

Capitolo 10

Modelli per serie storiche circolari

Maria Maddalena Barbieri

Università degli Studi Roma Tre

Abstract

In molti contesti applicativi sono presenti variabili le cui realizzazioni sono rappresentabili attraverso angoli in un cerchio e per questa ragione vengono dette circolari. Esempi tipici riguardano la direzione del vento, delle onde o del percorso di un animale, o anche l'orario di un evento fisiologico o comportamentale, come l'insorgenza di un sintomo o l'inizio di un'attività. La natura ciclica della scala di misurazione rende non applicabile la quasi totalità degli indicatori di sintesi comunemente adottati per dati a valori reali. Nel lavoro passiamo in rassegna i principali metodi di analisi adottabili in questo contesto con riferimento principalmente a serie storiche circolari, cioè a sequenze di osservazioni su una variabile circolare raccolte nel tempo, tipicamente in istanti equidistanti.

In many applied contexts, there are variables whose realizations can be represented as angles; for this reason, they are referred to as circular variables. Typical examples include wind direction, wave direction, or an animal's movement direction, as well as the timing of physiological or behavioral events, such as the onset of a symptom or the start of an activity. The cyclical nature of the measurement scale makes most of the summary measures commonly used for real-valued data inapplicable. In this work, we review the main analytical methods suitable for this setting, with particular focus on circular time series, that is, sequences of observations of a circular variable collected over time, typically at equally spaced time points.

Parole-chiave: Dati circolari, Modelli autoregressivi circolari, Modelli hidden Markov, Modelli threshold autoregressivi, Serie storiche circolari

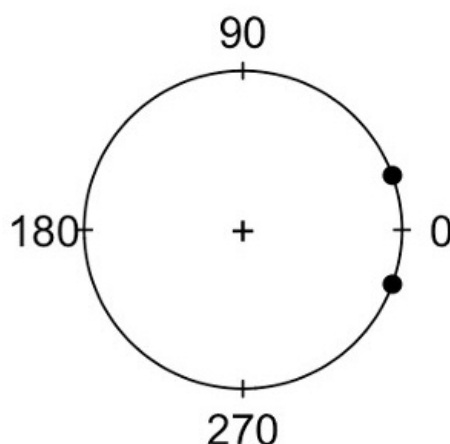
Keywords: Circular data, Circular autoregressive models, Hidden Markov models, Threshold autoregressive models, Circular time series.

1. Introduzione

I dati circolari rappresentano il caso più semplice di dati direzionali e la loro analisi costituisce un ambito peculiare della statistica.

A differenza delle variabili tradizionali, che assumono valori sulla retta reale, le variabili circolari hanno un campo di variazione in cui l'estremo superiore coincide con quello inferiore e tendono a presentarsi seguendo una dinamica ciclica. In altre parole, le loro realizzazioni vengono tipicamente espresse come angoli in un cerchio o come punti su una circonferenza di raggio unitario, misurati dopo aver scelto in modo opportuno un'origine di riferimento e una direzione di rotazione. Per questa ragione spesso ci si riferisce a dati di questo genere chiamandoli circolari.

La direzione del vento, quella del moto ondoso o del percorso di un animale, l'orario di un evento, la fase di un ritmo biologico o l'orientamento di una spaccatura in una roccia sono esempi di variabili circolari. In questi casi, l'estremo inferiore del campo di variazione della variabile coincide con l'estremo superiore. Ad esempio, nel caso di angoli misurati in radianti, il punto 0 coincide con il punto 2π ; analogamente, nel caso di un istante del giorno, l'orario 00:00 coincide con le 24:00 del giorno precedente. Questa caratteristica ha conseguenze pratiche importanti e rende non appropriate molte delle procedure di analisi usualmente adottate per dati a valori reali. Infatti, uno spostamento tra due punti posizionati ai margini del campo di variazione di una variabile circolare, cioè sui due lati opposti dell'origine, rappresenta una variazione minima sul cerchio, ma può apparire come una brusca discontinuità se misurato nella metrica lineare. Se calcolassimo la media aritmetica dei due valori con la formula tradizionale otterremmo come risultato un valore che non risulterebbe compreso tra i due originari, ma anzi sarebbe posizionato sul lato opposto della circonferenza. Per fare un esempio, supponiamo di aver osservato una variabile circolare misurata in gradi in due circostanze diverse e che i due valori risultanti siano 20 e 340.

Fig. 1 – Rappresentazione grafica di due osservazioni circolari misurate in gradi

In Fig. 1 abbiamo rappresentato graficamente ciascuno dei due dati come un punto sul cerchio. Dal grafico si vede che, tenendo conto del fatto che lo spazio di osservazione è il cerchio, i due punti non sono distanti e che la media dei due valori è rappresentata dal valore 0. Se però trattassimo i due valori come osservazioni da una variabile lineare il valore della loro media aritmetica sarebbe 180.

Queste difficoltà sono ben note nella statistica dei dati circolari, un ambito consolidato che fornisce metodi adeguati a descrivere e analizzare osservazioni di questo tipo.

La letteratura sulla statistica circolare e direzionale si è sviluppata lungo un arco temporale di diversi decenni, passando da contributi fondativi a trattazioni sempre più specializzate e orientate a contesti applicativi complessi. Nel percorso, alcuni testi occupano una posizione centrale per il ruolo che hanno avuto nel definire il quadro teorico e concettuale e nel guidarne gli sviluppi successivi. In questo contesto, il volume di Mardia e Jupp (2000) rappresenta sicuramente uno dei punti di riferimento più solidi e duraturi nello studio dei dati circolari e, più in generale, dei dati direzionali. Il testo introduce in modo sistematico i concetti chiave della statistica circolare, quali media direzionale, dispersione, correlazione e modelli parametrici fondamentali, enfatizzando il ruolo centrale della geometria dello spazio campionario. La sua importanza non risiede soltanto nella completezza del trattamento teorico, ma soprattutto nella capacità di fornire un quadro concettuale unificante per una vasta classe di problemi statistici che emergono quando le osservazioni non hanno una scala lineare, ma una

struttura geometrica come il cerchio o la sfera. Gli autori raccolgono e organizzano i principali contributi sull'argomento in un impianto coerente, mostrando come essi differiscano sostanzialmente dalle loro controparti lineari.

Mardia e Jupp (2000) è inoltre fondamentale per il modo in cui collega teoria e applicazioni. Gli esempi tratti da meteorologia, biologia, geologia e neuroscienze mostrano come le tecniche circolari non siano esercizi matematici astratti, ma risposte concrete a problemi reali. Questo approccio applicativo ha contribuito a diffondere l'uso dei metodi circolari al di fuori della statistica pura, influenzando campi come l'ecologia del movimento, la cronobiologia e le scienze ambientali. Infine, la longevità del libro è essa stessa un indicatore della sua centralità. A più di vent'anni dalla pubblicazione, continua a essere citato come riferimento standard, sia in lavori metodologici sia in applicazioni.

Un'impostazione in parte complementare è utilizzata in Jammalamadaka e Sengupta (2001), che rappresenta un altro testo di riferimento nella letteratura. Questi ultimi autori adottano un approccio più orientato a specifici problemi statistici e a sviluppi metodologici mirati, inclusa una maggiore attenzione ad aspetti inferenziali e ad alcune estensioni dinamiche. Il loro volume approfondisce misure di dipendenza, test di ipotesi ed estensioni metodologiche rilevanti anche per dati longitudinali e temporali.

Un ulteriore punto di riferimento è rappresentato dalla rassegna di Pewsey e García-Portugués (2021), che offre una sintesi aggiornata degli sviluppi più recenti nel campo, evidenziandone le estensioni e le direzioni future. Questo contributo si distingue per il suo taglio metodologico e per l'attenzione rivolta a temi associati ad aspetti computazionali. La rassegna mostra chiaramente come i concetti introdotti nei testi classici siano stati estesi e adattati per affrontare problemi sempre più complessi.

Per una introduzione all'argomento rimandiamo anche a Cremers e Klugkist (2018), che forniscono anche spiegazioni intuitive e suggerimenti operativi.

Uno dei contesti rilevanti nelle applicazioni riguarda l'analisi di serie storiche circolari, cioè di sequenze di dati risultato dell'osservazione di una variabile circolare nel tempo. Anche in questo caso la natura dei dati solleva questioni non banali a cominciare dalla definizione di persistenza, passando per la specificazione del modello, l'identificazione e la stima dei parametri,

e la successiva interpretazione dei risultati.

Negli ultimi decenni, diversi filoni di ricerca, accomunati dall'obiettivo di descrivere come una quantità ciclica evolve nel tempo, hanno cercato di estendere i modelli per serie storiche tradizionali al contesto circolare, adattandoli a una geometria non lineare. Tuttavia, questi modelli, pensati per situazioni regolari, faticano a rappresentare fenomeni che mostrano comportamenti meno regolari e difficilmente rappresentabili attraverso un'unica schematizzazione, cioè nei quali periodi di stabilità si alternano a cambiamenti improvvisi o nel tempo si susseguono periodi in cui il fenomeno sembra presentarsi con caratteristiche diverse. Sono così stati proposti modelli che descrivono la persistenza angolare, modelli capaci di rappresentare cambiamenti di regime, modelli più flessibili di tipo non parametrico e metodi di analisi che fanno uso di processi latenti utilizzati adottando prevalentemente l'impostazione inferenziale frequentista, ma anche quella bayesiana. Molti di questi approcci sono stati generalizzati al caso multivariato ed anche alla situazione comune in cui una variabile circolare viene osservata insieme ad una variabile lineare, cioè quando la serie viene anche detta cilindrica, poiché la coppia di osservazioni riferite a ciascun istante può essere rappresentata geometricamente come un punto sulla superficie di un cilindro. Esempi tipici sono rappresentati da direzione e velocità del vento o del percorso di un animale, direzione e altezza del moto ondoso.

Nel resto del lavoro cercheremo di offrire una panoramica accessibile e non formalizzata dei principali modelli per serie storiche circolari, evitando gli aspetti tecnici e mettendo l'accento sulle idee di base, al fine di aiutare il lettore a orientarsi tra le diverse proposte, comprendendone la logica e le potenzialità. Verranno illustrati i diversi modi in cui la dipendenza temporale può essere incorporata nei modelli, i contesti applicativi in cui ciascun approccio risulta più adatto e le principali questioni interpretative.

2. Serie storiche circolari: definizione e modelli

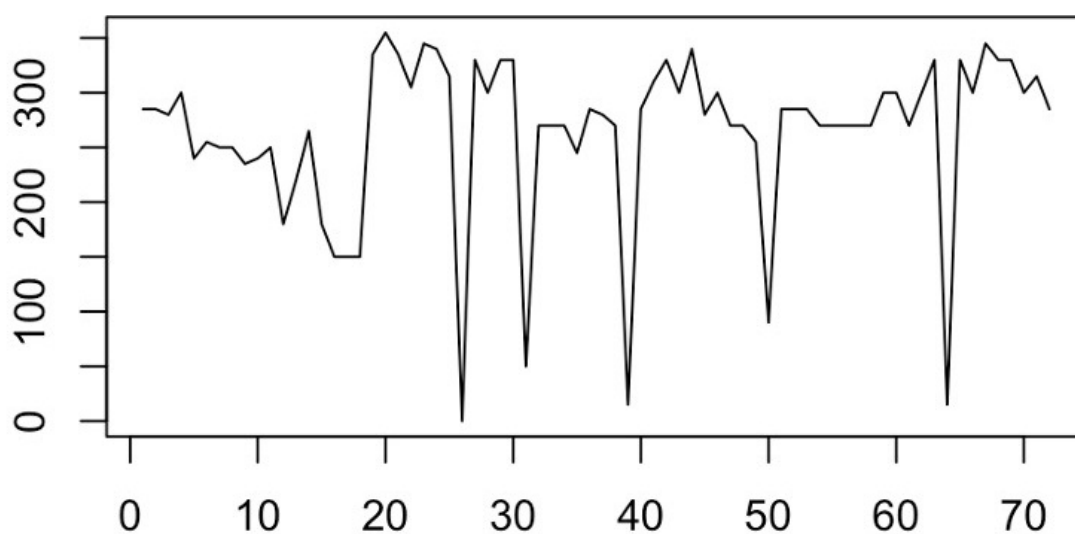
Una *serie storica circolare* è una sequenza di dati osservati nel tempo, a valori su uno spazio circolare, di solito caratterizzate da dipendenza seriale. Nelle situazioni più comuni i dati sono osservati a cadenza regolare, cioè in istanti equidistanti o su intervalli di tempo equispaziati.

Preliminarmente all'analisi di qualunque tipologia di dati si ricorre ad una opportuna rappresentazione grafica delle informazioni raccolte. Tuttavia, una questione centrale nell'analisi di serie storiche circolari risiede nel fatto che nessuna rappresentazione grafica disponibile sembra essere in grado di far rilevare simultaneamente tutte le caratteristiche potenzialmente rilevanti.

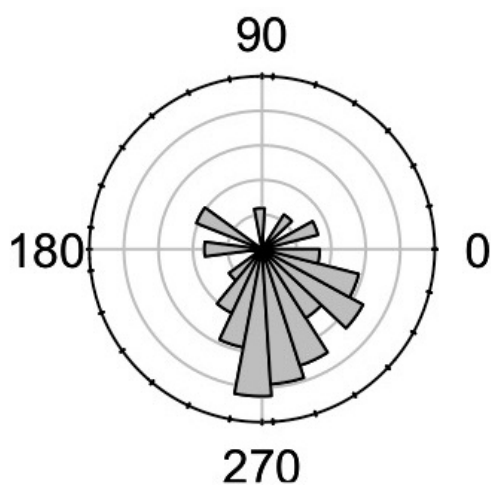
Come è bene noto, una serie storica lineare viene convenzionalmente rappresentata graficamente su un piano cartesiano posizionando sull'asse delle ascisse il tempo e sull'asse delle ordinate il valore della variabile misurata, unendo i punti riferiti ad osservazioni consecutive attraverso un segmento di retta. Il grafico risultante rappresenta il metodo standard e più efficace per mostrare tendenze, ciclicità, presenza di stagionalità o di anomalie, quali outlier o cambi di livello, nell'andamento di una serie lineare. La stessa rappresentazione viene comunemente utilizzata anche per serie storiche circolari, poiché non è possibile definirne una analoga sul cerchio. Come esempio, in Figura 2 riportiamo l'andamento di una serie storica di 72 osservazioni, che rappresentano la direzione del vento misurata ogni ora per quattro giorni consecutivi in un sito posizionato sulla Black Mountain in Australia. La serie è stata analizzata in Fisher e Lee (1994). Il grafico è utile per mettere in evidenza le principali difficoltà che si presentano nell'utilizzo di uno strumento di questo genere per dati circolari e per sottolineare la necessità di una interpretazione opportuna dei risultati.

Nel grafico in Figura 2 notiamo alcune fluttuazioni tra valori vicini agli estremi del campo di variazione $[0, 360)$, che possono apparire come la presenza di dati anomali, ma la cui presenza è indotta dall'utilizzo della metrica lineare per rappresentare graficamente la variabile circolare. Notiamo infatti che in realtà le fluttuazioni si riferiscono ad osservazioni che sono vicine se consideriamo una metrica che rispetti la topologia circolare. In particolare, il primo picco in basso presente nel grafico corrisponde all'istante 26. Il valore corrispondente, pari a zero, sulla circonferenza è relativamente vicino al precedente, pari a 315, ed al successivo, pari a 330.

Nella Fig. 3 è rappresentata la distribuzione dei dati osservati attraverso un diagramma a rosa, l'equivalente sul cerchio di un usuale istogramma. Il grafico consente di valutare la posizione e la dispersione della distribuzione, ma ovviamente non dà informazioni sulla dinamica temporale del fenomeno.

Fig. 2 – Direzione del vento misurata ogni ora

Fonte dei dati: Fisher e Lee, 1994

Fig. 3 – Diagramma a rosa della direzione del vento

Fonte dei dati: Fisher e Lee, 1994

Come già accennato nell'introduzione, la peculiarità dei dati circolari risiede nel fatto che operazioni fondamentali come la media, la varianza e l'autocorrelazione non possono essere definite in modo coerente utilizzando gli strumenti standard della statistica lineare (Mardia e Jupp, 2000). In particolare, la nozione di persistenza deve essere riformulata tenendo

conto della periodicità intrinseca della variabile.

La misura della relazione tra due variabili circolari non è così diretta come nel caso di due variabili a valori reali e questo essenzialmente per la mancanza di una relazione naturale tra variabili circolari. Di conseguenza, sono state fatte numerose differenti proposte, ciascuna corrispondente ad uno specifico tipo di associazione tra variabili. In analogia al caso lineare hanno tutte la denominazione di coefficiente di correlazione circolare. Le definizioni e le proprietà di molti di essi vengono riportati in Mardia e Jupp (2000) e in Jammalamadaka e Sengupta (2001).

I principali modelli per serie storiche circolari sono il risultato dei tentativi di trasferire le idee classiche dell'analisi delle serie storiche nel contesto dei dati circolari. Ciò che distingue questi modelli dalle controparti lineari è il modo in cui la dipendenza temporale viene adattata alla geometria del cerchio e la dinamica definita in modo coerente con la ciclicità della variabile, evitando discontinuità artificiali.

Una prima famiglia di modelli è costituita da versioni circolari degli usuali modelli autoregressivi, che, nel caso stazionario, rappresentano quello che potenzialmente accade in un istante in funzione lineare di quanto verificatosi negli istanti precedenti, assumendo che l'errore abbia distribuzione normale. Gli adattamenti al caso circolare ovviamente adottano una distribuzione definita sul cerchio per la componente di variabilità casuale. I diversi approcci proposti in questo ambito permettono di sfruttare strumenti ben noti dell'analisi delle serie storiche, mantenendo una connessione con i modelli ARMA classici. Tuttavia, l'interpretazione dei parametri risulta spesso meno immediata, la dinamica sul cerchio è descritta in modo indiretto e i metodi utilizzabili per adattarli a serie osservate meno immediati da utilizzare. Tra l'altro non risultano disponibili software che ne consentano l'implementazione. Probabilmente a causa di tutte queste ragioni, tali modelli sono più diffusi in contesti metodologici che applicativi.

Breckling (1989) formula due proposte di modelli per serie storiche circolari, che chiama processo di von Mises e modello autoregressivo Wrapped. I due modelli sono collegati, nell'ordine, alla distribuzione di von Mises e alla distribuzione Wrapped Normal. In particolare, la distribuzione di von Mises dipende da un parametro di posizione e uno di concentrazione (il reciproco della dispersione); mentre la distribuzione Wrapped Normal è ottenuta arrotondando intorno sul cerchio la distribuzione normale definita

nello spazio euclideo. Entrambe le distribuzioni hanno una funzione di densità con forma simmetrica, che approssima quella della normale se la variabilità è bassa, e godono di importanti proprietà, che le rendono le due più importanti distribuzioni di probabilità definite sul cerchio (Mardia e Jupp, 2000; Jammalamadaka e SenGupta, 2001). La distribuzione di von Mises è più utilizzata, grazie alla possibilità di ricavare analiticamente gli stimatori di massima verosimiglianza dei parametri, mentre nel caso della Wrapped Normal è necessario ricorrere a metodi numerici di ottimizzazione.

Nel processo di von Mises viene utilizzata una struttura autoregressiva che riguarda un vettore di concentrazione. È quindi attraverso quest'ultimo che passa la dipendenza da quanto accaduto in precedenza. Mentre il modello autoregressivo Wrapped è ottenuto avvolgendo sul cerchio un modello autoregressivo lineare latente (cioè non osservabile) di ordine opportuno e errori con distribuzione normale.

Il modello autoregressivo Wrapped è ottenuto partendo da un modello autoregressivo di ordine opportuno, che rappresenta una serie storica latente (cioè non osservabile) definita sulla retta reale. Si arriva al modello Wrapped trasformando la serie lineare in una serie circolare tramite una riduzione che consiste nel calcolare l'operazione di modulo il periodo, in cui il periodo è pari, ad esempio, a 2π nel caso in cui la variabile sia misurata in radianti o a 24 se in ore del giorno. In parole povere, il modello Wrapped si ottiene prendendo valori reali e "avvolgendoli" su un cerchio, facendo coincidere tutti quelli che differiscono per un multiplo del periodo.

Breckling (1989) propone di ricorrere all'algoritmo EM (Expectation-Maximisation - Dempster, Laird e Rubin, 1977) per la stima dei parametri di un modello di ordine 1 e analizza le condizioni sotto le quali un modello autoregressivo Wrapped risulti approssimato da un processo di von Mises. La validità di tali condizioni risolve i problemi computazionali necessari a individuare le stime di massima verosimiglianza dei parametri del modello Wrapped quando l'ordine è maggiore di 1.

Il modello autoregressivo Wrapped viene ripreso da Fisher e Lee (1994), che generalizzano al caso di modelli di ordine maggiore di 1 la proposta di ricorrere all'algoritmo EM (Expectation-Maximisation - Dempster, Laird e Rubin, 1977) per la stima dei parametri. Tuttavia, l'algoritmo risulta inefficiente dal punto di vista computazionale poiché il passo E dell'algoritmo stesso coinvolge rapporti di somme infinite, che devono essere ap-

prossimati in ogni iterazione. Coles (1998) considera lo stesso modello, ma adotta un'impostazione bayesiana e propone un algoritmo MCMC con un passo di data augmentation, per risolvere i problemi computazionali nell'operazione di stima.

Fisher e Lee (1994) introducono anche altri tipi di modelli, sempre sulla base di trasformazioni opportune di modelli lineari autoregressivi, quali la proiezione di un processo bivariato lineare sulla circonferenza di raggio unitario o l'utilizzo di una funzione link per passare dallo spazio circolare a quello lineare. Le proposte vanno sotto il nome di modelli autoregressivi Linked e autoregressivi Inversi.

Più di recente Erdem e Shi (2011) discutono vari modelli, ancora nell'ambito della classe ARMA, per rappresentare congiuntamente una variabile circolare e una lineare.

Il principale punto di forza dei modelli descritti finora è il loro equilibrio tra realismo e semplicità. Essi catturano la continuità temporale tipica di molti fenomeni circolari senza introdurre strutture eccessivamente complesse e i parametri hanno spesso un'interpretazione accessibile anche per utenti non specialisti. D'altra parte, questi modelli assumono che il fenomeno considerato mantenga un comportamento relativamente omogeneo in tutto il periodo di osservazione. In molte applicazioni reali, tuttavia, questa ipotesi risulta troppo restrittiva. Il comportamento del sistema può cambiare in modo sostanziale a seconda di condizioni esterne, fasi interne o contesti non direttamente osservabili. Ad ulteriore conferma, la distribuzione delle osservazioni spesso presenta più di una moda. Questa caratteristica suggerisce che i dati possano essere stati generati da una mistura di distribuzioni differenti, associate a regimi diversi. In tali casi, è naturale passare a modelli più articolati, come quelli a stati latenti.

Per dare una motivazione intuitiva della opportunità di considerare modelli di questo tipo possiamo riferirci al fatto che molti fenomeni ciclici mostrano comportamenti qualitativamente diversi in momenti diversi. Si pensi alla direzione del vento in condizioni meteorologiche stabili rispetto a situazioni perturbate, o all'orientamento di un animale durante la migrazione rispetto alle fasi di esplorazione locale. In questi casi, non è tanto o almeno non solo la singola osservazione precedente a determinare l'evoluzione futura, quanto il regime in cui il sistema opera.

I modelli a stati latenti permettono di rappresentare questa eteroge-

neità temporale introducendo livelli distinti di comportamento, ciascuno caratterizzato da una propria dinamica circolare. L'idea centrale di questi modelli è che la dinamica osservata sia governata, almeno in parte, da uno o più processi nascosti, che evolvono nel tempo e influenzano le osservazioni circolari. L'angolo osservato non è quindi descritto solo in funzione del passato immediato, ma anche dello stato in cui il sistema si trova in quel momento.

Tra i modelli a stati latenti più noti vi sono quelli nella classe dei modelli hidden Markov (Holzmann, Munk, Suster e Zucchini, 2006; Zucchini MacDonald e Langrock, 2016). In questi approcci, si assume che il sistema possa trovarsi in un numero finito di stati non osservabili, che cambiano nel tempo secondo regole probabilistiche, a ciascuno dei quali è associata una specifica distribuzione circolare. Dal punto di vista interpretativo, questi modelli sono particolarmente interessanti perché consentono di individuare regimi distinti nel comportamento dei dati. Ad esempio, una serie di direzioni può alternare periodi in cui è fortemente concentrata attorno a un valore medio a periodi in cui è molto più dispersa o orientata in modo diverso.

L'idea ha riscosso parecchio successo per la sua flessibilità ed è stata generalizzata in numerosi articoli, anche per dati cilindrici. In quest'ultimo caso le proposte includono anche l'obiettivo di rilassare l'ipotesi di indipendenza condizionata allo stato latente tra la variabile circolare e quella lineare, che è presente nell'approccio tradizionale (Lagona, Picone e Maruotti 2015, Mastrantonio, Maruotti e Jona-Lasinio 2015).

Nella letteratura, modelli di questo tipo sono stati applicati con successo sia all'analisi di dati ambientali, quali vento e onde, che allo studio dei movimenti animali, dove gli stati latenti sono spesso interpretati come modalità comportamentali diverse, come esplorazione, spostamento diretto o riposo (Holzmann, Munk, Suster e Zucchini, 2006; Langrock, King, Matthiopoulos, Thomas, Fortin, e Morales, 2012; Lagona, Picone e Maruotti 2015, Mastrantonio, Maruotti e Jona-Lasinio 2015; Zucchini MacDonald e Langrock, 2016).

Tutti gli articoli menzionati finora assumono che le osservazioni successive siano indipendenti, condizionatamente allo stato occupato, cioè che la relazione dinamica possa essere spiegata essenzialmente tramite cambiamenti di regime. Tuttavia, questa assunzione risulta spesso non realistica,

dal momento che, anche condizionatamente al regime occupato, nelle serie storiche osservate di solito è presente un legame forte tra osservazioni vicine.

Per rispondere a questa criticità, è possibile ricorrere a modelli hidden Markov in cui a ciascuno stato è associata una specifica struttura di dipendenza temporale. In particolare (Ailliot e Monbet, (2012) e Ailliot, Bessac, Monbet e Pene (2015) suggeriscono l'utilizzo di diversi modelli autoregressivi Markov Switching per la direzione del vento, anche quando si voglia tener conto della velocità. L'idea di base è che i dati seguano un modello autoregressivo (nel quale il valore presente dipende dai valori passati) i cui parametri non sono fissi, ma cambiano a seconda del *regime* in cui il sistema si trova. Il passaggio da un regime all'altro è governato da una catena di Markov, cioè la probabilità di essere in un certo regime in un istante dipende solo dal regime occupato nell'istante precedente e non da quello nell'intera storia passata. In altre parole, la classe dei modelli autoregressivi Markov switching, introdotta in origine nel caso di una generica serie temporale da Hamilton (1989), generalizza sia i modelli hidden Markov che i modelli autoregressivi. Lo fa combinando diversi modelli autoregressivi utili a rappresentare l'evoluzione della serie storica, con le transizioni tra uno e l'altro governate da una catena di Markov, come nei modelli hidden Markov. Più nello specifico, Ailliot, Bessac, Monbet e Pene (2015) considerano serie storiche bivariate della direzione e della velocità del vento e assumono che il processo Markov Switching sia non omogeneo, cioè che la probabilità di passaggio da un regime ad un altro dipenda anche dalla direzione del vento nell'istante precedente. Condizionatamente al regime e alle precedenti condizioni del vento, il modello per la direzione del vento è un processo di von Mises (Breckling, 1989), mentre per la velocità del vento un autoregressivo lineare con errori normali. Si assume che gli ordini sia del processo di von Mises che del modello autoregressivo lineare siano gli stessi in ogni regime, ma che il numero di regimi e la classificazione delle osservazioni nei diversi regimi possa essere diversa nelle due componenti della serie. Attraverso le applicazioni, Ailliot, Bessac, Monbet e Pene (2015) mettono in evidenza come i modelli da loro proposti portino ad una migliore rappresentazione delle caratteristiche delle serie bivariate e una migliore descrizione dell'evoluzione congiunta della velocità e della direzione del vento rispetto ai modelli hidden Markov che non considerano la struttura auto-

regressiva all'interno dei regimi.

L'idea di base delle proposte appena ricordate viene ripresa da Zhan, Ma, Liu e Shimizu (2018), che per l'analisi della serie delle direzioni del vento propongono un modello Markov Switching con una struttura autoregressiva circolare di tipo Linked.

Barbieri, Battaglia e Cucina (2025, 2026) affrontano il problema della costruzione in un modello per una serie storica con due componenti, una circolare e una lineare, con un numero potenziale di regimi maggiore di uno, in grado di tener conto dell'eventuale presenza di correlazione all'interno di ciascun regime, al fine anche di migliorare la previsione. In tali lavori vengono considerati diversi modelli. In alcuni il passaggio da un regime all'altro è regolato da una catena di Markov, come nei modelli hidden Markov, in altri è determinato dai valori precedenti come nei modelli autoregressivi a soglia (threshold - Tong e Lim, 1980). Quest'ultima caratteristica, cioè il fatto che la classificazione dei dati avvenga in modo deterministico sulla base dei valori delle osservazioni raccolte un certo numero di istanti precedenti, fa sì che i modelli di tipo autoregressivi threshold se usati a fini previsivi abbiano prestazioni migliori rispetto ai modelli di tipo hidden Markov, in cui il regime corrispondente a ciascuna osservazione è non osservato e aleatorio. Per i modelli di tipo hidden Markov, vengono proposte diverse specificazioni basate su differenti distribuzioni congiunte riguardanti le componenti cilindrica e lineare della serie ed impiegati varie strategie per l'adattamento dei modelli, tutte basate sull'algoritmo EM. Con riguardo al threshold, l'adattamento dei modelli viene effettuato impiegando un algoritmo genetico (Holland, 1975) e vengono proposte due diverse definizioni di soglia bivariate, cioè definite sulla base dello spazio di definizione di entrambe le variabili (Tsay, 1998). Inoltre, come in Ailliot, Bessac, Monbet e Pene (2015), si assume che sia possibile che nel modello per la componente circolare si possa arrivare a una suddivisione delle variabili diverse. In ogni caso, l'ordine autoregressivo può differire in base al regime.

Di recente Perera, Hui, Huston e Welsh (2025) hanno formulato una proposta alternativa che tiene conto della eventuale presenza di regimi diversi, combinando un modello hidden Markov con una distribuzione circolare condizionata al regime latente. Per la stima dei parametri suggeriscono il ricorso al metodo della massima verosimiglianza, basato sull'impiego di opportune formule ricorsive. Il modello proposto considera

anche la correlazione seriale e quindi risulta adatto alla previsione.

Un diverso approccio è quello proposta da Harvey e Palumbo (2023) e da Harvey, Hurn, Palumbo e Thiele (2024) che generalizzano al caso di serie storiche circolari e cilindriche l'approccio detto score-driven, in origine proposto per serie storiche lineari in ambito econometrico (Harvey, 2013). In Harvey, Hurn, Palumbo e Thiele (2024) si tiene conto della non stazionarietà presente nella serie calcolando preliminarmente la differenza tra dati successivi, mentre Harvey e Palumbo (2023) considerano esplicitamente la possibile presenza di più regimi. Il loro approccio è finalizzato anche ad effettuare previsioni, dal momento che tiene conto della dinamica all'interno dei regimi.

3. Conclusioni

Nel lavoro abbiamo richiamato le caratteristiche più rilevanti dei dati circolari e ricordato che per la loro analisi non sia opportuno ricorrere ai metodi usualmente impiegati nel caso di variabili definite sulla retta reale. Si possono invece impiegare gli strumenti introdotti nella letteratura degli ultimi decenni come risposta alla necessità di studiare fenomeni che si presentano con una struttura ciclica. Dalla direzione del vento agli orientamenti comportamentali, dagli orari delle attività quotidiane alle fasi di processi biologici ed economici, molti dati non si dispongono lungo una linea retta, ma lungo un cerchio. Riconoscere questa caratteristica non è un dettaglio tecnico, bensì un passaggio concettuale fondamentale per una corretta interpretazione dei dati. In particolare, abbiamo esplorato le principali diverse tipologie di modelli per serie storiche circolari, adatte a contesti e obiettivi differenti. Le generalizzazioni dei modelli autoregressivi lineari offrono una descrizione semplice della persistenza nel tempo e sono utili in tutte le situazioni in cui il fenomeno tende a presentarsi in modo approssimativamente omogeneo nel tempo. Mentre i modelli a stati latenti permettono di rappresentare cambiamenti di regime e comportamenti eterogenei, consentendo di tener conto di dinamiche complesse e irregolari, a fronte di una maggiore complessità computazionale.

La ricerca sull'argomento continua ad essere attiva e la letteratura in espansione, anche con lo sviluppo di software idoneo. Come in tutti i con-

testi, la possibilità di disporre di idonei programmi di calcolo immediatamente utilizzabili è particolarmente importante per consentire la diffusione e l'applicazione delle nuove proposte.

Bibliografia

- Ailliot, P., Bessac, J., Monbet, V., Pene, F. (2015). Non-homogeneous hidden Markov-switching models for wind time series. *Journal of Statistical Planning and Inference*, 160, 75-88.
- Ailliot, P., e Monbet, V. (2012). Markov-switching autoregressive models for wind time series. *Environmental Modelling & Software*, 30, 92-101.
- Barbieri, M.M., Battaglia, F. e Cucina, D. (2026). Models for environmental cylindrical time series. In corso di pubblicazione in: *Environmental Statistics: Innovative Methods and Applications*. Eds: A. Bekker, J. Ferreira, P. Nagar, A. Ramoelo, B. Erasmus, CRC Press.
- Barbieri, M.M., Battaglia, F. e Cucina, D. (2025). Multi-regime autoregressive models for cylindrical time series. Sottoposto per la pubblicazione.
- Breckling, P. (1989). *The Analysis of Directional Time Series: Applications to Wind Speed and Direction* (Vol. 61). Springer.
- Coles, S. (1998). Inference for circular distributions and processes. *Statistics and Computing*, 8(2), 105-113.
- Cremers, J., e Klugkist, I. (2018). One direction? A tutorial for circular data analysis using R with examples in cognitive psychology. *Frontiers in psychology*, 9, 2040.
- Dempster, A.P., Laird, N.M., e Rubin, D.B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the royal statistical society: series B (methodological)*, 39(1), 1-22.
- Erdem, E., e Shi, J. (2011). ARMA based approaches for forecasting the tuple of wind speed and direction. *Applied Energy*, 88(4), 1405-1414.
- Fisher, N.I. e Lee, A.J. (1994). Time series analysis of circular data. *Journal of the Royal Statistical Society, Series B*, 56(2), 327-339.
- Hamilton, J.D. (1989). A new approach to the economic analysis of non-stationary time series and the business cycle. *Econometrica: Journal of the econometric society*, 357-384.
- Harvey, A.C. (2013). *Dynamic models for volatility and heavy tails: with applications to financial and economic time series* (Vol. 52). Cambridge University Press.
- Harvey, A., Hurn, S., Palumbo, D., e Thiele, S. (2024). Modelling circular time series. *Journal of Econometrics*, 239(1), 105450.
- Harvey, A., e Palumbo, D. (2023). Regime switching models for circular and linear time series. *Journal of Time Series Analysis*, 44(4), 374-392.

- Holland, J.H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.
- Holzmann, H., Munk, A., Suster, M., e Zucchini, W. (2006). Hidden Markov models for circular and linear-circular time series. *Environmental and Ecological Statistics*, 13(3), 325-347.
- Lagona, F., Picone, M., e Maruotti, A. (2015). A hidden Markov model for the analysis of cylindrical time series. *Environmetrics*, 26(8), 534-544.
- Langrock, R., King, R., Matthiopoulos, J., Thomas, L., Fortin, D., e Morales, J.M. (2012). Flexible and practical modeling of animal telemetry data: hidden Markov models and extensions. *Ecology*, 93(11), 2336-2342.
- Jammalamadaka, S.R. e Sengupta, A. (2001). *Topics in Circular Statistics*. World Scientific.
- Mardia, K.V. e Jupp, P.E. (2000). *Directional Statistics*. Wiley.
- Mastrantonio, G., Maruotti, A., e Jona-Lasinio, G. (2015). Bayesian hidden Markov modelling using circular-linear general projected normal distribution. *Environmetrics*, 26(2), 145-158.
- Perera, A.A.P.N.M., Hui, F.K., Huston, C., e Welsh, A.H. (2025). A circular hidden Markov model for directional time series data. *Japanese Journal of Statistics and Data Science*, 1-25.
- Pewsey, A. e García-Portugués, E. (2021). Recent advances in directional statistics. *TEST*, 30(1), 1-58.
- Tong, H., e Lim, K.S. (1980). Threshold autoregression, limit cycles and cyclical data. *Journal of the Royal Statistical Society: Series B (Methodological)*, 42(3), 245-268.
- Tsay, R.S. (1998). Testing and modeling multivariate threshold models. *Journal of the American Statistical Association*, 93(443), 1188-1202.
- Zhan, X., Ma, T., Liu, S., e Shimizu, K. (2018). Markov-switching linked autoregressive model for non-continuous wind direction data. *Journal of Agricultural, Biological and Environmental Statistics*, 23(3), 410-425.
- Zucchini, W., MacDonald, I.L. e Langrock, R. (2016). *Hidden Markov Models for Time Series*. Chapman & Hall.