



Università
Ca' Foscari
Venezia
Facoltà
di Economia

Corso di Laurea
in Statistica per l'Impresa

Prova finale di Laurea

Modellazione Statistica per
l'Ottimizzazione di Sistemi
Energetici

Relatrice

Prof. Irene Poli

Laureando

Denis Baldo

Matricola 815758

Anno Accademico

2011-2012

INDICE

Indice	i
Elenco delle tabelle	iii
Elenco delle figure	v
1 La costruzione di modelli statistici per l'ottimizzazione dei sistemi energetici	1
2 I sistemi per la realizzazione del comfort abitativo	7
2.1 Cenni storici	11
2.2 Sistemi e caratteristiche	11
2.2.1 Sistema di climatizzazione di base	13
2.2.2 Sistema di climatizzazione a zone	13
2.2.2.1 Sistemi totalmente ad aria	14
2.2.2.2 Sistemi aria-acqua	15
2.2.2.3 Sistemi totalmente ad acqua	16
2.2.2.4 Sistemi unitari basati su refrigeranti	16
2.2.3 Scelta del sistema di climatizzazione	16
2.2.4 Standard energetici	18
2.3 Il problema dell'ottimizzazione di un edificio ad uso uffici	19
2.3.1 Le variabili del problema	20
2.3.2 Le funzioni obiettivo e i vincoli	22
3 Metodi e modelli per lo studio delle serie storiche di dati HVAC	25
3.1 Alcuni metodi descrittivi per le serie storiche	26
3.1.1 Livellamento esponenziale	27
3.1.1.1 Livellamento esponenziale semplice	28
3.1.1.2 Metodo di Holt-Winters	28
3.1.1.3 Metodo di Holt-Winters stagionale	29
3.2 I modelli stocastici di classe ARMA	29
3.2.1 Elementi essenziali per la definizione dei processi stocastici	30
3.2.2 Alcuni modelli stocastici di classe ARMA	33
3.2.2.1 I modelli autoregressivi (AR)	33
3.2.2.2 I modelli media mobile (MA)	34
3.2.2.3 I modelli ARMA	35

3.2.2.4	I modelli ARIMA	37
3.2.2.5	I modelli ARIMA stagionali	37
3.2.2.6	I modelli ARMA con variabili esogene	38
3.2.3	La costruzione del modello secondo l'approccio di Box e Jenkins	40
3.2.3.1	Identificazione del Modello	40
3.2.3.2	Stima del Modello	41
3.2.3.3	Controllo Diagnostico	42
3.2.4	Previsione	42
4	Il problema dell'ottimizzazione di sistemi energetici	45
4.1	Un'introduzione ai criteri di ottimalità	46
4.2	Classificazione dei problemi di ottimizzazione	48
4.2.1	Problemi di ottimizzazione discreta	49
4.2.2	Problemi di ottimizzazione continua	49
4.3	Criteri multi-obiettivo per l'ottimalità	50
4.3.1	Ottimizzazione pesata	51
4.3.2	Ottimizzazione di pareto	52
4.4	Approcci di risoluzione	54
4.4.1	Metodo enumerativo	55
4.4.2	Metodi deterministici	55
4.4.3	Metodi stocastici	57
4.5	Algoritmi evolutivi	59
4.5.1	Algoritmi genetici	62
4.5.2	Ant colony optimization	66
4.5.3	Particle swarm optimization	69
5	La costruzione del modello statistico per la realizzazione del comfort abitativo	75
5.1	Modelli previsivi per le variabili climatiche esterne	75
5.1.1	La temperatura esterna	76
5.1.2	L'irraggiamento	80
5.1.3	La luminosità esterna	84
5.1.4	L'umidità esterna	88
5.2	Le variabili del sistema e le loro interrelazioni	92
5.3	Modelli previsivi per le variabili climatiche interne	93
5.4	Modelli previsivi per l'ottimizzazione del consumo di energia	96
5.5	Modelli previsivi per l'ottimizzazione dell'emissione di biossido di carbonio	99
5.6	Il modello stimato per la realizzazione dell'ottimo abitativo	103
	Conclusione	113
	Bibliografia	115

ELENCO DELLE TABELLE

2.1	Variabili del problema	20
2.2	Livelli delle variabili controllabili	21
2.3	Zone Comfort DGI	22
5.1	Sintesi del modello stimato sulla temperatura esterna	79
5.2	Confronto valori osservati e stimati della temperatura esterna	79
5.3	Sintesi del modello stimato sull'irraggiamento	83
5.4	Confronto valori osservati e stimati dell'irraggiamento	84
5.5	Sintesi del modello stimato sulla luminosità esterna	87
5.6	Confronto tra valori osservati e stimati della luminosità esterna	88
5.7	Sintesi del modello stimato sull'umidità esterna	91
5.8	Confronto tra valori osservati e stimati dell'umidità esterna	91
5.9	Sintesi del modello stimato sulla temperatura interna	95
5.10	Confronto tra valori osservati e stimati della temperatura interna	96
5.11	Sintesi del modello stimato sul consumo del sistema HVAC	98
5.12	Confronto tra valori osservati e stimati del consumo	99
5.13	Sintesi del modello stimato sulle emissioni di CO_2	102
5.14	Confronto tra valori osservati e stimati della CO_2 emessa	103
5.15	Codifica di Gray per una stringa di 4 bit	108

ELENCO DELLE FIGURE

2.1	I fattori che influenzano il Comfort Termico(14)	8
2.2	Consumo energetico degli edifici ripartito in base all'utilizzo finale (16)	10
2.3	Consumo energetico degli edifici ripartito in base alla fonte energetica (16)	10
2.4	Sistema di climatizzazione di base	13
2.5	Sistema di riscaldamento a volume costante	15
2.6	Sistema a volume d'aria variabile (VAV)	15
3.1	Esempio di un processo autoregressivo del secondo ordine	34
3.2	Esempio di un processo a media mobile del secondo ordine	36
4.1	Non-dominated Set, Pareto Front e vettori ideali in un problema di minimizzazione con due obiettivi f_1 e f_2 .	54
4.2	Tecniche ottimizzazione multi-obiettivo(23)	55
4.3	Rappresentazione di un generico flowchart degli algoritmi evolutivi	61
4.4	Un esempio di ricombinazione genetica: il 'Single Crossover'	65
4.5	Un esempio di mutazione nella codifica binaria	66
4.6	Comportamento delle formiche in natura	67
4.7	Esempio di come si evolvono le particelle	71
4.8	Esempi di topologie usate nel PSO	72
4.9	Rappresentazione di un generico flowchart del PSO	72
5.1	Temperatura esterna massima, media e minima rilevata	77
5.2	Analisi Classica sulla Temperatura Esterna	77
5.3	Analisi preliminare della temperatura esterna	78
5.4	Diagnostica del modello e previsione per la serie della temperatura esterna	80
5.5	L'influenza dell'atmosfera terrestre sull'irraggiamento	81
5.6	Intensità della radiazione solare in funzione delle condizioni atmosferiche	81
5.7	Irraggiamento rilevato	82
5.8	Analisi Classica sull'Irraggiamento	83
5.9	Analisi preliminare dell'irraggiamento	83
5.10	Diagnostica del modello e previsione per la serie dell'irraggiamento	84
5.11	Luminosità rilevata	85
5.12	Analisi deterministica sulla serie della luminosità esterna	86

5.13	Analisi preliminare sulla serie della luminosità esterna	87
5.14	Diagnostica del modello e previsione per la serie della luminosità esterna	88
5.15	Umidità esterna rilevata	89
5.16	Analisi deterministica sulla serie dell'umidità esterna	90
5.17	Analisi preliminare sulla serie dell'umidità esterna	90
5.18	Diagnostica del modello e previsione per la serie dell'umidità esterna	91
5.19	Matrice di correlazione delle variabili	92
5.20	Temperatura interna rilevata	93
5.21	Analisi deterministica sulla serie della temperatura interna	94
5.22	Analisi preliminare sulla serie della temperatura interna	94
5.23	Diagnostica del modello e previsione per la serie della temperatura interna	95
5.24	Rilevazioni del consumo del sistema HVAC	97
5.25	Analisi preliminare sulla serie del consumo	98
5.26	Diagnostica del modello e previsione per il consumo del sistema HVAC	99
5.27	Rilevazioni dell'anidride carbonica emessa dal sistema HVAC	100
5.28	Analisi deterministica sulla serie della CO ₂ emessa	101
5.29	Analisi preliminare sulla serie della CO ₂ emessa dal sistema	101
5.30	Diagnostica del modello e previsione per la serie della CO ₂ emessa	102
5.31	Possibile realizzazione del cromosoma	108
5.32	Rappresentazione della funzione Schwefel per n=2	110
5.33	Diagnostica del modello e previsione per il consumo del sistema HVAC	111

Capitolo 1

LA COSTRUZIONE DI MODELLI STATISTICI PER L'OTTIMIZZAZIONE DEI SISTEMI ENERGETICI

Il crescente interesse rivolto ai sistemi di riscaldamento, ventilazione e climatizzazione dell'aria (HVAC) da parte della ricerca e da parte dei responsabili del settore è dovuto agli sviluppi ottenuti negli ultimi anni.

Il mondo sta affrontando la necessità di ridurre drasticamente il consumo di energia e il tema coinvolge queste tipologie di sistemi, principalmente poiché la maggior parte del consumo di energia viene impiegata negli edifici. La tematica assume rilevanza soprattutto nei paesi industrializzati, dove il parco immobiliare esistente è molto vasto e le ristrutturazioni degli edifici sono molto costose; in questi paesi i nuovi progetti di costruzione possono contribuire solo in parte alla riduzione del consumo di energia per cui il controllo dell'efficienza energetica degli edifici esistenti diventa importante.

Un contributo fondamentale deriva dallo sviluppo moderno delle tecnologie che ha permesso di incrementare notevolmente la capacità di calcolo dei sistemi così come ha reso disponibile nel mercato sensori a basso costo. Tutti questi miglioramenti hanno consentito nuovi sviluppi nel controllo del consumo energetico per il comfort abitativo all'interno degli edifici.

Attualmente, la metodologia maggiormente utilizzata per il controllo del com-

fort è detta "Rule Based Control" (RBC); si tratta di controllori che si basano su regole del tipo "se condizione, allora azione" e sono costituiti da un numero molto elevato di vincoli e parametri che, per poter descrivere le condizioni e le azioni, devono essere preventivamente determinati. Le prestazioni di questi "controller" dipendono strettamente dalla qualità dei parametri scelti; inoltre bisogna considerare che, a causa della crescente complessità dei sistemi HVAC e del desiderio di integrare ulteriori informazioni, una buona scelta diventa molto difficile da realizzare. In riferimento all'integrazione di ulteriore informazioni, un possibile approccio nell'ottimizzazione di questi sistemi prevede di utilizzare le informazioni sulle condizioni climatiche esterne e delle rispettive previsioni. I limiti di questa metodologia, che hanno spinto la ricerca di nuove soluzioni per il controllo del comfort di un edificio, sono principalmente due fattori: la capacità di utilizzare le previsioni meteorologiche nella modellazione al fine di ridurre il consumo energetico degli edifici e la capacità di gestire in modo più efficiente la complessità crescente dei sistemi HVAC.

Questa metodologia può essere migliorata attraverso l'uso di controllori che si basano sulla previsione delle informazioni, quali ad esempio i controllori MPC "Model Predictive Control". L'utilizzo delle previsioni all'interno di un sistema MPC è stato motivato dai tempi (generalmente elevati) con cui le variabili osservate (per esempio la temperatura interna) rispondono alle variazioni dei parametri del sistema HVAC. Includendo la previsione del carico termico, l'algoritmo MPC può avviare azioni di controllo in anticipo e in modo da garantire il compromesso ottimale tra comfort e consumi energetici. Ad esempio, se si prevede un aumento della temperatura esterna verso il mezzogiorno, il MPC inizierà il raffreddamento della stanza prima di allora, poiché non sarà in grado di soddisfare i vincoli comfort degli occupanti se reagisce troppo tardi.

Il comportamento di molti sistemi dinamici può essere rappresentato dalla

seguinte legge di aggiornamento:

$$x(t + 1) = Ax(t) + Bu(t)$$

dove t è un istante temporale, x è lo stato attuale del sistema, u è il segnale di ingresso e i termini A, B sono costanti.

Il principio fondamentale di un sistema MPC è quello di individuare un segnale di controllo per minimizzare la funzione obiettivo. Il modello del sistema di cui sopra viene quindi utilizzato come vincolo nel problema di ottimizzazione. L'algoritmo MPC è solitamente impostato per calcolare il segnale di controllo ottimale per un certo periodo di tempo (N step in avanti), ma solo il segnale di controllo del primo step viene applicato; al passo successivo, un nuovo problema di ottimizzazione viene risolto e un nuovo segnale di controllo viene generato. Un tipico problema di ottimizzazione MPC si può presentare come:

$$\min_{x \in \mathcal{R}^N, u \in \mathcal{R}^{N-1}} f(x, u)$$

subject to

$$x(k + 1) = Ax(k) + Bu(k), \quad \forall k \in \{t, t + 1, \dots, t + N - 2\}$$

$$x(t) = x_t \text{ (Stato Attuale del Sistema)}$$

$$u(k) \leq u_{ub}, \quad \forall k \in \{t, t + 1, \dots, t + N - 2\}$$

$$u(k) \geq u_{lb}, \quad \forall k \in \{t, t + 1, \dots, t + N - 2\}$$

Altri vincoli possono essere aggiunti per soddisfare le esigenze del sistema di controllo.

L'utilizzo delle previsioni meteorologiche nel controllo del comfort negli edifici è stato studiato in diversi lavori (10) (17), che hanno dimostrato come le strategie di tipo predittivo risultano essere più efficienti rispetto alle strategie convenzionali (non predittive).

Nel presente lavoro si focalizza l'attenzione sull'automazione integrata di una stanza, intendendo il controllo integrato di un sistema di riscaldamento, ventilazione e climatizzazione dell'aria (HVAC), di un sistema di posizionamento delle tende e di un sistema di illuminazione automatico. L'obiettivo del controllo integrato è quello di mantenere la temperatura della stanza, i livelli di CO_2 emessa e l'illuminamento in un intervallo predefinito di comfort (Capitolo 2), il che può essere realizzato tramite un insieme di diversi attuatori. Gli attuatori differiscono in termini di tempi di risposta e di efficacia, nella loro dipendenza dalle condizioni atmosferiche e dai costi energetici risultanti. L'obiettivo è quello di scegliere in modo ottimale le impostazioni dell'attuatore a seconda delle condizioni climatiche future al fine di soddisfare le esigenze di comfort e di minimizzazione dei costi energetici.

A causa della complessa interdipendenza dell'intero sistema HVAC i metodi tradizionali generalmente si basano su ipotesi che non possono realizzarsi nella pratica poiché tali sistemi sono complessi, non lineari e dinamici. È quindi necessario suggerire un metodo che fornisca delle impostazioni ottimali per differenti operazioni in risposta ai carichi dinamici di raffreddamento e alle condizioni atmosferiche variabili. Per gestire tale dinamicità del sistema, la modellazione sulle serie storiche delle variabili è stata realizzata utilizzando i modelli di Box e Jenson, nella particolare versione che permette di includere nel modello l'informazione proveniente da variabili esterne (Capitolo 3). Le variabili esterne in questo caso sono rappresentate dalle variabili climatiche, forti del fatto che diversi studi in letteratura hanno confermato la bontà della modellazione di tali variabili nel breve periodo (20).

I modelli stimati non sono ottimizzabili mediante le tecniche classiche che si basano sull'analisi del gradiente della funzione obiettivo, in quanto siamo in presenza di funzioni molto complesse e le classiche ipotesi non vengono rispettate. Un approccio di simulazione metaeuristico è stato quindi sviluppato, utilizzando

gli algoritmi di programmazione evolutiva (Capitolo 4), che dimostrano essere in grado di gestire efficacemente problemi di tipo discreto, non lineari e problemi di ottimizzazione estremamente vincolati, come ad esempio quelli relativi ai sistemi HVAC (15).

L'applicazione di quanto proposto viene realizzata sulla base dei dati a disposizione, fornendo un possibile approccio risolutivo al problema (Capitolo 5).

Capitolo 2

I SISTEMI PER LA REALIZZAZIONE DEL COMFORT ABITATIVO

Il termine HVAC è l'acronimo inglese di Heating, Ventilation e Air Conditioning, ossia riscaldamento, ventilazione e condizionamento dell'aria. Il contesto HVAC è molto vasto in quanto nel momento in cui si considera un sistema di questo tipo si può avere a che fare con una semplice stufa a legna, utilizzata per il riscaldamento di comodità, così come un sistema estremamente affidabile di condizionamento globale dell'aria utilizzato nei sottomarini e nelle navette spaziali. Considerando le apparecchiature di raffreddamento, esse variano dalla piccola unità domestica fino alle macchine frigorifere utilizzate nei processi industriali.

Il principale obiettivo di un sistema HVAC consiste nel garantire un certo comfort ambientale ai fruitori degli spazi in cui viene realizzato. Tuttavia un sistema efficiente costituisce un fattore determinante alla riduzione dei consumi di energia negli ambienti oltre che ad una riduzione delle emissioni di anidride carbonica. Infatti è statisticamente provato che più del 30% delle emissioni di CO_2 nell'ambiente siano prodotte dagli edifici in cui viviamo(3).

Il **comfort ambientale** è definito come "that condition of mind which expresses satisfaction with the thermal environment and is assessed by subjective evaluation" (1), ossia una condizione mentale che esprime soddisfazione in funzione delle percezioni sensoriali di un individuo nei confronti dell'ambiente che lo

circonda. Poiché stiamo parlando di uno stato d'animo, non esiste un metodo per poterlo misurare; è stato quindi creato un modello numerico delle condizioni di comfort sulla base di dati ottenuti intervistando molte persone in diverse condizioni sperimentali. Considerando un generico individuo sono stati individuati sette fattori che influenzano il comfort termico personale, aggregabili in tre gruppi che sono rappresentati in Figura 2.1.

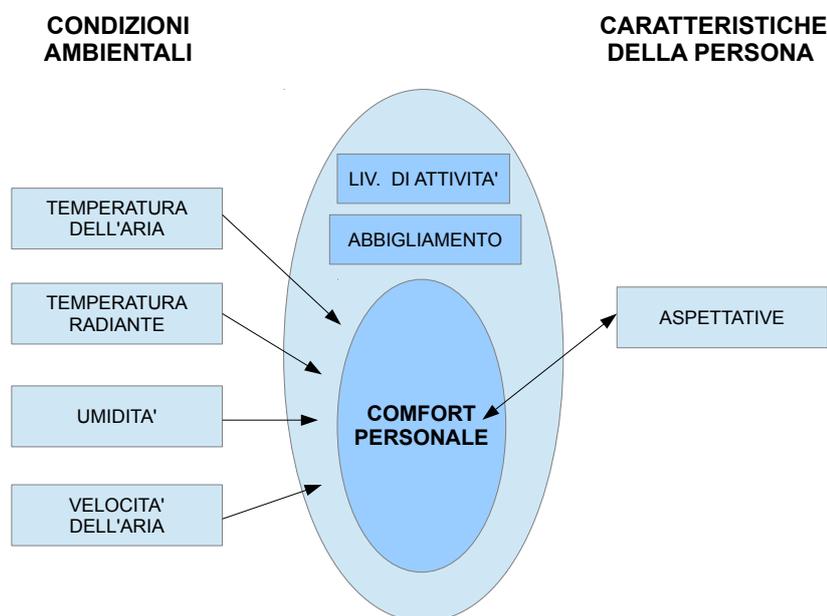


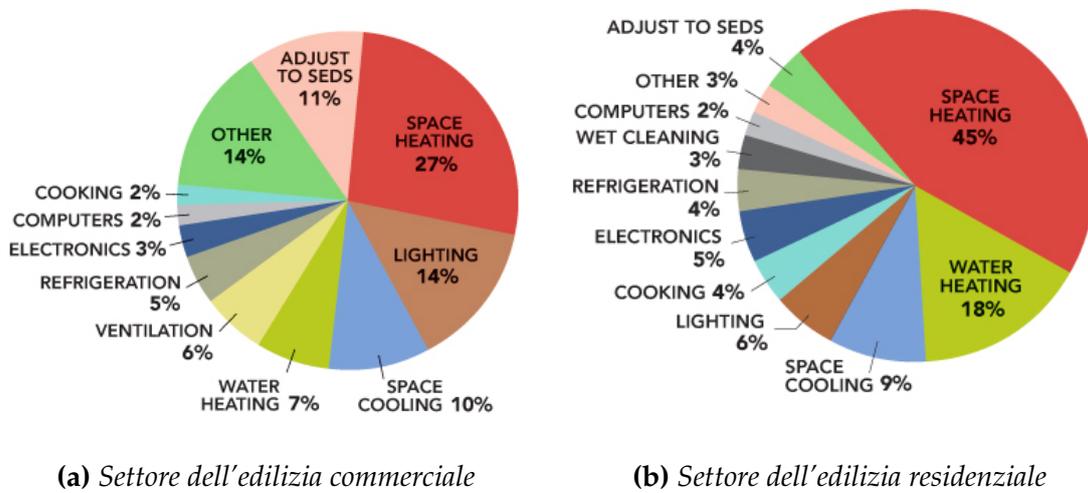
Figura 2.1: I fattori che influenzano il Comfort Termico(14)

Tra i fattori personali troviamo il livello di attività, poiché il corpo umano produce continuamente calore attraverso un processo chiamato metabolismo ed al crescere dell'attività della persona questo calore aumenta (una persona seduta produce meno calore di una persona che corre); anche l'abbigliamento è un fattore importante in quanto agisce come isolante, rallentando la perdita di calore dal corpo; al fine del raggiungimento del comfort bisogna valutare quale potrebbe essere l'abbigliamento degli occupanti e, a causa della grande varietà di materiali,

pesi e trame dei tessuti, si possono considerare solo delle stime approssimative. Nelle caratteristiche della persona rientra anche il fattore aspettativa, poiché influisce sulla percezione di comfort in relazione al tipo di locale in cui si trova (una persona che entra in un prestigioso hotel ha delle aspettative differenti rispetto ad una che entra in un condominio economico).

Infine, ma per questo non meno importanti, ci sono i fattori ambientali. La temperatura dell'aria, ossia la temperatura del locale in cui si trova la persona; la temperatura radiante, cioè quel calore che viene trasmesso da un corpo caldo ad un corpo più freddo senza effetto sulla spazio intermedio (per esempio l'irraggiamento solare); l'umidità che ha degli effetti sull'uomo ma anche all'ambiente stesso, infatti per alcune persone un tasso di umidità troppo basso può causare dei problemi e al contrario livelli troppo alti nell'aria possono causare seri problemi di muffa all'edificio ed al suo contenuto; infine la velocità dell'aria che influisce sulla temperatura percepita dall'uomo, poiché tanto maggiore è la velocità sul corpo di una persona tanto maggiore è l'effetto di raffreddamento percepito; inoltre una velocità troppo elevata dell'aria può dare fastidio.

Il concetto di **risparmio energetico** è strettamente collegato ai sistemi HVAC. Infatti, prendendo in considerazione l'utilizzo di energia negli edifici commerciali negli USA nel 2012(16), è emerso come circa un quarto del consumo totale di energia primaria è dovuto al riscaldamento, mentre illuminazione e raffreddamento impegnano ciascuno circa un settimo del totale. Se consideriamo anche la ventilazione con il suo 6% di energia, l'HVAC diventa nel suo insieme il più grande utilizzatore di energia negli edifici commerciali Figura 2.2. Considerando l'utilizzo di energia negli edifici residenziali emerge come il riscaldamento impegna circa la metà del consumo di energia (Figura 2.2), seguita dal riscaldamento dell'acqua sanitaria (18%) e dal raffreddamento degli ambienti (9%).

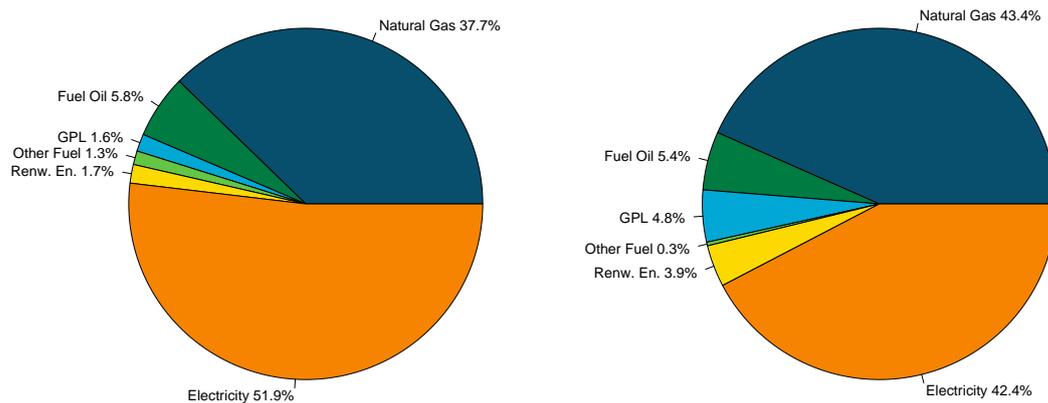


(a) Settore dell'edilizia commerciale

(b) Settore dell'edilizia residenziale

Figura 2.2: Consumo energetico degli edifici ripartito in base all'utilizzo finale (16)

In una situazione di questo tipo, è facile comprendere come l'energia elettrica sia una fonte presente costantemente nei consumi degli edifici commerciali e residenziali. A conferma di ciò uno studio ha evidenziato come nel 2007 nei 27 stati membri dell'Unione Europea i consumi di energia elettrica dei sistemi HVAC sono stati circa 313 TWh, circa l' 11% del totale (2800 TWh) di quell'anno (9).



(a) Settore dell'edilizia commerciale

(b) Settore dell'edilizia residenziale

Figura 2.3: Consumo energetico degli edifici ripartito in base alla fonte energetica (16)

2.1 Cenni storici

Volendo ripercorrere l'evoluzione dei sistemi HVAC si può dire che semplici sistemi sono stati utilizzati fin dai tempi remoti, basti pensare come le persone utilizzavano il fuoco per il riscaldamento. Intorno al 100 dC i romani impiegarono primi impianti di riscaldamento centrale, che garantivano la conduzione di aria riscaldata da forni attraverso tubature sotto i pavimenti o nelle pareti. L'esigenza di una ventilazione adeguata per gli occupanti divenne evidente, fino alla fine del 1880 quando delle regole pratiche per la progettazione dell'aerazione sono stati sviluppate e utilizzate in molti paesi. La rivoluzione industriale è stata il principale precursore all'invenzione ed innovazione dei componenti dei sistemi HVAC. La prima macchina per la refrigerazione è stata brevettata nel 1851 grazie al Dr. John Gorrie e nell'anno seguente Lord Kelvin ha sviluppato il concetto di pompa di calore. Dal 1880 la refrigerazione è divenuta disponibile a fini industriali, inizialmente come congelamento della carne per il trasporto e per la produzione di ghiaccio e successivamente con gli utilizzi che conosciamo tutt'ora. Nel 1900 c'è stata una nuova iniziativa che mirava a garantire il comfort negli edifici: nel 1902 il Dott. Willis Carrier H. (1876-1950) scopre la relazione tra temperatura e umidità e come controllarle; inoltre nello stesso anno viene installato il primo sistema di climatizzazione nel New York Stock Exchange. Da questo momento storico in poi, grazie alle scoperte della scienza e in ingegneria, la tecnologia delle componenti HVAC si è evoluta a dismisura fino ai componenti dei nostri giorni.

2.2 Sistemi e caratteristiche

Finora abbiamo considerato il sistema HVAC come un tutt'uno anche se esso è composto dalle tre componenti fondamentali, ossia la componente che si occupa del riscaldamento, quella della ventilazione ed infine quella della climatizzazione. Una quarta componente, il Controllo, influisce direttamente sulle altre com-

ponenti poiché determina come devono operare per soddisfare gli obiettivi di progettazione precedentemente citati.

Il **riscaldamento** dell'ambiente può essere ottenuto riscaldando l'aria all'interno, attraverso sistemi di alimentazione dell'aria, oppure mediante il riscaldamento diretto degli occupanti tramite la radiazione, attraverso l'irradiazione a pavimento o a parete oppure attraverso pannelli radianti.

La **ventilazione** mantiene "pulita" l'aria che respira nell'ambiente attraverso una corretta miscela di gas, il controllo degli odori e la rimozione di contaminanti, garantendo in questo modo gli occupanti sani e produttivi. La ventilazione può essere realizzata passivamente attraverso la ventilazione naturale, sfruttando le differenze di pressione dell'aria che esistono tra l'interno e l'esterno di un edificio, oppure attivamente attraverso la ventilazione meccanica (sistemi di distribuzione alimentati da ventole).

Con il termine **climatizzazione** facciamo riferimento al raffreddamento sensibile e latente dell'aria. Il raffreddamento sensibile riguarda il controllo della temperatura dell'aria mentre il raffreddamento latente comporta il controllo dell'umidità nell'aria. L'aria di un ambiente viene raffreddata attraverso il trasferimento di calore tra ambienti o respingendo il calore all'esterno tramite dispositivi raffreddati ad aria o ad acqua. Tuttavia l'umidità influisce sulla temperatura percepita dagli occupanti e quindi sul comfort, per cui l'aria va deumidificata/umidificata in base alle caratteristiche climatiche della zona in cui è situato l'edificio. Il termine climatizzazione è progressivamente mutato dal significato proprio di raffreddamento fino a quello di controllo globale di temperatura, umidità nell'aria, fornitura di aria esterna per la ventilazione, filtrazione di particelle e circolazione dell'aria nell'ambiente. D'ora in poi useremo il termine "climatizzazione" per includere tutti questi problemi mentre il termine "HVAC" sarà usato qualora solo alcuni degli elementi appena citati sono controllati.

2.2.1 Sistema di climatizzazione di base

In un sistema di base l'aria viene richiamata dall'ambiente, miscelata con dell'aria proveniente dall'esterno ed infine condizionata prima di essere reimpressa nell'ambiente. Questo sistema viene comandato da un termostato ambiente che, in base alla temperatura target impostata, comanda e controlla la quantità di calore o di raffreddamento fornita dalla serpentina di riscaldamento o raffreddamento. La ventilazione dell'ambiente è garantita dall'aria esterna che viene miscelata con quella del sistema mentre dei filtri garantiscono la fornitura di aria filtrata. Un igrostatò nell'ambiente permette di monitorare anche l'umidità dell'aria ed eventualmente intervenire.

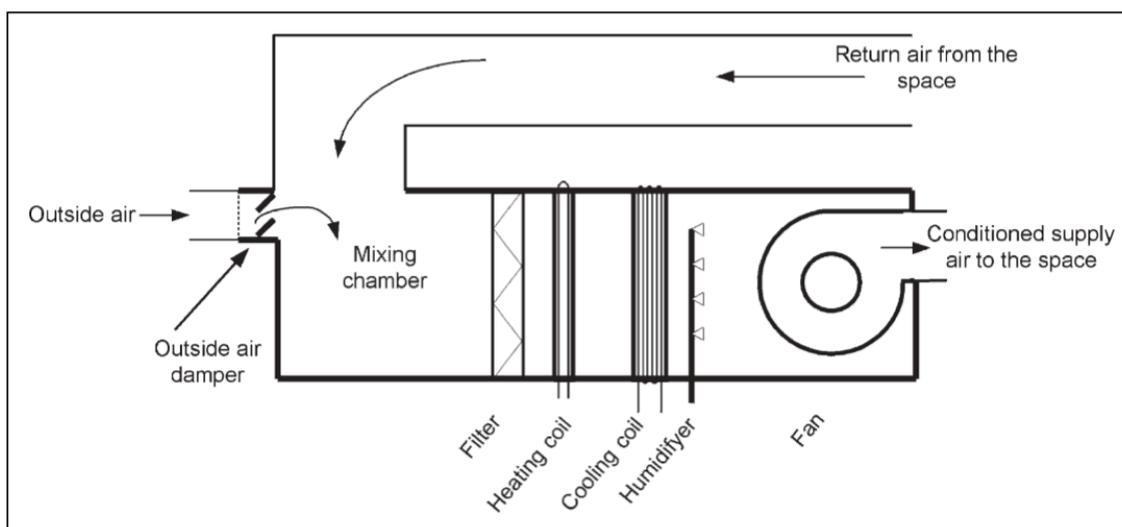


Figura 2.4: Sistema di climatizzazione di base

2.2.2 Sistema di climatizzazione a zone

Il sistema finora considerato fornisce una singola fonte di aria che mantiene una temperatura uniforme per l'intero spazio e viene controllata da un termostato e da un igrostatò. Tuttavia, in molti edifici sono presenti una varietà di spazi con diversi occupanti e diversi carichi termici. Questi carichi variabili possono essere dovuti a diversi tipi di impiego all'interno dei locali oppure dovute al cambia-

mento dei carichi di raffreddamento perché l'irraggiamento solare è presente per alcuni locali e non per altri. Pertanto il nostro sistema semplice, che fornisce una singola fonte di riscaldamento o raffreddamento, deve essere opportunamente modificato per poter fornire indipendentemente una fonte di raffreddamento o riscaldamento variabile per ogni spazio.

Quando un sistema è progettato per fornire il controllo indipendente in diversi spazi, ogni spazio viene definito "zona". Una zona può essere una stanza chiusa così come può anche essere una parte di un grande spazio (per esempio considerando un teatro, il palco teatrale può essere una zona mentre dove siede il pubblico abbiamo una seconda zona).

Questa esigenza di suddivisione in zone ci porta a quattro grandi categorie di sistemi e la considerazione di come ognuno può fornire raffreddamento e riscaldamento a zone.

2.2.2.1 Sistemi totalmente ad aria

I sistemi ad aria forniscono aria condizionata mediante un flusso di aria temperata agli spazi; in questi sistemi è necessaria una canalizzazione dell'aria per ogni zona, e questo può richiedere l'occupazione di un notevole spazio. All'interno dei sistemi ad aria si possono distinguere due tipologie di sistemi.

Il sistema più semplice ma con la minima efficienza energetica è il **sistema di riscaldamento a volume costante**. Supponendo che il sistema dell'aria principale fornisca aria fredda a sufficienza da poter soddisfare tutti i possibili carichi di raffreddamento e che in ogni condotto di ciascuna zona sia presente un radiatore. Il radiatore viene controllato da un termostato per mantenere la zona alla temperatura prefissata (Figura 2.5).

Un altro sistema viene chiamato **sistema a volume d'aria variabile (VAV)** poiché varia il volume di aria fornita a ciascuna zona. Questa tipologia di sistemi è più efficiente dei sistemi di riscaldamento. Anche in questo caso, si supponga che

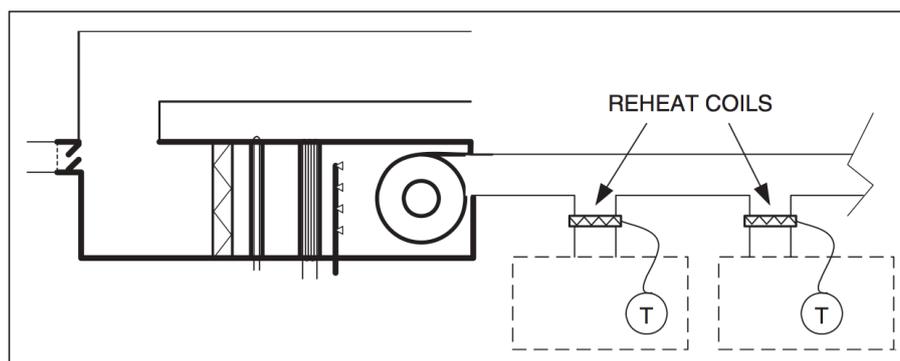


Figura 2.5: Sistema di riscaldamento a volume costante

il sistema di base fornisca aria fresca abbastanza da poter soddisfare tutti i possibili carichi di raffreddamento. Nelle zone che necessitano solo di raffreddamento, il condotto di ciascuna zona può essere dotato di una valvola di controllo (controllata da un termostato) che viene chiusa per ridurre il flusso d'aria e mantenere quindi la temperatura desiderata (Figura 2.6).

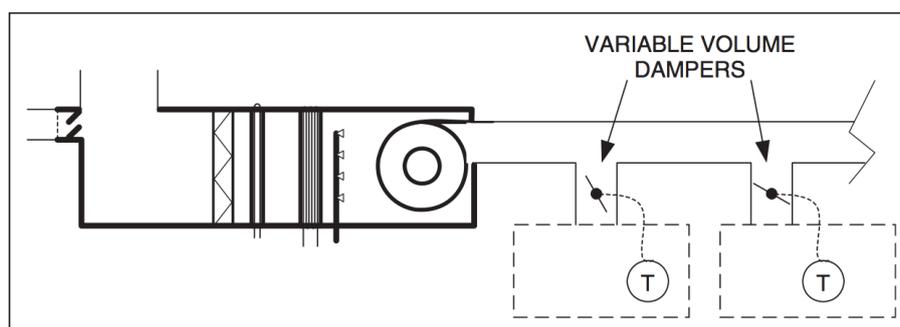


Figura 2.6: Sistema a volume d'aria variabile (VAV)

2.2.2.2 Sistemi aria-acqua

Nei sistemi ad aria e acqua un sistema centrale fornisce tutta la ventilazione primaria dell'aria mentre le unità locali forniscono un ulteriore condizionamento. Il sistema di ventilazione primaria fornisce anche la maggior parte, se non tutto, del controllo dell'umidità attraverso il condizionamento dell'aria di ventilazione. Le unità locali vengono normalmente fornite da un sistema centrale con acqua calda per il riscaldamento e acqua refrigerata per il raffreddamento. Questi sistemi so-

no particolarmente efficaci in spazi perimetrali, dove si verificano elevati carichi di riscaldamento e raffreddamento.

2.2.2.3 Sistemi totalmente ad acqua

I sistemi solo ad acqua sono particolarmente utilizzati qualora la ventilazione è naturale, attraverso l'apertura di finestre o altri mezzi, e non è necessario disporre di un condotto di ventilazione dell'aria. Questo fa sì che tutti i processi della climatizzazione tranne la ventilazione sono forniti dalle unità locali alimentate con acqua calda e/o refrigerata da un impianto centralizzato. Il più grande gruppo di sistemi totalmente ad acqua è costituito dagli impianti di riscaldamento.

2.2.2.4 Sistemi unitari basati su refrigeranti

L'ultima tipologia sono i sistemi unitari basati su refrigeranti, che utilizzano apparecchiature locali di refrigerazione e riscaldamento per fornire la climatizzazione. Il condizionatore a finestra è l'esempio più semplice di questo tipo di sistema. La ventilazione dell'aria può essere introdotta dall'unità, attraverso l'apertura di finestre oppure da un sistema centrale di ventilazione dell'aria. Il nome del sistema deriva dal fatto che il raffreddamento in loco è basato sull'uso di refrigerante.

2.2.3 Scelta del sistema di climatizzazione

Ciascuna delle quattro tipologie di sistemi di climatizzazione ha numerose varianti, quindi la scelta di un sistema non è un compito semplice. Inoltre gli edifici non sono tutti uguali, non solo per tipologia d'uso e dimensioni ma anche e soprattutto per orari di occupazione, criteri di comfort e condizioni ambientali. Per questo, riuscire a trovare la soluzione adatta per ognuna di queste problematiche rappresenta una vera e propria sfida.

I fattori o parametri che influenzano la scelta del sistema possono essere suddivisi in cinque gruppi. Il primo gruppo include i fattori relativi al **disegno del-**

L'**edificio** e sono quelli che hanno un'influenza maggiore sulla scelta del sistema. Troviamo poi i fattori relativi all'**ubicazione dell'edificio**, poiché determina le condizioni meteorologiche che influenzano l'edificio stesso e i suoi occupanti. Per una specifica località i fattori da considerare potrebbero essere le condizioni ambientali, le condizioni estreme di raffreddamento estivo, i valori di umidità in estate, le condizioni di punta di riscaldamento invernale, la velocità del vento, le ore di sole, ecc. Inoltre la posizione dell'edificio determina quali specifiche regolamentazioni nazionali o locali devono essere seguite. Altri fattori rientrano tra i **beni utili** poiché la scelta del sistema può essere fortemente influenzata da beni di utilità già disponibili e dai relativi costi per la fornitura e l'uso. Un'altro gruppo di fattori molto importante è rappresentato dai **requisiti interni** tra cui le esigenze degli occupanti, la produzione di calore dovuta all'illuminazione, l'inquinamento prodotto (per esempio in cucina), la disposizione interna degli spazi e i relativi impieghi. Assieme agli effetti dovuti alla posizione dell'edificio fornisce tutte le informazioni necessarie al calcolo del carico per il sistema. Infine, ma non di minore importanza, troviamo i fattori relativi al **cliente** di cui il principale fattore è il denaro speso. Di conseguenza si deve prendere in considerazione le esigenze del cliente, sia per la costruzione del sistema che per i costi in uso.

Dopo aver esaminato questi fattori va detto che non esiste una procedura ben definita e riconosciuta per la scelta del sistema. Tuttavia è possibile individuare una procedura guidata dalla ragionevolezza e dal buon senso. A proposito, il primo passo consiste nel calcolare i carichi del sistema e stabilire il numero e la dimensione delle zone. La comprensione dei carichi generalmente porta all'eliminazione di alcuni sistemi presi in considerazione. Ad un certo punto si ha la necessità di fare un confronto punto a punto per fare una scelta definitiva. In aiuto di ciò viene una matrice denominata **System Choice Matrix** che consiste in un elenco dei fattori rilevanti che influenzano la scelta del sistema e un metodo tabulare di confronto dei sistemi in esame. Questo metodo consente di valutare

metodicamente delle alternative di sistema.

2.2.4 Standard energetici

La promozione dell'efficienza energetica è uno degli obiettivi principali delle politiche energetiche in quanto migliora la gestione delle risorse e riduce il consumo energetico e il suo impatto ambientale. Ad oggi, la maggior parte dei paesi sviluppati includono una sezione in materia di efficienza energetica all'interno delle politiche di pianificazione energetica, di solito realizzate mediante di una serie di leggi, codici, strategie, regolamenti e sistemi di certificazione.

La normativa energetica degli edifici, nota come codice energetico dell'edificio, emerse nel 1970 (dopo la crisi petrolifera) come uno strumento essenziale per migliorare l'efficienza energetica e minimizzare il consumo di energia negli edifici. Fondamentalmente, lo scopo era quello di fissare i requisiti minimi di efficienza energetica per poter conseguire la progettazione ad alta efficienza energetica negli edifici di nuova costruzione. Ora, la normativa energetica imposta i requisiti di efficienza energetica per la progettazione e la costruzione di nuovi edifici e per l'adeguamento degli edifici (sia estensioni che ristrutturazioni). In generale, mira a ridurre al minimo il consumo finale di energia o di altre grandezze correlate (energia primaria, le emissioni di gas, i costi energetici), senza compromettere il comfort o la produttività degli occupanti.

Per poter perseguire una linea comune, oltre che alle sezioni nazionali, si sono create delle associazioni professionali a livello continentale con l'impegno di incoraggiare lo sviluppo e l'applicazione sia del risparmio energetico che delle fonti di energia rinnovabili, supportando la sostenibilità degli edifici come un mezzo per fornire un ambiente interno sicuro, sano, confortevole e contemporaneamente limitare l'impatto sulle risorse naturali della Terra.

Prendendo in considerazione solo le principali, in Europa troviamo il REHVA, Federation Of European Heating And Air-Conditioning Association mentre in

America è presente l'ASHRAE, American Society of Heating, Refrigerating and Air Conditioning Engineers.

A conferma della linea comune da perseguire, ASHRAE e REHVA durante l'ASHRAE Winter Meeting del 2007 hanno firmato un nuovo protocollo d'intesa chiedendo il rafforzamento della cooperazione tra le due associazioni. Con questo protocollo le associazioni si impegnano a operare insieme per trovare la migliore tecnologia sostenibile per un migliore ambiente interno degli edifici. Inoltre, il coordinamento e la cooperazione sono necessari affinché la duplicazione degli sforzi sia evitata e le risorse limitate vengano utilizzate nel modo più efficiente.

In Europa, l'obiettivo per il risparmio energetico è fissato al 20% entro il 2020¹. ASHRAE sta portando avanti un obiettivo simile, con la speranza di raggiungere un consumo di energia netto pari a zero entro il 2020.

2.3 Il problema dell'ottimizzazione di un edificio ad uso uffici

Il problema oggetto di studio riguarda l'ottimizzazione di un edificio situato a Rovereto adibito ad uffici.

Il caso di studio prevede l'utilizzo di tre stanze differenti. Ogni stanza implementa livelli di controllo differenti dell'impianto HVAC e del sistema di illuminazione. La prima stanza è controllata direttamente dall'uomo e nessun sistema automatico è presente. La seconda stanza prevede il controllo del sistema di riscaldamento e climatizzazione. La terza stanza, quella più automatizzata, prevede la possibilità di controllare il sistema HVAC, l'impianto di illuminazione e l'apertura delle tende attraverso un algoritmo di ottimizzazione. La diversi-

¹Con il COM/2008/0772 def. del 13.11.2008 i leader europei si sono impegnati a ridurre del 20% il consumo di energia primaria rispetto alle previsioni per il 2020

tà di queste tre stanze permetterà di confrontare le performance degli approcci sviluppati. In questo lavoro ci concentreremo sulla terza stanza.

L'obiettivo perseguito è quello di minimizzare il consumo totale di energia, mantenendo la temperatura interna (comfort termico) e l'illuminazione ad un livello desiderabile sulla base della regolazione di quattro punti di controllo impostati: l'apertura delle tende, il comando centrale e due regolatori dell'illuminazione artificiale della stanza.

Questi valori di riferimento dovrebbe essere ottimizzati per ridurre il consumo dell'energia di funzionamento garantendo il comfort abitativo dell'edificio. Il problema di ottimizzazione è formalizzato attraverso la definizione delle variabili del problema, delle funzioni obiettivo e dei vincoli.

2.3.1 Le variabili del problema

Di seguito sono riportate le variabili problema per il sistema HVAC oggetto di studio:

Tipo Variabili	Variabili	Unità di Misura	Codifica
Variabili Controllabili	Dimmer 1	Lux	d_1
	Dimmer 2	Lux	d_2
	Blinds Opening	N°	b
	F.C. command	N°	f_c
Variabili Non Controllabili	Temperatura Interna	°C	v_1
	Umidità	g/m^3	v_2
	Velocità dell'aria	m/s	v_3
	Temp. Centrale Rad. Media	°C	v_4
	Luminosità Ovest	Lux	v_5
	Luminosità Est	Lux	v_6
	CO_2	g	v_7
	Occupazione	Bool	v_8
	Temperatura Esterna	°C	v_9
	Irraggiamento Esterno	W/m^2	v_{10}
	Illuminazione Esterna	Lux	v_{11}
	Umidità Esterna	g/m^3	v_{12}
Variabili di Comfort	PMV	N°	x_1
	DGI	N°	x_2
Variabili di Risposta del Sistema	En. Totale	kw/h	y
	En. Imp. Termico	kw/h	y_{load}

Tabella 2.1: Variabili del problema

Le **variabili controllabili** sono le variabili su cui si può agire per ottimizzare il sistema e permettono di regolare l'intensità dell'illuminazione artificiale, di regolare le tende oppure di agire sull'impianto termico. I livelli che possono assumere le variabili controllabili sono rappresentati nella Tabella 2.2; in totale sono possibili 972 combinazioni di queste variabili.

Attuazione	Variabile	Livelli
Dimmer 1	d_1	da 400 a 600 Lux (passo 25)
Dimmer 2	d_2	da 400 a 600 Lux (passo 25)
Blinds Opening	b	0,1,2
F.C. command	f_c	0,1,2,3

Tabella 2.2: Livelli delle variabili controllabili

Le **variabili non controllabili** sono tutte quelle variabili che sono dovute allo stato del sistema negli istanti precedenti, in quanto il sistema non reagisce istantaneamente alle variazioni ma si riflette anche nei valori rilevati nei periodi successivi. In particolare le variabili temperatura, l'umidità e temperatura centrale media radiante sono influenzate dai valori passati del sistema HVAC. Le altre variabili, quali la luminosità ad est e ovest, la velocità dell'aria e la CO_2 sono influenzate maggiormente dai valori temporalmente vicini. Inoltre, è presente una variabile che riporta la presenza o assenza di individui nella stanza. Infine, variabili che non risultano influenzate per niente dai valori passati del sistema sono quelle relative alle condizioni climatiche esterne.

Le **variabili di comfort** sono costituite da due indici per la valutazione del comfort termico e luminoso della stanza (permettono di sintetizzare la qualità del comfort attraverso un unico valore numerico). Il Predicted Mean Vote (PMV) è il risultato di un'equazione che mette in relazione funzioni dell'abbigliamento della persona (isolamento degli indumenti e rapporto tra superficie coperta e superficie nuda), funzioni di attività della persona (produzione di calore metabolico e produzione metabolica di energia libera), variabili dell'ambiente (temperatura dell'aria, temperatura media radiante, velocità relativa dell'aria e la pressione del

vapore acqueo) e considera inoltre se la pelle è bagnata. Il valore ottimale dell'indice è 0, allontanandosi negativamente aumenta la sensazione del calore percepito fino al valore di -3, allontanandosi positivamente aumenta la sensazione di fresco fino al valore di +3 (freddo) (1).

Il Daylight Glare Index (DGI) o indice di abbagliamento viene calcolato per ciascuna porzione di sorgente vista attraverso la finestra (cielo, ostruzioni, terreno, ...) e mette in relazione l'illuminazione della fonte, l'illuminazione media dello sfondo, la dimensione angolare della fonte in steradiani come percepita dall'occhio e l'angolo solido sotteso dalla fonte, modificato per effetto della posizione dell'osservatore rispetto alla sorgente in steradiani²

Nella Tabella 2.3 sono riportati i valori dell'indice per le varie zone di comfort³.

Zona	Regione	DGI
Discomfort zone	intolerable	>28
	just intolerable	28
	uncomfortable	26
	just uncomfortable	24
Comfort zone	acceptable	22
	just acceptable	20
	noticeable	18
	just perceptible	16

Tabella 2.3: Zone Comfort DGI

Infine, le variabili di risposta del sistema sono quelle che esprimono il consumo di energia cumulato.

2.3.2 Le funzioni obiettivo e i vincoli

I valori di riferimento della strategia di controllo di supervisione sono ottimizzati al fine di ridurre il consumo energetico e migliorare il comfort termico.

La funzione obiettivo del problema è la funzione che quantifica il consumo di energia dell'impianto termico. La modellazione di tale funzione viene realizzata mediante un approccio basato sulle serie storiche con variabili esogene (3);

²Lo steradiano (*sr*) è l'unità di misura del Sistema internazionale per l'angolo solido, il corrispondente tridimensionale del radiante.

³Fonte: www.new-learn.info/packages/clear/visual/people/performance/glare/glare_daylight.html

dato il legame delle variabili rispetto ai valori passati, questo approccio permette di modellare i valori passati considerando inoltre l'informazione proveniente da variabili esterne.

Il problema è inoltre soggetto a dei vincoli, che sono dovuti a restrizioni sul funzionamento del sistema HVAC oppure possono riguardare la capacità di progettazione delle componenti del sistema. Essi sono di tipo duale, in quanto fissano i limiti inferiore e superiore delle seguenti variabili :

- Temperatura interna
- PMV
- Luminosità Ovest
- Luminosità Est
- DGI
- CO₂
- Energia Elettrica Totale

Capitolo 3

METODI E MODELLI PER LO STUDIO DELLE SERIE STORICHE DI DATI HVAC

Molti fenomeni presentano andamenti nel tempo caratterizzati da una certa regolarità o costanza strettamente legate alla posizione occupata dall'osservazione nella sequenza di dati osservati. Fenomeni di questo tipo sono rappresentabili tramite una serie storica x_t , $t = 1, 2, \dots, n$, ossia una successione di dati numerici nella quale ogni dato è associato ad un indice temporale t .

L'ordine degli elementi di una serie storica non è accidentale, come nel caso di un campione casuale di osservazioni, e il riconoscimento di tale ordine è una delle caratteristiche più importanti della serie stessa. Un'altra caratteristica delle serie storiche è la periodicità delle osservazioni: generalmente le serie storiche sono mensili, trimestrali o annuali tuttavia, in alcuni ambiti, si incontrano serie che presentano una maggiore frequenza delle osservazioni (ad esempio serie settimanali, giornaliere oppure orarie).

È lecito presumere che vi sia una sorta di dipendenza tra osservazioni successive e che essa sia legata alla posizione dell'osservazione nella sequenza. Lo studio di tale dipendenza e la conseguente possibilità di sfruttarla per effettuare delle previsioni costituiscono lo scopo primario dell'analisi delle serie storiche.

Di seguito andiamo ad approfondire i principali approcci all'analisi delle serie

storiche: l'analisi classica e quella stocastica (o moderna).

3.1 Alcuni metodi descrittivi per le serie storiche

Il metodo di analisi classica, che risale agli anni 20, è stato il primo approccio di analisi delle serie storiche e costituisce ancora oggi un valido strumento. Esso suppone che la realizzazione del fenomeno studiato si può esprimere come la somma di una parte sistematica (o deterministica) e di una parte casuale (errore). La componente deterministica viene scomposta in tre componenti: il trend, il ciclo e la stagionalità. Tutto ciò che non viene spiegato dalla parte deterministica è ritenuto un residuo dovuto al caso e costituisce la parte casuale.

Il **Trend (T)** è la componente che spiega l'andamento della serie riferito al lungo periodo; questa componente sintetizza la struttura e l'evoluzione sistematica della serie, a prescindere da fattori contingenti o irrilevanti. Per quanto riguarda l'analisi del trend va sottolineato che non esistono tecniche sempre valide per evidenziarlo. Spesso la sola analisi visiva della serie permette di evidenziare la presenza di un trend, altrimenti si può far affidamento a funzioni specifiche; generalmente viene rappresentato con una funzione matematica di tipo semplice.

Il **Ciclo (C)** è la componente che rappresenta le fluttuazioni pluriennali attorno al trend dovute alle espansioni e contrazioni congiunturali dell'attività economica. Generalmente è meno nitido delle altre componenti in quanto spesso si confonde col trend; in questi casi può essere ignorato come elemento autonomo e viene incorporato nel trend.

La **Stagionalità (S)** è costituita da variazioni sistematiche del fenomeno che tendono a ripetersi in modo analogo in periodi successivi (è presente solo per dati con frequenza sub-annuale). Questo fenomeno è legato solo alla periodicità della serie storica (ad esempio, se la serie è mensile avremo 12 coefficienti stagionali relativi ai mesi). L'eliminazione della stagionalità avviene attraverso l'utilizzo di

medie mobili (medie di termini successivi di una serie storica) che consentono di ridurre l'intensità della variazione di un insieme di dati e vengono utilizzate per lisciare la serie.

Esistono tre tipologie di modelli decompositivi, in base alle relazioni che sussistono tra le componenti di una serie storica (7). Il **modello additivo** assume l'ipotesi che le componenti, espresse tutte nella stessa unità di misura della serie, siano indipendenti tra loro, per cui non possono influenzarsi a vicenda; l'espressione analitica del modello è: $x_t = T_t + C_t + S_t + \varepsilon_t$.

Il **modello moltiplicativo** invece assume l'ipotesi che solo il trend sia espresso nella stessa unità di misura di x_t , mentre le altre componenti si configurino come numeri puri e che siano legate al trend da una relazione di proporzionalità (presenza di dipendenza tra le componenti). Tale modello può essere espresso anche in forma additiva ricorrendo alla trasformazione logaritmica. Il modello può essere espresso come: $x_t = T_t C_t S_t \varepsilon_t$.

In alternativa ai due modelli precedenti, si può considerare una via intermedia composta dal **modello misto**. Questo modello è caratterizzato da relazioni tra le componenti sia additive che moltiplicative. Se la componente stagionale è di tipo additivo abbiamo $x_t = (T_t C_t) + S_t + \varepsilon_t$ mentre se è di tipo moltiplicativo abbiamo $x_t = (T_t C_t S_t) + \varepsilon_t$.

3.1.1 Livellamento esponenziale

All'interno dei modelli deterministici rientrano anche tecniche che si basano sul livellamento esponenziale, un modello semplice che viene utilizzato per descrivere l'andamento di una serie storica e che permette di effettuarne delle previsioni a breve termine. Il modello si fonda su una logica di "aggiornamento sequenziale" (*updating*) secondo cui il dato che osserveremo nel prossimo istante è la risultante di una componente generale, funzione del passato, e di un contributo specifico indotto dalla nuova osservazione.

3.1.1.1 Livellamento esponenziale semplice

Questo metodo permette di prevedere il valore del fenomeno un periodo in avanti, al tempo $t + 1$, ponderando le osservazioni in ragione della loro distanza dal tempo corrente, assumendo che le osservazioni più recenti presentino un contenuto informativo più elevato a fini previsivi. Per ottenere questo effetto i coefficienti di ponderazione sono definiti come una progressione geometrica di ragione (1λ) , dove λ è una costante di livellamento compresa tra 0 e 1, $w_j = \lambda(1\lambda)^j$. La previsione si può esprimere come:

$$\hat{y}_{t+1|t} = \lambda y_t + (1 - \lambda)\hat{y}_{t|t-1}$$

Il problema fondamentale sta nella determinazione di λ . Essa può essere effettuata minimizzando la somma dei quadrati degli errori di previsione. Questo metodo si può applicare a serie storiche senza trend e senza stagionalità.

3.1.1.2 Metodo di Holt-Winters

È più generale del precedente, poiché si applica a serie che presentano anche un trend, cioè una tendenza di fondo a crescere o a decrescere e che localmente possiamo assumere lineare. Il valore stimato risulta composta da due elementi: m_t rappresenta il livello del trend al tempo t mentre $b_t = \beta$ rappresenta l'incremento (costante); la generalizzazione del modello mediante le formule di Holt & Winters viene espressa come:

$$\begin{aligned}\hat{y}_{t+1|t} &= m_t + b_t \\ m_t &= \lambda_0 y_t + (1 - \lambda_0)(m_{t-1} + b_{t-1}) \\ b_t &= \lambda_1 (m_t - m_{t-1}) + (1 - \lambda_1)b_{t-1}\end{aligned}$$

con valori iniziali $m_0 = \alpha$ e $b_0 = \beta$. Le due costanti di livellamento, λ_0 e λ_1 , sono comprese tra 0 e 1 e possono essere determinate minimizzando la somma dei quadrati degli errori di previsione. La formalizzazione di questo metodo facendo

dipendere λ_0 e λ_1 da un unico parametro ω è nota in letteratura come livellamento esponenziale doppio.

3.1.1.3 Metodo di Holt-Winters stagionale

Consideriamo ora una serie che, oltre a presentare un trend presenti anche una componente stagionale di periodo s e prendiamo a riferimento il modello di scomposizione moltiplicativo dove g_t è un fattore stagionale che misura l'espansione o la contrazione del fenomeno nelle stagioni dell'anno. La previsione $l = 1, 2, \dots$ periodi in avanti al tempo t sarà:

$$\begin{aligned}\hat{y}_{t+l|t} &= (m_t + b_t l) g_{t+l-s} \\ m_t &= \lambda_0 (y_t / g_t) + (1 - \lambda_0) (m_{t-1} + b_{t-1}) \\ b_t &= \lambda_1 (m_t - m_{t-1}) + (1 - \lambda_1) b_{t-1} \\ g_t &= \lambda_s (y_t / m_t) + (1 - \lambda_s) g_{t-s}\end{aligned}$$

Le tre costanti di livellamento, λ_0, λ_1 e λ_s , sono comprese tra 0 e 1 e possono essere determinate minimizzando la somma dei quadrati degli errori di previsione.

Nel caso il modello sia additivo si avrà:

$$\begin{aligned}\hat{y}_{t+l|t} &= m_t + b_t l + g_{t+l-s} \\ m_t &= \lambda_0 (y_t - g_t) + (1 - \lambda_0) (m_{t-1} + b_{t-1}) \\ b_t &= \lambda_1 (m_t - m_{t-1}) + (1 - \lambda_1) b_{t-1} \\ g_t &= \lambda_s (y_t - m_t) + (1 - \lambda_s) g_{t-s}\end{aligned}$$

3.2 I modelli stocastici di classe ARMA

Prenderemo in considerazione quello che, forse con terminologia inappropriata, è conosciuto come approccio "moderno" delle serie temporali, il cui elemento di differenziazione sta nel considerare la serie x_t come una realizzazione finita di un processo stocastico. Il problema inferenziale consiste nel risalire da x_t al processo

generatore. L'obiettivo finale dell'analisi è quello di comprendere la struttura probabilistica sottostante ai dati e di tentare di riprodurre il processo stocastico che si suppone abbia generato la serie storica. In tale contesto, la modellistica ARIMA, proposta da Box e Jenkins nel 1976, semplifica il problema mediante una restrizione della classe dei processi stocastici.

3.2.1 Elementi essenziali per la definizione dei processi stocastici

Un *processo stocastico*, $\{X_t\}$, può essere definito come una successione ordinata di variabili casuali X_t indicizzate dal parametro t appartenente ad un insieme parametrico τ . Poiché nel seguito ci limiteremo a considerare la classe dei processi stocastici continui a parametro discreto, avremo $\tau = 1, 2, \dots$ e $\{X_t\} = \{X_1, X_2, \dots\}$.

Il processo stocastico è noto se è nota la densità congiunta di ogni evento nello spazio reale a τ dimensioni. Nelle applicazioni si dispone, per ogni t , di una singola realizzazione della v.c. x_t per cui il processo inferenziale presenterebbe complicazioni insuperabili se non venissero imposte due classi di restrizioni sulle caratteristiche del processo: la stazionarietà e l'ergodicità.

Definizione 3.1 (P.S. stazionario in senso forte). *Un processo stocastico è stazionario in senso forte se la distribuzione di probabilità congiunta di $\{X_t, X_{t+1}, \dots, X_{t+h}\}$ è indipendente da t per ogni valore di h . La condizione necessaria e sufficiente perché ciò si verifichi è che tutti i momenti della v.c. multipla $\{X_t, X_{t+1}, \dots, X_{t+h}\}$ siano finiti ed indipendenti dat. La struttura dinamica così ottenuta è invariante nel tempo. Siamo però di fronte a una condizione ideale e irraggiungibile nella maggior parte dei casi reali.*

Definizione 3.2 (P.S. stazionario in senso debole). *Un processo stocastico è stazionario in senso debole (o in covarianza) se i suoi momenti fino al secondo sono finiti ed indipendenti da t e se l'auto-covarianza fra le componenti del processo stocastico è funzione*

solo del lag h :

$$\mathbb{E}[X_t] = \mu$$

$$\mathbb{E}[X_t - \mu]^2 = \gamma(0) < \infty$$

$$\mathbb{E}[(X_t - \mu)(X_{t-k} - \mu)] = \gamma(k) < \infty$$

Per un processo stocastico gaussiano le condizioni di stazionarietà in senso debole garantiscono anche la stazionarietà in senso forte, in quanto la sua funzione densità congiunta dipende esclusivamente dai primi due momenti.

La seconda caratteristica che un processo stocastico deve avere per poter giungere a risultati utili disponendo di un insieme limitato di informazioni è l'ergodicità, termine originario delle scienze fisiche.

Definizione 3.3 (P.S. Ergodico). *Un processo stocastico $\{X_t\}$ è ergodico rispetto ad un parametro se la stima temporale del parametro, ottenuta da una serie storica, converge in media quadratica a quel parametro.*

Un'altra caratterizzazione del processo stocastico deriva dall'invertibilità o meno dello stesso. Infatti un processo X_t viene detto invertibile se può essere espresso come una funzione delle variabili aleatorie precedenti più una componente erratica. L'invertibilità è connessa alla possibilità di formulare previsioni nell'ambito dei processi stocastici, infatti ci permette di esplicitare il processo X_t tramite una funzione delle variabili casuali precedenti l'istante t .

Definizione 3.4 (Teorema di Wold). *Alla classe dei processi stazionari si applica un importante risultato noto come teorema di Wold: esso afferma che ogni p.s. stazionario (in senso debole) può essere scomposto in due processi stocastici mutualmente incorrelati, uno dei quali è deterministico, $c(t)$, mentre l'altro (indeterministico) è una sequenza infinita di variabili causali incorrelate:*

$$X_t = c(t) + \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + \dots$$

con $\sum |\psi_j| < \infty$ e $\mathbb{E}[c_t \varepsilon_{t-j}] = 0 \forall t, j$. Il termine $\varepsilon_t \sim WN$ rappresenta l'errore di previsione uniperiodale ed è anche detto innovazione: $\varepsilon_t = X_t - \mathbb{E}[X_t | X_{t-1}, X_{t-2}, \dots]$.

Definizione 3.5 (White Noise). Il processo stazionario più elementare è costituito da una sequenza di variabili casuali incorrelate a media nulla e varianza costante: esso è denominato white noise (o rumore bianco), e viene indicato con $\varepsilon_t \sim WN(0, \sigma^2)$, dove $\mathbb{E}(\varepsilon_t) = 0$, $\mathbb{E}(\varepsilon_t^2) = \sigma^2$ e $\mathbb{E}(\varepsilon_t \varepsilon_{t-h}) = 0$ per $k \neq 0$.

Non vengono fatte a priori ipotesi sulla distribuzione di $\varepsilon_1, \varepsilon_2, \dots$ ma qualora si supponga che, per ogni t , ε_t sia anche una v.c. Normale, allora si parla di Processo White Noise Gaussiano e, poiché l'incorrelazione di v.c. Normali implica l'indipendenza, un processo WN Gaussiano è a componenti indipendenti.

Uno strumento utile per la caratterizzazione di un processo stazionario nel dominio temporale è rappresentato dalla funzione di autocorrelazione (ACF) che è definita come

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}$$

e gode delle seguenti proprietà *i)* $\rho(0) = 1$, *ii)* $|\rho(h)| < 1$, *iii)* $\rho(h) = \rho(h)$.

Di un processo stocastico si possono calcolare anche i coefficienti di correlazione parziali, che esprimono una misura dell'associazione lineare tra X_t e X_{t-h} "depurata" della correlazione dovuta alle v.c. intermedie $X_{t-1}, \dots, X_{t-h+1}$. Nel caso di p.s. stazionari il calcolo di questi coefficienti risulta alquanto semplice. Il coefficiente di correlazione parziale è dato dal rapporto fra il determinante della matrice $P(h)$, che corrisponde alla matrice di Toeplitz di ordine h , e il determinante della matrice $P_{(h)}^*$, che è la stessa matrice alla cui ultima colonna è stato sostituito il vettore composto dai valori della funzione di autocorrelazione fino al lag h .

$$\pi_h = \frac{\mathbb{E}[X_t X_{t-h} | X_{t-1}, X_{t-2}, \dots, X_{t-h+1}]}{\gamma(0)} = \frac{|P_{(h)}^*|}{|P(h)|} \text{ con } h = 1, 2, \dots$$

Il grafico dei valori $\{\pi_0, \pi_1, \pi_2, \dots, \pi_h\}$ rispetto ad h , viene detto correlogramma parziale, mentre la sequenza π_h è la funzione di autocorrelazione parziale (PACF). Ovviamente, $\pi_0 = 1$ e $\pi_1 = \rho(1)$.

3.2.2 Alcuni modelli stocastici di classe ARMA

Uno strumento molto importante è l'operatore ritardo (Lag o Backward), indicato con L , che ci permette di traslare nel tempo un processo stocastico intervenendo direttamente sull'indice temporale del processo:

$$LX_t = X_{t-1}$$

L'operatore ritardo gode di alcune interessanti proprietà, ad esempio se viene applicato ad una costante, la lascia invariata mentre l'iterazione dell'applicazione dell'operatore viene indicata con la scrittura: $L^n X_t = X_{t-n}$. Inoltre, l'operatore ritardo è un operatore lineare: dati due processi Y_t e X_t e due costanti a, b vale l'uguaglianza $L(aY_t + bX_t) = aLY_t + bLX_t = aY_{t-1} + bX_{t-1}$.

3.2.2.1 I modelli autoregressivi (AR)

Introdotta per la prima volta da Yule nel 1926(13), un processo autoregressivo di ordine p , $AR(p)$, è un processo nel quale il valore della serie al tempo t è linearmente dipendente dai propri p valori passati più uno shock casuale contemporaneo:

$$X_t = m + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + \varepsilon_t$$

con $\varepsilon_t \sim WN(0, \sigma^2)$. Utilizzando l'operatore ritardo L , assumendo $m = 0$ e indicando con $\phi(L) = 1 - \phi_1 L - \phi_2 L^2 - \dots - \phi_p L^p$ il polinomio autoregressivo di ordine p , il processo si può riscrivere sinteticamente come:

$$\phi(L)X_t = \varepsilon_t$$

Un processo autoregressivo di ordine p è stazionario se le p radici del polinomio $\phi(L)$ sono in modulo superiori all'unità mentre è sempre invertibile. Le

caratteristiche di un processo $X_t \sim AR(p)$ stazionario sono:

$$\mathbb{E}(X_t) = \frac{m}{1 - \phi_1 - \dots - \phi_p}$$

$$\gamma(h) = \begin{cases} \phi_1\gamma(h-1) + \dots + \phi_p\gamma(h-p) + \sigma^2 & h = 0 \\ \phi_1\gamma(h-1) + \dots + \phi_p\gamma(h-p) & h > 0 \end{cases}$$

$$\rho(h) = \begin{cases} 1 & h = 0 \\ \phi_1\rho(h-1) + \dots + \phi_p\rho(h-p) & h > 0 \end{cases}$$

In un processo $AR(p)$ stazionario la funzione di autocorrelazione globale decresce al crescere di h con un andamento che dipende dal segno dei parametri (si dice che ha memoria infinita) mentre la funzione di autocorrelazione parziale è identicamente nulla per $h > p$.

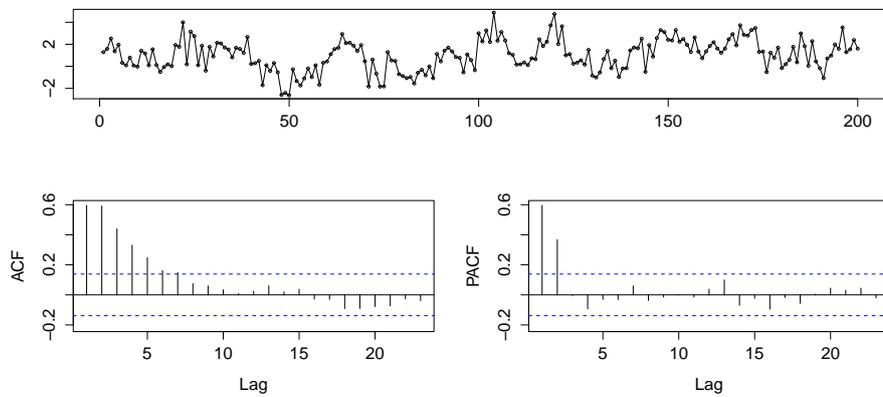


Figura 3.1: Esempio di un processo autoregressivo del secondo ordine

3.2.2.2 I modelli media mobile (MA)

Presentato per la prima volta da Slutsky nel 1937(13), un processo a media mobile di ordine q , $MA(q)$, si ottiene dalla rappresentazione di Wold assumendo $\psi_j = \theta_j$ per $j \leq q$ e $\psi_j = 0$ per $j > q$. Pertanto si può rappresentare come:

$$X_t = \mu + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \dots + \theta_q\varepsilon_{t-q}$$

In un processo di questo tipo, il valore osservato al tempo t è ottenuto come una combinazione lineare dei valori di una componente casuale rilevati nello stesso istante di tempo e in q istanti immediatamente precedenti.

Utilizzando l'operatore ritardo L e indicando con $\theta(L) = (1 + \theta_1 L + \dots + \theta_q L^q)$ il polinomio a media mobile di ordine q , il processo si può esprimere come:

$$X_t = \theta(L) \cdot \varepsilon_t$$

Un processo a media mobile è sempre stazionario poiché media, varianza e covarianza sono costanti e non dipendono dal tempo. Affinché sia invertibile è necessario che le radici dell'equazione $\theta(L) = 0$ siano in modulo maggiori di uno.

Le caratteristiche di un processo $X_t \sim MA(q)$ sono:

$$\mathbb{E}(X_t) = \mu$$

$$\gamma(h) = \begin{cases} \sigma^2 (1 + \sum_{i=1}^q \theta_i^2) & h = 0 \\ \sigma^2 (\theta_h + \theta_1 \theta_{h+1} + \dots + \theta_{q-h} \theta_q) & h = 1, \dots, q \\ 0 & h > q \end{cases}$$

$$\rho(h) = \begin{cases} 1 & h = 0 \\ \frac{(\theta_h + \theta_1 \theta_{h+1} + \dots + \theta_{q-h} \theta_q)}{1 + \theta_1^2 + \dots + \theta_q^2} & h = 1, \dots, q \\ 0 & h > q \end{cases}$$

In un processo a media mobile di ordine p la funzione di autocorrelazione globale si interrompe bruscamente per $h > q$ mentre la funzione di autocorrelazione parziale non si annulla mai e tende a 0 al crescere di h .

3.2.2.3 I modelli ARMA

Il processo AutoRegressivo a Media Mobile di ordine p, q , che è stato combinato da Wold nel 1938(13), si può pensare come una generalizzazione di un p.s. $AR(p)$ con innovazioni che seguono un processo $MA(q)$, ovvero come un

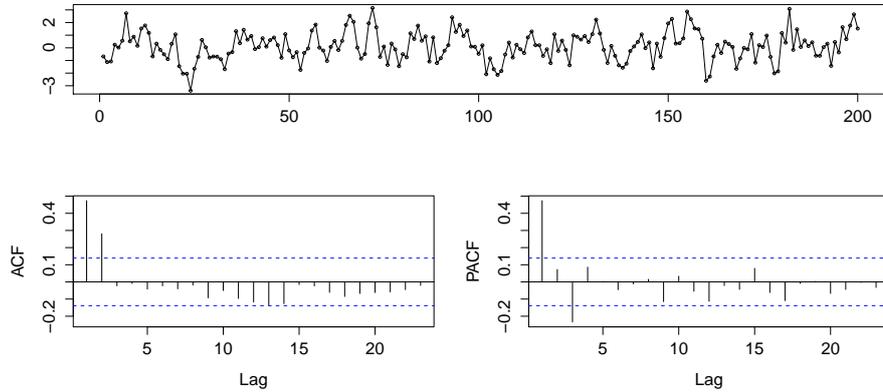


Figura 3.2: Esempio di un processo a media mobile del secondo ordine

processo $MA(q)$ che dipende ulteriormente dai suoi p valori passati:

$$X_t = m + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \dots + \theta_q \varepsilon_{t-q}$$

Applicando l'operatore ritardo e utilizzando il polinomio autoregressivo ed a media mobile si può rappresentare come :

$$\phi(L)X_t = m + \theta(L)\varepsilon_t$$

. Le condizioni sotto le quali il processo è stazionario sono le stesse per le quali il processo autoregressivo è stazionario, vale a dire le p radici del polinomio $\phi(L)$ devono essere esterne al cerchio di raggio unitario. Allo stesso modo, il processo è invertibile se le q radici del polinomio $\theta(L)$ sono esterne al cerchio di raggio unitario.

Le caratteristiche di un processo $X_t \sim ARMA(p, q)$ sono:

$$\gamma(h) = \begin{cases} \phi_1 \gamma_{h-1} + \dots + \phi_p \gamma_{h-p} + \mathbb{E}[X_{t-h} \varepsilon_t] - \theta_1 \mathbb{E}[X_{t-h} \varepsilon_{t-1}] - \dots - \theta_q \mathbb{E}[X_{t-h} \varepsilon_{t-q}] & h = 0, 1, \dots, q \\ \phi_1 \gamma_{h-1} + \dots + \phi_p \gamma_{h-p} & h > q \end{cases}$$

$$\rho(h) = \begin{cases} \frac{\phi_h}{\phi_0} & h = 0, 1, \dots, q \\ \phi_1 \rho_{h-1} + \dots + \phi_p \rho_{h-p} & h > q \end{cases}$$

La funzioni di autocorrelazione globale e parziale presentano un comportamento che rappresenta una mistura di quelli che caratterizzano processi puramente AR e MA e non si annullano mai; si può inoltre mostrare che l'ACF tende a 0 a partire dal lag q , mentre la PACF a partire dal lag p .

3.2.2.4 I modelli ARIMA

Finora sono stati considerati solo processi stazionari. Tuttavia, parte delle serie storiche presentano un andamento di fondo che rende la media non costante per tutto l'intervallo di osservazione. Per ovviare a questo, è stata proposta un'estensione dei processi ARMA, che possono essere resi stazionari attraverso opportune trasformazioni. Un processo ARMA non è stazionario quando le radici dell'equazione caratteristica $\phi(L)X_t = 0$ non sono tutte esterne al cerchio di raggio unitario. Indicando con $\varphi(L)$ il polinomio autoregressivo contenente radici unitarie, questo può essere fattorizzato in

$$\varphi(L) = \phi(L)(1 - L)^d = \phi(L)\Delta^d$$

dove d è la molteplicità della radice unitaria e $\phi(L)$ è il polinomio autoregressivo stazionario.

Sulla base di quanto appena detto, il processo non stazionario autoregressivo integrato a media mobile, *ARIMA*(p, d, q) si può rappresentare come:

$$\phi(L)\Delta^d X_t = m + \theta(L)\varepsilon_t$$

3.2.2.5 I modelli ARIMA stagionali

La classe dei processi ARIMA può essere ampliata anche a processi stagionali, ossia che presentano la ripetizione di un comportamento simile ogni s ritardi inferiori all'anno (s indica la lunghezza del periodo stagionale). Qualora la componente stagionale che descrive l'evoluzione del fenomeno ai ritardi $t, t-s, t-2s, \dots$ sia di natura stocastica e correlata con la componente regolare, essa può essere descritta da un processo di tipo ARIMA stagionale che interagisce con il processo ARIMA non stagionale. I processi stagionali, oltre all'usuale autocorrelazione ai ritardi $t, t-1, t-2, \dots, t-h$, presentano correlazione ai ritardi stagionali $t, t-s, t-2s, \dots, t-hs$. La correlazione tra X_t e X_{t-j} può essere spiegata

da un modello ARIMA(p,d,q) del tipo

$$\phi_p(L)(1 - L)^d X_t = \theta_q(L)b_t$$

dove b_t non è un processo white noise in quanto contiene la correlazione non spiegata dovuta alla stagionalità. Tale correlazione può essere modellata attraverso un processo $ARIMA(P, D, Q)_s$ del tipo

$$\Phi_P(L^s)(1 - L^s)^D b_t = \Theta_Q(L^s)\varepsilon_t$$

dove $\varepsilon_t \sim WN(0, \sigma^2)$ e $1 - L^s)^D$ sono le differenze stagionali di ordine D necessarie a rendere il processo stagionale stazionario in media. Il processo ARIMA stagionale di frequenza s , $ARIMA(p, d, q)(P, D, Q)_s$ si può rappresentare come

$$\Phi_P(L^s)\phi_p(L)(1 - L^s)^D(1 - L^d)X_t = \theta_q(L)\Theta_Q(L^s)\varepsilon_t$$

dove $\varepsilon_t \sim WN(0, \sigma^2)$, il polinomio autoregressivo stagionale di ordine P è $\Phi_P(L^s) = 1 - \Phi_1(L^s) - \dots - \Phi_P(L^{Ps})$ e il polinomio media mobile stagionale di ordine Q è $\Theta_Q(L^s) = 1 - \Theta_1(L^s) - \dots - \Theta_Q(L^{Qs})$

I modelli $ARIMA(p, d, q)(P, D, Q)_s$ presuppongono l'esistenza di una relazione moltiplicativa, ossia di dipendenza, tra la componente regolare e la componente stagionale. Tale relazione assicura che le condizioni di stazionarietà e invertibilità dei processi possano essere valutate separatamente per le componenti regolare e stagionale.

3.2.2.6 I modelli ARMA con variabili esogene

Il modello ARMAX è un raffinamento di un ARMA in quanto riduce la varianza dell'ARMA con l'introduzione nel modello di una o più variabili esogene (o variabili in input). Questo accade per il motivo che le variabili esterne contengono informazioni aggiuntive rispetto al "passato" della serie in grado di contribuire nella spiegazione della variabilità del modello. Modelli di questo tipo sono stati identificati con diversi nomi quali modelli con funzioni di trasferimento, modelli

di regressione con errori di tipo ARMA oppure modelli con regressione dinamica (19).

Al fine di una corretta descrizione del modello, introduciamo il concetto di **funzione di trasferimento** (TFM) che costituisce un modello statistico che descrive la relazione tra una variabile in uscita Y ed una o più variabili in ingresso X .

Nella maggior parte delle applicazioni, per descrivere la relazione viene utilizzata un'equazione lineare che permette di ottenere un modello in funzione del tempo. Per modello lineare si intende che il rapporto tra X e Y è lineare e X e Y sono processi lineari(24). La dipendenza dinamica di Y_t rispetto ai valori attuali e passati di X si può esprimere come:

$$\begin{aligned} Y_t &= v_0 X_t + v_1 X_{t-1} + v_2 X_{t-2} + \dots, \\ &= v(L) X_t \end{aligned}$$

dove v_0, v_1, \dots sono costanti che descrivono l'impatto di X_{t-j} su Y_t (a volte vengono anche detti funzione di risposta all'impulso del sistema) e $v(L) = v_0 + v_1 B + v_2 B^2 + \dots$.

Una modellazione più interessante si ha quando si applica il concetto della modellazione ARIMA per descrivere la funzione $v(L)$ (21). Si assume che $v(L)$ sia un polinomio razionale in L come:

$$Y_t = \frac{\omega(L) L^b}{\delta(L)} X_t$$

dove b è un intero non negativo e rappresenta il tempo di ritardo (o tempo morto) del sistema; $\omega(L) = \omega_0 + \omega_1 B + \omega_2 B^2 + \dots + \omega_s B^s$ e $\delta(L) = 1 - \delta_1 B - \delta_2 B^2 - \dots - \delta_r B^r$ sono i polinomi di ordine finito in L e $\omega_0 \neq 0$.

Possiamo quindi introdurre il modello ARMAX come:

$$W_t = \mu + \sum_i \frac{\omega_i(L)}{\delta_i(L)} L^{ki} X_{i,t} + \frac{\theta(L)}{\phi(L)} \varepsilon_t$$

dove W_t rappresenta la variabile risposta oppure una sua differenziazione; $X_{i,t}$ costituisce l' i -esima serie storica in input oppure la differenziazione dell' i -esima serie storica in input al tempo t (per renderle stazionaria se non lo fosse); k_i è il tempo di ritardo dovuto all'effetto dell' i -esima serie in input mentre $\omega_i(L)$ e $\delta_i(L)$ sono il numeratore e denominatore della funzione di trasferimento per l' i -esima serie in input.

Se esprimiamo i pesi della funzione di trasferimento per l' i -esima serie come $\Psi_i(L) = (\omega_i(L)/\delta_i(L))L^{k_i}$ il modello si può riscrivere in modo più compatto come:

$$W_t = \mu + \sum_i \Psi_i(L)X_{i,t} + \frac{\theta(L)}{\phi(L)}\varepsilon_t$$

3.2.3 La costruzione del modello secondo l'approccio di Box e Jenkins

Box e Jenkins hanno proposto una metodologia per l'adattamento di un modello ARIMA alla serie temporale, X_t , che consiste nell'iterare le fasi di Identificazione del Modello, Stima e Controllo Diagnostico.

3.2.3.1 Identificazione del Modello

La fase di identificazione mira in primo luogo a determinare la trasformazione della serie che induce la stazionarietà attraverso l'analisi grafica della serie; si procede dunque all'identificazione di eventuali valori anomali ed alla ricerca delle trasformazioni più adeguate a rendere stazionaria la serie (calcolo delle differenze e uso della trasformazione Box-Cox). Successivamente si passa a selezionare il modello ARMA (gli ordini p e q ed eventualmente quelli stagionali P e Q) sulla base delle funzioni di autocorrelazione globale e parziale della serie trasformata. I correlogrammi forniscono lo strumento più importante di identificazione: dal confronto dei correlogrammi stimati con quelli teorici che caratterizzano i p.s. al variare degli ordini p e q si ottengono indicazioni circa i medesimi.

3.2.3.2 Stima del Modello

Fino ad ora abbiamo considerato che il processo stocastico che sovrapponiamo ai dati per interpretarli fosse governato da parametri noti. Se questi ultimi noti non sono (e nella realtà non lo sono mai), si possono utilizzare delle loro stime. La tecnica di base per la stima dei parametri di un processo ARMA è il metodo della massima verosimiglianza¹. Di solito si assume che il processo sia normale, in modo che la forma della funzione di densità delle osservazioni sia nota e trattabile. Quando osserviamo una realizzazione di un processo stocastico (o, per meglio dire, una serie storica) x_1, \dots, x_T , la funzione di verosimiglianza non è altro che la funzione di densità congiunta della parte di processo osservata, ossia la funzione di densità marginale del vettore aleatorio (x_1, \dots, x_T) , calcolata nei valori osservati. Nel caso di un processo ARMA del tipo

$$\phi(L)x_t = \mu + \theta(L)\varepsilon_t$$

essa dipenderà dal vettore di parametri $\psi = \{\mu; \phi_1, \dots, \phi_p; \theta_1, \dots, \theta_q; \sigma^2\}$. Se supponiamo (come generalmente si fa) che il processo sia gaussiano, la funzione di verosimiglianza non è che la funzione di densità di una normale multi-variata:

$$\mathcal{L}(\psi) = f(x; \psi) = (2\pi)^{-\frac{T}{2}} |\Sigma|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2}(x - k)' \Sigma^{-1}(x - k) \right\}$$

dove x è il vettore (x_1, \dots, x_T) delle T osservazioni; k e Σ sono i suoi momenti primi e secondi, che dipendono da ψ . È possibile dimostrare che gli stimatori di massima verosimiglianza di processi ARMA gaussiani sono consistenti, asintoticamente normali ed asintoticamente efficienti. Inoltre, sotto condizioni piuttosto blande, le proprietà di consistenza e normalità asintotica vengono conservate anche quando la vera distribuzione del processo non sia normale (si parla in questo caso di stime di quasi-massima verosimiglianza).

¹La verosimiglianza è la funzione di densità del campione, calcolata nel punto corrispondente al campione osservato. Essa dipenderà da un vettore ψ di parametri incogniti, che ne determinano la forma. Per questo si scrive $\mathcal{L}(\psi)$. Massimizzando questa funzione rispetto a y si ottiene la stima di massima verosimiglianza.

3.2.3.3 Controllo Diagnostico

In questa fase del lavoro si valuta l'adeguatezza del modello stimato attraverso l'applicazione di test di significatività sui parametri e l'analisi dei residui per verificare se sono una realizzazione campionaria di un processo white noise a componenti gaussiane. Si tratta di sottoporre a verifica tre proprietà: normalità, omoschedasticità e incorrelazione. Qualora le assunzioni fatte in fase di definizione del modello non siano soddisfatte dai residui, è necessario ripartire dalla prima fase per adattare un modello appropriato ai dati.

3.2.4 Previsione

Data la disponibilità della serie temporale fino al tempo T , ci proponiamo ora di prevedere il fenomeno l periodi in avanti. Il set informativo a cui facciamo riferimento è denotato da $\mathcal{I}_T = \{x_1, x_2, \dots, x_{T-1}, x_T\}$.

Sfrutteremo il risultato fondamentale secondo il quale il previsore ottimale di x_{T+l} è il suo valore atteso condizionato all'informazione \mathcal{I}_T :

$$\hat{x}_{T+l|T} = \mathbb{E} [x_{T+l} | \mathcal{I}_{T+l|T}]$$

L'ottimalità va intesa nel senso che \hat{x}_{T+l} minimizza l'errore quadratico medio di previsione. Inoltre, l'errore di previsione, $x_{T+l} - \hat{x}_{T+l|T}$, ha valore atteso nullo e varianza che denotiamo $Var(\hat{x}_{T+l|T})$.

La funzione di previsione verrà ottenuta sotto le seguenti assunzioni:

1. $\varepsilon_t \sim NID(0, \sigma^2)$
2. I parametri $\phi_1, \phi_2, \dots, \phi_p, \theta_1, \theta_2, \dots, \theta_q, \mu, \sigma^2$ sono noti
3. Si dispone di una realizzazione infinita: $\mathcal{I}_T = \{x_T, x_{T-1}, \dots, x_1, x_0, x_{-1}, \dots\}$

L'assunzione 1 è più forte di $\varepsilon_t \sim WN(0, \sigma^2)$ poiché postula l'indipendenza (oltre che l'incorrelazione) tra le v.c. ε_t ; la seconda implica che i parametri sono

stimati senza errore mentre la terza assunzione assume rilievo quando si ha a che fare con processi MA o misti ed implica che le realizzazioni passate e corrente della v.c. ε_t siano note.

La previsione da modelli ARIMA viene effettuata applicando le seguenti regole generali, che discendono direttamente dalle assunzioni precedenti:

$$\hat{x}_{T+j|T} = x_{T+j} \quad \text{per } j \leq 0$$
$$\hat{\varepsilon}_{T+j|T} = \begin{cases} \varepsilon_{t+j} & \text{per } j \leq 0 \\ 0 & \text{per } j > 0 \end{cases}$$

Infatti, quando $j > 0$, il miglior previsore lineare non distorto dei valori futuri di ε_t è la media incondizionata, $\hat{\varepsilon}_{T+j|T} = \mathbb{E}[\varepsilon_{T+j} | \mathcal{I}_T] = 0$.

Capitolo 4

IL PROBLEMA DELL'OTTIMIZZAZIONE DI SISTEMI ENERGETICI

In questo capitolo verranno introdotti i concetti e le definizioni fondamentali dell'ottimizzazione (o programmazione matematica), la branca della matematica applicata che studia teoria e metodi per la ricerca dei punti di massimo e minimo di una funzione matematica.

Problemi d'ottimizzazione sorgono in tutte le aree della scienza, dell'ingegneria e del business; ci si può ricondurre ad un problema di ottimizzazione ogni qualvolta si vuole formalizzare matematicamente un problema di decisione ed individuare per esso una soluzione ottima o subottima.

In pratica, a partire da un'analisi del problema si estrapolano le grandezze di interesse, che sono quindi associate ad opportune variabili. Queste variabili costituiscono le incognite del problema. Occorre poi esprimere quantitativamente gli eventuali legami esistenti tra le variabili e le limitazioni (derivanti da considerazioni di carattere matematico, fisico o economico) da imporre sulle variabili. Tali legami e limitazioni definiscono i vincoli del problema; l'insieme dei valori delle variabili per cui i vincoli sono soddisfatti costituirà la regione delle soluzioni ammissibili (o dominio dei vincoli). Infine, ma non meno importante, c'è la rappresentazione matematica del problema attraverso la definizione formale del-

la funzione obiettivo (o di costo o di guadagno in base al contesto) che si intende minimizzare o massimizzare.

Bisogna sottolineare come questo passaggio ricopre un ruolo importante, in quanto uno stesso problema reale può essere tradotto in una funzione matematica in diversi modi, ognuno dei quali può presentare in fase risolutiva aspetti positivi o negativi a seconda dell'approccio impiegato per la soluzione dello stesso.

Un problema di ottimizzazione quindi consiste nell'individuare il minimo di una funzione $f(x)$, dove $f : X \subseteq \mathbb{R}^n \rightarrow \mathbb{R}$. Il problema viene sinteticamente rappresentato come:

$$\min_{\mathbf{x} \in X} f(\mathbf{x}). \quad (4.1)$$

4.1 Un'introduzione ai criteri di ottimalità

Per una migliore comprensione di quanto trattato di seguito, verranno introdotti ora alcuni concetti fondamentali.

Si dice **vincolo** una qualunque condizione imposta ad un sistema materiale che impedisce di assumere una generica posizione e/o atto di moto. La presenza di un vincolo di posizione impedisce al sistema materiale di assumere certe configurazioni, e quindi di compiere certi spostamenti.

Un punto $\mathbf{x} = (x_1, \dots, x_n)^T$ che soddisfi tutti i vincoli è detto punto ammissibile e l'insieme di tutti i punti ammissibili (X) generalmente è definito da un set di equazioni e disequazioni, ed è chiamato regione ammissibile.

Nella realtà i punti di ottimo ottenuti possono essere distinti tra punti di ottimo locale e di ottimo globale. I punti di ottimo locale sono quei punti sul dominio della funzione obiettivo che hanno la peculiarità di essere migliori di tutti i punti a loro vicini.

Definizione 4.1 (Minimo Locale). *Dato l'insieme $X \subseteq \mathbb{R}^n$ e la funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$*

il punto $x_0 \in X$ è un punto di minimo locale (o relativo) della $f(x)$ su X , se esiste una sfera aperta di centro x_0 e raggio $\rho > 0$, $S(x_0, \rho) = \{x \in X : \|x - x_0\| \leq \rho\}$, tale che:

$$f(x_0) < f(x), \forall x \in S(x_0, \rho) \cap X.$$

Se la disuguaglianza precedente è verificata in senso stretto, per ogni $x \in X$, $x \neq x_0$, si dirà che il punto x_0 è un punto di **minimo locale stretto** della $f(x)$ su X .

In maniera analoga è possibile definire punti di massimo locale e massimo locale stretto semplicemente invertendo il segno della disuguaglianza nella definizione precedente.

Definizione 4.2 (Massimo Locale). Dato l'insieme $X \in \mathbb{R}^n$ e la funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ il punto $x_0 \in X$ è un punto di massimo locale (o relativo) della $f(x)$ su X , se esiste una sfera aperta di centro x_0 e raggio $\rho > 0$, $S(x_0, \rho) = \{x \in X : \|x - x_0\| < \rho\}$, tale che:

$$f(x_0) \geq f(x), \forall x \in S(x_0, \rho) \cap X.$$

Se la disuguaglianza precedente è verificata in senso stretto, per ogni $x \in X$, $x \neq x_0$, si dirà che il punto x_0 è un punto di **massimo locale stretto** della $f(x)$ su X .

Di seguito verranno introdotte le definizioni relative ai punti di minimo e massimo globale per la funzione $f(x)$. Con il termine punto di ottimo (massimo o minimo) globale si fa riferimento al punto migliore fra tutti quelli del dominio della funzione.

Definizione 4.3 (Minimo Globale). Dato l'insieme $X \in \mathbb{R}^n$ e la funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ il punto $x_0 \in X$ è un punto di **minimo globale** della $f(x)$ su X , se:

$$f(x_0) \leq f(x), \forall x \in X.$$

Se la disuguaglianza precedente è verificata in senso stretto, per ogni $x \in X$, $x \neq x_0$, si dirà che il punto x_0 è un punto di **minimo globale unico** della $f(x)$ su X .

Definizione 4.4 (Massimo Globale). Dato l'insieme $X \in \mathbb{R}^n$ e la funzione $f : \mathbb{R}^n \rightarrow \mathbb{R}$ il punto $x_0 \in X$ è un punto di **massimo globale** della $f(x)$ su X , se:

$$f(x_0) \geq f(x), \forall x \in X.$$

Se la disuguaglianza precedente è verificata in senso stretto, per ogni $x \in X, x \neq x_0$, si dirà che il punto x_0 è un punto di **massimo globale unico** della $f(x)$ su X .

Dopo aver considerato i diversi punti di ottimo che si possono incontrare, nel proseguo si farà riferimento ai soli problemi di minimizzazione in quanto (per convenzione) un problema di massimizzazione può essere convertito in uno di minimizzazione semplicemente negando la funzione obiettivo:

$$\min_{\mathbf{x} \in X} f(\mathbf{x}) \Leftrightarrow \max_{\mathbf{x} \in X} -f(\mathbf{x}). \quad (4.2)$$

In un problema di minimizzazione, la funzione da ottimizzare può assumere diverse forme, più o meno complesse. Una funzione da ottimizzare è detta *funzione uni-modale* qualora presenti un solo punto di ottimo all'interno della regione ammissibile. Nel caso siano presenti più punti di ottimo viene detta *funzione multi-modale*.

La maggior parte delle funzioni che si incontrano nella realtà sono *multi-modali*; presentano infatti un numero più o meno elevato di minimi locali, numero che generalmente dipende dal numero di dimensioni del problema e che tende ad aumentare con esso.

4.2 Classificazione dei problemi di ottimizzazione

I problemi di ottimizzazione (o Programmazione Matematica) possono essere classificati a seconda delle proprietà delle funzioni coinvolte (funzione obiettivo e funzioni che definiscono i vincoli) e a seconda delle variabili coinvolte.

Considerando le caratteristiche delle variabili, i problemi di ottimizzazione possono essere suddivisi in **problemi di ottimizzazione discreta** in cui le varia-

bili sono vincolate ad essere numeri interi, **problemi di ottimizzazione continua** in cui l'insieme ammissibile è un sottoinsieme di \mathbb{R}^n e **problemi misti** in cui solo parte delle variabili sono soggette a vincoli di interezza.

4.2.1 Problemi di ottimizzazione discreta

Gli usi più tipici delle variabili intere comprendono, in particolare, la rappresentazione di quantità indivisibili, la condizione di appartenenza a un insieme finito, l'espressione di alternative tra un numero finito di possibili decisioni, la rappresentazione di vincoli di tipo logico, la rappresentazione di strutture combinatorie quali i grafi e le reti (molti problemi di ottimizzazione su grafi rientrano in questa categoria di problemi).

Tra i problemi di ottimizzazione discreta possiamo considerare **problemi di ottimizzazione a numeri interi**, in cui l'insieme ammissibile è un sottoinsieme dello spazio Z^n delle n -uple di numeri interi e **problemi di ottimizzazione combinatoria o Booleana**, in cui l'insieme ammissibile è un sottoinsieme di $\{0, 1\}^n$, ossia dell'insieme dei vettori a n componenti, i cui valori possono essere esclusivamente 0 oppure 1.

4.2.2 Problemi di ottimizzazione continua

I problemi di tipo continuo definiti su \mathbb{R}^n possono essere classificati in base alla struttura dell'insieme ammissibile ed alle ipotesi sulla funzione obiettivo f .

Se l'insieme ammissibile corrisponde all'intero spazio \mathbb{R}^n il problema rientra nella categoria dei *problemi di ottimizzazione non vincolata*

$$\min f(x), \quad x \in \mathbb{R}^n.$$

In tal caso f è necessariamente una funzione non lineare da \mathbb{R}^n in \mathbb{R} ; è facile verificare, infatti, che se f fosse lineare il problema non ammetterebbe soluzione.

Tra i problemi vincolati la classe più comune di problemi è quella in cui l'insieme ammissibile S è descritto attraverso un insieme finito di vincoli di egua-

gianza e disequaglianza:

$$S = \{x \in \mathbb{R}^n : l(x) \leq 0, h(x) = 0\}$$

in cui $l : \mathbb{R}^n \rightarrow \mathbb{R}^m$ e $h : \mathbb{R}^n \rightarrow \mathbb{R}^p$ sono vettori di funzioni assegnate.

Il problema di ottimizzazione vincolata scritto in forma standard può essere espresso con notazioni vettoriali come:

$$\begin{aligned} \min \quad & f(x) \\ & l(x) \leq 0, \\ & h(x) = 0, \\ & x \in \mathbb{R}^n \end{aligned} \tag{4.3}$$

Facendo riferimento alla proprietà di linearità della funzione obiettivo e dei vincoli possiamo distinguere *problemi di Programmazione Lineare* (PL), in cui l'obiettivo è una funzione lineare e i vincoli sono espressi da un sistema di equazioni e disequazioni lineari, *problemi di Programmazione Quadratica* (PQ), in cui l'obiettivo è una funzione quadratica e i vincoli sono lineari e *problemi di Programmazione Non Lineare* (PNL), in cui l'obiettivo oppure i vincoli non sono tutti lineari.

I problemi di Programmazione Non Lineare corrispondono alla classificazione più generale e comprendono anche i problemi non vincolati.

Un'ulteriore classificazione basata sulle proprietà di convessità potrebbe essere considerata. In tal caso, problemi di minimo in cui la funzione obiettivo è convessa e l'insieme ammissibile è un insieme convesso sono detti *problemi di Programmazione Convessa* (PC). Una caratteristica particolare di questi problemi è quella per cui ogni soluzione che sia di ottimo locale è anche di ottimo globale.

4.3 Criteri multi-obiettivo per l'ottimalità

I problemi di ottimizzazione discussi finora presentano un solo criterio da ottimizzare (problemi definiti sinteticamente con l'acronimo SOP, *Single Optimization*

Problem). Nella realtà si ha spesso a che fare con problemi in cui bisogna ottimizzare simultaneamente due o più obiettivi contrastanti soggetti a determinati vincoli. Per esempio, se vogliamo ottimizzare le prestazioni di un prodotto e contemporaneamente ridurre i costi di produzione, abbiamo a che fare con un problema di ottimizzazione multi-obiettivo. Questi problemi sono chiamati MOP, dall'inglese *Multiobjective Optimization Problem*, e sono formulati come segue:

$$\min_{\mathbf{x} \in X} g(\mathbf{x}) = [f_1(\mathbf{x}), f_2(\mathbf{x}), \dots, f_C(\mathbf{x})] \quad (4.4)$$

dove $l_k(\mathbf{x}) \leq 0, k = 1, \dots, p$ e $h_k(\mathbf{x}) = 0, k = p + 1, \dots, m,$

dove $\mathbf{x} = (x_1, x_2, \dots, x_D)$ rappresenta il vettore delle variabili nello spazio delle soluzioni D -dimensionale X , C il numero degli obiettivi e $l(\mathbf{x})$ e $h(\mathbf{x})$ gli eventuali vincoli posti al problema. Il valore ottimo x_0 sarà quello che ottimizza simultaneamente i trade-off fra tutti i C obiettivi del problema.

$$g(\mathbf{x}_0) = (f_1(\mathbf{x}_0), f_2(\mathbf{x}_0), \dots, f_C(\mathbf{x}_0)) \quad (4.5)$$

A differenza dei problemi a singolo obiettivo, i problemi multi-obiettivo non necessariamente presentano una soluzione ottima che minimizza contemporaneamente tutti gli obiettivi. Spesso, differenti obiettivi sono in conflitto tra di loro e i parametri ottimali di alcuni obiettivi non portano a valori ottimali di altri obiettivi (a volte li rendono peggiori). Pertanto, tra questi obiettivi spesso contrastanti, è necessario scegliere un compromesso o un certo equilibrio tra gli obiettivi. Se ciò non è possibile, bisogna scegliere una lista di preferenze che indichi quali obiettivi dovrebbero essere raggiunti prima.

4.3.1 Ottimizzazione pesata

Uno degli approcci più diffusi da utilizzare per risolvere questo tipo di problemi è quello dell'Ottimizzazione Pesata (o Aggregazione Lineare). Questa tecnica consiste nell'individuare una funzione $g(\mathbf{x})$ che rappresenti le combinazioni pon-

derate o l'ordine di preferenza di tutte le funzioni obiettivo. Un modo semplice per costruire la funzione è quello di utilizzare la somma pesata

$$g(\mathbf{x}) = u(f_1(\mathbf{x}), \dots, f_C(\mathbf{x})) = \sum_{i=1}^C \alpha_i f_i(\mathbf{x}) \quad (4.6)$$

dove α_i sono i coefficienti (o pesi) di ponderazione. Ogni singola funzione obiettivo f_i viene quindi moltiplicata per un peso α_i che ne rappresenta la relativa importanza. Solitamente si assume che $\sum_{i=1}^C \alpha_i = 1$.

In contrasto con la semplicità di questa tecnica vi è il principale limite, ossia quello della determinazione a priori del peso dei singoli obiettivi; questa scelta va effettuata anche in parziale o totale assenza di informazioni riguardanti quale sia effettivamente l'obiettivo più importante, al fine di pervenire ad una buona soluzione finale. Questo non accade in tutti i problemi, infatti per alcuni l'importanza dei criteri può essere ben definita dal problema stesso mentre per altri, dove ciò non è possibile, questa tecnica può portare a dei risultati non soddisfacenti.

4.3.2 Ottimizzazione di pareto

Nel 1896 Vilfredo Pareto¹ sviluppò un metodo denominato *Pareto Optimization* che non richiede la definizione a priori dell'importanza degli obiettivi da raggiungere, superando in questo modo il limite principale del metodo precedente. Questo metodo non fornisce un'unica soluzione, bensì fornisce un insieme di soluzioni definite *Pareto Optimal Set*. Per descrivere questo metodo è necessario introdurre il concetto di dominanza tra soluzioni nel senso di Pareto.

Definizione 4.5 (Dominanza di Pareto). *In un problema di minimizzazione, una soluzione x_1 domina un'altra soluzione x_2 ($x_1 \preceq x_2$) se la soluzione x_1 è migliore di x_2 in*

¹Vilfredo Federico Damaso Pareto (Parigi, 15 luglio 1848 - Cligny, 19 agosto 1923) è stato un ingegnere, economista, filosofo e sociologo italiano. Ha realizzato diversi contributi all'economia, in particolar modo nello studio della distribuzione del reddito e nell'analisi delle scelte degli individui. Ha anche contribuito ai campi della sociologia e della matematica, dove ha introdotto il concetto di efficienza Pareto. Inoltre ha contribuito a sviluppare il campo della microeconomia.

almeno un obiettivo e non è peggiore di x_2 nelle rimanenti. Ovvero:

$$x_1 \preceq x_2 \text{ se } \begin{cases} \exists j \in 1, \dots, C & f_j(\mathbf{x}_1) < f_j(\mathbf{x}_2) \\ f_i(\mathbf{x}_1) \leq f_i(\mathbf{x}_2) & \forall i \in 1, \dots, C \end{cases} \quad (4.7)$$

Se una delle due condizioni viene violata la soluzione x_1 non domina la soluzione x_2 . Questo non comporta che $x_2 \preceq x_1$, in quanto la relazione di dominanza non è simmetrica; per esempio, entrambe le soluzioni x_1, x_2 potrebbero essere non dominate. Poiché la relazione di dominanza non è simmetrica non può essere neppure antisimmetrica, inoltre non è riflessiva mentre gode della proprietà transitiva(4).

La precedente definizione di dominanza fa riferimento alla relazione di dominanza in senso debole. La dominanza in senso forte è definita come :

Definizione 4.6 (Dominanza forte). *Una soluzione x_1 domina in senso forte una soluzione x_2 ($x_1 \prec x_2$), se la soluzione x_1 è strettamente migliore della soluzione x_2 in tutti i C obiettivi.*

$$f_i(\mathbf{x}_1) < f_i(\mathbf{x}_2) \quad \forall i \in 1, \dots, C \quad (4.8)$$

Tipicamente, l'obiettivo è individuare uno specifico ottimo di Pareto. Per pervenire a questo è necessario individuare, tra l'insieme delle soluzioni trovate, quelle che non sono dominate da nessun'altra. L'insieme di queste soluzioni è detto *Non-dominated Set* e la sua rappresentazione sullo spazio delle soluzioni genera una superficie (detta anche *Pareto Front* per la forma che assume), in cui tutte le soluzioni presenti risultano equivalenti e incomparabili ai fini del problema(25).

Viene quindi lasciato all'utente finale (o decisore) il compito di selezionare o caratterizzare la soluzione considerata "migliore". In base al ruolo svolto dal decisore, i metodi risolutivi possono essere suddivisi in quattro grandi categorie:

- Metodi senza preferenze, nei quali il decisore non ha nessun ruolo e si considera soddisfacente l'aver trovato un qualunque ottimo di Pareto.

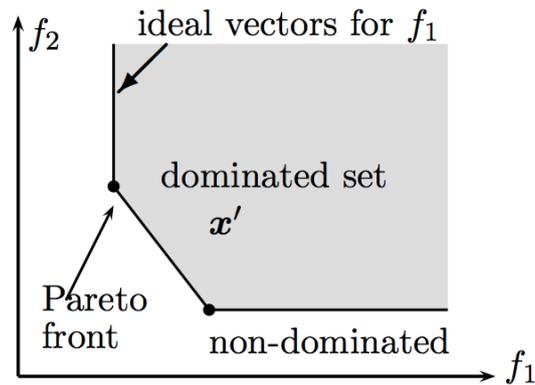


Figura 4.1: Non-dominated Set, Pareto Front e vettori ideali in un problema di minimizzazione con due obiettivi f_1 e f_2 .

- Metodi a posteriori, nei quali si genera l'insieme di tutti gli ottimi di Pareto e poi lo si presenta al decisore che sceglie la soluzione per lui migliore.
- Metodi a priori, nei quali il decisore specifica le sue preferenze prima che abbia inizio il processo risolutivo. In base alle informazioni avute dal decisore viene direttamente trovata la soluzione "migliore", senza dover (necessariamente) generare tutti gli ottimi di Pareto.
- Metodi interattivi, nei quali il decisore specifica le sue preferenze mano a mano che il processo risolutivo procede, guidando in tal modo il processo stesso verso la soluzione per lui più soddisfacente.

4.4 Approcci di risoluzione

Come si è potuto osservare, i problemi di ottimizzazione a singolo obiettivo costituiscono solo un sottoinsieme dei problemi multi-obiettivo, che rappresentano la maggior parte dei problemi che si possono incontrare nella reale. Focalizzando quindi l'attenzione ai problemi multi-obiettivo, le metodologie che permettono di affrontare questi problemi possono essere classificate in tre classi: metodo enumerativo, metodi deterministici e metodi stocastici. Ognuno di questi metodi presenta alcuni vantaggi e alcuni svantaggi.

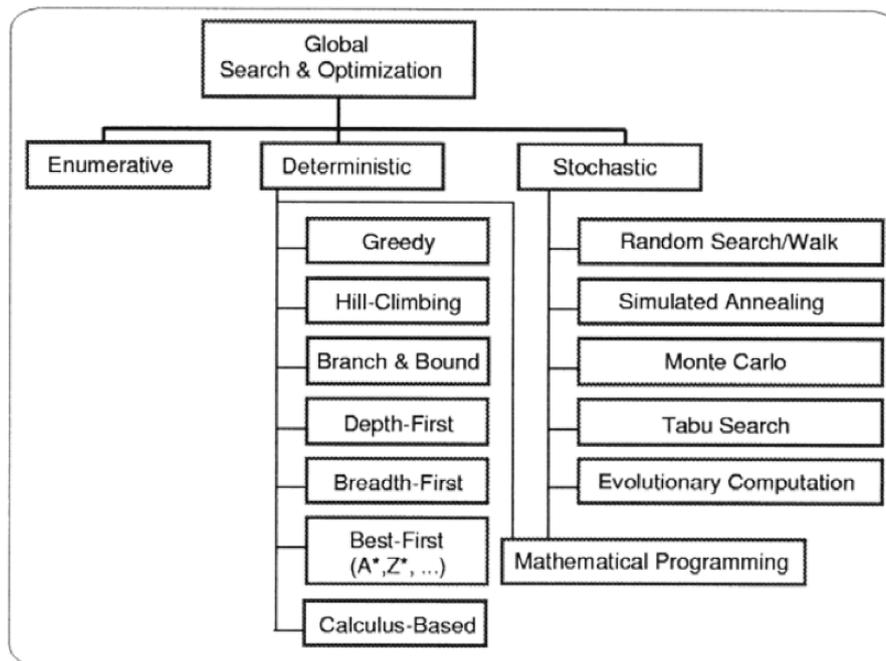


Figura 4.2: Tecniche ottimizzazione multi-obiettivo(23)

4.4.1 Metodo enumerativo

Il metodo enumerativo rappresenta l'approccio di ottimizzazione meno efficiente ma robusto ed efficace; esso consiste nella valutazione di tutte le possibili soluzioni. Per contro, questo approccio è efficace solo in caso di problemi a variabili discrete, dal momento che la valutazione di tutti le possibili soluzioni è realizzabile solo se il numero di soluzioni è finito. Potrebbe essere applicato anche a problemi con variabili continue ma richiede in questo caso una preventiva discretizzazione delle variabili. In pratica, a causa dell'esplosione combinatoriale e della complessità temporale risultante, tale approccio può essere applicato solo a casi in cui il dominio è contenuto (composto da poche variabili discretizzate con una risoluzione molto bassa).

4.4.2 Metodi deterministici

I metodi deterministici, detti anche metodi esatti, sono algoritmi di ottimizzazione che esaminano in modo esaustivo lo spazio di ricerca fino ad arrivare alla sua

precisa identificazione. L'uso di algoritmi deterministici è tipicamente basato sul presupposto per cui la funzione obiettivo è caratterizzata da talune proprietà; per tale motivo l'uso efficace di questi metodi è limitato a un piccolo sottoinsieme dei problemi di ottimizzazione multi-obiettivo. Queste assunzioni/limitazioni garantiscono che il punto di minimo trovato sia sicuramente di minimo globale per la funzione nello spazio di riferimento. Alcune delle ipotesi restrittive che generalmente vengono fatte riguardano:

- funzioni obiettivo lineari
- funzioni obiettivo convesse
- spazio di ricerca limitato
- spazio di ricerca continuo
- funzioni obiettivo unimodali
- interazioni lineari tra i parametri
- funzioni obiettivo derivabili

Quando la forma del problema non è nota a priori, l'algoritmo **Branch and Bound** (letteralmente *ramifica e delimita*) sembra essere lo strumento più promettente per trovare la soluzione globale in modo efficiente(12). Si tratta di una tecnica di enumerazione implicita, che ricerca la soluzione ottima in passi successivi, eliminano dall'analisi ampi sottoinsiemi di soluzioni ammissibili. L'idea alla base di questa tecnica è quella di ricondurre la risoluzione di un problema difficile a quella di sottoproblemi più semplici effettuando una partizione (ricorsiva) della regione ammissibile. La denominazione branch and bound, fa riferimento ai due passi che si alternano in qualsiasi implementazione della metodologia. Il passo branch (ramifica) definisce una partizione dello spazio di riferimento. Nel passo bound (delimita) per ogni sottoproblema, viene ricercato il valore ottimo,

se presente e confrontato con il valore ottimo finora trovato oppure viene dimostrato che quel sottoproblema non ammette soluzione; nel caso il sotto-problema non venga risolto si esegue il passo *branch*. L'applicazione dei due passi ricorsivamente definisce una struttura ad albero (l'albero di ricerca) i cui i nodi sono i sottoinsiemi dello spazio di riferimento. Il principale svantaggio di questa tecnica emerge nei problemi ad alta dimensionalità, in cui sono necessari tempi computazionali molto elevati per pervenire alla soluzione.

4.4.3 Metodi stocastici

I metodi stocastici, a differenza di quelli deterministici, sono algoritmi di ottimizzazione che si caratterizzano per avere al loro interno delle componenti probabilistiche che influenzano, in modo più o meno rilevante, la direzione di ricerca del punto di ottimo. Questi metodi non richiedono ipotesi restrittive o informazioni ausiliarie sulla funzione obiettivo, infatti tipicamente richiedono solo il valore obiettivo della funzione. Per la totale assenza di restrizioni e per la capacità di poter operare sia in spazi continui che discreti, questo tipo di approccio può essere applicato in generale.

Nonostante esistano algoritmi di ottimizzazione stocastica molto diversi tra loro, alcune caratteristiche basilari sono l'ispirazione alle regole che governano alcuni fenomeni naturali (natural problem solving) e l'indipendenza dal problema specifico trattato.

Per contro, i principali svantaggi di tali metodi sono la mancanza della certezza che la soluzione trovata sia di ottimo globale e che, spesso, viene fornita solo un'approssimazione numerica della soluzione.

I metodi stocastici, per essere efficaci, devono tuttavia garantire una convergenza in probabilità all'ottimo:

$$\lim_{t \rightarrow \infty} P(|x_t - x^*| \leq \varepsilon) = 1. \quad (4.9)$$

All'aumentare del numero di generazioni t , la probabilità che il candidato ottimo x_t sia realmente l'ottimo del sistema x^* deve aumentare sempre più, fino a venire approssimata ad x^* nel caso si siano effettuate un numero di generazioni molto elevate. Il simbolo ε rappresenta un valore piccolo a piacimento. Una delle caratteristiche più importanti di questi metodi è quella di generare soluzioni accettabili in un tempo relativamente ristretto.

L'approccio stocastico più semplice è il Pure Random Search (o Metodo di Monte Carlo); in questa tecnica lo spazio di ricerca viene campionato in modo casuale N volte: se il punto campionato è migliore della soluzione migliore finora trovata, viene mantenuto, altrimenti viene abbandonato. La soluzione migliore al termine degli N campionamenti sarà considerata la soluzione migliore e quindi l'ottimo globale per il problema.

Una classe di questi metodi rientra nella categoria degli *algoritmi euristici*, ossia algoritmi che utilizzano l'informazione attualmente in possesso per decidere quale possibile soluzione candidata dovrebbe essere esaminata al passo successivo. Gli algoritmi euristici si suddividono in due grandi famiglie: gli algoritmi euristici costruttivi e gli algoritmi di ricerca locale. Gli algoritmi appartenenti alla prima famiglia partono da una soluzione vuota e determinano di passo in passo quali sono i nuovi elementi da aggiungere alla soluzione base, fino ad ottenere una soluzione completa. Gli algoritmi di ricerca locale partono da una soluzione ammissibile e, successivamente di passo in passo, cercano di migliorarla effettuando semplici modifiche alla soluzione corrente; questi algoritmi terminano quando non esistono più modifiche del tipo prescelto in grado di migliorare la soluzione corrente (minimo locale).

Per aumentare l'applicabilità di una euristica è possibile dotarla di parametri che permettano di personalizzare meglio il suo operato relativamente al particolare tipo di problema affrontato. In questi casi si parla di *meta-euristiche* e sono dei metodi utili a risolvere una grande quantità di problemi. Esse, per essere efficaci,

necessitano di un accurato settaggio dei diversi parametri che la compongono.

Una classe molto importante di meta-euristiche probabilistiche è quella dei metodi di Evolutionary Computation (EC). Questa classe comprende tutti gli algoritmi che si basano su un insieme di soluzioni candidate multiple (chiamato popolazione) che sono iterativamente raffinati.

4.5 Algoritmi evolutivi

Attualmente esistono più di trenta tecniche di programmazione matematica per problemi di ottimizzazione multi-obiettivo. Tuttavia, queste tecniche tendono a generare uno alla volta gli elementi dell'insieme di Pareto; inoltre, la maggior parte di essi sono molto sensibili alla forma della curva di Pareto (infatti non funzionano quando la Curva di Pareto è concava). In risposta a questi problemi è stata sviluppata una sotto-categoria degli algoritmi stocastici che si basa su quanto viene proposto e offerto dalla natura: studi hanno evidenziato come la natura (che ha goduto di un periodo lunghissimo a disposizione, come 3,7 miliardi di anni⁽⁴⁾) fornisca una vastissima gamma di esempi di capacità di ottimizzazione; gli algoritmi evolutivi si basano sul principio dell'evoluzione delle specie, teorizzato per la prima volta da Charles Darwin nella seconda metà del diciannovesimo secolo⁽⁶⁾.

La **selezione naturale** si basa sul principio secondo cui le specie animali e vegetali che meglio si adattano all'ambiente che le circonda sono quelle che hanno la maggior probabilità di riprodursi. Di conseguenza, sono gli individui che meglio si adattano all'habitat esterno a trasmettere alle generazioni successive le loro caratteristiche genetiche distintive, le stesse che hanno permesso loro di vincere la sfida per la sopravvivenza. Questo principio dà luogo al processo di evoluzione delle specie, grazie al quale ogni specie vivente può sopravvivere e far fronte con successo alle modifiche dell'ambiente esterno da cui dipendono.

Ogni individuo nasce con un corredo genetico unico, che si basa, per la maggior parte, su quello dei genitori e, in parte minore, su alcune mutazioni casuali. Entrambi gli apporti genetici sono responsabili delle nuove caratteristiche di cui l'individuo è portatore. Le caratteristiche che sopravvivono alle generazioni sono quelle che portano un miglioramento nella capacità degli individui di sopravvivere, mentre quelle che portano ad un peggioramento non vengono trasmesse. I cambiamenti fra una generazione e l'altra sono minimi, ma il loro accumularsi lungo un arco di tempo sufficientemente ampio porta a modifiche che sono decisive nella specie. La teoria dell'evoluzione delle specie si basa su tre presupposti: la selezione, la riproduzione e la variazione. La selezione è il processo nel quale vengono scelti per la riproduzione gli individui che meglio si sono distinti in termini di adattamento all'ambiente. La riproduzione (o ricombinazione) consiste nel trasmettere le caratteristiche genetiche da due individui detti genitori, ad un nuovo individuo, detto figlio. Tramite la variazione (o mutazione) avviene la modifica casuale del patrimonio genetico, in modo che possibili miglioramenti possano determinarsi.

Gli algoritmi basati sul principio evolutivo simulano ciò che accade in natura. Per fare questo va definita la popolazione iniziale di possibili soluzioni, la quale viene, di generazione in generazione, modificata seguendo una linea evolutivistica che porta la popolazione sempre più ad adattarsi alla funzione di fitness, che rappresenta l'obiettivo da perseguire. Questo avviene grazie all'uso di particolari regole, dette euristiche, che garantiscono la sopravvivenza alle soluzioni migliori, guidando così l'algoritmo verso la soluzione migliore. In linea generale, gli algoritmi che sfruttano il principio evolutivo seguono il diagramma di flusso esposto in Figura 4.3.

La potenzialità degli algoritmi evolutivi nell'ottimizzazione (in particolare in quella multi-obiettivo) è stata suggerita da Rosenberg nella sua tesi di dottorato nel 1960(4), anche se per la prima implementazione pratica bisogna aspettare

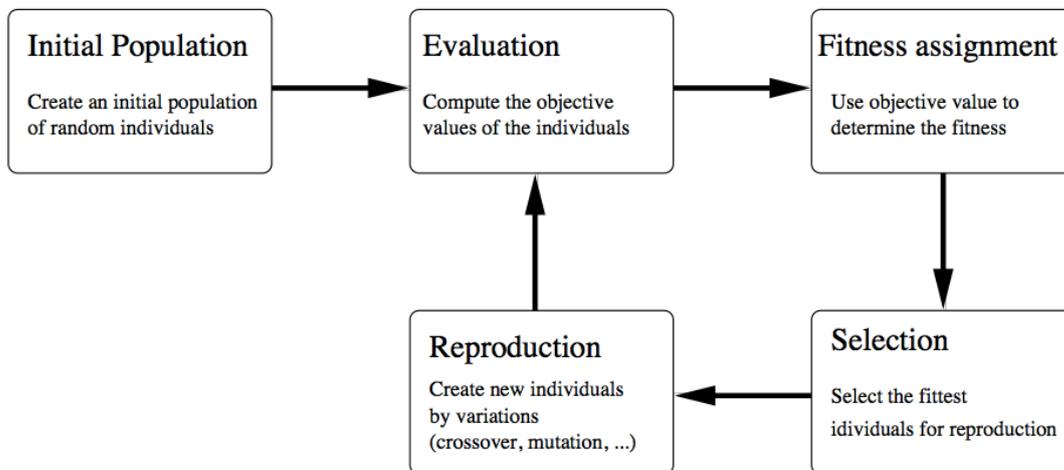


Figura 4.3: Rappresentazione di un generico flowchart degli algoritmi evolutivi

la tesi di dottorato svolta da David Shaffer nel 1984(23). Nel decennio seguente il campo rimase praticamente inattivo, ma dalla metà degli anni 1990 abbiamo assistito ad una crescita notevole, grazie allo sviluppo di diverse tecniche e applicazioni(5).

Gli algoritmi evolutivi sembrano particolarmente adatti a risolvere problemi di ottimizzazione multi-obiettivo, perché trattano contemporaneamente con una serie di possibili soluzioni (la cosiddetta popolazione). Ciò permette di trovare numerosi membri dell'insieme di Pareto ottimale in una singola esecuzione dell'algoritmo, senza dover eseguire una serie di prove separate come avviene nel caso delle tradizionali tecniche di programmazione matematica. Inoltre, gli algoritmi evolutivi sono meno suscettibili della forma o della continuità del fronte di Pareto (permettono di gestire i Fronti di Pareto discontinui o concavi, vincoli non lineari o variabili discrete).

Per contro, bisogna dire che non c'è la garanzia della ricerca di soluzioni ottimali in un ammontare di tempo limitato, anche se spesso vengono utilizzate delle prove di convergenza asintotica. La complessità computazionale di questi algoritmi è molto elaborata e la messa a punto dei parametri avviene per lo più

mediante prove.

Nei paragrafi che seguono vengono descritti i principali algoritmi che sfruttano il paradigma genetico-evolutivo appena descritto.

4.5.1 Algoritmi genetici

L'algoritmo genetico (GA, Genetic Algorithm) fa parte degli Algoritmi Evolutivi ed è stato sviluppato per la prima volta nel 1975 da John Henry Holland, professore dell'università del Michigan anche se venne delineato già nel 1950 da Alan Turing(22). Il suo funzionamento è ispirato alla teoria Darwiniana dell'evoluzione delle specie, fenomeno naturale per cui solo gli individui più adatti e meglio dotati sopravvivono. L'algoritmo genetico non opera su una singola soluzione, come avviene con la maggior parte degli algoritmi, bensì su un insieme di soluzioni.

Ogni soluzione (o individuo) viene rappresentato mediante il proprio patrimonio genetico in una stringa detta cromosoma, la quale è suddivisa in tante unità, dette geni, che sono utilizzate per rappresentare le variabili del problema. Esse vengono codificate in modo che ad ogni punto dello spazio campionario corrisponda una sola stringa di codifica, denominata appunto cromosoma. Gli algoritmi genetici operano quindi su questa popolazione di cromosomi che vengono fatti riprodurre selettivamente in base alle prestazioni dei fenotipi a cui danno origine rispetto al problema da risolvere. Durante il processo riproduttivo le repliche dei cromosomi degli individui migliori vengono accoppiate casualmente e parte del materiale genetico viene scambiato, mentre alcune piccole mutazioni alterano localmente la struttura del codice genetico. Le nuove strutture genetiche vanno quindi a rimpiazzare quelle dei loro genitori dando luogo ad una nuova generazione di individui. Lo scopo dell'algoritmo è quello di "spostare", di generazione in generazione, la popolazione verso una composizione ottimale degli individui che la formano in relazione al problema in esame. Gli algoritmi gene-

tici si basano quindi su tre operatori principali: la riproduzione selettiva degli individui migliori, la ricombinazione genetica (crossover) e le mutazioni casuali dei cromosomi. Prima di analizzare con maggior dettaglio il funzionamento di questi operatori, osserviamo che vi sono due importanti strumenti necessari allo sviluppo di un algoritmo genetico: la codifica genetica e la funzione di fitness.

Codifica

Con il termine “codifica genetica” ci si riferisce al tipo di rappresentazione che viene utilizzata per identificare le soluzioni del problema nei cromosomi. Un tipo di codice usato molto frequentemente è la codifica binaria Gray con lunghezza e ordine delle stringhe prestabilita, in modo che ogni soluzione sia rappresentabile da un vettore di bit $\{0,1\}$ di lunghezza fissa e che ogni gene possa essere univocamente individuato. La codifica binaria Gray si differenzia dalla semplice codifica binaria in quanto il passaggio da un intero al successivo avviene modificando un solo bit della codifica. Questa soluzione permette di avere distanze tra un intero e il successivo pari di 1, in quanto un solo bit varia tra due codifiche di interi adiacenti. La scelta del tipo di codifica è però importante per il tipo di problema che si vuole risolvere: non esiste una codifica che vada bene per tutti i problemi, né esistono delle regole generali che permettano di fare delle scelte ottimali. Il problema della rappresentazione genetica e delle regole di decodificazione da genotipo a fenotipo è quindi molto importante per poter sfruttare al meglio le potenzialità di ricerca dell'algoritmo genetico.

Funzione di Fitness

La funzione di valutazione serve per giudicare le prestazioni di ciascun fenotipo rispetto al problema che si vuole risolvere: essa fornisce un valore numerico per ciascun individuo proporzionale alla bontà della soluzione offerta. Di conseguenza gli individui che presentano valori di fitness più elevati sono i più adatti all'ambiente e quindi hanno maggiori probabilità di sopravvivere e di riprodur-

si. Va ricordato che la funzione di fitness non deve essere per forza la funzione da massimizzare in quanto potrebbe essere una trasformazione della stessa (per esempio una trasformata monotona crescente).

Riproduzione Selettiva

Una volta definito il tipo di rappresentazione genetica e la funzione di fitness, il primo passo consiste nella creazione di una popolazione iniziale di stringhe genetiche. Solitamente la popolazione iniziale è composta da stringhe casuali. Ciascuna stringa di questa generazione viene a turno decodificata e valutata in base alla funzione di fitness. Il processo di riproduzione selettiva consiste nello scegliere gli individui (detti genitori) che partecipano alla creazione di nuovi individui nella popolazione (detti figli). La loro scelta viene guidata dal grado di adattamento di ciascun individuo all'ambiente circostante, in linea con il principio naturale dove a riprodursi sono generalmente gli individui più forti. Quindi maggiore sarà l'adattamento, maggiore sarà la probabilità di scelta del cromosoma. In questo modo è auspicabile che le nuove soluzioni risultino di maggiore qualità (in termini della funzione di fitness) rispetto a quella attribuibile a chi li ha generati. Questo processo può avvenire in molti modi. Possono venire scelti, ad esempio, solo cromosomi figli, oppure sia genitori che figli; si può avere la certezza o meno del campionamento degli individui che attualmente risultano migliori; la loro selezione può essere o meno proporzionale al grado di adattamento, oppure può venire fatta all'interno di gruppi chiusi in cui a vincere è uno soltanto.

Crossover e Mutazione

La fase di crossover è quella in cui avviene lo scambio delle informazioni genetiche dai genitori ai figli. In questo modo i figli vengono dotati di un corredo genetico basato su quello dei genitori. Il ruolo di questo operatore è quello di ricombinare (crossing-over appunto) parte della stringa cromosomica dei due

genitori al fine di formare due nuove stringhe. In questo modo il patrimonio genetico dei genitori viene trasmesso a figli ricombinato, nella speranza che risulti migliore.

Per ciascuna delle coppie viene scelto un punto d'incrocio casuale attorno al quale viene effettuato uno scambio reciproco di materiale genetico, come rappresentato in Figura 4.4. Se, invece, la soglia di separazione con cui vengono rimescolati i cromosomi dei genitori è più di una, ad esempio due, si parla di crossover a due punti di taglio. L'operazione di crossover, non introduce nuovi

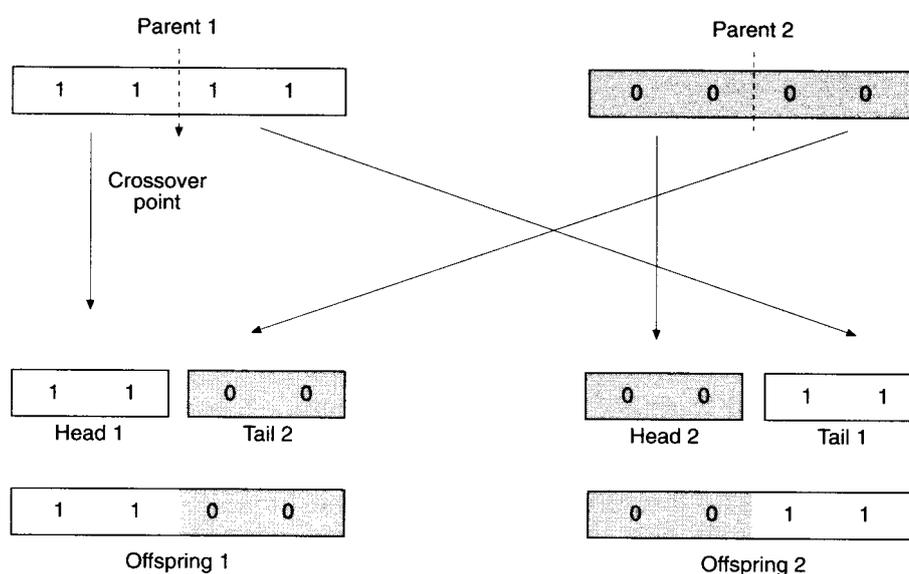


Figura 4.4: Un esempio di ricombinazione genetica: il 'Single Crossover'

valori dei geni ma si limita solamente a rimescolare quelli presenti. Questo comporta dei rischi per l'algoritmo. A tal proposito risulta fondamentale l'operatore di mutazione, per quanto riguarda la rapidità nell'esplorare lo spazio di ricerca; la mutazione infatti porta un po' di casualità in essa in modo che nessun punto dello spazio abbia probabilità nulla di essere esaminato. In pratica si concretizza con la scelta, nei cromosomi figli, di uno o più geni a cui vengono assegnati dei valori in modo casuale, diversi da quelle che avrebbero ricevuto dai genitori (Figura 4.5).

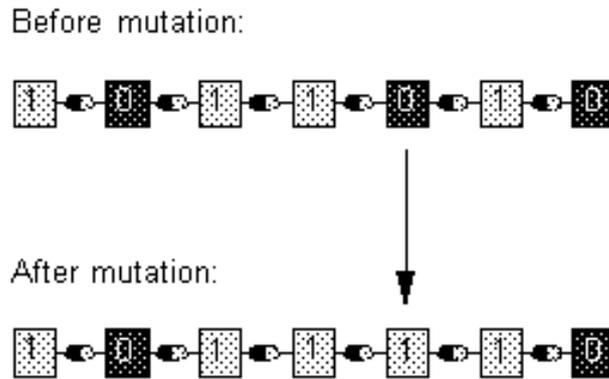


Figura 4.5: Un esempio di mutazione nella codifica binaria

A volte in aggiunta è necessario effettuare un'operazione di filtro (Repair Method), che consiste in una mutazione particolare che trasforma una soluzione inammissibile in una soluzione ammissibile.

La nuova popolazione di stringhe rimpiazza parzialmente o completamente le vecchie stringhe. Il processo di codifica, valutazione, riproduzione selettiva crossover e mutazione si ripete ed ogni ciclo dell'algoritmo è detto generazione e fornisce un'insieme di possibili soluzioni al problema.

4.5.2 Ant colony optimization

Ant Colony Optimization è una tecnica di ottimizzazione ideata da Marco Dorigo nel 1991 sulla base dello studio delle colonie di formiche. Si tratta di una tecnica che viene usata per risolvere problemi di ricerca di percorso minimo in un grafo. Essa si basa sul concetto di stigmergia, ossia quel meccanismo di comunicazione che si verifica attraverso la modifica dell'ambiente.

In natura, il comportamento di ogni formica è quello di muoversi a caso deponendo una traccia di feromone (si tratta di un particolare marcatore chimico percepito dalle formiche); se trova il cibo, ritorna al nido deponendo un'ulteriore traccia di feromone. Le altre formiche seguono uno dei sentieri in modo casuale (preferiscono quello con maggiore traccia di feromone). Dal momento che le

formiche sul percorso più breve depongono tracce di feromone più velocemente, questo percorso viene rinforzato con più feromone, rendendolo più attraente alle formiche future. Per le formiche diventa sempre più probabile seguire il percorso più breve, poiché viene costantemente rinforzato con una quantità maggiore di feromone, mentre il feromone dei percorsi più lunghi evapora (infatti il feromone evapora nel tempo, in modo tale che se il percorso non viene rinforzato dalle formiche, la traccia scompare). In Figura 4.6 viene rappresentato il comportamento delle formiche in natura.

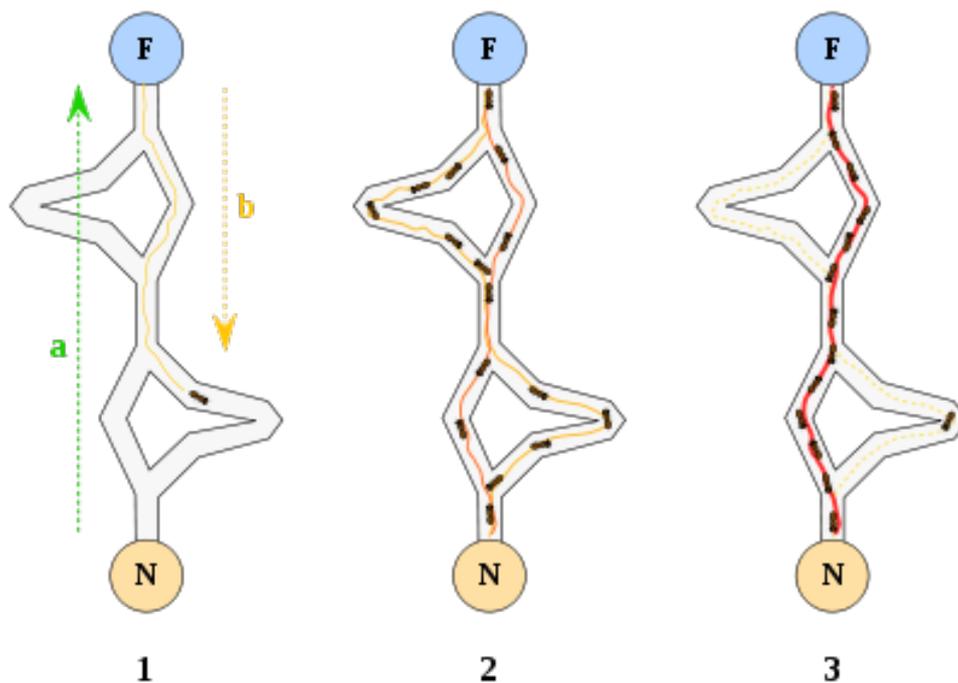


Figura 4.6: *Comportamento delle formiche in natura*

La sua realizzazione avviene attraverso gli elementi dei grafi (nodi ed archi), dove i nodi rappresentano gli stati mentre gli archi le transizioni ammissibili. Ogni “formica” parte da uno stato iniziale ed ha una condizione di terminazione del percorso. Nel suo viaggio, passo dopo passo in base alle decisioni prese ad ogni nodo, crea e percorre un certo percorso di cui tiene traccia, data la peculiarità per cui dispone di una memoria.

I passi di funzionamento della tecnica consistono nell'inizializzazione delle formiche e degli stati; successivamente viene scelto la transizione successiva in maniera probabilistica in relazione alla sua attrattività e visibilità. Ogni formica mantiene inoltre un elenco delle transizioni che non sono possibili in quell'iterazione. Infine viene aggiornata l'attrattiva delle transizioni in base al numero di formiche che ci sono passate attraverso.

Due sono gli aspetti fondamentali di questa tecnica di ottimizzazione e riguardano la probabilità di scegliere un percorso e la velocità di evaporazione del feromone. Sono presenti diversi modi per la definizione di questi aspetti, che tuttavia rappresenta un campo di ricerca molto attivo.

Attualmente, il miglior metodo in circolazione(25) prevede che la probabilità di una formica, situata in particolare nodo i , di scegliere il percorso che va dal nodo i al nodo j è data da:

$$p_{ij} = \frac{\phi_{ij}^\alpha d_{ij}^\beta}{\sum_{ij=1}^n \phi_{ij}^\alpha d_{ij}^\beta}$$

dove $\alpha > 0$ e $\beta > 0$ sono i parametri di influenza e i loro valori tipici sono $\alpha \approx \beta \approx 2$. La concentrazione di feromone sull'arco tra i e j è rappresentata da ϕ_{ij} mentre l'opportunità (o attrazione) dello stesso arco è d_{ij} . Spesso si utilizza della conoscenza a priori sul grafo, ad esempio si utilizza la distanza tra due nodi s_{ij} in modo che $d_{ij} \propto 1/s_{ij}$, il che implica che percorsi più brevi saranno selezionati a causa del minor tempo di percorrenza e di conseguenza la concentrazione di feromone sarà maggiore. La concentrazione di feromone in un arco può variare con il tempo a causa dell'evaporazione dello stesso. Questo può essere un fattore positivo, in quanto l'evaporazione del feromone può evitare di rimanere intrappolati in ottimi locali. Considerando un tasso costante γ di evaporazione o decadimento, la concentrazione di feromone di solito varia in modo esponenziale con il tempo:

$$\phi(t) = \phi_0 e^{-\gamma t},$$

dove ϕ_0 rappresenta la concentrazione iniziale di feromone e t rappresenta il tempo. Se $\gamma_t \ll 1$ si ottiene $\phi(t) \approx (1 - \gamma_t)\phi_0$ e considerando un incremento unitario del tempo l'evaporazione si può approssimare come $\phi^{t+1} \leftarrow (1 - \gamma)\phi^t$. Pertanto, una formula semplificata per l'aggiornamento del feromone si può scrivere come:

$$\phi_{ij}^{t+1} = (1 - \gamma)\phi_{ij}^t + \delta\phi_{ij}^t$$

dove $\gamma \in [0, 1]$ rappresenta il tasso di evaporazione del feromone. L'incremento $\delta\phi_{ij}^t$ rappresenta la quantità di feromone depositata lungo il percorso che porta da i a j al tempo t quando una formica percorre una distanza L . Generalmente $\delta\phi_{ij}^t \propto 1/L$. Se non sono presenti formiche su quel percorso, la quantità di feromone depositata sarà pari a zero.

I vantaggi principali di questa tecnica sono la sua capacità di risolvere alcuni problemi NP-hard in tempo polinomiale ed il fatto che è approssimativamente convergente alla soluzione ottima. Per contro, è limitata a problemi che possono essere simulati da grafi e poi ottimizzati.

4.5.3 Particle swarm optimization

La Particle Swarm Optimization (PSO) venne ideata nel 1995 dagli statunitensi James Kennedy e Russel Eberart, rispettivamente psico-sociologo e ingegnere elettrico, e deriva dall'analisi dei meccanismi di interazione tra gli individui appartenenti ad un gruppo, i quali risultano particolarmente interessanti qualora sia presente un obiettivo comune.

Il comportamento di uno stormo è molto complesso in quanto non esiste un leader che decide come muoversi, tutti gli individui sono leader in quanto subentrano dei fattori di comunicazione e di imitazione. Un modello semplice si basa sulla **separazione**, in quanto ogni individuo deve evitare la collisione con i vicini (repulsione a corto raggio), sull'**allineamento** poiché l'individuo punta verso la posizione media dello stormo, ed infine sulla **coesione** dovuta al fatto che

ogni individuo dovrebbe puntare verso la posizione media dei vicini (attrazione a lungo raggio).

Quando un individuo, nel suo movimento, scorge una fonte di cibo si trova a dover scegliere se allontanarsi dal gruppo per raggiungerlo (individualismo) oppure se rimanere in gruppo (socialità). Qualora più di un individuo si dirige verso il cibo, anche altri membri vicini possono cambiare la loro direzione per sfruttare la stessa fonte di nutrimento; gradualmente, il gruppo cambia direzione verso le zone più promettenti in quanto pian piano l'informazione si propaga a tutti gli individui del gruppo.

La strategia di ricerca può essere espressa come bilanciamento tra exploration ed exploitation. Con exploration si intende la capacità di esplorare un'area diversa dello spazio di ricerca, al fine di individuare un nuovo valore di ottimo. Exploitation invece è la capacità di concentrare la ricerca intorno a un'area promettente, con l'obiettivo di definire una soluzione candidata. Ogni individuo (detto anche particella) tramite queste fasi si muove nello spazio ed ha la capacità di ragionare sulla base di due informazioni: la propria posizione migliore e la posizione migliore globale o del vicinato.

Vediamo ora nel dettaglio il funzionamento di questa tecnica. Consideriamo una particella come una configurazione dei parametri che si muove in uno spazio reale a D dimensioni; questa particella i al tempo t assume una posizione $X_i^t \in \mathbb{R}^D$ e una velocità $V_i^t \in \mathbb{R}^D$. Esiste poi una funzione di fitness $f : S \rightarrow \mathbb{R}$, ove $S \subset \mathbb{R}^D$, che valuta la bontà della posizione delle particelle. Ogni particella dispone di una memoria in cui tiene traccia della migliore posizione raggiunta (P_i^t) e, grazie alla comunicazione con i vicini, della migliore posizione raggiunta dai suoi vicini (B_i^t).

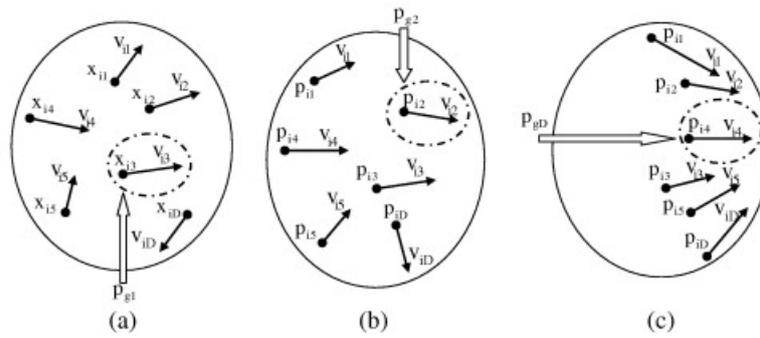


Figura 4.7: Esempio di come si evolvono le particelle

Sulla base delle informazioni a disposizione la particella modifica la propria posizione e velocità per il tempo $t + 1$ sulla base delle formule seguenti:

$$V_{ij}^{t+1} = V_{ij}^t + \phi_1 r_1 (P_{ij}^t - X_{ij}^t) + \phi_2 r_2 (B_{ij}^t - X_{ij}^t)$$

$$X_{ij}^{t+1} = X_{ij}^t + V_{ij}^{t+1}$$

Nella formula dell'aggiornamento della velocità sono presenti due termini riconducibili alle fasi di exploitation ed exploration, infatti $(P_{ij}^t - X_{ij}^t)$ è il termine cognitivo mentre $(B_{ij}^t - X_{ij}^t)$ è quello sociale. I termini ϕ_1 , ϕ_2 sono i parametri di importanza, che definiscono se dare più importanza al termine cognitivo piuttosto che a quello sociale; in letteratura è diffuso l'uso di $\phi_1 = \phi_2 = 2$. Infine r_1, r_2 sono numeri casuali nell'intervallo $[0, 1]$. Bisogna infine determinare il metodo in cui due particelle possono comunicare tra di loro attraverso la definizione di un grafo di "vicinanza".

La topologia del grafo può essere statica oppure dinamica. Tra le topologie statiche più diffuse troviamo la Global Best che considera il grafo completo (tutte le particelle comunicano tra loro) e la Local Best in cui si considera un anello (ogni particella comunica solo con le particelle direttamente collegate nel grafo). In Figura 4.8 sono rappresentate graficamente alcune di queste topologie.

Lo schema di un generico algoritmo di PSO viene rappresentato in Figura 4.9.

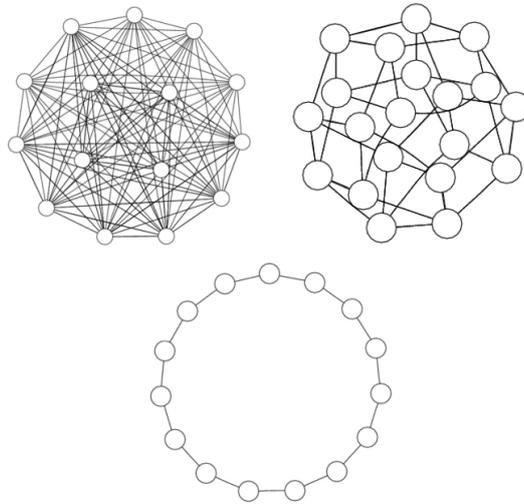


Figura 4.8: Esempi di topologie usate nel PSO

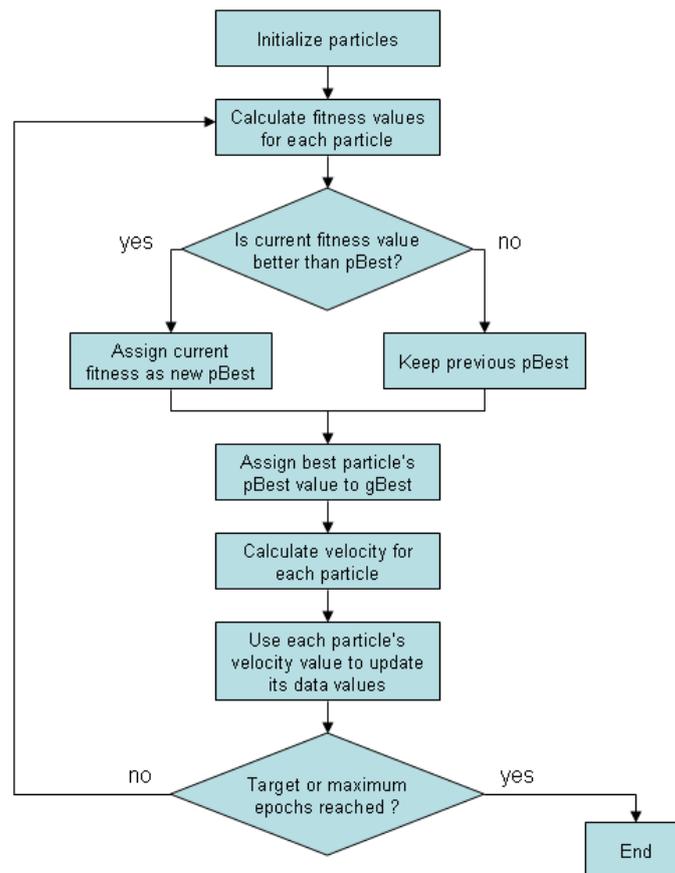


Figura 4.9: Rappresentazione di un generico flowchart del PSO

Una variante dell'algoritmo è stata presentata da Shi e Eberhart nel 1998. La variante aggiunge un parametro w definito Inertia Weight che assume un ruolo chiave nell'ottica di garantire equilibrio tra la ricerca globale e quella locale. L'introduzione di w ha portato da subito un miglioramento significativo in un gran numero di applicazioni. Questo nuovo parametro va a moltiplicare la velocità v^t nella formula precedente modificandola nel modo seguente:

$$V_{ij}^{t+1} = wV_{ij}^t + \phi_1 r_1 (P_{ij}^t - X_{ij}^t) + \phi_2 r_2 (B_{ij}^t - X_{ij}^t)$$

Il valore di questo parametro generalmente diminuisce nel tempo.

In letteratura sono presenti molteplici versioni, che meglio si adattano ad un problema specifico. Le principali fanno uso di topologie di vicinanza dinamiche oppure di regole differenti per l'aggiornamento della velocità. Sono presenti delle versioni apposite per problemi di ottimizzazione discreta.

Si tratta di una tecnica che è stata applicata in svariati problemi, poiché esiste la possibilità di modificare la versione "originale" apportando degli accorgimenti in modo tale che si adatti al problema specifico. Tuttavia l'ambito naturale di applicazione del PSO è l'ottimizzazione numerica.

Capitolo 5

LA COSTRUZIONE DEL MODELLO STATISTICO PER LA REALIZZAZIONE DEL COMFORT ABITATIVO

In questo capitolo verranno formalizzati ed applicati gli argomenti teorici precedentemente introdotti. I dati utilizzati in questo lavoro si riferiscono a due distinti periodi di osservazione. In un primo momento, dal 31 Gennaio 2011 al 21 Aprile 2012, sono stati rilevate alcune variabili climatiche di interesse mentre in un secondo momento, dal 26 Giugno 2012 al 23 Settembre 2012, in aggiunta alle variabili precedenti sono state rilevate anche alcune variabili inerenti al sistema HVAC. Nel periodo, i valori sono stati rilevati ogni 15 minuti.

5.1 Modelli previsivi per le variabili climatiche esterne

In primo luogo, è necessario delineare il ruolo dei fattori ambientali caratteristici della posizione in cui è situato l'edificio nei confronti del sistema HVAC. L'obiettivo consiste nella modellazione delle variabili climatiche esterne al fine di ottenere le previsioni delle stesse nei periodi successivi. In tale modo, i valori delle previsioni saranno utilizzati in una fase successiva nella modellazione del consumo energetico del sistema HVAC.

5.1.1 La temperatura esterna

La temperatura esterna è un dato facilmente rilevabile e risulta molto utile nel problema, infatti influenza direttamente il raggiungimento del comfort di un edificio. Basti pensare alla muratura perimetrale che viene a contatto con l'ambiente interno da una parte e con l'ambiente esterno dall'altra, anche se viene utilizzato il miglior sistema isolante in commercio, sarà presente una dispersione termica che influenzerà la temperatura interna.

A fine descrittivo, abbiamo rappresentato il valore medio giornaliero della serie originalmente rilevata con il valore massimo e minimo osservato¹ (Figura 5.1). Questo ha permesso di evidenziare l'andamento della temperatura nel lungo periodo, mettendo in risalto l'influenza del periodo di osservazione (stagione), con un andamento sinusoidale abbastanza marcato. La figura pone in evidenza anche la variabilità dell'escursione termica dovuta ad una caratterizzazione climatica della zona, infatti le temperature minima e massima tendono ad essere più lontane nei mesi caldi, tra aprile ed agosto.

Tuttavia, l'interesse è rivolto al breve periodo per cui abbiamo impostato un'analisi delle serie storiche focalizzano l'attenzione nelle oscillazioni infra-giornaliere, considerando le osservazioni a frequenza oraria. Un primo approccio descrittivo si è realizzato tramite l'impostazione di un'analisi classica volta alla scomposizione della serie in trend, stagionalità ed componente erratica. Il risultato dell'analisi, svolta su insieme di osservazioni rilevate in un periodo di 30 giorni, è visualizzato in Figura 5.2

La figura evidenzia come la componente stagionalità assuma un'importanza rilevante. Questa analisi viene utilizzata come base di partenza per l'analisi stocastica della serie. Infatti, l'andamento del trend evidenziato in figura congiunto al test KPSS² ha confermato la non stazionarietà della serie.

¹I valori espressi si riferiscono alla prima serie rilevata

²Il test KPSS (Kwiatkowski-Phillips-Schmidt-Shin) viene utilizzato per testare l'ipotesi nulla che la serie storica osservata sia stazionaria intorno ad un trend deterministico.

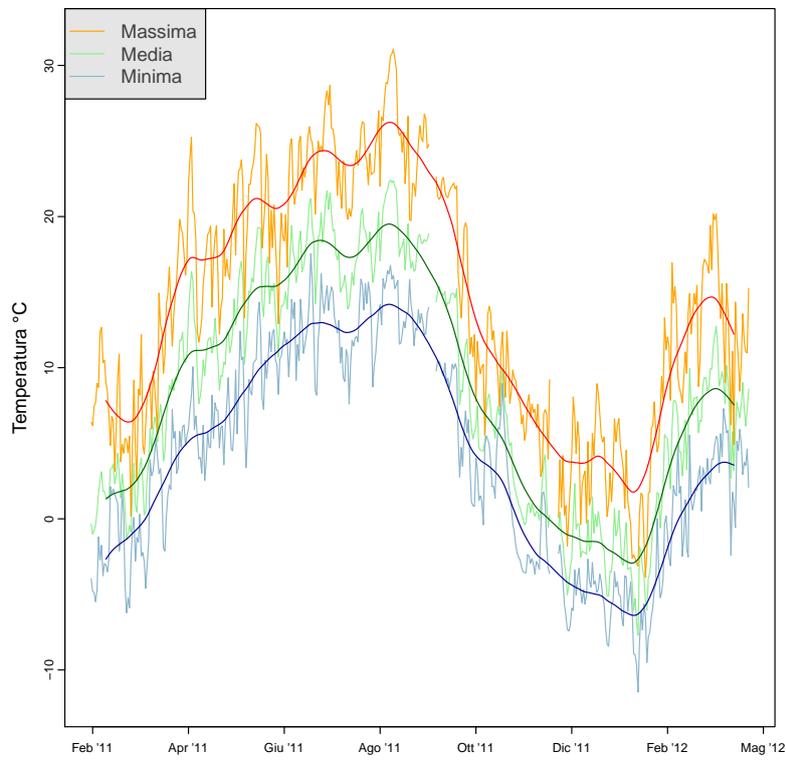


Figura 5.1: Temperatura esterna massima, media e minima rilevata

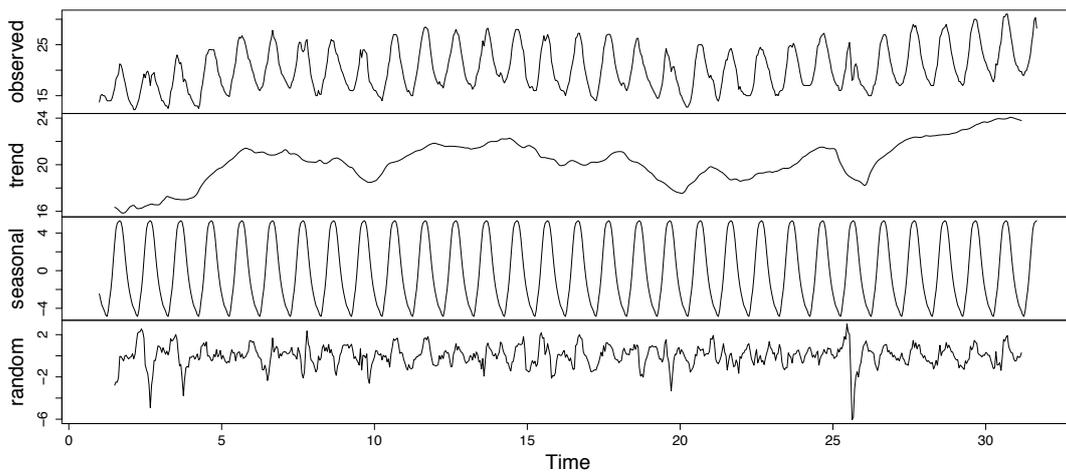


Figura 5.2: Analisi Classica sulla Temperatura Esterna

Una corretta individuazione del grado degli ordini del modello ARIMA (o eventualmente SARIMA) si basa sullo studio dell'autocorrelazione globale e parziale 5.3. Nella stessa figura è stato affiancato un grafico che mette in evidenza la presenza di una struttura di autocorrelazione all'interno di una serie storica, mediante la rappresentazioni dei valori al tempo t rispetto i valori a diversi lag precedenti.

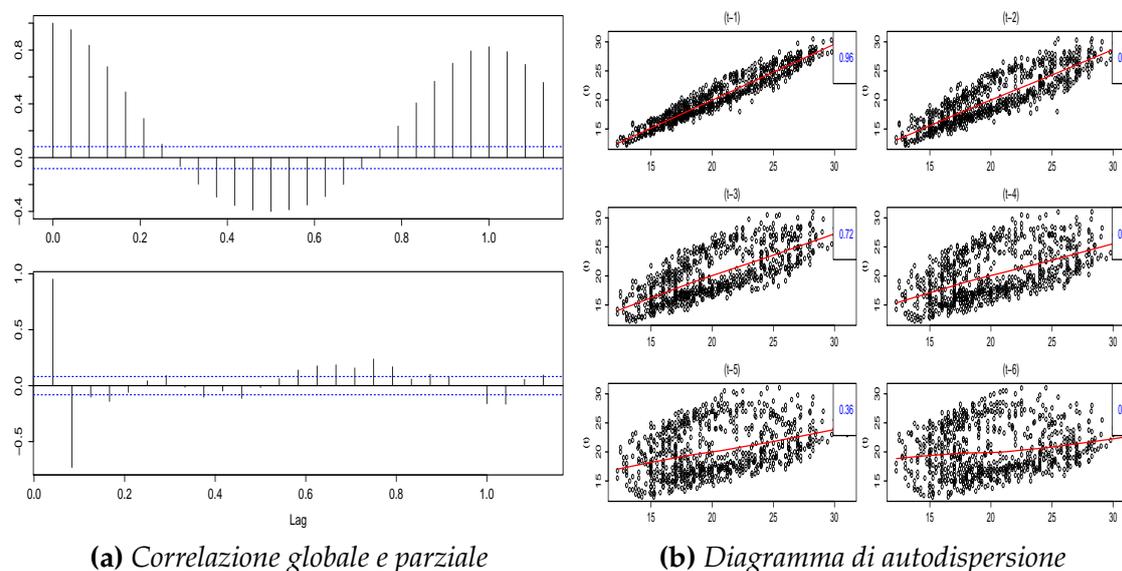


Figura 5.3: *Analisi preliminare della temperatura esterna*

Per una corretta determinazione degli ordini della componente stagionale del modello abbiamo inoltre utilizzato il correlogramma globale e parziale sulla differenziazione della serie di ordine pari alla frequenza dei dati (nel nostro caso 24). I valori dei coefficienti ai diversi lag rispecchiano la stessa informazione dei valori di un modello ARIMA (3), solo che in questo caso si riferiscono alla modellazione dei parametri della componente stagionale.

Sulla base dei risultati dei grafici appena descritti, sulla significatività dei parametri ottenuti e sul valore dell'AIC³ del modello, il modello finale sulla

³ L'AIC (Akaike Information Criterion) è una misura della bontà di adattamento di un modello statistico. Nel caso generale è definito come $-2 \log(\mathcal{L}) + 2k$, dove k è il numero di parametri del modello statistico e \mathcal{L} è il valore della funzione di verosimiglianza per il modello stimato.

temperatura esterna è un $SARIMA(1, 1, 0)(2, 0, 2)_{24}$. I coefficienti stimati sulla un sottoinsieme della prima serie delle rilevazioni sono rappresentati nella Tabella 5.1.

Coefficients:						
	ar1	sar1	sar2	sma1	sma2	
	0.3434	0.1844	0.8088	-0.0946	-0.7360	
Std. Err.	0.0090	0.0042	0.00438	0.0092	0.0089	
Error Measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
	6.014e-04	5.989e-01	4.259e-01	1.811e+00	1.263e+01	4.996e-01

Tabella 5.1: Sintesi del modello stimato sulla temperatura esterna

Il parametri del modello stimato risultato tutti significativi ed anche la diagnostica del modello conferma la bontà dello stesso. Infatti, il correlogramma dei residui non evidenzia la persistenza di un legame residuo mentre dal test di Ljung-Box⁴ emerge come sia rimasta ancora una certa correlazione rispetto al periodo precedente (Figura 5.4).

I valori previsti sono stati confrontati con quelli realmente osservati, Tabella 5.2, confermando la bontà del modello trovato ed evidenziando come lo scostamento massimo raggiunto nella previsione di sei ore in avanti risulti di °C 1.38 .

Tempo	Previsione	Osservato	Differenza
t+1	28.16	28.00	0.16
t+2	27.15	28.00	0.85
t+3	25.98	27.14	1.16
t+4	24.46	25.84	1.38
t+5	22.94	23.75	0.81
t+6	21.71	22.45	0.74

Tabella 5.2: Confronto valori osservati e stimati della temperatura esterna

⁴ Il test di Ljung-Box (da Greta M. Ljung e George EP Box) è un test per verificare se un qualsiasi gruppo di autocorrelazioni di una serie storica è diverso da zero. Tuttavia, invece di testare la casualità ad ogni singolo ritardo, esso verifica la casualità "globale" sulla base di un certo numero di ritardi.

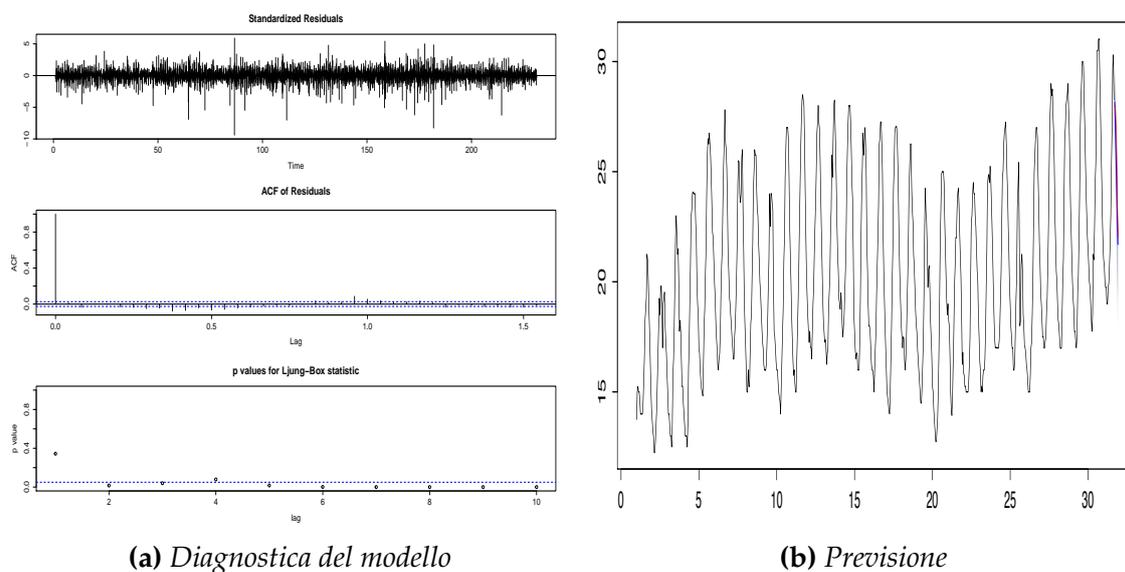


Figura 5.4: *Diagnostica del modello e previsione per la serie della temperatura esterna*

5.1.2 L'irraggiamento

Prima di procedere all'analisi dei dati relativi all'irraggiamento vediamo di cosa stiamo parlando. Come tutti i corpi caldi, il sole emette una radiazione elettromagnetica che si propaga nello spazio. La potenza del campo elettromagnetico incidente su una superficie di area unitaria prende il nome di Irraggiamento Solare (o Irradianza) ed è espresso in W/m^2 (watt al metro quadrato). La presenza dell'atmosfera ha diversi effetti sulla radiazione che arriva sulla superficie terrestre. Infatti, una parte di radiazione viene riflessa verso lo spazio, una parte è diffusa in tutte le direzioni dalle molecole dei gas atmosferici e dal vapore acqueo, una parte viene assorbita dalle molecole dell'atmosfera e da queste riemessa come radiazione infrarossa, (Figura 5.5).

L'assorbimento e la diffusione atmosferica hanno l'effetto di ridurre l'intensità della radiazione. La parte di irraggiamento che raggiunge direttamente il suolo costituisce la radiazione diretta, mentre la parte rimanente, che arriva dall'alto in ogni direzione, costituisce la radiazione diffusa. A queste va infine aggiunta la radiazione riflessa, che rappresenta la percentuale di radiazione diretta e

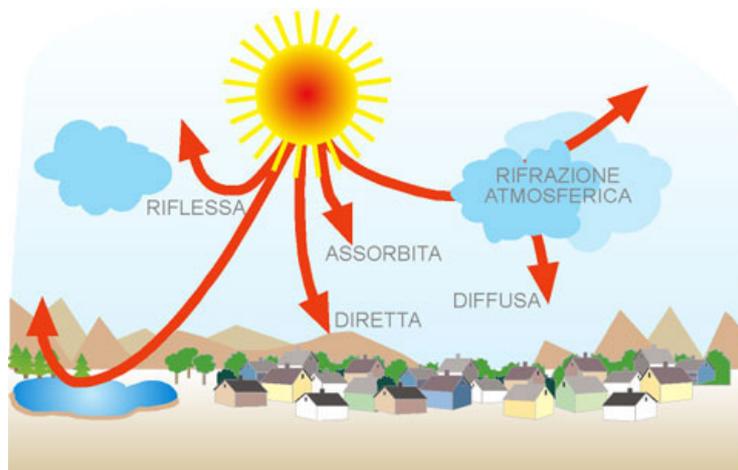


Figura 5.5: L'influenza dell'atmosfera terrestre sull'irraggiamento

diffusa che viene riflessa dal suolo o dalle superfici circostanti sulla superficie considerata. La radiazione diretta, che in una giornata di sole può essere valutata sul 80-90% di quella globale che investe direttamente la superficie orizzontale è preponderante rispetto alla diffusa, tende a ridursi all'aumentare dell'umidità e della nuvolosità presente nell'aria, fino ad annullarsi in condizioni di cielo completamente coperto (Figura 5.6⁵).

Radiazione solare	Condizioni atmosferiche							
	Cielo sereno	Nebbia	Nuvoloso	Disco solare giallo	Disco solare bianco	Sole appena percettibile	Nebbia fitta	Cielo coperto
globale	1000 W/m ²	600 W/m ²	500 W/m ²	400 W/m ²	300 W/m ²	200 W/m ²	100 W/m ²	50 W/m ²
diretta	90%	50%	70%	50%	40%	0%	0%	0%
diffusa	10%	50%	30%	50%	60%	100%	100%	100%

Figura 5.6: Intensità della radiazione solare in funzione delle condizioni atmosferiche

L'irraggiamento è quindi altamente correlato con il periodo dell'anno e con la condizione atmosferica. In Figura 5.7, sono raffigurati i valori dell'irraggiamento rilevato ogni 15 minuti, che confermano quanto appena detto anche per i dati rilevati; infatti l'andamento dovuto al periodo è ben definito, con i valori più

⁵Fonte: www.itishertz.it/solare/radiazione.htm

elevati osservati nel periodo estivo quando la posizione del sole è prossima alla terra e l'angolo di incidenza solare è maggiore (oltre alla relativa maggior esposizione), tenendo in considerazione il fatto che in figura emergono di fatto i valori giornalmente osservati maggiori in quanto, per la numerosità della serie, gli altri ne sono contenuti.

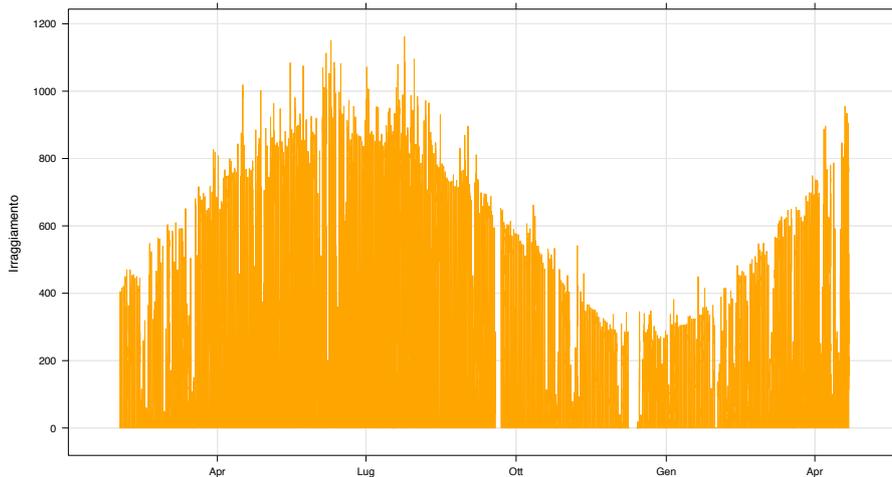


Figura 5.7: *Irraggiamento rilevato*

È stata impostata un'analisi classica con l'obiettivo di descrivere le singole componenti della serie, focalizzandoci nel breve periodo. Il risultato di questa analisi, svolto su sottoinsieme della serie, viene raffigurato in Figura 5.8. La componente stagionale assume notevole importanza nella determinazione del valore osservato.

Dopo questa descrizione della serie, un'analisi della correlazione tra i valori è d'aiuto per la determinazione del modello da applicare a questa serie. Questa analisi (Figura 5.9) congiunta all'analisi della correlazione per evidenziare la stagionalità, ha portato all'individuazione di un modello $SARIMA(1, 0, 0)(2, 0, 2)_{24}$.

Il modello conferma quanto intravisto con l'analisi classica, infatti la stagionalità assume un ruolo fondamentale ai fini del modello. I coefficienti stimati dei parametri del modello e i relativi errori sono presentati in Tabella 5.3.

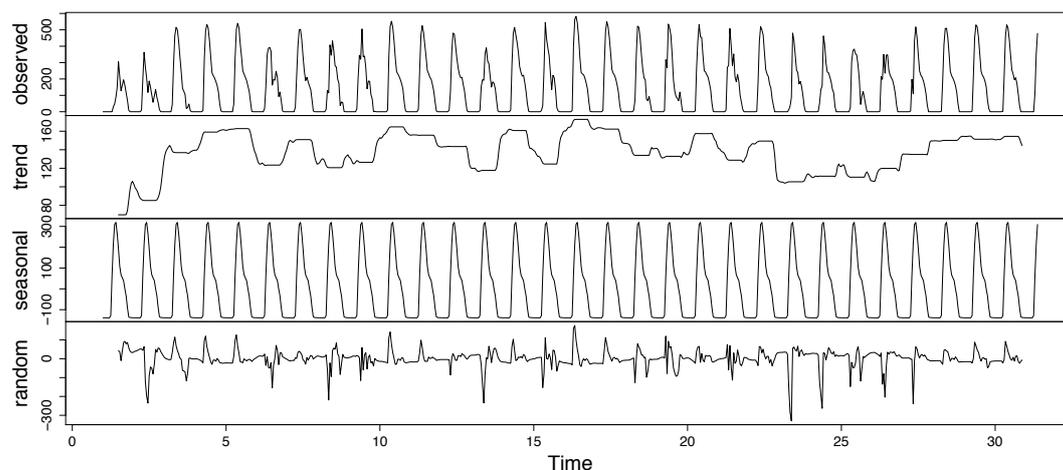
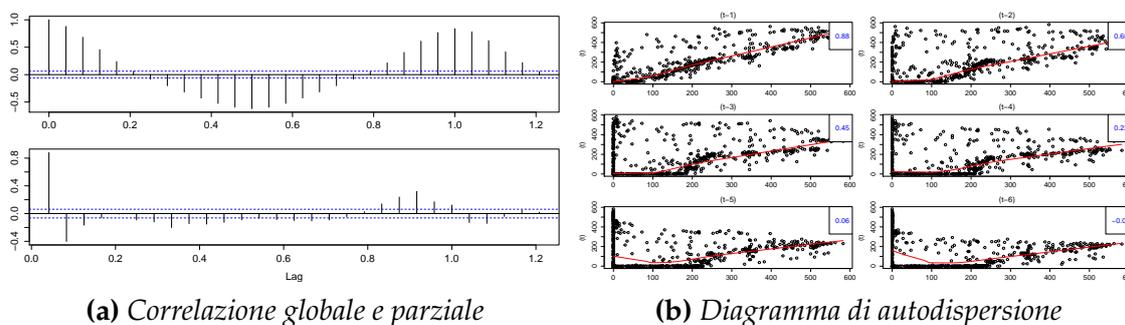


Figura 5.8: Analisi Classica sull'Irraggiamento



(a) Correlazione globale e parziale

(b) Diagramma di autodispersione

Figura 5.9: Analisi preliminare dell'irraggiamento

Coefficients:						
	ar1	sar1	sar2	sma1	sma2	
	0.6336	0.2315	0.7755	-0.0912	-0.7687	
Std. Err.	0.0249	0.0283	0.0287	0.0286	0.0271	
Error Measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
	0.8805	49.1173	24.0617	-	Inf	0.5305

Tabella 5.3: Sintesi del modello stimato sull'irraggiamento

La validazione del modello, realizzata tramite strumenti di diagnostica (5.10), evidenzia come i residui risultino incorrelati tra loro mentre l'indipendenza dei residui non viene rispettata.

Le previsioni effettuate dal modello risultano in linea con i valori reali e, se

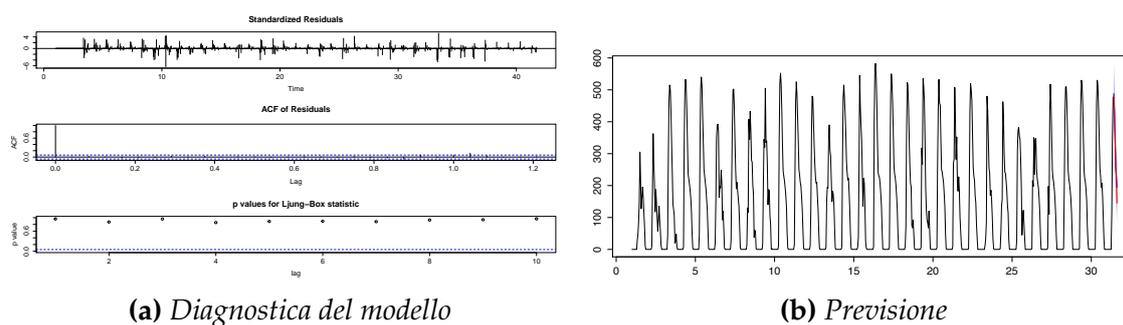


Figura 5.10: Diagnostica del modello e previsione per la serie dell'irraggiamento

considerate fino al quinto periodo di previsione, sono pressoché identiche. I valori sono riportati nella Tabella 5.4.

Tempo	Previsione	Osservato	Differenza
t+1	488.51	482.50	6.01
t+2	435.98	440.00	4.02
t+3	351.13	345.00	6.13
t+4	255.56	255.00	0.56
t+5	224.94	227.50	2.56
t+6	194.67	145.00	49.67

Tabella 5.4: Confronto valori osservati e stimati dell'irraggiamento

5.1.3 La luminosità esterna

La luminosità (o illuminamento) è una grandezza fotometrica risultante dal rapporto tra il flusso luminoso emesso da una sorgente e la superficie dell'oggetto illuminato. L'illuminamento prodotto da un flusso luminoso di 1 Lumen che cade in modo uniforme su una superficie di 1m^2 è pari a 1 Lx (Lux). È chiaro quindi come questa unità di misura sia riferita all'oggetto illuminato e non alla sorgente.

L'illuminamento è dovuto all'illuminazione naturale e all'illuminazione artificiale. L'**illuminazione naturale**, ossia i valori rilevati nella serie che tratteremo di seguito, proviene dal sole (di cui l'irraggiamento ne è l'unità di misura), dalla volta celeste e dagli oggetti che la riflettono; essa presenta una luminanza variabile in funzione delle condizioni meteorologiche, dell'ora e del giorno

dell'anno considerati. Di conseguenza, l'illuminazione naturale all'interno di un ambiente varia non solo in relazione al punto in cui si valuta ma anche nel tempo. L'**illuminazione artificiale** viene emessa da sorgenti artificiali, ossia apparecchiature che generalmente sfruttano l'energia elettrica. Le tipologie di sorgenti luminose artificiali possono essere classificate in lampade a incandescenza, lampade a scarica in gas e lampade a induzione. A differenza dell'illuminazione naturale, le sorgenti artificiali mantengono costante nel tempo la loro luminanza.

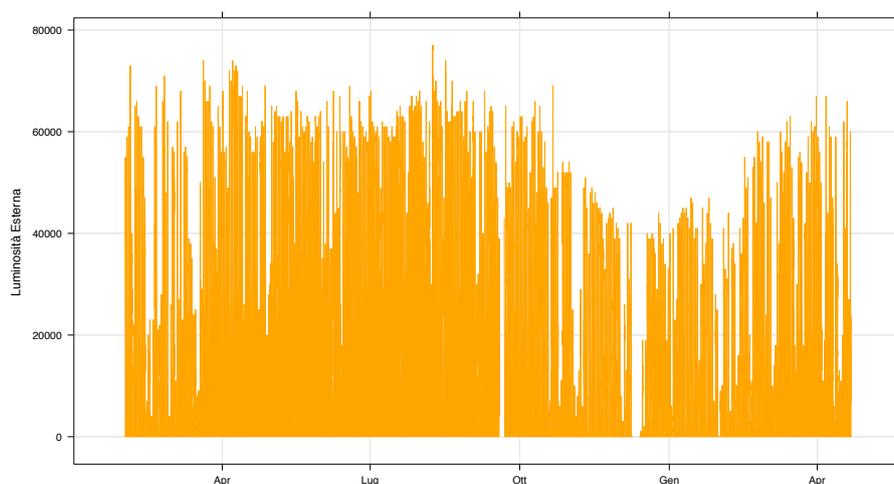


Figura 5.11: *Luminosità rilevata*

Tra i valori rilevati nel corso del periodo, rappresentati in Figura 5.11, risaltano in particolar modo i valori massimi poiché gli altri valori, essendo contenuti all'interno, sono difficilmente identificabili ad un primo impatto. Per descrivere meglio la struttura dei dati, è stata svolta un'analisi di tipo deterministico. L'analisi ha permesso di determinare come le diverse componenti contribuiscono a determinare il valore osservato. L'esito dell'analisi è rappresentato in Figura 5.12.

Come nella serie precedente, anche qui troviamo una marcata significatività della componente stagionale. Tra i residui spiccano alcuni picchi negativi, dovuti a picchi nei valori rilevati in giornate non completamente soleggiate. Inoltre bisogna considerare che durante la notte i valori della luminosità, a meno della

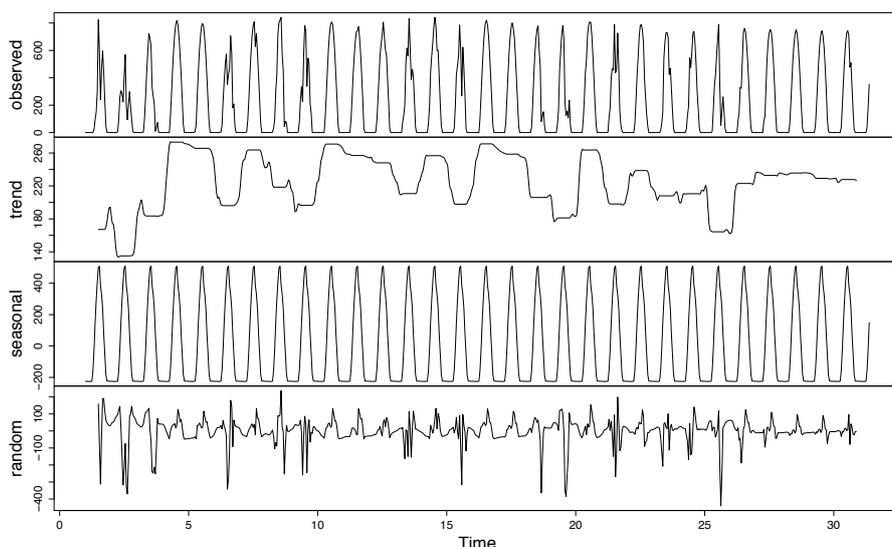


Figura 5.12: *Analisi deterministica sulla serie della luminosità esterna*

presenza di fonti di illuminazione artificiale, sono pari a zero e quindi il trend è influenzato anche da questi valori. Potrebbe essere studio di analisi futura, l'analisi della serie tralasciando i valori notturni ed i problemi correlati alla decisione di quando interrompere e riprendere la serie affinché l'analisi risulti coerente con il problema.

Per la determinazione degli ordini del modello ARIMA (o SARIMA eventualmente) sono stati utilizzati i correlogrammi globali e parziali dei dati e il diagramma di autodispersione dei dati per evidenziare la presenza di dipendenza rispetto ai valori passati Figura 5.13. Per la determinazione degli ordini della parte stagionale del modello sono stati usati i correlogrammi della serie differenziata tante volte quanto la frequenza delle osservazioni.

Sulla base delle informazioni a disposizione, per la luminosità esterna è stato adattato un modello $SARIMA(1, 1, 1)(2, 0, 2)_{24}$. La stima dei parametri del modello viene rappresentata nella Tabella 5.5.

Il parametri del modello stimato risultato tutti significativi ed anche la diagnostica del modello conferma la bontà dello stesso. Infatti i residui del modello non sembrano evidenziare la presenza di alcuna forma di dipendenza e il te-

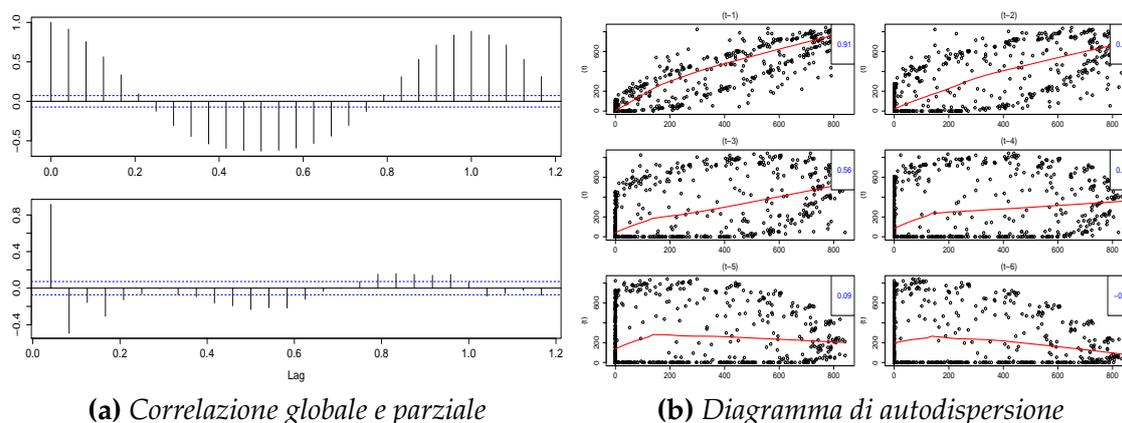


Figura 5.13: Analisi preliminare sulla serie della luminosità esterna

Coefficients:						
	ar1	ma1	sar1	sar2	sma1	sma2
	0.6447	-1.0000	0.4798	0.5192	-0.3628	-0.5458
Std. Err.	0.0240	0.0013	0.0069	0.0070	0.0287	0.0307
Error Measures:						
	ME	RMSE	MAE	MPE	MAPE	MASE
	3.9547	70.1960	36.6816	-	Inf	0.5093

Tabella 5.5: Sintesi del modello stimato sulla luminosità esterna

st di Ljung-Box conferma l'indipendenza degli stessi rispetto i valori precedenti
 Figura 5.14.

In Figura 5.14 viene anche rappresentata la previsione dei valori della luminosità esterna. Tali valori sono riportati anche nella Tabella 5.6 per il confronto con i valori realmente osservati. Il confronto mette in risalto la bontà della previsione che, almeno fino al quinto periodo in avanti, risulta molto accurata.

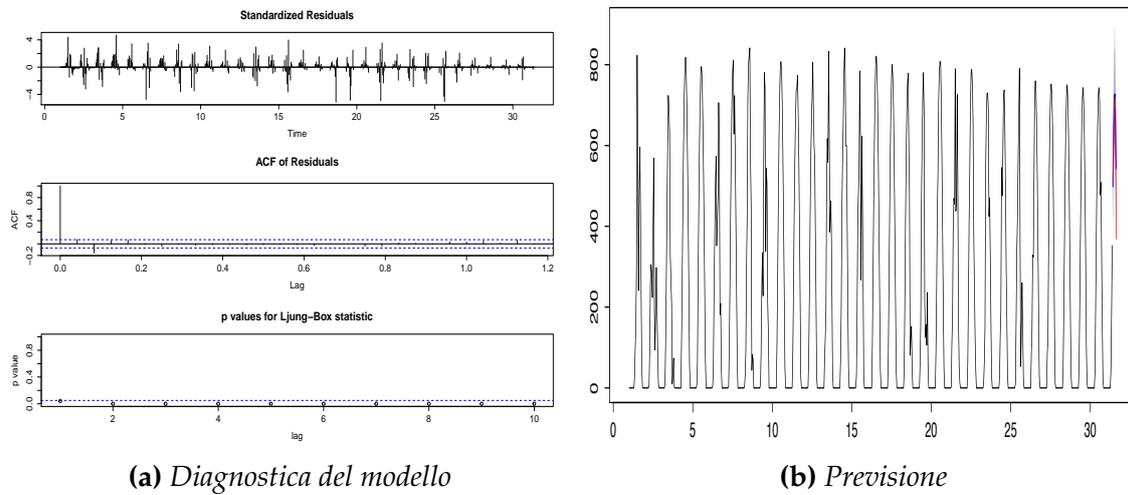


Figura 5.14: Diagnostica del modello e previsione per la serie della luminosità esterna

Tempo	Previsione	Osservato	Differenza
t+1	498.50	508.26	9.77
t+2	614.58	630.48	15.90
t+3	714.77	692.52	22.25
t+4	727.42	715.79	11.64
t+5	625.94	690.39	64.45
t+6	543.59	368.98	174.61

Tabella 5.6: Confronto tra valori osservati e stimati della luminosità esterna

5.1.4 L'umidità esterna

L'aria che ci circonda è una miscela di aria secca e vapore acqueo e la quantità di acqua presente nell'aria si indica col termine umidità. La quantità massima di umidità che può essere contenuta nell'aria dipende dalla temperatura, infatti decresce al diminuire della temperatura e diventa quasi nulla a $-40\text{ }^{\circ}\text{C}$.

L'umidità si può esprimere attraverso tre grandezze:

- Umidità assoluta, esprime il rapporto tra la massa di vapore acqueo ed il volume dell'aria $[g/m^3]$;
- Umidità specifica, esprime il rapporto tra la massa di vapore acqueo e la massa d'aria $[g/kg]$;
- Umidità relativa, esprime il rapporto tra la massa di vapore acqueo con-

tenuta nell'aria e quella massima che potrebbe contenere alla temperatura attuale [%].

I valori dell'umidità esterna rilevati nel periodo sono espressi come umidità relativa e vengono riportati in Figura 5.15. Da questa rappresentazione non si riesce a trarre molta informazione sulla serie, se non che la maggior parte dell'umidità rilevata è compresa tra il 20 e l'80%, e che nell'ultimo mese di rilevazione (settembre) tende ad assumere dei valori più elevati.

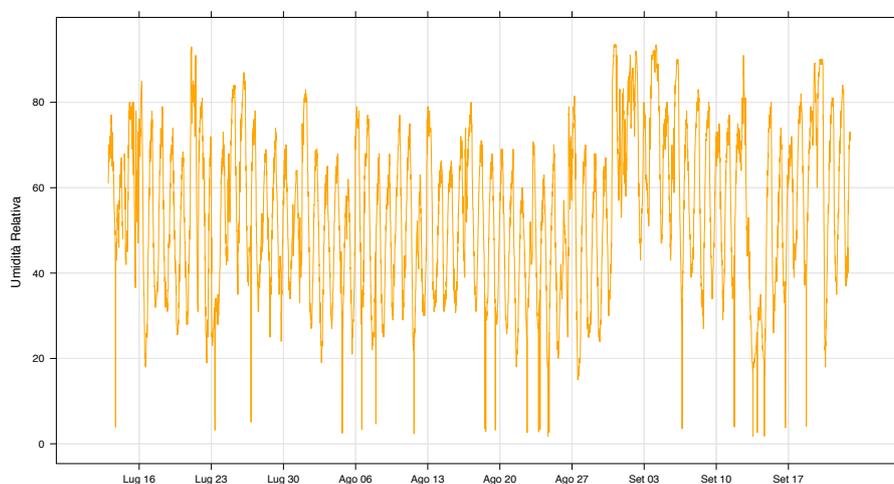


Figura 5.15: Umidità esterna rilevata

Per una miglior comprensione dei dati è stato intrapreso un approccio deterministico. Questo approccio è portato alla scomposizione della serie nelle componenti trend, stagionalità ed errore (Figura 5.16).

Dall'analisi emerge come la componente stagionale ottenuta assume un certo rilievo nella determinazione del valore, contribuendo fino ad un massimo del 30% al valore dell'umidità relativa rilevato. Quest'analisi è usata come punto di partenza per l'impostazione di un'analisi di tipo stocastico. Per una corretta individuazione del modello, sono state analizzate le funzioni di autocorrelazione globale e parziale delle serie rilevate e del diagramma di autodispersione (Figura 5.17). Inoltre, la valutazione della serie dopo averci applicato una differenzia-

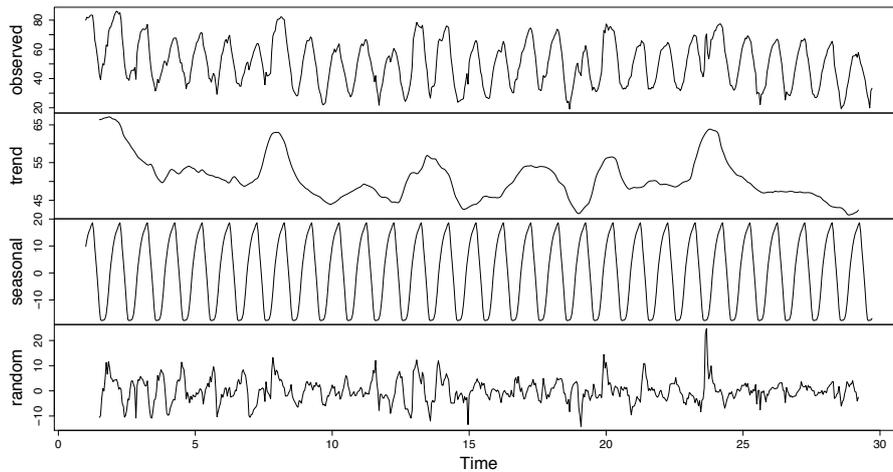
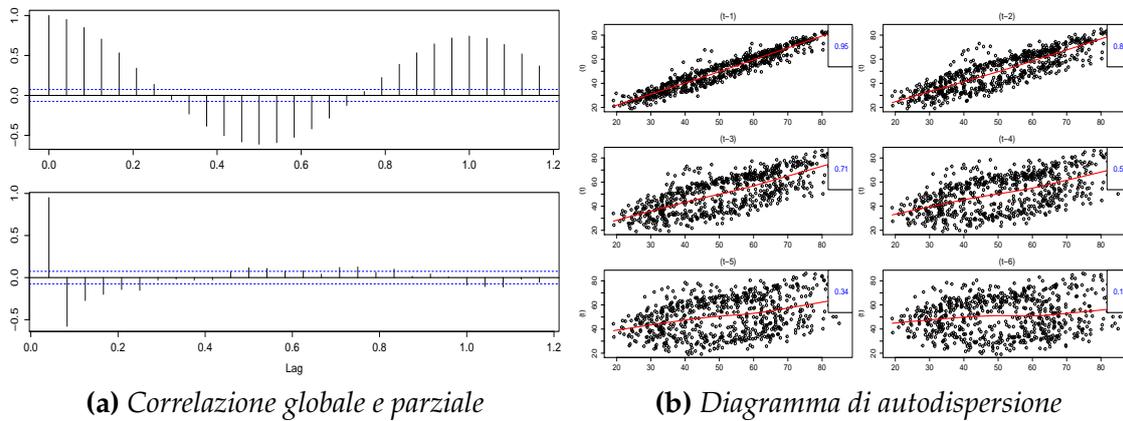


Figura 5.16: Analisi deterministica sulla serie dell'umidità esterna

zione di ordine 24, è stata effettuata per una corretta individuazione degli ordini della parte stagionale del modello.



(a) Correlazione globale e parziale

(b) Diagramma di autodispersione

Figura 5.17: Analisi preliminare sulla serie dell'umidità esterna

Sulla base di ciò, abbiamo deciso di adattare alla serie dell'umidità esterna un modello $SARIMA(2, 0, 2)(2, 0, 2)_{24}$. I parametri stimati sono rappresentati nella Tabella 5.7.

La diagnostica del modello ha evidenziato l'assenza di autocorrelazione tra i residui mentre, per quanto l'indipendenza degli stessi, l'andamento del p-value del test di Box-Ljung assume un andamento decrescente all'aumentare del numero di ritardi (Figura 5.18).

Coefficients:

	ar1	ar2	ma1	ma2	sar1	sar2	sma1	sma2	intercept
	-0.0751	0.9028	1.165	0.1650	0.3182	0.6815	-0.2462	-0.7191	46.2247
Std. Err.	0.0028	0.0041	0.038	0.0379	0.1426	0.1426	0.1535	0.1504	8.9980

Error Measures:

	ME	RMSE	MAE	MPE	MAPE	MASE
	-0.2267	2.9714	2.008	-1.0706	4.5419	0.5434

Tabella 5.7: Sintesi del modello stimato sull'umidità esterna

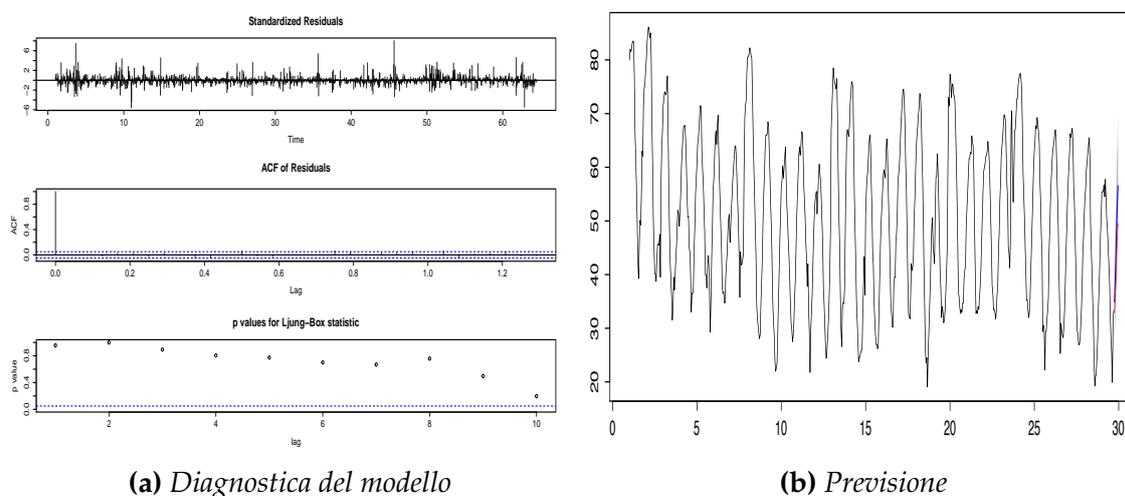


Figura 5.18: Diagnostica del modello e previsione per la serie dell'umidità esterna

Le previsioni effettuate sulla base del modello stimato sono rappresentate in Figura 5.18 ed i valori sono riportati anche nella Tabella 5.8, dove vengono confrontati con i valori effettivamente rilevati.

Tempo	Previsione	Osservato	Differenza
t+1	34.90	32.90	2.01
t+2	38.73	36.00	2.73
t+3	43.05	38.75	4.30
t+4	48.58	43.75	4.83
t+5	53.00	48.00	5.00
t+6	56.50	49.50	7.00

Tabella 5.8: Confronto tra valori osservati e stimati dell'umidità esterna

5.2 Le variabili del sistema e le loro interrelazioni

Dopo aver modellato le variabili climatiche, è possibile ora fornire una modellazione delle variabili che sono strettamente collegate al sistema HVAC. LA maggior parte di queste variabili sono delle variabili di risposta del sistema e pertanto non direttamente controllabili ma, assodato che l'ambiente non reagisce istantaneamente alle modifiche del sistema HVAC, alcune di queste sono modellabili sulla base dei valori passati, eventualmente integrati con delle informazioni sulle variabili climatiche.

Il set dati a disposizione non include tutte le variabili del problema presentate nel Capitolo 2, ma solo un sottoinsieme di esse. Si cercherà quindi di elaborare un primo approccio all'ottimizzazione del sistema HVAC sulla base di quanto a nostra disposizione.

Prima di procedere alla modellazione, un'analisi delle relazioni esistenti tra le variabili risulta fondamentale al fine di garantire la comprensibilità e accuratezza dei modelli che saranno stimati.

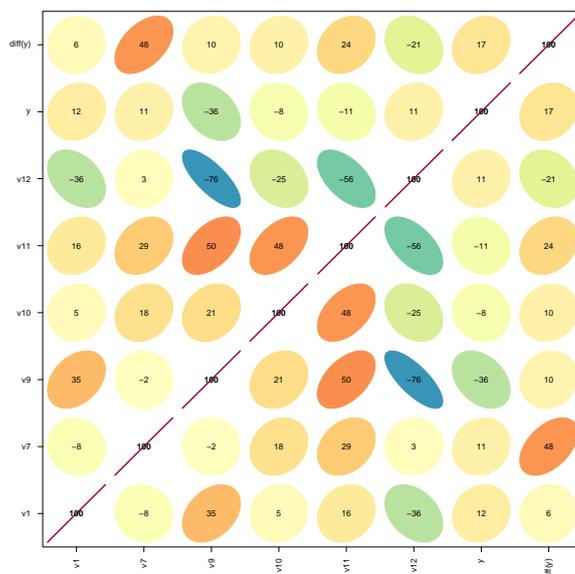


Figura 5.19: *Matrice di correlazione delle variabili*

In Figura 5.19 viene rappresentata graficamente la matrice di correlazione del-

le variabili. Le correlazioni sono espresse in forma percentuale, in base alla forma della figura in cui sono contenute (una figura sferica corrisponde ad una correlazione pari a 0, mentre all'assottigliarsi della sfera la correlazione cresce fino al raggiungimento della forma rettilinea) ed in base alla tonalità del colore. La figura evidenzia alcune peculiarità delle variabili, quali per esempio la forte correlazione tra l'umidità e la temperatura esterna oppure la relazione tra temperatura interna ed esterna.

5.3 Modelli previsivi per le variabili climatiche interne

Un primo studio di modellazione viene svolto sulla temperatura interna, variabile che influisce direttamente sul comfort termico. I valori rilevati sono rappresentati in Figura 5.20, evidenziando un andamento ripetitivo nell'arco della giornata poiché ci aspettiamo che durante la notte la temperatura scenda di qualche grado, non necessitando di una temperatura particolarmente confortevole.

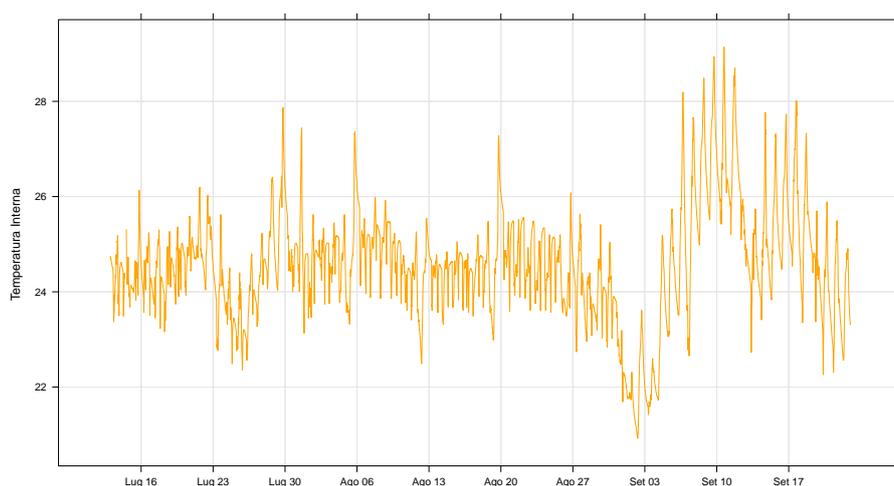


Figura 5.20: *Temperatura interna rilevata*

Una primo approccio deterministico viene impostato per indagare le caratteristiche delle componenti della serie storica. Viene evidenziata in questo modo

la ricorrenza stagionale nell'arco della giornata, emersa anche nella mera rappresentazione grafica della serie. L'output dell'analisi (Figura 5.21) evidenzia inoltre l'assenza di un andamento di fondo costante.

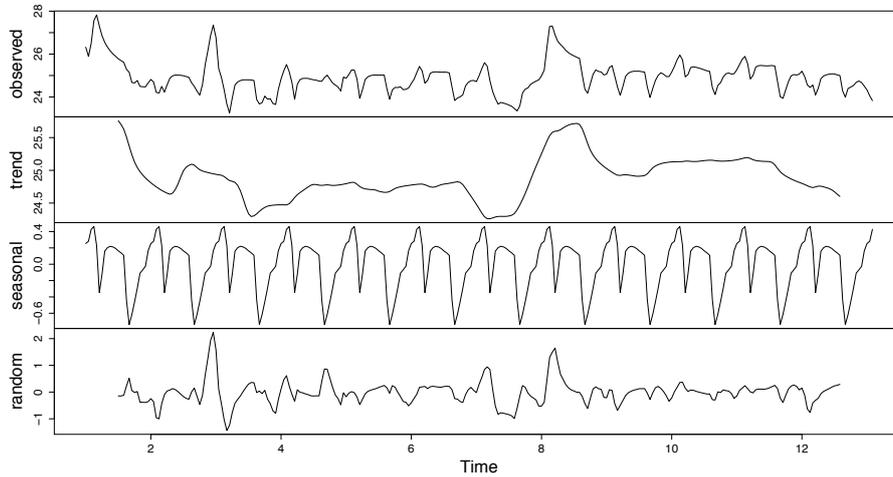
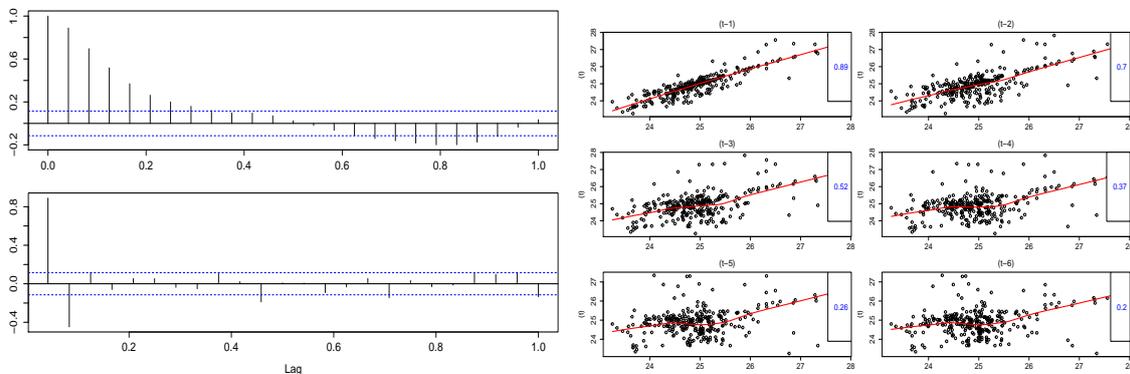


Figura 5.21: *Analisi deterministica sulla serie della temperatura interna*

Data la natura dei sistemi HVAC negli edifici, la temperatura interna è influenzata da molti fattori quali la velocità dell'aria, il valore dell'umidità, la temperatura esterna ed il comando imposto al sistema al tempo corrente e passato. Poiché è collegata anche ai propri valori passati (Figura 5.22), si è impostato una modellazione di tipo autoregressivo integrato a media mobile con variabili esogene ARIMAX (Capitolo 3).



(a) *Correlazione globale e parziale*

(b) *Diagramma di autodisperzione*

Figura 5.22: *Analisi preliminare sulla serie della temperatura interna*

Sulla base delle funzioni di autocorrelazione globale e parziale della serie e della matrice di autocorrelazione delle variabili a disposizione, il modello applicato è un $ARIMAX(2, 0, 2)(1, 0, 1)_{24}\Phi(v_{9t}, v_{9t-1}, v_{9t+1}, v_{10t}, v_{10t-1}, v_{10t+1})$; stime ed errori dei parametri sono rappresentati nella Tabella 5.9.

Coefficients:

	ar1	ar2	ma1	ma2	sar1	sma1	intercept	$v_9(t)$	$v_9(t-1)$	$v_9(t+1)$	$v_{12}(t)$	$v_{12}(t-1)$	$v_{12}(t+1)$
	1.20653	-0.31858	0.33784	0.08331	0.96379	-0.76987	21.44777	0.08137	0.03680	0.05024	0.00033	0.00085	-0.00052
Std. Err.	0.18974	0.17276	0.19192	0.12143	0.02617	0.08198	0.83533	0.02145	0.01922	0.01978	0.00022	0.00025	0.00025

Error Measures:

	ME	RMSE	MAE	MPE	MAPE	MASE
	-0.0104	0.2154	0.1512	-0.0491	0.6082	0.6428

Tabella 5.9: Sintesi del modello stimato sulla temperatura interna

Nella definizione del modello vengono utilizzate le informazioni della serie agli istanti precedenti unitamente ai valori di temperatura e umidità esterna all'istante precedente, corrente e futuro. Tramite l'analisi diagnostica (Figura 5.23) si evidenzia come il modello non riesce a catturare tutta l'informazione contenuta, infatti l'indipendenza dei residui è garantita solo dopo un certo numero di ritardi. La mancata modellazione di questa informazione può essere dovuta all'assenza di alcune variabili utili al modello nel set di dati a disposizione.

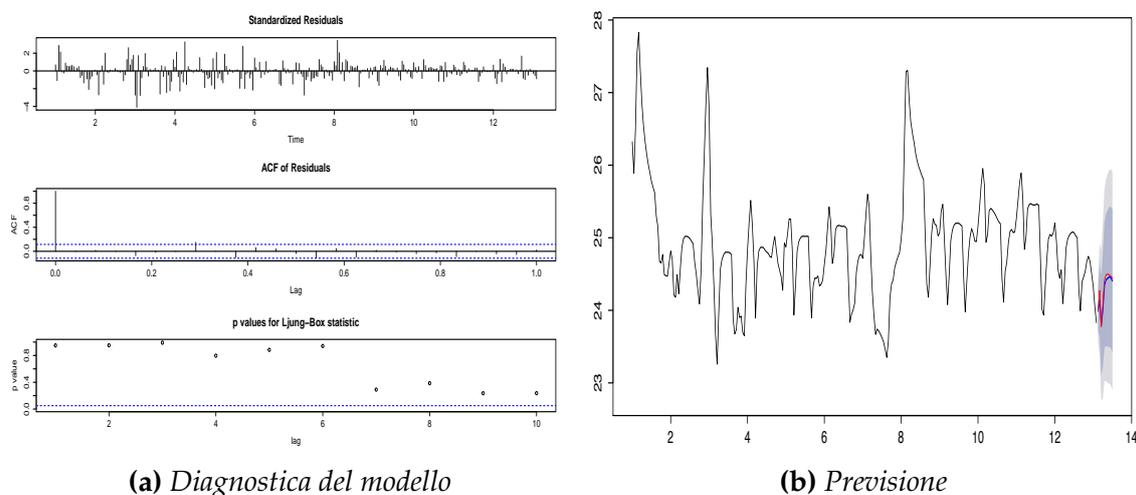


Figura 5.23: Diagnostica del modello e previsione per la serie della temperatura interna

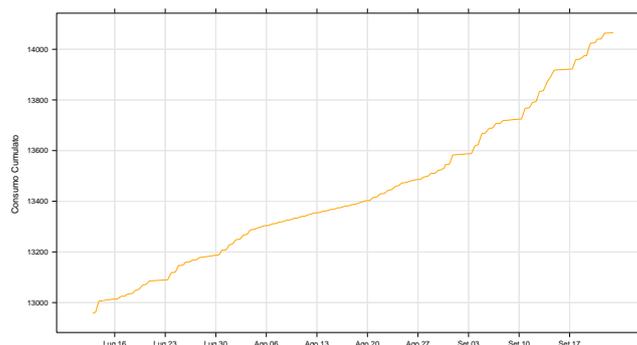
Il modello, pur basandosi su un sottoinsieme delle variabili del sistema, permette di ottenere delle buone previsioni della temperatura interna. In Figura 5.23 è rappresentata la rappresentazione grafica delle previsioni, riportata anche in Tabella 5.10 e confrontata con i valori realmente osservati. Il confronto evidenzia come lo scarto massimo tra il valore previsto e quello osservato sia nettamente inferiore al grado centigrado.

Tempo	Previsione	Osservato	Differenza
t+1	24.00	23.98	0.02
t+2	24.14	24.27	0.13
t+3	23.79	23.77	0.02
t+4	24.00	24.07	0.06
t+5	24.34	24.42	0.07
t+6	24.42	24.49	0.07
t+7	24.44	24.50	0.06
t+8	24.46	24.49	0.03
t+9	24.45	24.47	0.02
t+10	24.41	24.44	0.03

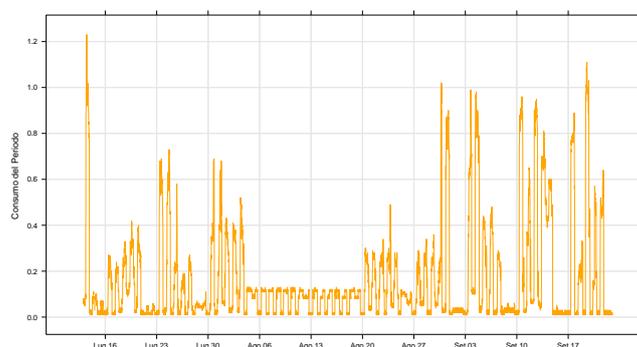
Tabella 5.10: Confronto tra valori osservati e stimati della temperatura interna

5.4 Modelli previsivi per l'ottimizzazione del consumo di energia

Il consumo di energia del sistema HVAC è la variabile più importante da modellare, in quanto una sua modellazione ci permette di capire come varia il consumo del sistema in base alle diverse configurazioni delle variabili. Come già detto in questa sede, non abbiamo a disposizione l'intero set di variabili ma solo un sottoinsieme di esse, per cui si imposterà una prima modellazione del consumo in base a quanto in nostro possesso. Le rilevazioni del consumo totale di energia sia cumulato che riferite ad ogni intervallo di rilevazione sono presentate in Figura 5.24.



(a) Consumo totale cumulato



(b) Consumo totale in ogni intervallo

Figura 5.24: Rilevazioni del consumo del sistema HVAC

Dalla natura del consumo di energia, è ragionevole aspettarsi che sia legato ai comandi imposti alla centrale termica, allo stato attuale del sistema ed alle condizioni climatiche. In base alle relazioni tra le variabili analizzate in Figura 5.19, il modello presentato si basa sui valori passati del consumo unitamente alle variabili climatiche esterne di temperatura, illuminazione e umidità.

Per la determinazione di quanti valori passati considerare abbiamo fatto affidamento alla rappresentazione grafica delle funzione di autocorrelazioni globale e parziale unitamente al grafico di auto-dispersione (Figura 5.25). Si è deciso quindi di utilizzare i valori relativi ai due istanti precedenti e il valore osservato il giorno precedente allo stesso istante.

Un $ARIMAX(2, 1, 2)(1, 0, 1)_{24}\Phi(v_9(t-1), v_9(t), v_9(t+1), v_{11}(t-1), v_{11}(t), v_{11}(t+1), v_{12}(t-1), v_{12}(t), v_{12}(t+1))$ è stato applicato ai dati e le stime dei parametri sono riportate nella Tabella 5.11.

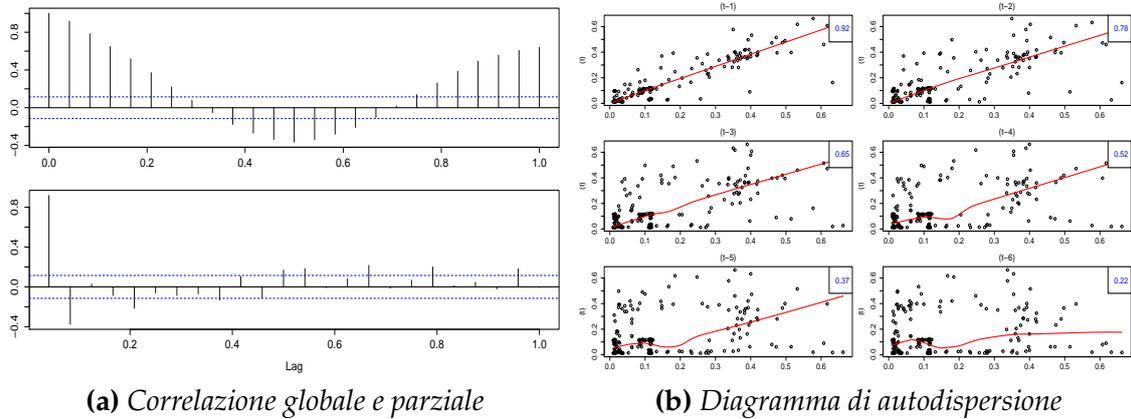


Figura 5.25: Analisi preliminare sulla serie del consumo

Coefficients:															
ar1	ar2	ma1	ma2	sar1	sma1	v9 _t	v9 _{t-1}	v9 _{t+1}	v11 _t	v11 _{t-1}	v11 _{t+1}	v12 _t	v12 _{t-1}	v12 _{t+1}	
-0.41122	-0.61445	0.73026	0.77307	0.90197	-0.68221	-0.00309	0.00432	-0.00442	0.00004	0.00003	0.00002	0.00005	0.00000	0.00003	
Std. Err. 0.12391 0.13928 0.13074 0.07727 0.05193 0.08963 0.00299 0.00331 0.00362 0.00008 0.00007 0.00008 0.00011 0.00009 0.00010															
Error Measures:															
ME	RMSE	MAE	MPE	MAPE	MASE										
-0.0005	0.0454	0.0256	-1.7537	43.1672	1.0389										

Tabella 5.11: Sintesi del modello stimato sul consumo del sistema HVAC

Il consumo non dipende dalle sole variabili incluse, ma ci si può aspettare dipenda anche dal comando inviato alla centrale termica, alla velocità dell'aria ed all'umidità interna, in quanto sono i fattori che influenzano direttamente il raggiungimento del comfort abitativo e quindi del consumo connesso. Poiché non avevamo a disposizione tutte le variabili di interesse, il modello stimato non riesce a cogliere tutta l'informazione evidenziando quindi nei residui di tale modello la presenza di relazione/correlazione ignota. Questo viene evidenziato nel grafico della statistica di Ljung-Box sui residui, dove per ogni ritardo i valori del *p-value* sono superiori alla soglia che garantisce l'indipendenza dei residui (Figura 5.26).

Il confronto tra i valori futuri previsti e quelli realmente osservati viene riportato nella Tabella 5.12. Le previsioni effettuate non risultano così lontane dai valori realmente osservati, tuttavia data la mancanza considerazione di tutte le varia-

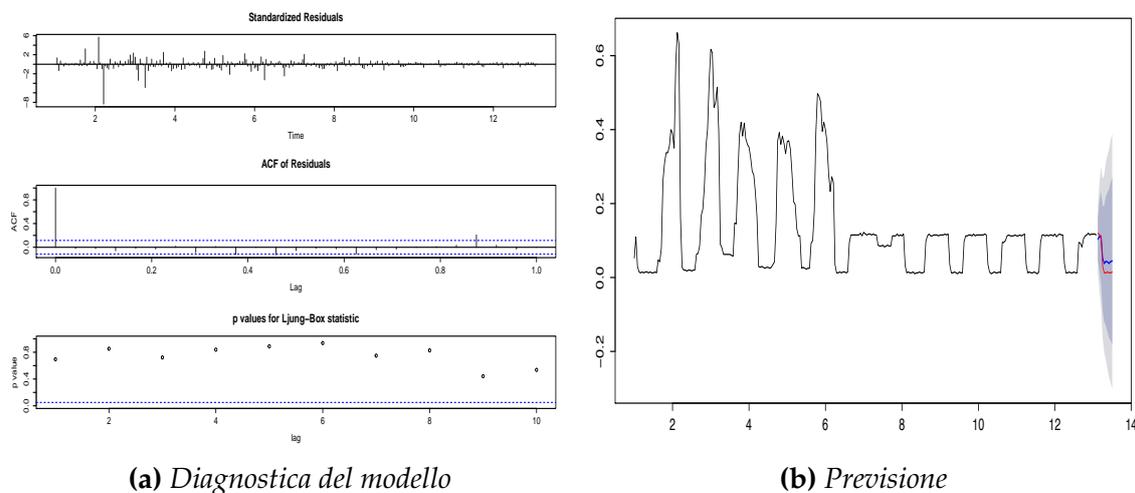


Figura 5.26: Diagnostica del modello e previsione per il consumo del sistema HVAC

bili che contribuiscono alla determinazione del consumo, è opportuno valutare e trattare tali risultati con cautela.

Tempo	Previsione	Osservato	Differenza
t+1	0.1035	0.1200	0.0165
t+2	0.1111	0.1150	0.0039
t+3	0.1086	0.1125	0.0039
t+4	0.0515	0.0300	0.0215
t+5	0.0376	0.0125	0.0251
t+6	0.0439	0.0125	0.0314
t+7	0.0420	0.0150	0.0270
t+8	0.0380	0.0125	0.0255
t+9	0.0432	0.0125	0.0307
t+10	0.0447	0.0150	0.0297

Tabella 5.12: Confronto tra valori osservati e stimati del consumo

5.5 Modelli previsivi per l'ottimizzazione dell'emissione di biossido di carbonio

L'anidride carbonica (o biossido di carbonio) è la sostanza risultante della combustione di un composto organico in presenza di una quantità di ossigeno suf-

ficiente a completarne l'ossidazione⁶. Nel caso specifico di un sistema HVAC, è il risultato della combustione di petrolio e carbone fossile prodotta dalla centrale termica. Le rilevazioni delle emissioni di anidride carbonica del periodo sono riportate in Figura 5.27.

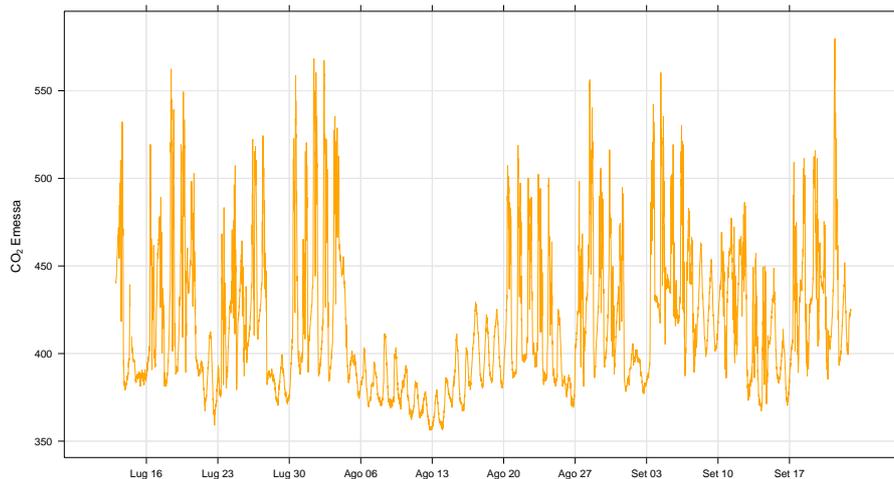


Figura 5.27: Rilevazioni dell'anidride carbonica emessa dal sistema HVAC

Si tratta di una variabile strettamente legata al consumo di energia della centrale centrale (y_{load}), variabile attualmente non a disposizione. Per ovviare a questo problema e cercando altresì la formulazione di un modello la variabile delle emissioni di CO_2 è stata relazionata al consumo totale di energia (y), poiché questa variabile comprende il consumo di energia della centrale termica che ne costituisce la maggior parte.

Per indagare la struttura della serie è stata svolta una prima analisi di tipo descrittivo. Tale analisi, rappresentata in Figura 5.28, ha evidenziato una certa ricorrenza nei valori osservati, infatti è presumibile che nei periodi notturni il funzionamento del sistema sia minimo.

Per la valutazione dei ritardi della serie da considerare, sono stati analizzate le rispettive funzioni di autocorrelazione globale e parziale (Figura 5.29).

⁶Fonte: Ministero dell'Ambiente e della Tutela del Territorio e del Mare www.minambiente.it

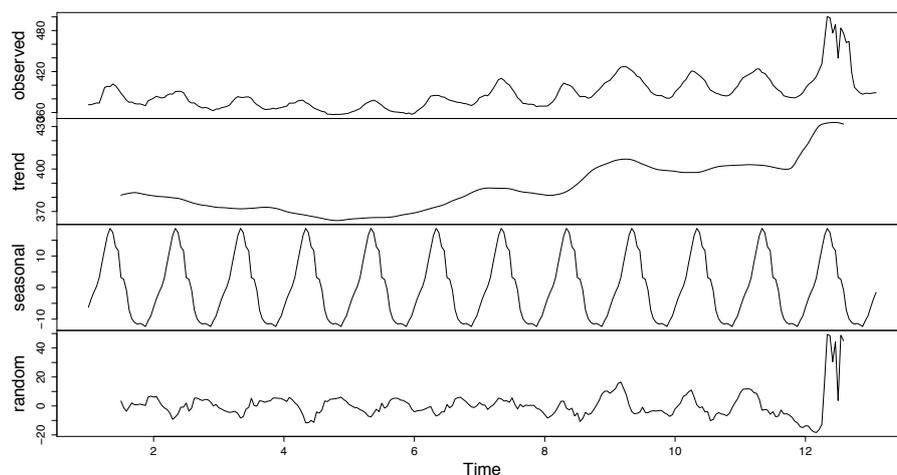
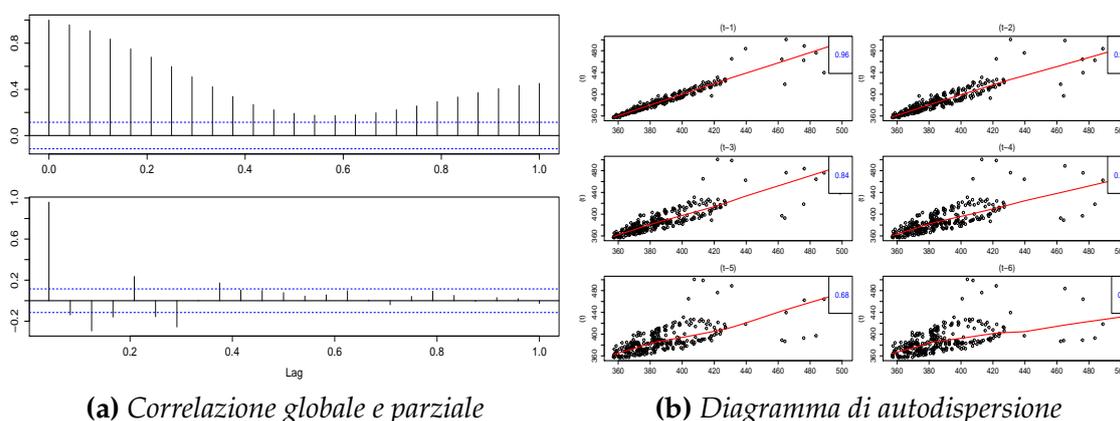


Figura 5.28: Analisi deterministica sulla serie della CO2 emessa



(a) Correlazione globale e parziale

(b) Diagramma di autodispersione

Figura 5.29: Analisi preliminare sulla serie della CO₂ emessa dal sistema

Nel modellazione è stata considerata anche la serie del consumo totale di energia del sistema HVAC, dopo essere stata opportunamente differenziata per poter disporre in tal modo del consumo riferito all'intervallo. Risulterebbe infatti difficile modellare le emissioni di anidride carbonica nell'intervallo considerando il consumo di energia totale cumulato. Per tale motivo, d'ora in poi si farà riferimento alla serie del consumo differenziato e riferito quindi ad ogni intervallo di rilevazione.

Il modello utilizzato per la modellazione delle emissioni di anidride carbonica è un $ARIMAX(1, 0, 1)(1, 0, 1)_{24}\Phi(y(t), y(t - 1))$, e le stime dei parametri sono riportate

nella Tabella 5.13.

Coefficients:							
	ar1	ma1	sar1	sma1	intercept	$y(t)$	$y(t - 1)$
	0.9421	-0.0540	0.8964	-0.6341	375.9195	140.0964	25.3994
Std. Err.	0.0203	0.0529	0.0567	0.1258	13.0411	19.0478	18.5536
Error Measures:							
	ME	RMSE	MAE	MPE	MAPE	MASE	
	0.1710	5.7202	3.0327	0.0267	0.7529	0.8554	

Tabella 5.13: Sintesi del modello stimato sulle emissioni di CO_2

La modellazione delle emissioni, pur basandosi sulle rilevazioni del consumo totale di energia e non al solo consumo di energia riferito alla centrale termica, risulta valida e permette di ottenere delle previsioni corrette. L'analisi diagnostica del modello conferma quanto detto, infatti i residui non evidenziano la presenza di correlazione rispetto ai valori passati (Figura 5.30); permane tuttavia una leggera dipendenza rispetto al valore osservato all'istante precedente, che il modello non è riuscito ad inglobare.

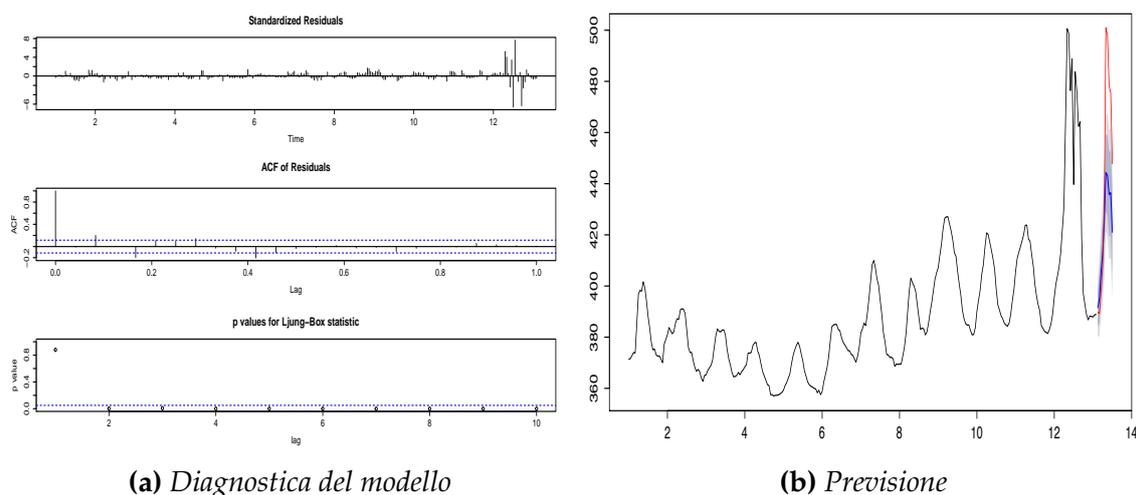


Figura 5.30: Diagnostica del modello e previsione per la serie della CO_2 emessa

Le previsioni effettuate sulla base del modello stimato risultano prossime ai valori osservati fino al quinto istante di previsione nel futuro, mentre per gli

istanti di previsione successivi tendono a divergere. La rappresentazione grafica della serie con i valori di previsione è riportata in Figura 5.30, mentre il confronto con i valori realmente osservati è visualizzato nella Tabella 5.14.

Tempo	Previsione	Osservato	Differenza
t+1	391.55	389.64	1.91
t+2	395.23	389.38	5.85
t+3	405.50	398.04	7.46
t+4	412.65	408.15	4.50
t+5	425.17	431.15	5.98
t+6	444.31	500.93	56.62
t+7	443.20	498.02	54.82
t+8	435.79	478.26	42.47
t+9	436.47	475.44	38.97
t+10	421.04	447.79	26.75

Tabella 5.14: Confronto tra valori osservati e stimati della CO_2 emessa

5.6 Il modello stimato per la realizzazione dell'ottimo abitativo

Sulla base dell'approccio di modellazione alle serie delle variabili a disposizione e presumendo di avere a disposizione anche le variabili mancanti, si propone ora una possibile formulazione pratica del modello e la relativa metodologia di soluzione.

La modellazione delle variabili climatiche permette di sfruttare i valori passati per stimare i possibili valori negli istanti futuri. Questo ha permesso di inglobare nella formulazione del problema valori passati e futuri delle variabili climatiche.

Per la corretta formulazione del problema è necessaria la determinazione dei modelli della temperatura, del consumo sostenuto dall'impianto termico e del consumo totale è necessaria. La modellazione della temperatura si può effettuare mediante l'estensione del modello precedentemente stimato, includendo anche le variabili velocità dell'aria e umidità interna. Per quanto riguarda la model-

lazione dei consumi (sia dovuti al solo impianto termico che totali) l'utilizzo di modelli auto-regressivi integrati a media mobile con variabili esogene potrebbe fornire degli ottimi risultati, sulla base di quanto evidenziato con le variabili a disposizione. Sono inoltre di interesse, per il mantenimento del comfort abitativo, le variabili PMV e DGI. Una volta determinati, mediante i modelli stimati, i valori agli istanti successivi, questi indici ci permettono di valutare se la possibile soluzione del modello garantisce la soddisfazione dei vincoli di comfort abitativo imposti.

La temperatura interna, sulla base di quanto evidenziato con le sole variabili a disposizione, si può modellare mediante un modello ARIMAX sulla base dell'intero dell'intero set di variabili. Potremmo esprimere il modello della temperatura come:

$$\hat{v}_1(t+1) = f_1(v_1(t), v_1(t-1), v_1(t-2), v_5(t), v_5(t-1), v_6(t), v_6(t-1), v_8(t), v_8(t-1), v_9(t-1), v_9(t), v_9(t+1), v_{10}(t-1), v_{10}(t), v_{10}(t+1), d_1(t), d_1(t+1), d_2(t), d_2(t+1), b(t), b(t+1), f_c(t), f_c(t+1))$$

Le variabili di comfort, già introdotte nel Capitolo 2, permettono di sintetizzare mediante un indice il comfort abitativo percepito dall'individuo. Queste variabili assumono un ruolo rilevante nella formulazione del problema in quanto sono soggette a vincolo, al fine di garantire un comfort minimo. Tuttavia, le rispettive funzioni analitiche(18), pur essendo ben definite, risultano complesse e richiedono un numero elevato di parametri per produrre un risultato valido.

Considerando il DGI (Daylight Glare Index), la rispettiva funzione analitica è definita come:

$$DGI = 10 \log \sum_{i=1}^n 0.478 \frac{L_s^{1.6} \Omega^{0.8}}{L_b + (0.07 \omega^{0.5} L_s)}$$

dove

- L_s = luminanza della fonte di luce;
- L_b = luminanza media delle superfici interne dell'ambiente che rientrano nel campo visivo dell'occupante;
- Ω = angolo solido sotteso dalla fonte di luce;

- ω = angolo solido totale sotteso dalla finestre;
- n = numero delle fonti di luce;

Per quanto riguarda il PMV (Predicted Mean Vote), la formula di calcolo si può esprimere come:

$$PMV = (0.303e^{-2.1 \times M} + 0.028) \times [(M - W) - H - E_c - C_{res} - E_{res}]$$

dove

- M = dispendio metabolico;
- W = lavoro;
- H = scambio di calore dovuto al contatto;
- E_c = scambio di calore dovuto all'evaporazione sulla pelle;
- C_{res} = scambio di calore dovuto alla convezione respiratoria;
- E_{res} = scambio di calore dovuto all'evaporazione respiratoria.

A sua volta, questi fattori sono influenzati da altri parametri, quali l'isolamento dell'abbigliamento, la temperatura dell'aria, la temperatura media radiante, la velocità relativa dell'aria, la pressione parziale di vapore acqueo e la temperatura superficiale dell'abbigliamento. Per una corretta modellazione di queste variabili sarebbe quindi necessario disporre di un numero notevolmente superiore di parametri.

Un approccio alternativo alla modellazione di queste variabili viene fornito da S. Atthajariyakul(2) nel suo lavoro di ricerca dove propone, con successo, una modellazione basata sulle reti neurali. Prendendo in considerazione l'indice PMV, nel lavoro citato esso viene legato a sei variabili facilmente reperibili, quali la temperatura dell'aria, la temperatura dell'aria a bulbo umido, la temperatura del pavimento, la velocità dell'aria, l'isolamento fornito dall'abbigliamento e l'attività umana.

Si possono quindi modellare le variabili PMV e DGI utilizzando una modellazione di questo tipo. La formulazione del modello si può esprimere come:

$$\widehat{PMV}(t+1)_{neural} = f_2(v_1(t-1), v_1(t), \hat{v}_1(t+1), v_4(t-1), v_4(t), v_3(t-1), v_3(t), \\ v_2(t-1), v_2(t), b(t), b(t+1), f_c(t), f_c(t+1))$$

$$\widehat{DGI}(t+1)_{neural} = f_3(v_5(t-1), v_5(t), v_6(t-1), v_6(t), v_{11}(t-1), v_{11}(t), v_{11}(t+1), \\ d_1(t), d_1(t+1), d_2(t), d_2(t+1), b(t), b(t+1))$$

Per quanto riguarda la modellazione delle variabili di risposta del sistema, si propone un modello ARIMAX, in quanto ha evidenziato dei buoni risultati con le variabili attualmente a disposizione e promette quindi di fornire dei risultati ancora migliori con la disponibilità del set completo di variabili. La modellazione si può esprimere come:

$$\hat{y}_{LOAD}(t+1) = f_4(y_{LOAD}(t-1), y_{LOAD}(t), v_1(t-1), v_1(t), \hat{v}_1(t+1), v_9(t-1), v_9(t), v_9(t+1), \\ v_{10}(t-1), v_{10}(t), v_{10}(t+1), f_c(t), f_c(t+1), b(t), b(t+1))$$

$$\hat{y}(t+1) = f_5(y_{LOAD}(t-1), y_{LOAD}(t), \hat{y}_{LOAD}(t+1), y(t-1), y(t), d_1(t), d_1(t+1), d_2(t), d_2(t+1), \\ b(t), b(t+1), v_{11}(t-1), v_{11}(t), v_{11}(t+1), v_5(t), v_5(t-1), v_6(t), v_6(t-1))$$

L'idea è quella di minimizzare il consumo dovuto all'impianto termico, valutando l'impatto che le variabili controllabili hanno nei periodi futuri; operando in questo modo, per ogni istante si applicherà la combinazione che garantisce il consumo minimo ma altresì la soddisfazione del comfort anche negli istanti successivi. Il problema ottenuto è vincolato, in quanto ci sono i vincoli di comfort oltre che di consumo massimo totale da rispettare, e mono-obiettivo dato che l'obiettivo consiste nella minimizzazione del consumo dovuto alla centrale termica.

Il problema di ottimizzazione si può formulare come:

$$\begin{aligned}
 & \min_{d_1(t+1), d_2(t+1), b(t+1), f_c(t+1)} y_{LOAD}(t+1) \\
 & \text{subject to:} \\
 & \hat{v}_1(t+1) = f_1(v_1(t), v_1(t-1), v_1(t-2), v_1(t-24), v_9(t-1), v_9(t), v_9(t+1), v_{10}(t-1), v_{10}(t), v_{10}(t+1), d_1(t), d_1(t+1), \\
 & \quad d_2(t), d_2(t+1), b(t), b(t+1), f_c(t), f_c(t+1), v_5(t), v_5(t-1)v_6(t), v_6(t-1), v_8(t), v_8(t-1)) \\
 & \widehat{PMV}(t+1) = f_2(v_1(t-1), v_1(t), \hat{v}_1(t+1), v_4(t-1), v_4(t), v_3(t-1), v_3(t), v_2(t-1), v_2(t), b(t), b(t+1), f_c(t), f_c(t+1)) \\
 & \widehat{DGI}(t+1) = f_3(v_5(t-1), v_5(t), v_6(t-1), v_6(t), v_{11}(t-1), v_{11}(t), v_{11}(t+1), d_1(t), d_1(t+1), d_2(t), d_2(t+1), b(t), b(t+1)) \\
 & \hat{y}_{LOAD}(t+1) = f_4(y_{LOAD}(t-1), y_{LOAD}(t), v_1(t-1), v_1(t), \hat{v}_1(t+1), v_9(t-1), v_9(t), v_9(t+1), v_{10}(t-1), v_{10}(t), \\
 & \quad v_{10}(t+1), f_c(t), f_c(t+1), b(t), b(t+1)) \\
 & \hat{y}(t+1) = f_5(y_{LOAD}(t-1), y_{LOAD}(t), \hat{y}_{LOAD}(t+1), y(t-1), y(t), d_1(t), d_1(t+1), d_2(t), d_2(t+1), b(t), b(t+1), \\
 & \quad v_{11}(t-1), v_{11}(t), v_{11}(t+1), v_5(t), v_5(t-1), v_6(t), v_6(t-1)) \\
 & T \text{ Comfort Setpoint Lower} \leq \hat{v}_1(t+1) \leq T \text{ Comfort Setpoint Upper} \\
 & PMV \text{ Setpoint Lower} \leq \widehat{PMV}(t+1) \leq PMV \text{ Setpoint Upper} \\
 & DGI \text{ Setpoint Lower} \leq \widehat{DGI}(t+1) \leq DGI \text{ Setpoint Upper} \\
 & \hat{y}(t+1) \leq \text{ELP Max Setpoint}
 \end{aligned}$$

Il problema così formulato presenta le funzioni da stimare come vincoli. Il problema si può riformulare includendo le determinazioni delle previsioni della temperatura interna e delle variabili di comfort nella funzione obiettivo; in tal caso abbiamo un problema di ottimizzazione multi-obiettivo.

Il problema di ottimizzazione presenta una funzione obiettivo complessa che non si può minimizzare utilizzando le tecniche convenzionali che si basano sull'utilizzo di metodi di analisi del gradiente della funzione. Si rende necessario quindi l'impiego di un algoritmo di tipo evolutivo per poter risolvere il problema di ottimizzazione. In questo studio, si propone l'utilizzo di un algoritmo genetico (Capitolo 4).

L'implementazione di un algoritmo genetico richiede che l'insieme delle variabili da manipolare durante il processo di ottimizzazione sia codificato come una stringa di lunghezza fissa.

Ciascuna delle variabili controllate da manipolare durante l'ottimizzazione sono state mappate in una stringa binaria di lunghezza fissa (4 bit), in quanto il numero massimo di attuazioni considerando le singole variabili risulta pari a 9 e si può mappare con una sequenza di 4 bit. Per considerare anche i limiti inferiori e superiori delle attuazioni la mappatura si realizza tramite la seguente formula che ripartisce il valore del parametro rispetto al numero di valori che si possono

mappare con l'intera stringa di bit:

$$Gene = \frac{\beta - \beta_{min}}{\beta_{max} - \beta_{min}}(2^4 - 1)$$

Ad esempio, un valore pari a 2 per la variabile f_c sarebbe codificato come

$$Gene_{f_c} = \frac{2 - 0}{3 - 0}(2^4 - 1) = \frac{2}{3} \times 15 = 10 = \{1010\}$$

Infine i geni sono concatenati per formare una stringa binaria a 16 bit detto cromosoma. Ogni cromosoma rappresenta quindi una soluzione realizzabile per il sistema HVAC. Una possibile realizzazione del cromosoma e dei singoli geni che lo compongono viene raffigurata in Figura 5.31.

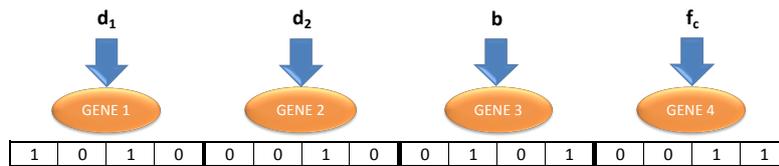


Figura 5.31: Possibile realizzazione del cromosoma

Alla codifica ottenuta viene applicata inoltre la codifica di Gray (Capitolo 4), in modo che il passaggio da un intero al successivo avvenga modificando un solo bit. Un esempio di tale codifica è riportato in Tabella 5.15.

Valore	Cod. Binaria	Cod. di Gray
0	0000	0000
1	0001	0001
2	0010	0011
3	0011	0010
4	0100	0110
5	0101	0111
6	0110	0101
7	0111	0100
8	1000	1100

Tabella 5.15: Codifica di Gray per una stringa di 4 bit

Il punto di partenza per la ricerca della migliore combinazione delle variabili è determinato generando in modo casuale una serie iniziale, o popolazione di stringhe di 16 bit. Ogni stringa binaria viene decodificata e i valori delle variabili risultanti vengono utilizzati nei modelli stimati per la determinazione e valutazione del consumo e degli indici di comfort abitativo. Questo processo viene ripetuto finché la bontà di ogni stringa binaria è stata determinata.

È stato dimostrato che la dimensione della popolazione dovrebbe essere tra i 25 ed i 100 individui(8). In questo caso abbiamo impostato una dimensione della popolazione pari a 50.

Non appena la popolazione iniziale o la nuova generazione è stata determinata, viene impiegato l'operatore di selezione per determinare quali individui e in quale quantità potranno partecipare al processo di creazione dei nuovi individui. Si utilizza il metodo di selezione a "Torneo " che prevede la selezione casuale di tre individui alla volta e la scelta individuo con il massimo valore della funzione di fitness(11). Successivamente viene applicato l'operatore di crossover con punto di taglio singolo, che viene determinato in modo casuale per tutti i cromosomi. La probabilità di crossover viene impostata all'80%.

Infine viene impiegato l'operatore di mutazione per fornire delle nuove informazioni alla popolazione ed impedire alla popolazione di diventare satura di cromosomi simili. Una posizione casuale del cromosoma viene selezionata e viene variata casualmente. Questo operatore non viene applicato a tutti le stringhe ma in modo casuale con una probabilità dell'10%.

Il processo di generazione viene ripetuto fino a quando il numero di generazioni supera il valore impostato (100 in questo caso) oppure fino al raggiungimento di un criterio di convergenza, quale il raggiungimento di un determinato valore da parte della fitness media della popolazione.

L'algoritmo genetico descritto è stato testato su una funzione matematica classica generalmente utilizzata per testare i sistemi di ottimizzazione. Per il test è

stata utilizzata la funzione multimodale Schwefel (Figura 5.32). Si tratta di una funzione ingannevole in quanto il minimo globale è geometricamente lontano, nello spazio dei parametri, dal minimo locale ottimo più vicino. Pertanto, gli algoritmi di ricerca sono potenzialmente inclini ad una convergenza nella direzione sbagliata.

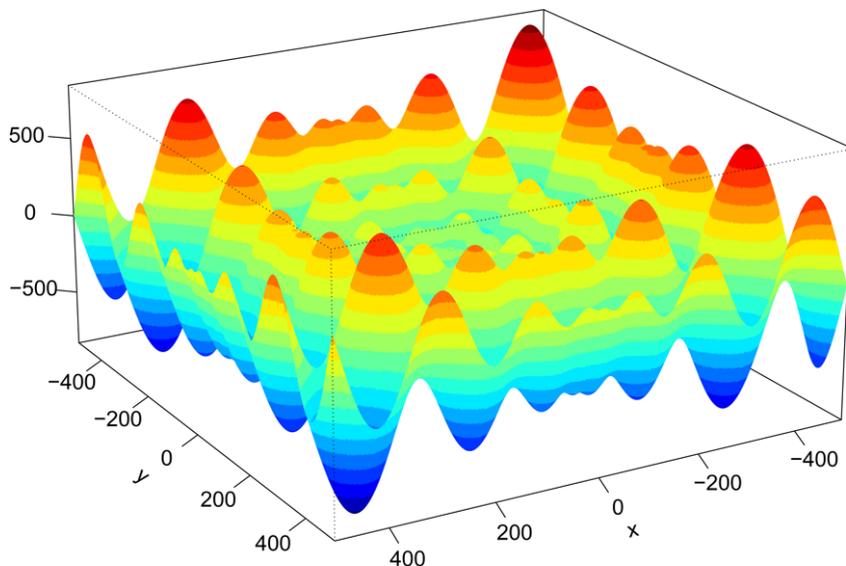


Figura 5.32: *Rappresentazione della funzione Schwefel per $n=2$*

La funzione si definisce come:

$$f(x) = \sum_{i=1}^n -x_i \cdot \sin\left(\sqrt{|x_i|}\right) \quad -500 \leq x_i \leq 500$$

Il minimo globale della funzione è pari a $f(x) = -n \cdot 418.9829$ e si ottiene nel punto $x_i = 420.9687$ dove $i = 1, 2, \dots, n$.

Al termine della 100-esima generazione l'algoritmo ha riportato -837.9629 come miglior valore della funzione nei punti $x_1 = 420.8833$, $x_2 = 420.8437$ mentre il valore medio considerando tutti i cromosomi della popolazione finale è di -785.7146. L'andamento del miglior valore e del valore medio ottenuto per ogni

iterazione viene evidenziato nella Figura 5.33, dove viene riportata la negazione del valore ottenuto.

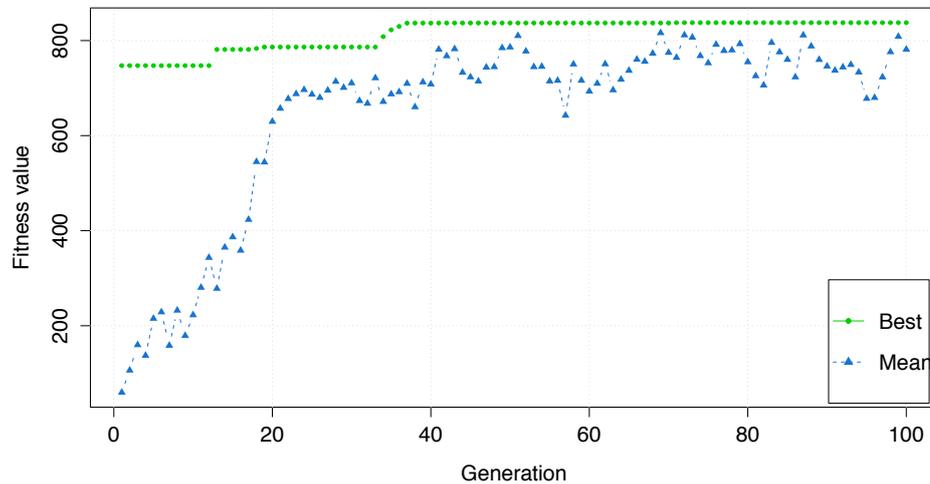


Figura 5.33: Diagnostica del modello e previsione per il consumo del sistema HVAC

L'algoritmo permette di raggiungere con una certa rapidità (entro la quarantesima generazione) il valore ottimale. Si evidenzia inoltre la velocità con cui il valore medio assunto da tutti i cromosomi appartenenti alla popolazione tende a migliorare nelle prime venti generazioni, a conferma della notevole capacità di adattamento dell'algoritmo proposto.

CONCLUSIONE

Nell'ambito dell'ottimizzazione del consumo del sistema HVAC si è considerato un approccio di analisi dinamico che permette di valutare non solo il semplice output di una serie di decisioni prese all'istante corrente, ma anche l'impatto che tali decisioni assumono negli istanti futuri. Questo ha permesso la modellazione di un sistema che non si limita all'analisi di un singolo istante temporale ma che punta oltre, formulando il problema per un periodo di tempo più esteso ed applicando solo i valori riferiti al primo istante temporale.

La modellazione è stata realizzata mediante l'impiego di un approccio basato sulle serie storiche e di un modello autoregressivo integrato a media mobile con variabili esogene. Nella modellazione è stato considerato anche l'effetto delle condizioni climatiche.

In una prima fase è stata realizzata la modellazione delle variabili climatiche; in una seconda fase, sulla base dei dati a disposizione, è stata realizzata la modellazione del consumo, che ha prodotto degli ottimi risultati in quanto permette di considerare congiuntamente sia i valori passati che i valori delle variabili esterne.

Infine, a causa della mancanza di alcuni dati, è stato fornito un possibile approccio alla soluzione del problema descritto, mediante la definizione dell'intero problema e l'impostazione di un algoritmo genetico per l'ottimizzazione dello stesso. L'algoritmo di ottimizzazione proposto è stato testato ed ha dimostrato di produrre dei risultati molto validi.

BIBLIOGRAFIA

- [1] Ashrae. *ANSI/ASHRAE Standard 55-2004, Thermal Comfort Conditions for Human Occupancy*. American Society of Heating, Air-Conditioning, and Refrigeration Engineers, Inc., 2004.
- [2] S. Atthajariyakul e T. Leephakpreeda. Neural computing thermal comfort index for HVAC systems. *Energy Conversion and Management*, 46(15-16):2553–2565, settembre 2005.
- [3] UNEP Sustainable Building e Climate Initiative. *Buildings and climate change: Summary for decision-makers*. Relazione tecnica, UNEP, 2009.
- [4] Carlos A. Coello Coello. A short tutorial on evolutionary multiobjective optimization. In *Proceedings of the First International Conference on Evolutionary Multi-Criterion Optimization, EMO '01*. Springer-Verlag, 2001.
- [5] Carlos A. Coello Coello, Gary B. Lamont, e David A. Van Veldhuizen. *Evolutionary Algorithms for Solving Multi-Objective Problems (Genetic and Evolutionary Computation)*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2006.
- [6] Charles Darwin. *The Origin of Species: by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life*. J. Murray, 1872.
- [7] T. Di Fonzo e F. Lisi. *Serie storiche economiche: analisi statistiche e applicazioni*. Carocci, 2005.

- [8] David E. Goldberg. *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley Longman Publishing Co., Inc., 1989.
- [9] Ian Knight. Assessing electrical energy use in hvac systems. *REHVA Journal*, 49:06–11, 2012.
- [10] Andrew Kusiak e Guanglin Xu. Modeling and optimization of HVAC systems using a dynamic neural network. *Energy*, 42(1):241–250, giugno 2012.
- [11] Alfred Jean Philippe Lauret, Harry Boyer, Carine Riviere, e Alain Bastide. A genetic algorithm applied to the validation of building thermal models. *CoRR*, abs/1212.5250, 2012.
- [12] Leo Liberti e Sergei Kucherenko. Comparison of deterministic and stochastic approaches to global optimization. *International Transactions in Operational Research*, 12, 2005.
- [13] S.G. Makridakis e M. Hibon. *Arma Models and the Box Jenkins Methodology*. INSEAD, 1995.
- [14] R. McDowall. *Fundamentals of HVAC Systems: Inch/Pound Edition*. ASHRAE learning. Elsevier Science & Technology, 2006.
- [15] Nabil Nassif, Stanislaw Kajl, e Robert Sabourin. Optimization of hvac control system strategy using two-objective genetic algorithm. *HVAC and R Research*, 11(3):459–486, 2005.
- [16] United States Department of Energy (DOE). *Buildings Energy Data Book*. D&R International, Ltd., 2011.
- [17] Frauke Oldewurtel, Alessandra Parisio, Colin N. Jones, Dimitrios Gyalistras, Markus Gwerder, Vanessa Stauch, Beat Lehmann, e Manfred Morari. Use of

- model predictive control and weather forecasts for energy efficient building climate control. *Energy and Buildings*, 45(0):15 – 27, 2012.
- [18] B. W. Olesen e K. C. Parsons. Introduction to thermal comfort standards and to the proposed new version of en iso 7730. *Energy and Buildings*, 34(6):537–548, luglio 2002.
- [19] A. Pankratz. *Forecasting with dynamic regression models*. Wiley series in probability and mathematical statistics: Applied probability and statistics. J. Wiley, 1991.
- [20] Mehmet Tektaş. Weather forecasting using anfis and arima models. *Environmental Research, Engineering and Management*, 51(1), 2010.
- [21] R.S. Tsay. *Analysis of Financial Time Series*. Wiley Series in Probability and Statistics. Wiley, 2005.
- [22] A. M. Turing. Computing machinery and intelligence. *Mind*, 59:433–460, 1950.
- [23] David Allen Van Veldhuizen. *Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations*. Tesi di Dottorato di Ricerca, Air Force Institute of Technology, Wright Patterson AFB, OH, USA, 1999. AAI9928483.
- [24] G.C.S. Wang e C.L. Jain. *Regression Analysis: Modeling & Forecasting*. Graceway Pub., 2003.
- [25] Xin-She Yang. *Introduction to Mathematical Optimization: From Linear Programming to Metaheuristics*. Cambridge International Science Publishing, 2008.