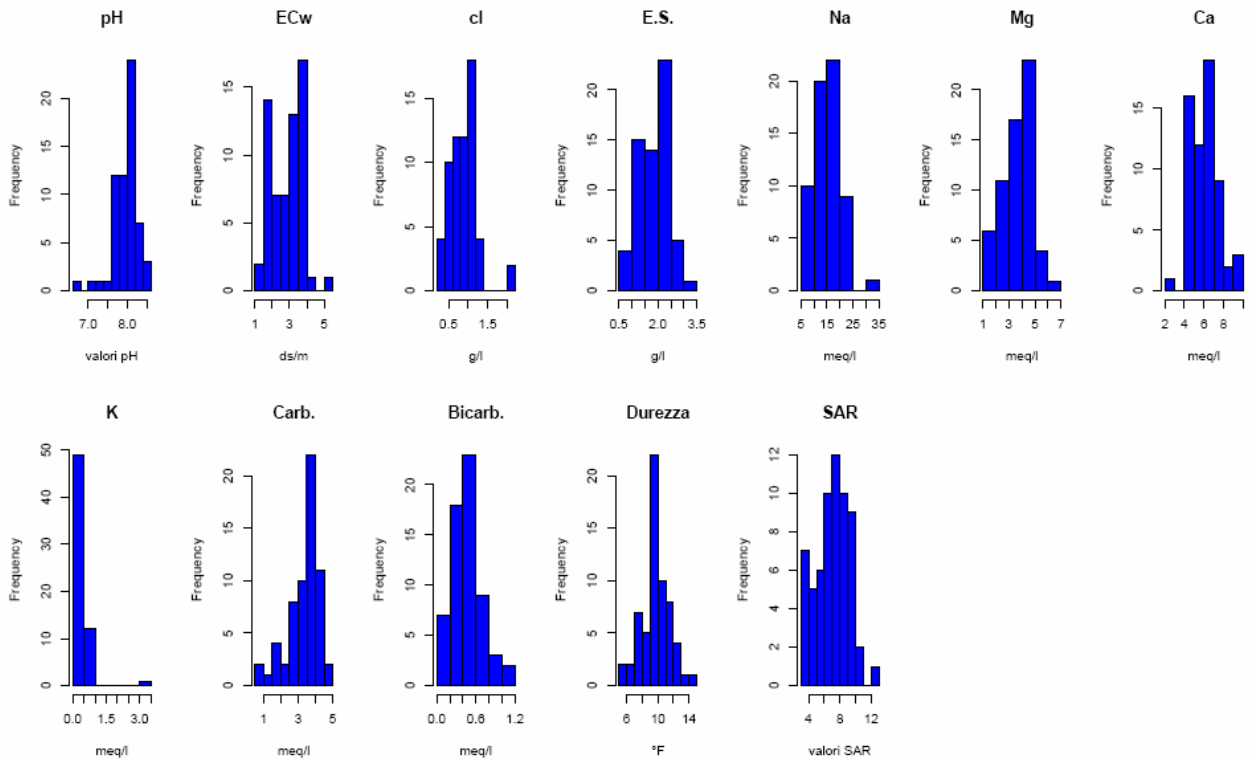
$$V_r^{Dur} = 0.4423665$$

$$V_r^{SAR} = 0.4390972$$

3.4 Idrostruttura del Salento

3.4.1 Sito BR1

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Dur	SAR
Min. :6.787	Min. :1.233	Min. : 5.883	Min. : 3.423
1st Qu.:7.790	1st Qu.:1.986	1st Qu.: 8.929	1st Qu.: 5.814
Median :8.015	Median :3.080	Median : 9.450	Median : 7.347
Mean :7.960	Mean :2.894	Mean : 9.681	Mean : 7.076
3rd Qu.:8.148	3rd Qu.:3.610	3rd Qu.:10.626	3rd Qu.: 8.397
Max. :8.515	Max. :5.370	Max. :14.848	Max. :12.845
Cl	E.S	Na	Mg
Min. :0.2900	Min. :0.842	Min. : 5.87	Min. :1.604
1st Qu.:0.6052	1st Qu.:1.357	1st Qu.:11.36	1st Qu.:2.762
Median :0.8860	Median :1.981	Median :15.65	Median :3.492
Mean :0.8841	Mean :1.845	Mean :15.61	Mean :3.624
3rd Qu.:1.0990	3rd Qu.:2.277	3rd Qu.:19.22	3rd Qu.:4.490
Max. :2.0920	Max. :3.300	Max. :35.00	Max. :6.869

Ca		K		NO3		SO4	
Min.	:2.395	Min.	:0.1280	0.0218:	1	tracce	:18
1st Qu.	:4.971	1st Qu.	:0.3197	0.023	: 1	0.0255	: 2
Median	:6.061	Median	:0.4210	0.0233:	1	0.044	: 2
Mean	:6.058	Mean	:0.4398	0.029	: 1	0.0475	: 2
3rd Qu.	:6.753	3rd Qu.	:0.4890	tracce:	58	0.051	: 2
Max.	:9.685	Max.	:3.4100			0.0734	: 2
						(Other)	:34

Carb.		Bic.	
Min.	:0.520	Min.	:0.0000
1st Qu.	:2.850	1st Qu.	:0.3600
Median	:3.600	Median	:0.4800
Mean	:3.338	Mean	:0.4781
3rd Qu.	:3.950	3rd Qu.	:0.6000
Max.	:4.840	Max.	:1.1200

Valori dello scarto quadratico medio calcolato per ciascun parametro:

$$\sigma_{pH} = 0.302973$$

$$\sigma_{ECw} = 0.883892$$

$$\sigma_{Cl} = 0.356511$$

$$\sigma_{E.S.} = 0.542272$$

$$\sigma_{Na} = 5.330993$$

$$\sigma_{Mg} = 1.104835$$

$$\sigma_{Ca} = 1.385224$$

$$\sigma_K = 0.406973$$

$$\sigma_{Bic.} = 0.245235$$

$$\sigma_{Carb.} = 0.897930$$

$$\sigma_{SAR} = 2.097020$$

$$\sigma_{Dur} = 1.791547$$

Valori dell'indice di variabilità relativo:

$$V_r^{pH} = 0.3754842$$

$$V_r^{ECw} = 0.4358736$$

$$V_r^{Cl} = 0.4208475$$

$$V_r^{E.S.} = 0.4488719$$

$$V_r^{Na} = 0.3879010$$

$$V_r^{Mg} = 0.4315461$$

$$V_r^{Ca} = 0.3800383$$

$$V_r^K = 0.4228749$$

$$V_r^{Carb} = 0.4364569$$

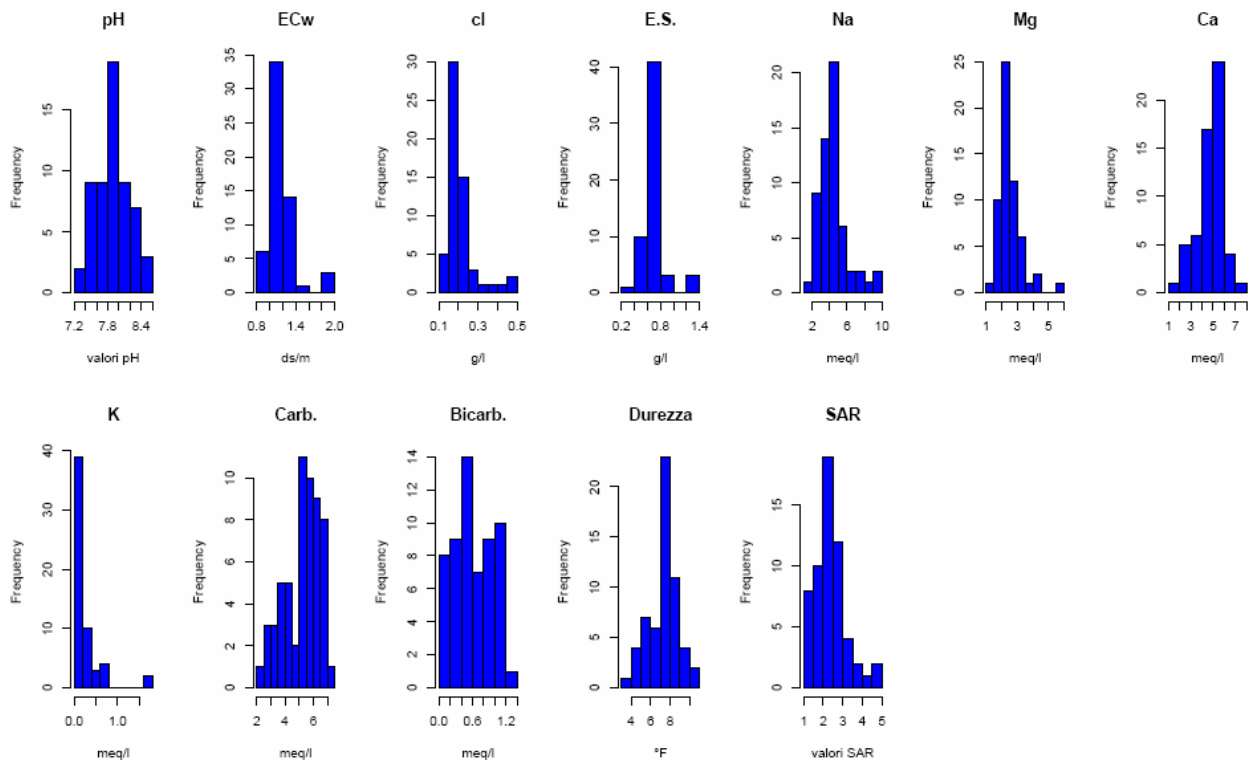
$$V_r^{Bic.} = 0.4426846$$

$$V_r^{Dur} = 0.4044147$$

$$V_r^{SAR} = 0.4567691$$

3.4.2 Sito BR2

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Dur	SAR
Min. :7.323	Min. :0.893	Min. : 3.728	Min. :1.049
1st Qu.:7.716	1st Qu.:1.038	1st Qu.: 6.747	1st Qu.:1.856
Median :7.898	Median :1.111	Median : 7.481	Median :2.277
Mean :7.904	Mean :1.164	Mean : 7.342	Mean :2.361
3rd Qu.:8.077	3rd Qu.:1.242	3rd Qu.: 8.114	3rd Qu.:2.649
Max. :8.580	Max. :1.938	Max. :10.945	Max. :4.645
Cl	E.S	Na	Mg
Min. :0.1340	Min. :0.3440	Min. :1.978	Min. :1.456
1st Qu.:0.1732	1st Qu.:0.6130	1st Qu.:3.631	1st Qu.:2.182
Median :0.1890	Median :0.6870	Median :4.128	Median :2.394
Mean :0.2100	Mean :0.7002	Mean :4.485	Mean :2.553
3rd Qu.:0.2130	3rd Qu.:0.7255	3rd Qu.:4.967	3rd Qu.:2.834
Max. :0.4830	Max. :1.3420	Max. :9.178	Max. :5.890
Ca	K	NO3	SO4
Min. :1.981	Min. :0.1160	tracce:58	tracce :26
1st Qu.:4.465	1st Qu.:0.1610		0.0249 : 2
Median :5.027	Median :0.1735		0.0514 : 2
Mean :4.789	Mean :0.2887		0.0165 : 1
3rd Qu.:5.523	3rd Qu.:0.2350		0.0167 : 1
Max. :7.305	Max. :1.7890		0.0178 : 1
			(Other):25

	Carb.	Bic.
Min.	:2.360	Min. :0.0000
1st Qu.	:4.240	1st Qu.:0.3600
Median	:5.460	Median :0.5800
Mean	:5.194	Mean :0.6234
3rd Qu.	:6.150	3rd Qu.:0.9100
Max.	:7.040	Max. :1.2400

Valori dello scarto quadratico medio calcolato per ciascun parametro:

$$\sigma_{pH} = 0.290577$$

$$\sigma_{ECw} = 0.2187381$$

$$\sigma_{Cl} = 0.073192$$

$$\sigma_{E.S.} = 0.1786032$$

$$\sigma_{Na} = 1.561762$$

$$\sigma_{Mg} = 0.7254819$$

$$\sigma_{Ca} = 1.210580$$

$$\sigma_K = 0.3206839$$

$$\sigma_{Bic.} = 0.361675$$

$$\sigma_{Carb.} = 1.2194330$$

$$\sigma_{SAR} = 0.781302$$

$$\sigma_{Dur} = 1.4823390$$

Valori dell'indice di variabilità relativo:

$$V_r^{pH} = 0.4636811$$

$$V_r^{ECw} = 0.4776154$$

$$V_r^{Cl} = 0.5082165$$

$$V_r^{E.S.} = 0.3735349$$

$$V_r^{Na} = 0.4553131$$

$$V_r^{Mg} = 0.3791738$$

$$V_r^{Ca} = 0.4554487$$

$$V_r^K = 0.6299268$$

$$V_r^{Carb} = 0.5331482$$

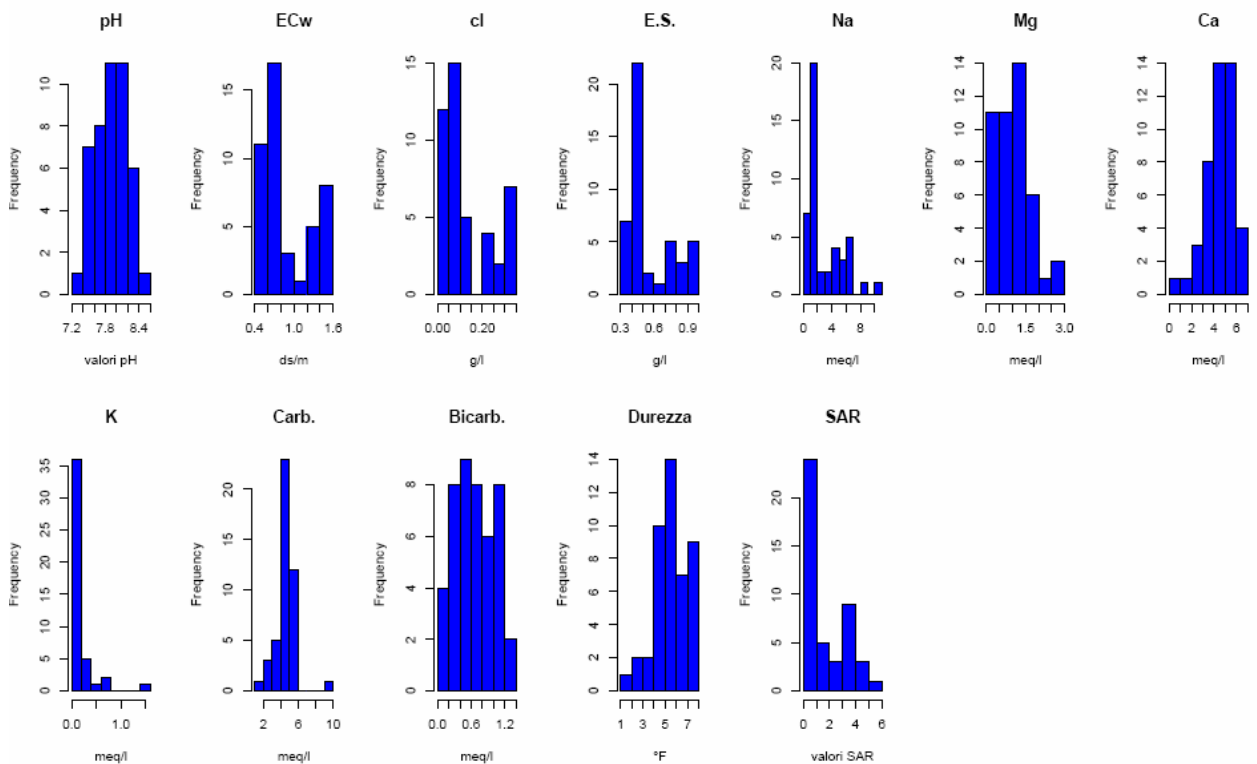
$$V_r^{Bic.} = 0.5833561$$

$$V_r^{Dur} = 0.4107913$$

$$V_r^{SAR} = 0.4513703$$

3.4.3 Sito BR3

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Dur	SAR
Min. :7.322	Min. :0.4850	Min. :1.412	Min. :0.3873
1st Qu.:7.718	1st Qu.:0.6080	1st Qu.:4.902	1st Qu.:0.6739
Median :7.932	Median :0.7050	Median :5.552	Median :0.8713
Mean :7.927	Mean :0.8776	Mean :5.541	Mean :1.7717
3rd Qu.:8.154	3rd Qu.:1.2150	3rd Qu.:6.710	3rd Qu.:3.1356
Max. :8.449	Max. :1.5300	Max. :7.886	Max. :5.3586
Cl	E.S	Na	Mg
Min. :0.0360	Min. :0.3260	Min. : 0.735	Min. :0.316
1st Qu.:0.0490	1st Qu.:0.4120	1st Qu.: 1.126	1st Qu.:0.503
Median :0.0710	Median :0.4340	Median : 1.452	Median :1.028
Mean :0.1315	Mean :0.5514	Mean : 2.911	Mean :1.044
3rd Qu.:0.2200	3rd Qu.:0.7320	3rd Qu.: 4.652	3rd Qu.:1.341
Max. :0.3380	Max. :0.9580	Max. :10.043	Max. :2.813
Ca	K	NO3	SO4
Min. :0.384	Min. :0.0180	0.0520: 1	tracce :18
1st Qu.:3.917	1st Qu.:0.0350	tracce:44	0.0149 : 3
Median :4.665	Median :0.0420		0.0141 : 2
Mean :4.497	Mean :0.1601		0.0100 : 1
3rd Qu.:5.279	3rd Qu.:0.1480		0.0125 : 1
Max. :6.562	Max. :1.4340		0.0130 : 1
			(Other):19

Carb.		Bic.	
Min.	:1.88	Min.	:0.0000
1st Qu.	:4.28	1st Qu.	:0.4000
Median	:4.64	Median	:0.7200
Mean	:4.58	Mean	:0.6693
3rd Qu.	:5.08	3rd Qu.	:0.9600
Max.	:9.52	Max.	:1.2800

Valori dello scarto quadratico medio calcolato per ciascun parametro:

$$\sigma_{pH} = 0.275434$$

$$\sigma_{ECw} = 0.340752$$

$$\sigma_{Cl} = 0.106681$$

$$\sigma_{E.S.} = 0.199681$$

$$\sigma_{Na} = 2.464154$$

$$\sigma_{Mg} = 0.621708$$

$$\sigma_{Ca} = 1.254823$$

$$\sigma_K = 0.254958$$

$$\sigma_{Carb.} = 1.118070$$

$$\sigma_{Bic.} = 0.358116$$

$$\sigma_{Dur} = 1.399779$$

$$\sigma_{SAR} = 1.441206$$

Valori dell'indice di variabilità relativo:

$$V_r^{pH} = 0.4900916$$

$$V_r^{ECw} = 0.6733119$$

$$V_r^{Cl} = 0.7596488$$

$$V_r^{E.S.} = 0.6595799$$

$$V_r^{Na} = 0.6254946$$

$$V_r^{Mg} = 0.5479429$$

$$V_r^{Ca} = 0.4305948$$

$$V_r^K = 0.5991952$$

$$V_r^{Carb} = 0.3061541$$

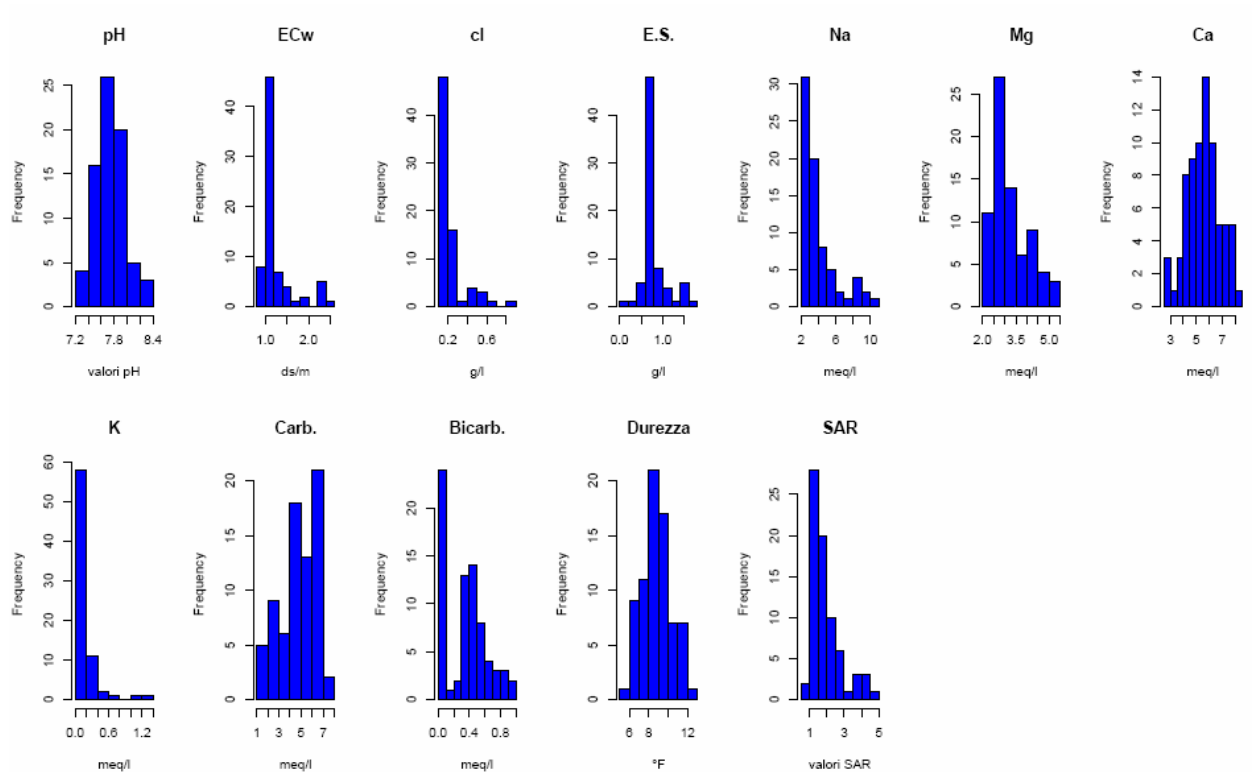
$$V_r^{Bic.} = 0.5601453$$

$$V_r^{Dur} = 0.4498495$$

$$V_r^{SAR} = 0.6467420$$

3.4.4 Sito LE1

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Dur	SAR
Min. :7.240	Min. :0.848	Min. : 5.886	Min. :0.967
1st Qu.:7.595	1st Qu.:1.059	1st Qu.: 7.969	1st Qu.:1.316
Median :7.745	Median :1.133	Median : 8.723	Median :1.652
Mean :7.760	Mean :1.258	Mean : 8.793	Mean :1.904
3rd Qu.:7.904	3rd Qu.:1.252	3rd Qu.: 9.716	3rd Qu.:2.169
Max. :8.372	Max. :2.540	Max. :12.129	Max. :4.592
Cl	E.S	Na	Mg
Min. :0.1200	Min. :0.1700	Min. : 2.087	Min. :2.106
1st Qu.:0.1545	1st Qu.:0.6800	1st Qu.: 2.733	1st Qu.:2.601
Median :0.1635	Median :0.7450	Median : 3.348	Median :2.974
Mean :0.2218	Mean :0.7986	Mean : 3.997	Mean :3.242
3rd Qu.:0.2130	3rd Qu.:0.8145	3rd Qu.: 4.434	3rd Qu.:3.799
Max. :0.8510	Max. :1.6400	Max. :10.173	Max. :5.462
Ca	K	NO3	SO4
Min. :2.629	Min. :0.0690	tracce:74	tracce :42
1st Qu.:4.766	1st Qu.:0.0780		0.0035 : 1
Median :5.631	Median :0.0860		0.0308 : 1
Mean :5.551	Mean :0.1639		0.0361 : 1
3rd Qu.:6.158	3rd Qu.:0.1520		0.0420 : 1
Max. :8.079	Max. :1.2370		0.0422 : 1
			(Other):27

	Carb.	Bic.
Min.	:1.640	Min. :0.0000
1st Qu.	:3.810	1st Qu.:0.0000
Median	:5.000	Median :0.4000
Mean	:4.823	Mean :0.3514
3rd Qu.	:6.270	3rd Qu.:0.5600
Max.	:7.360	Max. :0.9600

Valori dello scarto quadratico medio calcolato per ciascun parametro:

$$\sigma_{pH} = 0.2296298$$

$$\sigma_{ECw} = 0.38079$$

$$\sigma_{Cl} = 0.1392023$$

$$\sigma_{E.S.} = 0.25969$$

$$\sigma_{Na} = 1.9642640$$

$$\sigma_{Mg} = 0.82149$$

$$\sigma_{Ca} = 1.2559040$$

$$\sigma_K = 0.20605$$

$$\sigma_{Bic.} = 0.2870937$$

$$\sigma_{Carb.} = 1.62104$$

$$\sigma_{SAR} = 0.8768861$$

$$\sigma_{Dur} = 1.48658$$

Valori dell'indice di variabilità relativo:

$$V_r^{pH} = 0.4070689$$

$$V_r^{ECw} = 0.5253459$$

$$V_r^{Cl} = 0.5499341$$

$$V_r^{E.S.} = 0.3570911$$

$$V_r^{Na} = 0.5718770$$

$$V_r^{Mg} = 0.5172521$$

$$V_r^{Ca} = 0.4620892$$

$$V_r^K = 0.6455865$$

$$V_r^{Carb} = 0.5704523$$

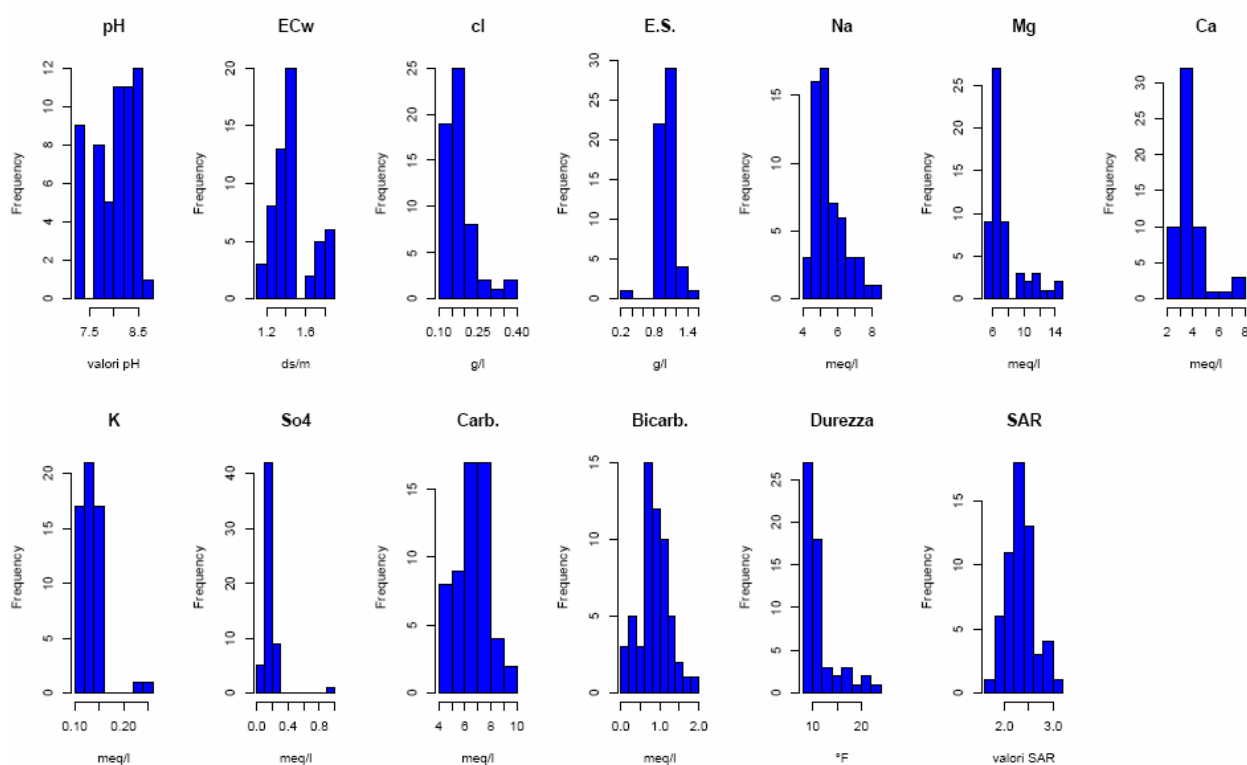
$$V_r^{Bic.} = 0.6208255$$

$$V_r^{Dur} = 0.4773652$$

$$V_r^{SAR} = 0.5524336$$

3.4.5 Sito LE2

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH	ECw	Dur	SAR
Min. :7.285	Min. :1.191	Min. : 8.382	Min. :1.772
1st Qu.:7.785	1st Qu.:1.349	1st Qu.: 9.466	1st Qu.:2.129
Median :8.096	Median :1.425	Median :10.143	Median :2.314
Mean :8.036	Mean :1.454	Mean :11.407	Mean :2.344
3rd Qu.:8.293	3rd Qu.:1.459	3rd Qu.:11.588	3rd Qu.:2.535
Max. :8.608	Max. :1.815	Max. :22.505	Max. :3.197
Cl	E.S	Na	Mg
Min. :0.1110	Min. :0.354	Min. :4.226	Min. : 5.273
1st Qu.:0.1370	1st Qu.:0.936	1st Qu.:4.956	1st Qu.: 6.153
Median :0.1630	Median :1.018	Median :5.347	Median : 6.737
Mean :0.1765	Mean :1.023	Mean :5.526	Mean : 7.572
3rd Qu.:0.1840	3rd Qu.:1.110	3rd Qu.:5.956	3rd Qu.: 7.428
Max. :0.3970	Max. :1.420	Max. :8.261	Max. :14.561
Ca	K	NO3	SO4
Min. :2.535	Min. :0.1010	0.0210: 1	Min. :0.0633
1st Qu.:3.144	1st Qu.:0.1195	0.0237: 1	1st Qu.:0.1109
Median :3.653	Median :0.1301	0.0280: 1	Median :0.1242
Mean :3.835	Mean :0.1336	0.0351: 1	Mean :0.1501
3rd Qu.:4.011	3rd Qu.:0.1430	tracce:53	3rd Qu.:0.1554
Max. :7.944	Max. :0.2487		Max. :0.9440

Carb.		Bic.	
Min.	:4.040	Min.	:0.0000
1st Qu.	:5.710	1st Qu.	:0.6800
Median	:6.840	Median	:0.8800
Mean	:6.554	Mean	:0.8795
3rd Qu.	:7.280	3rd Qu.	:1.1600
Max.	:9.200	Max.	:1.8400

Valori dello scarto quadratico medio calcolato per ciascun parametro:

$$\sigma_{pH} = 0.390296$$

$$\sigma_{ECw} = 0.188164$$

$$\sigma_{Cl} = 0.059737$$

$$\sigma_{E.S.} = 0.148047$$

$$\sigma_{Na} = 0.896965$$

$$\sigma_{Mg} = 2.377898$$

$$\sigma_{Ca} = 1.155214$$

$$\sigma_K = 0.026279$$

$$\sigma_{SO4} = 0.113167$$

$$\sigma_{Carb.} = 1.224512$$

$$\sigma_{Bic.} = 0.391829$$

$$\sigma_{Dur} = 3.371936$$

$$\sigma_{SAR} = 0.310475$$

Valori dell'indice di variabilità relativo:

$$V_r^{pH} = 0.5954817$$

$$V_r^{ECw} = 0.6105308$$

$$V_r^{Cl} = 0.4971901$$

$$V_r^{E.S.} = 0.2872995$$

$$V_r^{Na} = 0.4756733$$

$$V_r^{Mg} = 0.5932395$$

$$V_r^{Ca} = 0.4997817$$

$$V_r^K = 0.4289711$$

$$V_r^{SO4} = 0.4311318$$

$$V_r^{Carb.} = 0.4747749$$

$$V_r^{Bic.} = 0.4263151$$

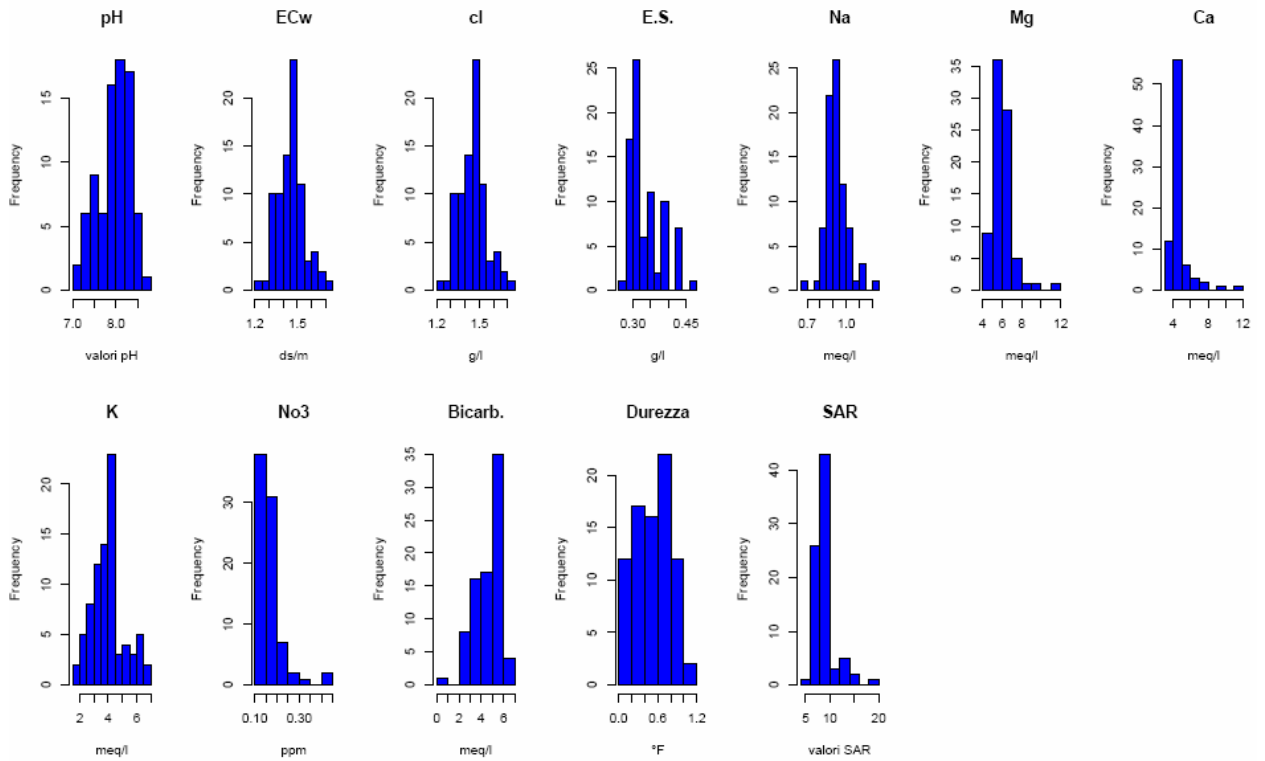
$$V_r^{Dur} = 0.5819310$$

$$V_r^{SAR} = 0.4442663$$

3.5 Idrostruttura Arco Jonico

3.5.1 Sito TA1

Distribuzioni di frequenza per ogni singola caratteristica rilevata:



Risultati del comando “summary”:

pH		ECw		Dur		SAR	
Min.	:7.140	Min.	:1.204	Min.	:0.0000	Min.	: 5.618
1st Qu.	:7.709	1st Qu.	:1.394	1st Qu.	:0.3600	1st Qu.	: 7.801
Median	:8.014	Median	:1.462	Median	:0.5600	Median	: 8.456
Mean	:7.953	Mean	:1.457	Mean	:0.5284	Mean	: 8.767
3rd Qu.	:8.218	3rd Qu.	:1.502	3rd Qu.	:0.7600	3rd Qu.	: 8.972
Max.	:8.635	Max.	:1.712	Max.	:1.2000	Max.	:18.333

Cl		E.S		Na		Mg	
Min.	:1.204	Min.	:0.2760	Min.	:0.6860	Min.	: 4.565
1st Qu.	:1.394	1st Qu.	:0.3050	1st Qu.	:0.8840	1st Qu.	: 5.347
Median	:1.462	Median	:0.3190	Median	:0.9220	Median	: 5.956
Mean	:1.457	Mean	:0.3379	Mean	:0.9297	Mean	: 6.022
3rd Qu.	:1.502	3rd Qu.	:0.3600	3rd Qu.	:0.9540	3rd Qu.	: 6.347
Max.	:1.712	Max.	:0.4640	Max.	:1.2400	Max.	:11.130

Ca		K		NO3		SO4	
Min.	: 3.340	Min.	:1.851	Min.	:0.1240	tracce:81	
1st Qu.:	4.253	1st Qu.:	3.318	1st Qu.:	0.1410		
Median :	4.524	Median :	3.947	Median :	0.1520		
Mean :	4.764	Mean :	4.003	Mean :	0.1684		
3rd Qu.:	4.853	3rd Qu.:	4.426	3rd Qu.:	0.1650		
Max.	:11.681	Max.	:6.652	Max.	:0.4320		

Carb.		Bic.	
tracce :	43	Min.	:0.440
0.0449 :	2	1st Qu.:	3.920
0.0607 :	2	Median :	4.960
0.0616 :	2	Mean :	4.614
0.0010 :	1	3rd Qu.:	5.400
0.0050 :	1	Max.	:6.480
(Other):	30		

Valori dello scarto quadratico medio calcolato per ciascun parametro:

σ_{pH} = 0.368048	σ_{ECw} = 0.09279
σ_{Cl} = 0.092789	$\sigma_{E.S.}$ = 0.04452
σ_{Na} = 0.082650	σ_{Mg} = 1.03104
σ_{Ca} = 1.194491	σ_K = 1.15130
σ_{NO3} = 0.053894	$\sigma_{Bic.}$ = 1.15733
σ_{Dur} = 0.29791	σ_{SAR} = 2.09808

Valori dell'indice di variabilità relativo:

V_r^{pH} = 0.4942598	V_r^{ECw} = 0.3653189
V_r^{Cl} = 0.3653189	$V_r^{E.S.}$ = 0.5039661
V_r^{Na} = 0.3005625	V_r^{Mg} = 0.3779363
V_r^{Ca} = 0.3806176	V_r^K = 0.4822164
V_r^{NO3} = 0.4981072	$V_r^{Bic.}$ = 0.4147112
V_r^{Dur} = 0.5000858	V_r^{SAR} = 0.3822593

3.6 Considerazioni sulla variabilità dei parametri chimico-fisici nelle idrostrutture pugliesi.

Osservando gli istogrammi, si nota che la distribuzione delle frequenze presenta per molti parametri (Na, Bicarbonati, K, Mg) un'asimmetria verso destra, per alcuni (Carbonati) un'asimmetria verso sinistra. In diversi casi, inoltre, le distribuzioni di frequenza risultano bimodali. Sebbene, generalmente, l'analisi delle distribuzioni di frequenza venga approfondita mediante la loro interpolazione tramite una funzione matematica idonea a rappresentare la distribuzione del fenomeno osservato, l'analisi grafica condotta evidenzia che, in alcuni casi, la presenza di classi di valori a frequenza nulla non consente di approssimare correttamente gli istogrammi mediante una curva continua.

I livelli del pH riscontrati nelle acque di falda delle varie idrostrutture presentano un range piuttosto omogeneo compreso tra 6,7 e 8,7 rientrando nei limiti di accettabilità per scopi irrigui.

Il problema legato alla velocità d'infiltrazione dell'acqua nel terreno, tenendo conto dei parametri EC_w e S.A.R., non è rilevabile in nessuna delle falde esaminate.

Le stazioni di BA1, BA3, BA5, FG1, FG5 e LE2 si caratterizzano per la presenza non in tracce di solfati durante tutto (o quasi) il periodo di rilevazione, mentre la falda jonica registra valori non trascurabili di nitrati, sicuramente dovuti agli scarichi delle industrie presenti nel territorio tarantino.

Idrostruttura delle Murge

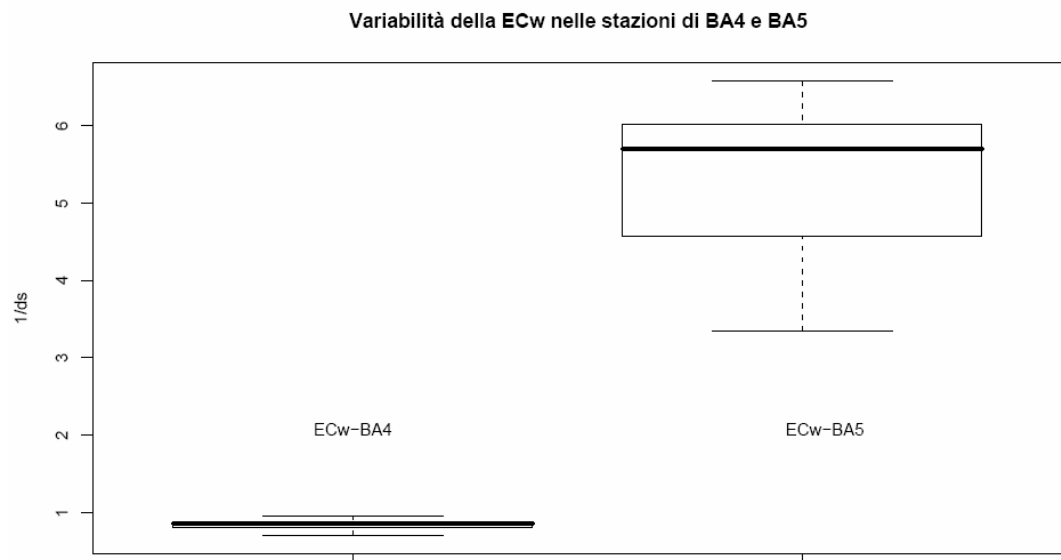
I valori dell'ECw , per i siti BA1, BA3, BA5, superano i limiti di salinità imposti dalla classificazione proposta dalla F.A.O., in base ai quali si consiglia una severa limitazione all'uso irriguo. Gli indici relativi calcolati indicano che, ad eccezione del sito BA4, i parametri ECw, Cl, Na e S.A.R. presentano una variabilità più elevata. Ciò è legato, quasi con certezza, all'effetto esercitato dall'intrusione marina, infatti le falde in questione si situano in prossimità delle zone costiere ed hanno una profondità modesta che va dai 20 metri, per la stazione BA5, ai 50 di BA3.

La profondità della falda è un fattore che influisce sulla variabilità dei parametri chimico-fisici presi in esame.

Nel seguente *boxplot* si confrontano le variabilità della Conducibilità Elettrica, rilevate presso le vicine falde di Rutigliano (BA4) e Mola (BA5) che presentano, rispettivamente, profondità di 200

e 20 metri. Le due falde, inoltre, si situano ad una diversa lontananza dal mare e sono distanti, in linea d'area, appena 10 Km.

I due siti selezionati si caratterizzano per il fatto che presentano una minore (0,394 per BA4) e maggiore (0,571 per BA5) variabilità relativa del parametro di Conducibilità Elettrica.



Dal *boxplot* si evince nettamente che la profondità di falda e la distanza dal mare sono determinanti sulla variabilità e sul valore medio del parametro.

Idrostruttura della Capitanata

I valori dell'ECw, per i siti FG1 ed FG5, superano i limiti di salinità imposti dalla classificazione proposta dalla F.A.O.. L'utilizzo delle acque di queste falde ai fini irrigui deve essere, quindi, limitato.

La ECw presenta un indice di variabilità relativo pressoché omogeneo per gruppi di stazioni. Per i siti FG1, FG3 e FG5 l'indice misura intorno allo 0,4, mentre per FG2 ed FG3 si attesta intorno allo 0,3.

Idrostruttura salentina

Le falde salentine considerate si caratterizzano per una spiccata variabilità dei parametri rilevati, in particolare della ECw e dei Cloruri. Le motivazioni alla base di ciò possono ricondursi all'eccessivo emungimento delle acque di falda per uso non solo irriguo, ma anche abitativo e turistico. I valori massimi di tali parametri si registrano, infatti, proprio nei mesi estivi quando le falde sono sottoposte a maggior stress idrico.

CAPITOLO 4

L'ANALISI CLASSICA DELLE SERIE STORICHE: ASPETTI TEORICI

4.1 Premessa

Fiumi, laghi, ghiacciai, sorgenti e acque del sottosuolo rappresentano solo il 3% delle acque del globo⁸, ma la loro importanza per la nostra sopravvivenza è enorme. Ogni forma di vita dipende, infatti, in larga misura dalle riserve di acqua dolce presenti, in varie forme, nelle aree continentali.

Ad eccezione delle acque provenienti dalle zone profonde della

⁸ NEVIANI I., PIGNOCCHINO C., Geografia generale, Torino, Società Editrice Internazionale, 1996.

litosfera, che vengono liberate per effetto di processi endogeni, tutte le acque che circolano in superficie o nel sottosuolo derivano dalle precipitazioni atmosferiche, che cadono in quantità diverse e con diverse modalità nelle varie regioni della Terra.

Delle acque meteoriche, una frazione viene assorbita dal suolo in quantità maggiore o minore in relazione con la maggiore o minore porosità e permeabilità del suolo; una frazione ritorna nell'atmosfera per evaporazione o per traspirazione, dopo essere stata assorbita e utilizzata dalle radici delle piante; una frazione, infine, scorre sul terreno, come acque di dilavamento prima e di scorrimento poi. Nel complesso quindi sul suolo cade una quantità d'acqua maggiore di quanta ne scorra poi in superficie.

L'acqua nel sottosuolo rappresenta la più importante riserva potenziale di acqua dolce per l'uomo, dal momento che ammonta a più del 90% dell'acqua dolce esistente sulla Terra⁹. Proviene quasi esclusivamente dalle acque meteoriche, che penetrano nel suolo e vengono trattenute negli strati porosi della crosta, mantenendosi comunque sempre sopra il livello del mare.

⁹NEVIANI I., PIGNOCCHINO C., Geografia generale,.....ecc., op.cit

La presenza di acque nel sottosuolo dipende da numerosi fattori che variano da luogo a luogo sulla superficie terrestre. Fra questi:

- la natura e la permeabilità delle rocce e del suolo;
- il clima, che influenza l'intensità dell'evaporazione e il regime di precipitazioni;
- la presenza di vegetazione, che sottrae acqua dal terreno;
- la conformazione topografica del territorio, che può facilitare o impedire uno scorrimento veloce dell'acqua (che implica sempre un minore assorbimento).

La distribuzione delle acque sotterranee non è, quindi, omogenea e la quantità d'acqua profonda può variare anche in uno stesso luogo con il passare del tempo. Le riserve d'acqua nel sottosuolo sono influenzate anche dalle perdite che inevitabilmente si verificano: in parte l'acqua riemerge alimentando corsi d'acqua o finendo direttamente nel mare, in parte viene prelevata dall'uomo per soddisfare il suo fabbisogno idrico.

Occorre considerare, inoltre, che l'acqua è un ottimo solvente per tutte le sostanze polari, produce una grande varietà di miscugli e facilita lo svolgimento di molte reazioni chimiche. Per questo l'acqua è il più importante agente esogeno coinvolto nel modellamento della superficie terrestre: partecipa ai processi di degradazione delle rocce e, grazie alla forza meccanica che possiede, quando è in movimento, fornisce un

mezzo di trasporto per la gran quantità di sostanze disciolte e per i detriti.

Nella precedente analisi si è constatato, infatti, che l'acqua delle falde non è veramente pura, poiché contiene in soluzione ioni, sostanze organiche e inorganiche, asportate dalle rocce, che conferiscono alle acque del sottosuolo caratteristiche chimiche specifiche.

A seguito di queste considerazioni ci si chiede, quindi, come e quanto l'evento meteorico possa influenzare i parametri chimico-fisici delle acque di falda monitorate.

In particolare, attraverso la metodologia statistica, si individueranno le eventuali relazioni di dipendenza tra le altezze di pioggia e i rilevamenti della Conducibilità Elettrica che, come già visto, rappresentano uno dei più importanti indicatori sullo stato di qualità ad uso irriguo delle acque di falda.

Prima di effettuare l'analisi è opportuno soffermarsi su alcuni aspetti teorici relativi alla metodologia statistica che sarà applicata in seguito.

4.2 Il coefficiente di correlazione lineare di Bravais-Pearson

La misura della relazione esistente fra due variabili viene comunemente definita, in Statistica, correlazione. Fra due caratteri X e Y, questa potrebbe essere valutata attraverso la covarianza:

$$Cov(X, Y) = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{n}$$

Tale misura è idonea a porre in luce l'esistenza di una relazione lineare fra due caratteri¹⁰, nel senso che:

- se $Cov(X, Y) < 0$ esiste discordanza (correlazione negativa);
- se $Cov(X, Y) = 0$ c'è assenza di relazione lineare o indipendenza;
- se $Cov(X, Y) > 0$ esiste concordanza (correlazione positiva).

Tuttavia, tale indice presenta due difetti: l'uno è quello di essere vincolato alle unità di misura dei due caratteri, l'altro consiste nel fatto che aumentando o diminuendo il numero n delle coppie dei valori si può far variare anche la $Cov(X, Y)$.

¹⁰ DELVECCHIO F., Statistica per la ricerca sociale, Cacucci, Bari, 2002.

Detta grandezza non ha, infatti, né un limite superiore né uno inferiore. Per ovviare a quest'inconveniente, si rapporta la $Cov(X, Y)$ al suo massimo e si ottiene così il coefficiente di correlazione di Bravais-Pearson:

$$r = \frac{Cov(X, Y)}{MaxCov(X, Y)} = \frac{Cov(X, Y)}{\sigma(X)\sigma(Y)} = \frac{Cov(X, Y)}{\sqrt{Dev(X) \cdot Dev(Y)}}$$

con $-1 < r < 1$.

Il largo uso del coefficiente di correlazione, almeno per scopi descrittivi, è legato al fatto che esso è un numero puro o invariante (non dipende, in pratica, né dalle unità di misura con cui sono espresse X e Y né dall'origine), perciò, se occorre, si possono anche trasformare linearmente i dati in quanto il valore di r non muta.

4.3 Funzione di correlazione incrociata

La funzione di correlazione incrociata tra due processi stocastici, y_t e x_t , è definita come:

$$\lambda_{xy}(t, t-h) = corr\{y_t, x_{t-h}\}$$

ovvero, misura la dipendenza lineare esistente tra i due processi a vari istanti di tempo. Sotto un'ipotesi di 'stazionarietà congiunta' dipende solo da h e può essere stimata dai coefficienti campionari di covarianza incrociati a ritardo h:

$$\frac{1}{n} \sum_{i=h+1}^n (y_i - \bar{y}) (x_{i-h} - \bar{x})$$

È bene precisare che, soltanto nell'ipotesi in cui le due serie storiche in esame siano generate da un processo stocastico puramente aleatorio (*white noise*), è possibile verificare la significatività dei valori della funzione di correlazione incrociata¹¹.

A tale fine, è necessario eseguire, in genere, un'operazione di filtraggio di ciascuna serie, allo scopo di trasformarla in una successione di residui, per i quali siano accettabili le ipotesi poste. Il criterio solitamente impiegato è quello di adattare ai valori di ogni serie un opportuno modello univariato.

¹¹ ZANI S., Osservazioni sulle serie storiche multiple e l'analisi dei gruppi, Relazione al Convegno Nazionale di studio sull' "Analisi Moderna delle Serie Storiche" (a cura di Piccolo D.), Franco Angeli Editore, Napoli, 19-22 maggio 1981.

4.4 L'analisi delle serie storiche

Per comprendere cosa sia l'analisi delle serie storiche, conviene anzitutto precisarne gli scopi.

L'analisi statistica di una serie storica si propone di chiarire il meccanismo casuale che l'ha generata, o per dare una descrizione succinta delle caratteristiche della serie, oppure per prevedere l'evoluzione del fenomeno osservato, di cui è nota la storia passata. Volendo schematizzare, possiamo attribuire all'analisi delle serie storiche i seguenti obiettivi.

- **Descrizione:** il primo obiettivo è dato dalla descrizione sintetica dell'andamento del fenomeno. Uno strumento adatto a tale scopo è senza dubbio il grafico della serie rispetto al tempo. Nel grafico di una serie, infatti, sono spesso chiaramente visibili andamenti regolari nella dinamica temporale del fenomeno. Sempre con l'ausilio del grafico è possibile individuare valori anomali e/o eterogenei (outlier).
- **Spiegazione:** un altro obiettivo dell'analisi delle serie storiche è dato dalla spiegazione del fenomeno. Si intende, cioè, individuare il

meccanismo generatore della serie e, eventualmente, le relazioni che legano la variabile sotto studio ad altri fenomeni.

- **Previsione:** nell'analisi delle serie temporali assume notevole importanza il problema della previsione, ossia dell'inferenza su valori futuri del fenomeno d'interesse in base alla sua storia passata. Le problematiche previsive sono strettamente collegate ai problemi di filtraggio e di controllo, considerati ai punti successivi.
- **Filtraggio:** spesso si desidera usare i dati di una serie storica per stimare componenti non osservabili della serie stessa. Le tecniche di filtraggio servono appunto a questo scopo, secondo procedure molto simili a quelle proprie delle previsioni.
- **Controllo:** un altro obiettivo che è possibile perseguire con l'analisi delle serie storiche è il controllo di un processo produttivo. Di fatto, ciò avviene studiando la dinamica temporale di uno o più fenomeni, eventualmente correlati, legati alle caratteristiche di qualità del processo.

4.5 Approccio classico o moderno

Un modello stocastico abbastanza generale per descrivere il processo generatore dei dati di una serie storica $\{y_t\}_{t=1}^n$ relativa ad una variabile Y è dato da:

$$Y_t = f(t) + u_t \quad (1)$$

Nella (1) si assume che la serie osservata sia il risultato della composizione di:

- a) una sequenza completamente deterministica, $\{f(t)\}$, che costituisce la parte sistematica della serie;
- b) una sequenza di variabili casuali, $\{u_t\}$, che rappresenta la parte stocastica della serie ed obbedisce ad una determinata legge di probabilità.

Nel trattamento del modello (1) secondo l'approccio all'analisi delle serie storiche detto *classico* (o *tradizionale*) si suppone che esista una "legge di evoluzione temporale" del fenomeno, rappresentata da $f(t)$. La componente casuale u_t viene invece assunta a rappresentare l'insieme delle circostanze, ciascuna di entità trascurabile, che non si vogliono o non si possono considerare esplicitamente in Y_t . I *residui* di Y_t , non

spiegati da $f(t)$, vengono, pertanto, imputati al caso ed assimilati ad *errori accidentali*. Da un punto di vista statistico, ciò equivale ad ipotizzare che la componente stocastica del modello (1) sia generata da un processo *white noise*, ossia da una successione di variabili casuali indipendenti, di media nulla e varianza costante. Per tale processo, sinteticamente indicato con la notazione $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, si ha:

$$E[\varepsilon_t] = 0 \quad \forall t$$

$$Var[\varepsilon_t] = \sigma_\varepsilon^2 \quad \forall t$$

$$E[\varepsilon_r \varepsilon_s] = 0 \quad \forall r, s, \quad r \neq s.$$

Una successione di v.c. $\{u_t\}$ per cui si abbia $Cov[u_r, u_s] = 0, \quad \forall r, s, \quad r \neq s$, indipendentemente dal fatto che sia o meno un white noise, viene detta *processo stocastico a componenti incorrelate*. Per contro, quando $Cov[u_r, u_s] \neq 0$ per qualche $r \neq s$, si ha un *processo a componenti correlate*.

In sintesi, nell'approccio classico l'attenzione viene concentrata su $f(t)$; u_t , essendo considerato un processo a componenti incorrelate, è dunque trascurabile. Nell'approccio *moderno* si ipotizza, invece, che $f(t)$ manchi o sia già stata eliminata (mediante stima o altri metodi).

L'attenzione viene posta quindi sulla componente stocastica u_t , che si ipotizza essere un processo a componenti correlate del tipo

$$u_t = g(Y_{t-1}, Y_{t-2}, \dots, \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) + \varepsilon_t,$$

che va trattato con opportune tecniche statistiche¹².

4.5.1 Le componenti di una serie storica

Le componenti di una serie storica di solito sono le seguenti:

- il trend (o componente tendenziale) rappresenta l'andamento di lungo periodo, noto come tendenza di fondo, del fenomeno allo studio. Tale componente è caratterizzata da un comportamento o monotonicamente o unimodale nel periodo di osservazione¹³;
- il ciclo rappresenta le fluttuazioni di medio periodo della serie storica. Tale componente si presenta tipicamente come una oscillazione irregolare della durata di alcuni anni. L'ampiezza delle fluttuazioni cicliche può essere costante o più frequentemente variabile: si parlerà in

¹² PICCOLO D., Introduzione all'analisi delle serie storiche, Roma, La Nuova Italia Scientifica, 1990.

¹³ FALIVA M., L'analisi delle serie storiche nel dominio delle frequenze, Relazione al Convegno Nazionale di studio sull' "Analisi Moderna delle Serie Storiche" (a cura di Piccolo D.), Napoli, 19-22 maggio 1981, Franco Angeli, Milano, 1983.

quest'ultimo caso di ciclo evolutivo. Il ciclo è tipico delle serie storiche che descrivono fenomeni economici in un periodo di osservazione piuttosto lungo¹⁴;

- la stagionalità rappresenta le fluttuazioni annuali della serie. Tale componente si presenta tipicamente come una oscillazione sostanzialmente regolare della durata di un anno. L'ampiezza delle fluttuazioni stagionali può essere costante o, più frequentemente, variabile: si parlerà in quest'ultimo caso di stagionalità evolutiva¹⁵;
- la componente erratica rappresenta le oscillazioni accidentali della serie storica.

4.5.2 Modelli di combinazione delle componenti

I più semplici modelli di combinazione delle componenti sono:

- a. il modello additivo:

$$Y_t = T_t + C_t + S_t + \varepsilon_t,$$

- b. il modello moltiplicativo:

$$Y_t = T_t \cdot C_t \cdot S_t \cdot \varepsilon_t,$$

¹⁴ RICCI V., Analisi delle serie storiche con R, 2004.

¹⁵ FALIVA M., L'analisi delle serie storiche nel dominio delle frequenze,ecc., op. cit..

c. il modello misto:

$$Y_t = T_t \cdot C_t \cdot S_t + \varepsilon_t,$$

dove si è indicato con T_t il valore del trend al tempo t , con C_t il valore del ciclo, con S_t il valore della stagionalità e con ε_t la componente accidentale.

Il modello moltiplicativo può ricondursi a quello additivo mediante trasformazione logaritmica (a condizione che le componenti siano sempre positive):

$$\log Y_t = \log T_t + \log C_t + \log S_t + \log \varepsilon_t.$$

Si noti che nel modello additivo le quattro componenti sono tutte espresse nella stessa unità di misura di Y_t . Nel modello moltiplicativo una sola componente, generalmente T_t , è espressa nella medesima unità di misura di Y_t , mentre le restanti sono espresse sotto forma di numeri puri che ne esprimono l'incidenza relativa al tempo t . Inoltre, nel modello moltiplicativo, le ipotesi sulla componente di disturbo vanno riferite a $\log \varepsilon_t$ e non a ε_t ¹⁶. Nel modello misto, infine, le componenti T_t e ε_t sono espresse nella stessa metrica di Y_t , mentre C_t e S_t assumono la forma di numeri indici.

¹⁶ DI FONZO T., LISI F., Complementi di statistica economica, Cleup Editrice, Padova, 2001.

Inoltre, nelle analisi congiunturali la componente tendenziale e quella ciclica vengono spesso considerate congiuntamente. La componente risultante viene denominata trend-ciclo ed il suo valore al tempo t viene indicato con TC_t .

4.6 Richiami all'analisi di regressione

Molte delle tecniche statistiche utilizzate per analizzare una serie storica sono applicazioni dirette oppure adattate dell'analisi di regressione nella sua formula più semplice, la teoria classica dei minimi quadrati. In questo paragrafo si espongono alcuni dei risultati dell'analisi di regressione che saranno poi utilizzati per l'analisi del trend.

Si consideri il seguente modello di regressione:

$$y_t = f(x_t; \beta) + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (2)$$

dove $f(x_t; \beta)$ è una funzione matematica di p variabili esplicative

$x_t = (x_{t_1}, x_{t_2}, \dots, x_{t_p})'$ e di m parametri ignoti $\beta = (\beta_1, \beta_2, \dots, \beta_m)'$,

mentre ε_t è un termine di disturbo casuale generato da un processo

white-noise, $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$. Si suppone, inoltre, che il vettore x_t sia

non stocastico, cioè è composto da valori dati. In altre parole, ciò significa che x_t è considerato fisso in campioni ripetuti. Le x_{t_i} sono dette *variabili indipendenti*, mentre y_t è detta *variabile dipendente*.

Dalle proprietà del modello (2) segue che, per t fissato, anche y_t è una variabile casuale, con media e varianza date rispettivamente da:

$$E[y_t] = f(x_t; \beta), \quad \text{Var}[y_t] = \sigma_\varepsilon^2, \quad t = 1, 2, \dots, n.$$

Una volta note le osservazioni sulle variabili del modello (2), il problema di stimare il vettore β dei parametri si risolve mediante la scelta di un vettore $\hat{\beta}$ tale che la funzione $f(x_t; \hat{\beta})$ sia “vicina” a y_t . Il criterio di “vicinanza” usualmente adottato si basa sulla distanza $\{y_t - f(x_t; \beta)\}^2$. Il vettore $\hat{\beta}$ che minimizza la funzione:

$$S(\beta) = \sum_{t=1}^n \{y_t - f(x_t; \beta)\}^2 \quad (3)$$

è chiamato stimatore dei minimi quadrati di β ¹⁷. Se come generalmente avviene, $S(\beta)$ è differenziabile, $\hat{\beta}$ si ottiene risolvendo il sistema di equazioni normali:

$$\frac{\delta S(\beta)}{\delta \beta} = 0.$$

¹⁷ DI FONZO T., LISI F., Complementi di statistica economica,ecc., op.cit..

4.6.1 Il modello di regressione lineare e la stima dei minimi quadrati

Si consideri il seguente modello:

$$y_t = \beta_1 x_{t_1} + \beta_2 x_{t_2} + \dots + \beta_p x_{t_p} + \varepsilon_t, \quad t = 1, 2, \dots, n, \quad (4)$$

dove $x_{t_1} = 1$, per $t = 1, 2, \dots, n$, se nel modello è presente l'intercetta. Il modello (4) in notazione matriciale diventa:

$$y = X\beta + \varepsilon, \quad (5)$$

con

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}.$$

Per il modello (5) si assume¹⁸ quanto segue:

- a. X è una matrice non stocastica;
- b. $\text{rango}(X) = p < n$;
- c. i termini di disturbo hanno media nulla, varianza costante e finita e sono incorrelati, ossia $E[\varepsilon] = 0$ e $E[\varepsilon\varepsilon'] = \sigma_\varepsilon^2 I_n$.

¹⁸ AZZALINI A., *Inferenza statistica*, Springer, Milano, 2001.

La funzione definita in (3) diventa quindi:

$$S(\beta) = (y - X\beta)'(y - X\beta) = y'y - 2y'X\beta + \beta'X'X\beta$$

ed il relativo sistema di equazioni normali ha la forma:

$$\frac{\delta S(\beta)}{\delta \beta} = -2X'y + 2X'X\beta = -2X'(y - X\beta) = 0,$$

da cui

$$X'X\beta = X'y.$$

Lo stimatore dei minimi quadrati $\hat{\beta}$ risulta dunque pari a:

$$\hat{\beta} = (X'X)^{-1} X'y.$$

Per stimare i termini di disturbo del modello (5) si usano i valori $e_t = y_t - \hat{y}_t$, per $t = 1, 2, \dots, n$, detti anche *residui*. I residui possono essere anche espressi in funzione degli ε_t come segue:

$$\begin{aligned} e &= y - \hat{y} = y - X\hat{\beta} = y - X(X'X)^{-1}X'y = [I_n - X(X'X)^{-1}X']y \\ &= [I_n - X(X'X)^{-1}X'](X\beta + \varepsilon) = [I_n - X(X'X)^{-1}X']\varepsilon. \end{aligned} \tag{6}$$

Essi risultano inoltre ortogonali rispetto alle variabili indipendenti:

$$X'e = X'(y - X\hat{\beta}) = X'y - X'X(X'X)^{-1}X'y = 0.$$

Usando questo criterio si ha:

$$\begin{aligned} y'y &= (\hat{y} + e)'(\hat{y} + e) = (X\hat{\beta} + e)'(X\hat{\beta} + e) \\ &= \hat{\beta}'X'X\hat{\beta} + e'e + 2\hat{\beta}'X'e = \hat{\beta}'X'X\hat{\beta} + e'e, \end{aligned} \quad (7)$$

da cui

$$e'e = y'y - \hat{\beta}'X'X\hat{\beta} = y'y - \hat{\beta}'X'\hat{y}.$$

L'espressione (7) corrisponde alla scrittura:

$$\sum_{t=1}^n y_t^2 = \sum_{t=1}^n \hat{y}_t^2 + \sum_{t=1}^n e_t^2$$

la quale, nel caso in cui nel modello sia presente l'intercetta, può essere ridefinita nella seguente forma, con le osservazioni espresse in termini di scarti dalla media generale:

$$\sum_{t=1}^n (y_t - \bar{y})^2 = \sum_{t=1}^n (\hat{y}_t - \bar{y})^2 + \sum_{t=1}^n e_t^2. \quad (8)$$

Il termine a primo membro rappresenta la *devianza totale* delle osservazioni (*Total Sum of Squares, TSS*), il primo termine al secondo membro rappresenta la *devianza spiegata* dalla regressione (*Explained*

Sum of Squares, ESS), mentre l'ultimo termine rappresenta la *devianza residua (Residual Sum of Squares, RSS)*¹⁹. L'espressione (8) diventa quindi:

$$TSS = ESS + RSS.$$

Una misura della bontà della regressione è data dal *coefficiente di determinazione multipla R^2* , così definito:

$$R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}, \quad 0 \leq R^2 \leq 1.$$

Il coefficiente R^2 può essere interpretato come misura della proporzione della devianza di y_t spiegata dalla regressione²⁰. Per tener conto del numero di variabili indipendenti del modello si può considerare una versione “corretta” di R^2 :

$$\bar{R}^2 = 1 - \frac{RSS/(n-p)}{TSS/(n-1)} = 1 - \left[\frac{n-1}{n-p} (1 - R^2) \right].$$

Infine, lo stimatore di σ_ε^2 risulta così definito:

$$s^2 = \frac{e'e}{n-p} = \sum_{i=1}^n \frac{e_i^2}{n-p}. \quad (9)$$

¹⁹ DI FONZO T., LISI F., Complementi di statistica economica,ecc., op.cit..

²⁰ GIRONI G., SALVEMINI T., Lezioni di statistica,.....ecc., op. cit..

4.6.2 Proprietà degli stimatori dei minimi quadrati

Per calcolare media e varianza di $\hat{\beta}$ si consideri la seguente espressione:

$$\hat{\beta} = (X'X)^{-1} X'y = (X'X)^{-1} X'(X\beta + \varepsilon) = \beta + (X'X)^{-1} X'\varepsilon. \quad (10)$$

Se vale l'assunto che $E[\varepsilon] = 0$, allora $\hat{\beta}$ è uno stimatore corretto di β .

Infatti:

$$E[\hat{\beta}] = \beta + (X'X)^{-1} X'E[\varepsilon] = \beta + 0 = \beta.$$

La matrice di covarianza di $\hat{\beta}$ è paria :

$$\text{Var}[\hat{\beta}] = E\left[(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\right] = (X'X)^{-1} X'E[\varepsilon\varepsilon']X(X'X)^{-1}.$$

Se vale l'assunto $E[\varepsilon\varepsilon'] = \sigma_\varepsilon^2 I_n$, si ha che:

$$\text{Var}[\hat{\beta}] = \sigma_\varepsilon^2 (X'X)^{-1}$$

e il singolo elemento di $\hat{\beta}$ avrà varianza paria $\text{Var}[\hat{\beta}_i] = \sigma_\varepsilon^2 a_{ii}$, dove a_{ii}

è il termine i -esimo sulla diagonale principale della matrice $(X'X)^{-1}$.

In base al teorema di Gauss-Markov²¹, si può affermare che gli

²¹ JOHNSTON J., Econometrica, Franco Angeli, Milano, 1993.

elementi di $\hat{\beta}$ sono i migliori stimatori lineari corretti dei corrispondenti elementi β , nel senso che ogni elemento di $\hat{\beta}$ possiede varianza minore rispetto allo stesso elemento di qualsiasi altro stimatore corretto e lineare in \mathbf{y} . Si può inoltre dimostrare che s^2 è una stima corretta di σ_ε^2 e quindi $\hat{\text{var}}[\hat{\beta}] = s^2(X'X)^{-1}$ è una stima corretta di $\text{Var}[\hat{\beta}]$.

Se vale l'assunto che i termini di disturbo ε_t siano normalmente distribuiti, si può ricavare la distribuzione degli stimatori $\hat{\beta}$ e s^2 . L'espressione (10) ci consente di osservare che il vettore $\hat{\beta}$ è una combinazione lineare di variabili distribuite normalmente ed ha dunque una distribuzione normale multivariata:

$$\hat{\beta} \sim N[\beta, \sigma_\varepsilon^2(X'X)^{-1}]$$

Per quanto riguarda s^2 , risultando dall'espressione (6) che e_t è funzione di ε_t ed essendo s^2 definito in (9) come la sommatoria dei valori e_t^2 , si ha che la quantità $\frac{s^2}{\sigma^2}$ è pari alla somma di $n-p$ variabili χ_1^2 indipendenti e, dunque,

$$s^2 \sim \frac{\sigma^2}{n-p} \chi_{n-p}^2$$

4.7 La determinazione della componente tendenziale

In questo paragrafo si considerano funzioni rappresentanti il trend, che siano stimabili mediante le procedure derivate nell'ambito del modello di regressione lineare. Va premesso che è sempre consigliabile procedere alla determinazione della componente tendenziale utilizzando dati annuali perché per tal via vengono eliminate tutte le difficoltà che sorgono quando si opera con dati aventi scadenza temporale inferiore all'anno²².

4.7.1 Procedimento grafico

Consiste nell'individuare la componente tendenziale tracciando a mano libera una linea continua passante tra i valori osservati.

Per ridurre al minimo il grado di arbitrarietà, occorre che la linea sia tracciata in modo che la somma complessiva degli scarti fra i valori osservati e i corrispondenti valori teorici sia zero o prossima allo zero e la somma delle aree al di sopra della linea di trend sia all'incirca uguale alla somma delle aree che si trovano al di sotto.

²² ALVARO G., *Contabilità nazionale e statistica economica*,.....ecc., op. cit..

4.7.2 Rappresentazione analitica

Quando si vuole definire la legge alla base della manifestazione della tendenza evolutiva di un fenomeno bisogna ricorrere ad una rappresentazione analitica della serie storica, ossia alla determinazione di una funzione o curva matematica che si accosti il più possibile ai valori della serie. Si tratta in pratica di effettuare una interpolazione per punti con una curva che passi il più possibile vicino ai dati osservati²³.

- **Trend polinomiale ed esponenziale**

Un trend polinomiale è essenzialmente uno strumento descrittivo. Esso riassume le caratteristiche salienti di una serie storica e per far questo in maniera utile deve essere di grado piuttosto basso. In molti casi, non c'è alcun significato sostanziale che possa essere dato ai coefficienti del polinomio. Questo è in realtà una funzione scelta come sostitutiva rispetto a qualche funzione del tempo non nota, ma certamente più sofisticata e complessa.

Si assuma che valga il modello:

$$Y_t = f(t) + \varepsilon_t$$

²³ ALVARO G., *Contabilità nazionale e statistica economica,.....ecc.*, op.cit.

con $\varepsilon_t \sim WN(0, \sigma_\varepsilon^2)$, e si supponga, per il momento, che la parte sistematica della serie sia composta dal solo trend.

Se si assume che la funzione deterministica $f(t)$ sia un polinomio di grado q del tipo

$$f(t) = \alpha_0 + \alpha_1 t + \dots + \alpha_q t^q,$$

il modello di regressione che ne deriva è dato da

$$y_t = \alpha_0 + \alpha_1 t + \dots + \alpha_q t^q + \varepsilon_t, \quad t = 1, 2, \dots, n,$$

che è lineare nei parametri e facilmente stimabile adottando il principio dei minimi quadrati. Attraverso questo procedimento si determinano i parametri di una curva interpolatrice che presenta il minimo valore della somma dei quadrati degli scostamenti (o scarti) fra i valori teorici forniti dalla curva e i corrispondenti valori osservati della serie storica.

Se, per esempio, il valore osservato al tempo t è dato da x_t e quello teorico, y_t , è fornito dalla retta $\alpha_0 + \alpha_1 t$, i valori dei parametri α_0, α_1 in corrispondenza dei quali la retta si approssima meglio ai valori della serie storica si ottengono imponendo la condizione:

$$\sum_{i=1}^n (y_{t_i} - x_{t_i})^2 = \sum_{i=1}^n [(\alpha_0 + \alpha_1 t) - x_{t_i}]^2 = \text{minimo}.$$

L'ordine q del polinomio dipende dal comportamento di fondo della serie storica. Casi particolari di modelli con trend polinomiale sono i seguenti:

$$Y_t = \alpha_0 + \varepsilon_t \quad \text{trend costante;}$$

$$Y_t = \alpha_0 + \alpha_1 t + \varepsilon_t \quad \text{trend lineare;}$$

$$Y_t = \alpha_0 + \alpha_1 t + \alpha_2 t^2 + \varepsilon_t \quad \text{trend parabolico.}$$

Il polinomio stimato può essere utilizzato tranquillamente a fini interpolativi, mentre va usato con molta cautela a fini previsivi, poiché non si sa quanto buona sia l'approssimazione polinomiale stimata, quando si esce dall'intervallo temporale di osservazione della serie storica²⁴.

Prendendo q abbastanza grande si può ottenere un accostamento molto stretto dei valori stimati y_t ai valori osservati x_t , ma il modello perde gradi di libertà all'aumentare di q .

La scelta del tipo di funzione non può essere effettuata senza un attento esame del grafico della serie storica. In alternativa o in aggiunta a tale esame si può analizzare l'andamento delle differenze prime dei valori osservati:

$$\Delta x_t = x_t - x_{t-1}$$

²⁴ DI FONZO T., LISI F., Complementi di statistica economica,ecc., op.cit..

in cui x_t rappresenta il valore della serie storica relativo al tempo t .

Se le differenze prime sono costanti, o moderatamente variabili, la curva da adottare è una retta.

Se le differenze prime non sono costanti, ma lo sono le differenze seconde:

$$\Delta^2 x_t = \Delta x_t - \Delta x_{t-1}$$

segue che la curva più idonea a rappresentare la componente tendenziale è un polinomio di secondo grado, ossia una parabola.

Invece, se risultano costanti o moderatamente variabili i rapporti fra i successivi termini di una serie storica:

$$z_t = \frac{x_t}{x_{t-1}}$$

oppure, il che è lo stesso, se risultano costanti o moderatamente variabili le differenze prime dei logaritmi dei successivi termini della serie:

$$\log z_t = \log x_t - \log x_{t-1} = \Delta \log x_t = (1 - B) \log x_t$$

si ha che la componente tendenziale può essere analiticamente rappresentata da una curva del tipo:

$$x_t = \alpha_0 \cdot \alpha_1^t$$

cioè, da una funzione esponenziale.

La funzione esponenziale può essere assunta per rappresentare il trend di quelle serie storiche che aumentano o diminuiscono secondo una progressione geometrica.

4.8 La determinazione della componente stagionale

Il modello di regressione può essere utilizzato anche per stimare la componente stagionale, che viene rappresentata tramite una funzione periodica $g(t)$. Sono dette periodiche quelle funzioni il cui valore all'istante t si riproduce esattamente ad intervalli costanti, la cui lunghezza s costituisce il periodo, ossia:

$$g(t) = g(t + s) = g(t + 2s) = g(t + 3s) = \dots$$

($s = 4$ per serie trimestrali, $s=12$ per serie mensili, ecc.).

Nel seguito vengono presi in considerazione diversi modi di trattare la componente stagionale nel contesto di un modello di regressione.

4.8.1 Le variabili dummy

Supponiamo che la funzione periodica $g(t)$ sia rappresentabile tramite

$$g(t) = \sum_{j=1}^s \gamma_j d_{jt}, \quad t = 1, 2, \dots, n,$$

dove $n = sN$ (la serie storica y_t viene così osservata su un arco di N anni completi) e d_{jt} è una variabile ausiliaria, detta anche variabile *dummy*²⁵, data da:

$$d_{jt} = \begin{cases} 1 & \text{nella stagione } j\text{-esima, } j = 1, 2, \dots, s \\ 0 & \text{altrimenti} \end{cases} .$$

Ad esempio, nel caso di una serie storica trimestrale osservata per N anni completi e in assenza della componente di trend, il modello di regressione associato è il seguente:

$$y = D\gamma + \varepsilon,$$

dove

$$D = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad \text{e} \quad \gamma = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \gamma_3 \\ \gamma_4 \end{bmatrix}$$

sono rispettivamente la matrice $(n \times 4)$, $n = 4N$, contenente le variabili ausiliare ed il vettore (4×1) contenente i parametri.

²⁵ PASSAMANI G., Analisi classica delle componenti di una serie storica, Materiale didattico, Università di Padova, 1990.

In base ai risultati visti nel paragrafo (4.6), il vettore dei coefficienti stimati col metodo dei minimi quadrati è:

$$\hat{y} = (D'D)^{-1} D'y,$$

per cui la stima della componente stagionale è data da

$$\hat{g}(t) = \sum_{j=1}^S \hat{\gamma}_j d_{jt}, \quad t = 1, 2, \dots, n,$$

ovvero, in notazione matriciale, $\hat{g} = D\hat{\gamma}$.

La serie *destagionalizzata*, y^d , in questo caso particolare coincide con la serie dei residui:

$$y^d = y - \hat{g} = y - D\hat{\gamma} = e.$$

4.8.2 L'uso di funzioni trigonometriche

Una serie armonica è data da:

$$\delta_t = A \cos \frac{2\pi t}{T} + B \sin \frac{2\pi t}{T}$$

che può essere scritta anche nella forma:

$$\delta_t = \sqrt{A^2 + B^2} \cos \left(\frac{2\pi t}{T} - \alpha \right)$$

dove $\alpha = \arctan \frac{B}{A}$ $\left(\alpha = \frac{\pi}{2} \text{ se } A = 0 \right)$. La quantità T è detta periodo

dell'armonica, il reciproco, $\frac{1}{T}$, prende il nome di frequenza, $\sqrt{A^2 + B^2}$

è l'ampiezza e α è l'angolo di fase. Le quantità A e B vengono solitamente chiamate componenti dell'armonica.

La componente stagionale può essere rappresentata anche da funzioni del tipo:

$$S_t = \sum_{i=1}^m A_i \cos\left(\frac{2\pi i}{s} t - \phi_i\right)$$

In questa caratterizzazione, la componente stagionale è una somma di m armoniche, il cui generico addendo ha periodo pari a $\frac{s}{i}$, frequenza angolare $\omega_i = \frac{2\pi i}{s}$, ampiezza A_i e angolo di fase ϕ_i , $i = 1, 2, \dots, m$.

Ad esempio, per dati mensili ($s = 12$), la prima armonica ($i = 1$), $\cos\left(\frac{\pi}{6} t - \phi_1\right)$, descrive un'onda cosinusoidale che completa il suo ciclo in 12 periodi di tempo; la seconda armonica ($i = 2$) completa il ciclo in 6 periodi, la terza in 4, la quarta in 3 e così via.

Solitamente non è necessario considerare tutte le possibili armoniche, poiché già le prime sono in grado di generare dinamiche stagionali complesse²⁶.

²⁶ DI FONZO T., LISI F., Complementi di statistica economica,ecc., op.cit..

4.9 La determinazione della componente ciclica

La rappresentazione delle fluttuazioni cicliche è sempre stata al centro degli studi statistici e statistico-econometrici, data la sua importanza a fini decisionali. Conoscere, infatti, l'ampiezza e il periodo di un movimento ciclico significa conoscere in anticipo quando e come si verificano le fasi di espansione e di depressione.

Contrariamente alle fluttuazioni stagionali che presentano un periodo costante, la ricorrenza della componente ciclica generalmente non si presenta con periodicità costante. Ciò in quanto un evento che si verifica in un dato luogo e tempo rappresenta in genere l'effetto di cause registratesi in tempi e luoghi diversi e, per tale motivo, difficilmente individuabili e, soprattutto, quantificabili.

Di qui, l'impossibilità di poter effettuare una rappresentazione analitica della componente ciclica mediante l'impiego di funzioni semplici, in particolare attraverso le funzioni trigonometriche.

Presentando periodo e ampiezza generalmente variabili, le serie storiche richiedono, per la rappresentazione del loro andamento ciclico, funzioni molto complesse, il cui impiego, oltre a non essere agevole,

difficilmente conduce a interpretazioni di apprezzabile contenuto e a risultati di diretta utilizzazione decisionale²⁷.

4.10 Il metodo delle medie mobili

La descrizione analitica della componente di fondo come funzione del tempo presenta degli indubbi vantaggi, ma trova un suo limite quando l'andamento del fenomeno si manifesta in modo irregolare. In questo caso, per ottenere indicazioni valide dal punto di vista statistico, è necessario ricorrere a polinomi di grado elevato, i cui risultati non sono di facile interpretazione.

Un modo alternativo di operare consiste nel ricercare, in modo empirico, la componente di fondo senza voler necessariamente evidenziare una legge di variazione. Uno strumento semplice e flessibile, adatto a questo scopo è la media mobile, data da una trasformazione lineare delle osservazioni. Le procedure fondate sulle medie mobili vengono usate principalmente per stimare il trend, destagionalizzare e eliminare o ridurre la componente erratica.

²⁷ ALVARO G., *Contabilità nazionale e statistica economica*,.....ecc., op.cit.

4.10.1 Le medie mobili

Si consideri una serie storica $\{y_t\}_{t=1}^n$ per cui valga un modello di aggregazione delle componenti di tipo additivo:

$$y_t = T_t + S_t + \varepsilon_t \quad t = 1, \dots, n.$$

Nel modello non compare la componente ciclica, che per semplicità si ipotizza assimilabile al trend T_t .

Una maniera semplice per determinare una delle componenti, per esempio il trend, consiste nell'applicare alla serie una trasformazione lineare g che conservi il trend ed annulli le altre componenti. Più precisamente, si indicherà con y_t^*, T_t^*, S_t^* e ε_t^* le trasformazioni ottenute tramite g della serie, rispettivamente, y_t, T_t, S_t e ε_t . In base alle ipotesi fatte su g , $S_t^* = 0$, $\varepsilon_t^* = 0$ e, dunque, $y_t^* = T_t^* = T_t$.

Ovviamente, la scelta della trasformazione g non è affatto semplice. Nella maggior parte delle situazioni concrete, non è possibile individuare una trasformazione che conservi esattamente il trend e che allo stesso tempo annulli la componente stagionale e la componente erratica. Al più, si può sperare di vedere soddisfatte queste proprietà in forma approssimata, ossia $T_t^* \cong T_t$, $S_t^* \cong 0$, $\varepsilon_t^* \cong 0$.

Caratteristiche desiderabili della trasformazione g sono le seguenti:

- a. i calcoli devono essere semplici;
- b. l'aggiornamento facile all'aumentare di n ;
- c. il metodo deve reagire bene a cambiamenti di regime.

4.10.2 La stima della componente tendenziale-ciclica

Si consideri una serie storica di n dati mensili. Una stima dei valori attribuibili al trend-ciclo può ottenersi in base alle seguenti considerazioni ²⁸:

- a. la componente stagionale per definizione, si compensa nell'arco di 12 mesi, per cui una media di 12 valori mensili consecutivi non ne è più influenzata;
- b. la componente erratica tende a compensarsi in una media di un congruo numero di termini successivi della serie.

Date n osservazioni di una serie storica, le medie mobili a k termini sono date da medie aritmetiche semplici di k osservazioni.

Quando k è dispari, ogni termine della media mobile corrisponde o, meglio, è centrato sul termine centrale delle k osservazioni. Per esempio, il termine della prima media mobile a tre termini è centrato sul secondo termine della serie osservata; quello della prima media mobile a cinque

²⁸ ZANI . S., Indicatori statistici della congiuntura, Loescher, Torino, 1982.

termini risulta centrato sul terzo della serie osservata. In generale, quindi, se k è dispari il termine della prima media mobile è centrato sul termine $\frac{k+1}{2}$ della serie osservata.

Quando k è pari, i termini delle medie mobili non risultano centrati rispetto ai termini della serie osservata. Per esempio, se $k = 4$, il termine della prima media mobile è dato da:

$$y_{(2,3)}^{**} = \frac{y_1 + y_2 + y_3 + y_4}{4}$$

che, come si può rilevare, si colloca tra il secondo e il terzo termine dei valori originari.

Di conseguenza, il secondo termine della media mobile, fornito da:

$$y_{(3,4)}^{**} = \frac{y_2 + y_3 + y_4 + y_5}{4}$$

si situa tra il terzo e il quarto termine della serie osservata.

Per poter ottenere la media mobile centrata sul terzo termine occorre eseguire le medie mobili a due termini sulle medie mobili calcolate. Nell'esempio in esame, la media mobile a 4 termini centrata sul terzo termine è data da:

$$\bar{y}_3^{**} = \frac{y_{(2,3)}^{**} + y_{(3,4)}^{**}}{2}$$

che, ricordando il valore di $y_{(2,3)}^{**}$ e $y_{(3,4)}^{**}$, si può anche scrivere:

$$\bar{y}_3^{**} = \frac{1}{4} \left(\frac{1}{2} y_1 + y_2 + y_3 + y_4 + \frac{1}{2} y_5 \right)$$

la quale si riferisce a cinque mesi successivi con pesi uguali a 0,5 per il primo e l'ultimo termine e pesi uguali a 1 per i rimanenti.

Conseguentemente, una stima dei valori attribuibili al trend-ciclo può essere ottenuta calcolando medie mobili a 12 termini centrate sul settimo termine che, in forma generale, sono date da:

$$y_t^{**} = \frac{1}{12} \left(\frac{1}{2} y_{t-6} + y_{t-5} + y_{t-4} + y_{t-3} + y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2} + y_{t+3} + y_{t+4} + y_{t+5} + \frac{1}{2} y_{t+6} \right)$$

Ogni serie ottenuta da medie mobili presenta un numero di termini inferiori a quello della serie osservata. Per esempio, nelle medie mobili a tre termini risultano mancanti, rispetto alla serie originaria, il primo e l'ultimo termine; nelle medie mobili a cinque termini mancano invece i primi e gli ultimi due termini.

In conclusione, occorre considerare che, se la serie storica presenta una componente di fondo non lineare, l'impiego delle medie mobili conduce a valori errati. In particolare, se la componente tendenziale è concava verso l'alto (il che significa che si tratta di una serie di valori che crescono a tassi crescenti), le medie mobili forniscono valori errati per eccesso; invece, se l'andamento è convesso verso l'alto

(ossia, la serie è crescente a tassi decrescenti), l'impiego delle medie mobili conduce a valori errati per difetto²⁹.

Questi risultati spingono alla conclusione che l'impiego delle medie mobili per l'eliminazione della componente ciclica e la conseguente determinazione di quella tendenziale deve essere effettuato con molta cautela ed accortezza. Soprattutto perché nella realtà le serie storiche presentano un profilo molto irregolare e comunque un andamento ciclico con periodo non costante, per cui diviene molto difficile, se non addirittura impossibile, determinare il numero dei termini in base al quale calcolare le medie mobili in grado di eliminare la componente ciclica.

4.10.3 Il trattamento della componente stagionale

L'impiego delle medie mobili centrate su 12 termini trova, invece, una giustificata utilità, quando si deve procedere alla eliminazione della componente stagionale, giacché tale componente presenta un andamento con un periodo di 12 termini.

In una serie storica espressa in termini mensili, i dati, relativi a due o più mesi consecutivi, non possono essere fra loro comparati.

²⁹ ALVARO G., *Contabilità nazionale e statistica economica*,.....ecc., op.cit

Per effettuare tale comparazione, occorre dapprima determinare l'influenza della componente stagionale sulla manifestazione del fenomeno per poterla eliminare dai valori osservati, in modo da ottenere serie storiche destagionalizzate.

Di conseguenza, se una serie di dati mensili è perequata con medie mobili centrate su 12 termini, si ottiene una nuova serie mensile da cui, risultano eliminate le oscillazioni stagionali; detto diversamente, si ottiene una nuova serie mensile, il cui andamento è caratterizzato dalle componenti tendenziale, ciclica e accidentale. In questo senso le medie mobili centrate su 12 termini possono anche essere denominate medie mobili centrate destagionalizzanti.

Le medie mobili centrate destagionalizzanti (nel caso di dati mensili) si calcolano come segue:

$$y_t^{**} = \frac{1}{12} \left(\frac{1}{2} y_{t-6} + y_{t-5} + y_{t-4} + y_{t-3} + y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2} + y_{t+3} + y_{t+4} + y_{t+5} + \frac{1}{2} y_{t+6} \right)$$

Se sui valori osservati mensilmente, y_t , si calcolano le medie mobili centrate destagionalizzanti, si ottiene una nuova serie di valori mensili che non presenta più fluttuazioni stagionali ed il cui andamento esprime l'effetto combinato delle due componenti tendenziale e ciclica.

Di conseguenza, una stima per l'insieme della componente stagionale e aleatoria, che si indicherà con $S\varepsilon_t$, può essere ottenuta calcolando:

$$S\varepsilon_t = y_t - y_t^{**} \quad \text{nel caso si tratti di modello additivo,}$$

$$S\varepsilon_t = \frac{y_t}{y_t^{**}} \quad \text{nel caso di modello moltiplicativo.}$$

Le quantità $S\varepsilon_t$ vengono dette indici o rapporti specifici di stagionalità e sono caratterizzate dalla presenza della componente accidentale.

Prima di procedere alla separazione della componente erratica da quella stagionale è opportuno verificare se quest'ultima sia significativamente presente nei valori $S\varepsilon_t$. Si formula l'ipotesi nulla di assenza di stagionalità nei valori $S\varepsilon_t$. Se essa è vera, le medie degli $S\varepsilon_t$, calcolate sui valori dello stesso mese nei diversi anni, non differiscono significativamente tra loro. Il metodo solitamente impiegato per la verifica dell'ipotesi nulla suddetta è l'analisi della varianza ad un criterio di classificazione.

Se l'ipotesi d'assenza di stagionalità viene rigettata, si perviene ad una serie di 12 valori, detti coefficienti grezzi di stagionalità:

$$S_j^* = \frac{1}{N} \sum_{T=1}^N S\varepsilon_{T,j}, \quad j = 1, \dots, 12,$$

dove N è il numero di anni in cui la serie storica è stata osservata.

Prima di poter utilizzare i risultati conseguiti, vi è un'ultima considerazione da fare. Nel caso in cui il modello d'aggregazione delle componenti scelto sia di tipo moltiplicativo, si ha che i coefficienti di stagionalità sono numeri puri. Ciò impone che a livello annuo si dovrebbe avere:

$$\sum_{j=1}^{12} S_j^* = 12$$

o, se si opera con delle percentuali,

$$\sum_{j=1}^{12} S_j^* = 1200$$

Poiché tali condizioni non sono, in genere, soddisfatte, si rende necessaria un'ulteriore elaborazione dei coefficienti grezzi che dà luogo ai coefficienti ideali:

$$\hat{S}_j^* = \frac{S_j^* \cdot 12}{\sum_{j=1}^{12} S_j^*} \quad j = 1, \dots, 12,$$

ovvero, nel caso di percentuali,

$$\hat{S}_j^* = \frac{S_j^* \cdot 1200}{\sum_{j=1}^{12} S_j^*} \quad j = 1, \dots, 12,$$

La serie dei dati destagionalizzati si ottiene, quindi, sottraendo o dividendo i dati originari y_t per \hat{S}_t .

4.11 Regressione non parametrica

In molti casi, a meno di un certo numero di parametri, la funzione $f(t)$, è nota; se, inoltre, essa è lineare nei parametri la (1) diventa un modello di regressione lineare facilmente stimabile. Se $f(t)$ non è lineare nei parametri, la procedura per la stima del modello si presenta più complessa.

In altri casi la funzione $f(t)$ non è nota, ma è approssimabile adeguatamente mediante una opportuna combinazione lineare di funzioni del tempo. Ad esempio, se il trend fluttua con ampi movimenti, lo si può rappresentare con un polinomio; se invece mostra una certa periodicità, esso può essere rappresentabile mediante una combinazione lineare di termini trigonometrici. In tutti questi casi la stima del modello (1) si presenta generalmente facile.

Il punto debole dei diversi approcci fin qui visti è che i risultati dipendono in maniera cruciale dalla capacità e dalla possibilità di scegliere in maniera appropriata le funzioni con cui interpolare il trend e

la stagionalità. Talvolta, infatti la funzione $f(t)$ non è nota e non è neppure approssimabile. In questi casi, per la stima delle componenti si usa un approccio più “flessibile” basato su procedure di lisciamiento (*smoothing*)³⁰.

4.11.1 Formulazione del problema

Si considerino dei dati bivariati del tipo:

$$\{(x_1, y_1), \dots, (x_n, y_n)\}$$

su due variabili X e Y;

La relazione tra le due variabili può essere scritta nella forma

$$y_i = f(x_i) + \varepsilon_i \quad (11)$$

dove $f(\cdot) = E(Y | X = x)$ mentre le ε_i sono delle variabili casuali a media nulla.

Il problema è che non si è in grado di specificare $f(\cdot)$ parametricamente (ad esempio, non è una retta, non è un polinomio, ...), ma risulta che $f(\cdot)$ è una funzione continua e senza oscillazioni particolarmente violente.

Ci si chiede, pertanto di costruire una stima di $f(\cdot)$ utilizzando i dati.

³⁰ DI FONZO T., LISI F., Complementi di statistica economica,ecc., op.cit..

4.11.2 Le medie locali

Si supponga che:

$$f(x) = \eta \quad \text{per qualsivoglia } x$$

dove η indica una costante coincidente con la media della variabile Y .

Si noti che, per la (11), si può scrivere $y_i = \eta + \varepsilon_i$ e che le ε_i hanno media nulla.

In questo caso degenere, si potrebbe stimare $f(\cdot)$ mediante:

$$\hat{f}(x) = \hat{\eta} = \bar{y} = \frac{1}{n} \sum_{i=1}^n y_i \quad \text{per qualsivoglia } x,$$

ovvero, semplicemente calcolando la media delle osservazioni di Y .

Tale procedura, però, non può essere la soluzione generale.

Se le oscillazioni di $f(\cdot)$ sono dolci, si può pensare di stimare $f(\cdot)$

mediante delle medie locali³¹ del tipo:

$$\hat{f}(x) = \text{media delle } y_i \text{ t.c. } |x_i - x| \text{ sia minore di una costante prescelta;}$$

$$\hat{f}(x) = \text{media ponderata delle } y_i \text{ con pesi costruiti in maniera che}$$

risultino “grandi” se $x_i \approx x$ e “piccoli” se x_i è “lontano” da x .

³¹MASAROTTO G., Analisi delle serie temporali, Materiale didattico, Facoltà di Scienze Statistiche, Università di Padova, 2003.

Questo conduce a degli stimatori del tipo:

$$\hat{f}(x) = \sum_{i=1}^n w_i(x) y_i \quad (12)$$

dove $w_i(x)$ è il peso che si assegna a y_i quando si calcola la stima di $f(\cdot)$ a x .

4.11.3 Pesi costruiti da un nucleo

Si supponga di scegliere una funzione $k(\cdot)$ non decrescente per $x < 0$ e crescente per $x > 0$ e tale che $k(x) \approx 0$ quando $|x|$ è sufficientemente grande. Una possibilità per generare i pesi consiste nel porre:

$$w_i(x) = \frac{k\left(\frac{x_i - x}{h}\right)}{\sum_{i=1}^n k\left(\frac{x_i - x}{h}\right)}$$

e, quindi,

$$\hat{f}(x) = \frac{\sum_{i=1}^n k\left(\frac{x_i - x}{h}\right) y_i}{\sum_{i=1}^n k\left(\frac{x_i - x}{h}\right)}$$

La funzione $k(\cdot)$ è usualmente indicata come nucleo (*kernel* in inglese) e lo stimatore risultante viene definito come basato sul metodo del nucleo³².

³² MASAROTTO G., Analisi delle serie temporali,ecc., op. cit.

4.11.4 Regressione locale

Una possibilità diversa consiste nell'utilizzare come stima di $f(x)$ il valore assunto ad x da un polinomio adattato utilizzando solo le osservazioni "vicine".

Uno degli stimatori più utilizzati basati su tale procedura è lo stimatore *loess* che stima $f(x)$ mediante

$$\hat{f}(x) = b_0(x) + b_1(x)x + \dots + b_p(x)x^p$$

dove i coefficienti $b_0(x), \dots, b_p(x)$, che, si osservi, dipendono da x , sono determinati minimizzando:

$$\sum_i k\left(\frac{x_i - x}{h(x)}\right) \left(y_i - b_0(x) - b_1(x)x - \dots - b_p(x)x^p\right)^2$$

con

$$k(x) = \begin{cases} \left(1 - |x|^3\right)^3 & \text{se } |x| \leq 1 \\ 0 & \text{altrove} \end{cases}$$

mentre, $h(x)$ è usualmente determinato in maniera tale che solo s osservazioni ricevono un peso maggiore di 0 (con s valore prefissato).

Anche questo stimatore è del tipo (12), anche se, non è assicurato che i pesi sommino ad 1 e che siano positivi.

4.11.5 Spline

Una *smoothing spline* è la soluzione del seguente problema: trovare la funzione $\hat{f}(\cdot)$ che minimizza, tra tutte le possibili funzioni $f : R \rightarrow R$, la seguente somma dei quadrati penalizzata:

$$SQ_p = \sum_{i=1}^n [y_i - f(x_i)]^2 + \nu \int_{x_{(1)}}^{x_{(n)}} [f''(x)]^2 dx$$

dove $x_{(1)} = \min(x_1, \dots, x_n)$ e $x_{(n)} = \max(x_1, \dots, x_n)$

Si osservi che:

- il primo addendo è una usuale somma dei quadrati degli scarti tra le osservazioni e i valori previsti dal modello; diventa piccolo ovviamente più il modello prevede bene le osservazioni ed, in particolare, diventa nullo per ogni funzione che interpola esattamente i dati stessi;
- il secondo addendo viceversa è una penalità che diventa grande più la derivata seconda è grande (in modulo), ovvero, più varia la derivata prima, cioè più $f(\cdot)$ si allontana da una retta (per una retta la derivata seconda è sempre nulla); penalizza, quindi, le funzioni “non lisce”, quelle con molte oscillazioni e cambi di pendenza;
- ν è un coefficiente che controlla il peso relativo dei due addendi; in particolare, se ν è piccolo la penalizzazione non “pesa”; in questo

caso, otteniamo una $\hat{f}(\cdot)$ che tende ad interpolare molto bene anche a costo di essere troppo poco “liscia” (ovvero, può interpolare anche la componente erratica); viceversa se ν è grande la penalità pesa molto e quindi otteniamo una stima molto “liscia” (per $\nu \rightarrow \infty$ si ottiene, qualsiasi siano i dati, una retta visto che in ogni altro caso la penalità dominerebbe SQ_p).

Si noti che anche questo stimatore è del tipo (12).

4.11.6 Numero di parametri equivalenti

Gli stimatori visti hanno un parametro “aggiustabile” che controlla il grado di lisciamento (h nel caso dello stimatore del nucleo, s nel caso dello stimatore loess, ν nel caso delle spline).

Oltre ai dettagli formali, è inoltre possibile far vedere che una volta fissato h o s o ν lo stimatore che si ottiene ha, nella sostanza, la flessibilità di un modello di regressione con un certo numero di parametri, numero che viene usualmente chiamato numero di parametri equivalenti.

Ovviamente, più il numero di parametri equivalenti è grande, più lo stimatore è flessibile e viceversa.

Il numero di parametri equivalenti costituisce, quindi, una maniera unificata per fissare il grado di lisciamento desiderato.

Tra l'altro, stimatori diversi (ad esempio, "loess" o "spline") con un numero di parametri equivalenti uguali producono, di norma, stime molto simili.

4.11.7 Scelta del grado di lisciamiento

Le costanti che controllano il grado di lisciamiento (ad esempio, il "numero di parametri equivalenti") possono essere scelte "ad occhio" utilizzando un procedimento a prova d'errore e cercando di bilanciare il grado di lisciamiento con la capacità della curva stimata di descrivere i dati.

Esistono poi procedure più formali. Alcune tra le più popolari sono basate su tecniche di validazione incrociata. Nella forma più semplice la validazione incrociata consiste nel:

- dividere i dati (casualmente) in due sottoinsiemi;
- stimare $f(\cdot)$ sul primo sottoinsieme utilizzando vari valori per la costante che controlla il lisciamiento;
- utilizzare le varie stime per prevedere le osservazioni del secondo sottoinsieme;
- adottare il parametro di lisciamiento della "curva migliore", ovvero, quella che ha previsto in maniera migliore il secondo gruppo di

dati, per produrre la stima “finale” di $f(\cdot)$ che, ovviamente, sarà basata su tutti i dati.

Un approccio alternativo si basa sull'utilizzo di criteri del tipo:

$$n \log(\hat{\sigma}_p^2) + cp,$$

dove, indicata con $\hat{f}_p(\cdot)$ la stima di $f(\cdot)$ ottenuta utilizzando p parametri equivalenti, si ha che:

$$\hat{\sigma}_p^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{f}_p(x_i))^2,$$

mentre c è una appropriata costante positiva (eventualmente dipendente da n , ma non da p).

La scelta di p avviene minimizzando il criterio³³. Il primo addendo misura quanto la stima di $f(\cdot)$ “prevede” bene le osservazioni e quindi, usualmente, decresce al crescere di p (più p è grande più lo stimatore usato è flessibile, quindi meglio riesce a riprodurre i dati osservati). Il secondo addendo, invece, penalizza i valori grandi di p .

Noti criteri di questo tipo sono l'Akaike Information Criterion (AIC), che si ottiene ponendo $c = 2$, e lo Schwarz Information Criterion o Bayesian Information Criterion (BIC), che si ottiene ponendo $c = \log(n)$.

³³ MASAROTTO G., Analisi delle serie temporali,.....ecc., op. cit.

4.12 L'analisi dei residui

È fondamentale che si considerino alcune questioni connesse al trattamento della componente accidentale. Questa, come si è visto, viene sempre stimata come residuo (per differenza o rapporto), a conclusione del processo di stima delle componenti deterministiche della serie.

L'analisi dei residui stimati permette di verificare l'adeguatezza del modello scelto per stimare le componenti deterministiche e, quindi, di giustificarne l'uso ai fini previsivi. La verifica fornisce esito positivo se è lecito assumere che la serie dei residui sia stata generata da un *white noise*. Se invece tale serie presenta andamenti regolari o ulteriori componenti stagionali, i risultati ottenuti vanno ridiscussi.

Per la modellazione della componente stocastica u_t , conviene, quindi, ricorrere alla nozione di *processo stocastico stazionario*. Un processo stocastico può essere definito come “un fenomeno che evolve seguendo delle leggi probabilistiche”. Quanto alla stazionarietà, intuitivamente un processo stocastico si dice stazionario se non presenta sistematici cambiamenti nella media e/o nella varianza e variazioni strettamente periodiche.

Il processo *white noise* è uno dei più importanti processi stocastici. Si tratta di un processo puramente casuale, $\{\varepsilon_t\}$, che consiste di una

sequenza di v.c. indipendenti ed identicamente distribuite, di media nulla e varianza costante. Inoltre, se le ε_t hanno distribuzione normale si parla di *white noise gaussiano*.

Sorge, allora, il problema di cosa debba essere una successione di numeri per non rifiutare l'ipotesi nulla H_0 di *casualità*. In effetti, la casualità esiste solo rispetto ad alternative prefissate: contro la ciclicità, contro l'alternanza di segni, contro il trend, contro la correlazione seriale, ecc., per cui occorrerebbe effettuare una sequenza molto articolata di test, in modo da assicurarsi che nessuno di essi conduca al rifiuto di H_0 . In pratica, una volta ottenuta la componente residua, ci si limita ad effettuare alcuni test statistici.

4.12.1 Verifica sulla normalità dei residui

Un modo abbastanza semplice ed intuitivo per verificare la normalità della distribuzione degli errori è quello di ricorrere all'ausilio grafico con un istogramma e con un QQ-plot.

Per avere un risultato più affidabile bisogna, però, effettuare dei test di normalità. La letteratura fornisce vari test a tale scopo: il test di Jarque-Bera, il test del chi-quadro di Pearson, quello di Kolmogorov-Smirnov e quello di Shapiro-Wilk sono fra i più importanti ed utilizzati. In

particolare quest'ultimo è considerato uno dei test più potenti per la verifica della normalità, soprattutto per piccoli campioni.

4.12.2 Il test sulle autocorrelazioni

Test più efficienti si basano sui coefficienti di autocorrelazione ρ_k . È stato infatti dimostrato che i coefficienti di autocorrelazione campionari di una successione di n valori generati da un processo *white noise*, per n sufficientemente grande, si distribuiscono approssimativamente come una variabile casuale normale di media nulla, varianza pari a circa $1/n$ e non sono tra loro correlati.

In una serie di n valori si può dunque respingere l'ipotesi nulla di successione generata da un *white noise* quando si riscontrano valori dei coefficienti di autocorrelazione stimati $\hat{\rho}_k$, esterni all'intervallo $[-z/\sqrt{n}, +z/\sqrt{n}]$, ove z è una funzione del livello di significatività prescelto. In genere si pone $z = 1,96$, che corrisponde ad un livello di significatività di 0,05.

Questa metodologia è però soggetta a varie limitazioni: gli intervalli sono approssimati (la media viene assunta pari a zero e z viene assunto fisso, laddove è in generale variabile di k). Inoltre, la probabilità di ottenere *almeno* un coefficiente di autocorrelazione esterno all'

intervallo $[-1.96/\sqrt{n}, +1.96/\sqrt{n}]$ aumenta al crescere del numero di coefficienti di autocorrelazione considerati, cioè, al numero di ritardi (*lag*) k . Conseguentemente, quando soltanto 1 o 2 coefficienti risultano significativi, bisogna tener conto del numero di sfasamenti considerati prima di respingere l'ipotesi nulla di aleatorietà. Infatti, scelto un livello di significatività pari a 0,05, non vi sarebbe nulla di strano se, anche per dati generati da un processo stocastico *white noise*, fra i primi 20 coefficienti di autocorrelazione campionari uno risulti esterno all'intervallo sopra indicato³⁴.

4.12.3 I test di Ljung-Box e Box-Pierce

Una statistica che può essere utilizzata per verificare l'assenza di autocorrelazione è una opportuna combinazione lineare dei coefficienti di autocorrelazione dei residui $r(t)$ è data:

$$LB = n(n+2) \sum_{t=1}^k \frac{r^2(t)}{n-t}$$

dove k è un intero prescelto. Se è vera l'ipotesi nulla (assenza di autocorrelazione) la statistica LB si distribuisce asintoticamente come una chi-quadro con k gradi di libertà.

³⁴ DI FONZO T., LISI F., Complementi di statistica economica,ecc., op.cit..

Valori troppo grandi della statistica sono evidenza di una possibile autocorrelazione dei residui.

Un test, asintoticamente analogo a quello di Ljung e Box, si basa sulla statistica test proposta da Box e Pierce:

$$BP = n \sum_{t=1}^k r^2(t)$$

La differenza tra le due statistiche consiste semplicemente nella differente ponderazione adottata: nella prima il quadrato di $r(t)$ viene pesato con $n(n+2)/n-t$ mentre nella seconda con peso n .

CAPITOLO 5

ANALISI SULLA DIPENDENZA TRA PIOGGIA E CONDUCEBILITA' ELETTRICA

5.1 Una misura della dipendenza

Gli indici relativi, calcolati nel capitolo terzo, denotano come la profondità di falda influenzi significativamente sulla variabilità della ECw, oltre che sul valore medio.

Come premesso nel precedente capitolo, però, numerosi altri fattori incidono sul parametro di qualità. Ci si chiede, quindi, se esiste qualche relazione di dipendenza tra la pioggia e i parametri di qualità delle acque: ECw e S.A.R..

L'analisi è stata condotta sulle stazioni di Rutigliano e Mola che appartengono all'idrostruttura delle Murge, ma si differenziano per la diversa profondità. La scelta di questi due siti è stata dettata dalle seguenti considerazioni:

1. sono vicine fra loro (la distanza in linea d'aria tra i due pozzi è inferiore a 10 km) e le differenze climatiche e le variazioni meteorologiche si riducono;
2. l'andamento e la quantità delle precipitazioni sono pressoché identiche a Mola e a Rutigliano;
3. i parametri qualitativi registrati a Mola denunciano una maggiore variabilità rispetto a quelli rilevati nelle falde di Rutigliano.

Pertanto, l'analisi statistica effettuata per le falde di Mola e di Rutigliano, che risultano assai contrastate fra di loro, potrebbe essere generalizzata alle altre falde della regione.

Di seguito è riportata la matrice di correlazione tra la variabile pioggia e le variabili di qualità rilevate presso la stazione di Rutigliano:

	rain	ECw	SAR
rain	1.00000000	0.02964368	-0.1467289
ECw	0.02964368	1.00000000	0.1247887
SAR	-0.14672894	0.12478871	1.00000000

, e presso la stazione di Mola:

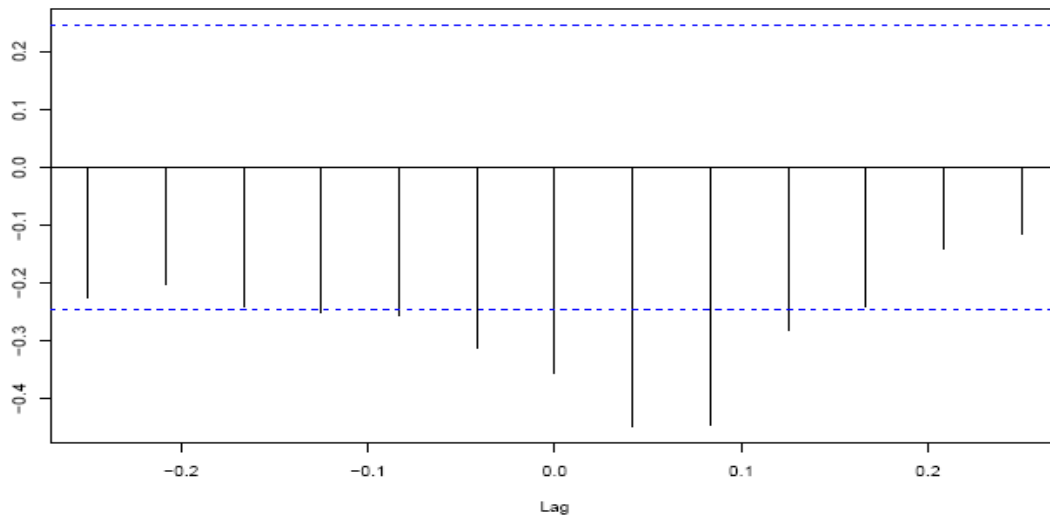
	Rain	ECw	SAR
Rain	1.0000000	-0.3569421	-0.2949086
ECw	-0.3569421	1.0000000	0.4399123
SAR	-0.2949086	0.4399123	1.0000000

I coefficienti mostrano come la correlazione negativa tra l'evento meteorico e i parametri di qualità sia più evidente per la stazione di Mola. Considerando la stazione di Rutigliano, sembra addirittura poter ipotizzare una indipendenza in media tra la ECw e l'evento meteorico. Gli indici calcolati, inoltre, rappresentano una dipendenza "contemporanea", ossia, una relazione, tra le variabili, valutata nello stesso istante temporale.

Da ciò nasce l'esigenza di un'analisi temporale delle variabili interessate al fine di individuare una possibile relazione dinamica tra le stesse.

Come illustrato nel paragrafo 3 del capitolo precedente, una misura della dipendenza lineare esistente tra due processi stocastici a vari istanti di tempo viene fornita dalla funzione di correlazione incrociata. Di seguito se ne riporta una rappresentazione grafica in cui sono stati considerati sei ritardi.

Correlazione incrociata tra la pioggia e la ECw



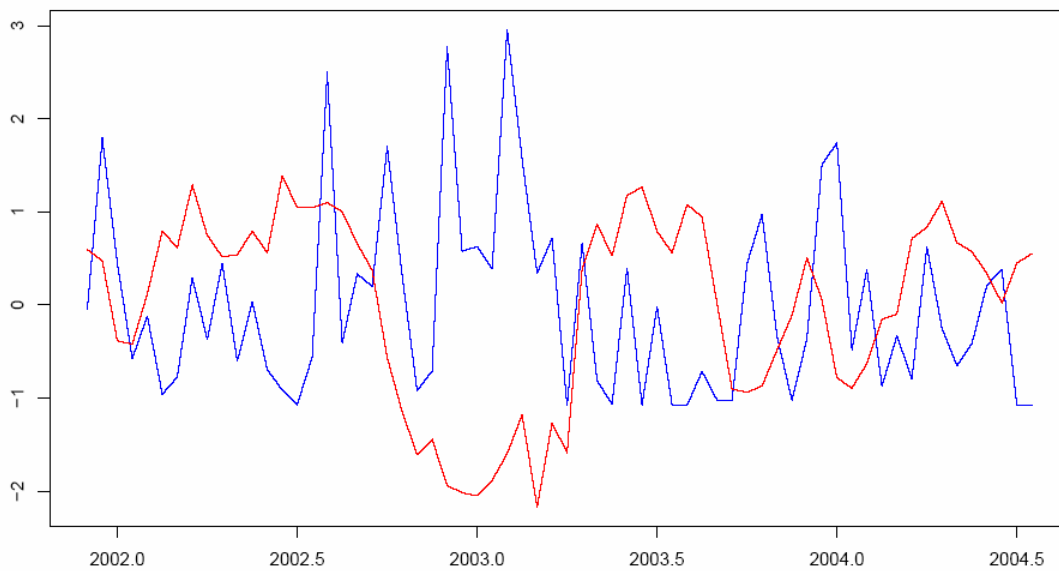
Dall'analisi grafica si denota che la correlazione negativa tra le due variabili si estende nel breve periodo a vari istanti di tempo, in particolare la correlazione assume valori maggiori (-0,45 circa) tra la ECw rilevata al tempo t e la pioggia riferita ai tempi $t-1$ e $t-2$. Tuttavia, la significatività dei coefficienti è verificata soltanto nell'ipotesi che le due serie storiche in esame siano generate da un processo stocastico puramente aleatorio (*white noise*).

A tale fine, si analizzeranno le due serie allo scopo di trasformarle in una successione di residui, per i quali siano accettabili le ipotesi poste.

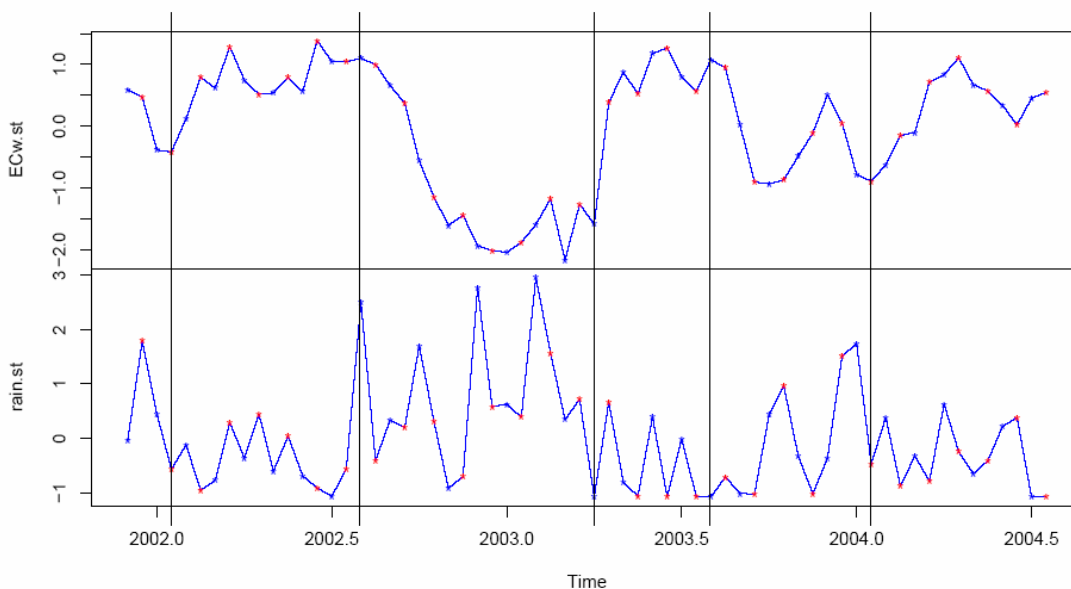
Di seguito è riportato l'andamento quindicinale della ECw (linea rossa), rilevata presso la stazione di Mola (BA5), nel periodo che va dal 1 dicembre 2001 al 15 luglio 2004, e quello relativo alle altezze di pioggia cumulate quindicinali (linea blu), rilevate presso la vicina

stazione di Rutigliano nello stesso periodo di riferimento. Ovviamente, per poter effettuare una comparazione tra le due variabili espresse in scale di misura differenti, si è dovuto preliminarmente standardizzarle³⁵.

Andamento pioggia e ECw standardizzate



Rappresentazione dei punti di svolta per la ECw

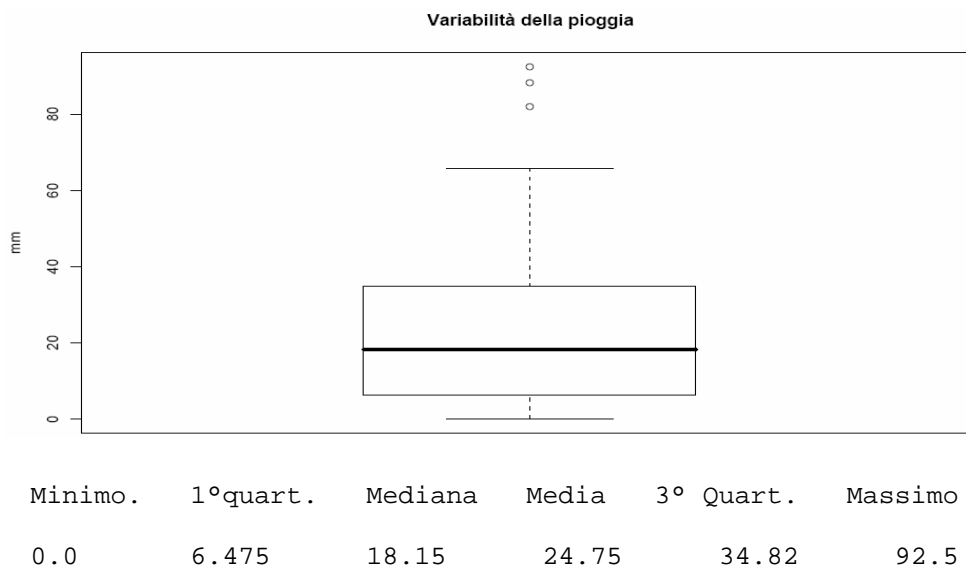


³⁵ cfr TAVOLA A1 in Appendice.

Dall'analisi grafica appare evidente che a periodi in cui si sono registrati massimi di pioggia corrispondono valori minimi del parametro ECw e, viceversa, massimi di ECw si riscontrano proprio nei periodi in cui è piovuto poco. Si sottolinea una tendenza all'aumento della salinità dall'inizio alla fine della stagione irrigua (da aprile a ottobre di ogni anno). Questo andamento si potrebbe giustificare, oltre che per la diminuzione delle precipitazioni, anche per un aumento dei prelievi da falda per l'irrigazione (è il periodo di massima richiesta di acqua da parte delle colture orticole della zona) e per il soddisfacimento del fabbisogno idrico ad uso abitativo e turistico che aumenta durante il periodo estivo.

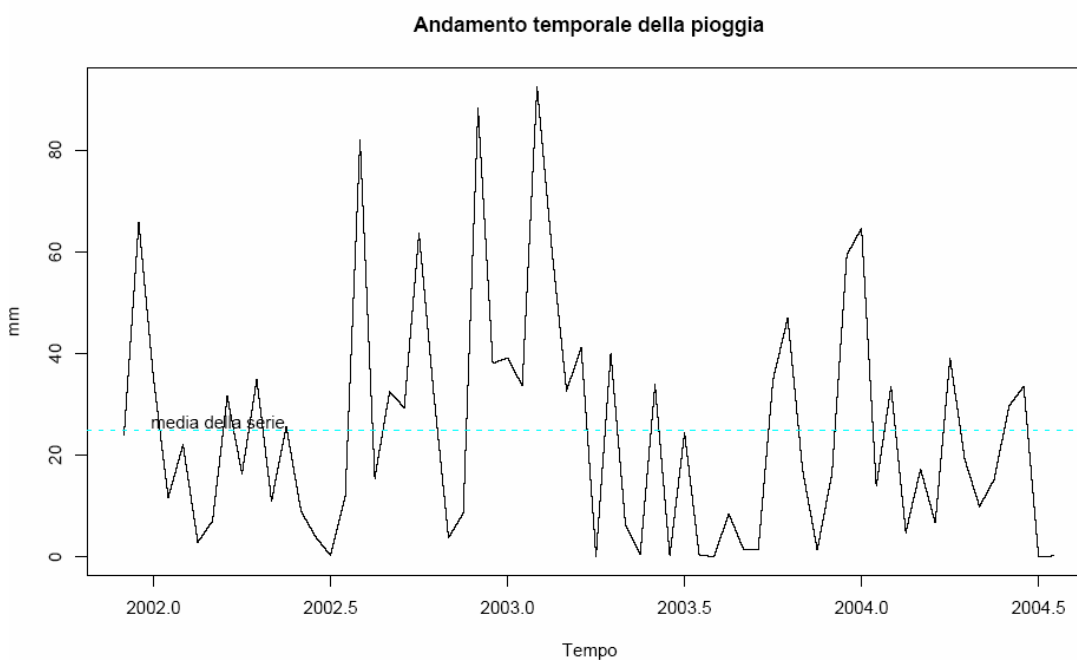
5.2 Analisi temporale della pioggia

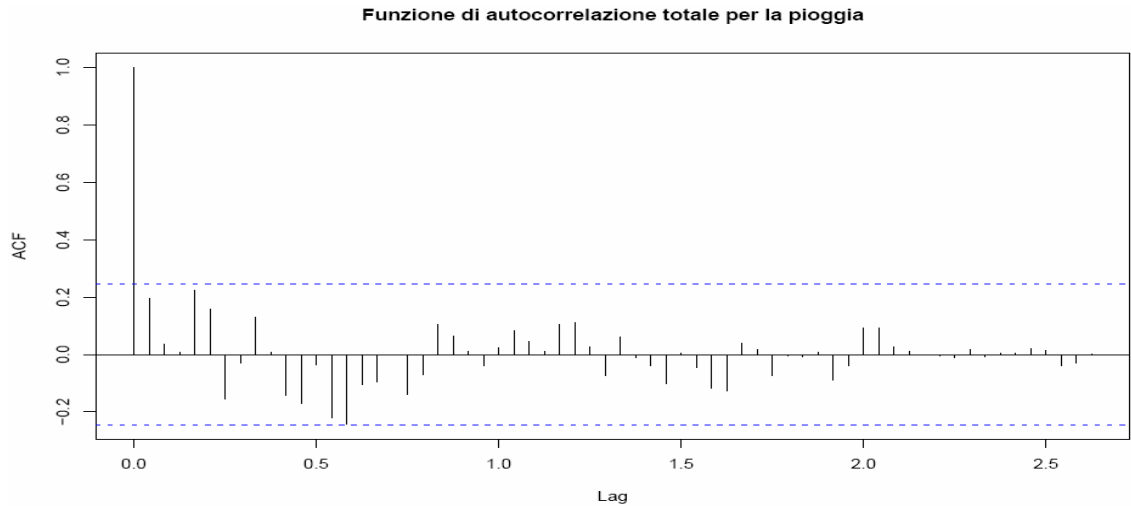
Nell'analisi che segue i valori di pioggia si riferiscono alle altezze di pioggia giornaliere, cumulate per un periodo quindicinale, presso la stazione di Rutigliano. La prima osservazione, associata al 1° dicembre 2001, rappresenta il totale di pioggia caduta, in millimetri, dal 16 al 30 novembre 2001 inclusi; la seconda, i valori cumulati dal 1° al 15 dicembre, e così via. Si osservi il *boxplot* della pioggia:



Si evince la presenza di tre valori anomali (*outliers*), classificati al di sopra degli 80 mm.

Osservando l'andamento temporale della serie si nota che due tra questi valori sono stati registrati nei mesi invernali del 2003, mentre il terzo si è verificato nell'estate del 2002.





Il correlogramma mostra che la serie è stazionaria: la funzione di autocorrelazione totale indica assenza di correlazioni significative a qualsiasi ritardo (*lag*). Ciò è confermato anche dai valori che restituisce il test di Ljung-Box calcolato a vari ritardi:

```
Box.test(rain, lag=1, type="L")
X-squared = 2.5307, df = 1, p-value = 0.1117

> Box.test(rain, lag=2, type="L")
X-squared = 2.614, df = 2, p-value = 0.2706

> Box.test(rain, lag=3, type="L")
X-squared = 2.6194, df = 3, p-value = 0.4541

> Box.test(rain, lag=12, type="L")
X-squared = 14.9694, df = 12, p-value = 0.2431

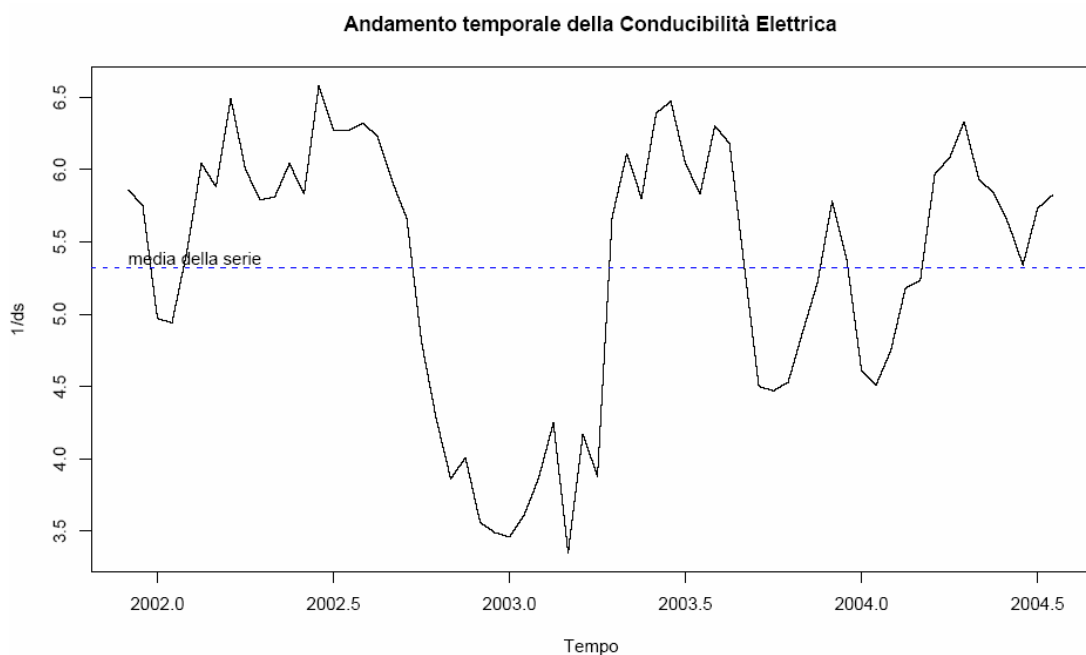
> Box.test(rain, lag=24, type="L")
X-squared = 29.608, df = 24, p-value = 0.1981
```

Si conclude, quindi, che la procedura di verifica sulle autocorrelazioni non porta a respingere l'ipotesi nulla che i 64 valori di pioggia siano generati da un *white noise*. Tutto ciò, relativamente al breve periodo di tempo considerato, la dice lunga su quanto possa essere difficile modellare il fenomeno pluviometrico, soprattutto nelle regioni mediterranee dove si manifesta molto variabile non mostrando alcuna costanza legata, quantomeno, al cambio delle stagioni.

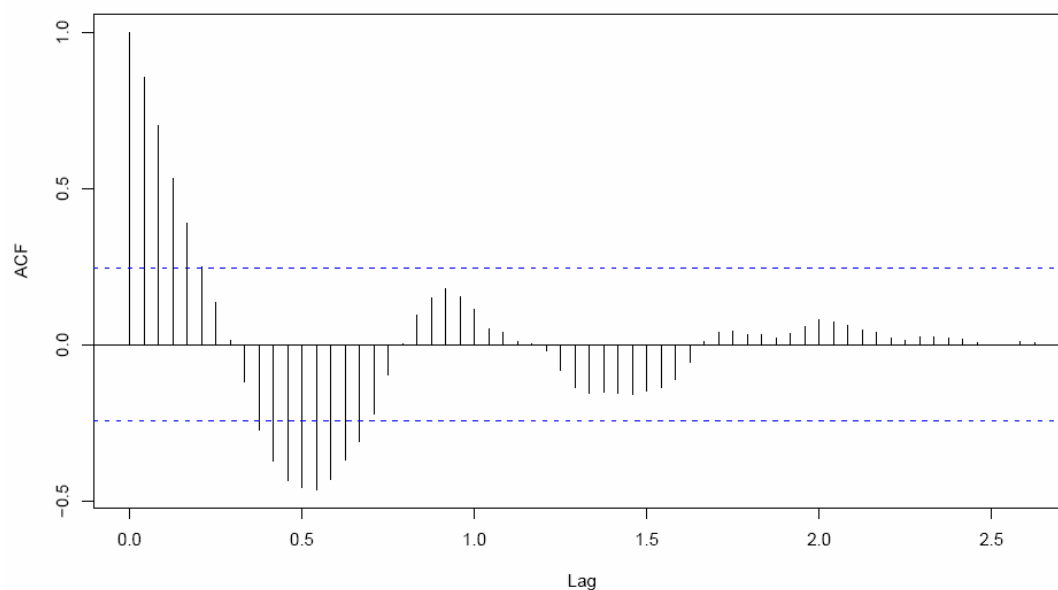
5.3 Analisi temporale della ECw

La rilevazione della ECw analizzata è stata effettuata presso la stazione di Mola con periodicità quindicinale. La prima osservazione è riferita al 1° dicembre 2001.

La serie temporale dei dati originali sarà indicata d'ora in poi con ECw.



Funzione di autocorrelazione totale per la ECW



In questo caso, la stima del correlogramma mostra una correlazione seriale tra le osservazioni; il processo non è assimilabile ad un “rumore bianco”. I primi coefficienti di autocorrelazione sono positivi e molto vicini ad uno. Il correlogramma poi si avvicina progressivamente, in maniera piuttosto lenta e regolare allo zero (senza, ad esempio, il “salto” che si era osservato nel caso della pioggia), e, continuando fino al lag 12 (6 mesi), diventa negativo, anche in questo caso progressivamente. Dal lag 12, sempre con un andamento molto lento e regolare, le correlazioni ritornano, progressivamente, ad essere positive intorno al ritardo 24 (1 anno). Occorre precisare che, oltre il sedicesimo ritardo non è rilevabile alcuna correlazione significativa.

L'onda sinusoidale che si smorza lentamente indica la presenza di una componente stagionale. Infatti, anche se non significative, le osservazioni distanti un anno o due presentano correlazioni positive, mentre quelle distanti sei mesi o un anno e mezzo mostrano della dipendenza negativa. L'andamento sinusoidale può essere spiegato osservando che, ad esempio, osservazioni riferite allo stesso periodo in anni diversi tendono a stare dalla stessa parte rispetto alla media di tutte le osservazioni e, quindi, che quasi tutti gli addendi che entrano nel calcolo dell'autocovarianza ai vari ritardi hanno un "segno prevalente" facilmente determinabile e anche prevedibile.

5.4 Scelta del modello di aggregazione

Il primo problema che si presenta riguarda la scelta tra modello moltiplicativo e modello additivo. Non sempre tale scelta è possibile in termini netti³⁶.

Per quanto riguarda il caso esaminato, si è scelto il modello moltiplicativo.

³⁶ MUTTARINI L., *Aspetti di una scomposizione di serie storiche demografiche*, Relazione al Convegno Nazionale di studio sull' "Analisi Moderna delle Serie Storiche" (a cura di Piccolo D.), Napoli, 19-22 maggio 1981, Franco Angeli, Milano, 1983.

5.5 Scelta di uno stimatore del trend

Al fine di evidenziare la componente tendenziale si è proceduto dapprima alla eliminazione della componente stagionale. Poiché tale componente presenta, nel caso in questione, un andamento con un periodo di 24 termini, si è applicata ai dati originari una media mobile ponderata a 25 termini in modo da eliminare la stagionalità e mettere in risalto solo la componente di fondo. Per il generico termine:

$$y_t^{**} = \frac{1}{24} \left(\frac{1}{2} y_{t-12} + \dots + y_{t-3} + y_{t-2} + y_{t-1} + y_t + y_{t+1} + y_{t+2} + y_{t+3} + \dots + \frac{1}{2} y_{t+12} \right)$$

Ciò è stato automatizzato attraverso il comando *sfilter* (*ECw*) messo a disposizione dal *package ast*.

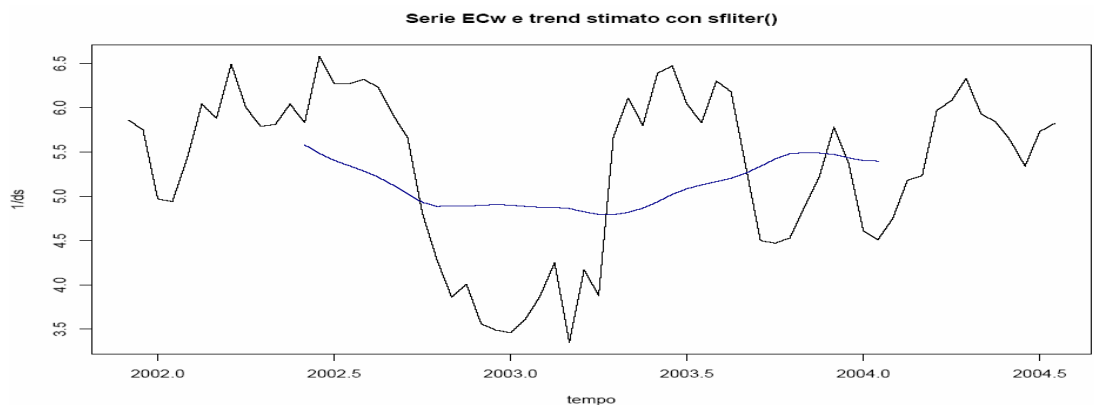
Di seguito, è riportato il comando inserito nel software e l'output relativo:

```
ECw.pre<-sfilter(ECw)
ECw.pre->Time Series: Start = c(2001, 23); End = c(2004, 14); Frequency = 24

[1]      NA      NA      NA      NA      NA      NA      NA      NA      NA
[9]      NA      NA      NA      NA  5.578750  5.483750  5.405208  5.346042
[17] 5.285833 5.216042 5.126042 5.025000 4.932500 4.885833 4.889583 4.890833
[25] 4.897500 4.906875 4.899792 4.885833 4.876250 4.874792 4.861667 4.825417
[33] 4.794167 4.792292 4.818750 4.865208 4.936667 5.022083 5.085208 5.127917
[41] 5.165000 5.202708 5.261250 5.337917 5.421250 5.480833 5.490833 5.487917
[49] 5.472917 5.433542 5.403542 5.396875      NA      NA      NA      NA
[57]      NA      NA      NA      NA      NA      NA      NA      NA      NA
```

Come si evince mancano i primi e gli ultimi 12 termini.

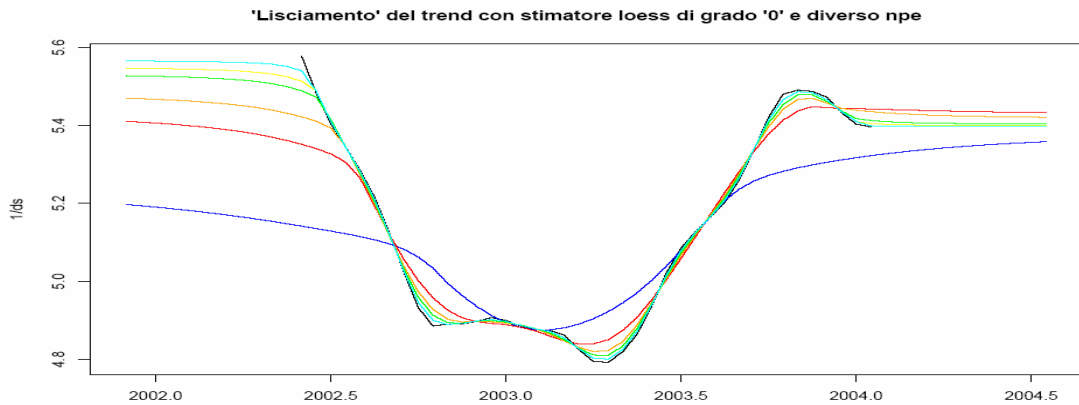
A seguire, viene rappresentata la serie temporale e il trend stimato (linea blu) con le medie mobili tramite *sfilter*:



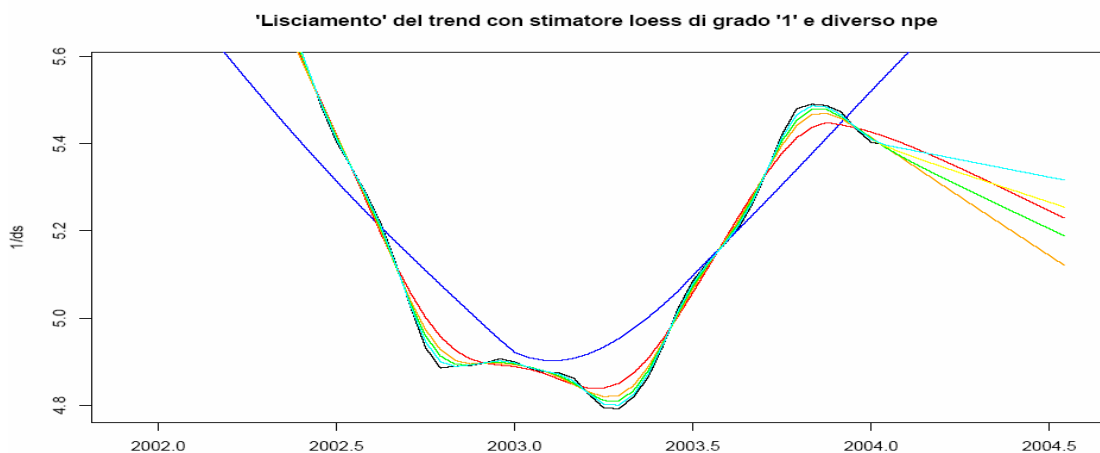
Appare evidente che la stima del trend non può essere effettuata utilizzando polinomi di primo o di secondo grado. Si è ritenuto indispensabile, quindi, utilizzare un approccio flessibile basato su procedure di lisciamiento. In particolare, si è utilizzato lo stimatore *loess*.

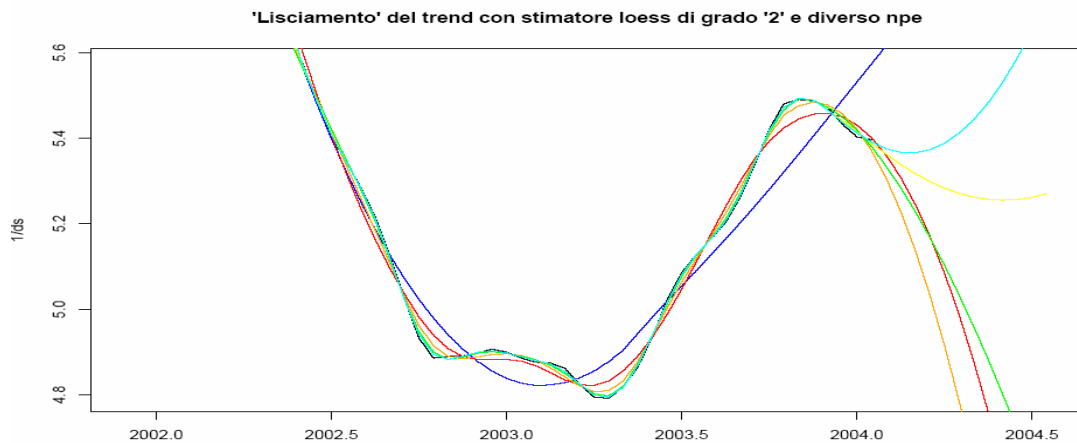
La scelta del grado di lisciamiento e del numero di parametri equivalenti è stata effettuata con un procedimento a prova di errore.

Di seguito sono riportati alcuni esempi di *smoothing* applicati al caso in questione. Le linee blu indicano un lisciamiento con 5 parametri equivalenti, quelle rosse con 10; le arancione sono associate a 15 parametri; le verdi a 20, le gialle a 25 e le celeste a 30.



Guardando al grafico sopra indicato, si nota come, agli estremi, le curve “perdano” la tendenza stimata. Ciò è legato al fatto che lo stimatore loess di grado 0 si basa su una media locale e, quindi, non riesce a prolungare la linea osservata. Utilizzando, invece, una regressione locale con delle rette (loess di grado 1) e numero di parametri equivalenti maggiore di 10 si nota un migliore adattamento alla curva tendenziale, ma la linea osservata sembra essere prolungata in maniera troppo rigida.





Lo stimatore loess di grado 2, basato su una regressione locale con un polinomio di secondo grado, restituisce risultati molto diversi secondo il numero di parametri equivalenti scelti; sembrerebbe però che, nel caso in cui si scelgano 25 parametri equivalenti, ci sia una buona capacità della curva stimata di descrivere i dati.

5.6 Scelta di uno stimatore per la componente stagionale

Attraverso il comando³⁷ *smoothts()*:

```
>ECw.ptrend<-smoothts(ECw.pre~lo(2,25))
```

si ottiene una stima della componente di trend per la serie in esame.

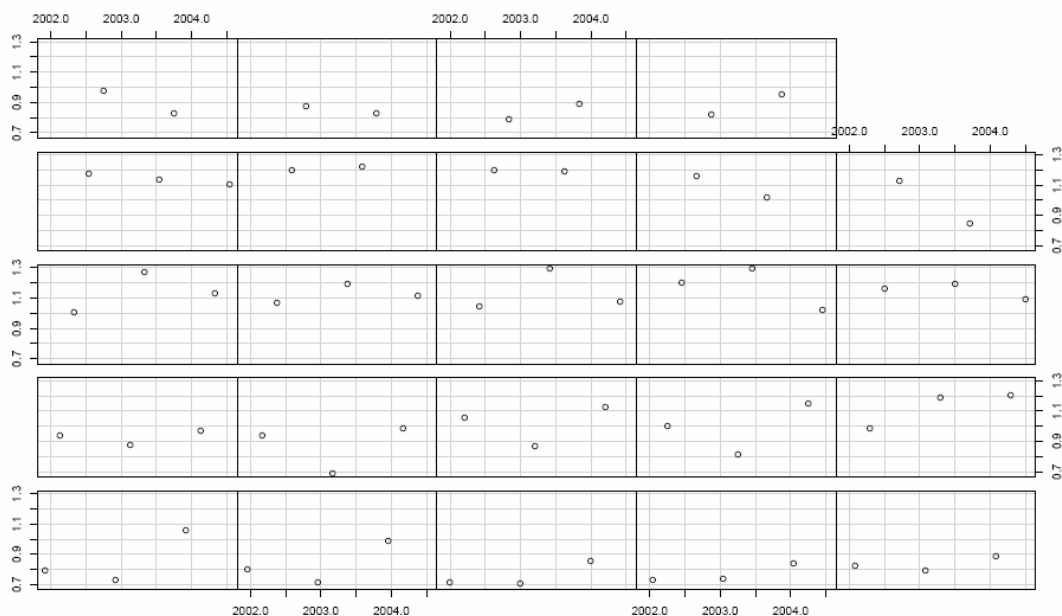
³⁷ maggiori chiarimenti sulla funzione *smoothts()* sono riportate in TAVOLA A2 dell'Appendice

Avendo ipotizzato un modello di aggregazione delle componenti di tipo moltiplicativo,

```
>ECw.ps<-ECw/ECw.ptrend
```

rappresenta una versione della serie da cui è stato, preliminarmente, eliminato il trend.

Di seguito, sono riportate le sotto-serie stagionali da stimare:

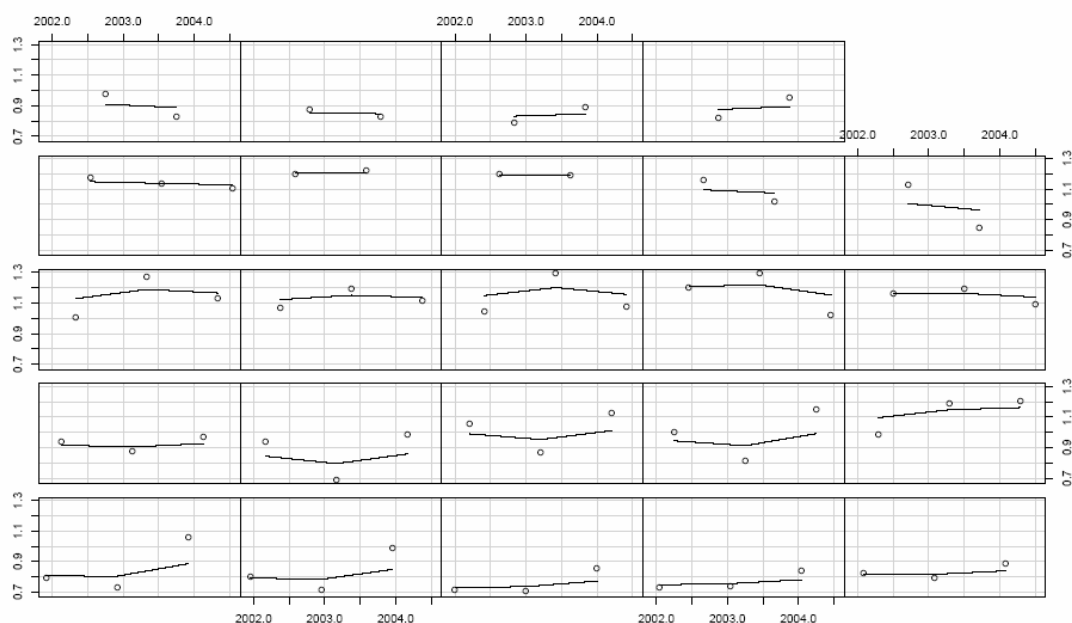


Il grafico si legge da sinistra a destra, dal basso verso l'alto.

La stagionalità, relativamente al breve periodo considerato, varia nel tempo. Ad esempio, il primo pannello in basso a sinistra, concernente le rilevazioni effettuate al primo dicembre 2001, 2002 e 2003, aumenta negli anni passando da un valore inferiore ad uno (che nel caso di stagionalità moltiplicativa indica un valore “sotto la media”) ad un

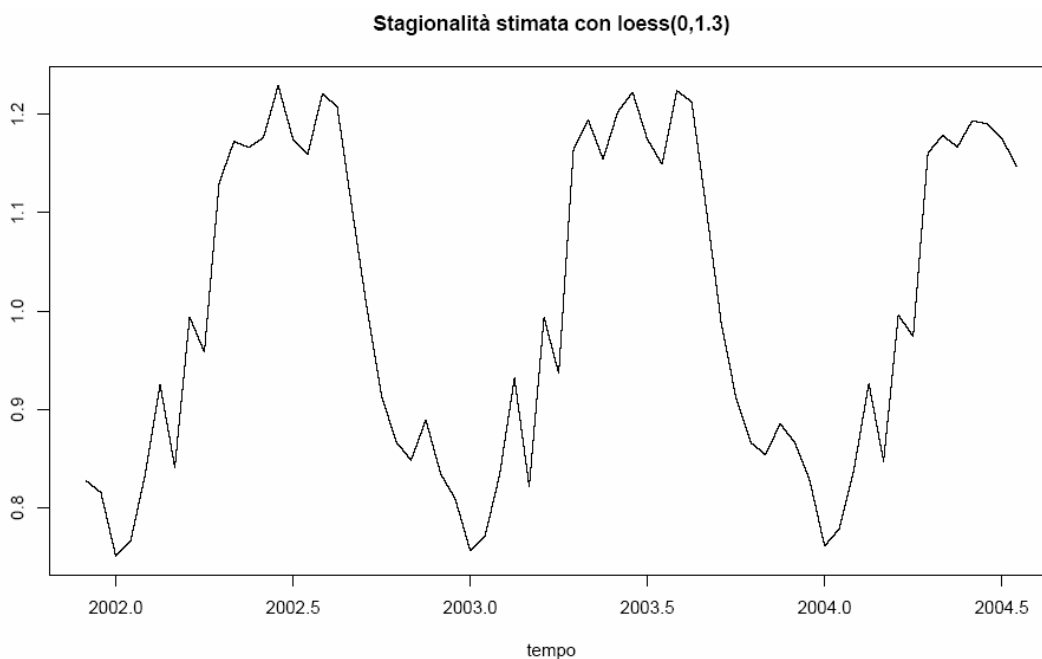
valore superiore ad uno (ovvero “sopra la media”). Analoghe variazioni si vedono in altri pannelli.

Anche in questo caso, per “interpolare” ciascuna sotto-serie quindicinale in maniera adeguata si è utilizzato uno stimatore di tipo loess basato su una regressione di ordine zero e 1.3 parametri equivalenti. Di seguito se ne riporta la rappresentazione grafica:



La scelta del grado di lisciamento, in questo caso, è stata limitata dal numero di osservazioni presenti per ogni sotto serie quindicinale. Utilizzando, infatti, un loess con 2 parametri equivalenti si sarebbe sovrastimata la stagionalità nei sottoperiodi che presentano solo 2 osservazioni. È bene ricordare che i dati della precedente configurazione sono da riferirsi non solo alla componente stagionale, ma anche a quella erratica; un modello troppo flessibile rischierebbe di cogliere anche una

parte dei residui, che potrebbero assumere un comportamento “antistagionale”. Di seguito, viene riportata la stima della componente stagionale ottenuta con lo stimatore loess:



Si osservi che, avendo scelto un modello di aggregazione di tipo moltiplicativo, i valori della stagionalità stimata sono numeri puri, da interpretarsi, quindi, come variazioni dal livello della serie secondo che siano maggiori o minori di uno.

5.7 Stima del modello

L'analisi sperimentale di stima delle componenti suggerisce di adottare un modello del tipo³⁸:

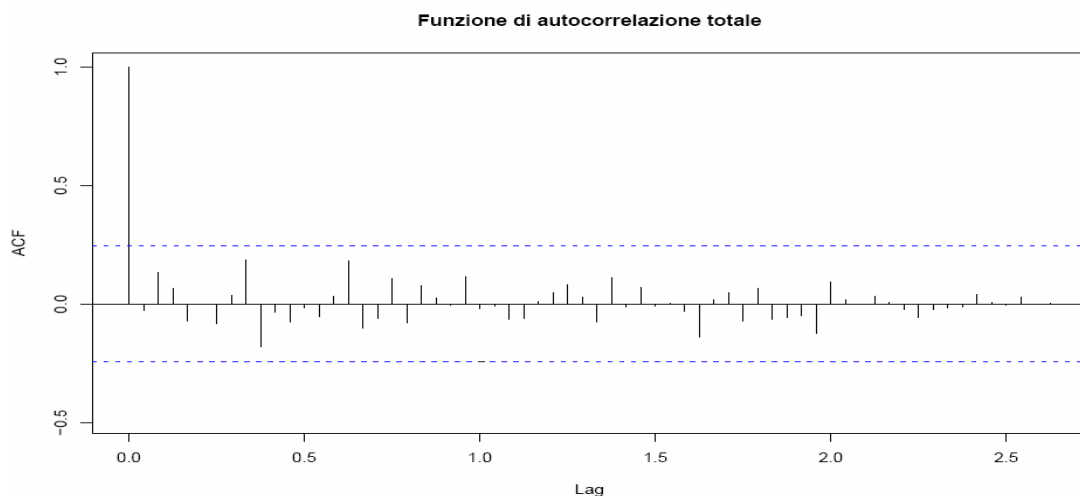
```
m<-tsr(ECw~lo(2,25)*lo(0,1.3))
```

Le componenti stimate (vedi TAVOLA A4 in Appendice) sono tutte positive.

Ciò consente, se necessario, di poter ricondurre il modello d'aggregazione stimato da moltiplicativo ad additivo applicando una semplice trasformazione in scala logaritmica. È opportuno, però, analizzare prima i residui del modello.

- **Test di autocorrelazione**

Si osservi il correlogramma dei residui del modello stimato:



³⁸ maggiori approfondimenti sul comando `tsr` sono riportati in TAVOLA A3 dell'Appendice.

I coefficienti di autocorrelazione dei residui, calcolati per tutti i ritardi, si dispongono tutti all'interno delle bande di confidenza (ad un livello del 95%) e seguono un andamento del tutto casuale, indicando quindi assenza di correlazione seriale.

È utile approfondire l'analisi dei residui con opportuni test statistici, meglio noti, come test di specificazione del modello. Il venire meno di alcune delle ipotesi che si testeranno potrebbe, infatti, inficiare la validità del modello adottato.

Prima di procedere nella verifica dei test, la letteratura consiglia di operare sui logaritmi dei residui del modello stimato³⁹.

Con la trasformazione, però, i residui sarebbero espressi nella stessa unità di misura della variabile in esame, per cui si è ritenuto adeguato operare una standardizzazione per avere a che fare con numeri puri.

- **Test di Box-Pierce:**

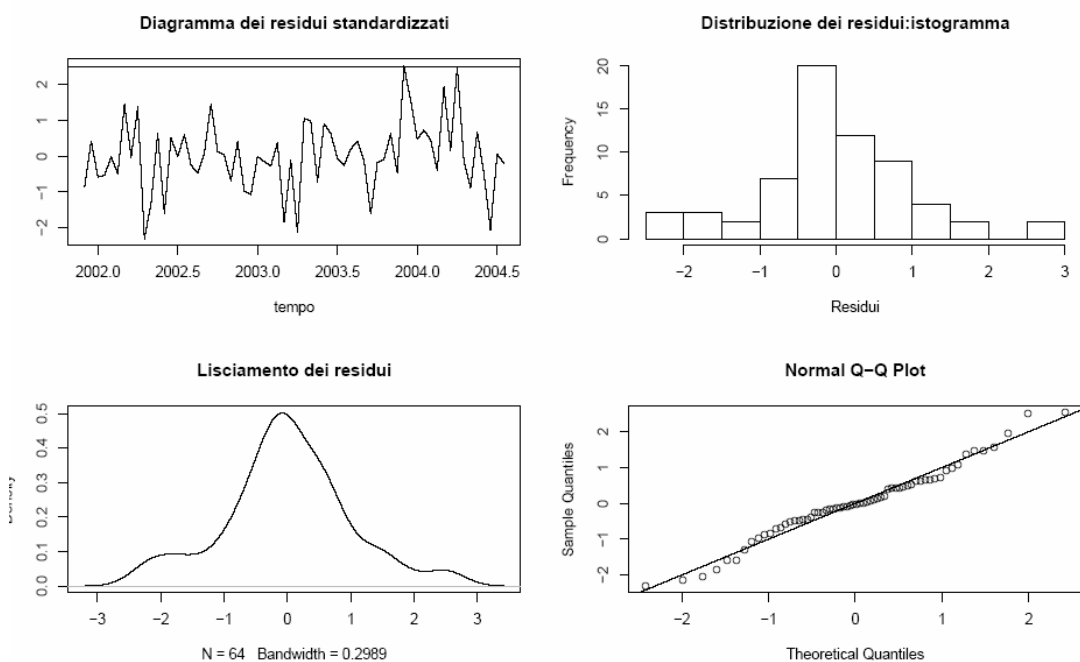
```
Box.test(log.res.st,lag=1)
X-squared = 0.0595, df = 1, p-value = 0.8073

> Box.test(log.res.st,lag=24)
X-squared = 28.0129, df = 24, p-value = 0.2595
```

³⁹ DI FONZO T., LISI F., Complementi di statistica economica,ecc., op.cit..

Il risultato dei test non consente di rifiutare l'ipotesi nulla di incorrelazione dei residui in quanto il livello di significatività osservato (p-value) è maggiore di quello assunto pari a 0,05. Si conclude che tra i residui del modello non esiste alcuna correlazione seriale e quindi può supporre che agli errori sottenda un processo stocastico di tipo *white noise*.

- **Test di normalità degli errori**



Dal diagramma (il primo in alto a sinistra) emerge che due sole osservazioni (15 novembre 2003 e 15 marzo 2004) sono al limite dall'essere considerate anomale, poiché prossime alla banda di confidenza del 99% (banda compresa tra -2,5 e +2,5). D'altra parte, non

c'è da stupirsi se si considera la variabilità insita nel fenomeno allo studio. I restanti grafici, mostrano una buona indicazione per una probabile distribuzione pressoché normale dei residui. Per avere un risultato statisticamente più affidabile, si è effettuato il test di Shapiro-Wilk:

```
shapiro.test(log.res.st)
      Shapiro-Wilk normality test
data:  log.res.st
W = 0.9753, p-value = 0.2258
```

Il p-value è concretamente elevato rispetto ai livelli di significatività a cui di solito si fa riferimento: ciò fa propendere per l'ipotesi nulla, ovvero, la normalità della distribuzione degli errori.

5.8 I risultati dell'analisi sulla serie della ECw

L'analisi dei residui stimati permette di verificare l'adeguatezza del modello scelto per stimare le componenti deterministiche e, quindi, di giustificare l'uso ai fini previsivi. La verifica ha fornito esito positivo, ovvero è lecito assumere che la serie dei residui sia stata generata da un *white noise*, che è anche *gaussiano*.

Una misura della bontà di adattamento della parte della serie osservata spiegata dal trend e dalla stagionalità è stata calcolata utilizzando il coefficiente di determinazione R^2 calcolato come segue:

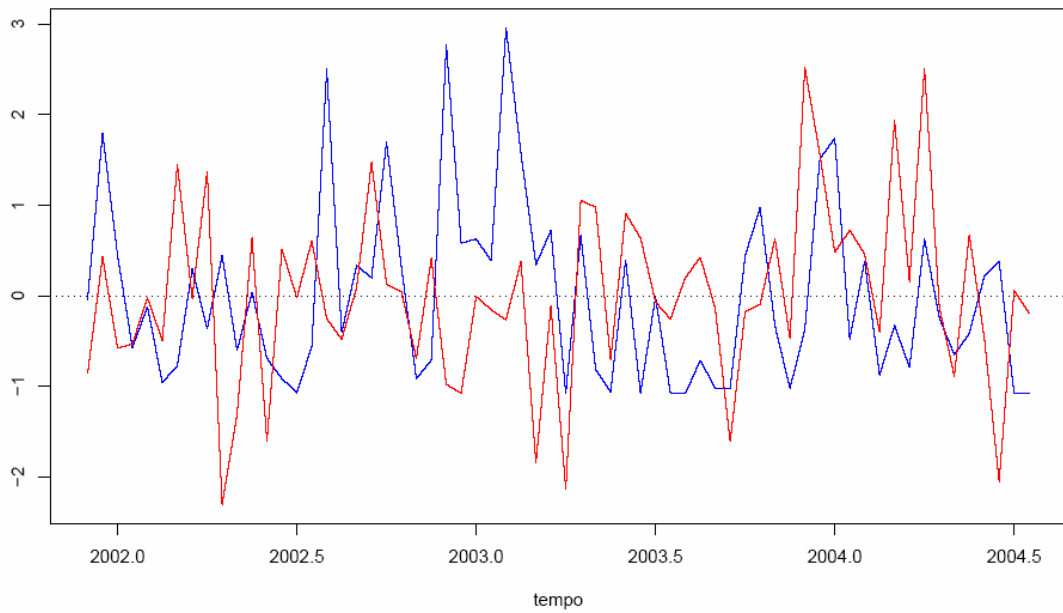
```
squareR<-1-var(ECw-fitted(m))/var(ECw)
squareR
[1] 0.9702803
```

Con riferimento ai risultati ottenuti (analisi dei residui e bontà di adattamento ai dati) la scomposizione della serie può ritenersi globalmente accettabile almeno per scopi descrittivi. L'estensione a fini previsivi potrebbe essere un po' azzardata considerando sia il periodo di osservazione del fenomeno, di soli 3 anni, sia la variabilità che lo stesso presenta nel corso del tempo.

5.9 Le relazioni di dipendenza tra la pioggia e la ECw

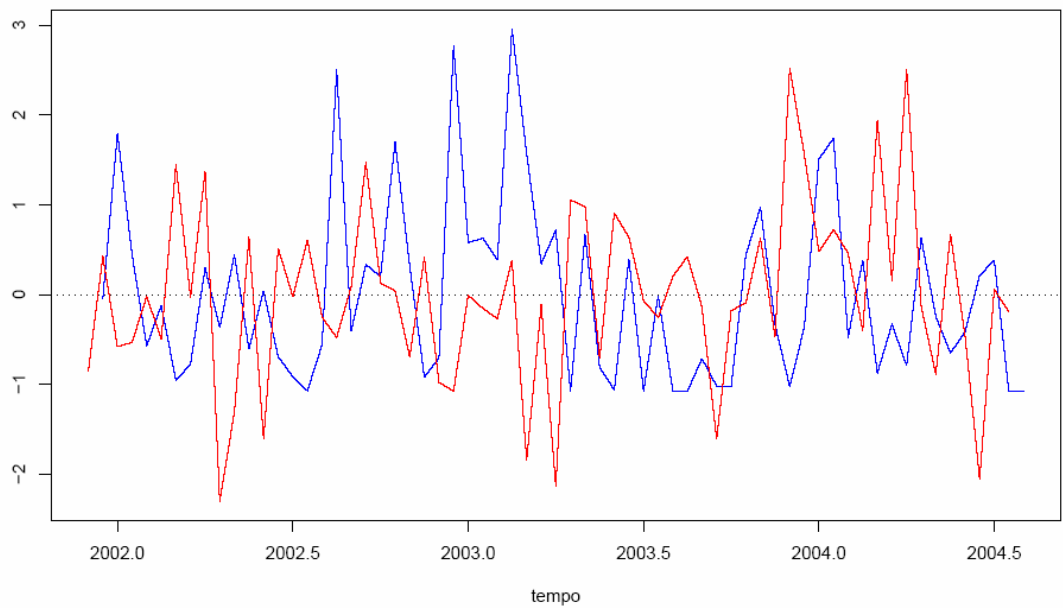
Avendo trovato che la componente erratica nel modello di aggregazione sia un *white noise* e avendo valutato l'andamento della pioggia come assimilabile ad un processo dello stesso tipo, ovvero, stocastico puramente casuale, si sono confrontate le due serie di dati. Nelle figura seguente, vengono riportati gli andamenti delle due variabili, preliminarmente standardizzate per renderle omogenee.

Andamento della pioggia(linea blu) e dei residui del modello m (linea rossa)

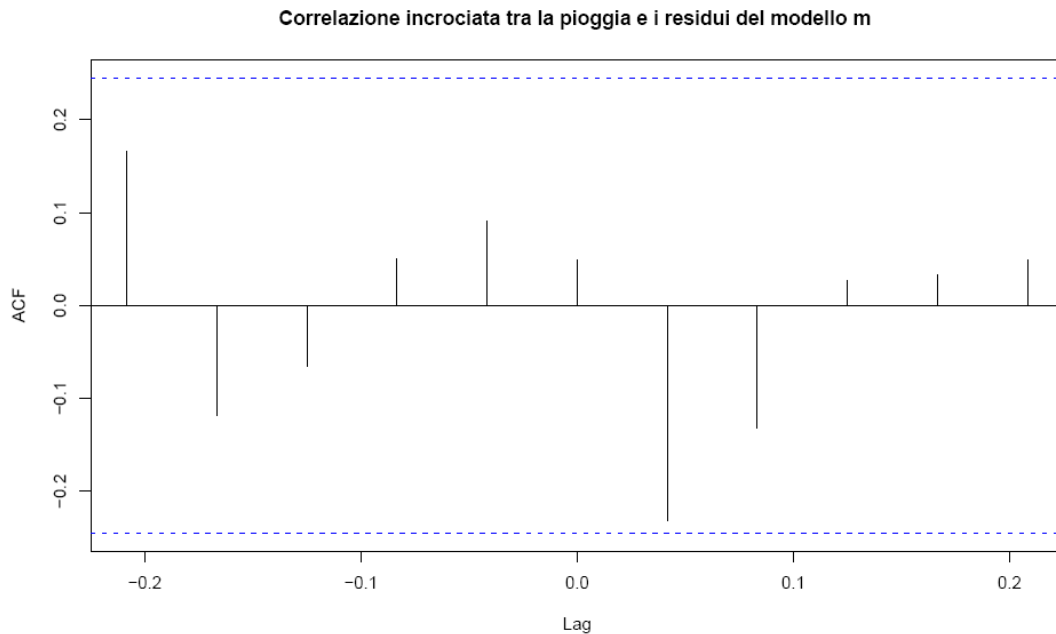


Un'attenta analisi grafica mostra come gli andamenti delle due serie, pur essendo state definite casuali, manifestano delle evidenti asimmetrie. Ciò è ancor più evidente se si considera la pioggia con uno sfasamento temporale di un'unità (quindici giorni nel caso in esame).

Andamento della pioggia(linea blu)ritardata e dei residui del modello m (linea rossa)



Anche la funzione di correlazione incrociata mostra una correlazione negativa tra i residui del modello d'aggrgazione stimato per la ECw valutati al tempo t e la pioggia valutata ad un ritardo (lag=1).



Tale correlazione risulta al limite della significatività indicata dalle bande di confidenza (linee blu), che, come ampiamente illustrato nel paragrafo 4.12 del capitolo precedente, presentano varie limitazioni. Inoltre, nelle stesse condizioni, qualche osservazione in più avrebbe ridotto l'ampiezza dell'intervallo tra le bande e la correlazione individuata sarebbe risultata graficamente significativa.

Ipotizzare una relazione di dipendenza tra l'evento meteorico e la Conducibilità Elettrica può essere testata valutandola attraverso un modello lineare di regressione.

Il modello di regressione classico $Y_t = c + X_t + \varepsilon_t$ consente di modellare una dipendenza (lineare) contemporanea, ma non dinamica: X_t potrebbe influenzare Y_t , oltre che al tempo t , anche al tempo $t+1$, $t+2, \dots$

Per cogliere una dipendenza (lineare) dinamica si può considerare un modello del tipo:

$$Y_t = c + X_t + X_{t-1} + X_{t-2} + \dots + X_{t-n} + \varepsilon_t$$

Ovviamente, non si può trascurare che la Conducibilità Elettrica al tempo t dipenda da quella al tempo $t-1$, $t-2$, L'analisi della serie storica ha, infatti, mostrato come l'ECw sia non stazionaria, ovvero, abbia 'memoria' del passato.

Dopo aver opportunamente standardizzato le variabili, indicando con:

r.t.st : variabile pioggia nel periodo 1/01/2002 – 15/07/2004;

r.t1.st : variabile pioggia nel periodo 15/12/2001 – 1/07/2004;

r.t2.st : variabile pioggia nel periodo 1/12/2001 – 15/06/2004;

e con:

ECw.t.st : variabile ECw nel periodo 1/01/2002 – 15/07/2004;

ECw.t1.st : variabile ECw nel periodo 15/12/2001 – 1/07/2004;

ECw.t2.st : variabile ECw nel periodo 1/12/2001 – 15/06/2004;

si è utilizzato, quindi, un modello lineare di regressione, considerando *ECw.t.st* come variabile indipendente (o risposta) e le rimanenti come variabili esplicative.

Di seguito, si espongono i risultati dell'analisi del modello di regressione⁴⁰ valutato con tutte le variabili predette:

```

modello<-lm(ECw.t.st~ECw.t1.st+ECw.t2.st+r.t.st+r.t1.st+r.t2.st)
> summary(modello)

lm(formula = ECw.t.st ~ ECw.t1.st + ECw.t2.st + r.t.st + r.t1.st +
    r.t2.st)
Residuals:
    Min       1Q   Median       3Q      Max
-1.1404 -0.2889 -0.0284  0.3142  1.4896

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.725e-16   6.182e-02  2.79e-15   1.0000
ECw.t1.st    8.197e-01   1.345e-01   6.093 1.07e-07 ***
ECw.t2.st   -1.048e-01   1.229e-01  -0.852   0.3977
r.t.st       -9.466e-02   6.741e-02  -1.404   0.1658
r.t1.st     -1.633e-01   6.695e-02  -2.439   0.0179 *
r.t2.st     -9.023e-02   7.086e-02  -1.273   0.2081

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4867 on 56 degrees of freedom
Multiple R-Squared:  0.7825,    Adjusted R-squared:  0.7631
F-statistic:  40.3 on 5 and 56 DF,  p-value: < 2.2e-16

```

⁴⁰ cfr. TAVOLA A5 dell'Appendice per maggiori dettagli sulla funzione `lm()`.

e, quello con le sole variabili risultate, precedentemente, significative:

```
modellol<-lm(ECw.t.st~ -1 + ECw.tl.st+r.tl.st)

>summary(modellol)

lm(formula = ECw.t.st ~ -1 + ECw.tl.st + r.tl.st)

Residuals:
    Min       1Q   Median       3Q      Max
-1.08712 -0.29426 -0.05551  0.38525  1.44666

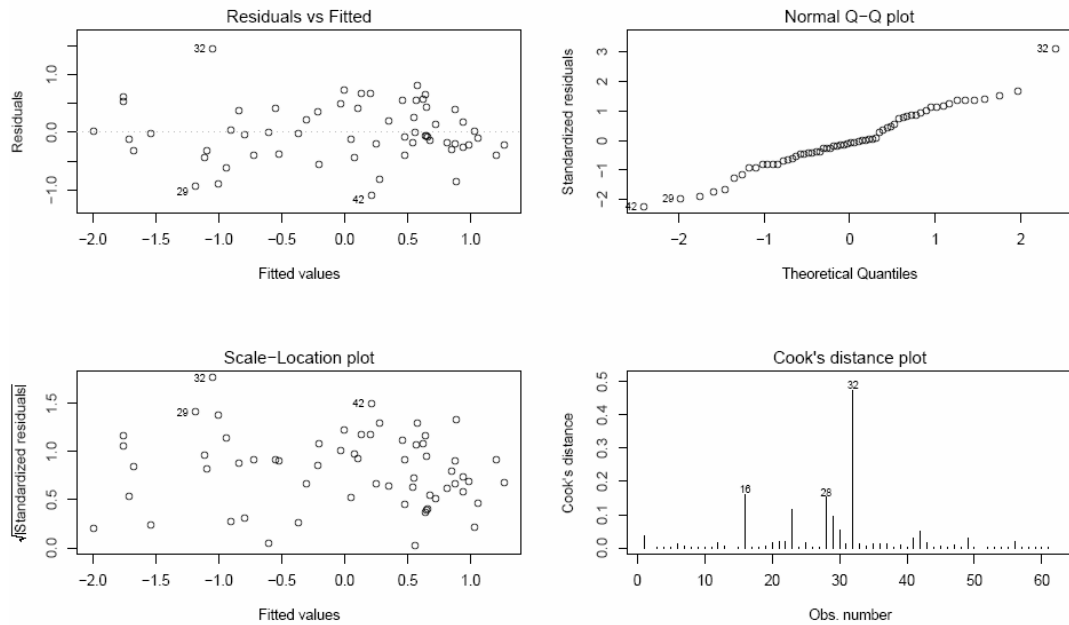
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
ECw.tl.st  0.79872     0.06672  11.971  <2e-16 ***
r.tl.st   -0.17393     0.06672  -2.607  0.0115 *

Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4877 on 60 degrees of freedom
Multiple R-Squared:  0.766,    Adjusted R-squared:  0.7582
F-statistic: 98.21 on 2 and 60 DF,  p-value: < 2.2e-16
```

Dai risultati dell'analisi, soltanto due fra i regressori inseriti nel modello presentano un coefficiente statisticamente significativo, ovvero, diverso da zero: la ECw valutata al tempo $t-1$, con un coefficiente positivo molto alto (0,79872), e la pioggia valutata al tempo $t-1$ con un coefficiente negativo, di misura pari a $-0,17393$. L' R^2 'aggiustato' indica che i regressori inseriti nel modello spiegano circa i $3/4$ della variabilità

legata alla variabile indipendente. Prima di convalidare i risultati ottenuti occorre analizzare i residui.



L'analisi grafica consente di convalidare il modello adottato accreditando la forma funzionale scelta per spiegare la dipendenza e l'assunzione sulla normalità della distribuzione dei residui (i punti giacciono in prossimità della linea q-q). Dal diagramma a punti dei residui (il primo in alto a sinistra), in particolare, si rileva come i punti si dispongano in una fascia di ampiezza costante, parallela all'asse delle ascisse. Ciò dimostra che è corretto ipotizzare una relazione lineare tra le variabili del modello. Inoltre, sembrerebbe che non siano state omesse covariate significative alla spiegazione della variabile risposta.

CAPITOLO 6

CONSIDERAZIONI CONCLUSIVE

Il lavoro di tesi ha riguardato l'analisi dei parametri qualitativi delle acque sotterranee presenti in Puglia ed utilizzate, prevalentemente, per scopi irrigui.

Dopo aver descritto gli aspetti territoriali e ambientali della Puglia, si è passati a sottolineare la scarsità delle risorse idriche della regione. La Puglia, infatti, manca di una vera e propria idrografia superficiale e le precipitazioni sono scarse e con frequenza (numero di giorni di pioggia in un anno) piuttosto bassa.

Tuttavia, la maggior parte delle attività agricole della regione sono legate, in misura crescente, ai sistemi colturali irrigui che garantiscono un reddito più elevato. Per far fronte alla carenza di risorse idriche, gli agricoltori continuano a scavare pozzi, spesso in modo abusivo.

I continui e talvolta illeciti prelievi provocano il degrado delle risorse idriche sotterranee, da una parte perché essi superano la quantità

di acqua di precipitazione che alimenta le falde, dall'altra perché, in prossimità delle coste, si verificano spesso fenomeni di intrusione di acqua di mare.

Le informazioni sul fenomeno di degrado delle acque sotterranee in Puglia non sono comprovate da un monitoraggio sistematico sul territorio. Le prime fonti di dati sono state prodotte dall'Istituto Sperimentale Agronomico del Consiglio per la Ricerca e sperimentazione in Agricoltura (C.R.A.) di Bari. Si tratta di rilevazioni eseguite ad intervalli regolari (ogni quindici giorni) su una ventina di pozzi distribuiti sulla superficie regionale. Tale monitoraggio si prefiggeva l'obiettivo di quantificare il rischio di salinizzazione (o di degrado in genere) delle acque di falda, in una situazione agricola assai delicata, perché basata su colture irrigue che utilizzano una risorsa, erroneamente, ritenuta "rinnovabile". Si prevede che questa situazione, in futuro, diventerà ancora più drammatica a causa dei cambiamenti climatici ipotizzati per l'Italia mediterranea.

Dall'analisi dei primi dati risulta che le falde più interne, collocate ad una profondità elevata, superiore ai 100 m, non subiscono pesanti cambiamenti nel tempo, circa i livelli di salinità. Al contrario, forti variazioni si rilevano, analizzando le falde più superficiali e situate in prossimità della costa.

Lo studio si è, quindi, soffermato su due falde spazialmente vicine: quella di Rutigliano, più profonda (200 m) e interna e quella di Mola, più superficiale (20 m) e vicino alla costa. Per queste caratteristiche i due pozzi hanno presentato degli indici di variabilità relativi molto contrastanti, pur appartenendo alla stessa idrostruttura delle Murge.

Prendendo in considerazione l'evoluzione temporale della Conducibilità Elettrica dell'acqua (indice agronomico della salinità) nel periodo che va da dicembre 2001 a luglio 2004, si sono osservate oscillazioni legate alla stagionalità: i valori più bassi si sono registrati in inverno e quelli più elevati in estate.

Parallelamente, sono stati esaminati i valori di pioggia ottenuti cumulando le altezze giornaliere di pioggia registrate tra una rilevazione del parametro indice di qualità e quella successiva (ogni quindici giorni).

In prima analisi, dal confronto degli andamenti temporali delle due variabili (opportunamente standardizzate), si è rilevata una possibile relazione di causa-effetto tra i due fenomeni: periodi di pioggia a elevata intensità sono accompagnati da una diminuzione del livello di salinità delle acque di falda, e, viceversa, periodi di incremento del valore di Conducibilità Elettrica sono associati a periodi di quasi siccità.

Lo studio di un'eventuale relazione di dipendenza tra le variabili non poteva basarsi sul coefficiente di correlazione lineare di Bravais-

Pearson, in quanto avrebbe fornito una misura ‘contemporanea’, ovvero valutata nello stesso istante temporale.

Ai fini della rassomiglianza fra serie storiche, occorre tener conto anche degli sfasamenti temporali. Le relazioni fra due serie storiche sono valutate abitualmente mediante la funzione di correlazione incrociata (*cross-correlation*), la significatività dei coefficienti della quale è verificata solo nell’ipotesi che le due serie siano generate da un processo stocastico puramente aleatorio (*white noise*).

L’analisi condotta sulle precipitazioni ha portato ad accettare l’ipotesi di stazionarietà; d’altra parte, l’evento meteorologico è difficilissimo da modellare, soprattutto in ambiente mediterraneo e in fase di cambiamenti climatici caratterizzati da una crescente variabilità del fenomeno.

Al contrario, l’analisi condotta sulla conducibilità elettrica ha evidenziato una non stazionarietà della serie. Al fine di valutare la significatività dei coefficienti di correlazione incrociata, si è resa necessaria un’operazione di filtraggio della serie, in modo da trasformarla in una successione di residui per i quali siano accettabili le ipotesi poste.

Il criterio impiegato è stato quello di adattare ai valori della serie un opportuno modello. L'individuazione del modello è stata conseguita tramite l'approccio classico delle serie storiche.

In questo caso si è adottato un modello di aggregazione di tipo moltiplicativo e le componenti (trend e stagionalità) sono state stimate attraverso opportune procedure di lisciamiento. La componente irregolare del modello stimato è risultata assimilabile ad un *white noise gaussiano*.

Dai coefficienti di correlazione incrociati, calcolati tra i residui del modello di scomposizione della Conducibilità Elettrica e i valori di precipitazione, è emersa una relazione di dipendenza negativa tra il parametro di salinità valutato al tempo t e la pioggia riferita al tempo $t-1$.

Ciò ha spinto l'analisi ad individuare un modello di regressione lineare, capace di spiegare la dipendenza tra la Conducibilità Elettrica, valutata al tempo t (variabile risposta), e una serie di variabili esplicative, riguardanti lo stesso parametro e le altezze di pioggia quindicinali riferite a diversi istanti temporali. L'analisi del modello adottato ha portato alla conclusione che il livello di salinità al tempo t dipende, positivamente, dal livello di salinità rilevato in precedenza, negativamente, dalle altezze di pioggia caduta rilevate al tempo $t-1$. Detto in altri termini, si può pensare che la Conducibilità Elettrica della falda di Mola sia influenzata, oltre che dal livello di salinità precedente, anche dalle precipitazioni che

si verificano a Rutigliano nel periodo che va dai 15 ai 30 giorni antecedenti la rilevazione. Le acque meteoriche che alimentano la falda non vi giungono direttamente, ma molto probabilmente scorrono nel sottosuolo fino a raggiungere lo strato impermeabile della falda.

In conclusione, il modello di analisi adottato per Mola potrebbe essere proposto per le altre falde della Puglia, al fine di una sua validazione. Dalle calibrazioni locali, si potrebbe, infatti, ipotizzare un modello “territoriale” che consenta di prevedere il comportamento della qualità delle acque di falda, a partire dalle sequenze di precipitazione, e di evidenziare le anomalie legate allo sfruttamento abusivo delle acque sotterranee. Le previsioni, se usate in modo appropriato, potranno servire a prevenire e scongiurare eventi disastrosi per l’ambiente e per l’economia della regione, nell’ottica di uno sviluppo sostenibile, tanto più necessario in quanto il prezzo da pagare, in termini di rischio di salinizzazione e desertificazione, è altissimo.

APPENDICE

TAVOLA A1 - Script della funzione *stand* utilizzata in R per standardizzare le variabili

```
> stand<-function(x){m=mean(x)
+   s=(var(x)^0.5)
+   z=(x-m)/s
+   return(z)}
```

TAVOLA A2 - Sulla funzione *smoothts* per ‘lisciare’ una serie temporale

La funzione `smoothts()` permette di ‘lisciare’ una serie temporale in una varietà di modi. La sua sintassi è:

```
(output)<-smoothts((input)~("lisciatore"))
```

dove `input` è la serie temporale univariata che si vuole ‘lisciare’ e `output` è la serie storica ‘lisciata’.

I ‘lisciatori’, ovvero gli stimatori, riconosciuti sono:

-**constant**: la serie ‘lisciata’ assume valore uguale alla media della serie osservata per ogni istante di tempo;

-**poly(r)**: la serie osservata viene interpolata con un polinomio di grado `r`;

-**loess(r,g)**: la serie osservata è ‘lisciata’ utilizzando una regressione locale di tipo loess; `r` è il grado del polinomio utilizzato; `g` è (approssimativamente) il numero di parametri equivalenti desiderati;

-**gauss(r,g)**: la serie osservata è ‘lisciata’ utilizzando una regressione locale con pesi gaussiani (ovvero la funzione peso è la densità di una normale di media nulla); `r` è il grado del polinomio utilizzato; `g` è (approssimativamente) il numero di parametri equivalenti desiderati;

-**spline(g)**: la serie osservata è ‘lisciata’ utilizzando una spline con (approssimativamente) `g` parametri equivalenti.

I ‘lisciatori’ da utilizzare possono essere tranquillamente abbreviati (`c`, `p`, `lo`, `g`, `s`).

TAVOLA A3 - Sulla funzione *tsr* per la decomposizione di una serie storica

La libreria *ast* contiene la funzione *tsr* che permette di stimare una serie di modelli del tipo: (serie osservata)=f (trend, stagionalità, residuo).

La sua sintassi è data:

```
nome.mod.stimato<-tsr(data~formula)
```

dove *data* è la serie osservata che si vuole decomporre, mentre *formula* indica gli stimatori da utilizzare per il trend e la stagionalità (nell'ordine) e come le componenti interagiscono tra loro (+ indica una stagionalità additiva, mentre * indica una stagionalità moltiplicativa). Gli stimatori possono essere specificati come in *smoothts()* (vedi Tavola A1).

Ad esempio: *m <-tsr(x~p(1)+c)* indica che si vuole scomporre la serie temporale *x* nella somma di un polinomio di primo grado (il trend viene stimato tramite una retta) e di una componente stagionale costante.

È importante considerare che la funzione *tsr* permette di stimare simultaneamente le componenti utilizzando un approccio basato sul cosiddetto algoritmo di *backfitting*⁴¹.

Si supponga che la serie osservata Y_t presenti sia una componente di trend che una componente stagionale che si combinino tra loro in accordo al modello moltiplicativo.

Si supponga, inoltre, di avere a disposizione una stima preliminare della componente di trend (indicata con \hat{T}_t^0).

L'algoritmo di *backfitting* si concretizza nei seguenti passi:

1. Si pone $i=1$;
2. Si calcola una versione della serie "senza trend", $a_t^{(i)} = y_t / \hat{T}_t^{(i-1)}$, e poi una stima della componente stagionale, \hat{S}_t^i , 'lisciando' le sottoserie stagionali di $a_t^{(i)}$;

⁴¹ MASAROTTO G., Analisi delle serie temporali,.....ecc., op. cit.

3. si calcola una versione della serie destagionalizzata $b_t^{(i)} = y_t / \hat{S}_t^{(i)}$ e una stima della componente di trend, \hat{T}_t^i , 'lisciando' $b_t^{(i)}$;
4. si pone $i=i+1$ e si ritorna al passo 2 a meno che l'algoritmo non sia arrivato a convergenza, ovvero, \hat{T}_t^i non sia sufficientemente vicino a $\hat{T}_t^{(i-1)}$.

Se l'algoritmo viene bloccato dopo i iterazioni, allora si utilizzano le ultime stime prodotte, ovvero \hat{T}_t^i e \hat{S}_t^i come stime delle componenti di trend e stagionalità. La componente irregolare diventa, quindi, $\hat{I}_t = y_t / (\hat{T}_t^{(i)} \cdot \hat{S}_t^{(i)})$.

Ovviamente, nel caso in cui il modello di composizione sia additivo, si può utilizzare l'algoritmo semplicemente ridefinendo:

$$a_t^{(i)} = y_t - \hat{T}_t^{(i-1)}$$

$$b_t^{(i)} = y_t - \hat{S}_t^{(i)}$$

$$\hat{I}_t = y_t - \hat{T}_t^{(i)} - \hat{S}_t^{(i)}$$

Inoltre, si può iterare l'algoritmo partendo con una stima preliminare della componente stagionale e procedendo in maniera analoga a quanto illustrato semplicemente invertendo i passi 2 e 3.

La funzione `tsr` ritorna un oggetto che nella sostanza include una serie temporale multivariata le cui componenti sono il risultato della scomposizione richiesta della serie originale:

`trend (nome.mod.stimato)`: restituisce la stima del trend;

`seasonal (nome.mod.stimato)`: restituisce la stima della stagionalità;

`residuals (nome.mod.stimato)`: restituisce la stima della componente irregolare;

`deseasonal(nome.mod.stimato)`: restituisce la stima della serie destagionalizzata;

`detrend(nome.mod.stimato)`: restituisce la stima della serie a cui è stato eliminato il trend;

`fitted(nome.mod.stimato)`: restituisce i valori previsti dal modello.

Le componenti stimate possono essere visualizzate contemporaneamente richiamando semplicemente l'oggetto creato da `tsr` (ovvero, `nome.mod.stimato`) e rappresentate graficamente tramite il comando `plot (nome.mod.stimato)`

TAVOLA A4 - Risultati ottenuti dalla scomposizione attraverso la funzione *tsr* applicata alla serie ECw.

Rappresentazione analitica

m

Call=tsr(f = ECw ~ lo(2, 25) * lo(0, 1.3))

Time Series:

Start = c(2001, 23)

End = c(2004, 14)

Frequency = 24

	data	trend	seasonal	remainder
2001.917	5.86	7.264305	0.8273464	0.9750259
2001.958	5.75	6.968939	0.8150977	1.0122586
2002.000	4.97	6.735985	0.7506471	0.9829229
2002.042	4.94	6.554346	0.7657042	0.9843205
2002.083	5.43	6.514152	0.8342865	0.9991409
2002.125	6.04	6.624319	0.9253108	0.9853898
2002.167	5.88	6.701737	0.8411872	1.0430313
2002.208	6.49	6.537559	0.9937748	0.9989439
2002.250	6.00	6.016061	0.9583276	1.0406986
2002.292	5.79	5.481317	1.1302144	0.9346152
2002.333	5.81	5.150632	1.1723889	0.9621526
2002.375	6.04	5.084081	1.1663835	1.0185517
2002.417	5.83	5.194640	1.1765892	0.9538679
2002.458	6.58	5.272980	1.2295278	1.0149190
2002.500	6.27	5.342455	1.1747292	0.9990539
2002.542	6.27	5.313646	1.1595587	1.0176119
2002.583	6.32	5.213719	1.2214481	0.9924175
2002.625	6.23	5.234452	1.2073400	0.9857964
2002.667	5.92	5.348998	1.1041256	1.0023764
2002.708	5.66	5.405017	1.0034160	1.0436103
2002.750	4.81	5.247978	0.9133571	1.0034886
2002.792	4.28	4.933157	0.8667418	1.0009885
2002.833	3.86	4.644493	0.8483368	0.9796721
2002.875	4.01	4.455735	0.8894096	1.0118665
2002.917	3.56	4.391340	0.8344333	0.9715413
2002.958	3.49	4.455597	0.8085133	0.9687960
2003.000	3.46	4.578810	0.7560205	0.9995163
2003.042	3.61	4.707798	0.7704688	0.9952547
2003.083	3.87	4.679033	0.8338042	0.9919523
2003.125	4.25	4.511909	0.9317150	1.0109869
2003.167	3.35	4.304182	0.8216559	0.9472492
2003.208	4.17	4.211863	0.9936154	0.9964226
2003.250	3.88	4.411565	0.9365023	0.9391396
2003.292	5.67	4.719791	1.1651819	1.0310187
2003.333	6.11	4.973929	1.1941354	1.0286983
2003.375	5.80	5.130194	1.1546050	0.9791760
2003.417	6.39	5.175686	1.2025551	1.0266631
2003.458	6.47	5.195399	1.2226421	1.0185587
2003.500	6.04	5.153446	1.1746041	0.9978096

2003.542	5.83	5.111654	1.1494406	0.9922487
2003.583	6.30	5.117596	1.2241949	1.0055971
2003.625	6.18	5.034443	1.2127799	1.0121737
2003.667	5.34	4.863113	1.1028604	0.9956492
2003.708	4.50	4.774834	0.9881628	0.9537307
2003.750	4.47	4.928447	0.9119642	0.9945340
2003.792	4.53	5.245518	0.8661586	0.9970395
2003.833	4.88	5.612435	0.8539301	1.0182306
2003.875	5.22	5.977942	0.8855133	0.9861063
2003.917	5.78	6.204484	0.8656735	1.0761381
2003.958	5.37	6.200480	0.8276706	1.0463849
2004.000	4.61	5.975003	0.7609943	1.0138679
2004.042	4.51	5.675105	0.7782857	1.0210890
2004.083	4.75	5.583997	0.8395263	1.0132443
2004.125	5.18	5.659764	0.9263666	0.9879808
2004.167	5.23	5.836246	0.8469376	1.0580756
2004.208	5.97	5.966841	0.9961696	1.0043766
2004.250	6.08	5.803709	0.9737553	1.0758410
2004.292	6.33	5.474903	1.1608605	0.9959723
2004.333	5.93	5.165286	1.1785359	0.9741313
2004.375	5.84	4.907973	1.1669600	1.0196584
2004.417	5.63	4.781770	1.1934904	0.9865084
2004.458	5.34	4.763818	1.1905804	0.9415152
2004.500	5.73	4.865570	1.1759704	1.0014390
2004.542	5.82	5.101963	1.1475164	0.9940925

Rappresentazione grafica :

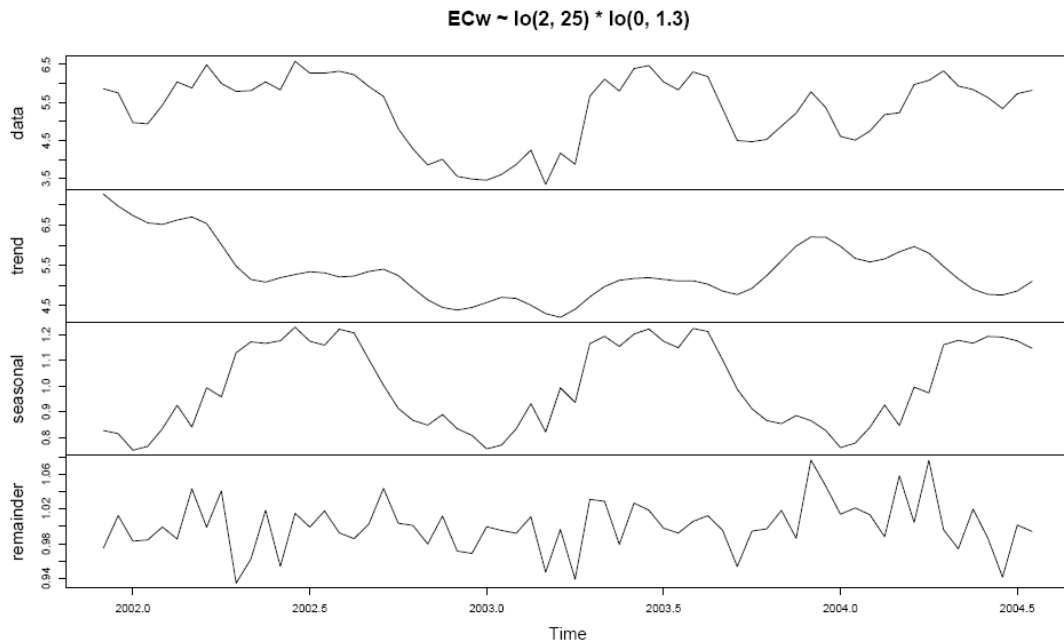


TAVOLA A5 - Sulla funzione *lm* per l'analisi dei modelli lineari

La funzione `lm()` è utilizzata per l'adattamento di modelli lineari generali, ovvero per l'analisi di regressione e per l'analisi della varianza e della covarianza.

Si utilizza il comando `nome.mod.stimato<-lm(formula,data,.....)` dove `data` è il data frame che contiene le variabili da analizzare e `formula` specifica l'espressione del modello lineare in termini simbolici, tipicamente nella forma:

`y~x1+.....+xp`

dove `y`, `x1`,, `xp` sono rispettivamente i vettori che contengono le osservazioni della variabile risposta (quantitativa) e di `p` variabili esplicative (non necessariamente quantitative). L'intercetta viene automaticamente inclusa nel modello stimato, e qualora si voglia ometterla bisogna specificare il modello nella forma:

`y~-1+x1+.....+xp`

Il risultato della funzione `lm()`, ovvero l'oggetto `nome.mod.stimato`, è una lista composta da numerosi elementi che riguardano tutti la stima del modello. I principali sono i vettori:

`nome.mod.stimato$coefficient`, che contiene la stima dei coefficienti del modello,

`nome.mod.stimato$residuals`, che contiene i residui,

`nome.mod.stimato$fitted.values`, che contiene i valori teorici stimati con il modello di regressione.

Vi sono diverse funzioni che possono essere applicate agli oggetti prodotti tramite la funzione `lm()`, in particolare, nel presente lavoro si è fatto uso del comando:

`plot(nome.mod.stimato$residuals)` che restituisce i grafici diagnostici della bontà della stima del modello.

BIBLIOGRAFIA

1. AA.VV. (a cura di P.SCANDELLA – G. MECELLA), *Irrigazione sostenibile: la buona pratica irrigua*, Ministero delle Politiche Agrarie e Forestali, Progetto Editoriale PANDA, Volume No 5, L'Informatore Agrario, Verona, 2004.
2. AA.VV., *Atlante geografico De Agostini*, Istituto Geografico De Agostini S.p.a., Novara, 2003.
3. ALVARO G., *Contabilità nazionale e statistica economica*, Cacucci, Bari, 1999.
4. AZZALINI A., *Inferenza statistica, Una presentazione basata sul concetto di verosimiglianza*, Springer, Milano, 2001.
5. BORTOT P. – VENTURA L. – SALVAN A., *Inferenza statistica, Applicazioni con S-PLUS e R*, Cedam, Padova, 2002.
6. DELVECCHIO F., *Statistica per la ricerca sociale*, Cacucci, Bari, 2002.
7. DI FONZO T. – LISI F., *Complementi di statistica economica, Analisi delle serie storiche univariate*, Cleup, Padova, 2001.

8. FALIVA M., *L'analisi delle serie storiche nel dominio delle frequenze*, Relazione al Convegno Nazionale di studio sull' "Analisi Moderna delle Serie Storiche" (a cura di Piccolo D.), Napoli, 19-22 maggio 1981, Franco Angeli, Milano, 1983.
9. JOHNSTON J., *Econometrica*, Franco Angeli, Milano, 1993.
10. GIRONE G. – SALVEMINI T., *Lezioni di statistica*, Cacucci, Bari, 2002.
11. GIUSTI F. – VITALI O., *Statistica Economica*, Cacucci, Bari, 1990.
12. IACUS S. – MASAROTTO G., *Laboratorio di statistica con R*, McGraw-Hill, Milano, 2003.
13. MASAROTTO G., *Analisi delle serie temporali*, Materiale didattico, Facoltà di Scienze Statistiche, Università di Padova, 2003.
14. 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.
15. MUTTARINI L., *Aspetti di una scomposizione di serie storiche demografiche*, Relazione al Convegno Nazionale di studio sull'

- “Analisi Moderna delle Serie Storiche” (a cura di Piccolo D.),
Napoli, 19-22 maggio 1981, Franco Angeli, Milano, 1983.
16. NEVIANI I. – PIGNOCCHINO FEYLES C., *Geografia generale*,
Torino, Società Editrice Internazionale, 1996.
 17. PASSAMANI G., *Analisi classica delle componenti di trend e
periodiche di una serie storica*, Materiale didattico, Università di
Padova, 1990.
 18. PICCOLO D., *Introduzione all’analisi delle serie storiche*, Roma,
La Nuova Italia Scientifica, 1990.
 19. POLLICE A., *Esercitazioni con R*, Materiale didattico,
Dipartimento di Scienze Statistiche, Università degli Studi di Bari,
2004.
 20. TAYLOR J.R., *Introduzione all’analisi degli errori, Lo studio
delle incertezze nelle misure fisiche*, Bologna, Zanichelli, 1990.
 21. VIANELLI S., *L’analisi delle serie temporali nello sviluppo
storico e metodologico della statistica*, Relazione al Convegno
Nazionale di studio sull’ “Analisi Moderna delle Serie Storiche”
(a cura di Piccolo D.), Napoli, 19-22 maggio 1981, Franco Angeli,
Milano, 1983.
 22. VILLAVECCHIA V., *Trattato di chimica analitica applicata*,
Ulrico Hoepli, vol. 1, Milano 1967.

23. SNEYERS R., *On the statistical analysis of series observation*, Technical note n° 143, Secretariat of the World Meteorological Organization, Geneva, Switzerland, 1990.
24. ZANI S., *Indicatori statistici della congiuntura*, Loescher, Torino, 1982.
25. ZANI S., *Osservazioni sulle serie storiche multiple e l'analisi dei gruppi*, Relazione al Convegno Nazionale di studio sull' "Analisi Moderna delle Serie Storiche" (a cura di Piccolo D.), Napoli, 19-22 maggio 1981, Franco Angeli, Milano, 1983.

Siti internet consultati:

<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>

<http://www.sirio.stat.unipd.it/files/ts04-05>

<http://www.statistica.unipd.it/servizi/matdid.asp?idins=7#appunti>