

UNIVERSITÀ DEGLI STUDI DI PADOVA

Corso di Laurea Magistrale in Scienze Statistiche



**CORRELAZIONE TRA SERIE STORICHE CLIMATICHE :
ANALISI SU DATI EQUISPAZIATI E NON EQUISPAZIATI**

**CORRELATION BETWEEN CLIMATE TIME SERIES :
ANALYSIS OF EVENLY AND UNEVENLY SPACED DATA**

Relatore: Ch.mo Prof. Silvano Bordignon

DIPARTIMENTO DI SCIENZE STATISTICHE

Laureando: Boscaro Manuel

Matricola: 621114-SST

Anno Accademico 2012/2013

Indice Generale

1	Introduzione.....	3
2	Il Danubio.....	5
2.1	Descrizione del percorso dalla sorgente alla foce.....	5
2.2	Idrografia e Idrologia.....	8
2.3	Storia geologica.....	11
3	La stazione Vostok.....	13
3.1	Il Sito e la storia delle ricerche.....	13
3.2	Elementi raccolti.....	15
3.2.1	Anidride carbonica.....	15
3.2.2	Delta deuterio.....	16
3.3	Teoria di MILANKOVIĆ (cenni).....	17
4	Serie equispaziate - caso DANUBIO.....	21
4.1	Introduzione e definizione di serie storica climatica.....	21
4.2	La correlazione tra due serie storiche climatiche.....	23
4.2.1	Moving block bootstrap a coppie.....	23
4.3	Applicazione al caso Danubio.....	26
4.3.1	L'insieme dei dati.....	26
4.3.2	Analisi descrittive.....	28
4.3.3	Stima della parte deterministica e degli outliers.....	32
4.3.3.1	Parte deterministica	32
4.3.3.2	Outliers	35
4.3.4	Stima modello AR(p) per la parte stocastica.....	38
4.3.5	Correlazione tra coppie di stazione successive.....	43
5	Serie non equispaziate - caso Vostok.....	51
5.1	Introduzione serie climatiche non equispaziate	51
5.2	Analisi spettrale per serie non equispaziate.....	54

5.2.1	Stima dello spettro per serie storiche non equispaziate.....	55
5.2.2	Stima del filtro armonico.....	57
5.3	Modelli AR(1) per serie storiche non equispaziate.....	58
5.4	Correlazione tra serie storiche non equispaziate.....	64
5.4.1	Binned correlation	65
5.5	Applicazione al caso Vostok.....	68
5.5.1	L'insieme dei dati.....	68
5.5.2	Analisi preliminari.....	70
5.5.3	Analisi spettrale.....	72
5.5.4	Stima del filtro armonico.....	74
5.5.5	Modello AR(1) per stima della persistenza temporale.....	75
5.5.6	Stima della correlazione.....	78
6	Conclusioni.....	79
7	Ringraziamenti.....	81
8	Bibliografia.....	83

1 Introduzione

Negli ultimi anni, il manifestarsi nel nostro pianeta, di eventi climatici in maniera congiunta, sempre più intensi e talvolta dannosi, ha spinto diversi studiosi a cercare di misurare il grado di interazione (o correlazione) tra gli stessi, mettendo a punto metodi di indagine basati su strumenti statistici.

Lo scopo di questo lavoro è quello di visualizzarne alcuni appositamente creati per stimare la misura della relazione tra osservazioni di due fenomeni differenti, raccolte a determinati istanti temporali.

Nello specifico vengono presentate (ed applicate) due tecniche di stima della correlazione, le quali fanno riferimento a due differenti tipologie di scala temporale, ossia quella equispaziata e quella non equispaziata.

Per quanto riguarda il primo caso, il contesto applicativo è rappresentato dalla serie delle portate d'acqua mensile del fiume Danubio in quattro successive stazioni di monitoraggio, ed i capitoli riservati alla trattazione di tale argomento sono il secondo ed il quarto.

Il capitolo 2 fornisce diverse informazioni geografiche sul Danubio, infatti inizialmente ne viene descritto in modo dettagliato il suo percorso, il suo bacino idrografico, i suoi principali affluenti ed i differenti regimi di portata nelle diverse aree attraversate; infine vien dato qualche cenno sulle origini del fiume stesso.

Nel capitolo 4 invece si definiscono le componenti di una serie storica climatica equispaziata e si enuncia la procedura di stima per la correlazione tra stazioni successive. L'applicazione alle serie delle portate di quanto considerato viene svolta nel terzo ed ultimo paragrafo.

Per il caso delle serie storiche non equispaziate sono state considerate le informazioni provenienti dalla carota di ghiaccio di Vostok, in Antartide e precisamente si è focalizzata l'attenzione sulle osservazioni riguardanti il livello di anidrite carbonica e la variazione del Deuterio nel passato, tutto ciò è trattato nei capitoli 3 e 5.

Nel secondo capitolo (come nel caso precedente) si offre una descrizione delle principali caratteristiche del sito e la cronistoria delle ricerche nello stesso, si elencano

gli elementi raccolti ed in particolare vengono approfondite le peculiarità della CO_2 e del δD ; infine si riportano i punti fondamentali della **teoria di Milankovic**, la quale è legata alla ricostruzione del clima nel passato.

Il capitolo 5 presenta l'insieme delle tecniche statistiche per l'analisi e la modellazione di tali serie storiche; in particolare viene suggerito un metodo sia per la stima dello spettro sia per la stima della correlazione, lasciando l'ultimo paragrafo all'applicazione di tali metodi alle due serie considerate.

2 Il Danubio

2.1 Descrizione del percorso dalla sorgente alla foce

Il Danubio con i suoi ben 2.902 km è il secondo corso d'acqua più lungo del continente (dopo il Volga), ed il più lungo fiume navigabile dell'Unione Europea.

Le sue fonti risiedono nella *Foresta Nera* (situata nella Germania sud-occidentale), esattamente da due torrenti discendenti dalla stessa, ossia il *Brigach* ed il *Berg*. Questi ultimi si incontrano nella cittadina di *Donaueschingen*, dove nel parco del castello della medesima, è presente una fontana monumentale del XIX secolo che simboleggia la sorgente ufficiale del fiume.

Nel primo tratto del percorso, esso bagna la città di *Ulma*, nella quale avviene l'incontro con uno dei suoi principali affluenti di destra, ovvero l'*Iller*; poi attraversa la Baviera passando per *Ratisbona* (incontrando un suo affluente di sinistra, il *Regen*) e per *Passavia*, dove vi confluiscono sia le acque dell'*Inn* (il quale sorge dalle Alpi) sia quelle dell'*Ilz* (proveniente da una zona paludosa della Foresta Bavarese); compiendo complessivamente 687 km nel solo suolo tedesco, prima di approdare in Austria. Quest'ultima nazione è molto legata al Danubio non solo a livello storico, dove al grande Impero austro-ungarico (il cui territorio si estendeva per 1300 km lungo il fiume) veniva dato l'appellativo di 'Monarchia del Danubio', ma anche a livello idrologico, in quanto praticamente tutti i fiumi forniscono acqua allo stesso, il quale, inoltre annovera la presenza di ben 12 dighe idroelettriche ubicate nei pressi del proprio corso fluviale, in tutta la nazione. Immediatamente dopo l'entrata nel territorio austriaco, il corso del fiume compie letteralmente un giro di 180° (tecnicamente ciò viene chiamato *ansa*) presso la cittadina di *Schlögen* e proseguendo, attraversa le città di *Linz* e *Mauthausen* fino ad arrivare ad *Enns* (nella quale avviene la confluenza con altri due affluenti come il *Grein* e l'omonimo della medesima cittadina) raggiungendo qui il suo punto più profondo in Austria; scorrendo verso est, esso sfocia in una valle, chiamata *Walchau*, la quale rappresenta uno dei più pittoreschi paesaggi associati al Danubio, tanto da meritare la tutela da parte dell'UNESCO come *Patrimonio Mondiale dell'Umanità*. Il termine del suo tragitto in Austria coincide con l'arrivo a *Vienna*, nella quale incrocia ancora un altro affluente di destra, il *Wienfluss*, da cui proviene il nome della città; proprio per la capitale austriaca, il Danubio ha contribuito in modo non

indifferente allo sviluppo ed alla crescita delle proprie attività economiche, tanto che ancor tutt'oggi costituisce un'importante rotta commerciale tra Oriente ed Occidente e per limitare gli effetti negativi dovuti a piene o addirittura alluvioni, le acque del fiume in questo tratto sono state artificialmente regolamentate.

All'ingresso in Slovacchia, la prima grande città bagnata è proprio la capitale, ossia *Bratislava*, in cui vi confluiscono dalla sponda sinistra le acque del *Morava*; successivamente il Danubio transita a *Komárno*, ove trova (sempre da affluente di sinistra) il *Váh* (il più grande fiume slovacco), per poi scorrere lungo il confine con l'Ungheria, nella quale vi entra totalmente quando valica le montagne della *Dorsale ungherese*, più precisamente lasciandosi sulla destra i *Monti Gerecse* e *Pilis*, sulla sinistra invece i *Monti Börzsöny*. Una volta eseguito questo attraversamento, giunge nella città di *Visegrád*, laddove il fiume nel virare completamente corso **da est verso sud**, disegna una vera e propria curva a 90° gradi (detta in ungherese *Dunakanayar*, che significa letteralmente *ginocchio del Danubio*), la quale si sviluppa per una lunghezza complessiva di ben 20 km e termina il suo raggio poco dopo la cittadina di *Vác*. Puntando dunque in direzionale meridionale, esso perviene nella capitale Budapest, la quale è la più grande città incontrata lungo il suo cammino, entrando così nella *Grande pianura ungherese*, dove il Danubio stesso ne funge da confine naturale con la parte occidentale; e dopo aver lambito altre città minori, lascia il suolo magiaro per addentrarsi in quello croato.

Il transito del fiume in Croazia è contraddistinto principalmente da due tappe: una prima è il porto fluviale di *Batina* (frazione del comune di *Draž*), collocato al crocevia tra Ungheria, Serbia e Croazia stessa; una seconda è invece l'innesto del suo secondo affluente principale di destra, la *Drava*, distante esattamente 20 km dal passaggio per la città di *Osijek*.

Dopo aver percorso una parte del confine nord-occidentale tra la Croazia e la Serbia, entra definitivamente nel territorio di quest'ultima, voltando il suo corso nuovamente **verso est**, nelle vicinanze di *Bačka-Palanka* e proseguendo fino a *Novi Sad*, ove è stato costruito il cosiddetto *Ponte della Libertà* in sostituzione di quello gravemente danneggiato durante la *Guerra del Kosovo*, nel 1999. Procedendo **verso sud-est**, il Danubio prima arriva nella capitale serba, *Belgrado*, la quale rappresenta la terza città più grande bagnata dallo stesso e dove vi trova la confluenza con il maggiore affluente

sulla sua riva destra, il *Sava* ; poi a *Pančevo*, una città industriale, in cui incrocia le acque del *Tibisco* (suo principale affluente in assoluto) e successivamente prosegue il suo corso verso *Golubac*, cittadina serba che in pratica segna l'inizio del confine con la Romania. Quest'ultimo è caratterizzato da una vera e propria 'gola' tortuosa e stretta (denominata *le Porte di Ferro*), che il fiume percorre per ben 130 km, nell'aprirsi un varco tra i *Balcani* ed i *Carpazi Meridionali* ; causando un forte abbassamento del livello delle acque (nell'ordine di 20 m) e da un'altrettanto brusca riduzione del letto dello stesso, il quale passa da 1,5 km a solo 170 m, rendendo la navigazione molto impervia ed allo scopo di ridurne i rischi, vi è stato costruito un canale laterale (sulla riva serba).

Inserendosi totalmente in territorio romeno, il Danubio lambisce *Orșova* ed in seguito *Drobeta-Turnu Severin*, da dove devia il suo corso **verso sud** per passare in successione *Gruia*, *Pristol*, *Cetate e Calafat* ; ritorna in seguito a muoversi ancora **verso est**, puntando la parte settentrionale della Bulgaria, dove scorrendo per quasi 500 km vi funge da confine attraversando *Giurgiu* (situata di fronte alla città bulgara di *Ruse*), *Svištov* (raggiunge il suo punto più meridionale) ed infine *Oltenița* (incontra le acque del fiume *Argeș*) per poi lasciarlo e voltare **a nord** per giungere nelle città più grandi di *Brălia* e *Galați*, dove in quest'ultima si innestano le acque del *Siret* ed un paio di km più avanti quelle del *Prut*, varcando una piccola porzione della Moldavia.

In procinto di concludere il suo percorso sfociando nel *Mar Nero*, il Danubio, non appena abbandonato il suolo moldavo, sopraggiunge a *Tulcea* e si **divide in tre rami**, in quello che è denominato *Delta del Danubio*, il quale è anch'esso divenuto (nel 1991) *Patrimonio mondiale dell'umanità*, tutelato dall'UNESCO. Una prima diramazione (quella situata più a nord), chiamata *Chilia* e lunga 104 km, rappresenta il confine tra la Romania e l'Ucraina ed una volta incontrato i porti di *Izmail*, *Kilia* e *Vilkovo*, si getta nel mare interamente in territorio ucraino; le rimanenti due diramazioni sono da attribuirsi completamente al suolo romeno, e sono denominati rispettivamente *Sulina* (lungo 72 km, rappresenta il ramo centrale e viene continuamente bonificato in modo tale da consentire la navigazione del fiume stesso) e *Sfintu Gheorge* (il più lungo con 112 km e situato più a sud dei due precedenti), i quali sfociano anch'essi nel *Mar Nero*; complessivamente il *Delta del Danubio* infine, si estende su un'area di 3446 km².

2.2 Idrografia e Idrologia

Il Danubio presenta due caratteristiche che lo rendono unico a fronte degli altri grandi fiumi europei; una prima è che il corso delle proprie acque muove **da occidente verso oriente**, mentre la seconda riguarda la rilevazione della propria lunghezza, in quanto il punto iniziale (lo zero) è posto ufficialmente nei pressi della città romena di *Sulina* (la quale si affaccia sul *Mar Nero* ed appartiene al ramo centrale del delta), quindi in pratica partendo dalla foce risalendo alla sorgente e non appunto viceversa.

Il *bacino idrografico* del Danubio (la superficie, o meglio l'area topografica che raccoglie tutte le acque confluenti verso il fiume stesso) si estende su un area di ben 817000 km², i quali sono percentualmente così ripartiti (in ordine decrescente) fra le nazioni lambite dal suo corso: Romania (28,9%), Ungheria (11,7%), Austria (10,3%), Germania (7,5%), Slovacchia (5,8%), Serbia e Bulgaria (ambedue con il 5,2%), Croazia (4,5%), Ucraina (3,8%) ed infine Moldavia (1,7%). Esso però comprende anche altri ben 9 paesi, i quali posseggono corsi d'acqua che vanno poi a confluire nel Danubio stesso; anche in questo caso vengono elencate proporzionalmente e sempre dal più grande al più piccolo: Bosnia ed Erzegovina (4,8%), Repubblica Ceca (2,6%), Slovenia (2,2%), Svizzera (0,32%), **Italia** (0,15%), Polonia (0,09%) ed infine Montenegro, Repubblica di Macedonia ed Albania (ciascuna con il 0,03%). Come si può notare, anche una, seppur molto esigua, parte del bacino idrografico del fiume appartiene al territorio italiano; tale affermazione è supportata da queste tre motivazioni: una prima è legata al comune lombardo di *Livigno* (situato oltre la dorsale alpina), anzi precisamente nella *Val di Livigno*, dove nasce e scorre il torrente *Spöl*, del quale corso sfocerà nell'*Inn* (esattamente a *Zernez*, in Svizzera) che è uno dei principali affluenti del Danubio ed in virtù di questa ragione l'Italia ha ottenuto i diritti di navigazione su quest'ultimo a partire dal 1948 con la Convenzione di Belgrado; una seconda invece è connessa al fatto che nel comune di *San Candido/Innichen* (in provincia di *Bolzano*) si trovano le sorgenti della *Drava*, il secondo più grande affluente del Danubio; una terza ed ultima risiede in una porzione del territorio della *Val Canale* (la parte assegnata al comune di *Tarvisio*, ubicato in provincia di *Udine* e precisamente tra le *Alpi Carniche* ed *Alpi Giulie*), nella quale scorre lo *Slizza*, le cui acque alimentano il *Gail*, il quale quest'ultimo si innesta nella *Drava* (a *Vilach*, in Austria), che come detto in precedenza

è affluente del Danubio.

La *portata* (il volume d'acqua che passa in una sezione di un fiume in un secondo) del Danubio invece, non risulta assolutamente omogenea durante tutto il lungo percorso del fiume, e la causa di ciò è da attribuire alla notevole differenza di *regime* (l'insieme delle variazioni della portata di un fiume durante l'anno) tra gli affluenti dello stesso. Ve ne sono riconosciute quattro tipologie: **regime pluvio oceanico** (nella Baviera occidentale), **nivo-pluviale di montagna** (prevalentemente in Austria), **pluvio-nivale di pianura** (in Ungheria), **nivale di pianura** (in Valacchia-Moldavia). La prima porta essenzialmente una sola peculiarità, ovvero quella che i valori massimi di portata coincidono con la distribuzione delle piogge nell'annata; in particolare, nelle regioni interessate ciò coincide tipicamente con il periodo stagionale primaverile e quello autunnale. La seconda e la terza classificazione hanno una particolarità in comune, ossia quella di registrare il massimo valore della portata nel periodo primaverile-estivo, nel quale si verifica prima lo scioglimento delle nevi e poi quello dei ghiacciai, mentre differiscono per la preponderanza delle precipitazioni piovose (in questo caso si parla di regime pluvio-nivale) o per la prevalenza di quelle a carattere nevoso (ciò viene definito invece regime nivo-pluviale); i rilevamenti fatti a *Linz* ed a *Vienna* corroborano quanto detto inizialmente, in quanto esse presentano livelli minimi attorno al mese di dicembre e livelli massimi attorno al periodo di maggio-giugno (a Vienna la portata media di quest'ultimo periodo è pari a $2237 \text{ m}^3/\text{s}$), le medesime caratteristiche vengono mantenute anche nei passaggi a *Bratislava* ed a *Budapest* (e qui il fiume si mantiene ancora gonfio per l'inizio della stagione estiva con piogge, tipica *dell'Europa centrale*) ovviamente con valori diversi. Eccezioni alla situazione appena descritta vengono a verificarsi seguendo il fiume prima verso sud e dopo verso est : la prima si manifesta a seguito della confluenza delle acque del *Tibisco* e del *Sava*, i quali anticipano le grandi portate in primavera (generalmente nei mesi di aprile-maggio) per decrescere successivamente verso le portate minime a partire da giugno fino a settembre, tale caratteristica è rilevata sino alle *Porte di Ferro*; la seconda, dove il fiume riceve forti influenze dal clima della prima parte del territorio romeno (propriamente quello della *steppa russa*) e presenta dunque bassi flussi d'acqua nel periodo estivo. La quarta ed ultima caratterizzazione invece, mette in evidenza sostanzialmente che la propria portata massima del fiume è raggiunta nel periodo estivo o più precisamente nel mese di

giugno, a causa dello scioglimento delle nevi dei *Carpazi meridionali* (o *Alpi Transilvaniche*), dove le precipitazioni invernali avvengono in maniera molto copiosa. Oltre a queste quattro tipologie si segnala inoltre, che presso diversi tratti del percorso il Danubio, a causa del *clima continentale*, (nella stagione invernale si possono arrivare a toccare temperature rigidissime, vicine ai -30°) le acque dello stesso possono ghiacciare, per poi nella fase di disgelo dar luogo a devastanti *inondazioni* (ad esempio una delle più importanti è quella del marzo del 1956). Per quanto riguarda invece i periodi in cui esso ha fatto registrare i valori massimi assoluti della propria portata vi sono sicuramente il giugno del 1965 ed il giugno 1899, con una media rispettivamente di $8000 \text{ m}^3/\text{s}$ e $9000 \text{ m}^3/\text{s}$. Il contributo in termini di portata d'acqua, delle diverse nazioni che si affacciano (direttamente ed indirettamente) sul fiume stesso, è così percentualmente diviso (in ordine decrescente): Austria (22,1%), Romania (17,6%), Germania (14,5%), Serbia (11,3%), Bosnia ed Erzegovina (8,8%), Croazia (6,4%), Ungheria ed Ucraina (entrambe con il 4,3%), Bulgaria (3,7%), Slovenia (3,1%), Slovacchia (1,9%), Repubblica Ceca (1,2%) ed infine Moldavia (0,7%).

Concludendo il paragrafo, viene elencata la lista dei principali affluenti del fiume Danubio, in ordine di portata media: *Sava* ($1564 \text{ m}^3/\text{s}$), *Tibisco* ($794 \text{ m}^3/\text{s}$), *Inn* ($735 \text{ m}^3/\text{s}$), *Drava* ($577 \text{ m}^3/\text{s}$), *Siret* ($240 \text{ m}^3/\text{s}$), *Morava* ($232 \text{ m}^3/\text{s}$), *Lech* ($200 \text{ m}^3/\text{s}$), *Enns* ($195 \text{ m}^3/\text{s}$), *Olt* ed *Isar* (ambedue con $174 \text{ m}^3/\text{s}$), *Váh* (con $161 \text{ m}^3/\text{s}$), *Traun* (con $150 \text{ m}^3/\text{s}$), *Prut* (con $110 \text{ m}^3/\text{s}$), *Jiu* (con $86 \text{ m}^3/\text{s}$), *Argeş* (con $71 \text{ m}^3/\text{s}$), *Iller* (con $70 \text{ m}^3/\text{s}$), *Rába* (con $63 \text{ m}^3/\text{s}$), *Hron* (con $55 \text{ m}^3/\text{s}$), *Iskar* (con $54 \text{ m}^3/\text{s}$), *Naab* (con $49 \text{ m}^3/\text{s}$), *Jantra* e *Timiş* (entrambi con $47 \text{ m}^3/\text{s}$), *Ialomiţa* (con $45 \text{ m}^3/\text{s}$), *Sió* (con $39 \text{ m}^3/\text{s}$), *Regen* (con $35 \text{ m}^3/\text{s}$), *Timok* (con $31 \text{ m}^3/\text{s}$) ed infine *Altmühl* e *Ilpel'* (tutt'e due con $22 \text{ m}^3/\text{s}$).

2.3 Storia geologica

Dal punto di vista geologico, le origini del Danubio sono direttamente associabili a quelle di un altro grande fiume europeo, il Reno, ed il fatto che il bacino idrografico di quest'ultimo confina con quello del Danubio stesso, (precisamente nella parte sud-occidentale della Germania) comporta la spiegazione 'storica' di una particolare peculiarità che il fiume attualmente presenta nella prima parte del suo corso.

Risalendo fino al periodo antecedente la *glaciazione Riss* (spazia tra un arco temporale che va dai 200000 ai 130000 anni fa), si è trovato che le acque ed i torrenti delle Alpi, confluenti oggi nel Reno, erano invece indirizzate verso est dal fiume Danubio, il quale transitava più a nord lambendo nel suo percorso gran parte dell'Alta Baviera (precisamente nel territorio compreso tra i comuni di Wellheim e Riedenburg) e ciò avvenne sino all'inizio del periodo glaciale menzionato in precedenza.

Successivamente alla *glaciazione Würm* (avvenuta in un periodo che va dai 100000 ai 12000 anni circa fa e l'ultima che si è verificata) nel Reno (che nel periodo precedente la stessa aveva le sue sorgenti nella Foresta Nera) cominciarono a convogliare molte acque provienti dalle Alpi, in quanto esso stava formando una parte dell'attuale pianura presente presso la propria parte superiore (la quale si affaccia sul Mare del Nord); ciò, nel tempo ha portato alla lenta sparizione del letto del fiume Danubio presso le gole del massiccio del Giura, nel quale le proprie acque attualmente vi penetrano e si infiltrano nei calcari porosi per poi raggiungere il Reno situato più a valle. Si prevede inoltre che, vista la continua azione erosiva del calcare da parte di queste grandi quantità d'acqua, la parte superiore del Danubio stesso andrà nel tempo a scomparire completamente in favore del fiume Reno; ad esempio vicino al comune di *Immendingen* (situato nel sud della Germania) il fiume è quasi completamente asciutto perché l'acqua filtra nella terra e, attraverso grotte sotterranee raggiunge l'*Achtopf* (la più grande sorgente carsica presente in Germania), la quale alimenta il *Lago di Costanza*, e quindi indirettamente il Reno; il fenomeno appena spiegato prende il nome di *scomparsa del Danubio* (*Donauversickerung*).

3 La stazione Vostok

3.1 Il Sito e la storia delle ricerche

La base **Vostok** è situata nel cuore del *continente antartico* e precisamente in una zona nella quale la calotta di ghiaccio raggiunge i 3700 m di spessore. Il sito fu realizzato dai russi nel dicembre del 1957, in occasione dell'*Anno geofisico internazionale*, il quale fu indetto e promosso per coordinare su scala mondiale un insieme di ricerche volte ad una maggiore conoscenza delle proprietà fisico-chimiche della Terra e le interazioni della stessa con il Sole, nel corso della sua storia. La caratteristica climatica principale della stazione è ovviamente data dalle temperature rigidissime, le quali spaziano da una massima con media mensile di -28°C registrata a Dicembre, ad una minima con media di $-75,4^{\circ}\text{C}$ in Agosto, mese che appartiene alla stagionale invernale, nell'emisfero australe; in particolare proprio Vostok detiene il record della temperatura minima più bassa registrata sul pianeta, ossia $-89,2^{\circ}\text{C}$ nel luglio del 1983. Condizioni simili hanno reso (e rendono tutt'ora) l'attività dei ricercatori molto complicata, ma tale aspetto non ha impedito agli stessi di raccogliere elementi, che rappresentano il cuore pulsante della ricostruzione della storia climatica e biologica del nostro pianeta. La collaborazione tra gli stessi, in particolare tra i russi ed i francesi, gettò le basi per arrivare passo dopo passo a scoperte di fondamentale importanza sia per la paleoclimatologia che per la geofisica moderna. Tutto ciò ebbe inizio nel 1955, quando **Claude Lorius**, al tempo giovane ricercatore presso l'Università di Besançon, decise di rispondere ad un annuncio sulla possibilità di partecipare alle campagne di ricerca organizzate per l'Anno Geofisico Internazionale e partì dunque per le spedizioni in Antartide, con l'obiettivo di lavorare sulla ricostruzione delle temperature del ghiaccio tramite i rapporti isotopici. Nel 1965, egli ebbe però l'intuizione più significativa, ovvero che le bolle d'aria, presenti nel ghiaccio, potevano ricostruire lo stato dell'atmosfera nel passato e perciò decise di mettere a punto una raffinata tecnica di analisi chimica del ghiaccio, presso il *Laboratoire de Glaciologie a Grenoble*. A metà degli anni '70, Lorius diventò Direttore delle spedizioni polari francesi e ritornò in Antartide per proseguire le ricerche e, dopo varie perforazioni, che non rivelarono oltre 35000 anni di storia del clima, riuscì ad ottenere il contatto proprio con la spedizione russa di Vostok, la quale era impegnata in

un carotaggio profondo del ghiaccio. Qui vi giunse nel 1984 e grazie al rapporto d'amicizia sviluppato con il direttore della stazione Volodya Kotlyakov, ottenne la possibilità di poter portare a Grenoble i campioni di ghiaccio ottenuti sino a quel periodo, al fine di poterli esaminare. Esattamente 3 anni più tardi, i risultati dell'analisi delle carote, comparvero su *Nature* (una tra le più importanti riviste di geoscienza) rivelando non solo 160000 anni di clima del pianeta, ma confermando l'aumento della concentrazione atmosferica dell'anidride carbonica e di metano, dopo la fine dell'ultimo ciclo glaciale, esattamente 12000 anni fa. Oltre a ciò, vennero ritrovate tracce del piombo della benzina ed il fallout delle esplosioni nucleari, i quali furono la prova dell'alterazione dell'ambiente ad opera dell'uomo. Questa pubblicazione funse da rampa di lancio per la scuola glaciologica francese ed al fianco di Lorius, lavorava un team di ricercatori che sarebbe divenuto inscindibile dai risultati della paleoclimatologia polare; al contrario la stazione di Vostok stava attraversando il periodo più buio della sua storia, rischiando la chiusura a causa della crisi dell'Unione Sovietica, ma fu fortunatamente scongiurata grazie all'intervento americano. Le perforazioni e le ricerche poterono così proseguire e nel 1998 fu finalmente estratto l'ultimo segmento di una carota di ghiaccio che raggiunse i 3623 m, la quale rimarrà sino al 2003 (dopo la perforazione ad Epica Dome) la più lunga al mondo. Gli esiti delle analisi della carota furono pubblicati (sempre su *Nature*) nel 1999, in un famoso articolo firmato da ben diciotto autori (tra i quali vi fu anche Claude Lorius) , dal titolo "*Climate and atmospheric history of the past 420000 years from the Vostok Ice Core, Antartica*", il quale è stato inserito tra i primi tre articoli più importanti per la geofisica contemporanea, in quanto permise di arrivare a determinare le concentrazioni di diversi elementi, nel corso degli ultimi quattro cicli glaciali. La carota di ghiaccio non raggiunse però il fondo roccioso e la sorpresa derivò dal fatto che dopo oltre tre chilometri e mezzo di calotta glaciale, venne individuato (grazie alle prospezioni radar ed al tipo di ghiaccio) un gigantesco invaso d'acqua liquida, il quale fu battezzato **Lago Vostok**, che è uno dei 140 laghi subglaciali presenti in Antartide. In seguito rilevazioni sempre più dettagliate portarono a determinare la posizione e l'area della superficie dello stesso, che si attesterebbero rispettivamente attorno ai 4000 metri di profondità ed ai 14000 kmq. Nel mese di Febbraio del 2012, i russi hanno comunicato di essere riusciti a raggiungere (senza contaminarle) le acque del lago; che rappresentano un interessante oggetto di ricerca dei

prossimi anni.

3.2 Elementi raccolti

Nell'ultima carota di ghiaccio estratta a Vostok nel 1998, diventata poi così famosa ed importante nella comunità geofisica grazie alla pubblicazione dei risultati della sua analisi, sono state ritrovate (e misurate) concentrazioni di molti elementi presenti in natura. Essi sono rilevanti in quanto vengono considerati *proxy* di variabili climatiche, quindi nel momento in cui si registrano delle variazioni nei loro livelli, automaticamente sono associate a cambiamenti (avvenuti nel corso degli ultimi 420000 anni) del clima terrestre. Nello specifico vi sono stati ritrovati *polveri atmosferiche*, alle quali sono associati cambiamenti nel corso dell'attività dei venti; *il sodio*, il quale è associato alla concentrazione di particelle di sale marino nell'atmosfera, che influenzerebbero la formazione delle nuvole e quindi delle piogge; la variazione dell'*isotopo 18 dell'ossigeno* (^{18}O), il quale è responsabile della variazione del volume globale dei ghiacci e del ciclo idrologico. Oltre a questi, sono stati individuati altri tre importanti elementi per lo studio dell'evoluzione del clima nel passato, ossia la concentrazione di metano (CH_4), di anidride carbonica (CO_2), i quali sono responsabili del cosiddetto effetto serra, e la variazione del livello di Deuterio (δD) che è direttamente associabile ad un cambiamento termico. Queste due ultime proxy saranno oggetto di analisi nel capitolo 5, motivo per cui se ne dà, nei sottoparagrafi seguenti, una descrizione delle caratteristiche di entrambe.

3.2.1 Anidride carbonica

La formazione (a livello chimico) del biossido di carbonio è data dall'unione di due atomi di ossigeno ed uno di carbonio e fisicamente si presenta nell'atmosfera allo stato gassoso. La concentrazione nella stessa del livello di CO_2 è misurata in *parti per milione volume*, ossia usando due grandezze omogenee che stanno tra loro in un rapporto pari ad uno su un milione (ad esempio mg/kg) moltiplicandolo per il rapporto fisso tra una mole di gas ed il volume occupato da essa in condizioni normali. Nonostante tale quantità sia relativamente molto piccola e garantisca la vita di piante ed animali, negli ultimi due secoli (soprattutto a causa della rivoluzione industriale)

l'attenzione verso il suo costante incremento hanno assunto rilevanza mondiale, poiché l'anidride carbonica (assieme al vapore acqueo ed al metano) è in grado di intrappolare la radiazione infrarossa della luce solare e di rifletterla poi sulla superficie terrestre, impedendo alla stessa di raffreddarsi e causando il famoso (e già citato) *effetto serra*, il quale a sua volta provoca un complessivo aumento (e progressivo) della temperatura media globale, tale da rendere via via sempre più difficoltoso l'adattamento degli ecosistemi. Infatti, rilevazioni effettuate dall'istituto statunitense *NOAA (National Oceanic and Atmospheric Administration)* indicano per l'anno 2011, una concentrazione media pari a **391,57 ppmv** contro un valore che prima dell'avvento delle macchine a vapore risultava pari a **280 ppmv**, dove le uniche fonti di produzione erano prevalentemente legate ai processi di fermentazione svolti da moltissimi microrganismi presenti in natura ed all'attività vulcanica.

3.2.2 Delta deuterio

Il significato fisico-chimico attribuito alle variazioni di Deuterio (il quale è un isotopo stabile dell'idrogeno) è esattamente lo stesso di quello assegnato alle variazioni dell'isotopo 18 dell'ossigeno, ovvero rappresentano l'indicatore della *variazione di temperatura dell'aria*. Nello specifico si tratta di una misura utilizzata per confrontare il peso complessivo (tramite la formazione di un rapporto) di due molecole d'acqua, una prima (posta al numeratore) nella quale è presente l'isotopo più pesante dell'ossigeno (denotata con $H_2^{18}O$ oppure $HD^{16}O$) ed una seconda (posta al denominatore) dove vi è l'isotopo più leggero dell'ossigeno (in notazione $H_2^{16}O$); chiamando il rapporto così costruito con $R_{CAMPIONE}$, in quanto calcolato sugli elementi raccolti. Per avere la variazione (δ) è però necessario disporre di un valore di confronto e nel caso relativo ai due isotopi dell'ossigeno (^{18}O e ^{16}O) esso corrisponde a quello del *Vienna-Standard Mean Ocean Water (V-SMOW)*; tutto ciò viene espresso secondo la seguente formulazione:

$$\delta D = [(R_{CAMPIONE} - R_{V-SMOW}) / R_{V-SMOW}] * 1000$$

Dove la moltiplicazione per mille è necessaria vista la notevole presenza di numeri decimali, facilitando così la lettura del valore finale. In virtù di quanto finora enunciato, si può affermare che tale variazione sarà superiore quando prevarrà la presenza del ^{18}O

(l'isotopo più pesante e che evapora meno facilmente) quindi ciò viene associato ad una **temperatura dell'aria bassa** e, in paleoclimatologia, ad un **periodo glaciale** della Terra. Viceversa una prevalenza del ^{16}O (l'isotopo dell'ossigeno più leggero) conduce alla conclusione completamente opposta, ossia viene collegata ad un regime di **temperature più miti** e quindi ad un **periodo interglaciale**.

3.3 Teoria di MILANKOVIĆ (cenni)

L'astronomo e geofisico serbo **Milutin Milanković**, nel primo decennio del '900, studiando le opere del climatologo austriaco Julius Von Hann, aveva notato un problema significativo, il quale sarebbe divenuto poi il suo argomento di studio nei successivi anni, ovvero la ricerca della spiegazione delle cosiddette **ere glaciali**, che hanno caratterizzato il clima del nostro pianeta nel corso della sua storia. Egli si concentrò in particolar modo sull'influenza della radiazione solare, riuscendo a determinarne la distribuzione della stessa sulla superficie terrestre, nel lavoro dal titolo "*La distribuzione della radiazione solare sulla superficie della Terra*" pubblicato nel giugno del 1913; tuttavia lo scienziato non si fermò ai risultati trovati per la Terra, ma proseguì gli studi anche per altri pianeti, in particolare su Marte ("*Indagine del clima del pianeta Marte*" pubblicato nel 1916). Successivamente, grazie ai suoi studi in ingegneria, iniziò ad occuparsi di meccanica celeste (ossia la disciplina che studia i moti dei pianeti, dei satelliti e degli asteroidi) e ciò si rivelerà fondamentale per la formulazione della teoria. Egli infatti, sfruttando quanto già scoperto da Keplero in merito all'eccentricità delle orbite planetarie, ricostruì la meccanica dei movimenti dei pianeti basandosi fondamentalmente su due vettori: un primo specifica il piano orbitale, il senso di rotazione del pianeta ed il parametro dell'ellisse formata con l'orbita stessa; un secondo invece l'eccentricità dell'orbita e l'asse della stessa, nel proprio piano. Furono dunque proprio tali studi a garantirgli la possibilità di elaborare una teoria (pubblicata poi nel 1930) basata principalmente su tre movimenti orbitali del nostro pianeta, dei quali egli ne mise in risalto la ciclicità. Essi risultarono noti come **Cicli di Milanković**, i quali lo stesso scienziato collegò direttamente ai cambiamenti geologici e climatici della Terra nel corso dei millenni passati, ma tale nesso fu definitivamente accettato solo a partire dal 1976, nel lavoro "*Variazioni dell'orbita della Terra: Pacemaker delle ere glaciali*"

degli scienziati Hays, Imbrie e Shackleton, che formularono la *Teoria Astronomica del clima*, perno della geologia moderna.

I *Cicli di Milanković*, riguardano nello specifico tre caratteristiche dei movimenti orbitali della Terra, ossia *l'eccentricità, l'obliquità e la precessione*. Tali peculiarità sono in seguito descritte distintamente, nell'ordine appena menzionato:

· **ECCENTRICITA'**:

Il primo dei tre cicli di Milanković, riguarda l'orbita terrestre intorno al Sole e come si sa, essa non è circolare, ma forma un'ellisse. L'eccentricità orbitale viene definita come la misura della forma ellittica rispetto ad un'orbita completamente circolare, o in generale come deviazione dalla stessa. Tale misura (contrassegnata con ϵ), in sede di valutazione orbitale, varia tra 0 ed 1, e nel caso del nostro pianeta, essa va da un'orbita quasi circolare ($\epsilon=0,005$) a un'orbita discretamente ellittica ($\epsilon=0,058$), attualmente essa si attesta attorno ad un valore pari a 0,017. Secondo Milanković, la più grande delle variazioni della stessa (con un entità pari a $\pm 0,012$) avviene con una periodicità di **413000 anni** e viene denominata **grande eccentricità**, mentre le altre (di valore complessivamente minore e comprese nell'intervallo tra -0,03 e +0,02) si manifestano con periodicità di **100000 anni circa** ed è chiamata **piccola eccentricità**. Milanković teorizzò anche che se la Terra fosse l'unico pianeta che ruota attorno al Sole, l'eccentricità non varierebbe nel tempo, ma la causa comunque di tali variazioni, risiede nell'interazione con i campi gravitazionali di Giove e Saturno, i due pianeti più massivi del sistema solare.



Figura 3.1: L'eccentricità dell'orbita terrestre attorno al Sole

· **OBLIQUITA'**

Il secondo ciclo invece riguarda l'asse di rotazione della Terra, il quale è noto non essere perpendicolare al piano della sua orbita. I valori di tale inclinazione coprono un intervallo che ha un minimo di $22,1^\circ$ ed un massimo di $24,5^\circ$, i quali, secondo Milanković, vengono raggiunti con una periodicità di **41000 anni**, alternandosi in due fasi: una **crescente** (spostamento dell'asse verso l'obliquità massima) e una **discendente** (direzione verso il valore minimo). Nei record di tutte le proxy appartenenti alla carota di ghiaccio di **EPICA Dome**, emerge che nel periodo che va dai 400000 anni agli 800000 anni B.P., essa fu la causa principale dei cambiamenti di valore delle stesse. Infatti si nota come l'aumento dell'inclinazione assiale che corrisponde ad una maggiore insolazione, abbia dunque aumentato nel nostro pianeta la concentrazione di elementi come l'anidride carbonica, polveri atmosferiche, sodio, ecc..., portandolo verso un clima complessivamente più mite e caldo; viceversa la diminuzione dell'inclinazione e dell'irraggiamento solare ha condotto verso le ere glaciali, dove le concentrazioni di tali elementi diminuirono. Oggi l'inclinazione della Terra ha valore pari a $23,45^\circ$ (quasi a metà del percorso) e si trova nella fase discendente del ciclo, che raggiungerà il minimo tra circa 20000 anni.

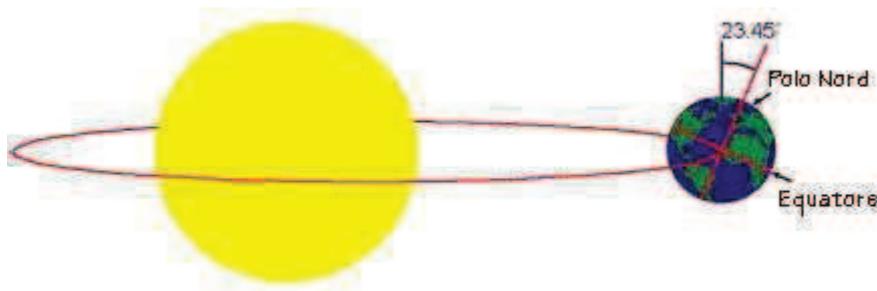


Figura 3.2: L'attuale inclinazione dell'asse terrestre rispetto all'orbita attorno al Sole

· **PRECESSIONE:**

L'ultimo dei tre cicli teorizzati dallo scienziato serbo, riguarda, come il precedente, l'asse di rotazione terrestre. Esso infatti oltre all'inclinazione rispetto ai raggi solari, subisce il cosiddetto *moto di precessione*, ovvero una rotazione completa attorno all'ipotetica retta perpendicolare al piano orbitale (movimento simile a quello di una trottola), il quale è stato calcolato da Milanković essere approssimativamente vicino ai **26000 anni** circa. L'origine di tale caratteristica è l'azione che esercitano le forze gravitazionali della Luna e del Sole, le quali agendo prevalentemente all'Equatore (a causa della forma non perfettamente sferica del nostro pianeta) cercano di allineare la Terra al proprio piano orbitale. Tuttavia, l'azione gravitazionale esercitata dalla Luna e dal Sole non è sempre costante, in quanto esse non si trovano sempre nello stesso piano orbitale e di conseguenza ciò rende il moto di precessione non regolare nel tempo. Essa infatti può distinguersi in due differenti tipologie ossia la *precessione corta*, la quale ha durata **19000 anni** e la precessione lunga che invece è lunga **23000 anni**.

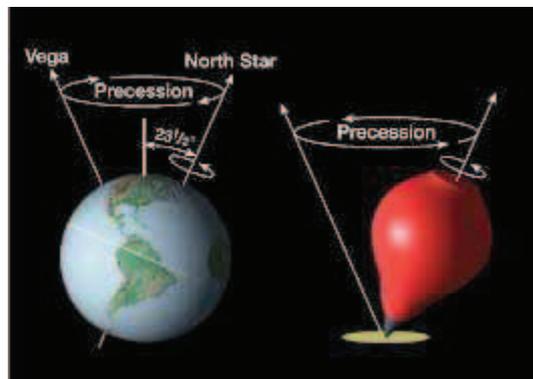


Figura 3.3: Il moto di precessione dell'asse terrestre rispetto alla perpendicolare orbitale

4 Serie equispaziate - caso DANUBIO

4.1 Introduzione e definizione di serie storica climatica

L'esplorazione e la comprensione di molti fenomeni che mutano nel corso di un arco temporale, presuppongono la disposizione di osservazioni empiriche, le quali avendo come criterio ordinatore il tempo stesso (inteso come progressione cronologica), compongono quella che è comunemente definita *serie storica o temporale*. Rilevazioni in istanti temporali successivi su svariati fatti e fenomeni geofisici, astronomici e sociali (anche se non catalogate propriamente sotto forma di serie storiche) ebbero già inizio diversi secoli fa, precisamente a Londra ed in Francia, dove si cominciarono a raccogliere settimanalmente informazioni ufficiali a carattere demografico, come ad esempio il numero di nascite, di matrimoni e di morti.

La percezione della variazione, di tutti i fenomeni nel corso del tempo e le crescenti necessità di conoscenza, analisi ed in alcuni casi di controllo degli stessi, hanno rappresentato l'impulso per ricercare, elaborare e provare metodologie, le quali, (arrivando via via ad esser sempre più sofisticate), hanno convertito l'obiettivo principale da una semplice descrizione del fenomeno, a quello della previsione dello stesso. La notevole esigenza di quest'ultima, ha assunto un ruolo chiave nell'estensione dei metodi di analisi a molteplici contesti applicativi, in primis quello economico-finanziario, ma anche demografico-sociale e non ultimo pure in ambito *fisico-climatico*, il quale riguarda direttamente il lavoro di questo capitolo e nello specifico si riferisce al campo *idrologico*, in quanto si focalizza l'attenzione su serie storiche di portate di un fiume. Nel suddetto campo, l'analisi statistica delle serie temporali è diventata, negli ultimi decenni, uno degli strumenti più utilizzati per la valutazione delle risorse idriche. E' utilizzata, in particolare, per costruire modelli matematici per generare *serie idrologiche sintetiche* (termine utilizzato in contrapposizione alle *serie idrologiche osservate* determinate dai valori realmente misurati); le quali sono utilizzate per la previsione di specifici eventi idrologici, per individuare trend e scostamenti nelle serie, per completare ed ampliare serie lacunose, ecc. In generale i processi idrologici, come ad esempio i deflussi delle acque, le precipitazioni, ecc, evolvono secondo una scala temporale nel *continuo*. Per esempio le altezze di deflusso misurate da una stazione

idrometrica evolvono nel tempo t con continuità. Tuttavia, i più importanti processi idrologici d'interesse sono definiti (e misurati) secondo una *scala temporale discreta ed equispaziata*. Una serie temporale discreta può quindi derivare dal campionare il processo idrologico continuo in tempi discreti, ovvero può essere costruita dall'integrazione dell'evento continuo in intervalli discreti consecutivi ed equidistanti tra loro. Si può dunque, ad esempio, parlare di serie di portate d'acqua di un fiume espresse come portate giornaliere, misurandole i deflussi una sola volta al giorno, oppure, quando si dispongono di osservazioni più frequenti, come portate medie giornaliere corrispondenti al volume complessivo defluito nel giorno. Normalmente le serie idrologiche sono infatti riferite ad intervalli temporali discreti e possono essere orari, giornalieri, settimanali ed anche mensili.

Una serie storica climatica X , e quindi anche idrologica, intesa come processo stocastico, può essere scritta nel seguente modo:

$$X_t = T_t + O_t + N_t \quad t = 1, \dots, n \quad (4.1)$$

Dove:

T_t è la componente di trend che include la parte *deterministica* del processo in esame, la quale si può presentare in tre differenti tipologie:

- incremento lineare o comunque associabile ad una funzione polinomiale;
- cambio netto di livello o di direzione dell'andamento della serie;
- segnale stagionale e/o periodico modellabile con *armoniche* (seno-coseno).

Tutte le caratteristiche elencate sinora sono associate ad una componente nel lungo periodo in maniera non casuale, ma perfettamente prevedibile, tuttavia essa può essere anche aleatoria e tale comportamento prende il nome di *trend stocastico*.

O_t è invece la parte che rappresenta l'insieme degli eventi con un valore assoluto estremamente elevato, usualmente rari e sicuramente ben distanti dalla media del processo.

N_t rappresenta la parte stocastica ed aleatoria ed è sovente assunta essere *debolmente stazionaria* e con una propria *struttura di autocorrelazione*. In

campo climatologico, con l'obiettivo di ricavare quanto possono essere influenti i valori di istanti temporali passati su quello contemporaneo, per poter così provare a stimare il *meccanismo stocastico* che ne governa la dinamica. Per fare ciò, il modello più adatto in tale contesto è il processo autoregressivo di ordine p (denotato con $AR(p)$) e spesso se ne utilizza un suo caso particolare ossia il processo $AR(1)$.

4.2 La correlazione tra due serie storiche climatiche

Lo studio completo dell'evoluzione temporale dei fenomeni climatici (idrologici) non può certamente limitarsi all'analisi delle singole serie storiche univariate, ma essendo presenti congiuntamente all'interno di un sistema complesso, diventa quindi importante studiarne le relazioni che possono intercorrere tra di esse. Si possono trovare molteplici esempi di studi circa le relazioni tra variabili del suddetto campo, uno su tutti è l'analisi della relazione tra il livello del diossido di carbonio (CO_2) e le temperature (il quale verrà rivisitato in seguito nell'ambito dell'analisi delle serie storiche non equispaziate), vi sono studi sulla relazione tra il numero delle macchie solari e le temperature dell'emisfero Nord, si è esaminato anche la relazione tra la velocità del flusso d'acqua nel fondo dell'Oceano Atlantico (nel tratto tra l'Islanda e la Scozia) e l'indice NAO (North Atlantic Oscillation Index), trovando importanti risultati sulla risposta dello stesso Atlantico ai cambiamenti climatici. In ambito idrologico si può essere interessati sia ad osservare set di diversi processi in un stesso sito ed un esempio è dato dal voler studiare la relazione che intercorre tra i deflussi e la quantità di pioggia caduta, sia dal considerare presso due siti differenti uno stesso processo, e sarà l'esempio analizzato dove verrà studiata (e stimata) la correlazione tra le portate d'acqua di due stazioni dislocate su uno stesso fiume.

4.2.1 Moving block bootstrap a coppie

La tecnica di ricampionamento *bootstrap* (Efron 1979) è stata introdotta come evoluzione del metodo *jackknife*, ed è stata fondata per risolvere importanti problematiche inferenziali, soprattutto quando la distribuzione campionaria delle osservazioni non è riconducibile a distribuzioni di probabilità note (in ambito climatologico ciò accade molto spesso) basandosi fortemente su aspetti computazionali,

in particolare di simulazione. L'idea di base di questa tecnica è di simulare dall'unico campione a disposizione (estraendo con reinserimento dallo stesso) B campioni casuali semplici, tutti di numerosità n . Per strutture di dati, non tra loro indipendenti, tale procedimento così come illustrato in precedenza, non può essere applicato. Un esempio è offerto dalle serie storiche, dove le osservazioni mostrano la loro dipendenza proprio attraverso il tempo ed appunto una tale elaborazione comporterebbe la perdita della stessa. Nel caso di serie storiche climatiche (ed anche idrologiche), il bootstrap rappresenterebbe una tecnica di vitale importanza, in quanto molto spesso ci si imbatte con variabili, le cui forme distributive sono distanti da quelle note, quindi fare inferenza in modo corretto su una qualsiasi quantità statistica d'interesse (nel caso in analisi, il coefficiente di correlazione), risulterebbe parecchio problematico. A questo scopo, sono state implementate varianti di tale tecnica per le serie storiche, nelle quali si usa campionare dalle stesse suddividendole in *blocchi*, nell'intento di preservare la struttura di persistenza temporale; una di queste è il *Moving Block Bootstrap*, la quale dopo una breve introduzione, verrà descritta nel dettaglio per le serie bivariate, impiegate nelle successive analisi. Il *Moving Block Bootstrap* è proprio una tecnica di ricampionamento non parametrica per serie storiche, la quale sostanzialmente campiona blocchi di n osservazioni della lunghezza l desiderata (possibilmente con $n/l=k$ che risulti una quantità intera), ottenendo B campioni bootstrap di serie temporali aventi n osservazioni ciascuno; i blocchi (di numerosità pari a $n-l+1$) possono ovviamente sovrapporsi, ma in modo da garantire il mantenimento dell'ordine temporale della serie. Per chiarire meglio il funzionamento di tale tecnica, si prenda in considerazione una serie $\mathbf{x} = (x(t_1), x(t_2), \dots, x(t_n))$, della quale si costruiscono blocchi di lunghezza l , ad esempio $b_1=[x(t_1), x(t_2)], b_2=[x(t_2), x(t_3)], \dots, b_n=[x(t_{n-l+1}), x(t_n)]$; a questo punto si possono ricampionare i blocchi b_i ed inserirvi man mano i loro valori, sino a che non si ritrova all'interno di uno degli stessi, il valore ultimo ($x^*(t_n)$), che ferma la procedura e designa x_1^* , tutto ciò è ripetuto B volte; come si può notare vengono colte alcune parti della dipendenza, altre invece si perdono quando si collegano tra loro i blocchi. L'introduzione della suddetta procedura è premessa importante per capire la successiva evoluzione nell'impiego su serie bivariate, ossia il *Moving Block Bootstrap a coppie* (*Pairwise Moving Block Bootstrap*). Quest'algoritmo è in generale utilizzato per stimare intervalli di confidenza bootstrap per il coefficiente di correlazione (contemporaneo) tra

due serie storiche, quando ovviamente la distribuzione empirica delle stesse non è riconducibile ad una distribuzione di probabilità nota. Il primo passo di questa procedura consiste nel calcolo del coefficiente di correlazione di Pearson r_{xy} tra le due serie $\{x_t, y_t\}$ e la trasformazione dello stesso tramite la z di Fisher ($z = \tanh^{-1}(r_{xy})$), ai fini di ottenere successivamente una distribuzione delle stime bootstrap confrontabili con quelli di una normale. Nel secondo step invece, si stima il coefficiente di autocorrelazione per la componente stocastica di entrambe le serie temporali, in modo tale da trovare la lunghezza ottimale del blocco per il ricampionamento, tramite la formula:

$$l_{opt} = NINT \left(\left[6^{1/2} * \hat{a}_{xy} / (1 - \hat{a}_{xy}^2) \right]^{2/3} * n^{1/3} \right)$$

dove $\hat{a}_{xy} = [\hat{\phi}_{1,x} \hat{\phi}_{1,y}]^{1/2}$. Il terzo passo prevede il ricampionamento a blocchi (pari a B volte) della serie bivariata, il quale avviene collegando le stesse, tramite un indice che ordina un blocco di osservazioni di x_t in contemporanea a quelle di y_t , ciò è utile per mantenere la struttura di persistenza di entrambe. Si ha ora a disposizione B campioni bootstrap da n osservazioni ciascuno della serie bivariata $\{x_t, y_t\}$, dalle quali si può ottenere le B stime della z -Fisher del coefficiente di correlazione di Pearson $\{z^*\}$, calcolando così l'intervallo di confidenza bootstrap prescelto per lo stesso. Sfruttando infine la funzione inversa ($r_{xy}^* = \tanh(z^*)$) si può trasformare ciò che è stato trovato al passo precedente, per arrivare all'IC bootstrap per il coefficiente di correlazione di Pearson r_{xy}^* , chiudendo così la procedura *Moving Block Bootstrap a coppie*. Uno fra i possibili *intervalli di confidenza bootstrap* (ma più usati in ambito climatologico), sono quelli del tipo *t-Student*, il quale si presenta nella seguente forma:

$$[\hat{\theta} + t_v(\alpha) * s\hat{e}_{\hat{\theta}^*}, \hat{\theta} - t_v(\alpha) * s\hat{e}_{\hat{\theta}^*}]$$

Dove $t_v(\alpha)$ è il quantile di livello α della distribuzione t con v gradi libertà, $s\hat{e}_{\hat{\theta}^*}$ è la deviazione standard delle stime bootstrap ed ha la seguente formulazione:

$$s\hat{e}_{\hat{\theta}^*} = \left\{ \sum_{b=1}^B [\hat{\theta}^{*b} - \langle \hat{\theta}^{*b} \rangle]^2 / (B-1) \right\}^{1/2}$$

Nella quale $\langle \hat{\theta}^{*b} \rangle = \sum_{b=1}^B \hat{\theta}^{*b} / B$ è invece la media delle stime bootstrap.

4.3 Applicazione al caso Danubio

4.3.1 L'insieme dei dati

Le serie temporali di riferimento per questo lavoro sono state reperite dal *Global River Discharge Database* e precisamente all'indirizzo web www.sage.wisc.edu/riverdata/. Nel dettaglio, essi riguardano la misura della portata d'acqua mensile (espressa in m^3/sec) del fiume Danubio presso quattro osservatori dislocati lungo il percorso dello stesso e l'arco temporale coperto va dal mese di Gennaio dell'anno 1901 fino al mese di Dicembre del 1984, per un totale di 1008 osservazioni. Le città (più la nazione di provenienza) delle stazioni di rilevamento prese in considerazione, vengono elencate in seguito in ordine di vicinanza alla sorgente del fiume (situata a *Donaueschingen*, Germania), assieme alla loro posizione in **Figura 4.1**:

- 1) *Hofkirchen (Germania)*
- 2) *Bratislava (Slovacchia)*
- 3) *Nagyymaros (Ungheria)*
- 4) *Drobeta-Turnu Severin (Romania)*



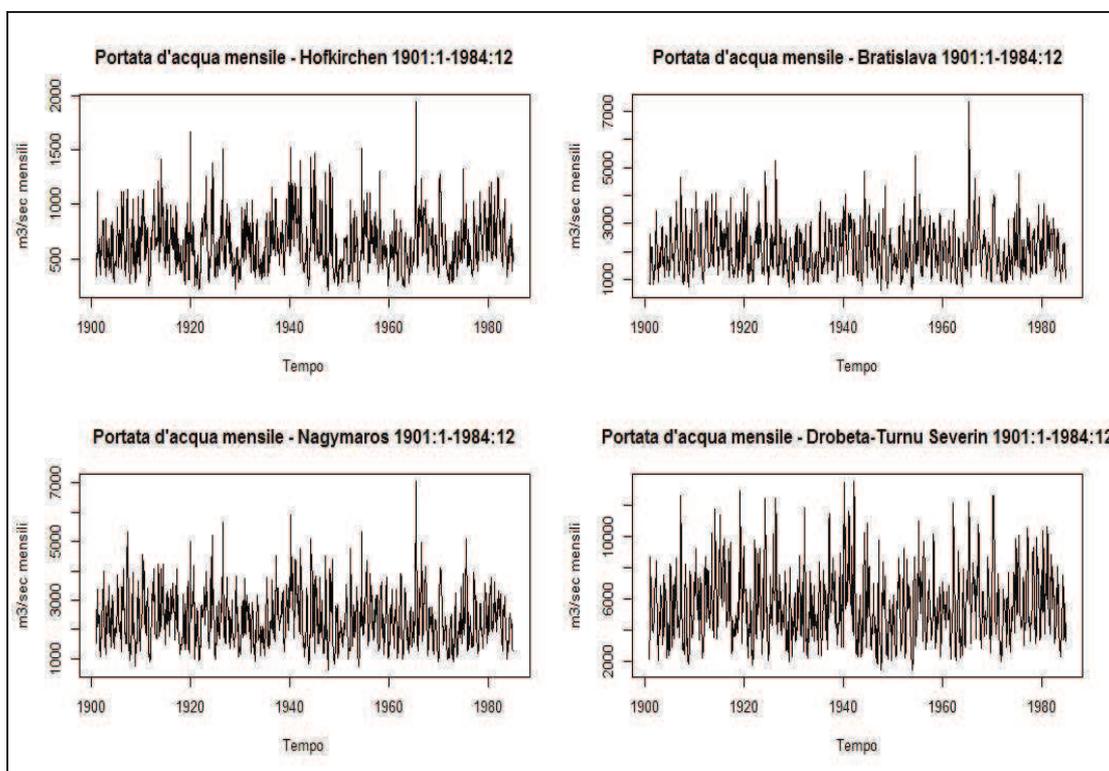
Figura 4.1 Posizione delle stazioni lungo il Danubio

La distanza che intercorre tra la sorgente e la prima stazione di rilevamento (che è posta come si vede in figura 4.1 ad *Hofkirchen*) si attesta sull'ordine dei **477 km**; quest'ultima è situata nella Baviera orientale, la quale precede di poco il confine austriaco, ciò significa che il primo tratto del fiume monitorato, è interamente imputabile al suolo teutonico. La seconda città, della quale si trovano osservazioni, ossia *Bratislava*, dista precisamente **394 km** dalla stazione precedente; la percorrenza di questa seconda parte, si può maggiormente attribuire al suolo austriaco, in quanto la capitale slovacca si trova quasi sul confine tra l'Austria e la Slovacchia stessa. Il terzo centro di monitoraggio, ubicato a *Nagyvaros*, si trova a **203 km** dalla seconda stazione e questa frazione (la più breve) funge da confine naturale tra la parte sud-occidentale del territorio slovacco e quella nord-occidentale di quello ungherese. La quarta ed ultima stazione d'osservazione, collocata a *Drobeta-Turnu Severin*, dista invece **777 km** dalla precedente stazione; questo tratto (indubbiamente il più lungo) interessa diverse nazioni: in primis tutta l'Ungheria centrale, successivamente un'esigua parte della Croazia e del proprio confine con la Serbia; quest'ultima, a sua volta, ne possiede una buona quota, compreso il proprio confine con la Romania, dove poi il Danubio approda per l'ultimo tratto prima della foce, del quale purtroppo non si dispone a sufficienza di osservazioni per l'analisi.

4.3.2 Analisi descrittive

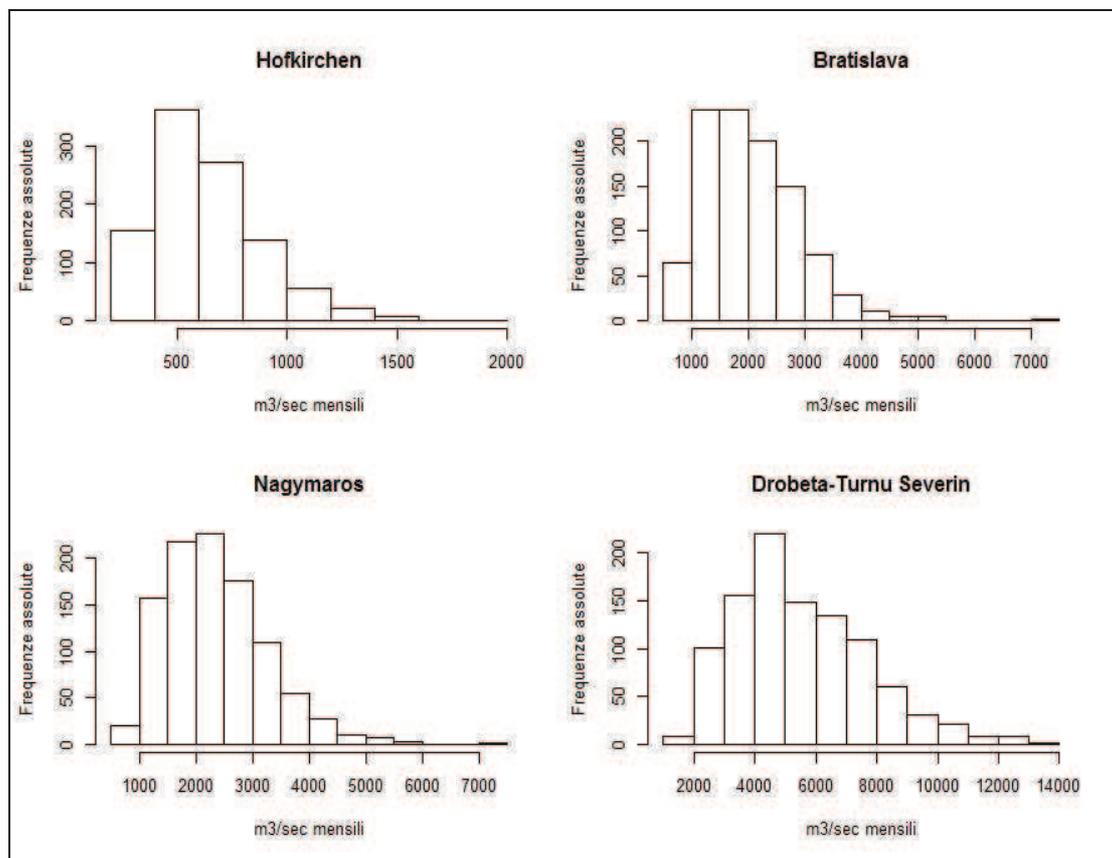
Una prima indagine sulle serie storiche mensili della portata d'acqua del Danubio, riguarda la presentazione in successione dei *grafici*, degli *istogrammi* per ciascuna delle quattro stazioni monitorate, al fine di esplorarne (a carattere puramente grafico) l'andamento nel periodo temporale e la distribuzione dei valori. Oltre a ciò, si mostra anche una tabella che contiene i valori dei principali *indici di posizione*, *di variabilità*, *di curtosi* ed *asimmetria* delle quattro distribuzioni empiriche, i quali forniscono una prima, ma importante, *informazione descrittiva* del fenomeno della portata d'acqua, a supporto di quello grafico. Infine, scendendo nel dettaglio, si utilizzano i *boxplot* mensili, con i quali viene svolta un'analisi della distribuzione per mese dei valori della portata d'acqua, ai fini di individuarne per ciascuna stazione le caratteristiche 'stagionali' nell'arco temporale di un anno. I grafici ed i relativi commenti vengono presentati in modo congiunto per tutte e quattro le stazioni.

Grafici delle portate mensili:



Il grafico riguardante tutte e quattro le serie storiche esaminate, mette in luce un *andamento sostanzialmente stazionario* per quanto riguarda il valore medio (il quale cresce considerando le stazioni ordinate dalla più vicina alla più lontana dalla sorgente), non evidenziando quindi importanti e significative tendenze nel lungo periodo del fenomeno osservato, ma pare soltanto modesti e nel breve periodo; viceversa la conclusione sembra non essere la stessa per quanto concerne la varianza. Quest'ultima considerazione risiede nel fatto che sono presenti alcuni *picchi*, i quali rappresentano livelli massimi di portata del tutto eccezionali e quindi risultano 'anomali' rispetto alla tendenza naturale del fenomeno; per la precisione si può evidenziare che il valore massimo è comune alle serie di Hofkirchen, Bratislava e Nagymaros. Per la stazione di Drobeta- Turnu Severin invece si nota che questi valori sembrano essere un po' più presenti e non coincidono con i livelli massimi delle stazioni precedenti, aumentandone la variabilità.

Istogrammi delle portate mensili:



Dall'istogramma dei valori riguardanti tutte e quattro le serie temporali, emerge, in modo molto evidente, la *non normalità della distribuzione* degli stessi. Un'ulteriore palese aspetto è l'*asimmetria della distribuzione* e precisamente essa lo è verso *sinistra* (maggiormente per Bratislava), anche se quella relativa ai valori per la stazione di Drobeta-Turnu Severin lo sembra essere di meno rispetto alle altre tre; si può inoltre notare l'analogia tra la distribuzione dei valori per le stazioni di Bratislava e di Nagymaros (le stazioni meno distanti tra di loro), dove si registra anche il valore anomalo comune, il quale risulta per entrambe oltre i $7000 \text{ m}^3/\text{sec}$.

Valori degli indicatori descrittivi principali:

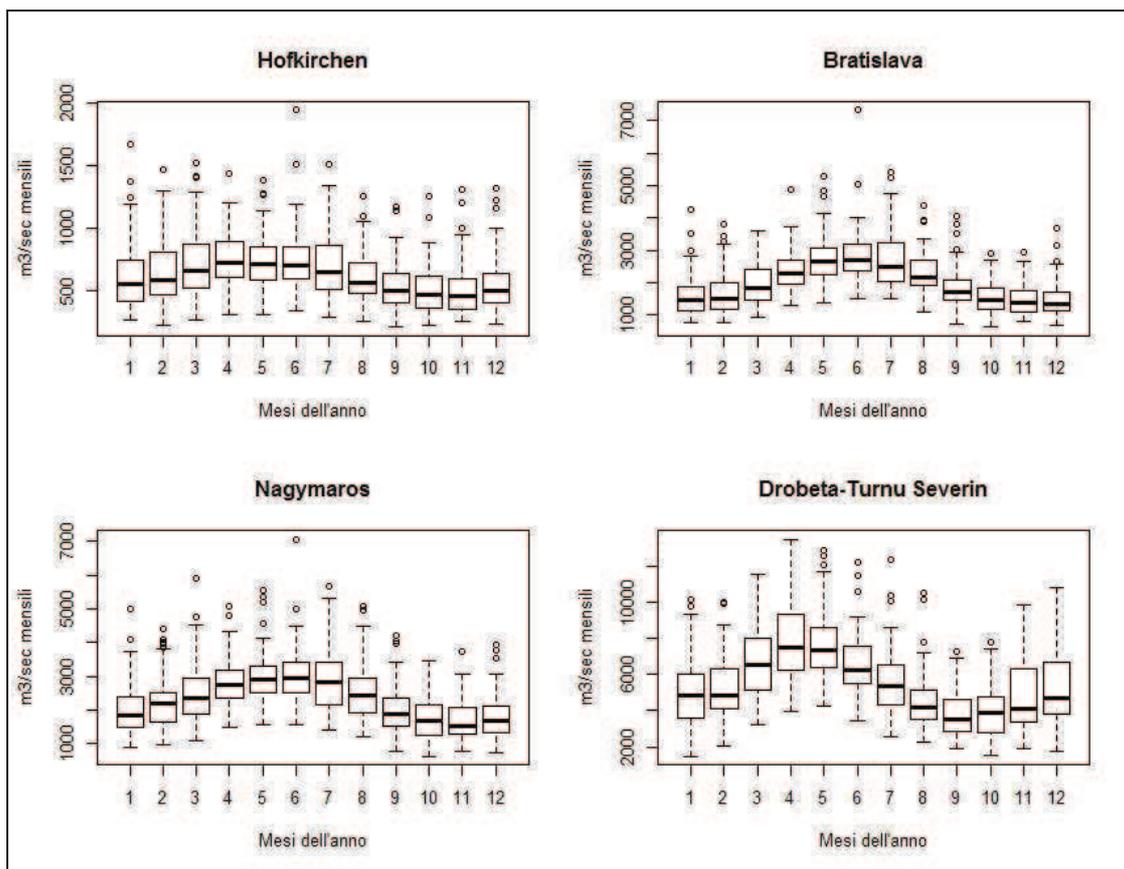
	<i>Hofkirchen</i>	<i>Bratislava</i>	<i>Nagymaros</i>	<i>Drobeta – Turnu Severin</i>
<i>Media</i>	635,46	2044,98	2350,63	5539,39
<i>Mediana</i>	590,5	1924	2227	5205
<i>Varianza</i>	59303,31	66985,6	79001,71	480395,9
<i>Deviazione Standard</i>	243,52	818,44	888,83	2191,79
<i>Asimmetria</i>	0,99	1,03	0,84	0,75
<i>Curtosi</i>	1,43	2,16	1,14	0,44
<i>Valore massimo</i>	1940	7324	7056	13500
<i>Valore minimo</i>	212	633	628	1470
<i>1° Quartile</i>	461	1425	1693	3920
<i>3° Quartile</i>	770	2555	2888,25	6887,5

Tabella 4.1 Statistiche descrittive per stazione di monitoraggio

Dall'analisi di alcuni indicatori descrittivi in *Tabella 4.1*, si può trovare la conferma di quanto è stato osservato e commentato in precedenza, a livello grafico. L'esempio più evidente è dato proprio dalla crescita dei valori di media, mediana e varianza (e di conseguenza anche della deviazione standard) dalla prima stazione monitorata fino alla quarta, in quest'ultima il valore addirittura diventa estremamente più grande rispetto alle precedenti. Un'altra dimostrazione, delle impressioni raccolte graficamente, viene data dai valori degli indici di asimmetria e curtosi, i quali risultano più elevati proprio per

Bratislava, dove la distribuzione è la più asimmetrica verso sinistra, e soprattutto in quest'ultima direzione ha la coda più pesante; conclusioni totalmente opposte vengono formulate per Drobeta-Turnu Severin, anzitutto per l'influenza meno marcata dei valori estremi.

Boxplot mensili delle portate:



Una prima informazione riportata dal grafico dei boxplot mensili delle quattro serie temporali, è sicuramente quella riguardante la differenza dei *valori mediani* delle distribuzioni mensili tra le serie stesse, i quali risultano essere maggiori nei mesi di Aprile e Maggio per le stazioni di Hofkirchen e Drobeta-Turnu Severin (in quest'ultima lo sono in maniera più netta) mentre per i centri di Bratislava e Nagymaros tale caratteristica è rilevata nel mese di Giugno. Una seconda considerazione invece emerge dalla distinzione di due andamenti opposti nel corso dell'anno, ovvero se generalmente nel primo semestre se ne ritrova uno crescente, nel secondo invece si ha la tendenza opposta, in particolare ciò è molto evidente per la stazione di Drobeta-Turnu Severin, mentre lo è di meno per Hofkirchen. Focalizzando l'attenzione sui mesi, si può

affermare che Marzo e Luglio risultano, per le prime tre stazioni, avere valori della portata meno stabili invece per l'ultima oltre a Marzo, lo sono anche Aprile, Novembre e Dicembre, mettendo così in luce differenti regimi. Infine, si può vedere che il valore eccezionalmente massimo, comune ai primi tre centri, si è avuto nel mese di Giugno.

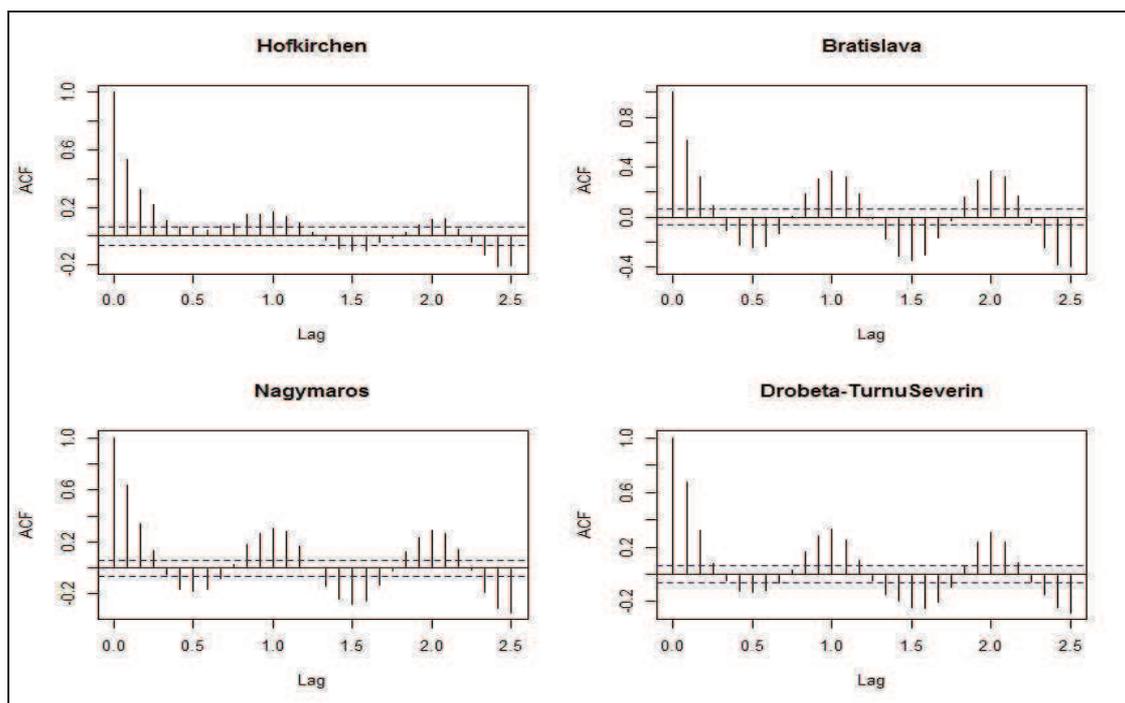
4.3.3 Stima della parte deterministica e degli outliers

Dopo l'aver esplorato, a livello preliminare, ogni singola serie temporale, e l'aver tratto alcune utili informazioni, l'obiettivo di questo paragrafo è la modellazione della componente sistematica della serie e la ricerca dei valori eccezionalmente anomali, al fine di valutarne la loro significatività nel modello per una serie climatica, nello specifico, idrologica.

4.3.3.1 Parte deterministica

Una serie storica climatica, come già introdotto nel 4.1, può presentare la componente deterministica sotto forma di trend o segnale stagionale. Al fine di trarre informazioni circa la periodicità e/o la tendenza di fondo del fenomeno osservato, vengono inizialmente illustrate le funzioni di autocorrelazione globale (ai primi 30 ritardi) per ciascuna serie:

Grafici delle autocorrelazioni globali:



Dalla visione delle funzioni di autocorrelazione globale delle serie storiche inerenti alle quattro stazioni monitorate, balza molto chiaramente agli occhi, la presenza in generale di una forte *componente periodica*, la quale si esibisce sotto forma di *andamento sinusoidale*. Tale aspetto però, sembra essere più marcato per le ultime tre stazioni, in accordo con quanto visto in precedenza nei boxplot mensili. Scendendo nel dettaglio, si può osservare anche che il comportamento periodico della prima serie si manifesta a partire circa dal sesto ritardo in poi, contrariamente a quanto accade per le successive tre stazioni, le quali lo presentano già dal terzo ritardo. Tutto ciò, spinge dunque a stimare la componente sistematica avvalendosi dell'uso delle sole funzioni trigonometriche, escludendo l'impiego di una funzione polinomiale ed ipotizzando per la stessa tale espressione:

$$T_t = \sum_{j=1}^m (\beta_{j1} \cos \omega_j t + \beta_{j2} \sin \omega_j t)$$

Trattandosi chiaramente di serie storiche mensili, si può disporre esattamente fino ad un massimo di 6 *armoniche*, dove al fine di cogliere tale periodicità, verranno inizialmente utilizzate tutte nella stima della parte deterministica di ogni serie, ma delle quali vengono riportati soltanto i valori di quelle risultate statisticamente significative (p-value <0.05%). Quanto appena detto, viene presentato congiuntamente per le quattro stazioni, a pagina seguente:

Tabelle con stime della componente deterministica:

<i>Stazione di Hofkirchen</i>		
	Stima	P-value
<i>Intercetta</i>	635,465	< 2e-16
<i>Cos(2*pi*1/12*t)</i>	-89,87	< 2e-16
<i>Sin(2*pi*1/12*t)</i>	83,41	6.18e-16
R-quadro Multiplo = 0,1269		
R-quadro Aggiustato = 0,1251		

<i>Stazione di Bratislava</i>		
	Stima	P-value
<i>Intercetta</i>	2044,98	< 2e-16
<i>Cos(2*pi*1/12*t)</i>	-694,02	< 2e-16
<i>Sin(2*pi*1/12*t)</i>	62,56	0,0304
<i>Cos(2*pi*1/6*t)</i>	128,96	8,76e-06
R-quadro Multiplo = 0,3752		
R-quadro Aggiustato = 0,3734		

<i>Stazione di Nagymaros</i>		
	Stima	P-value
<i>Intercetta</i>	2350,63	< 2e-16
<i>Cos(2*pi*1/12*t)</i>	-642,11	< 2e-16
<i>Sin(2*pi*1/12*t)</i>	232,34	4,78e-12
<i>Cos(2*pi*1/6*t)</i>	71,49	0,0316
R-quadro Multiplo = 0,2986		
R-quadro Aggiustato = 0,2965		

<i>Stazione di Drobeta-Turnu Severin</i>		
	Stima	P-value
<i>Intercetta</i>	5539,39	< 2e-16
<i>Cos(2*pi*1/12*t)</i>	-990,71	< 2e-16
<i>Sin(2*pi*1/12*t)</i>	1335,78	< 2e-16
<i>Cos(2*pi*1/6*t)</i>	293,91	0,000225
<i>Sin(2*pi*1/6*t)</i>	-584,82	3,62e-13
<i>Cos(2*pi*1/4*t)</i>	301,94	0,000151
R-quadro Multiplo = 0,3423		
R-quadro Aggiustato = 0,339		

Le precedenti tabelle evidenziano in primo luogo come la stima della componente periodica colga meglio la parte deterministica per la serie delle portate di Bratislava, di Nagymaros e Drobeta-Turnu Severin, in quanto l'R quadro (sia multiplo che aggiustato) risulta di gran lunga maggiore rispetto a quello della prima stazione di Hofkirchen. Nello specifico, si può vedere come la frequenza associata al periodo semestrale ($f=1/6$) sia altamente significativa per l'ultima stazione, parzialmente per le due stazioni intermedie (Bratislava e Nagymaros) e totalmente assente per la prima, confermando quanto si era già visto nei boxplot mensili e nei grafici delle autocorrelazioni globali. Infine, una nota merita la significatività di uno dei due coefficienti relativi alla periodicità quadrimestrale (la frequenza relativa è $f=1/4$) per la stazione di Drobeta-Turnu Severin, la quale non si era del tutto dedotta dall'ispezione dei boxplot mensili.

4.3.3.2 Outliers

Gli istogrammi ed i boxplot mensili illustrati nel paragrafo precedente, hanno evidenziato la presenza di alcuni valori massimi eccezionalmente anomali, i quali, in una serie storica di questo tipo, sono comunemente presenti. Trattandosi in tutte le stazioni monitorate di *outliers di tipo additivo*, dove l'effetto prodotto è una repentina variazione nel valore delle portate ad un istante temporale ma immediatamente assorbita a quello successivo, si è pensato di usare come strumento di modellazione le *variabili dummy*, costruite seguendo il criterio:

$$outlier_t = \begin{cases} 1 & \text{se l'outlier è l'i-esima ossevezione dell'istante temporale } t \\ 0 & \text{altrimenti} \end{cases}$$

Le quali verranno inserite nel modello assieme alla parte deterministica, per valutarne non solo la significatività dei coefficienti ma anche quanta variabilità campionaria (in termini percentuali) esse riescono a cogliere. I risultati delle stime ed i relativi commenti, vengono presentati congiuntamente, sempre seguendo (come svolto nelle precedenti analisi) l'ordine di vicinanza alla sorgente. Nel dettaglio, questi valori eccezionalmente massimi, dopo l'ispezione dei dataset e con l'ausilio dei boxplot mensili, risultano essere Giugno 1965 per le prime tre stazioni, Gennaio 1920 per la stazione di Hofkirchen ed Aprile 1940 e 1942 per l'ultima stazione, ossia quella di

Drobeta-Turnu Severin. Le variabili dummy utilizzate per individuare tali valori, sono denominate rispettivamente con *max_Hof_dummy* e *max_Hof_dummy1*, per i massimi della prima stazione, con *max_Bra_dummy* per quello di Bratislava, *max_Nag_dummy* individua quello di Nagymaros ed infine *max_DTS_dummy* e *max_DTS_dummy1* per i massimi della stazione di Drobeta-Turnu Severin.

Tablelle con stima simultanea della componente deterministica e degli outliers:

<i>Stazione di Hofkirchen</i>		
	Stima	P-value
<i>Intercetta</i>	633,193	< 2e-16
<i>Cos(2*pi*1/12*t)</i>	-89,298	< 2e-16
<i>Sin(2*pi*1/12*t)</i>	82,346	< 2e-16
<i>max Hof dummy</i>	1217,509	5,63e-08
<i>max Hof dummy1</i>	1072,969	1,64e-06
R-quadro Multiplo = 0,1709		
R-quadro Aggiustato = 0,1676		

<i>Stazione di Bratislava</i>		
	Stima	P-value
<i>Intercetta</i>	2040,53	< 2e-16
<i>Cos(2*pi*1/12*t)</i>	-685,14	< 2e-16
<i>Sin(2*pi*1/12*t)</i>	62,56	0,0266
<i>Cos(2*pi*1/6*t)</i>	120,08	2,27e-05
<i>max Bra dummy</i>	4478,4	3,09e-12
R-quadro Multiplo = 0,4048		
R-quadro Aggiustato = 0,4024		

<i>Stazione di Nagymaros</i>		
	Stima	P-value
<i>Intercetta</i>	2346,58	< 2e-16
<i>Cos(2*pi*1/12*t)</i>	-634,02	< 2e-16
<i>Sin(2*pi*1/12*t)</i>	232,34	2,59e-12
<i>max Bra dummy</i>	4075,4	3,09e-12
R-quadro Multiplo = 0,3162		
R-quadro Aggiustato = 0,3142		

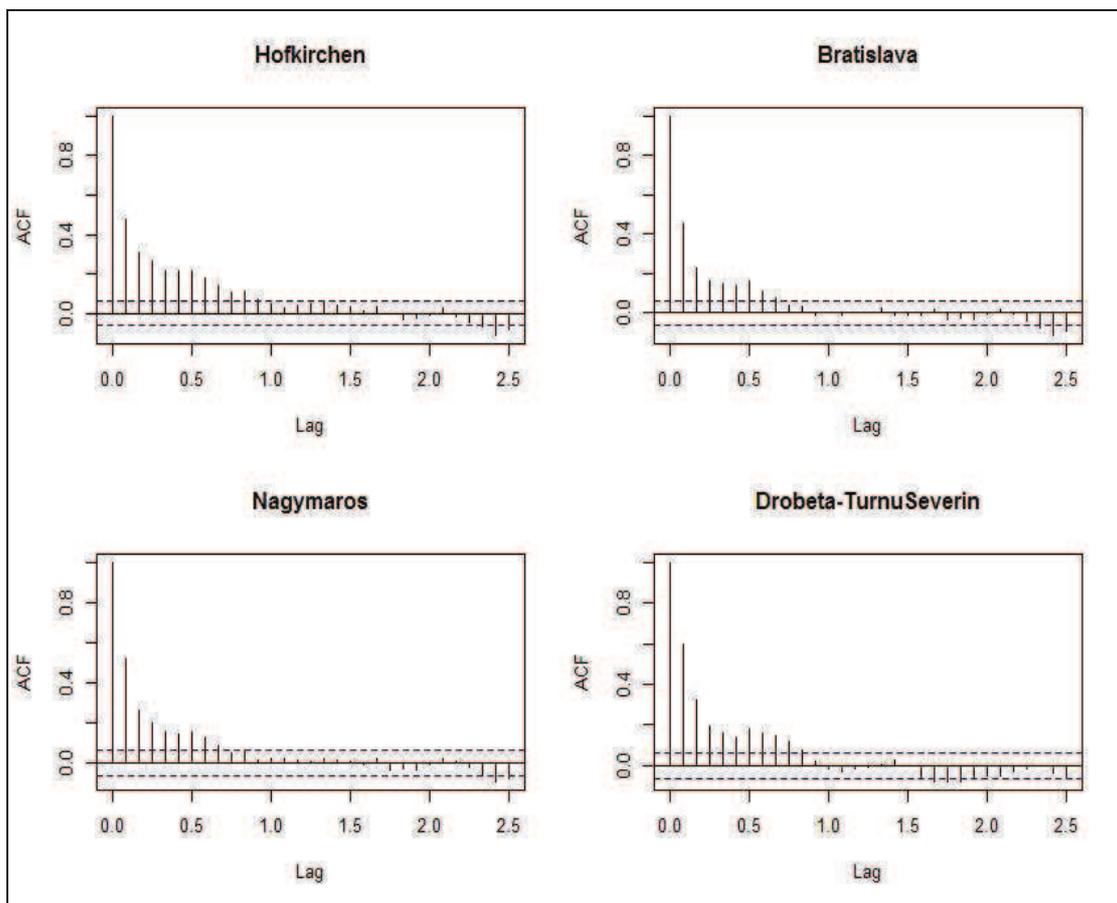
<i>Stazione di Drobeta-Turnu Severin</i>		
	Stima	P-value
<i>Intercetta</i>	5528,13	< 2e-16
<i>Cos(2*pi*1/12*t)</i>	-979,45	< 2e-16
<i>Sin(2*pi*1/12*t)</i>	1316,28	< 2e-16
<i>Cos(2*pi*1/6*t)</i>	305,17	0,000112
<i>Sin(2*pi*1/6*t)</i>	-565,32	1,39e-12
<i>Cos(2*pi*1/4*t)</i>	279,42	0,000410
<i>max DTS dummy</i>	5625,8	0,0001546
<i>max DTS dummy1</i>	5725,8	0,0001273
R-quadro Multiplo = 0,3554		
R-quadro Aggiustato = 0,3509		

Dalle tabelle riportate si possono formulare due considerazioni: una prima porta ad affermare che i coefficienti relativi alle variabili dummy in tutte le regressioni considerate sono statisticamente molto significativi (p-value inferiore all'1%); una seconda invece emerge dagli indici di determinazione (R^2 Multiplo ed R^2 Aggiustato), il quale è più alto rispetto alla stima del modello con la sola componente deterministica per tutte e 4 le serie delle stazioni considerate, ma non in modo omogeneo. Infatti per il primo centro (Hofkirchen) tale aumento è il più significativo tra le stazioni in esame ed è nell'ordine del 5% circa, viceversa i due valori massimi trovati per Drobeta-Turnu Severin accrescono entrambi gli R^2 in maniera più modesta (1,3%) e la causa è la presenza di altre osservazioni nella serie di entità non troppo differente a tali valori; per le stazioni intermedie l'introduzione delle dummy per i valori massimi provoca un aumento degli R^2 rispettivamente del 3% (Bratislava) e del 2% (Nagymaros). Infine, una particolarità è data proprio dalla stazione di Nagymaros, in quanto l'inserimento del modello della dummy per il valore massimo del mese di Giugno del 1965, rende contemporaneamente non significativo il coefficiente relativo al coseno della seconda armonica (la quale si riferisce alla periodicità semestrale) e viene dunque rimosso dal modello.

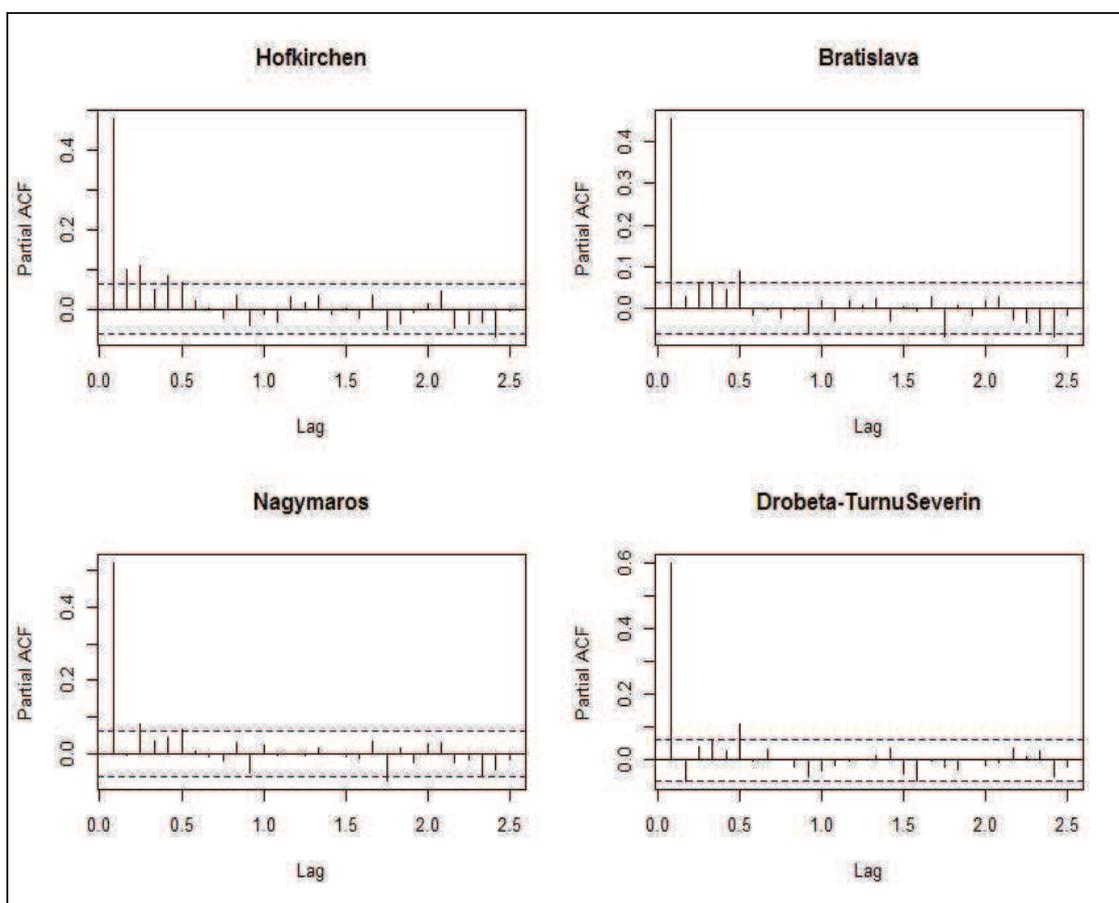
4.3.4 Stima modello $AR(p)$ per la parte stocastica

Nel paragrafo precedente, si sono stimati sia la parte deterministica che gli outliers, al fine di poter notare quanto essi potessero incidere in ognuna delle serie temporali delle quattro stazioni di rilevamento. Tuttavia, l'obiettivo principale è riuscire a cogliere le dinamiche della componente stocastica, la quale viene ottenuta togliendo dalle serie originali entrambe le parti stimate in precedenza; quindi considerando l'espressione 4.1, rimane soltanto N_t . Successivamente, se ne riportano le funzioni di autocorrelazione globale e parziale (ai primi 30 ritardi), al fine di individuare l'ordine della *parte autoregressiva* (p) che meglio potrebbe cogliere la dinamica di N_t . Infine vengono presentati i risultati della stima del modello $AR(p)$ ed il grafico delle autocorrelazioni globali e parziali dei residui (con i successivi commenti) considerando congiuntamente le quattro stazioni, ma rispettando l'ordine di vicinanza alla sorgente.

Grafici delle autocorrelazioni globali delle parti stocastiche delle serie:



Grafici delle autocorrelazioni parziali delle parti stocastiche delle serie:



Dai grafici delle funzioni di autocorrelazione globale della parte aleatoria, si evince chiaramente che il valore della stessa decade con un *andamento pseudo-esponenziale*, per tutte e quattro le stazioni di monitoraggio; mentre da quelli delle funzioni di autocorrelazione parziale si nota (sempre per tutte e quattro le stazioni) la dimensione nettamente più grande del primo ritardo, rispetto a quelli successivi; tutto ciò farebbe propendere verso l'impiego di un $AR(1)$ per la modellazione della componente stocastica di tutte le serie considerate. Andando a stimare dunque tale modello (senza considerare la costante ϕ_0 in quanto N_t ha media pari a zero), si riportano i risultati nelle tabelle sottostanti:

Tabella stima modello AR(1) per le parti stocastiche delle serie considerate:

Stazione di Hofkirchen	
Stima Modello AR(1)	
<i>Stima coefficiente</i>	<i>Standard Error</i>
0.4798	0.0276
Valore Criterio <i>AIC</i> = 2599.81	
Valore Criterio <i>BIC</i> = 2609.64	

Stazione di Bratislava	
Stima Modello AR(1)	
<i>Stima coefficiente</i>	<i>Standard Error</i>
0.4558	0.0280
Valore Criterio <i>AIC</i> = 2628.93	
Valore Criterio <i>BIC</i> = 2638.76	

Stazione di Nagymaros	
Stima Modello AR(1)	
<i>Stima coefficiente</i>	<i>Standard Error</i>
0.5204	0.0269
Valore Criterio <i>AIC</i> = 2545.12	
Valore Criterio <i>BIC</i> = 2554.95	

Stazione di Drobeta-Turnu Severin	
Stima Modello AR(1)	
<i>Stima coefficiente</i>	<i>Standard Error</i>
0.6021	0.0252
Valore Criterio <i>AIC</i> = 2410.85	
Valore Criterio <i>BIC</i> = 2420.69	

La prima considerazione che emerge dalla visione delle tabelle precedenti è che il valore della stima del coefficiente ϕ_1 è più alto per la parte stocastica della serie riguardante la stazione di Drobeta-Turnu Severin, a conferma della forte variabilità dei valori di quest'ultima. Una seconda ed ultima considerazione viene calcolando i valori

delle statistiche test per i coefficienti ϕ_1 di ogni serie, ottenuti dal rapporto tra le stime degli stessi ed i loro relativi standar error, riassunti nella seguente tabella:

<i>Stazioni considerate</i>	<i>Valori della statistica test per ϕ_1</i>
<i>Hofkirchen</i>	<i>18.04</i>
<i>Bratislava</i>	<i>16.28</i>
<i>Nagymaros</i>	<i>19.35</i>
<i>Drobeta-Turnu Severin</i>	<i>23.89</i>

Dove si nota, che i coefficienti risultano tutti significativi, in quanto i valori della statistica test sono sicuramente maggiori se confrontati con il livello soglia del 5% . Successivamente si va ad osservare le funzioni di autocorrelazione globale e parziale dei residui per tutte e quattro le stazioni:

Grafico delle autocorrelazioni globali dei residui del modelloAR(1):

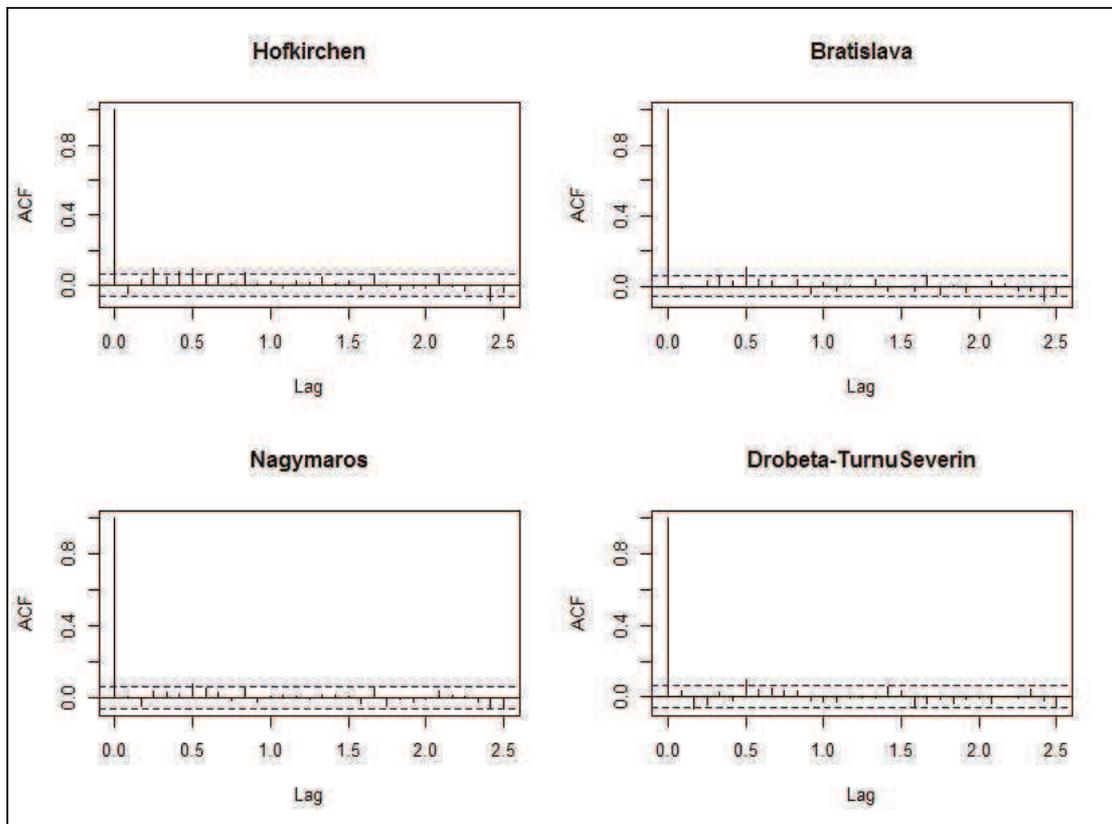
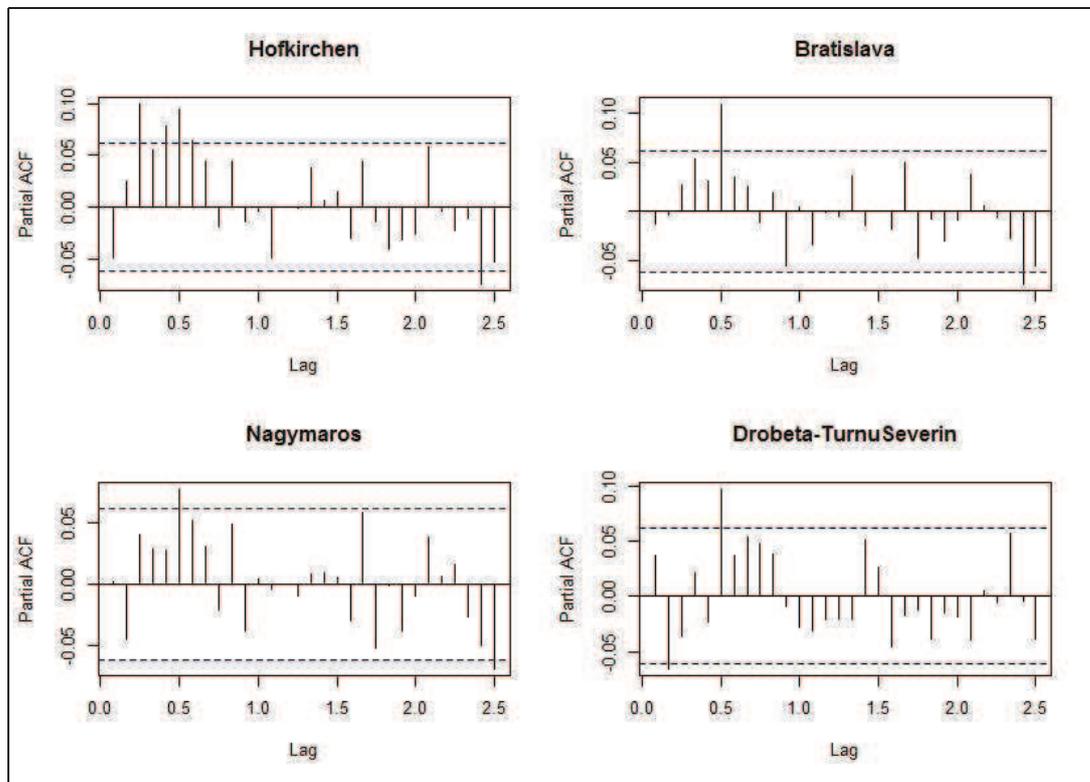


Grafico delle autocorrelazioni parziali dei residui del modello AR(1):



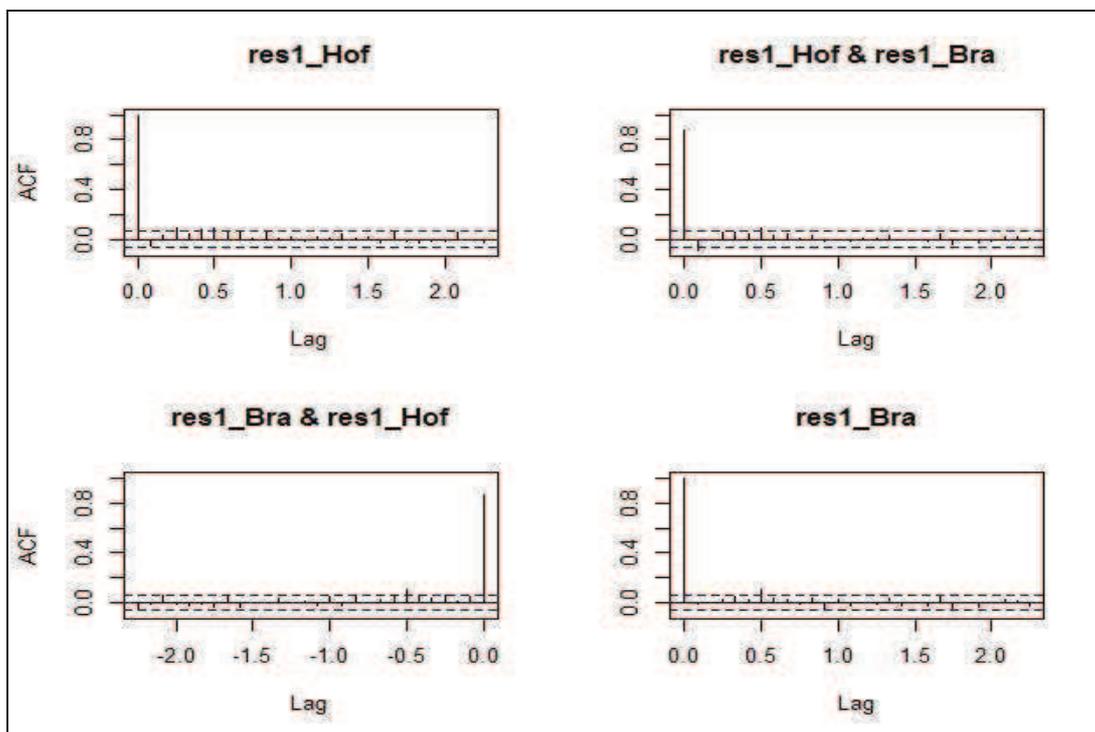
Dall'andamento delle funzioni di autocorrelazione globale e parziale dei residui (primi 30 ritardi) si deduce che sostanzialmente il modello AR(1) coglie abbastanza bene le dinamiche della parte stocastica delle serie considerate. Focalizzando l'attenzione sui grafici delle autocorrelazioni parziali si trova che per le serie di Bratislava, Nagymaros e Drobeta-Turnu Severin, si ha un solo ritardo abbastanza significativo e per la precisione è il sesto, ciò non è comunque sufficiente per cercare un ordine p maggiore di 1. Per la stazione di Hofkirchen invece si ha una situazione un po' differente dalle altre, in quanto vi sono altri due ritardi prima del sesto che risultano significativi; tuttavia la ricerca di modelli AR(p) con ordini del p superiori al primo, rendeva distorta la stima dei valori della funzione di autocorrelazione, inducendo altra correlazione seriale nei residui. In definitiva dunque si è adottato come processo generatore delle osservazioni di ciascuna componente stocastica, il modello AR(1).

4.3.5 Correlazione tra coppie di stazione successive

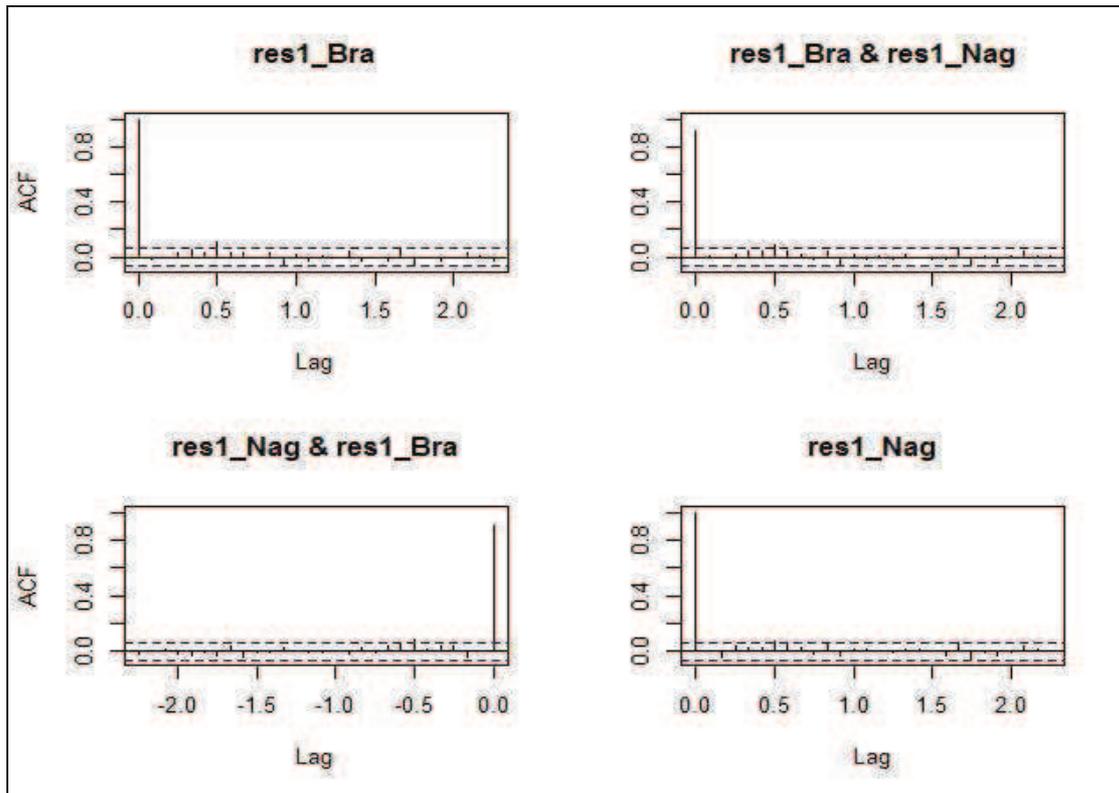
Nei sottoparagrafi precedenti l'obiettivo principale era di analizzare separatamente, o meglio, a livello *univariato* le quattro serie storiche idrologiche, stimandone inizialmente la componente sistematica T_t e la parte legata ai valori eccezionalmente elevati O_t , per arrivare poi all'analisi della struttura di dipendenza della componente stocastica N_t , stimando il modello AR(1), il quale ne coglie bene le dinamiche. Tale componente dunque servirà nello studio della correlazione tra due serie idrologiche successive, nella quale si considera il legame tra una stazione e la sua immediata successiva, vero obiettivo di tutto lo studio. L'operazione iniziale è quella di andare a ricavare la struttura della funzione di auto e cross-correlazione delle serie residue dei modelli AR(1) considerate a coppie, dove se ne vuole osservare il comportamento con l'obiettivo di andare a scoprire la presenza di possibili dipendenze tra le stesse a ritardi temporali differenti e valutare la cross-correlazione contemporanea. I risultati di tale procedura saranno illustrati per ciascuna coppia di stazioni successive, mediante l'ausilio grafico, mentre il commento complessivo si trova al termine.

Grafici delle funzioni di auto e cross-correlazione globali dei residui modello AR(1):

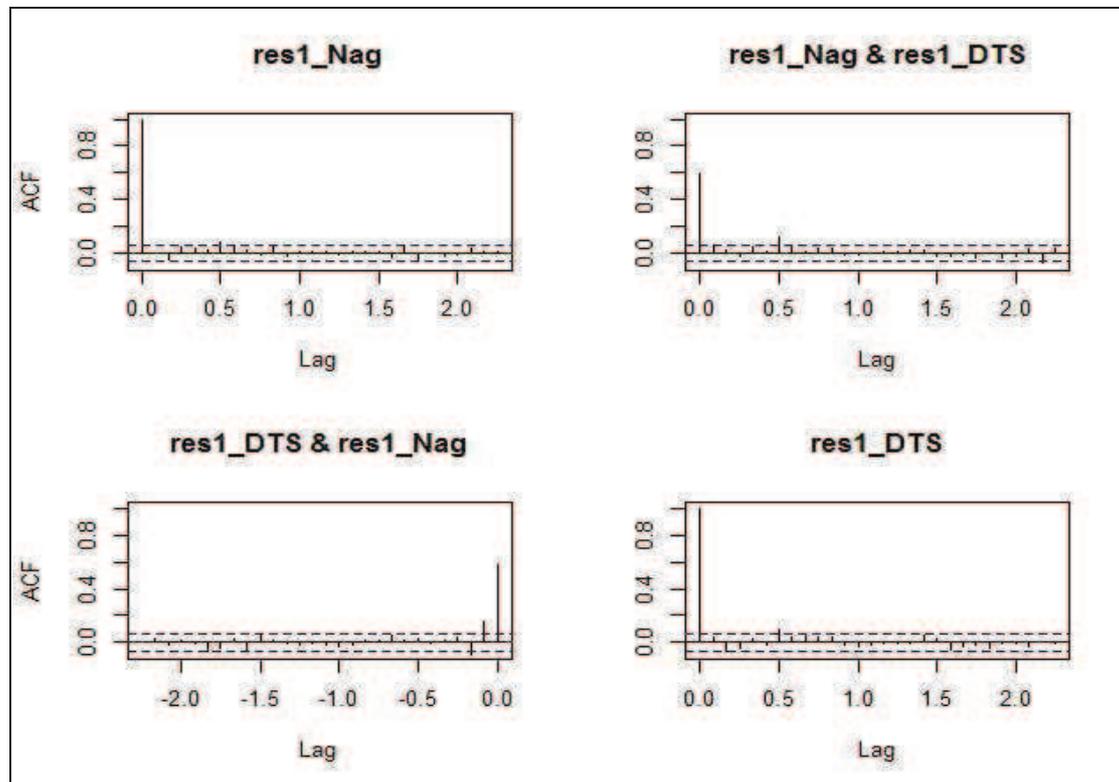
Stazioni di Hofkirchen e Bratislava



Stazioni di Bratislava e Nagymaros



Stazioni di Nagymaros e Drobeta-Turnu Severin



Dai grafici delle auto e cross-correlazioni globali dei residui del modello AR(1) per le componenti stocastiche di tutte le serie storiche riguardanti le stazioni considerate in coppia, si evince chiaramente la sostanziale assenza di relazioni d'anticipo e/o ritardo tra esse (relazioni di *feedback*), nonostante la presenza di qualche ritardo leggermente significativo (relazione tra le ultime due stazioni considerate). Risulta invece molto marcata la cross-dipendenza contemporanea, la quale però presenta un valore inferiore per Nagymaros e Drobeta-Turnu Severin; aspetto che rientra comunque nelle aspettative, in quanto esse risultano le più distanti, considerandole a coppie di stazioni immediatamente successive.

Dunque, una volta notato che la struttura di cross-correlazione tra coppie di serie, non ha evidenziato alcuna relazione di anticipo o di ritardo, o entrambe, l'attenzione ora sarà riservata al coefficiente di correlazione contemporaneo, dandone oltre ad una valutazione puntuale, anche una stima intervallare esatta. Nello svolgere quest'ultima operazione si trova una chiara problematica d'inferenza, perché le distribuzioni di frequenza dei valori delle portate per ciascuna stazione, non risultano associabili a distribuzioni di probabilità note; quindi diviene necessario mettere in campo tecniche che (in questo esempio nel calcolo dell'intervallo di confidenza per il coefficiente di correlazione) prescindano da assunzioni legate alla distribuzione di probabilità, ma che permettano di ricondursi il più possibile alla stessa. Questa tecnica prende il nome di *moving block bootstrap a coppie*, dove il criterio la lunghezza del blocco per il ricampionamento bootstrap (la quale si basa sulla struttura di autocorrelazione di entrambe le serie storiche) viene determinato secondo la formula vista nel 3.2.2, ma adatta al caso bivariato:

$$l_{opt} = NINT \left(\left[6^{1/2} \hat{\phi}_{1,xy} / (1 - \hat{\phi}_{1,xy}^2) \right]^{2/3} n^{1/3} \right)$$

dove $\hat{\phi}_{1,xy} = [\hat{\phi}_{1,x} \hat{\phi}_{1,y}]^{1/2}$ e $NINT ()$ sceglie l'intero più vicino, $\hat{\phi}_{1,x}$ e $\hat{\phi}_{1,y}$ sono i due coefficienti di autocorrelazione stimati per la parte stocastica delle due serie temporali di interesse. Nel voler dare però una visione più precisa di quanto è stato fatto, risulta necessario elencare i passi principali di tale tecnica, applicati sulle serie di riferimento. Inizialmente calcolando il coefficiente di correlazione di Pearson (denotato con r_{xy}) tra le parti stocastiche delle due serie temporali d'interesse e lo si è normalizzato tramite la *trasformazione z-Fisher* ($z = \tanh^{-1}(r_{xy})$). Successivamente si è

determinato la lunghezza ottimale (l_{opt}) del blocco per il ricampionamento, secondo il criterio visto in precedenza, sostituendo a $\hat{\phi}_{1,x}$ e $\hat{\phi}_{1,y}$ il valore trovato nella stima del modello AR(1) per le parti stocastiche delle serie esaminate. A questo punto, si è eseguito il *ricampionamento bootstrap a blocchi* sulle parti stocastiche della serie bivariata, il quale è stato effettuato in due cicli: un primo svolto con **B pari a 2000** (partendo dalle parti stocastiche delle serie originali) ed in seconda battuta un altro (utilizzando le serie ricampionate dopo il primo ciclo) con **B pari a 1000**, dove **B** viene inteso come il numero di ricampionamenti della serie bivariata, in quanto *l'intervallo di confidenza bootstrap* (per $z = \tanh^{-1}(r_{xy})$) calcolato sarà del tipo *t di Student*. Il penultimo passaggio è il calcolo di tale intervallo, il quale è stato fatto usando lo standard error delle stime bootstrap della z dopo il 2° ciclo; infine utilizzando la relazione inversa $r_{xy} = \tanh(z)$ si può ottenere l'intervallo cercato, ossia quello per il coefficiente di correlazione di Pearson.

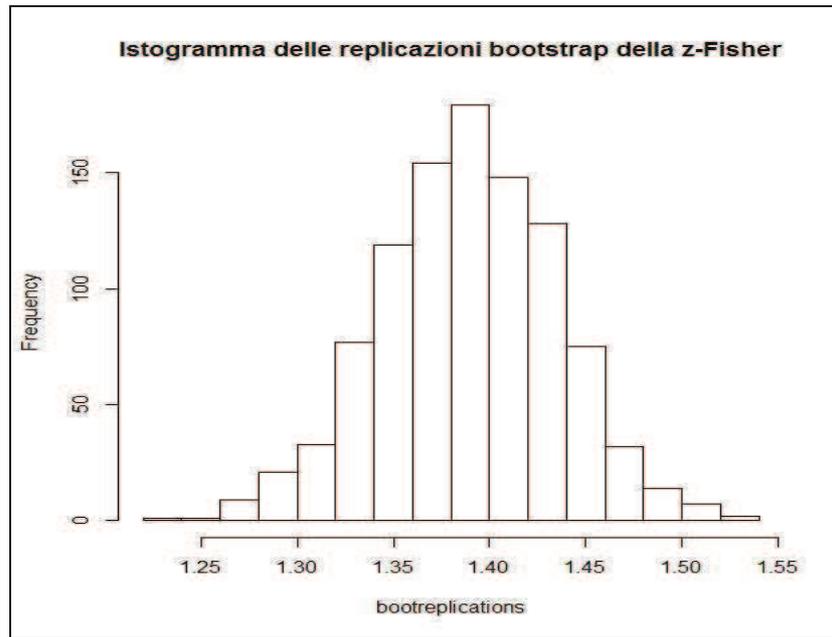
In relazione al primo passo di tale procedura, i coefficienti di correlazione di Pearson r_{xy} e le relative trasformazioni *z-Fisher* (denotate con $z(r_{xy})$) trovati per le parti stocastiche delle serie storiche d'interesse, considerate a coppie sono riassunti nella seguente tabella:

<i>Stazioni considerate</i>	r_{xy}	$z(r_{xy})$
Hofkirchen e Bratislava	0,865	1,313
Bratislava e Nagymaros	0,917	1,57
Nagymaros e Drobeta-Turnu Severin	0,675	0,819

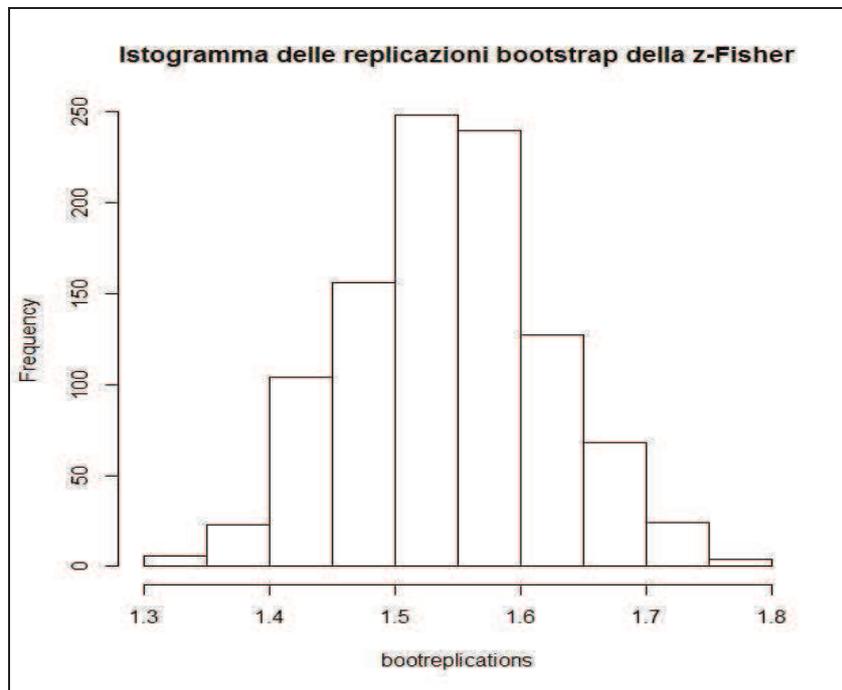
La lunghezza ottimale del blocco per il ricampionamento (secondo il criterio elencato prima) delle parti stocastiche delle serie bivariate è risultata pari a **14** sia per la coppia Hofkirchen-Bratislava che per quella Bratislava-Nagymaros, mentre per l'ultima si è trovata pari a **18**.

Dopo il secondo ciclo di ricampionamento si sono ottenute le seguenti distribuzioni delle stime bootstrap per la *z-Fisher* (z^{*b}) del coefficiente di correlazione, riportate a pagina seguente:

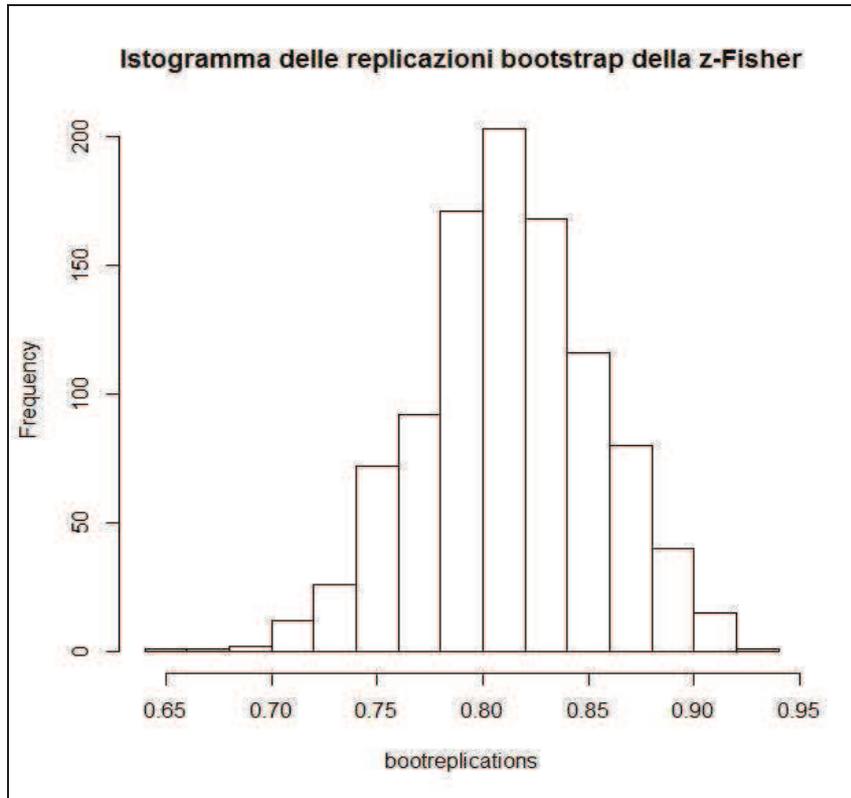
Stazioni di Hofkirchen e Bratislava



Stazioni di Bratislava e Nagymaros



Stazioni di Nagymaros e Drobeta-Turnu Severin



Oltre alla forma distributiva, si riportano le tabelle inerenti alla media, varianza e deviazione standard di tali stime bootstrap, per ogni coppia di stazioni considerata:

	<i>Valore tra Hofkirchen e Bratislava</i>
<i>Media Stime Bootstrap z-Fisher</i>	<i>1,389004</i>
<i>Varianza Stime Bootstrap z-Fisher</i>	<i>0,0020961</i>
<i>Deviazione Standard Stime Bootstrap z-Fisher</i>	<i>0,0457836</i>

	<i>Valore tra Bratislava e Nagymaros</i>
<i>Media Stime Bootstrap z-Fisher</i>	<i>1,543073</i>
<i>Varianza Stime Bootstrap z-Fisher</i>	<i>0,0062972</i>
<i>Deviazione Standard Stime Bootstrap z-Fisher</i>	<i>0,0793547</i>

	<i>Valore tra Nagymaros e Drobeta-Turnu Severin</i>
<i>Media Stime Bootstrap z-Fisher</i>	<i>0,8122129</i>
<i>Varianza Stime Bootstrap z-Fisher</i>	<i>0,0017388</i>
<i>Deviazione Standard Stime Bootstrap z-Fisher</i>	<i>0,0416989</i>

Gli intervalli di confidenza bootstrap del tipo t-di Student per la trasformazione z-di Fisher del coefficiente di correlazione di Pearson (nella forma $[\hat{z} + t_v(\alpha) * s\hat{e}_{\hat{z}}; \hat{z} - t_v(\alpha) * s\hat{e}_{\hat{z}}]$) sono ottenuti usando la deviazione standard delle stime bootstrap per la stessa (per la formula si rimanda al 4.2.1) presentata in tabella. La *t di Student*, nella quale si va a ricercare il quantile (di livello $\alpha=0.05$) da inserire nell'intervallo di confidenza bootstrap, ha $\nu=2n-5$ gradi di libertà; quindi ciò significa che per tutti e tre gli intervalli considerati essi sono pari a 2011 (tutte le serie hanno numerosità campionaria pari a 1008), un valore molto elevato, per il quale la *t-di Student* assume le caratteristiche di una distribuzione normale. Dunque, nel computo degli intervalli di confidenza bootstrap, verrà impiegato il tipico quantile di una normale standard a livello 0,95, ovvero 1,96. Considerando quanto detto, gli intervalli di confidenza bootstrap del tipo di Student per la z-di Fisher trovati per ciascuna delle tre coppie (parti stocastiche delle serie bivariate) di stazioni, vengono riassunti nella seguente tabella:

<i>Coppie di Stazioni considerate</i>	<i>IC bootstrap di tipo t z-Fisher</i>
Hofkirchen e Bratislava	<i>(1,223264 ; 1,402736)</i>
Bratislava e Nagymaros	<i>(1,414465 ; 1,725535)</i>
Nagymaros e Drobeta-Turnu Severin	<i>(0,737270 ; 0,900730)</i>

Ed adoperando la trasformazione inversa si trova infine l'intervallo per ciascuna coppia del coefficiente di correlazione di Pearson r_{xy} , i quali sono presentati anche qui in tabella:

<i>Coppie di Stazioni considerate</i>	<i>IC bootstrap di tipo t per r_{xy}</i>
Hofkirchen e Bratislava	<i>(0,840614;0,885942)</i>
Bratislava e Nagymaros	<i>(0,888439;0,938526)</i>
Nagymaros e Drobeta-Turnu Severin	<i>(0,627493;0,716653)</i>

5 Serie non equispaziate - caso Vostok

5.1 Introduzione serie climatiche non equispaziate

Diversamente da quanto considerato nel capitolo precedente, lo studio svolto in questa sezione focalizzerà la sua attenzione sulle *serie temporali non equispaziate*, nelle quali viene considerato un intervallo di tempo intercorrente tra un'osservazione e quella immediatamente precedente, non omogeneo in tutto il campione di dati considerati, i.e. $t_i - t_{i-1} = d_i \neq \text{costante}$. Esempi di dati, aventi questa peculiarità, sono presenti in svariati campi delle scienze naturali, in particolare nella *paleoclimatologia*, disciplina nella quale viene ricostruita la storia dei cambiamenti climatici del pianeta, attraverso osservazioni che risalgono fino ad un milione di anni fa. Record di questo tipo però ovviamente (e ragionevolmente) non sono possono essere stati rilevati dall'uomo agli istanti temporali di riferimento, ma l'informazione necessaria per affrontare questo tipo di studio è contenuta nei cosiddetti “*Archivi Naturali*”, ossia successioni di materiali geologici o paleo-biologici, nei quali sono registrati parametri connessi alle condizioni climatiche presenti al momento della loro formazione. Da questi è possibile estrapolare una serie di dati indiretti (*proxy data*), i quali sono il risultato dell'interazione di più parametri ambientali legati al sistema climatico e la cui interpretazione sta alla base della paleoclimatologia stessa; essi tuttavia contengono anche segnali climatici, che a volte possono essere di lieve entità o sono influenzati da altri segnali estranei (non climatici), i quali risulterebbero però irrilevanti ai fini della caratterizzazione paleoclimatica. Le *proxy variables*, associate agli “archivi naturali”, sono prelevate in diversi punti del pianeta ed il loro valore informativo (a livello paleoclimatico) è fortemente dipendente dal minimo intervallo di campionamento e dalla risoluzione della datazione di ciascun elemento, i quali rappresentano i due principali fattori che determinano il grado di dettaglio ottenibile dalle stesse, oltre che di differenza di orizzonte temporale nel quale sviluppare gli argomenti ed obiettivi di analisi; la tabella riportata a pagina seguente riporta i principali “*Archivi Naturali*” e le relative *proxy variables*, con gli intervalli di tempi coperti dal campione di rilevamento e risoluzione della datazione di ciascuna, espresse in decine di anni (e relative potenze degli stessi)

<i>Climate archive</i>	<i>Location</i>	<i>Time range (a)</i>	<i>Proxy variable</i>	<i>Resolution (a)</i>	<i>Climate variable</i>
Marine sediment core	Eastern equatorial Pacific	10 ⁶	$\delta^{18}\text{O}_2$, benthic foraminifera	10 ³	Ice volume, bottom water temperature
Ice core	Antarctica	10 ⁵	CO ₂ , air bubbles	10 ³	CO ₂ , atmosphere
	Greenland	10 ⁵	δD , ice	10 ²	Air temperature
			SO ₄ content, ice	10 ⁹	Volcanic activity
			Ca content, ice	10 ⁹	Aeolian dust, wind
			Dust content, ice	10 ⁹	Aeolian dust, wind
Conductivity, ^a ice	10 ⁹	Soluble material, wind			
Na content, ice	10 ⁹	Seasalt, wind			
Tree-rings	Worldwide	10 ⁴	$\Delta^{14}\text{C}$, wood	10 ⁹	Solar irradiance, ocean circulation
Lake sediment core	Boston area	10 ³	Varve thickness	10 ⁹	Wind ^b
Speleothem	Southern Oman	10 ⁴	$\delta^{18}\text{O}$, carbonate	10 ³	Monsoon rainfall
Documents	Weikinn source texts	10 ³		10 ⁹	Floods, river Elbe
Climate model	Hadley Centre, HadCM3	10 ²		10 ⁹	River runoff
Direct measurements	Siberia, North Atlantic	10 ²		10 ⁻¹	Surface temperature

Tabella 5.1: Principali “archivi climatici” e proxy variables (Mudelsee, 2010)

Si nota inoltre che, i dati relativi a ciascuna *proxy variable* vengono associati ad una variabile climatica e fra tutti quelli menzionati nella tabella, indubbiamente le rilevazioni fatte nelle calotte polari in Groenlandia ed in Antartide, costituiscono infatti i più completi archivi di informazioni climatico-ambientali. Perforazioni profonde, grazie alle quali si raccolgono lunghissimi campioni, comunemente chiamati *carote di ghiaccio*, hanno consentito di ottenere relazioni dettagliate su molti parametri paleoclimatici registrati contemporaneamente nei due emisferi; tutto ciò ne ha permesso un confronto, che si è rivelato di grandissima importanza per la verifica di ipotesi su fenomeni accaduti in passato e sulle loro cause. Tra le principali analisi che vengono eseguite sulle carote di ghiaccio vi sono quelle di tipo *chimico e geochimico* (svolte negli isotopi stabili dell’acqua ed in quelli cosmogenici), quelle di tipo *atmosferico*, che riguardano sia i gas contenuti nelle bolle d’aria isolate nel ghiaccio (livelli di CO₂ e CH₄) sia le polveri atmosferiche (o aerosol) presenti (attraverso l’analisi di proprietà fisiche e mineralogiche). Le prime, hanno dato luogo a proxy molto importanti nelle

ricerche di questa scienza, infatti gli isotopi stabili dell'acqua ($\delta^{18}O$ e δD) sono utilizzati nelle ricostruzioni delle *paleotemperature*; mentre le seconde, in particolar modo quelle relative al CH_4 , rendono possibile la datazione e la sincronizzazione dei records nei due emisferi; tuttavia queste ultime rappresentano uno dei principali problemi per lo studio delle carote di ghiaccio, a causa della difficoltà di stabilire una relazione tra profondità ed età degli strati. Nell'intento dunque di avere una corretta datazione degli stessi, sono stati sviluppati diversi approcci, tra i quali i più utilizzati sono sostanzialmente tre: un primo riguarda il *conteggio degli strati* della carota, qualora esso fosse possibile, in quanto le variazioni di temperatura, la presenza di elementi in tracce e gli isotopi stabili dell'ossigeno e dell'idrogeno, presentano ciclicità stagionale (tutto ciò però necessita di accumuli decisamente elevati); un secondo riguarda *la costruzione di modelli di scorrimento dei ghiacciai*, i quali tengono conto congiuntamente di molteplici caratteristiche tra le proprietà fisiche del ghiaccio, topografia del substrato roccioso, temperatura ed accumulo dello stesso; infine un ultimo metodo è rappresentato dal confronto e dalla sincronizzazione con altri paleorecords, come descritto prima per gli isotopi stabili dell'acqua. Queste metodologie hanno rappresentato un'importante risoluzione ai problemi connessi alla datazione degli strati, in quanto l'interpolazione (tramite un opportuno polinomio) di una serie non equispaziata $\{t_i, x_i\}_{i=1}^n$, al fine di ottenere una serie equispaziata $\{t(i)', x(i)'\}_{i=1}^{n'}$, può portare a vantaggi tanto quanto a svantaggi. I primi sono rappresentati sia dal fatto che operare con osservazioni temporali della prima tipologia richiede metodi statistici più sofisticati rispetto alle serie regolari, sia nel caso ci si ritrovi a lavorare con una numerosità del nuovo campione (n') inferiore a quella originaria (n), in quanto non vi sarebbero problemi di *dipendenza seriale* 'aggiuntiva'. Per i secondi invece, una motivazione risiede nel fatto che se si ha la situazione contraria, ovvero che la nuova numerosità campionaria n' è maggiore della precedente n , si riscontra un forte aumento di dipendenza seriale, rispetto alla normale; un'altra invece è data dalla scelta della forma del polinomio, in quanto la nuova serie x'_i può mostrare importanti variazioni rispetto alla serie originale x_i in termini di variabilità o di proprietà della componente stocastica.

5.2 Analisi spettrale per serie non equispaziate

La stima dello spettro di una serie storica paleoclimatica è spesso resa problematica in quanto le osservazioni non spaziate regolarmente nel tempo, devono essere opportunamente considerate. Nel caso equispaziato, il calcolo dello stesso per una serie temporale X_t è basato sull'assunzione di *stazionarietà* del processo stocastico, il quale ha le seguenti proprietà statistiche:

$$E[X_t] = \text{costante} \quad ,$$

$$E[X_t^2] < \infty \quad ,$$

$$E[X_t X_{t-j}] = \rho_j \quad .$$

Per passare dunque da un dominio temporale t a quello delle frequenze f , viene fatto uso della *trasformata discreta di Fourier*, la quale è così espressa:

$$F(f_t) = (\sqrt{2}/n) \sum_{t=1}^n x_t \exp(2\pi i f_t t) \quad (5.1)$$

Il collegamento con le funzioni periodiche seno-coseno, si può ottenere usando la relazione di Eulero ($\exp(ix) = \cos(x) + i \text{sen}(x)$), con la quale si può scrivere la 5.1, in una forma del tutto equivalente:

$$F(f_t) = (\sqrt{2}/n) \sum_{t=1}^n x_t (\cos(2\pi f_t t) + i \text{sen}(2\pi f_t t)) \quad (5.2)$$

Con le quali si possono stimare lo spettro, che in definitiva ha la seguente formulazione:

$$I(f_t) = |F(f_t)|^2 \quad (5.3)$$

Tuttavia tale stima non può valere per serie non equispaziate, in quanto sovente si ritrova un processo stocastico non stazionario, che non presenta le caratteristiche elencate precedentemente. Tale problematica è affrontata proprio nel successivo sottoparagrafo, dove se ne elenca una procedura di stima.

5.2.1 Stima dello spettro per serie storiche non equispaziate

Nel lavoro dal titolo “*Estimation of periods from unequally spaced observations*” (Ferraz-Mello, 1981) viene suggerita una procedura di stima per tali serie, nella quale viene introdotta una funzione di pesi $w_i = w(t_i)$, in maniera tale da tener conto della spaziatura non regolare tra le osservazioni rilevate nell'arco temporale di riferimento.

Si consideri un processo periodico $X_t = g_t + u_t$, nel quale g_t è una funzione periodica mentre u_t viene assunto essere un processo Gaussiano. Siano inoltre $x(t_1), x(t_2), \dots, x(t_n)$ le osservazioni campionate da tale processo periodico, alle quali viene loro addizionata una costante, in modo tale da soddisfare la condizione:

$$\sum_{i=1}^n \tilde{x}(t_i) = 0$$

Data la periodicità del processo X_t , la funzione periodica g_t comunemente scelta nella regressione ai minimi quadrati è quella sinusoidale; a tal fine dunque si considerano le seguenti funzioni per una data frequenza f : $K_0(t) = 1$, $K_1(t) = \cos(2\pi f t)$, $K_2(t) = \sin(2\pi f t)$; e sfruttando la proprietà di ortogonalità di queste ultime, esse vengono successivamente ortonormalizzate secondo le classiche formule di Gram-Schmidt ed assumono le seguenti espressioni:

$$k_0 = a_0 K_0(t) ;$$

$$k_1 = a_1 K_1(t) - a_1 k_0 [k_0, K_1(t)] ;$$

$$k_2 = a_2 K_2(t) - a_2 k_0 [k_0, K_2(t)] - a_2 k_1 [k_1, K_2(t)]$$

Dove $[,]$ indica il prodotto incrociato tra le funzioni ed i relativi pesi per ogni osservazione i , cioè:

$$\sum_{i=1}^n w_i g_1(t_i) g_2(t_i) \tag{5.4}$$

Ed i coefficienti a_0, a_1 e a_2 sono determinati in tale maniera (a pagina seguente)

$$a_0 = 1/\sqrt{\left(\sum_{i=1}^n w_i\right)} \quad ;$$

$$a_1 = 1/\sqrt{\left([\mathcal{K}_1(t), \mathcal{K}_1(t)] - a_0^2[\mathcal{K}_1(t)]^2\right)} \quad ;$$

$$a_2 = 1/\sqrt{\left([\mathcal{K}_2(t), \mathcal{K}_2(t)] - a_0^2[\mathcal{K}_2(t)]^2 - a_1^2[\mathcal{K}_1(t), \mathcal{K}_2(t)]^2\right)}$$

E nello specifico, utilizzando la 5.4 e ponendo $v_i = 2\pi f t_i$ si ha che:

$$[\mathcal{K}_1(t), \mathcal{K}_1(t)] = \sum_{i=1}^n w_i \cos^2(v_i) \quad ;$$

$$[\mathcal{K}_1(t)]^2 = \left(\sum_{i=1}^n w_i \cos(v_i)\right)^2 \quad ;$$

$$[\mathcal{K}_2(t), \mathcal{K}_2(t)] = \sum_{i=1}^n w_i \sin^2(v_i) \quad ;$$

$$[\mathcal{K}_2(t)]^2 = \left(\sum_{i=1}^n w_i \sin(v_i)\right)^2 \quad .$$

A questo punto si possono combinare le funzioni ortonormalizzate k_0, k_1 e k_2 con le osservazioni $\tilde{x}(t_i)$ ed ottenere i coefficienti per la regressione c_0, c_1 e c_2 per una determinata frequenza f :

$$c_0 = [\tilde{x}(t_i), k_0] = 0 \quad ;$$

$$c_1 = [\tilde{x}(t_i), k_1] = a_1 [\tilde{x}(t_i), \mathcal{K}_1(t)] \quad ;$$

$$c_2 = [\tilde{x}(t_i), k_2] = a_2 [\tilde{x}(t_i), \mathcal{K}_2(t)] - a_1 a_2 c_1 \{[\mathcal{K}_1(t), \mathcal{K}_2(t)] - a_0^2[\mathcal{K}_1(t), \mathcal{K}_2(t)]\}$$

Sfruttando sempre la 5.4 e considerando si ha che :

$$[\tilde{x}(t_i), \mathcal{K}_1(t)] = \sum_{i=1}^n w_i \tilde{x}(t_i) \cos(v_i) \quad ;$$

$$[\tilde{x}(t_i), \mathcal{K}_2(t)] = \sum_{i=1}^n w_i \tilde{x}(t_i) \sin(v_i) \quad ;$$

$$[\mathcal{K}_1(t), \mathcal{K}_2(t)] = \sum_{i=1}^n w_i \cos(v_i) \sin(v_i) \quad .$$

Riprendendo dunque la 5.2, considerando al posto di n , $\sum_{i=1}^n w_i$, e sostituendo a $x(t_i)$, $\tilde{x}(t_i)$ si può affermare che lo spettro per il processo periodico X_t (per una determinata frequenza f) è uguale alla somma dei due coefficienti non nulli della regressione ai minimi quadrati, ossia c_1 e c_2 . Tutto ciò, può essere formalizzato secondo la seguente espressione:

$$I(f) = |F(f)| = c_1^2(f) + c_2^2(f)$$

Con la quale si può stimare lo spettro di potenza per serie storiche non equispaziate per tutte le l frequenze di interesse.

5.2.2 Stima del filtro armonico

Nel lavoro di *Ferraz-Mello*, oltre ad un suggerimento per la stima dello spettro per le serie storiche non equispaziate, si è tenuto conto anche della possibilità che un processo periodico con tali caratteristiche, possa presentare oltre alle usuali onde sinusoidali legate alla principale frequenza f , altre periodicità riconducibili a frequenze che possono risultare o come *alias* di quella principale o comunque non propriamente distinguibili a livello grafico. A tal proposito, viene proposto uno strumento per depurare la serie da quelle periodicità che si ritiene possano non risultare d'interesse, eliminando le onde sinusoidali generate dalla frequenza associata alla stessa. Tale strumento è nominato **filtro armonico** ed in seguito se ne descriveranno le peculiarità e la procedura di formazione.

Si consideri una determinata frequenza \bar{f} , associata alla periodicità avente le caratteristiche elencate in precedenza, allora il relativo filtro armonico ha la seguente espressione:

$$d(\bar{f}, t) = c_1(\bar{f})k_1(\bar{f}, t) + c_2(\bar{f})k_2(\bar{f}, t)$$

La quale può essere sostituita da una del tutto equivalente, ma più utile per la stima, in quanto associabile ad una regressione lineare:

$$d(\bar{f}, t) = d_0 + d_1 \cos(\bar{v}_i) + d_2 \text{sen}(\bar{v}_i) \quad (5.5)$$

Dove i coefficienti d_0, d_1 e d_2 vengono così formulati, anche qui si pone $\bar{v}_i = 2\pi \bar{f} t_i$:

$$d_0 = -a_0^2 \left(d_1 \sum_{i=1}^n w_i \cos(\bar{v}_i) + d_2 \sum_{i=1}^n w_i \sin(\bar{v}_i) \right) ;$$

$$d_1 = a_1 c_1 + d_2 a_1^2 \left[a_0^2 \left(\sum_{i=1}^n w_i \cos(\bar{v}_i) \right) \left(\sum_{i=1}^n w_i \sin(\bar{v}_i) \right) - \sum_{i=1}^n w_i \cos(\bar{v}_i) \sin(\bar{v}_i) \right] ;$$

$$d_2 = a_2 c_2$$

Dove a_0, a_1, a_2 e c_1, c_2 sono i coefficienti usati nella stima dello spettro per serie storiche non equispaziate.

La serie filtrata dalla periodicità riconducibile alla frequenza \bar{f} risulterà dunque:

$$x(\hat{t}_i) = x(\tilde{t}_i) - d(\bar{v}_i) \quad (5.6)$$

E nel caso si voglia eliminare j periodicità, associate alle relative frequenze, si può scrivere:

$$x(\hat{t}_i) = x(\tilde{t}_i) - \sum_{l=1}^j d_l(\bar{v}_i) \quad (5.7)$$

Dove $d_l(\bar{v}_i)$ sono gli j filtri armonici.

5.3 Modelli AR(1) per serie storiche non equispaziate

Nell'analisi delle serie storiche climatiche e più precisamente nell'esaminare la componente stocastica della stessa (denotata con N_i per differenziarla dal precedente caso equispaziato), l'interesse è spesso rivolto al valore della funzione di autocovarianza

tra osservazioni di due istanti temporali, dove l'uno è immediatamente precedente all'altro ed è fornita dal calcolo della seguente espressione:

$$E[N_i N_{i-1}] \quad i=2, \dots, n \quad (5.8)$$

La motivazione della concentrazione degli studi sul valore della 5.8, risiede nel fatto che la parte stocastica di diverse variabili climatiche, esibisce degli scatterplots al ritardo 1 (interpolazione lineare tra i valori di N_{i-1} e N_i) dove i punti sono pressoché allineati lungo la retta di regressione di N_{i-1} su N_i , come dimostra la figura successiva:

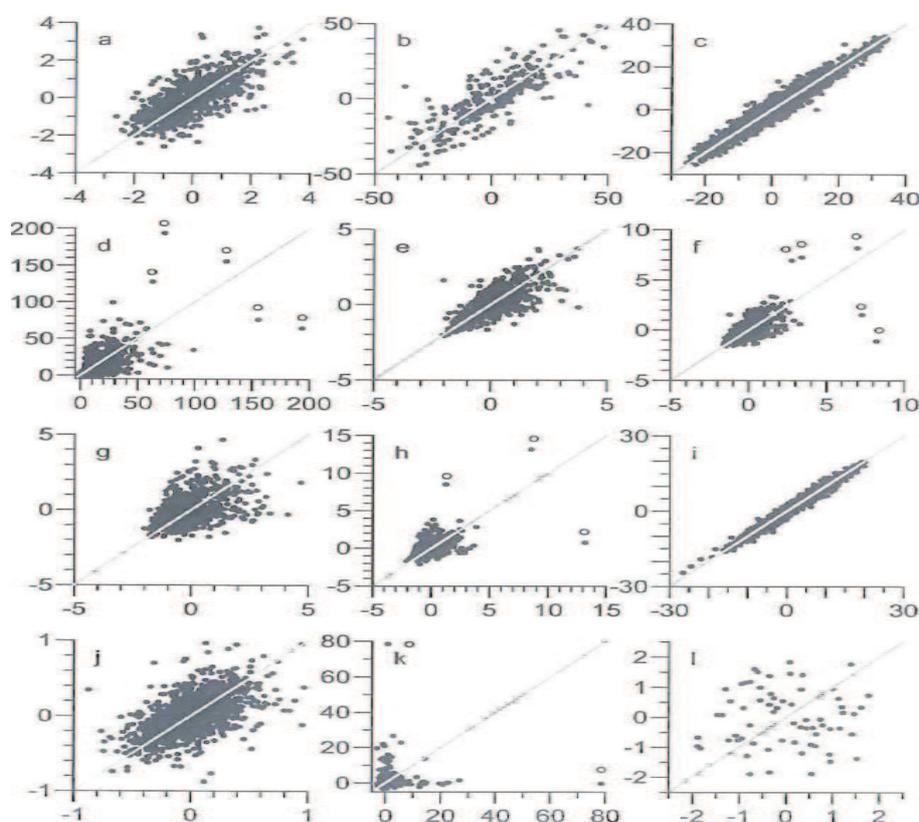


Figura 5.1: Scatterplots al ritardo 1 delle parti stocastiche di proxy climatiche (Mudelsee 2010)

Nella quale il grafico *a* rappresenta lo scatterplot per il livello dell'isotopo 18 dell'ossigeno misurato nell'Oceano Pacifico nei gusci calcarei dei foraminiferi betontici, *b* e *c* sono legati ai record del livello di CO_2 e δD rilevati nella carota di ghiaccio di Vostok; dal *d* all'*h* vi sono quelli relativi agli elementi acquisiti nella perforazione del ghiaccio in Groenlandia, ossia rispettivamente misure dei livelli di solfato, calcio, polveri atmosferiche, conduttività elettrica e sodio; il grafico *i* è invece associato al

livello del radiocarbonio presente negli anelli degli alberi; j è un'altra rilevazione per l'isotopo 18 dell'ossigeno presente nelle stalagmiti di Q5 nel sud dell'Oman; k è legato allo spessore delle varve rilevato nei sedimenti del Lower Mystic Lake presso Boston; infine l è collegato al livello di deflusso delle acque del fiume Artico.

Come menzionato nel paragrafo precedente, le serie storiche a disposizione nel contesto di uno studio paleoclimatico, però, non necessariamente presentano una regolare spaziatura tra l'istante temporale associato ad un'osservazione ed il suo immediato successivo. Dunque, tenendo in considerazione questa peculiarità, la forma di *un modello AR (1) per la parte stocastica di serie storiche non equispaziate* è fornita dalla seguente espressione:

$$N_i = \exp\{-[t_i - (t_i - 1)]/\tau\} * N_{i-1} + \varepsilon_i \quad (5.9)$$

Dalla quale si evince che, rispetto alla situazione di equispaziatura tra istanti temporali, la “perdita di memoria” del processo stesso è legata anche all'ampiezza della distanza temporale (denotata anche con d_i) che intercorre tra due osservazioni immediatamente successive; in particolare si può notare che un aumento di quest'ultima, causa un aumento della prima. Gli errori ε_i sono sempre provenienti da una distribuzione normale, la quale presenta una media nulla, però contrariamente al caso di spaziatura regolare, la varianza della stessa non è costante per ogni t ; essa invece è uguale a $1 - \exp\{-2[t - (t - 1)]/\tau\}$, facendo risultare così gli errori stessi, *eteroschedastici*.

La funzione di autocorrelazione del processo, viene ricavata dividendo la funzione di autocovarianza per la varianza prima menzionata, ottenendo quindi:

$$CORR[N_i N_{i-1}] = \rho_k = \begin{cases} 1 & \text{per } k=0 \\ \exp\{-[(t+k) - t]/\tau\} & \text{per } k > 0 \text{ e } < 0 \end{cases}$$

L'espressione della funzione di autocorrelazione conferma, di conseguenza, quanto è stato precedentemente menzionato, ossia l'aumento della perdita di memoria del processo, in presenza di osservazioni molto distanti tra loro a livello cronologico, anzi questo effetto viene amplificato, da k , che indica il ritardo al quale esse vengono calcolate.

Nella formulazione 5.9, il coefficiente che quantifica la dipendenza delle osservazioni

all'istante t da quelle all'istante $t-1$ viene denotato con τ ed è molto impiegato in campo climatologico in quanto esso esprime la dipendenza, nell'unità di misura temporale con la quale sono stati ordinati i dati (spesso in migliaia di anni); per questo motivo a τ viene dato l'appellativo di *persistenza temporale*.

La 5.9, in realtà, non è l'unica espressione possibile di un modello AR(1) per serie temporali non equispaziate. τ infatti è legato al coefficiente a tramite la seguente relazione biunivoca:

$$a = \exp(-1/\tau) \quad (5.10)$$

La quale, ci permette dunque di scrivere il modello per un processo autoregressivo del primo ordine (in termini di a), con un'espressione del tutto equivalente alla 5.9:

$$N_i = a^{[t_i - (t_i - 1)]} N_{i-1} + \varepsilon_i \quad (5.11)$$

La stima di a è più complessa rispetto alla condizione di equispaziatura tra le osservazioni delle serie temporali e la motivazione è l'inserimento del parametro stesso in una relazione *non lineare*, come risulta quella precedente. In generale, il metodo con il quale essa viene ottenuta, è però corrispondente a quello utilizzato in condizioni di regolarità della spaziatura tra osservazioni ordinate cronologicamente, ed è il *metodo dei minimi quadrati*. L'obiettivo della stima del parametro d'interesse, converge dunque con quello di rendere minima la seguente quantità:

$$S(\tilde{a}) = \sum_{i=2}^n [N_i - \tilde{a}^{[t_i - (t_i - 1)]} N_{i-1}]^2 \quad (5.12)$$

o equivalentemente, in termini di persistenza temporale o tempo di decadimento τ :

$$S(\tilde{\tau}) = \sum_{i=2}^n [N_i - \exp\{-[t_i - (t_i - 1)]/\tilde{\tau}\} N_{i-1}]^2 \quad (5.13)$$

Ed il valore del parametro (sia esso a o τ) che minimizza tali quantità viene dunque considerato lo stimatore per lo stesso, e ha come formulazione $\hat{a} = \operatorname{argmin}[S(\tilde{a})]$ o $\hat{\tau} = \operatorname{argmin}[S(\tilde{\tau})]$. Data la non linearità della relazione nella quale a e τ sono

coinvolti, la ricerca del valore, che rende minima la 5.12 e la 5.13, viene svolta con metodi numerici per la risoluzione delle equazioni non lineari e nello specifico si fa largo impiego del *metodo di Brent*, il quale a sua volta funge da catalizzatore di altri ossia *il metodo della bisezione, della secante e della parabola inversa*. In situazione di serie temporali equispaziate ($t_i - (t_i - 1) = d$ per ogni i) lo stimatore ai minimi quadrati per a (sfruttando la 5.10) corrisponde a $\hat{a} = \exp(-d/\hat{\tau})$; inoltre si può altresì derivare lo stimatore per l'usuale coefficiente di autocorrelazione del modello AR(1), ossia ϕ_1 , poiché a è legato ad esso dalla seguente relazione:

$$a = \phi_1^{1/d} \quad (5.15)$$

In virtù della quale, si può affermare che in caso di spaziatura regolare, esso risulta:

$$\hat{\phi}_1 = \left(\sum_{i=2}^n N_i N_{i-1} \right) / \left(\sum_{i=2}^n N_{i-1} \right) \quad (5.16)$$

Kendall in un lavoro del 1954, ha dimostrato che lo stimatore per a così trovato, risulta però distorto, ovvero che, il valore atteso dello stesso non è equivalente al valore vero del parametro, formalmente accade che $E[\hat{a}] \neq a$. Egli inoltre quantificò il valore della distorsione (o bias), ovvero sia la differenza tra il valore atteso degli stimatori per a ed il loro valore vero, ricavandone per ambedue la medesima espressione:

$$E[\hat{\theta}] - \theta = \theta - (1 + 3\theta)/(n - 1) \quad (5.17)$$

Dove al posto di θ e $\hat{\theta}$, vi si può inserire a e \hat{a} , arrivando così alla stima del bias per ciascuno dei due stimatori. Considerando la 5.17, è possibile anche giungere al calcolo dello stimatore non distorto per θ , infatti sostituendo al valor vero del parametro stesso, $\hat{\theta}$ (che rappresenta lo stimatore corretto), ed al valore atteso $E[\hat{\theta}]$ lo stimatore distorto $\hat{\theta}$, si ottiene la formulazione:

$$\hat{\theta} = (1 + (n - 1)\hat{\theta})/(n - 4) \quad (5.18)$$

E logicamente l'inserimento di \hat{a} e \hat{a} al posto rispettivamente di $\hat{\theta}$ e $\hat{\theta}$, permette di

avere lo stimatore corretto per il parametro a .

Nel 1966 Stuart e lo stesso Kendall riuscirono a trovare, attraverso esperimenti numerici effettuati con serie storiche simulate, un risultato ancor più importante, ovvero la dimostrazione che il valore della distorsione per gli stimatori del parametro a trovato nel caso di equispaziatura delle osservazioni di una serie temporale, è molto simile a quello ottenuto per serie storiche irregolarmente distanziate. Ciò, di conseguenza consente dunque di replicare quanto trovato in precedenza e, sfruttando la 5.17, vengono inseriti rispettivamente al posto di $\hat{\theta}$ e $\hat{\theta}$, lo stimatore corretto per serie temporali non equispaziate $\hat{\theta}$ e quello distorto $\hat{\theta}$ ottenendo:

$$\hat{\theta} = (1 + (n-1)\hat{\theta}) / (n-4) \quad (5.19)$$

La sostituzione di \hat{a} e \hat{a} al posto di $\hat{\theta}$ e $\hat{\theta}$, conduce allo stimatore corretto per \bar{a} . La relazione 5.15, ossia quella che lega a all'usuale coefficiente di autocorrelazione ϕ_1 , in seguito al risultato appena menzionato, è utilizzabile anche in presenza di serie temporali con osservazioni distanziate tra di loro irregolarmente e la differenza con la 5.15, si riscontra nel presentare al posto di d , la media delle distanze temporali tra le osservazioni (\bar{d}), poiché, esso non è palesemente univoco, contrariamente al caso di equispaziatura; formalmente, l'espressione che racchiude tutto ciò è la seguente:

$$\hat{a} = \hat{\phi}_1^{1/\bar{d}} \quad (5.20)$$

o equivalentemente, in termini di $\hat{\phi}_1^{1/\bar{d}}$:

$$\hat{\phi}_1 = \hat{a}^{\bar{d}} \quad (5.21)$$

Dove \hat{a} e $\hat{\phi}_1$ sono rispettivamente gli stimatori per i parametri di autocorrelazione dei modelli AR(1) per le serie con osservazioni non distanziate regolarmente nel tempo.

5.4 Correlazione tra serie storiche non equispaziate

In paleoclimatologia, lo studio delle relazioni esistenti tra due variabili misurate da due proxy differenti ed indipendenti tra loro, è sovente reso complicato per la manifestazione di due problematiche: una prima è la precedentemente menzionata non equispaziatura delle distanze temporali di una singola serie; mentre la seconda riguarda la non contemporaneità tra osservazioni di due serie storiche appartenenti a due proxy differenti, formalmente, se si disponesse di una serie di valori per una proxy X e per un'altra Y , si verifica spesso che $T_X(i) \neq T_Y(j)$ per ogni $i = j$, ovvero si è in presenza di misurazioni effettuate con *scale temporali diseguali*. La sfida imposta da queste ultime, risiede nel fatto che la formula del coefficiente di correlazione di Pearson non può essere direttamente applicata per questa tipologia di serie temporali, di conseguenza non può stimare il grado della relazione fra gli andamenti di entrambe le variabili. Ciò che invece viene usualmente svolto è un riadattamento del metodo al caso delle scale temporali diseguali, interpolando le due serie storiche in una 'griglia temporale' comune ad entrambe e successivamente si calcola l'usuale stimatore per la correlazione, cioè il coefficiente di Pearson, denotato in maniera diversa dal precedente, con r_{xy} . I pericoli derivanti dall'adottare tale metodologia sono fondamentalmente due: un primo discende dalla libertà di scelta della griglia di tempo comune alle due serie, la quale si traduce automaticamente nell'arbitrarietà di formazione della numerosità del campione 'interpolato', con la conseguenza di incorrere nel rischio di determinare in maniera distorta il grado di relazione tra le stesse; un secondo proviene dalla dipendenza seriale causata dall'operazione di interpolazione, la quale deve essere presa in considerazione correttamente. In un lavoro del 1968, Hocking e Smith svilupparono questa metodologia, focalizzandosi sui *valori mancanti* (osservazioni di X che non sono presenti in Y e viceversa), i quali costituiscono un caso particolare nelle scale temporali diseguali, utilizzandoli però per fare inferenza sulle medie e deviazioni standard delle distribuzioni marginali delle due variabili d'interesse; le osservazioni comuni invece consentono la stima della covarianza. Quest'ultima ricerca è però di rilevanza limitata nelle applicazioni paleoclimatiche, sostanzialmente per almeno tre ragioni. Una prima risiede squisitamente nelle caratteristiche delle scale temporali di questi dati, in quanto spesso capita di avere una serie di una proxy X che non ha assolutamente alcun istante temporale in comune con la serie di una proxy Y , quindi non si può applicare alcuno

degli stimatori conosciuti; una seconda invece, emerge dalla possibile non stazionarietà del primo o del secondo momento delle serie storiche delle variabili esaminate, la quale può causare distorsione nella stima, in particolare l'eteroschedasticità può sottostimare le deviazioni standard e quindi far risultare il coefficiente di correlazione maggiore di 1. Infine una terza sta nelle assunzioni fatte sulla distribuzione congiunta dei valori di tali variabili, la quale è ben lontana da una normale bivariata e non vi è assenza di dipendenza seriale. Tutto ciò fa allontanare dall'utilizzo di tali procedure di stima, in quanto poco adatte alla realtà; diviene dunque necessario valutare altre vie per la stima della correlazione tra proxy paleoclimatiche. Mudelsee, nel libro *Climate Time Series Analysis*, suggerisce un nuovo adattamento del metodo dell'interpolazione per la stima della correlazione tra dati di proxy di variabili climatiche, il quale viene chiamato **Binned Correlation**. Tale metodo, contrariamente a quelli nominati in precedenza, si basa sul fatto che la *persistenza temporale* τ , che rappresenta verosimilmente una peculiarità comune a diverse proxy climatiche (si veda in merito il paragrafo antecedente) può consentire di recuperare 'l'informazione mista' proveniente dalle due variabili, nonostante le due scale temporali differiscano. La descrizione dettagliata della metodologia è rinviata al successivo sottoparagrafo.

5.4.1 Binned correlation

Si considerino due differenti serie temporali di proxy climatiche, aventi rispettivamente numerosità campionaria n_X e n_Y , con i relativi dati che seguono i processi stocastici denotati con le seguenti espressioni, una prima è data per la proxy X :

$$\{T_X(i), X(i)\}_{i=1}^{n_X}$$

ed una seconda per la proxy Y :

$$\{T_Y(j), Y(j)\}_{j=1}^{n_Y}$$

Su ciascuna delle serie storiche viene inizialmente stimata la persistenza temporale, la quale viene indicata rispettivamente con τ_X e τ_Y . Dopo quest'operazione, comincia la procedura che ha, come primo passo, l'individuazione degli estremi della nuova 'griglia del tempo' comune (denotato con $\bar{T}(k)$), o meglio la ricerca del minimo istante temporale tra entrambe (in notazione $\bar{T}_{min} = \min\{T_X(1), T_Y(1)\}$) ed il massimo,

sempre tra le stesse ($\bar{T}_{max} = \max\{T_X(n_X), T_Y(n_Y)\}$). Il secondo step è la selezione del numero di categorie temporali (*bins*), le quali presentano tutte lunghezza costante pari a $\bar{\tau}$, che viene scelto in base a tre semplici regole: una prima è data dalla somma delle persistenze temporali delle due serie ($\bar{\tau} = \tau_X + \tau_Y$); una seconda invece dalla ricerca del massimo tra le persistenze delle stesse ($\bar{\tau} = \max\{\tau_X, \tau_Y\}$); infine una terza (la quale è notevolmente migliore in termini di RMSE di $\hat{\rho}_{XY}$, in base agli esperimenti effettuati con metodi Monte Carlo) che si basa sulle *distanze temporali medie* e sui *coefficienti di autocorrelazione non distorti per serie storiche non equispaziate*, i quali portano al calcolo di $\bar{\tau}$ attraverso i seguenti passaggi, in successione:

- Calcolo di $\bar{d}_X = [T_X(n_X) - T_X(1)] / (n_X - 1)$, $\bar{d}_Y = [T_Y(n_Y) - T_Y(1)] / (n_Y - 1)$ e della distanza temporale media della nuova 'griglia temporale' comune, ovvero $\bar{d}_{XY} = [\bar{T}_{max} - \bar{T}_{min}] / (n_X + n_Y - 1)$;
- Calcolo dei coefficienti di autocorrelazione non distorti per ambedue le serie storiche (\hat{a}_X e \hat{a}_Y) e successivamente del coefficiente di autocorrelazione non distorto 'comune', tramite la formulazione $\hat{a}_{XY} = \sqrt{\hat{a}_X * \hat{a}_Y}$;
- Calcolo di $\bar{\tau}$, il quale sfruttando la relazione 5.10, ha espressione $\bar{\tau} = -\bar{d}_{XY} / \ln(\hat{a}_{XY})$.

Una volta concluso il calcolo di $\bar{\tau}$, il terzo passo della procedura prevede la valutazione di ogni singola 'categoria di tempo' (*time bin*), ovvero l'analisi di se e quante osservazioni di entrambe le proxy ricadono all'interno di ciascuna delle stesse (nella figura 5.2, si trova nella prima categoria temporale, $[\bar{T}_{min}; \bar{T}_{min} + \bar{\tau}]$, tre osservazioni per la proxy X contro una per la proxy Y); se questa diagnosi è positiva (vi sono osservazioni sia di X che di Y all'interno di una stessa 'categoria temporale') allora si forma la media delle osservazioni all'interno della stessa 'classe temporale' per X e per Y , denotandole rispettivamente con $\bar{X}(k)$ e $\bar{Y}(k)$, assegnando come istante temporale il valore centrale (medio) dello stesso nella categoria, ed ha la seguente successione:

$\bar{T}(1) = (\bar{T}_{min} + \bar{T}_{min} + \bar{\tau}) / 2 = \bar{T}_{min} + (1/2) * \bar{\tau}$ come istante iniziale, gli altri sono ottenuti in modo ricorsivo, addizionati ognuno con $\bar{\tau}$, ossia $\bar{T}(2) = \bar{T}(1) + \bar{\tau} = \bar{T}_{min} + (3/2) * \bar{\tau}$,

$\bar{T}(3) = \bar{T}(2) + \bar{\tau} = \bar{T}_{min} + (5/2) * \bar{\tau}, \dots, \bar{T}(k) = \bar{T}(k-1) + \bar{\tau} = \bar{T}_{min} + [(k+1)/2] * \bar{\tau}$, con $\bar{T}(0) \equiv \bar{T}_{min}$ e la serie storica risultante è dunque la seguente:

$$\{\bar{T}(k), \bar{X}(k), \bar{Y}(k)\}_{k=1}^{\bar{n}}$$

Il campione 'categorizzato' (binned) ha ora numerosità pari a \bar{n} . Si può dunque compiere l'ultimo passo della procedura, calcolando il coefficiente di correlazione di Pearson tra le due serie storiche 'binned': tale stima è denotata con \bar{r}_{XY} .

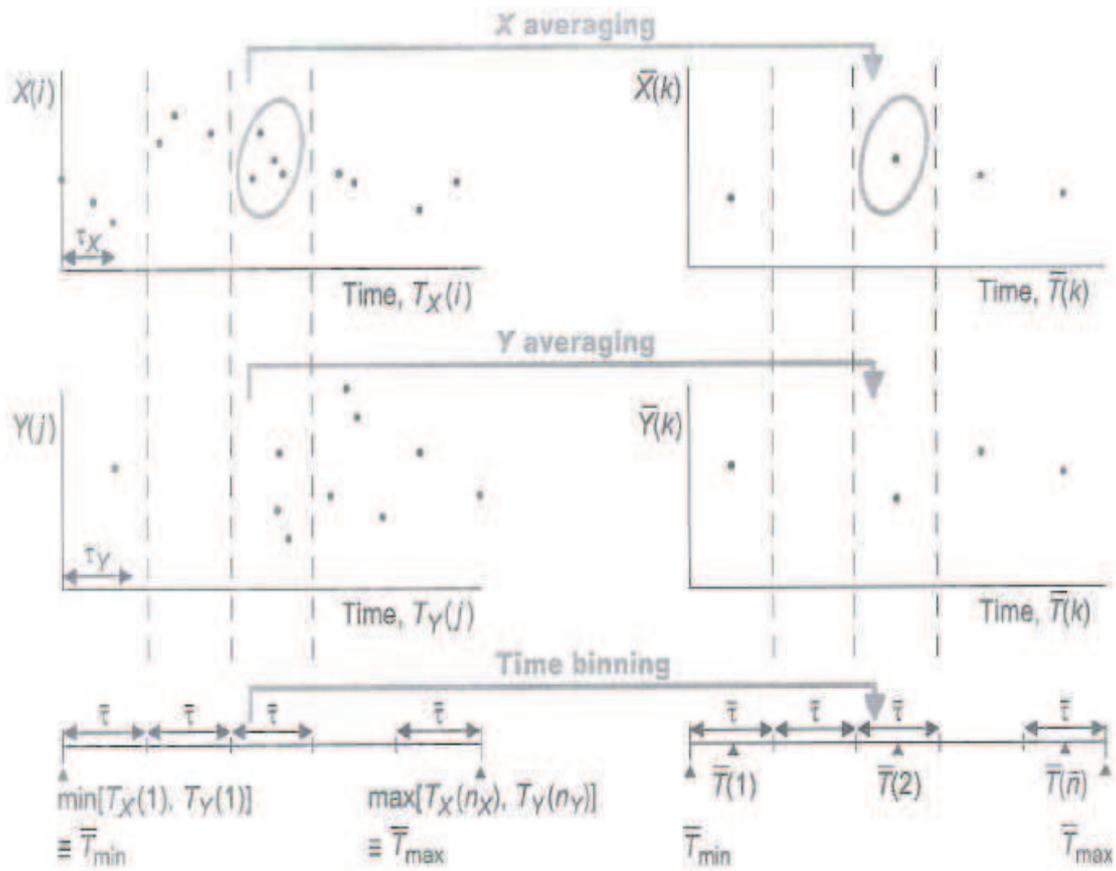


Figura 5.2: Binned correlation per due serie storiche con scale temporali diseguali

5.5 Applicazione al caso Vostok

5.5.1 L'insieme dei dati

Le due serie temporali di riferimento per questo lavoro sono state reperite a due indirizzi web differenti, in particolare le osservazioni sul livello di CO_2 sono stati scaricati al link <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/vostok/co2nat.txt>, mentre per il δD da <ftp://ftp.ncdc.noaa.gov/pub/data/paleo/icecore/antarctica/vostok/deutnat.txt>. Ambedue provengono precisamente dal *National Climatic Data Center* (una parte del più ampio *National Oceanic and Atmospheric Administration*), ovvero l'istituto che oltre a fornire servizi climatologici a tutti i settori dell'economia americana, possiede uno tra i più grandi archivi di dati climatici, i quali sono messi a disposizione per gli utenti di tutto il mondo. Prima di elencare le caratteristiche dei due dataset, è utile ricordare il luogo di provenienza di entrambi, ovvero la stazione di carotaggio e ricerca sul paleoclima di Vostok (già descritta nel dettaglio al 3.1), situata in Antartide, in una zona nella quale la calotta glaciale raggiunge circa i 3700 m di spessore. La seguente figura indica esattamente dove viene effettuata la perforazione:

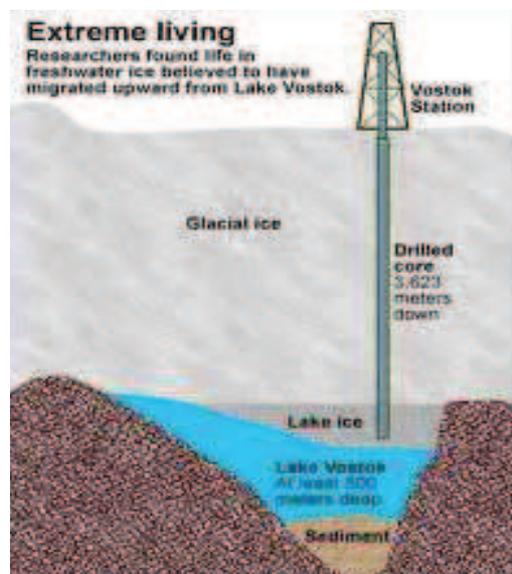


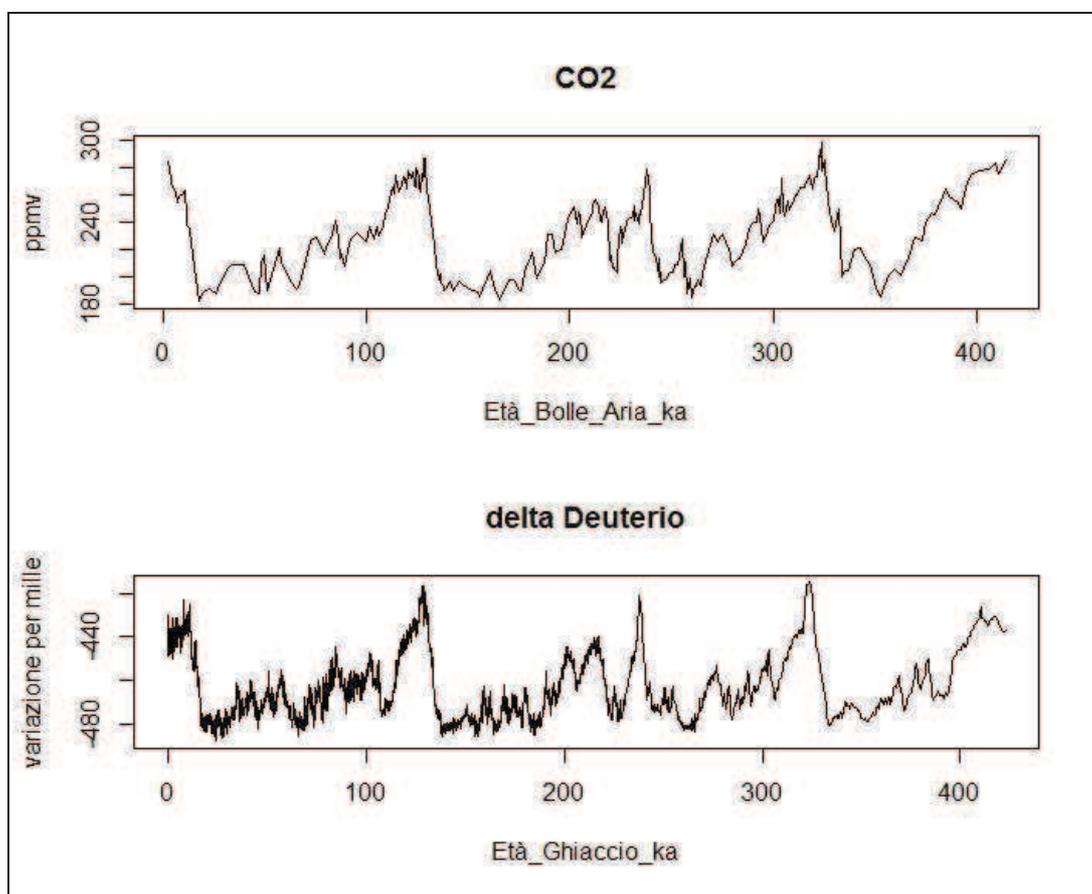
Figura 5.3: I differenti strati attraversati nella fase di carotaggio a Vostok

Lo studio sulla carota di ghiaccio, dalla quale poi sono state raccolte le informazioni su svariate proxy climatiche, è stato portato a termine nel Giugno 1999 (dopo altri effettuati in precedenza) da un progetto che ha coinvolto un gruppo di ricercatori (*Jean Robert Petit et al.*) provenienti da tre nazionalità differenti, nello specifico si tratta di Francia, Russia e Stati Uniti. I datasets a disposizione, contengono non soltanto il valore della proxy rilevato a varie profondità (la carota estratta è lunga *3310 m*), ma a ciascuno vi sono collegate anche le unità di misura temporali determinate con la cosiddetta *scala temporale GT4* e nel caso del δD sono presenti le profondità alle quali è stato rilevato il livello di tale proxy e la corrispondente variazione di temperatura in $^{\circ}C$. Le osservazioni sugli istanti temporali sono forniti in anni *Before Present*, ossia la distanza temporale di un evento direttamente da oggi.

5.5.2 Analisi preliminari

L'indagine preliminare sulle serie storiche non equispaziate del livello dell'anidride carbonica (CO_2) e della variazione del livello del Deuterio (δD), riguarda la presentazione in successione dei *grafici* e degli *istogrammi* di ciascuna di esse, con lo scopo di analizzare (a carattere puramente grafico) l'andamento nel periodo temporale di riferimento e la distribuzione dei valori di tali *proxy* climatiche. La presentazione dei grafici ed i relativi commenti vengono presentati in modo congiunto, esattamente come nel caso precedente:

Grafici delle proxy:

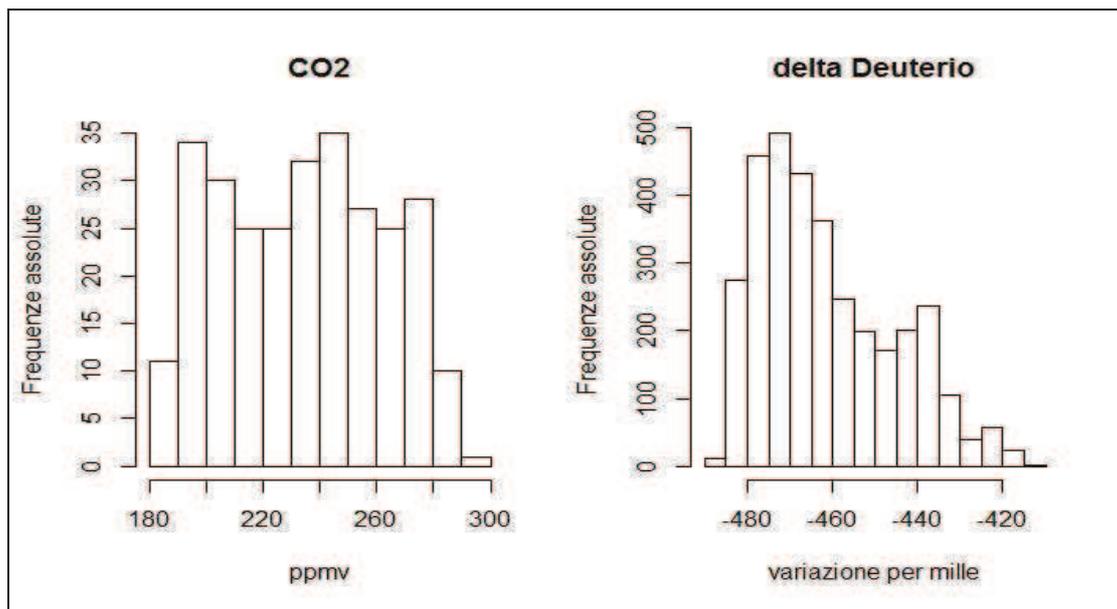


Prima di commentare i grafici appena ottenuti, si vuol evidenziare che i valori riportati in ascissa sono stati ottenuti dividendo per 1000 sia le informazioni sull'età delle bolle d'aria (comprese a varie profondità della carota di ghiaccio e contenenti il livello della CO_2) sia quelle sull'età del ghiaccio che le contiene (sul quale sono state direttamente

osservate le variazioni di Deuterio), in modo da ottenere l'indice temporale espresso in migliaia di anni (*ka*). Questi ultimi, come già detto in precedenza, sono anni *Before Present*, dunque le osservazioni meno recenti nel tempo si trovano, spostandosi verso destra nell'asse temporale.

La considerazione principale che emerge in modo molto chiaro dal grafico precedente, è la sostanziale non stazionarietà dei valori di entrambe le serie esaminate lungo l'asse temporale di riferimento e la caratteristica dominante che pare possano possedere è un andamento pressoché periodico. In particolare esso sembra risultare, prevalentemente, attorno all'ordine dei 100000 anni, che rappresenterebbe esattamente una delle tre principali peculiarità dell'orbita compiuta dal nostro pianeta attorno al Sole (descritte nella *teoria di Milanković* trattata nel 3.3), ossia *l'eccentricità*; chiaramente però una successiva analisi spettrale delle serie verificherà tale affermazione. Un'ulteriore considerazione è legata al fatto che la variazione del delta Deuterio è una proxy della variazione della temperatura, e quindi si può notare come siano presenti nel record ben 4 *cicli glaciali/interglaciali* attraversati dal pianeta. Infine si può affermare che le due serie abbiano un comportamento molto simile e l'obiettivo di questo lavoro sarà di quantificarne la relazione.

Istogrammi dei valori:

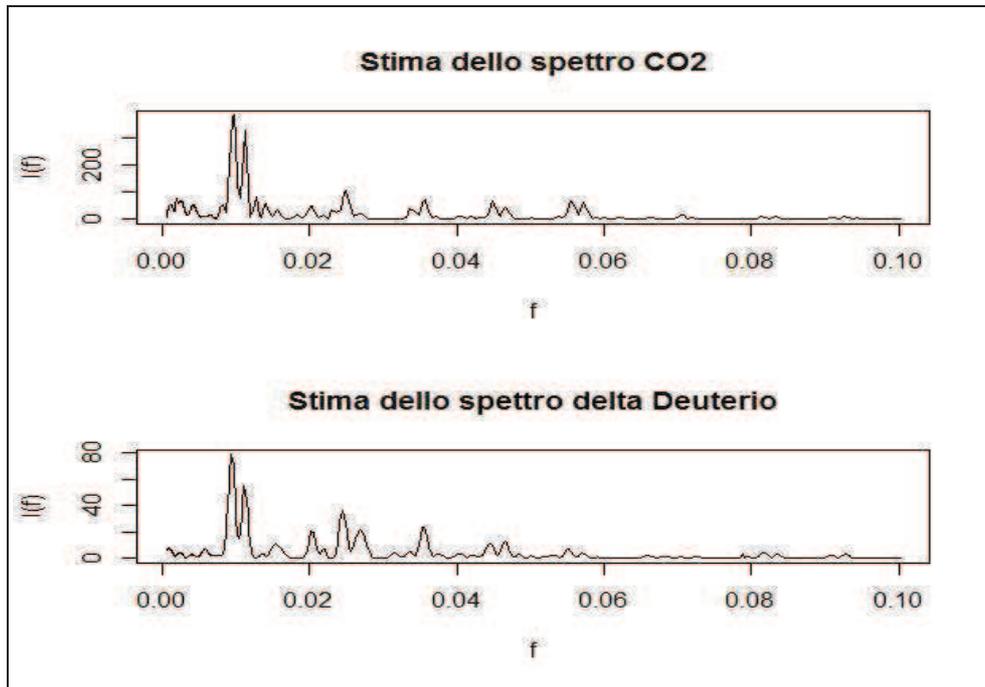


Ciò che si evince chiaramente dall'istogramma per le osservazioni di entrambe le serie, è la *non normalità della distribuzione* degli stessi e la presenza di ben 3 mode, particolarità, quest'ultima, molto più evidente per la distribuzione dell'anidride carbonica. Tale caratteristica inoltre, conferma (e rafforza) quanto era già emerso dai grafici precedenti, evidenziando da un lato l'alternanza periodica con la quale esse hanno raggiunto livelli massimi e minimi, e dall'altro che ciò non è avvenuto brevemente, ma con gradualità nel corso del tempo (in particolare nel pervenire a quelli più bassi). Una nota di differenza viene dall'*asimmetria verso sinistra* presentata dalla distribuzione della variazione del δD , però ciò non permette di esimerci dal concludere che la relazione tra le due proxy climatiche delle variabili in gioco, si dimostra ben evidente nell'arco degli oltre 400000 anni di informazioni arrivate dalla carota di ghiaccio di Vostok; le successive analisi oltre a ricercare un'eventuale connessione con la teoria sui cicli orbitali, mirano a quantificare tale relazione.

5.5.3 Analisi spettrale

Nel paragrafo precedente, la visione dei grafici dell'andamento di ambedue le serie storiche, aveva fatto emergere una certa periodicità delle stesse, attorno all'ordine del centinaio di migliaia di anni. Tale comportamento potrebbe essere ricondotto a quanto riportato nelle teorie sui cicli orbitali della Terra, sviluppate dallo scienziato serbo Milutin Milanković, delle quali sono stati enunciati i risultati fondamentali nel 3.3. Per poter verificare la bontà o meno di tale affermazione è necessario condurre un'analisi dello spettro delle due serie, il quale ovviamente deve essere stimato. Quest'ultima operazione dev'essere svolta tenendo in considerazione il fatto che le proxy, sulle quali si lavora, non hanno tra di loro la medesima distanza temporale, perciò la stima dello spettro è stata eseguita seguendo interamente quanto enunciato nel 5.2.1, nel quale, prima di effettuarla, è stata tolta la **media** delle osservazioni di ambedue le serie. I risultati ed i commenti di quanto trovato sono, anche in questo caso, esposti in maniera congiunta.

Spettri:



Nell'introdurre il commento di ciò che è stato ottenuto, si tiene a precisare che i pesi w_i (si veda il 5.2.1) utilizzati nella stima dello spettro, sono stati ottenuti in funzione dell'istante temporale di riferimento e precisamente si è voluto seguire il criterio

$$w_i = t(i) / \sum_{i=1}^n t(i), \text{ dalla quale discende ovviamente che } \sum_{i=1}^n w_i = 1.$$

La scelta di utilizzare tale criterio è stata eseguita in maniera da considerare, nella stima dello spettro, la caratteristica principale di questo tipo di osservazioni, ossia l'aver una scala temporale non equispaziata.

La prima ed immediata osservazione, guardando lo spettro di entrambe le serie (senza le medie) è la presenza di due elevati picchi (in particolare il primo) attorno alla frequenza 0,01, assieme ad altri di modesta entità, tra i quali il più alto si manifesta attorno ad una frequenza compresa tra il 0,02 ed il 0,03 (in particolare per la variazione del Deuterio). Nello specifico, andando a ricercare le frequenze corrispondenti ai due picchi più alti, si può successivamente calcolare le periodicità corrispondenti, le quali sono espresse in migliaia di anni (*ka*). Tale operazione ha portato come periodicità principale (quella inerente al primo picco) a **103,1 ka** per la CO_2 e **106,4 ka** per il δD , le quali fanno riferimento alle frequenze **0,0097** (anidride carbonica) e **0,0094** (variazione del

Deuterio). Per quanto riguarda il secondo picco, che rappresenta un *alias* di quello principale, sono state trovate periodicità rispettivamente di **88,5 ka** per il biossido di carbonio ed **89,3 ka** per il delta Deuterio, corrispondenti alle frequenze di **0,0113** (CO_2) e **0,0112** (δD). Per il terzo picco, si sono trovate periodicità pari a **40,3 ka** per l'anidride carbonica e **40,8 ka** per il δD , le quali si riferiscono rispettivamente alle frequenze **0,0248** e **0,0245**. Si può quindi affermare l'influenza dell'eccentricità (100000 anni) e dell'obliquità sul comportamento di entrambe le proxy, e quanto trovato svelerebbe dunque diversi processi fisico-chimici riguardanti la storia climatica del nostro pianeta. Infatti, quando esso si trovava in una posizione tale da avere una maggior esposizione solare, ebbe inizio la **deglaciazione**; dagli oceani più caldi, dalle torbiere e dal permafrost si liberava progressivamente CO_2 , contribuendo ad aumentare l'effetto serra e con esso la temperatura, guidando la Terra verso un'epoca più calda (periodi interglaciali). Millenni più tardi il fenomeno si invertiva: diminuzione dell'insolazione dovuta ai cicli orbitali, graduale assorbimento dell'anidride carbonica da parte degli oceani e materiali organici indecomposti, calo termico e pianeta verso la cosiddetta **era glaciale**.

5.5.4 Stima del filtro armonico

L'analisi spettrale, effettuata in precedenza ha messo dunque in luce (e confermato dopo l'ispezione grafica) che ambedue le serie mostravano un picco importante associato ad una periodicità temporale sull'ordine dei 100000 anni circa. In realtà essa ha mostrato ben di più, per tre ragioni: una prima, sta nella presenza di un picco (di valore inferiore ai precedenti, ma comunque importante) che dà origine ad una periodicità molto prossima ai 41000 anni, la quale non era emersa chiaramente dalla semplice visione del grafico. Una seconda invece, risiede nella presenza di un secondo picco vicino a quello principale ed avente valore quasi simile allo stesso, il quale però non rappresenta la periodicità reale delle due serie in quanto ne darebbe vita ad una di circa 90000 anni; questo fenomeno è tipico di serie aventi osservazioni la cui distanza temporale tra le stesse è elevata e non regolare, esso viene chiamato *aliasing*. Una terza ed ultima risiede nella presenza di altre periodicità inferiori, limitrofe alla principale, che contribuiscono

a distorcere non solo il segnale della stessa ma anche la successiva stima della persistenza temporale. Il filtro armonico descritto nel 5.2.2, è la tecnica con la quale le serie (a cui è stata tolta la loro media, come fatto per la stima dello spettro) vengono depurate da tali periodicità e le serie residue di entrambe avranno tutte formulazione come la 5.7. Quindi oltre alle due periodicità che fungono da *alias* alla principale (0,0113 per il biossido di carbonio e 0,0112 per il δD) si sono tolte dalle serie altre periodicità e precisamente esse corrispondevano a 4 frequenze per la CO_2 e 3 per la variazione del Deuterio. Nello specifico si tratta delle frequenze **0,0012**, **0,0019**, **0,0025**, **0,0128** per l'anidride carbonica e **0,0058**, **0,0154**, **0,0202** per il δD .

5.5.5 Modello AR(1) per stima della persistenza temporale

L'operazione svolta nel paragrafo precedente è stata quella di filtrare le serie (private della loro media), ovvero togliere dalle stesse le frequenze che potevano distorcere il valore della reale periodicità. Avendo a disposizione tali serie, viene dunque stimato per entrambe il modello AR(1), il quale fornisce una stima iniziale del coefficiente di autocorrelazione per le serie non equispaziate temporalmente. Successivamente, sapendo che tale stimatore è distorto, lo si corregge per il fattore trovato da Kendall (si veda il 5.3) ed infine tramite la relazione che lega α con τ , si giunge a stimare la persistenza temporale. I risultati sono presentati in modo distinto per ciascuna proxy, ai fini di aumentare la chiarezza espositiva:

- **Serie storica del livello di anidride carbonica:**

La motivazione della scelta di un modello AR(1) scaturisce principalmente dalla visione degli scatterplots ai primi 4 ritardi (grafico 5.1, a pagina seguente), dai quali si denota che la dipendenza temporale, per quanto riguarda la serie filtrata del livello di CO_2 , si manifesta maggiormente al primo ritardo.

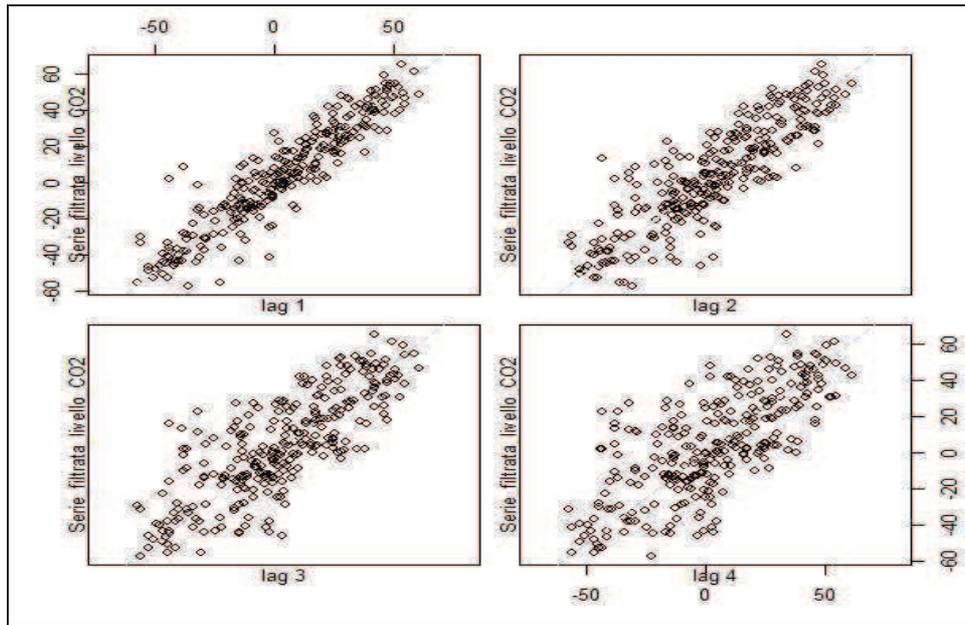


Grafico 5.1: Scatterplots ai primi 4 ritardi della serie filtrata del livello di biossido di carbonio

Il modello AR(1) per la serie filtrata (parte stocastica) del livello della CO_2 , assume la seguente espressione:

$$\hat{CO}_{2(i)} = a^{[t_i - t_{i-1}]} \hat{CO}_{2(i-1)} + \varepsilon_i$$

E per la stima di a , si va a recuperare la quantità 5.12, riadattandola al caso dell'anidride carbonica:

$$S(\tilde{a}) = \sum_{i=2}^n [\hat{CO}_{2(i)} - \tilde{a}^{[t_i - t_{i-1}]} \hat{CO}_{2(i-1)}]^2$$

Dove \tilde{a} viene ottenuto tramite la minimizzazione di tale quantità, con un algoritmo numerico (software R, comando *optim*) e nello specifico con *il metodo di Brent*, fornendo come valore iniziale $1/\exp(1)$. Il valore di \tilde{a} ottenuto con tale metodologia è pari a **0,9486**, il quale corretto per il fattore di distorsione (si veda la formula 5.18) si trova essere **0,9624**. Utilizzando infine la relazione inversa che lega a al valore che quantifica la persistenza usando l'unità di misura temporale, cioè τ , si ha che quest'ultimo per la serie riguardante la CO_2 , vale (dopo i successivi passaggi) $\tau = -\bar{d}/\ln(\hat{a})$, $\tau = -1.46/\ln(0,9624)$, $\tau = \mathbf{38,0951 ka}$.

• **Serie storica di variazione del Deuterio:**

La motivazione della scelta di un modello AR(1) scaturisce anche qui principalmente dalla visione degli scatterplots ai primi 4 ritardi (grafico 5.2), dai quali si denota che la dipendenza temporale, per quanto riguarda la serie filtrata di δD , si manifesta maggiormente al primo ritardo. Tuttavia, però vi è un elemento di discordanza rispetto alla proxy precedente, ossia che la correlazione sembra essere più elevata, in quanto ai 3 ritardi successivi, i valori rimangono ancora lungo la retta ideale di regressione.

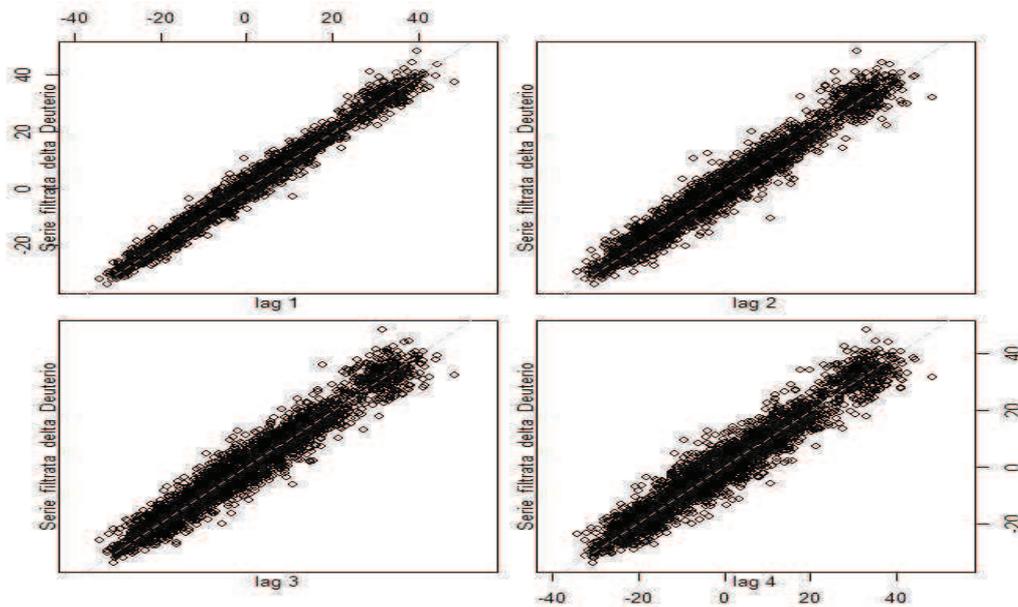


Grafico 5.2: Scatterplots ai primi 4 ritardi della serie filtrata della variazione del Deuterio

Il modello AR(1) per la serie filtrata (parte stocastica) del livello della variazione di Deuterio, assume la seguente espressione:

$$\delta \hat{D}_i = a^{[t_i - t_{i-1}]} \delta \hat{D}_{i-1} + \varepsilon_i$$

E per la stima di a , si va sempre a recuperare la quantità 5.12, riadattandola come nel caso dell'anidride carbonica:

$$S(\tilde{a}) = \sum_{i=2}^n [\delta \hat{D}_i - \tilde{a}^{[t_i - t_{i-1}]} \delta \hat{D}_{i-1}]^2$$

Dove \tilde{a} viene ottenuto ancora una volta tramite la minimizzazione di tale quantità, con un algoritmo numerico e nello specifico con *il metodo di Brent*, fornendo come valore iniziale $1/\exp(1)$. Il valore di \tilde{a} ottenuto con tale metodologia è pari a **0,99**, il quale corretto per il fattore di distorsione (si veda la formula 5.18) si trova essere **0,992**. Utilizzando infine la relazione inversa che lega a al valore che quantifica la persistenza usando l'unità di misura temporale, cioè τ , si ha che quest'ultimo per la serie riguardante la variazione di Deuterio, vale (dopo i successivi passaggi) $\tau = -\bar{d}/\ln(\hat{a})$, $\tau = -0.128/\ln(0,992)$, $\tau = 15,936 \text{ ka}$.

5.5.6 Stima della correlazione

L'ultimo sottoparagrafo è riservato alla quantificazione della correlazione tra l'anidride carbonica e della variazione del Deuterio, mediante l'applicazione della ***Binned Correlation***, suggerita nel libro *Climate Time Series Analysis* e rappresenta lo scopo principale dello studio.

La prima operazione di tale procedura consiste nel ricercare il minimo ed il massimo istante temporale comune, i quali sono risultati rispettivamente pari a **0** e **422,766 ka**. Nel secondo step si procede al calcolo di $\bar{\tau}$, il quale rappresenta la persistenza comune e la lunghezza delle categorie temporali (*bins*) entro le quali inserire le osservazioni di ambedue le serie; prima però si è determinato il valore della distanza e del coefficiente di autocorrelazione comune alle due serie (\bar{d}_{XY} e \hat{a}_{XY}), i quali risultando rispettivamente **0,1177** e **0,9771**, hanno dato luogo ad un valore per $\bar{\tau}$ pari a **5,1 ka**. Il terzo passo è quello di valutare le singole categorie di tempo (*time bin*), ovvero l'analisi di se e quante osservazioni di entrambe le serie sono contenute all'interno delle stesse. Una volta svolto ciò, si è dunque costruito il nuovo dataset dove sono presenti la nuova griglia temporale comune e le due nuove serie, ottenuti secondo quanto enunciato nel 5.4.1; in definitiva esso ha presentato una numerosità campionaria pari a **79**. Infine si è calcolato il grado di relazione tra le due proxy climatiche tramite il coefficiente di correlazione di Pearson ed è risultato pari a **0,85**.

6 Conclusioni

Questo lavoro ha voluto principalmente mostrare, sia da un punto di vista teorico che pratico, due differenti tecniche di stima della correlazione tra due serie storiche : il *Moving Block Bootstrap* a coppie per le osservazioni equispaziate e la *Binned Correlation* per quelle non equispaziate.

La prima metodologia è stata applicata alle parti stocastiche delle serie storiche a coppie di stazioni considerate successivamente e si è trovato che i due centri di monitoraggio intermedi (Bratislava e Nagymaros) presentano il coefficiente di correlazione più alto, mentre i due finali (Nagymaros e Drobeta – Turnu Severin) quello più basso. Tutto ciò rientra nelle aspettative, perchè in quel tratto del percorso il Danubio incrocia i tre suoi principali affluenti, la Drava, la Sava ed il Tibisco, i quali creano un'azione di disturbo al corso d'acqua tra le ultime due stazioni ed alimentano dunque il rischio di fenomeni alluvionali nella suddetta zona, peraltro già ben frequenti (e devastanti) anche per cause strutturali (dighe fatiscenti).

Il secondo (più interessante) metodo riguarda l'applicazione al contesto delle serie non equispaziate dell'anidride carbonica e della variazione di Deuterio nella carota di ghiaccio di Vostok. Il risultato trovato ha dimostrato non solo il grado di correlazione molto elevato tra la proxy dell'effetto serra e la variazione di temperatura, ma il fatto che tale relazione sia praticamente esistita da 420000 anni or sono, e di conseguenza amplificata nei secoli con l'avvento dell'era industriale. Lo studio inoltre ha permesso tramite l'analisi spettrale (*Ferraz-Mello,1981*) delle due proxy di validare un altro importante risultato, ossia l'influenza dei cicli orbitali compiuti dalla terra sul proprio clima (*Teoria Astronomica del Clima*). Ambedue le serie presentano significativi picchi in corrispondenza delle frequenze associate alle periodicità caratterizzanti l'*eccentricità* (100000 anni) e (con entità minore) l'*obliquità* (41000 anni).

Tale risultato consentirebbe di poter stimare quasi con certezza i futuri livelli delle due proxy e di conseguenza l'andamento del clima terrestre, ma un appuntamento con una nuova grande era glaciale (atteso fra 50000 anni) potrebbe anche essere rimandato, in quanto nel 2100 le emissioni di CO_2 raggiungeranno un valore compreso tra i 500 ed i 1000 ppmv (*Mercalli,2009*).

7 Ringraziamenti

Il ringraziamento più grande ovviamente non può che andare verso la mia famiglia che mi ha sopportato, sorretto e soprattutto spronato nei numerosi momenti di difficoltà che ho incontrato nel redigere tale elaborato, le tisane della mamma si sono rivelate fondamentali per mantenere la calma nelle ultime interminabili notti di lavoro. Un grazie particolare va al papà, per quello che mi ha trasmesso e fatto per me e non solo nelle ultime settimane.

Un sentito grazie va ovviamente a tutta la mia fantastica compagnia (che ora probabilmente mi vedrà uscire la sera un po' di più) per tutto quello che ha rappresentato in questi anni e non coincide soltanto come una mera valvola di sfogo e divertimento, ma ben molto di più.

Un ringraziamento speciale va al professor Bordignon, mio relatore, non solo per la pazienza e disponibilità con le quali mi ha seguito nello svolgimento di questa tesi, ma anche per avermi consigliato lo strumento imprescindibile per la sua realizzazione, ovvero il libro *Climate Time Series Analysis*, senza il quale questo lavoro probabilmente non sarebbe mai stato realizzato.

L'altro grazie inerente al suddetto libro va alla Biblioteca di Scienze Statistiche che in maniera celere e professionale è riuscita gentilmente a procurarmelo.

Altri ringraziamenti meritano tutti i miei amici e colleghi universitari, in particolare Federico, Emanuele ed Andrea con i quali ho costruito un solido rapporto d'amicizia ed hanno sicuramente contribuito a rendere quest'avventura difficilmente dimenticabile.

Infine, ma certamente non per ordine di importanza, un grazie ovviamente va a chi, mi ha aiutato a crescere e che da lassù sicuramente lo sta facendo ancora.

8 Bibliografia

Peter Bloomfield *Fourier Analysis of Time Series An Introduction* Second Edition, 1976

Lorenzo Caponovo, Oliver Scaillet, Fabio Trojani. *Robust Resampling Methods for time series*, May 2010

C.Codeiro, M. Neves. *The bootstrap methodology in the time series forecasting*

A.C. Davinson, D. Kuonen. *An Introduction to the bootstrap with applications in R*. Statistical Computing & Statistical Graphics. Newsletter. Vol 13 No.1

T. Di Fonzo, F.Lisi *Serie storiche economiche*, Carrocci editore, luglio 2005

A.C. Davison, D.V. Hinkley. *Bootstrap Methods and their Applications*, United States of America, Cambridge University PRESS 2005

John Fox, Sanford Weisberg. *Nonlinear Regression and Nonlinear Least Squares in R*. December 2010

Stefano Furlani. *Il Quaternario*, Università degli Studi di Padova. Appunti del corso di geomorfologia

Sergio Frasca. *Analisi Statistica dei Segnali cap. 6*, Dipartimento di Fisica, Università di Roma "La Sapienza"

Luke Kances, Alfred Lam, Eric Lin, Matthew Stringer. *Bootstrap Estimation of Variance*, Imperial College of London, June 2010

S. Ferraz-Mello. *Estimation of Periodes from unequally spaced observation* . The astronomical journal. Volume 86, number 4. April 1981

Manfred Mudelsee *Climate Time Series Analysis, classical statistical and bootstrap Methods*. Springer, 2010

Mazzola Claudia. Il Pleistocene medio-superiore nella carota di ghiaccio di Talos Dome (Antartide Orientale): *ricostruzione ad alta risoluzione del record delle polveri fini atmosferiche e loro significato paleoclimatico*.

Tesi di Dottorato di ricerca in Scienze della Terra, Università degli Studi di Milano – Bicocca, Anno Accademico 2008-2009

Luca Mercalli *Che tempo che farà. Breve storia del clima con uno sguardo al futuro*. Rizzoli Editore, 2009

Barbara Stenni. *Applicazioni degli isotopi stabili in paleoclimatologia: le carote di ghiaccio*. Studi Trentini Scienze Naturali, 2003

Claudio Smiraglia. *I ghiacciai polari, grandi archivi della storia del clima terrestre*. Comitato Geologico Italiano

J.R. Petit, J. Jouzel, D.Raynawd, N.I.Barkov, M. Barnola, I. Basile, M. Berder, J. Chappellaz, M. Davis, G. Delaygue, M. Delmotte, V.M. Kotlyakov, M. Legrand, V.Y. Lipenkov, C.Lonus, L. Pepin, C. Ritz, E. Saltzman, M. Stievenard, *Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antartica*. Nature, Vol. 399, June 1999

Pierluigi Pieruccini. *Le Variazioni Climatiche: il Paleoclima*. Dispense del corso di geografia di fisica. Università di Siena

Vito Ricci. *Analisi delle serie storiche con R*. 2005

Richard M. Vogel, Amy L. Shalleross. *The Moving Block Bootstrap versus parametric time series models*. Wother Resources Rescarch; June 1996

P.Yiou, C. Genthon, M. Ghil, J. Jouzel, H. Le Treut, M. Barnola, C. Lorius, and Y.N. Korotkevitch.

High Frequency Paleovariability in Climate and CO2 leves from Vostok ice cares records. Journal of geophysical Research, Vol. 96. November 1991

Siti web consultati, oltre a quelli citati nell'elaborato

<http://www.aari.nw.ru> , Arctic and Antarctic Research Institute

<http://www.wikipedia.com>

<http://www.nasa.gov/> National Aeronautics and Space Administration