

POLITECNICO DI TORINO

Corso di Laurea Magistrale  
in Ingegneria Matematica

Tesi di Laurea Magistrale

**Ottimizzazione del Sistema di Previsione Energetica:  
Strategie per la Riduzione delle Perdite**



**Relatore**  
prof. Paolo Brandimarte

**Candidata**  
Giulia Accossato

Anno Accademico 2023-2024

*Alla memoria della mia amata nonna.*

*Il suo amore incondizionato, la sua saggezza senza tempo  
e il suo instancabile sostegno  
hanno segnato profondamente la mia vita.*

*Questo traguardo è il mio modo di renderla orgogliosa,  
di onorarla e di ringraziarla  
per tutto ciò che ha fatto per me.*

*Con tutto il mio affetto e la mia gratitudine.*

# Sommario

L'elaborato si propone di offrire un'analisi dettagliata del mercato dell'energia elettrica in Italia esplorandone la struttura, le dinamiche e le sfide attuali. Lo studio inizia con un'indagine approfondita del contesto normativo che regola il mercato elettrico italiano, evidenziando i meccanismi di negoziazione e le strategie adottate per garantire un funzionamento efficace e competitivo. Vengono esaminati i principali attori coinvolti, inclusi gli operatori di mercato e le autorità di regolamentazione, con particolare enfasi sul loro ruolo nella promozione della concorrenza e della stabilità del sistema energetico nazionale.

Successivamente, l'analisi si concentra sulle metodologie di previsione dello sbilanciamento energetico, esplorando le variabili ambientali, socio-economiche e comportamentali dei consumatori, che influenzano la domanda e l'offerta di energia elettrica. Viene impiegato un algoritmo basato sui giorni simili, che tiene conto di pattern storici e condizioni simili, per prevedere l'errore percentuale.

Particolare attenzione viene dedicata allo studio dello sbilanciamento energetico a livello macrozonale, valutando l'impatto di variabili esterne, come le condizioni meteorologiche e le politiche energetiche, sulla direzione di tale squilibrio. Si implementano modelli predittivi avanzati, integrando tecniche di boosting del gradiente e modelli additivi generalizzati per prevedere il segno dello sbilanciamento macrozonale.

Infine, si introducono strategie per ottimizzare il sistema di previsione energetica con l'obiettivo di ridurre le perdite e migliorare l'efficienza complessiva; tali strategie si basano sui risultati ottenuti dall'analisi dello sbilanciamento. Si esplorano modelli decisionali e algoritmi di ottimizzazione stocastica, che consentono un aggiornamento in tempo reale delle previsioni e una riduzione degli errori. Si sviluppa una struttura basata sull'algoritmo genetico e sulla simulazione stocastica Monte Carlo. Per concludere si valuta l'impatto di tali strategie sulla riduzione dei costi e sull'efficienza del sistema energetico italiano.

Attraverso un approccio analitico rigoroso e attraverso l'applicazione di metodologie avanzate, l'elaborato propone, quindi, soluzioni pratiche e innovative per affrontare le sfide emergenti nel settore energetico italiano, contribuendo così alla sostenibilità e all'affidabilità dell'approvvigionamento energetico nel paese e al beneficio economico degli attori coinvolti.

# Summary

The thesis aims to provide a detailed analysis of the electricity market in Italy, exploring its structure, dynamics, and current challenges. The study begins with an in-depth analysis of the regulatory context governing the Italian electricity market, highlighting the trading mechanisms and strategies adopted to ensure effective and competitive operation. The main actors involved, including market operators and regulatory authorities, are examined with particular emphasis on their role in promoting competition and stability within the national energy system.

Subsequently, the thesis focuses on energy imbalance forecasting methodologies, exploring the environmental, socio-economic, and consumer behavior variables that influence electricity demand and supply. An algorithm based on similar days is employed, which takes into account historical patterns and similar conditions, to predict the percentage error.

Particular attention is given to the analysis of energy imbalance at the macro-zonal level, evaluating the impact of external variables, such as weather conditions and energy policies, on the direction of such imbalance. Advanced predictive models are implemented, integrating gradient boosting techniques and generalized additive models to predict the sign of macro-zonal imbalance.

Finally, strategies are introduced to optimize the energy forecasting system with the aim of reducing losses and improving overall efficiency; these strategies are based on the results obtained from the imbalance analysis. Decision models and stochastic optimization algorithms are explored, enabling real-time forecast updates and error reduction. A framework based on genetic algorithms and Monte Carlo stochastic simulation is developed. To conclude, the impact of these strategies on cost reduction and the efficiency of the Italian energy system is evaluated.

Through a rigorous analytical approach and the application of advanced methodologies, the thesis proposes practical and innovative solutions to address emerging challenges in the Italian energy sector, thereby contributing to the sustainability and reliability of the country's energy supply and to the economic benefit of the actors involved.

# Ringraziamenti

Vorrei dedicare un grande grazie a tutte le persone che mi hanno sostenuto durante questo percorso.

Desidero esprimere il mio sincero ringraziamento a Pierluigi, Davide, Sabrina, Sabrina e Marco per avermi accolto in azienda. La loro guida preziosa e il loro supporto mi hanno permesso di affrontare sfide stimolanti e di immergermi completamente nel mondo reale dei dati, un ambiente decisamente complesso e affascinante.

Desidero ringraziare il mio relatore, il professore Paolo Brandimarte, per la sua guida e il suo sostegno fondamentale durante la realizzazione di questa tesi. La sua competenza e disponibilità hanno contribuito al successo di questo lavoro.

Desidero ringraziare di cuore mia mamma Donata per essere stata sempre al mio fianco, soprattutto nei momenti più difficili. Per le sue telefonate dopo ogni esame, colme di gioia o di conforto. Il suo sostegno incondizionato e, diciamocelo, la sua pazienza, sono stati fondamentali per farmi arrivare fin qui. So di essere testarda, ma fortunatamente, ho preso da lei la determinazione che mi ha aiutato a perseguire i miei obiettivi con passione. Sono profondamente grata per questo, perché so che è il primo passo verso il raggiungimento dei migliori risultati.

Desidero ringraziare di cuore mio papà Lorenzo, che nonostante "Mi ricordi qual è l'argomento della tua tesi?", ha sempre saputo trovare il modo di sostenermi e farmi sentire al posto giusto. Con il suo incoraggiamento costante è riuscito a spronarmi a dare il meglio, dimostrandomi continuamente il suo orgoglio (e no, non parlo solo delle paghette dopo ogni esame andato bene). Grazie a lui ho imparato che ogni problema ha una soluzione e di non abbattersi se non la si trova subito. Questo insegnamento è stato fondamentale.

Desidero ringraziare mia sorella Valentina per aver pazientemente letto questo elaborato, anche se probabilmente non ci ha capito molto. Il suo supporto e il suo interesse dimostrano quanto è speciale per me. Grazie per essere sempre lì, anche quando le equazioni e le teorie sembravano più complicate di un rompicapo. Senza di te, questo traguardo non sarebbe stato lo stesso.

Desidero dedicare un sincero ringraziamento a Stefano per essere stato al mio fianco in questo percorso di crescita e realizzazione. Per aver creduto in me e nelle mie capacità anche quando non riuscivo a farlo da sola, per essere stato la mia principale valvola di sfogo (scusa!) e per il suo costante sostegno. Questo traguardo è anche merito suo. Guardo avanti con gioia e che questo sia uno fra i tanti successi che festeggeremo insieme.

Infine, desidero esprimere il mio profondo ringraziamento a tutta la mia famiglia e a tutti i miei amici. Il loro amore, affetto e sostegno sono stati fondamentali per superare ogni ostacolo e lo saranno sempre. Ciascuno di loro occupa un posto speciale nel mio cuore. Le lunghe telefonate in macchina per tenersi compagnia, le risate condivise per ogni successo raggiunto, i caffè prima di andare allo stage: questi momenti hanno reso tutto indimenticabile. Spero un giorno di poter ricambiare tutto ciò che mi hanno donato.

Un ringraziamento speciale, però, va a nonna Donata, la persona che più di tutti desiderava vedermi raggiungere questo traguardo. Anche se non è più con noi fisicamente, so che il suo spirito è con me ogni giorno, specialmente adesso. La ringrazio di cuore per ogni insegnamento prezioso che mi ha impartito, per ogni dolce ricordo condiviso e per le storie meravigliose che mi ha regalato. La ringrazio per avermi fatto sentire sempre amata e per avermi dato la forza di perseverare nei momenti difficili. Nonna, so con tutto il cuore che sei orgogliosa di me, e questo è uno dei tanti successi che avrei voluto festeggiare insieme a te. Il tuo amore e il tuo sostegno continuano a guidarmi, e per questo ti sarò eternamente grata.

# Indice

<b>Introduzione</b>	9
<b>1 Il Mercato dell'Energia Elettrica in Italia: Struttura e Dinamiche di Funzionamento</b>	11
1.1 Struttura e Dinamiche	12
1.1.1 Il Mercato del Giorno Prima	13
1.1.2 Il Mercato Infragiornaliero	14
1.1.3 Il Mercato dei Servizi di Dispacciamento	15
1.2 Tariffazione differenziata e gestione degli sbilanciamenti	15
1.2.1 Meccanismi di tariffazione e implicazioni economiche per le Unità Non Qualificate	18
<b>2 Analisi e Previsione dello Sbilanciamento Energetico Zonale</b>	22
2.1 Analisi dell'Influenza di Variabili Ambientali e Temporali nello Sbilanciamento Zonale	23
2.1.1 Analisi dei fattori influenzanti	25
2.1.2 Approccio adimensionale per l'applicazione dell'algoritmo	31
2.2 Modello di Stima: Algoritmo Giorni Simili	34
2.3 Analisi Risultati e Valutazione delle Prestazioni	35
<b>3 Analisi e Previsione del Segno Macrozonale di Sbilanciamento</b>	40
3.1 Esplorazione dello Sbilanciamento Macrozonale e delle Variabili Esogene	40
3.1.1 Analisi preliminare Serie Storiche e Test di Dipendenza Markoviana	44
3.2 Modello per la stima: Approccio con Modelli Generalizzati Additivi	48
3.2.1 Modello Generalizzato Additivo	48
3.2.2 <i>Boosting</i> del gradiente a componenti: Ottimizzazione dei Modelli Additivi Generalizzati	49
3.3 Valutazione e Risultati del Modello	52
3.3.1 Considerazioni sull'Importanza del Livello di Probabilità Previsto	56
<b>4 Ottimizzazione del Sistema di Previsione Energetica: Strategie per la Riduzione delle Perdite</b>	58
4.1 Problemi di Ottimizzazione: Caratteristiche e Applicazioni	59

4.1.1	Modelli di Programmazione Vincolata alla Casualità per Problemi Stocastici: Approcci e Gestione dei Vincoli . . . . .	63
4.2	Struttura Integrata di Ottimizzazione: Algoritmo Genetico e Simulazione Monte Carlo . . . . .	65
4.2.1	Simulazione stocastica per vincoli di probabilità . . . . .	66
4.2.2	Algoritmo genetico . . . . .	69
4.3	Valutazione dei Risultati e Analisi Economica . . . . .	75
4.3.1	Analisi della valorizzazione economica dello sbilanciamento energetico	80
	<b>Conclusione</b>	<b>83</b>



*La matematica è l'arte di trovare modelli,  
la scienza di scoprire cause  
e la tecnologia di usare queste scoperte.*

JOHN ALLEN PAULOS

# Introduzione

Nel complesso panorama del mercato elettrico italiano, caratterizzato non solo da una crescente domanda di energia, ma anche dalla necessità di garantire la sicurezza degli approvvigionamenti, emerge con chiarezza un cruciale bisogno di gestione efficiente del sistema.

Dal suo avvio alla fine degli anni '90, il mercato elettrico italiano ha subito un profondo cambiamento, motivato dalla necessità di promuovere la concorrenza, migliorare l'efficienza e assicurare una fornitura stabile di energia. Un elemento cruciale di questo mercato è il meccanismo di bilanciamento, che mira a mantenere l'equilibrio tra domanda e offerta di energia elettrica. Una previsione accurata del consumo di elettricità è vitale per un bilanciamento efficace, permettendo agli operatori di mercato di prendere decisioni informate e ottimizzare l'allocazione delle risorse. L'efficienza nell'allocazione delle risorse energetiche, la capacità di prevedere e adattarsi alle variazioni della domanda e dell'offerta, così come la promozione di pratiche e tecnologie a basso impatto ambientale, sono tutti elementi centrali che contribuiscono a una buona gestione.

## Obiettivi di Ricerca

Questo lavoro si propone di fornire un'analisi approfondita e critica del contesto normativo che regola il mercato elettrico italiano, esaminando i regolamenti e identificando le implicazioni per l'efficienza e la sostenibilità del sistema energetico nazionale. Attraverso un'indagine dei processi di negoziazione e delle dinamiche dei tre principali segmenti di mercato, si cerca di comprendere le opportunità presenti nel mercato elettrico italiano.

L'elaborato si concentra, dunque, sulle sfide della previsione e del bilanciamento del consumo di elettricità nel contesto del mercato elettrico italiano. Gli obiettivi principali sono duplici:

1. Prevedere lo sbilanciamento: Sviluppare modelli predittivi per la previsione dello sbilanciamento energetico, integrando metodologie di data mining, machine learning e modellazione statistica e tenendo conto dei molteplici fattori che influenzano la domanda.
2. Ottimizzare il meccanismo di bilanciamento: Utilizzare i modelli di previsione per fornire aggiustamenti e suggerimenti in tempo reale agli operatori di mercato, al fine di minimizzare gli squilibri e i costi associati, ottenendo previsioni di consumo più accurate e garantendo al contempo la stabilità del sistema.

## Metodologia

Per raggiungere tali obiettivi, la tesi adotta un approccio metodologico completo:

1. Raccolta e pre-elaborazione dei dati: Raccogliere dati storici sul consumo di elettricità e sui fattori influenti pertinenti. Pre-elaborare i dati per garantirne la qualità e la coerenza.
2. Analisi del meccanismo di bilanciamento: Esaminare il comportamento e le dinamiche del meccanismo di bilanciamento, inclusa l'identificazione dei fattori chiave che influenzano gli squilibri.
3. Modelli di previsione dello sbilanciamento: Sviluppare diversi modelli statistici e di apprendimento automatico per prevedere lo sbilanciamento tra previsione e consumo di elettricità a vari livelli (zonale, macrozonale, ecc.), valutandone l'accuratezza e l'affidabilità attraverso metodi di validazione incrociata.
4. Sistema di supporto decisionale: Progettare un sistema di supporto decisionale che integri i modelli di previsione e le analisi del meccanismo di bilanciamento per fornire suggerimenti in tempo reale per ottimizzare le operazioni di mercato, consentendo agli operatori di prendere decisioni rapide ed efficaci.

## Struttura dell'elaborato

L'elaborato è strutturato in quattro capitoli principali, ciascuno dei quali si focalizza su un aspetto specifico del mercato elettrico italiano e delle relative sfide.

Il capitolo introduttivo offre una panoramica del mercato dell'energia elettrica in Italia, esaminando la sua struttura, i segmenti principali e i ruoli chiave degli attori coinvolti, come il Gestore del Mercato Elettrico (GME) e gli Acquirenti Unici. Si approfondiscono la tariffazione differenziata e la gestione degli sbilanciamenti energetici. Il secondo capitolo si concentra sulla previsione dello sbilanciamento energetico zonale, presentando un approccio basato sull'algoritmo dei "giorni simili". Il terzo capitolo affronta la previsione del segno macrozonale dello sbilanciamento energetico, proponendo un modello basato su Modelli Generalizzati Additivi (GAM) e valutandone le prestazioni con metriche statistiche. Nel quarto capitolo, si propone una struttura integrata per ottimizzare il sistema di previsione energetico, con l'obiettivo di ridurre le perdite e migliorare l'efficienza complessiva, valutando le strategie proposte attraverso metriche di prestazione e attraverso un'analisi economica.

# Capitolo 1

## Il Mercato dell'Energia Elettrica in Italia: Struttura e Dinamiche di Funzionamento

Il Mercato Elettrico Italiano ha radici nel decreto legislativo 16 marzo 1999, n. 79 (d.lgs. n. 79/99), il quale si inserisce nel contesto del recepimento della direttiva comunitaria volta a creare un mercato interno dell'energia (96/92/CE). Secondo quanto analizzato da [Oggioni and Lanfranconi](#), tale iniziativa è stata motivata da due esigenze fondamentali, in linea con altre esperienze internazionali:

- La necessità di promuovere la concorrenza nelle attività di produzione e compravendita di energia elettrica, garantendo criteri di neutralità, trasparenza e obiettività attraverso l'istituzione di una "piazza del mercato".
- L'importanza di assicurare una gestione economica efficace per garantire una disponibilità adeguata dei servizi di dispacciamento.

Il Mercato Elettrico rappresenta un'arena telematica di grande importanza per il settore energetico italiano. Qui, gli operatori partecipanti negoziano l'energia elettrica all'ingrosso, determinandone il prezzo tramite l'equilibrio tra l'offerta e la domanda. Si tratta di un mercato fisico dove vengono definiti i programmi di immissione e prelievo dell'energia elettrica, basati su criteri di merito economico. È importante sottolineare che, sebbene la partecipazione alla Borsa Elettrica sia una prassi comune, non è obbligatoria; gli operatori hanno la possibilità di concludere contratti di compravendita al di fuori della piattaforma di borsa, mediante i cosiddetti contratti bilaterali (OTC). Questa flessibilità consente agli attori del mercato di adattare le proprie strategie commerciali alle esigenze specifiche del settore energetico e alle dinamiche di mercato in continua evoluzione.

## 1.1 Struttura e Dinamiche

Il panorama del mercato elettrico italiano si presenta articolato e complesso, caratterizzato da diverse componenti e sottosistemi che interagiscono per garantire l'approvvigionamento e lo scambio efficiente di energia elettrica su tutto il territorio nazionale.

Al centro di questo sistema si trovano il **Mercato Elettrico a Pronti** (MPE) e il **Mercato Elettrico a Termine** (MTE), due pilastri fondamentali che regolano le dinamiche di acquisto e vendita di energia. Nel MPE, la maggior parte delle transazioni avviene in tempo reale e si basa sull'equilibrio tra domanda e offerta del mercato stesso, stabilendo così il prezzo di equilibrio dell'energia elettrica; mentre nel MTE si negoziano contratti a lungo termine.

Il ruolo di controparte unica nel MPE è svolto dal **Gestore dei Mercati Elettrici S.p.A** (GME)<sup>1</sup>, che agisce come acquirente unico. Questo garantisce una gestione centralizzata del mercato elettrico italiano, consentendo a tutti gli operatori del settore di partecipare alla Borsa Elettrica italiana con libertà e trasparenza.

All'interno del Mercato a Pronti dell'Energia, si distinguono tre sottomercati chiave, che rappresentano il fulcro delle attività di trading sulla borsa elettrica :

1. Il **Mercato del Giorno Prima (MGP)**, dove gli attori del mercato (i produttori, i grossisti e i clienti finali idonei) possono vendere o acquistare energia elettrica per il giorno successivo, stabilendo i programmi di immissione e prelievo.
2. Il **Mercato Infragiornaliero (MI)**, che ha sostituito il preesistente Mercato di Aggiustamento e che consente agli operatori di apportare modifiche ai programmi stabiliti nel MGP attraverso ulteriori offerte, garantendo una maggiore flessibilità nell'adattamento alle variazioni della domanda e dell'offerta.
3. Il **Mercato del Servizio di Dispacciamento (MSD)**, essenziale per il gestore del Sistema di Trasmissione Italiano (TSO), **Terna S.p.A.**<sup>2</sup>, per garantire le risorse necessarie alla gestione e al controllo del sistema elettrico.

Parallelamente, il sistema elettrico italiano è suddiviso in porzioni di reti di trasmissione, definite zone di mercato, ciascuna con un prezzo zonale dell'energia. Questa suddivisione è fondamentale per gestire i flussi di energia lungo la penisola e garantire la sicurezza del sistema elettrico nazionale. Le zone sono state definite in base alle reti di trasmissione e ai limiti fisici di transito dell'energia, in modo da ottimizzare l'efficienza e la stabilità del sistema nel suo complesso.

La conformazione di tali zone è funzionale alle modalità di gestione dei transiti lungo la penisola adottate da Terna S.p.A. e si può sintetizzare in:

- 7 zone geografiche (Centro Nord, Nord, Centro Sud, Sud, Calabria, Sicilia, Sardegna) rappresentate nella Figura 1.1;

---

<sup>1</sup>Gestore dei Mercati Elettrici S.p.A.

<sup>2</sup>Terna S.p.A.

- 9 zone virtuali estere (Francia, Svizzera, Austria, Slovenia, Corsica, Corsica AC, Grecia, Malta, Montenegro);
- 5 zone virtuali estere, rappresentative dell'interconnessione dedicata al *market coupling* (Francia Coupling, Austria Coupling, Slovenia Coupling, Grecia Coupling, Italia Coupling)
- 4 zone virtuali nazionali che rappresentano poli di produzione limitata, ovvero zone costituite da sole unità di produzione, la cui capacità di interconnessione con la rete è inferiore alla potenza installata delle unità stesse.



Figura 1.1: Suddivisione in zone geografiche del mercato elettrico italiano.

### 1.1.1 Il Mercato del Giorno Prima

Il Mercato del Giorno Prima (MGP) rappresenta una tappa cruciale nel panorama del mercato elettrico italiano. Qui avviene lo scambio di energia elettrica all'ingrosso, dove vengono negoziati blocchi orari di energia per il giorno successivo. Il MGP va oltre la semplice determinazione di prezzi e quantità scambiate: è anche il momento in cui si definiscono i programmi di immissione e prelievo energetico per il giorno a venire. Questo aspetto assume particolare rilevanza, considerando l'importanza delle previsioni di consumo energetico nel processo decisionale. Gli operatori possono così adattare le proprie offerte in base alle esigenze previste, garantendo una gestione ottimale dell'energia elettrica sul territorio nazionale.

Il funzionamento del MGP si basa su un sistema di frazionamento del mercato, che si articola secondo un modello di asta implicita. È qui che si concentra la maggior parte delle transazioni di compravendita di energia elettrica. Tutti gli operatori che hanno ottenuto la qualifica di "operatore del mercato elettrico" possono partecipare al MGP, e il Gestore dei Mercati Elettrici S.p.A (GME) agisce, come già ribadito, da controparte centrale per tutte le operazioni di acquisto e vendita, garantendo trasparenza e equità nel processo.

Durante il periodo di apertura del MGP, gli operatori possono presentare le proprie offerte, specificando la quantità di energia e il prezzo massimo e minimo al quale sono disposti a comprare o vendere. Tuttavia, è importante sottolineare che le offerte devono essere coerenti con le capacità di immissione o prelievo del punto di offerta a cui si riferiscono e devono rispecchiare la volontà effettiva di immettere o prelevare energia elettrica. Una volta chiusa la seduta di mercato, le offerte vengono valutate in base al merito economico e nel rispetto dei limiti di transito tra le diverse zone. Si definiscono così i prezzi di equilibrio e i volumi di energia scambiati. L'algoritmo implementato dal GME genera due sistemi di prezzo distinti, a seconda degli operatori di mercato. Per l'energia destinata al consumo in Italia, viene applicato un prezzo unico di acquisto su base nazionale (**PUN**), calcolato come media ponderata dei prezzi di vendita zonali in base ai consumi zonali. Questo sistema assicura una gestione equa e trasparente del mercato elettrico, garantendo un approvvigionamento energetico efficiente e affidabile su tutto il territorio nazionale. Per i punti di immissione e prelievo appartenenti alle zone virtuali estere, invece, si applicano i prezzi zonali sia in vendita che in acquisto.

### **1.1.2 Il Mercato Infragiornaliero**

Una volta che il Mercato del Giorno Prima (MGP) ha chiuso le sue operazioni, gli operatori del settore energetico non si fermano. Al contrario, hanno la possibilità di aggiornare le loro offerte di vendita e acquisto, così come le loro posizioni commerciali. Questo aggiornamento costante tiene conto di molteplici fattori, tra cui le informazioni più recenti sulle condizioni di mercato, come le previsioni meteorologiche che possono influenzare la domanda di energia, eventi imprevedibili come guasti di impianti o interruzioni di forniture, e una più precisa definizione delle esigenze di consumo. Questo processo dinamico consente agli operatori di adattare le proprie strategie commerciali in tempo reale, ottimizzando così l'allocazione delle risorse energetiche e garantendo una gestione efficiente ed efficace del mercato.

Il Mercato Infragiornaliero è stato progettato proprio per fungere da prolungamento del MGP, seguendo le medesime regole e modalità operative. Anche questo mercato è organizzato secondo un sistema di aste implicite e tiene conto della stessa rappresentazione zonale del territorio. Tuttavia, a differenza del MGP, nel Mercato Infragiornaliero non viene calcolato il Prezzo Unico Nazionale (PUN). In questo contesto, tutte le transazioni di acquisto e vendita di energia sono valorizzate al prezzo zonale specifico di ciascuna area geografica.

Questa differenza sostanziale rispetto al MGP riflette la necessità di avere una maggiore flessibilità e adattabilità alle condizioni specifiche di ciascuna zona di mercato nell'ambito del Mercato Infragiornaliero. Inoltre, il Mercato Infragiornaliero consente agli operatori di rispondere in modo più reattivo alle variazioni delle condizioni di mercato, poiché le transazioni avvengono più vicino al momento effettivo di consumo energetico. Questo approccio dinamico contribuisce a garantire una gestione ancora più efficiente e ottimale del sistema energetico nel suo complesso.

### **1.1.3 Il Mercato dei Servizi di Dispacciamento**

Il Mercato per il Servizio di Dispacciamento (MSD) costituisce il fulcro operativo attraverso il quale Terna S.p.A., nell'ambito del suo ruolo di gestore della rete elettrica nazionale, si approvvigiona delle risorse necessarie per assicurare un funzionamento ottimale e affidabile del sistema energetico italiano in tempo reale. Questo mercato svolge un ruolo critico nell'affrontare le sfide operative quotidiane, come la gestione delle congestioni intrazonali, la creazione di riserve di energia per garantire la sicurezza di approvvigionamento e il bilanciamento dell'offerta e della domanda di energia.

In contrapposizione al Mercato del Giorno Prima (MGP) e al Mercato Infragiornaliero (MI), che si basano su una rappresentazione semplificata della rete elettrica suddividendo il territorio in zone, il MSD opera su una base più granulare e dinamica, consentendo a Terna S.p.A. di effettuare interventi mirati per gestire al meglio le fluttuazioni della domanda e risolvere eventuali congestioni o squilibri nel sistema.

Nel contesto del MSD, Terna agisce come fulcro delle negoziazioni, stipulando contratti di acquisto e vendita con gli operatori del settore energetico per assicurare un adeguato approvvigionamento di risorse per il servizio di dispacciamento. Inoltre, agisce come controparte centrale, garantendo trasparenza e affidabilità nel processo di negoziazione. Le offerte sul MSD sono limitate ai punti di offerta abilitati all'operatività su questo mercato e devono essere presentate direttamente dagli utenti del dispacciamento, che comprendono produttori, distributori e altri operatori del settore. Una volta che un'offerta di acquisto è accettata sul MSD e riferita a punti di offerta in prelievo, il Gestore dei Mercati Elettrici (GME) determina il corrispettivo di non arbitraggio che l'operatore è tenuto a pagare o a ricevere. Questo processo, noto come metodologia *pay-as-bid*, garantisce una remunerazione equa e trasparente per tutte le transazioni effettuate sul MSD.

In definitiva, il Mercato per il Servizio di Dispacciamento svolge un ruolo cruciale nell'assicurare un funzionamento efficiente e affidabile del sistema energetico italiano, consentendo a Terna S.p.A. di gestire in modo proattivo le sfide operative e di mantenere la sicurezza e la stabilità del sistema in ogni momento.

## **1.2 Tariffazione differenziata e gestione degli sbilanciamenti**

Il Mercato per il Servizio di Dispacciamento (MSD), come si è detto, riveste un ruolo cruciale nella gestione degli sbilanciamenti all'interno del sistema energetico italiano. Gli sbilanciamenti rappresentano la differenza tra il consumo o la produzione di energia previsti nel Mercato del Giorno Prima (MGP) e nel Mercato Infragiornaliero (MI) e quelli effettivamente verificatisi in tempo reale. Queste discrepanze possono derivare da molteplici fattori, come ad esempio la difficoltà degli attori del mercato nel prevedere con esattezza i consumi dei propri utenti o la produzione dei propri impianti, soprattutto nel caso delle fonti energetiche rinnovabili, la cui generazione è influenzata da variabili ambientali difficilmente prevedibili.



Rispetto al tema degli sbilanciamenti, fino a novembre 2014, il mercato italiano era diviso in quattro aree, chiamate macrozone:

- Nord;
- Centro-Sud;
- Sicilia;
- Sardegna.

Tuttavia, a partire da Novembre 2014, queste macrozone sono state ridotte a soli due distretti, rappresentati nella Figura 1.2:

- Nord, la vecchia macrozona Nord;
- Sud, l'unione delle altre tre.



*Figura 1.2: Suddivisione macrozone del mercato elettrico italiano.*

L'aggregazione degli sbilanciamenti delle singole imprese all'interno di una data macrozona definisce lo sbilancio macrozonale.

Per gestire efficacemente questi sbilanciamenti, Terna S.p.A. si avvale del MSD, che funge da piattaforma per acquistare o vendere l'energia necessaria al fine di riequilibrare il sistema energetico. Questo mercato opera attraverso l'utilizzo delle cosiddette Unità Qualificate (QU), che rappresentano unità di produzione con una potenza nominale superiore a 10 Mega Volt Ampere (MVA) e che forniscono una regolazione in tempo reale partecipando attivamente al MSD. Le QU svolgono un ruolo fondamentale nel garantire la stabilità e l'equilibrio del sistema energetico nazionale. Le unità di generazione con una potenza nominale inferiore a 10 MVA sono denominate Unità Non Qualificate (NQU). Queste Unità non possono partecipare al Mercato per il Servizio di Dispacciamento (MSD) poiché non dispongono delle capacità tecniche necessarie per fornire una regolazione in tempo reale del sistema energetico.

Considerando le regole attualmente applicate, il meccanismo di calcolo dei volumi e dei prezzi di sbilanciamento nonché il Periodo di Regolamento degli Squilibri (ISP) dipendono dalla categoria considerata, ovvero se si tratta di Unità Qualificate o Non Qualificate.

1. I volumi di sbilanciamento delle **Unità Qualificate** sono calcolati prendendo in considerazione la differenza tra il programma orario precedentemente comunicato attraverso il Gestore dei Mercati Elettrici (GME) e l'effettiva immissione fisica di energia sul sistema, misurata su base di 15 minuti. Questo processo di calcolo consente di valutare con precisione gli sbilanciamenti che si verificano nel sistema energetico nel corso del tempo. Le offerte orarie presentate dagli operatori sul mercato vengono convertite in step di 15 minuti utilizzando una tecnica di profilazione definita dal codice di rete nazionale. Questo passaggio è fondamentale per allineare le previsioni di consumo o produzione dell'energia con la realtà operativa del sistema, consentendo una gestione più precisa degli sbilanciamenti.

Le Unità Qualificate affrontano un sistema di doppia tariffazione per gli sbilanciamenti, il che significa che sono soggette a una penalizzazione economica se si discostano significativamente dai programmi orari comunicati in base al proprio sbilanciamento. Inoltre, il Periodo di Regolamento degli Squilibri (ISP) per queste unità è di 15 minuti, il che implica che gli sbilanciamenti devono essere corretti entro questo intervallo di tempo per evitare sanzioni o penalizzazioni ulteriori.

2. Per le **Unità Non Qualificate**, i volumi di sbilanciamento (sia in aumento che in diminuzione) e il Periodo di Regolamento degli Squilibri (ISP) sono stabiliti su base oraria. Questo significa che le discrepanze tra il consumo o la produzione previsti e quelli effettivi vengono valutate e corrette su base oraria.

Inoltre, gli sbilanciamenti delle Unità Non Qualificate sono trattati con un unico sistema di prezzi. Ciò implica che non sono soggetti a una tariffazione differenziata in base alla direzione o all'entità dello sbilanciamento. Viene applicato un unico prezzo per correggere gli sbilanciamenti nell'ambito di un determinato periodo orario. Questa uniformità nel trattamento degli sbilanciamenti per le Unità Non Qualificate semplifica il processo di gestione degli sbilanciamenti e contribuisce a garantire un approccio equo ed efficiente nella correzione delle discrepanze nel sistema energetico.

I diversi meccanismi di tariffazione applicati per gli sbilanciamenti delle Unità Qualificate e Non Qualificate rispecchiano i differenti ruoli che questi due gruppi di unità svolgono nel mercato del bilanciamento dell'energia. Le Unità Qualificate hanno un ruolo fondamentale nella garanzia della sicurezza e della stabilità dell'intero sistema energetico. Poiché sono gli unici fornitori di servizi nel Mercato per il Servizio di Dispacciamento (MSD), Terna S.p.A. si affida principalmente a loro per compensare gli sbilanciamenti e mantenere l'equilibrio del sistema. Per incentivare le Unità Qualificate a rispettare i programmi orari comunicati, viene applicato dunque il sistema di doppia tariffazione. Questo sistema implica che le Unità Qualificate sono soggette a penalizzazioni economiche se si discostano significativamente dai propri schedamenti, fornendo così un incentivo diretto

per mantenere la coerenza tra le previsioni e la produzione effettiva. D'altra parte, le Unità Non Qualificate non partecipano direttamente al mercato del bilanciamento e non forniscono servizi di regolazione in tempo reale. Di conseguenza, i loro sbilanciamenti sono trattati con un meccanismo di tariffazione unica e meno penalizzante rispetto al sistema di doppia tariffazione applicato alle Unità Qualificate.

In sintesi, mentre il sistema di doppia tariffazione applicato alle Unità Qualificate funge da incentivo per garantire una gestione precisa e affidabile degli sbilanciamenti, le Unità Non Qualificate sono soggette a una tariffazione più semplice poiché non svolgono un ruolo attivo nel mercato del bilanciamento dell'energia.

### **1.2.1 Meccanismi di tariffazione e implicazioni economiche per le Unità Non Qualificate**

In ciascuna delle macrozone precedentemente menzionate si trova un numero determinato di unità produttive e unità di consumo. Per ogni giorno  $t$  e per ogni periodo di carico  $h$ , lo sbilanciamento generato da ciascuna unità rappresenta la differenza tra il consumo o la produzione programmata e quella effettivamente realizzata. Lo sbilanciamento macrozonale è dato dalla somma algebrica dell'energia acquistata e venduta da Terna S.p.A. all'interno di una macrozona per affrontare lo squilibrio generato da ciascuna unità. Convenzionalmente, l'energia acquistata da Terna S.p.A. viene indicata con segno negativo.

Il segno dello squilibrio macrozonale, noto come segno macrozonale, riveste un'importanza fondamentale per le sue conseguenze economiche. Infatti, il prezzo al quale viene regolato lo sbilanciamento di una Unità Non Qualificata (NQU) nel Mercato per il Servizio di Dispacciamento (MSD), secondo il sistema di tariffazione unica, generalmente differisce da quello stabilito nel Mercato del Giorno Prima (MGP). La possibilità di trarre vantaggio (o di subire svantaggio) dipende in modo cruciale dal rapporto tra lo sbilanciamento macrozonale e il proprio sbilanciamento. In altre parole, per le NQU, la possibilità di realizzare un profitto o di avere una perdita, dipende esclusivamente dallo sbilanciamento aggregato macrozonale, mentre l'entità di tale profitto dipende dalla differenza di prezzo tra il MGP e il MSD, oltre che dal volume di sbilanciamento.

Per le Unità Non Qualificate (NQU) il sistema di tariffazione unica funziona come descritto nella Tabella 1.1, dove:

- $S_{t,h}^M$  è lo sbilanciamento aggregato macrozonale al giorno  $t$  e periodo di carico  $h$ ;
- $S_{t,h,i}$  è lo sbilanciamento zonale dell'unità  $i$  al giorno  $t$  e periodo di carico  $h$ ;
- $P_{MGP,t,h}$  è il prezzo  $MWh$  nel Mercato del Giorno Prima;
- $P_{MSD,t,h} \downarrow$ ,  $P_{MSD,t,h} \uparrow$  sono rispettivamente i prezzi medi in offerta e in acquisto  $MWh$  nel Mercato dei Servizi di Dispacciamento.

Sbilanciamento macrozonale	Prezzo valutazione sbilanciamento	Sbilanciamento dell'unità	Implicazione per l'unità
$S_{t,h}^M \geq 0$	$P_{t,h} = \min(P_{MGP,t,h}, P_{MSD,t,h} \downarrow)$	$S_{t,h,i} > 0$	Energia in eccesso venduta a $P_{t,h}$
		$S_{t,h,i} < 0$	Energia in difetto pagata $P_{t,h}$
$S_{t,h}^M \geq 0$	$P_{t,h} = \max(P_{MGP,t,h}, P_{MSD,t,h} \uparrow)$	$S_{t,h,i} > 0$	Energia in eccesso venduta a $P_{t,h}$
		$S_{t,h,i} < 0$	Energia in difetto pagata a $P_{t,h}$

Tabella 1.1: Sistema a tariffazione unica per lo sbilanciamento, applicato alle Unità Non Qualificate.

Nel dettaglio, si considerano i seguenti casi:

1. **Lo sbilanciamento macrozonale è positivo** (in gergo in tal caso lo sbilanciamento è detto *lungo*)  $S_{t,h}^M > 0$  e **lo sbilanciamento zonale dell'unità anche**  $S_{t,h,i} > 0$ .

Poiché, in queste condizioni, il prezzo nel Mercato del Giorno Prima (MGP) è normalmente maggiore rispetto al prezzo nel Mercato per il Servizio di Dispacciamento (MSD), l'energia extra immessa nel sistema attraverso il mercato dei servizi di dispacciamento viene venduta a un prezzo inferiore rispetto a quello pagato nel MGP. La penalità pagata a Terna viene calcolata sulla base della differenza di costo dell'energia extra (in valore assoluto) e, inoltre, si incappa in una situazione di perdita economica.

Questo meccanismo comporta che le unità di produzione e consumo debbano affrontare una doppia penalizzazione: da un lato, l'energia in eccesso viene venduta a un prezzo inferiore rispetto a quello di acquisto, e dall'altro lato, la penalità associata a tale squilibrio aumenta ulteriormente le perdite.

$$S_{t,h,i} \times (P_{t,h} - P_{MGP,t,h}) < 0$$

2. **Lo sbilanciamento macrozonale è positivo**  $S_{t,h}^M > 0$  **mentre lo sbilanciamento zonale dell'unità è negativo** (in gergo in tal caso lo sbilanciamento è detto *corto*)  $S_{t,h,i} < 0$ .

Poiché, in queste condizioni, il prezzo nel Mercato del Giorno Prima (MGP) è normalmente maggiore rispetto al prezzo nel Mercato per il Servizio di Dispacciamento (MSD), l'energia mancante viene acquistata a un prezzo inferiore rispetto a quello che sarebbe stato pagato nel MGP. La penalità pagata a Terna è calcolata sulla base del costo dell'energia mancante, ma in questo caso si è in una posizione di guadagno.

Questo scenario implica che, nonostante la penalità, l'unità di consumo o produzione riesca a ottenere un vantaggio economico grazie alla differenza di prezzo tra i due mercati. Di fatto, il costo minore dell'energia acquistata nel MSD compensa la penalità imposta, consentendo un risparmio rispetto al prezzo più alto del MGP.

$$S_{t,h,i} \times (P_{t,h} - P_{MGP,t,h}) > 0$$

3. **Lo sbilanciamento macrozonale è negativo**  $S_{t,h}^M < 0$  **e lo sbilanciamento zonale dell'unità anche**  $S_{t,h,i} < 0$ .

Poiché, in queste condizioni, il prezzo nel Mercato del Giorno Prima (MGP) è normalmente inferiore rispetto al prezzo nel Mercato per il Servizio di Dispacciamento (MSD), l'energia mancante viene acquistata a un prezzo superiore rispetto a quello che sarebbe stato pagato nel MGP. La penalità pagata a Terna viene calcolata sulla base della differenza di costo dell'energia mancante (in valore assoluto), e inoltre si è in una posizione di perdita economica.

Analogamente al primo caso, questo meccanismo comporta che le unità di produzione e consumo debbano affrontare una doppia penalizzazione: da un lato, l'energia mancante viene acquistata a un prezzo superiore rispetto a quello preventivato nel MGP, e dall'altro lato, la penalità associata a tale squilibrio aumenta ulteriormente le perdite.

$$S_{t,h,i} \times (P_{t,h} - P_{MGP,t,h}) < 0$$

4. **Lo sbilanciamento macrozonale è negativo**  $S_{t,h}^M < 0$  **mentre lo sbilanciamento zonale dell'unità è positivo**  $S_{t,h,i} > 0$ .

Poiché, in queste condizioni, il prezzo nel Mercato del Giorno Prima (MGP) è normalmente inferiore rispetto al prezzo nel Mercato per il Servizio di Dispacciamento (MSD), l'energia extra immessa nel sistema attraverso il MSD viene venduta a un prezzo superiore rispetto a quello pagato nel MGP. La penalità pagata a Terna viene calcolata sulla base della differenza di costo dell'energia extra, ma in questo caso si è in una posizione di guadagno.

In questo scenario, come nel secondo, nonostante la penalità imposta, l'unità di produzione o consumo riesce a ottenere un vantaggio economico grazie alla differenza di prezzo tra i due mercati. Il maggiore ricavo ottenuto dalla vendita dell'energia extra nel MSD compensa la penalità, consentendo un guadagno rispetto al prezzo inferiore del MGP.

$$S_{t,h,i} \times (P_{t,h} - P_{MGP,t,h}) > 0$$

È essenziale comprendere che il sistema di regolazione degli sbilanciamenti ha un impatto diretto sulla gestione economica delle unità di produzione e consumo. Le decisioni operative devono tenere conto non solo della quantità di energia prodotta o consumata, ma anche del contesto macrozonale in cui tali operazioni avvengono. Le NQU devono monitorare attentamente le dinamiche del mercato energetico e adeguare le loro strategie in base alle variazioni di prezzo tra il MGP e il MSD per massimizzare i profitti o meglio, minimizzare le perdite. Inoltre, l'analisi del segno macrozonale e la comprensione delle

sue implicazioni possono influenzare le decisioni di investimento nel settore energetico. Le aziende che operano in questo campo devono essere in grado di prevedere gli sbilanciamenti e gestire efficacemente le risorse per garantire la sostenibilità economica delle loro attività.

Tuttavia, è fondamentale sottolineare che le situazioni di guadagno, come il secondo e il quarto caso, non dovrebbero incentivare pratiche di squilibrio intenzionale. La gestione efficace degli sbilanciamenti deve mirare alla stabilità del sistema elettrico, piuttosto che al profitto ottenuto sfruttando differenze di prezzo. Le unità partecipanti al mercato devono impegnarsi a programmare con precisione le proprie attività di produzione e consumo per contribuire alla sostenibilità complessiva del sistema.

## Capitolo 2

# Analisi e Previsione dello Sbilanciamento Energetico Zonale

Il sistema di previsione del consumo elettrico mira a generare una stima accurata del fabbisogno energetico dei consumatori finali. Ogni società di distribuzione e vendita utilizza queste previsioni per due scopi principali: fornire quotidianamente a Terna S.p.A. una previsione del consumo a livello di zona per il giorno successivo e pianificare a medio/lungo termine l'acquisto dell'energia necessaria per soddisfare la domanda dei propri clienti.

Il processo di analisi dello sbilanciamento si propone di identificare uno o più intervalli di confidenza (a diversi livelli) sull'errore di previsione commesso dal sistema. POD, acronimo di "point of delivery", è il codice univoco identificativo del singolo contatore; questo codice rappresenta il collegamento essenziale con i dati di consumo. Un POD è gestito in modo differente in base alla tipologia di contatore e alla frequenza dei dati di misura. Esistono tre tipologie principali:

1. **MO (Misurati Orari, Trattati Orari o Generic Customer)**: contatori che rilevano misure ogni 15 minuti (trasmesse dal distributore al Sistema Informativo Integrato) e attivati in un periodo di tempo predefinito. Questi POD vengono gestiti a livello di previsione oraria.
2. **PRA (Prelievo Residuo di Area o Load Profile)**: contatori per cui vengono trasmesse misure mensili per fascia oraria.
3. **IP (Illuminazione Pubblica)**: POD per l'illuminazione pubblica, per i quali l'UdD riceve già una curva di consumi prevista.

La segmentazione dei clienti in categorie come MO, PRA e IP consente un'analisi mirata e una gestione differenziata della domanda energetica. I clienti con esigenze particolarmente elevate, come i Trattati Orari, richiedono strategie specifiche per garantire una fornitura costante e affidabile, mentre settori come l'illuminazione pubblica possono richiedere approcci differenti in termini di previsione e gestione della domanda.

L'indagine viene condotta a livello aggregato zonale, separando le tipologie di trattamento e escludendo l'illuminazione pubblica.

Per ciascuna di queste tipologie clienti, sono disponibili serie storiche orarie di consumo e le relative serie storiche di previsione. Un approccio comune suggerisce di trattare 24 serie storiche giornaliere diverse, una per ogni ora del giorno, anziché considerare le serie storiche a livello orario. L'adozione di serie storiche giornaliere, anziché orarie, per la previsione consente una rappresentazione più chiara e gestibile dello schema di consumo, consentendo una maggiore precisione nell'identificazione di trend e pattern. Questo approccio si rivela particolarmente utile in un contesto in cui i fattori che influenzano la domanda energetica sono sempre più eterogenei e complessi. Con lo sviluppo socio-economico moderno, infatti, i fattori che influenzano il consumo energetico sono sempre più diversificati.

Il modello utilizzato per stimare lo sbilanciamento zonale, in termini di errore percentuale, si basa sulla ricerca di giorni simili a livello orario. Questo avviene mediante l'utilizzo di una misura di similarità come la distanza euclidea, tra le variabili caratterizzanti, dopo un'attenta analisi ed elaborazione dei dati in ingresso. Questo approccio consente una previsione più accurata e adattabile alle variazioni complesse e dinamiche del consumo energetico.

## **2.1 Analisi dell'Influenza di Variabili Ambientali e Temporalis nello Sbilanciamento Zonale**

La selezione dei fattori che influenzano il consumo energetico riveste un ruolo fondamentale nell'accuratezza delle previsioni. Due dei principali fattori presi in considerazione includono la temperatura e il tipo di giorno.

1. La temperatura è un elemento chiave che influenza direttamente il consumo energetico. Le variazioni climatiche determinano l'attivazione di dispositivi di riscaldamento o di raffreddamento, come condizionatori d'aria e sistemi di riscaldamento, che influenzano significativamente la domanda di energia elettrica.
2. Il tipo di giorno svolge un ruolo determinante nelle variazioni del carico elettrico. Giorni diversi, come i giorni lavorativi e i giorni festivi, presentano modelli di consumo distinti. Ad esempio, durante i giorni lavorativi, si registra spesso un aumento del consumo energetico dovuto alla maggiore attività industriale e commerciale, mentre nei fine settimana si osserva una diminuzione della domanda energetica a causa della riduzione delle attività lavorative. Pertanto, comprendere e considerare il tipo di giorno è cruciale per ottenere previsioni accurate della domanda energetica.

Integrare sia i dati storici sulla temperatura che le informazioni sul tipo di giorno nelle previsioni del consumo energetico, consente di sviluppare modelli predittivi più robusti e accurati, in grado di adattarsi alle variazioni complesse e dinamiche del consumo energetico.



Le serie storiche orarie a disposizione di temperatura media, minima e massima forniscono preziose informazioni su come il consumo energetico possa variare in risposta alle condizioni termiche. Per ottenere una stima della temperatura per l'aggregata zonale, si utilizza la media ponderata delle temperature delle province di appartenenza dei POD. Questo significa che, anziché considerare semplicemente la temperatura media di ciascuna provincia, si tiene conto del consumo annuo dei POD in ciascuna provincia e si pesa di conseguenza la temperatura media. Inoltre, è essenziale disporre di previsioni accurate della temperatura per il giorno successivo al fine di integrare questo fattore nelle previsioni di consumo energetico.

La considerazione della temperatura e del tipo di giorno rappresenta un passo cruciale nell'analisi e nella previsione del consumo energetico. Per comprendere appieno l'impatto di questi fattori, vengono create specifiche variabili per ogni fascia oraria.

Innanzitutto, per quanto la temperatura, vengono generati due insiemi di variabili relative a giorni precedenti al giorno in esame ( $t$ ) per ogni fascia oraria. Queste variabili includono la differenza di temperatura media, minima e massima tra il giorno  $t$  e i giorni  $t - 2$  e  $t - 4$ . Questa analisi consente di cogliere le tendenze a breve e medio termine delle variazioni di temperatura, e di come queste influenzino il consumo energetico.

Passando all'analisi del tipo di giorno, vengono create diverse variabili, comunemente chiamate "*dummy di calendario*", per catturare le caratteristiche temporali del giorno  $t$  in esame:

- Mese;
- Anno;
- Giorno della settimana: lunedì, martedì, mercoledì . . . ;
- Giorno del mese: 1,2,3, . . . , 28/30/31;
- Giorno dell'anno: 1,2,3, . . . , 365;
- Giorno lavorativo: il giorno viene classificato come lavorativo, se il giorno cade durante la settimana e non è una festività, come festivo oppure in particolare come sabato non festivo;
- Giorno di "ponte": il giorno viene classificato come lavorativo, festivo, giorno di ponte pre festivo se è un giorno lavorativo che precede direttamente una festività, giorno di ponte post festivo se è un giorno lavorativo che segue direttamente una festività. . . ;
- Giorno appartenente ad una settimana particolare: il giorno viene classificato in base all'appartenza o meno ad una settimana particolare, come ad esempio le settimane centrali del mese di Agosto.

Queste variabili consentono di considerare in modo esaustivo e dettagliato il contesto temporale in cui si verifica il consumo energetico, permettendo al modello di previsione di catturare le complesse dinamiche che influenzano la domanda di energia elettrica

### 2.1.1 Analisi dei fattori influenzanti

Come primo step, si definiscono le serie storiche dello sbilanciamento orario e dell'errore percentuale, una per ciascuna ora del giorno utilizzando le serie storiche di consumo e di previsione specifiche per la tipologia di cliente finale. Nel contesto descritto, tutte le Figure successive si riferiscono alla tipologia MO per la zona NORD.

Per definire formalmente lo sbilanciamento zonale e l'errore percentuale, nel contesto delle previsioni del consumo energetico, utilizziamo le seguenti notazioni, fissata l'unità  $i$ :

- $C_{t,h,i}$  rappresenta il valore di consumo zonale al giorno  $t$  periodo di carico  $h$ .
- $F_{t,h,i}$  rappresenta il valore di previsione zonale al giorno  $t$  periodo di carico  $h$ .

Si definiscono lo sbilanciamento zonale e l'errore percentuale con le seguenti equazioni:

$$\begin{aligned} S_{t,h,i} &= F_{t,h,i} - C_{t,h,i} \\ E_{t,h,i} &= \frac{F_{t,h,i} - C_{t,h,i}}{F_{t,h,i}} * 100 \end{aligned} \quad (2.1)$$

Dove:

- $S_{t,h,i}$  è lo sbilanciamento zonale dell'unità  $i$  al giorno  $t$  e periodo di carico  $h$ . Questa misura quantifica la differenza tra il consumo previsto e quello reale per un determinato periodo di carico e giorno. Un valore positivo indica una sovrastima del consumo (previsione maggiore del consumo reale), mentre un valore negativo indica una sottostima (previsione minore del consumo reale).
- $E_{t,h,i}$  è l'errore percentuale dello sbilanciamento zonale dell'unità  $i$  al giorno  $t$  e periodo di carico  $h$ , esprimendo la differenza tra il valore di previsione e il consumo reale come percentuale della previsione. Viene espresso come percentuale del valore della previsione per motivi che saranno chiari in seguito. Questa misura è utile per normalizzare l'errore e renderlo comparabile tra diversi periodi di carico e giorni.

Osservando la Figura 2.1 si decide, per le analisi successive, di concentrarsi sulle seguenti fasce orarie: due fasce di picco, ovvero le 12:00 e le 18:00, e due fasce non di picco, ovvero le 00:00 e le 06:00. Nella Figura 2.2 sono riportate le serie storiche dello sbilanciamento relative alle fasce orarie sopra indicate.

Per garantire un'analisi più approfondita, è stata scelta infatti una rappresentazione grafica dettagliata che permette di individuare con precisione i momenti di maggiore e di minore sbilanciamento. Le fasce orarie di picco sono state selezionate perché rappresentano i periodi di maggiore attività, mentre le fasce non di picco rappresentano periodi di minore attività, permettendo così un confronto completo delle dinamiche dello sbilanciamento nel corso della giornata.

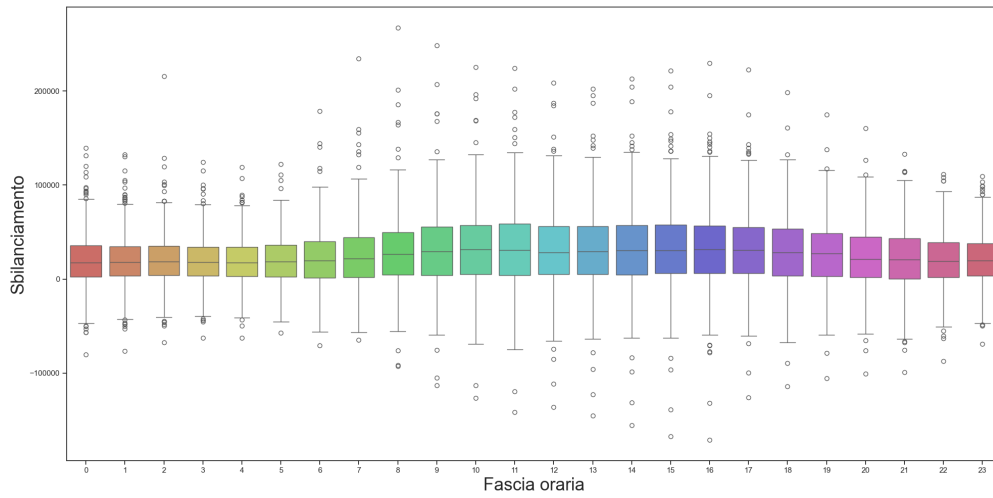


Figura 2.1: Distribuzione sbilanciamento storico per tipologia MO, in funzione della fascia oraria.

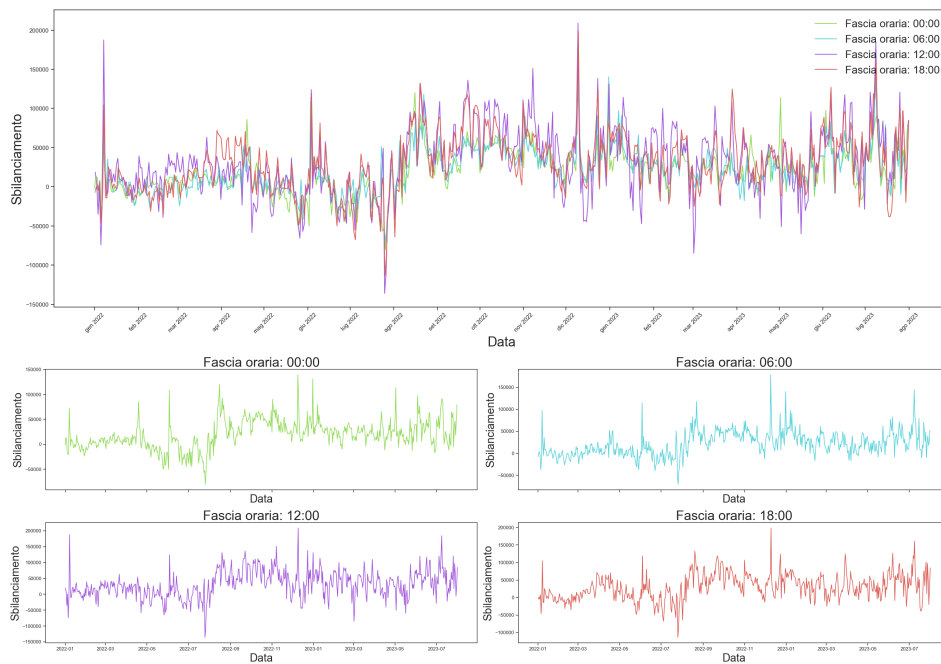


Figura 2.2: Sbilanciamento zonale storico, per tipologia MO, di alcune fasce orarie di picco e no.

Viene effettuata, dunque, un'analisi esplorativa degli altri dati a disposizione per valutare come, e in che misura, i diversi fattori sopracitati influenzano lo sbilanciamento.

Per quanto riguarda le variabili relative alla temperatura, individuata la linea di tendenza, la Figura 2.4a suggerisce una correlazione negativa nei mesi estivi, mentre la Figura 2.4b indica una correlazione positiva nei mesi invernali; dalla Figura 2.3 si osserva che le fasce "estreme" di temperatura influiscono maggiormente sullo sbilanciamento rispetto alle fasce intermedie.

Queste osservazioni sono fondamentali per comprendere l'impatto stagionale delle temperature sullo sbilanciamento e per identificare le condizioni in cui tali variabili hanno un'influenza più significativa.

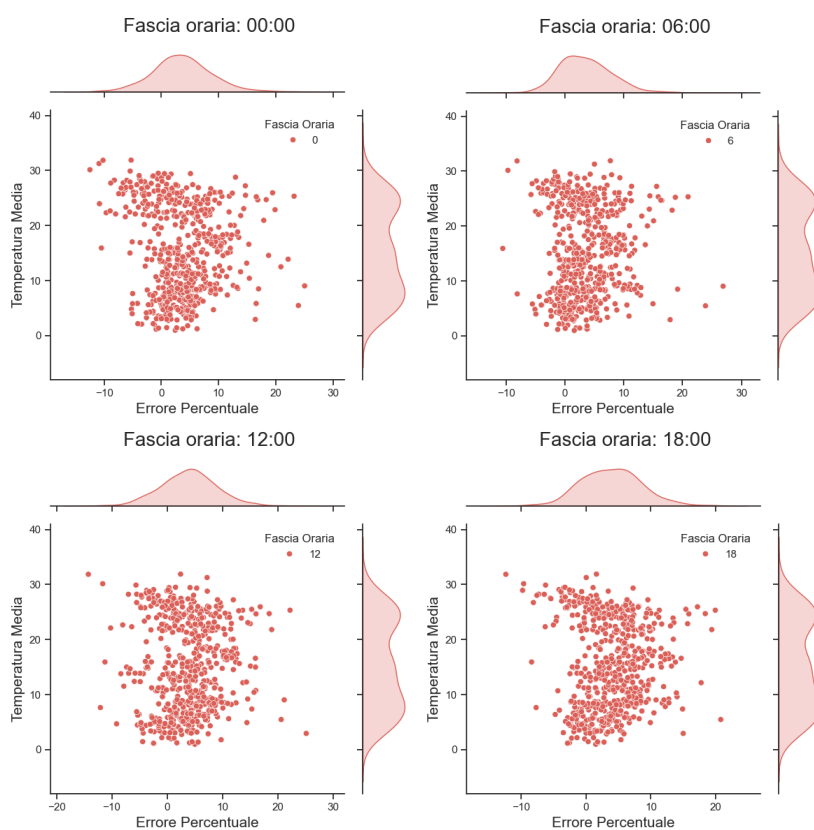
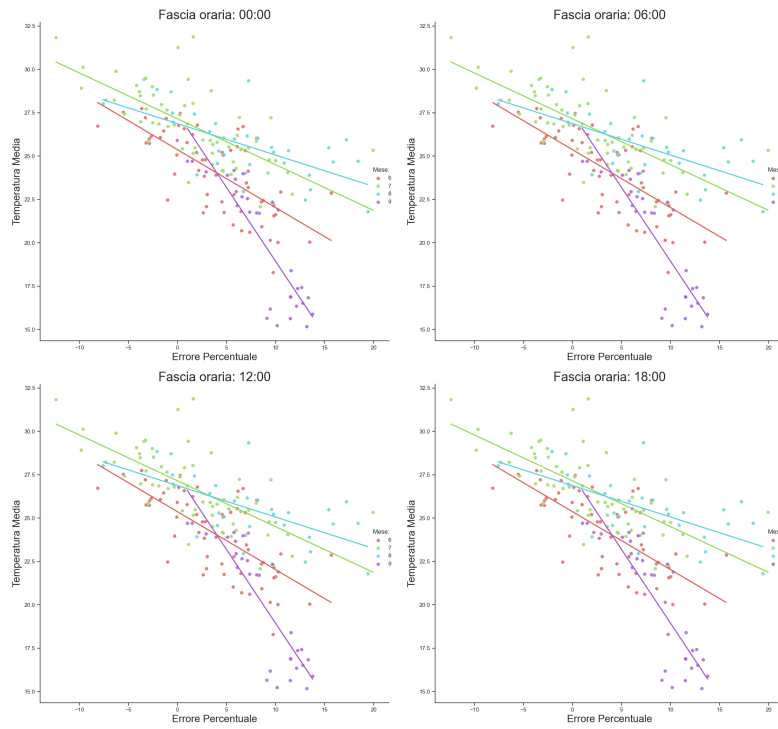
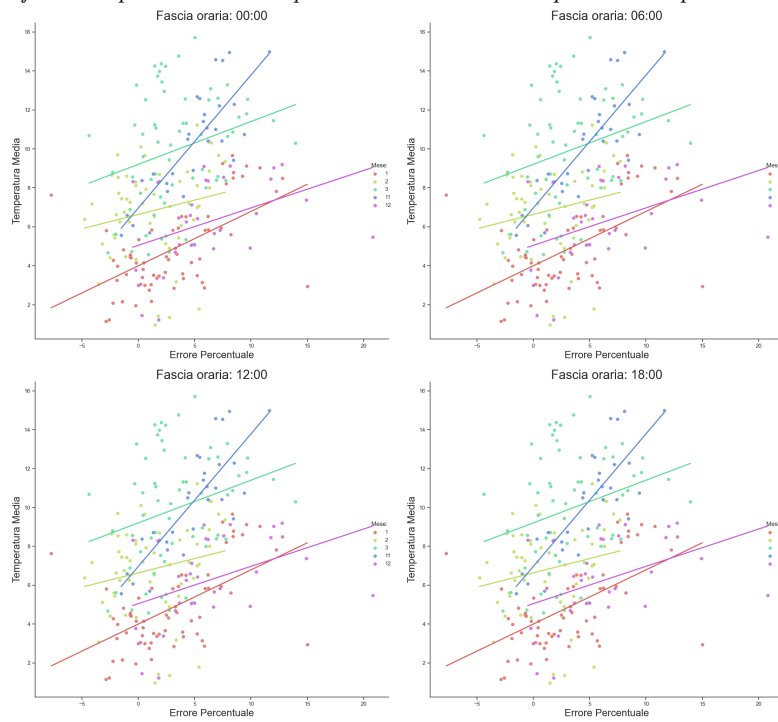


Figura 2.3: Grafico a dispersione tra temperatura media e errore percentuale.



(a) Grafico a dispersione tra temperatura media e errore percentuale per i mesi estivi.



(b) Grafico a dispersione tra temperatura media e errore percentuale per i mesi invernali.

Figura 2.4

Quando si considerano le variabili relative alle variazioni di temperatura, si osserva che ai bruschi cambiamenti di temperatura corrispondono altrettanti cambiamenti di sbilanciamento. Alcuni esempi di questi casi sono riportati nella Figura 2.5.

Questa relazione evidenzia come le variazioni repentine della temperatura possano avere un impatto immediato sullo sbilanciamento, suggerendo la necessità di monitorare attentamente tali cambiamenti per prevedere e mitigare eventuali effetti negativi. L'analisi di questi casi specifici fornisce ulteriori spunti per comprendere le dinamiche tra temperatura e sbilanciamento, rafforzando l'importanza di considerare le variabili climatiche nelle strategie.

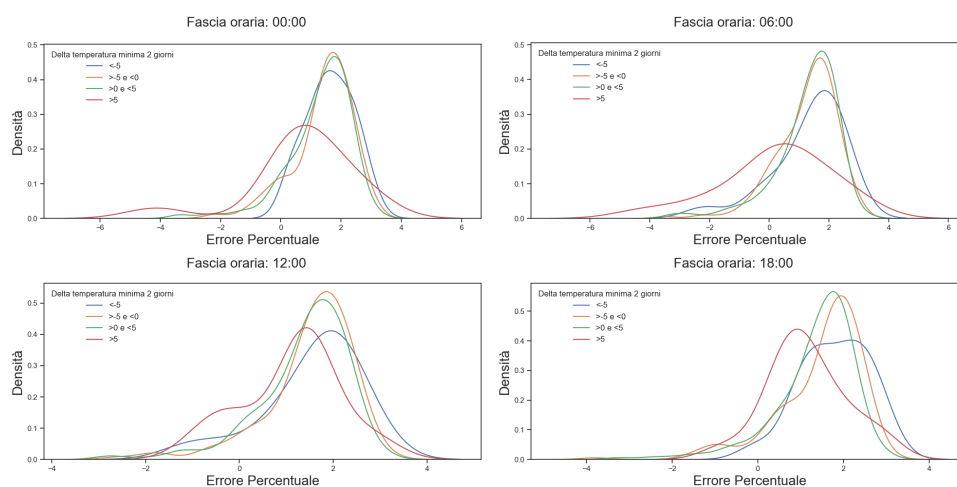


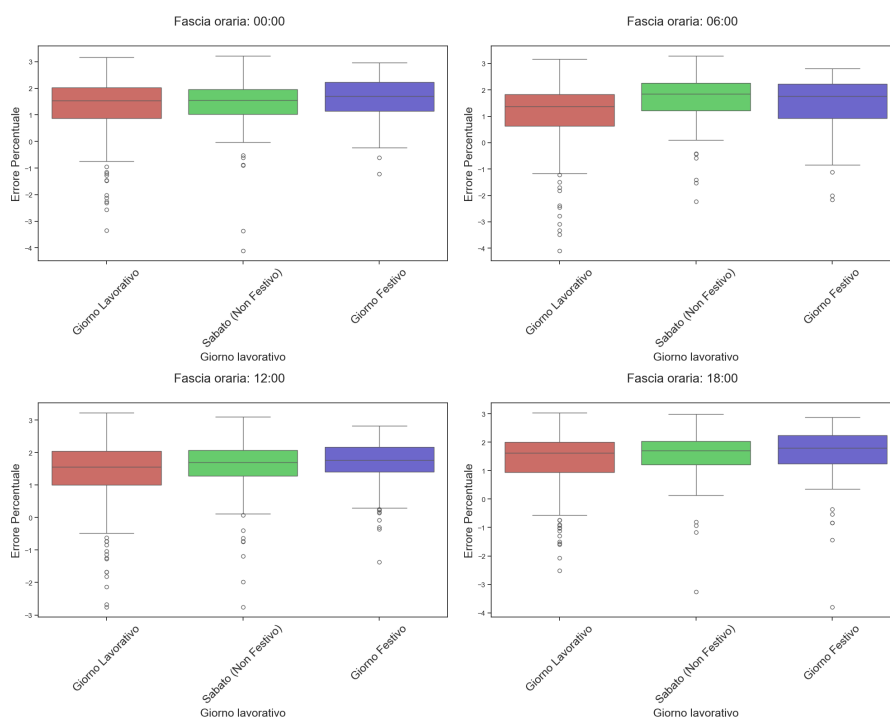
Figura 2.5: Distribuzione errore percentuale rispetto al delta temperatura.

Per comprendere appieno l'interazione tra il tipo di giorno e lo sbilanciamento, è fondamentale studiare la distribuzione dell'errore percentuale in relazione ai diversi valori che ciascuna variabile può assumere.

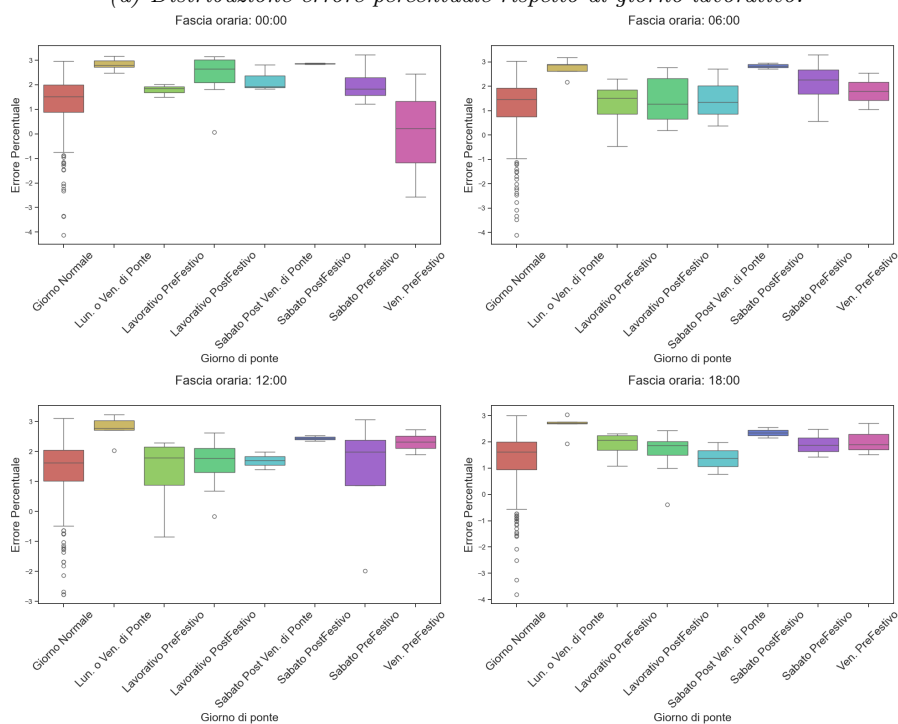
Nella Figura 2.6, vengono presentati i risultati relativi alle variabili considerate più significative. Emergono chiaramente le differenze di sbilanciamento tra i diversi tipi di giorno. Ad esempio, la Figura 2.6a, evidenzia uno sbilanciamento medio maggiore nei giorni festivi rispetto a quelli lavorativi, probabilmente a causa di una domanda energetica meno prevedibile durante i giorni di festa.

Osservando le Figure 2.6b e 2.6c, si nota una variazione significativa nella distribuzione dello sbilanciamento tra i giorni considerati normali e quelli specifici delle varie categorie. Questa analisi mette in luce l'importanza di considerare non solo il tipo di giorno generale, ma anche le categorie più specifiche, come ad esempio i giorni di ponte o quelli con caratteristiche particolari.

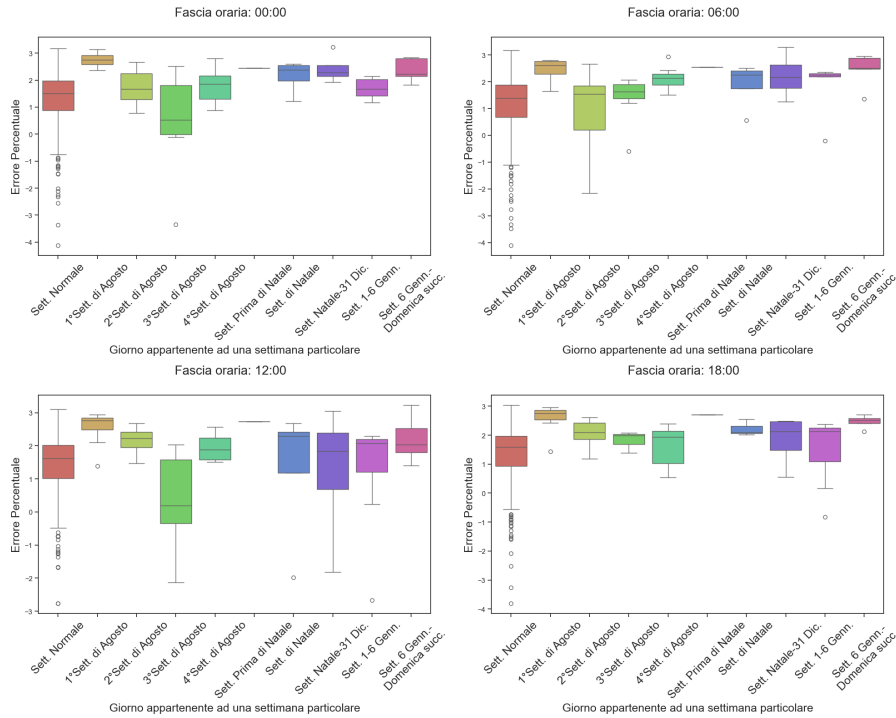
Comprendere queste differenze di sbilanciamento è cruciale per sviluppare modelli di previsione più accurati e adottare strategie di gestione flessibili che possano adattarsi alle variazioni della domanda energetica in base al tipo di giorno.



(a) Distribuzione errore percentuale rispetto al giorno lavorativo.



(b) Distribuzione errore percentuale rispetto al giorno di ponte.



(c) Distribuzione errore percentuale rispetto al giorno appartenente ad una settimana particolare.

Figura 2.6: Distribuzione errore percentuale rispetto alle variabili caratterizzanti il giorno tipo.

### 2.1.2 Approccio adimensionale per l'applicazione dell'algoritmo

Per applicare l'algoritmo dei giorni simili a livello orario, utilizzando la misura di similarità, è necessario saper confrontare le diverse caratteristiche, sia categoriche che non. Pertanto, è essenziale trattare ciascun fattore in modo adimensionale.

Solitamente, dopo aver codificato numericamente le variabili categoriche, tali caratteristiche dovrebbero essere mappate nell'intervallo  $[0,1]$ . Tuttavia, per enfatizzare il ruolo predominante di certi fattori nell'influencare lo sbilanciamento, si opta per una mappatura nell'intervallo  $[0, a]$ , dove  $a > 1$  [13]. La specifica mappatura, così come il valore di  $a$ , cambia per ogni variabile. Maggiore è l'influenza che la caratteristica ha sullo sbilanciamento, maggiore sarà il valore di  $a$ ; di conseguenza, tale fattore assume un'importanza maggiore nella determinazione della distanza euclidea.

Questo approccio consente di assegnare un peso differenziato alle caratteristiche, tenendo conto della loro rilevanza nell'analisi dello sbilanciamento. Ad esempio, se una determinata variabile categorica ha un impatto significativo sullo sbilanciamento, essa verrà mappata in un intervallo più ampio, riflettendo così la sua maggiore influenza nel calcolo della similarità tra giorni. Tale strategia migliora l'efficacia dell'algoritmo nell'identificare pattern e correlazioni significative nei dati orari, facilitando così la miglior riuscita della previsione.

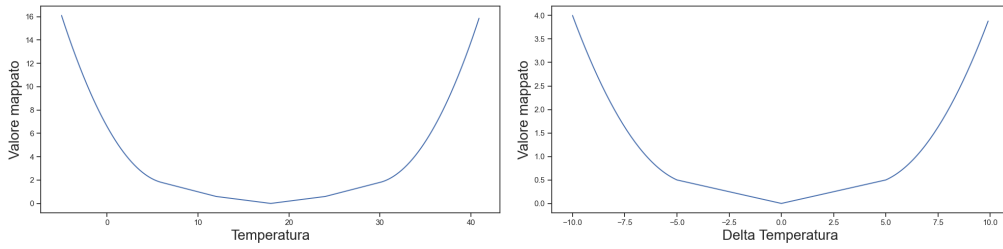


Per quanto riguarda la temperatura, osservando le Figure 2.3 e 2.5, si può notare chiaramente che diverse fasce di temperatura ( $<6$  °C, 6-12 °C, 12-24 °C, 24-30 °C,  $>30$  °C) influenzano lo sbilanciamento del sistema in modi distinti. Questa osservazione sottolinea l'importanza di applicare modalità di mappatura diverse per ciascuna fascia di temperatura, al fine di riflettere accuratamente l'impatto relativo delle varie condizioni termiche. È importante notare che tali modalità di mappatura sono uniformi per ciascuna delle 24 fasce orarie considerate nell'analisi.

- **Fascia "base" (12-24 °C).** Quando la temperatura si trova tra 12 e 24 gradi Celsius, si utilizza una rappresentazione lineare con cambiamenti di valore minimi. Questa fascia di temperatura è considerata ottimale e stabile, con un impatto relativamente basso sullo sbilanciamento del sistema. La rappresentazione lineare in questa fascia implica che piccoli cambiamenti nella temperatura non provocano variazioni significative nei valori rappresentati, riflettendone la prevedibilità.
- **Fasce intermedie (6-12 °C e 24-30 °C).** Per le temperature comprese tra 6 e 12 gradi Celsius e tra 24 e 30 gradi Celsius, viene adottata una rappresentazione lineare, ma con un cambiamento di valore più significativo rispetto alla fascia base. Queste temperature, pur non essendo estreme, possono avere un impatto maggiore sullo sbilanciamento, richiedendo quindi una risposta più sensibile. Ad esempio, un incremento di temperatura all'interno di queste fasce comporta variazioni più evidenti nei valori rappresentati.
- **Fasce "importanti" ( $<6$  °C e  $>30$  °C).** Quando la temperatura è al di sotto dei 6 gradi Celsius o al di sopra dei 30 gradi Celsius, viene adottata una rappresentazione non lineare. In queste condizioni estreme, ogni aumento, o diminuzione, di grado comporta un cambiamento sempre più significativo nei valori rappresentati. Per le temperature molto elevate, ad esempio, un piccolo incremento di temperatura può provocare un grande aumento nei valori di sbilanciamento, riflettendo l'impatto potenzialmente drammatico sul sistema. Analogamente, per le temperature molto basse, anche un lieve calo può comportare variazioni rilevanti, sottolineando la necessità di una gestione attenta.

Un ragionamento simile viene applicato anche ai delta di temperatura. Un delta di temperatura compreso tra 0 e 5 gradi Celsius (in valore assoluto) è considerato "base", indicando che tali variazioni sono generalmente gestibili e hanno un impatto limitato sul sistema. Un delta superiore a 5 gradi Celsius, invece, è considerato "importante", poiché variazioni di questa entità possono influenzare significativamente il bilancio del sistema. Questo approccio permette di distinguere meglio le variazioni di temperatura che potrebbero avere effetti più rilevanti, garantendo una gestione più accurata e responsiva alle condizioni ambientali variabili.

Per quanto riguarda le variabili caratterizzanti il giorno tipo (come giorno lavorativo, giorno di "ponte" e giorno appartenente a una settimana particolare) l'impatto di ciascuna classe della variabile sullo sbilanciamento è direttamente proporzionale al suo valore di mappatura. In altre parole, maggiore è l'importanza di una variabile caratterizzante



(a) Mappatura dei valori di temperatura. (b) Mappatura dei valori di delta temperatura.

Figura 2.7

giorno tipo, maggiore sarà il valore in cui viene mappata. Ad esempio, se si considera la variabile relativa al giorno lavorativo (2.6a), poiché è evidente che il giorno festivo ha un impatto significativo sullo sbilanciamento, viene mappato a 2; il giorno normale e il sabato, che hanno un impatto inferiore, vengono mappati rispettivamente a 0 e 0.2. Questa differenziazione nei valori di mappatura riflette l'entità dell'impatto di ciascuna classe e ciascun fattore sullo sbilanciamento del sistema.

Analogamente, viene gestita la variabile relativa al mese, con l'obiettivo di enfatizzarne l'importanza nella determinazione dello sbilanciamento. Le mappature per queste variabili vengono applicate uniformemente per ciascuna delle 24 fasce orarie considerate, garantendo così coerenza e nel processo di analisi e rappresentazione.

Per le variabili temporali, come il giorno della settimana, il giorno del mese e il giorno dell'anno, si utilizza una funzione di mappatura basata sulla media dello sbilanciamento per ogni valore della variabile. Ogni fascia oraria ha la propria funzione di mappatura, poiché dipende numericamente dalle serie storiche di sbilanciamento. Considerando il giorno della settimana come esempio, la funzione di mappatura è definita come segue:

$$\text{Dizionario di mappatura} = \left\{ i : \frac{\text{media}_i}{\max(\text{media}_j \text{ per } j = 1, 2, \dots, 7)} \quad \text{per } i = 1, 2, \dots, 7 \right\}$$

In questa equazione,  $i$  rappresenta ciascun giorno della settimana (da 1 a 7, ad esempio da Lunedì a Domenica). Per ogni giorno della settimana, si calcola la media dello sbilanciamento nel corso del tempo e si normalizza dividendo per il massimo tra le medie di tutti i giorni della settimana. Questo assicura che i valori mappati siano compresi tra 0 e 1, riflettendo l'importanza relativa di ciascun giorno della settimana nello sbilanciamento complessivo.

Questo processo di mappatura viene applicato sia ai valori delle serie storiche che ai valori delle variabili del giorno e del periodo di carico per i quali si desidera effettuare la stima dello sbilanciamento. Questo garantisce una coerenza nel trattamento delle variabili temporali e nella valutazione dell'impatto di ciascun valore sulla stima dello sbilanciamento.

## 2.2 Modello di Stima: Algoritmo Giorni Simili

Il modello utilizzato per stimare lo sbilanciamento zonale, espresso in termini di errore percentuale, si basa sulla ricerca dei giorni simili a livello orario, utilizzando la distanza euclidea tra le variabili che caratterizzano ciascun giorno e periodo di carico.

Una volta ottenuto il valore di mappatura delle caratteristiche di ciascun fattore, è possibile introdurre la misura di similarità tra due giorni diversi, fissato il periodo di carico. Questo consente di valutare quanto due giorni siano simili o meno tra loro in termini delle loro caratteristiche rilevanti. Per definire formalmente questa similarità, si utilizza l'indicatore quantitativo  $\mathbf{X}_i$  del giorno  $i$ , fissato il periodo di carico.  $\mathbf{X}_i$  è un vettore definito come  $\mathbf{X}_i = (x_{i,1}, x_{i,2}, \dots, x_{i,n})$ , dove  $x_{i,j}$  rappresenta il valore mappato delle  $n$  caratteristiche del giorno  $i$ . La misura di similarità tra due giorni  $i$  e  $k$ , ovvero la distanza euclidea, è calcolata come segue:

$$d_{i,k} = (\mathbf{X}_i, \mathbf{X}_k) = \sqrt{(x_{i,1} - x_{k,1})^2 + (x_{i,2} - x_{k,2})^2 + \dots + (x_{i,n} - x_{k,n})^2} \quad (2.2)$$

Qui,  $d_{i,k}$  rappresenta la distanza tra le caratteristiche dei due giorni nello spazio  $n$ -dimensionale. Un valore più piccolo di  $d_{i,k}$  indica una maggiore similarità tra i fattori del giorno  $i$  e quelli del giorno  $k$ . Viceversa, un valore più grande corrisponde a una minore somiglianza tra i due giorni.

Questo processo di calcolo della similarità tra i giorni fornisce una base importante per il modello di previsione dello sbilanciamento zonale, permettendo di identificare i giorni storici più rilevanti per la previsione accurata dello sbilanciamento futuro.

Data la serie storica di sbilanciamento in termini di errore percentuale di riferimento (2.1), data la tipologia di cliente finale e il periodo di carico  $h$ , si calcola la distanza euclidea tra il giorno  $t$  in esame e ogni giorno appartenente alla serie storica  $(0, \dots, t-2, t-1)$ . Il calcolo della distanza euclidea consente di valutare quanto ogni giorno storico sia simile al giorno di previsione, fornendo una misura della rilevanza per la previsione dello sbilanciamento futuro. Alcuni giorni storici con bassa similarità con il giorno di previsione vengono scartati; vengono considerati solo i giorni con elevata similarità. Questo processo di selezione dei giorni storici si basa su una soglia predefinita per la distanza euclidea. Per impostare questa soglia, i valori delle distanze eucldee vengono ridimensionati in modo che il loro intervallo sia compreso tra 1 e 10. Questo processo di mappatura può essere eseguito utilizzando una funzione di trasformazione lineare che scali i valori originali in modo che il minimo diventi 1 e il massimo diventi 10. In questo elaborato, la distanza di soglia viene impostata a  $\mathbf{d} = 2.75$ . In pratica, ciò significa che i giorni con una distanza euclidea maggiore di 2.75 rispetto al giorno di previsione vengono esclusi dall'analisi, mentre quelli con una distanza inferiore vengono considerati per la previsione dello sbilanciamento. Questo approccio consente di concentrarsi sui giorni storici più simili al giorno di previsione, migliorando così l'accuratezza della previsione dello sbilanciamento zonale.

Considerando l'insieme di giorni simili, è possibile predire l'errore percentuale del giorno  $t$  attraverso una media pesata degli errori percentuali presenti nel campione. Per

calcolare la media pesata degli errori percentuali, si utilizza il reciproco della distanza euclidea come coefficiente di peso del giorno nel campione. In pratica, ciò significa che i giorni più simili al giorno di previsione ricevono un peso maggiore nella media pesata degli errori; questo permette di attribuire maggiore importanza all'errore percentuale dei giorni più simili. In altre parole, il modello tiene conto della rilevanza dei giorni simili nella previsione dell'errore percentuale per il giorno successivo, fornendo stime più precise e informative.

Vengono quindi calcolati i percentili (25%, 50%, 75%) della distribuzione dell'errore percentuale dell'insieme di giorni simili, tenendo in considerazione lo stesso coefficiente di peso. Questo passaggio consente di valutare la variabilità dell'errore percentuale nei giorni simili e di identificare i valori di errore più rappresentativi e significativi per il giorno successivo. Per calcolare i percentili, si ordinano i valori distinti dell'errore percentuale in modo crescente  $y_1 < y_2 < \dots$ , e si associano i corrispondenti pesi a ciascun valore  $w_1, w_2, \dots$ . Successivamente, si calcolano le somme cumulative dei pesi fino a ciascun valore dell'errore percentuale. Si definisca  $s_j$  come la somma dei primi  $j$  pesi e  $W$  come la somma di tutti i pesi. Queste somme cumulative vengono utilizzate per determinare i punti di probabilità associati ai percentili desiderati. In particolare, per un punto di probabilità  $p$ , se  $pW$  cade strettamente tra le somme cumulative  $s_j$  e  $s_{j+1}$ , allora il quantile stimato corrisponde al valore dell'errore percentuale associato a  $s_{j+1}$ ,  $y_{j+1}$ . Nel caso in cui  $pW$  coincida esattamente con una somma cumulativa  $s_j$ , il quantile stimato è la media dei valori dell'errore percentuale corrispondenti a  $s_j$  e  $s_{j+1}$ ,  $\frac{(y_j + y_{j+1})}{2}$ . Se  $pW$  è inferiore al primo punto di probabilità  $p_1$ , il quantile stimato corrisponde al valore minimo dell'errore percentuale, ovvero  $y_1$ .

## 2.3 Analisi Risultati e Valutazione delle Prestazioni

### Caso di studio:

Nel caso di studio considerato, incentrato sull'unità  $i$  situata nella regione geografica Nord, una delle sette zone definite, si è avviata un'analisi dettagliata delle serie storiche a partire dal 1° Febbraio 2022. Queste serie storiche riguardano l'errore percentuale registrato per ogni ora del giorno, differenziato per ciascuna tipologia di cliente finale, escludendo l'illuminazione pubblica. L'obiettivo principale è di anticipare e quantificare lo sbilanciamento energetico in termini di errore percentuale per il mese di Agosto 2023.

Dopo aver interrotto la serie storica al 31 Luglio 2023, si è proceduto con un test retrospettivo (*backtest*) utilizzando dati coerenti con quelli disponibili nell'ambiente del sistema di previsione energetica. Questo processo è fondamentale per validare l'efficacia delle previsioni e garantire la coerenza dei risultati con le reali dinamiche di mercato.

I risultati di questo processo di previsione e validazione vengono rappresentati in Figure specifiche 2.8, dove è possibile osservare anche l'andamento previsto dello sbilanciamento energetico in relazione alla situazione effettiva. In particolare, per la stima dello sbilanciamento zonale dell'unità, si adotta un approccio che tiene conto della definizione dell'errore percentuale, moltiplicando la stima di tale errore per il valore di previsione corrispondente.

Per testare le prestazioni del modello di previsione, per confronto si sono testati diversi algoritmi di machine learning, tra cui LightGBM. Tuttavia, nonostante la complessità di alcuni modelli, i risultati ottenuti sono stati simili a un modello più semplice. Pertanto, si è scelto di utilizzare l'Algoritmo dei Giorni Simili per la sua accettabile precisione e per la sua comprensibilità. Inoltre questo algoritmo consente anche eventuali analisi specifiche sul campione di giorni simili ricavato.

Per valutare le prestazioni del modello di previsione si utilizzano due concetti distinti:

1. **Accuratezza della previsione:** questo concetto si riferisce alla capacità del modello di previsione di predire correttamente i valori effettivi delle variabili target. L'accuratezza della previsione è solitamente valutata attraverso metriche come l'errore assoluto medio (MAE), l'errore quadratico medio (RMSE), o l'errore percentuale medio (MAPE). Un modello con una bassa accuratezza della previsione (metriche di errore alte) produce previsioni che sono distanti dai valori effettivi, mentre un modello con un'alta accuratezza della previsione (metriche di errore basse) produce previsioni che sono vicine ai valori effettivi.
2. **Accuratezza Direzionale:** questo concetto si riferisce alla capacità del modello di previsione di prevedere correttamente la direzione dello sbilanciamento energetico rispetto alla situazione effettiva. Ad esempio, se il modello prevede un eccesso di produzione energetica e la situazione effettiva mostra effettivamente un eccesso, allora la previsione è corretta in termini direzionali. Allo stesso modo, se il modello prevede uno sbilanciamento negativo (cioè una carenza di produzione rispetto alla domanda) e la situazione effettiva mostra effettivamente uno sbilanciamento negativo, la previsione è accurata in termini direzionali. La previsione è errata negli altri casi.

Nelle Tabelle 2.1,2.2 e nella Figura 2.8 sono riportati anche i risultati di un modello semplice di base, cioè la media storica dell'errore percentuale, che è stata utilizzata come stima di riferimento per confrontare le prestazioni del modello di previsione implementato.

<b>Modello</b>	<b>MAE</b>	<b>Modello</b>	<b>MAE</b>
Alg. Giorni simili Err. Perc.	4.62	Alg. Giorni simili Err. Perc.	7.08
Alg. Giorni simili Sbilanciamento	32941.87	Alg. Giorni simili Sbilanciamento	11346.01
Media Storica Err. Perc.	5.41	Media Storica Err. Perc.	9.25
Media Storica Sbilanciamento	38594.03	Media Storica Sbilanciamento	15270.01

(a) MO

(b) PRA

Tabella 2.1: Accuratezza previsione

<b>Modello</b>	<b>Positivo/Positivo (True positive)</b>	<b>Negativo/Negativo (True Negative)</b>	<b>Accuratezza</b>
Alg. Giorni simili	537	70	0.82, 82%
Media Storica	545	0	0.73, 73%

(a) MO

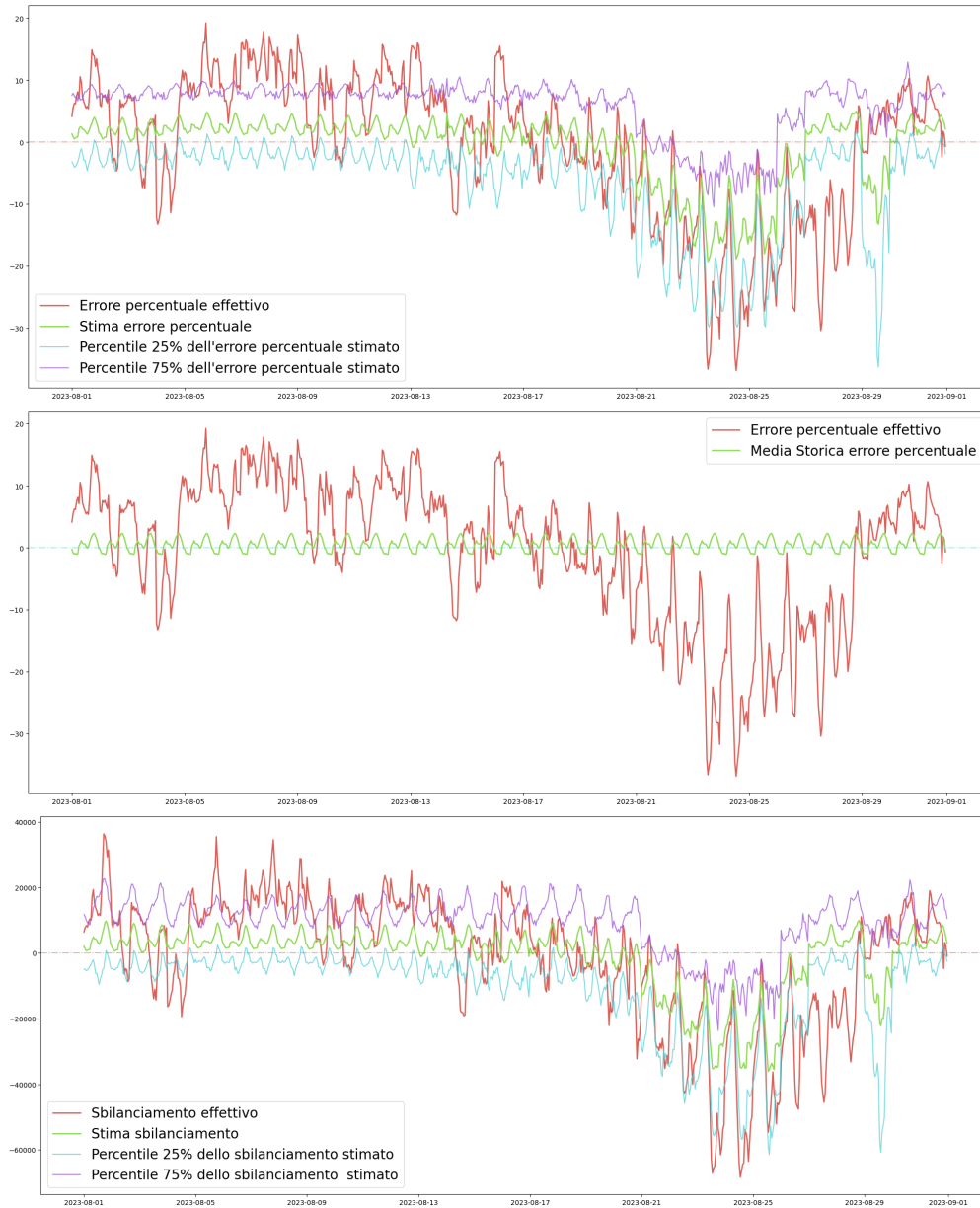
<b>Modello</b>	<b>Positivo/Positivo (True positive)</b>	<b>Negativo/Negativo (True Negative)</b>	<b>Accuratezza</b>
Alg. Giorni simili	372	175	0.74, 74%
Media Storica	279	103	0.52, 52%

(b) PRA

Tabella 2.2: Accuratezza direzionale previsione. (Errore percentuale stimato / Errore percentuale effettivo)



(a) MO



(b) PRA

Figura 2.8: Risultati stima errore percentuale e stima sbilanciamento per ciascuna delle tipologie cliente finale.



## Capitolo 3

# Analisi e Previsione del Segno Macrozonale di Sbilanciamento

Considerando una delle due macrozone in cui l'Italia è attualmente suddivisa (Nord e Sud), lo squilibrio macrozonale e quindi il segno di sbilanciamento macrozonale vengono determinati dall'aggregazione degli squilibri delle singole unità all'interno della stessa macrozona. Questo processo di aggregazione permette di ottenere una visione complessiva dello sbilanciamento energetico.

Per ciascuna delle macrozone è disponibile una serie storica oraria di sbilanciamento macrozonale. Questa situazione si differenzia dal capitolo precedente, dove erano disponibili separatamente le serie storiche di consumo e di previsione. Anche in questo contesto vengono considerate 24 serie storiche giornaliere, una per ogni ora del giorno. Tale approccio è preferito rispetto all'analisi delle serie storiche orarie poiché le serie presentano andamenti differenziati a seconda della fascia oraria, specialmente durante le ore di picco e quelle fuori picco. Analogamente al caso zonale, le ore di picco si riferiscono ai momenti di maggiore domanda energetica, durante la mattina e il tardo pomeriggio, mentre le ore fuori picco corrispondono ai periodi di minore domanda, la notte e le prime ore del pomeriggio.

Quindi, vengono esaminate le caratteristiche e la dinamica del segno di sbilanciamento macrozonale. Sulla base di tali analisi, si sviluppa un modello per la previsione del segno, assumendo che segua una distribuzione bernoulliana. Questo tipo di distribuzione è appropriata poiché il segno di sbilanciamento può assumere solo due valori: 0 (negativo) e 1 (positivo). La probabilità che il segno sia positivo viene stimata attraverso un'adeguata elaborazione dei dati in ingresso.

### 3.1 Esplorazione dello Sbilanciamento Macrozonale e delle Variabili Esogene

Essendo le serie storiche relative al consumo di energia elettrica, i fattori selezionati come variabili esogene nel modello di previsione sono simili a quelli considerati nel capitolo

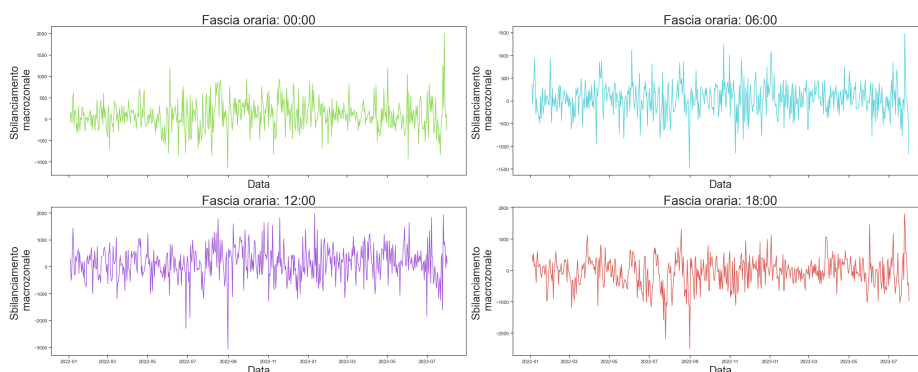


Figura 3.1: Sbilanciamento macrozonale storico per la macrozona Nord, relativo a diverse fasce orarie di picco e non.

precedente.

Tra queste variabili, particolare importanza è data alla temperatura e alle sue variazioni, oltre alle cosiddette *dummy di calendario*, che caratterizzano il tipo di giorno. Queste ultime includono:

- Giorno lavorativo;
- Giorno di "ponte";
- Giorno appartenente ad una settimana particolare.

Le Figure 3.2, 3.3, 3.4, illustrano esempi della distribuzione dello sbilanciamento orario macrozonale rispetto a queste variabili per la macrozona Nord. Viene riportata la frequenza del segno di sbilanciamento macrozonale al variare delle categorie di ciascuna delle variabili in esame. Questa rappresentazione visiva permette di comprendere come le diverse categorie di giorni influenzino lo sbilanciamento macrozonale.

Per quanto riguarda la temperatura sono disponibili serie storiche orarie che comprendono la temperatura minima, media e massima. In questo contesto, la temperatura di ciascuna macrozona è calcolata come la media delle temperature rilevate sull'intero territorio. Questo metodo garantisce una rappresentazione accurata delle condizioni climatiche che influenzano il consumo energetico. Nella Figura 3.5 è riportata la distribuzione della temperatura media rispetto al valore della variabile categorica segno macrozonale. Questa evidenzia come la temperatura media sia associata ai diversi segni di sbilanciamento macrozonale, fornendo utili indicazioni per l'analisi e la previsione del comportamento del sistema energetico.

Per utilizzare la temperatura e le relative variazioni come predittori nel modello è fondamentale disporre di previsioni meteorologiche accurate per i giorni di interesse. Questo permette di stimare in anticipo la temperatura e di integrarla nel modello come variabile esogena, migliorando così l'accuratezza delle previsioni per il giorno successivo o per il periodo futuro considerato. Le previsioni meteorologiche devono quindi essere raccolte e analizzate con attenzione per garantire la precisione del modello.

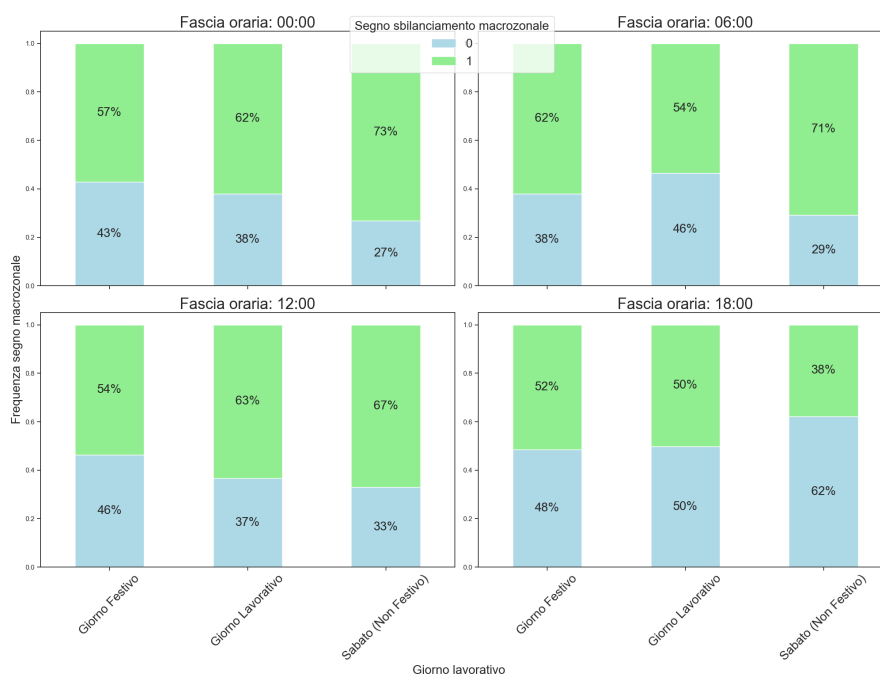


Figura 3.2: Frequenza del segno zonale rispetto ai valori di giorno lavorativo.

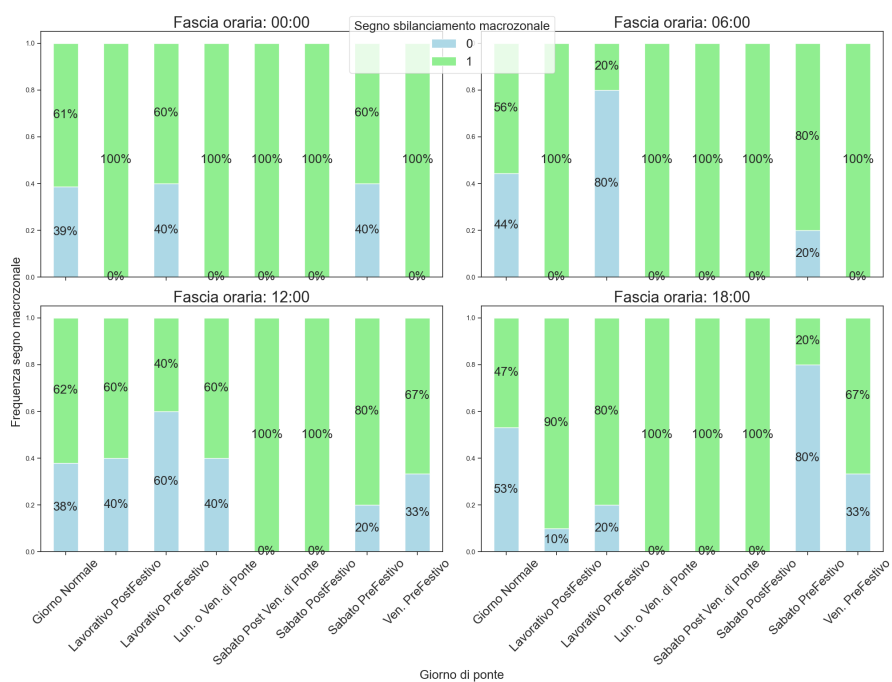


Figura 3.3: Frequenza del segno zonale rispetto ai valori di giorni di ponte.

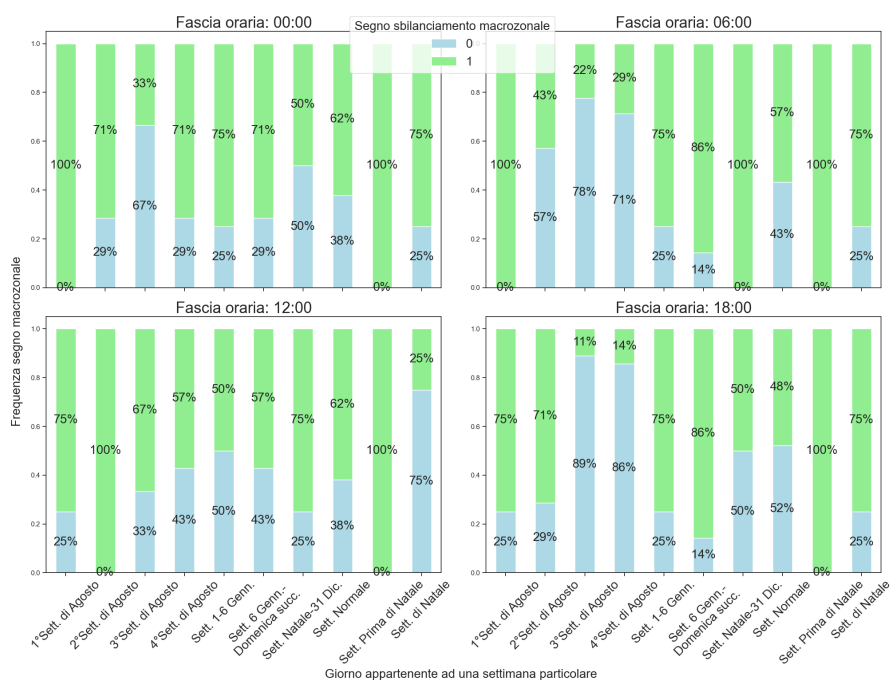


Figura 3.4: Frequenza del segno zonale rispetto ai valori delle settimane particolari.

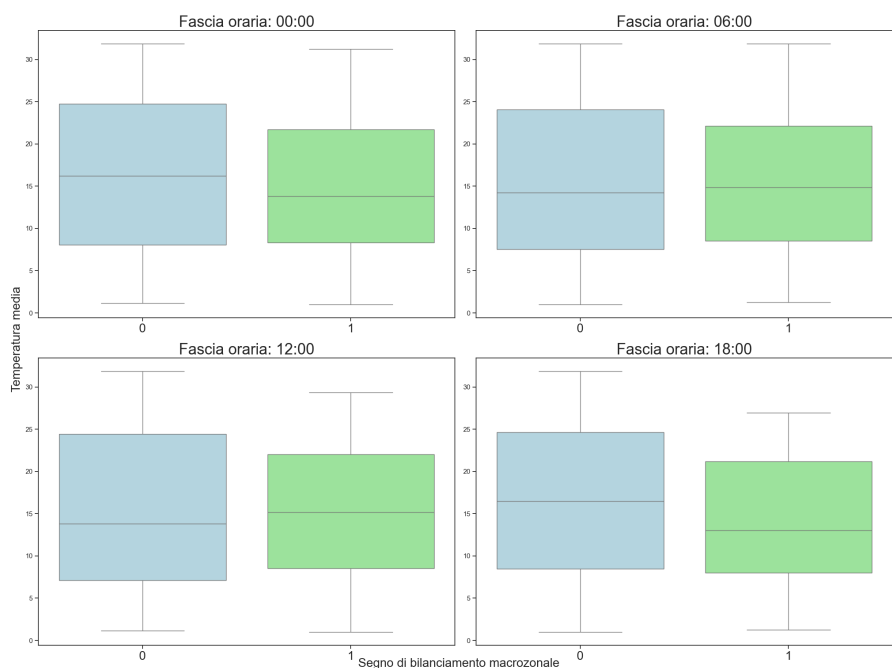


Figura 3.5: Distribuzione temperatura rispetto al segno di sbilanciamento della macrozona NORD.

### 3.1.1 Analisi preliminare Serie Storiche e Test di Dipendenza Markoviana

Prima di approfondire nuove analisi, è necessario formalizzare quanto descritto all'inizio del capitolo. Per una data macrozona, per ciascun giorno  $t$  e periodo di carico  $h$  si denota il volume di sbilanciamento macrozonale come  $S_{t,h}$ . Si omette quindi l'apice M (vedi capitolo 1), che denota la macrozona, e lo si utilizzerà solo quando necessario.

La variabile oggetto di studio è  $Y_{t,h}$  definita come

$$Y_{t,h} = \begin{cases} 1 & \text{se } S_{t,h} \geq 0 \\ 0 & \text{se } S_{t,h} < 0, \end{cases} \quad (3.1)$$

dove  $Y_{t,h}$  rappresenta la serie storica binaria del segno di sbilanciamento macrozonale.

Le analisi successive sono ispirate dall'articolo di [Lisi and Edoli](#). Per esaminare le caratteristiche delle serie storiche del segno macrozonale vengono condotti alcuni studi preliminari sull'intero periodo disponibile e su ogni periodo di carico della data macrozona.

Si denota come

$$\pi_{t,h} = \mathbb{P}(Y_{t,h} = 1) \quad (3.2)$$

la probabilità che il segno di sbilanciamento macrozonale sia positivo per il giorno  $t$ , il periodo di carico  $h$  e la data macrozona. Si indicano inoltre  $\pi_t$ ,  $\pi_h$  come le probabilità marginali rispetto al tempo e rispetto al periodo di carico.

Una stima della probabilità marginale di un evento positivo all'ora  $h$ , indicata con  $\pi_h$ , si ottiene calcolando la media di  $Y_{t,h}$  nel tempo, cioè

$$\hat{\pi}_h = \sum_{t=1}^T \frac{Y_{t,h}}{T} \quad (3.3)$$

Nella Figura 3.6, a sinistra, sono rappresentate le stime di  $\hat{\pi}_h$  per le 24 ore relative ad entrambe tutte le macrozone. Tale rappresentazione conferma che le probabilità di squilibri zonali positivi/negativi non rimangono costanti nel corso della giornata e che vi sono notevoli differenze tra le macrozone. Inoltre, un test statistico <sup>1</sup> per valutare l'ipotesi  $H_0 : \pi_h = 0.5$  conduce al rigetto di  $H_0$  a un livello di significatività del 5% nella maggior parte dei casi.

---

<sup>1</sup>**Test Z per Proporzioni:** il test Z è calcolato come la differenza tra la proporzione campionaria e la proporzione ipotizzata, divisa per la deviazione standard della proporzione campionaria. La formula è:

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 \cdot (1-p_0)}{n}}},$$

dove  $\hat{p}$  è la proporzione campionaria,  $p_0$  è la proporzione ipotizzata e  $n$  è la dimensione del campione. La formula è utilizzata per valutare se la differenza tra la proporzione campionaria e quella ipotizzata è statisticamente significativa.

Il pannello destro della Figura 3.6 illustra, invece, l'evoluzione nel tempo della probabilità  $\pi_t$ , stimata tramite finestre mobili di 60 giorni sovrapposte per 30 giorni, con un totale di 18 finestre, senza differenziazione oraria. Questa analisi temporale mostra che  $\pi_t$  ha subito variazioni significative nel corso degli anni e che le diverse macrozone presentano andamenti temporali differenti. Le divergenze tra le due macrozone potrebbero essere spiegate dalla distribuzione delle unità di produzione zonali, dalla congestione tra le zone, dalla posizione delle fonti di energia rinnovabile e dalla capacità di interconnessione tra le stesse.

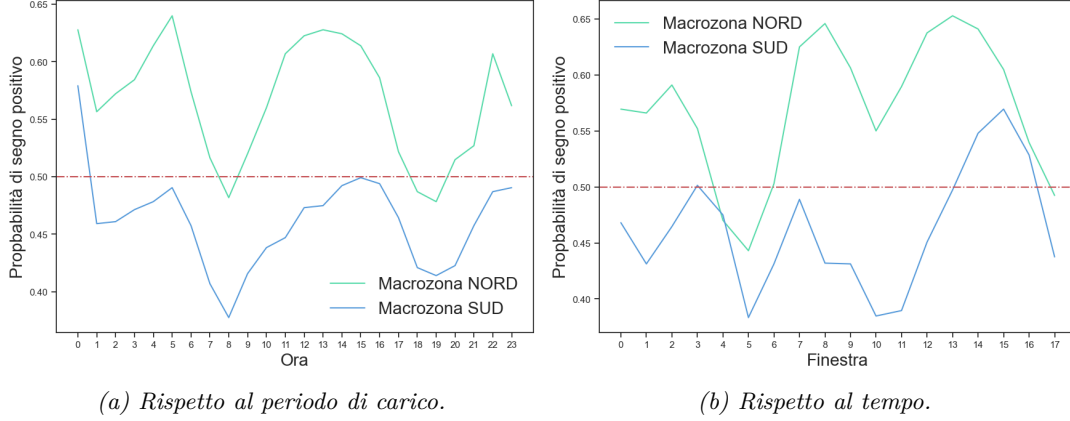


Figura 3.6: Stima della probabilità marginale di segno positivo

Per comprendere meglio la dinamica di lungo periodo delle serie temporali del segno di sbilanciamento macrozonale, si introduce la variabile esogena come fattore determinante.

Nel contesto dell'analisi della struttura di dipendenza delle serie temporali del segno dello sbilanciamento macrozonale nell'ambito della teoria delle catene di Markov, definiamo alcune probabilità chiave.

Denotiamo con  $\pi_u = P(Y_t = u)$  la probabilità marginale di  $Y_t$ , dove  $u$  può essere 0 o 1, indicando rispettivamente uno sbilanciamento negativo o positivo.

Le probabilità marginali congiunte sono indicate come  $\pi_{uv} = P(Y_t = u, Y_{t-1} = v)$  e  $\pi_{uvw} = P(Y_t = u, Y_{t-1} = v, Y_{t-2} = w)$ , dove  $u, v, w$  possono essere 0 o 1. Queste rappresentano le probabilità di osservare una certa sequenza di eventi nel tempo.

Inoltre, definiamo le probabilità condizionali rispetto all'ultima e alle ultime due osservazioni come  $\pi_{u|v} = P(Y_t = u | Y_{t-1} = v)$  e  $\pi_{u|vz} = P(Y_t = u | Y_{t-1} = v, Y_{t-2} = z)$ . Queste probabilità forniscono informazioni sullo stato corrente del sistema dato lo stato precedente o gli stati precedenti.

Queste definizioni ci consentono di analizzare in modo più dettagliato e approfondito la dinamica e le relazioni all'interno delle serie temporali dello sbilanciamento macrozonale.

Viene esaminata l'ipotesi di indipendenza rispetto all'alternativa di dipendenza markoviana del primo ordine:

$$\begin{cases} H_0 : \pi_{u|v} \neq \pi_u, \\ H_1 : \pi_{u|v} = \pi_u. \end{cases} \quad (3.4)$$

Questo sistema di ipotesi 3.4 viene testato mediante un test del rapporto di verosimiglianza (LR). Nell'ipotesi di indipendenza, la verosimiglianza è calcolata come

$$L_0 = (1 - \pi)^{n_{00}+n_{01}} \pi^{n_{10}+n_{11}}$$

mentre sotto l'ipotesi di dipendenza markoviana del primo ordine, si ottiene

$$L_1 = (1 - \pi_{01})^{n_{00}} \pi_{01}^{n_{01}} (1 - \pi_{11})^{n_{10}} \pi_{11}^{n_{11}}$$

dove  $n_{ij}$  indica la frequenza assoluta dell'occorrenza di  $(Y_{t,h} = j, Y_{t-1,h} = i)$  e  $p_{ij} = \frac{n_{ij}}{n_{i0}+n_{i1}}$ , con  $i, j \in \{0,1\}$ . Si sa che, in via asintotica, la statistica

$$LR_{ind} = -2 \log \frac{L_1}{L_0}$$

segue una distribuzione chi-quadro con un grado di libertà, che consente di effettuare il test. Per ottenere una comprensione più approfondita della dipendenza seriale presente nei dati, viene testata anche l'ipotesi di dipendenza markoviana del primo ordine rispetto all'alternativa di dipendenza markoviana del secondo ordine:

$$\begin{cases} H_0 : \pi_{u|vz} \neq \pi_{u|v}, \\ H_1 : \pi_{u|vz} = \pi_{u|v}. \end{cases} \quad (3.5)$$

Successivamente, con la stessa logica, sono considerati ordini crescenti di dipendenza markoviana. I test sono sempre di tipo LR con distribuzioni asintotiche chi-quadro e sono stati applicati alle 24 serie giornaliere dei segni di sbilanciamento per ciascun periodo di carico e per le due macrozone.

Il test del sistema di ipotesi 3.4 eseguito separatamente per ogni ora, ha portato al rigetto dell'ipotesi di indipendenza per tutti i periodi di carico e per tutte le macrozone. I valori  $p$  ottenuti per ogni singolo test sono generalmente inferiori a 0,001. Per quanto riguarda il sistema di ipotesi 3.5, l'ipotesi nulla è stata sempre respinta al livello di significatività del 5%, tranne che per alcuni periodi di carico. I risultati suggeriscono dunque una dipendenza markoviana del secondo ordine.

La situazione cambia quando la dipendenza markoviana del secondo ordine viene confrontata con la dipendenza del terzo ordine: sebbene in diversi casi l'ipotesi nulla venga respinta a favore di una dipendenza del terzo ordine al livello di significatività del 5%, in altri casi viene accettata.

Pertanto, basandosi su queste analisi preliminari, possiamo concludere che la probabilità  $\pi_{t,h}$  dipende sia dal periodo di carico che dalla macrozona, oltre che dal tempo. Inoltre, i segni di sbilanciamento macrozonale non sono serialmente indipendenti, ma presentano almeno una dipendenza markoviana del secondo ordine. Considerando questa dipendenza tra i segni di squilibrio, si adotta un approccio basato su modelli autoregressivi, utilizzando i valori precedenti come predittori. Questo permette di catturare la relazione sequenziale tra le osservazioni e di sfruttare l'informazione storica per fare previsioni future.

I dati relativi ai segni di sbilanciamento macrozonale orari ( $Y_{t,h}$ ) sono disponibili direttamente dal portale di Terna S.p.A. Tuttavia, questi dati vengono pubblicati solamente nel tardo pomeriggio del giorno  $t + 1$ , rendendo non immediatamente disponibili gli ultimi valori della serie storica al momento opportuno. Oltre ai segni di squilibrio passati, un'altra variabile importante che potrebbe essere inclusa nei modelli è la dimensione algebrica degli stessi squilibri ( $S_{t,h}$ ). È possibile che squilibri positivi (negativi) di grandi dimensioni abbiano un impatto più significativo sulla probabilità di uno squilibrio positivo (negativo) per il giorno successivo rispetto a squilibri positivi (negativi) di piccole dimensioni.

Quindi, per prevedere la probabilità del segno positivo per il tempo  $t$ , si ha a disposizione la serie storica fino a  $t - 3$ . Tuttavia, Terna S.p.A. mette a disposizione un valore preliminare di sbilanciamento macrozonale ( $\hat{S}_{t,h}$ ). Prima di utilizzare il valore preliminare di sbilanciamento macrozonale nel modello di previsione, è stata condotta un'analisi di correlazione <sup>2</sup> con il valore reale di sbilanciamento macrozonale. Questa analisi ha lo scopo di valutare la relazione tra i due valori e determinare se esista una correlazione significativa tra loro.

I risultati dell'analisi di correlazione evidenziano la presenza di una correlazione positiva tra il valore preliminare e il valore reale di sbilanciamento macrozonale. Questo suggerisce che il valore preliminare può essere un buon indicatore dello sbilanciamento macrozonale e può essere quindi utilizzato come variabile predittiva nel modello di previsione, ad introdurre autoregressività. La Figura 3.7 illustra graficamente questa correlazione.

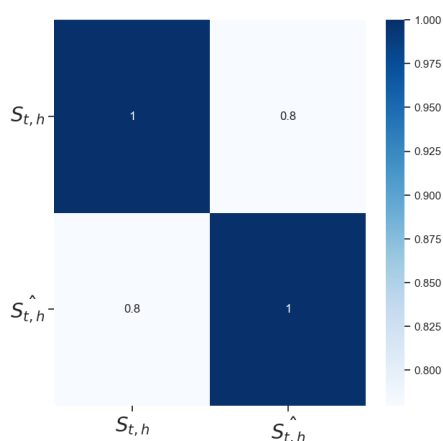


Figura 3.7: Correlazione di Pearson volume e volume preliminare di sbilanciamento macrozonale.

<sup>2</sup>**Correlazione Pearson:** Date due variabili statistiche  $X$  e  $Y$ , l'indice di correlazione di Pearson è definito come la loro covarianza divisa per il prodotto delle deviazioni standard delle due variabili:

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

dove  $\sigma_{XY}$  è la covarianza tra  $X$  e  $Y$  e  $\sigma_X, \sigma_Y$  sono le due deviazioni standard.

Il coefficiente assume sempre valori compresi tra  $-1$  e  $1$ .



## 3.2 Modello per la stima: Approccio con Modelli Generalizzati Additivi

Le sfide legate ai dati precedentemente analizzati per la stima del segno di sbilanciamento macrozonale includono relazioni non lineari e autocorrelazione delle variabili. Queste problematiche, insieme alle complessità della modellazione come l'*overfitting*, la selezione delle variabili e la previsione, rendono le analisi di regressione particolarmente impegnative.

Modelli flessibili, come i modelli generalizzati additivi (GAM), sono in grado di affrontare le problematiche legate ai dati. I GAM permettono di catturare relazioni non lineari tra le variabili, migliorando l'accuratezza delle previsioni. Inoltre, le tecniche di apprendimento automatico, come il *boosting* del gradiente, sono utili per risolvere le sfide della modellazione, inclusi l'*overfitting* e la selezione delle variabili.

Il pacchetto **mboost** in R [7], che implementa metodi di *boosting* per la stima di modelli di regressione e classificazione, è particolarmente utile per questo scopo. Il pacchetto permette di combinare la flessibilità dei GAM con la potenza del *boosting* del gradiente, offrendo un approccio robusto per la modellazione dei segni di sbilanciamento macrozonale.

Il pacchetto **mboost** consente di:

1. Gestire relazioni non lineari: utilizzando funzioni base non lineari, **mboost** può modellare relazioni complesse tra le variabili.
2. Prevenire l'*overfitting*: il *boosting* iterativo aiuta a ridurre l'*overfitting*, migliorando la generalizzazione del modello.
3. Selezionare in modo automatico le variabili: il processo di *boosting* include una forma di selezione delle variabili, migliorando l'efficienza del modello.

Pertanto, l'utilizzo di **mboost** per costruire modelli di previsione basati sui dati di sbilanciamento macrozonale può migliorare significativamente la qualità delle previsioni, affrontando al contempo le sfide tipiche dei dati e della modellazione.

### 3.2.1 Modello Generalizzato Additivo

Dato che i segni di sbilanciamento macrozonale non sono indipendenti e data la dipendenza da variabili esogene, il passo successivo consiste nel modellare queste dipendenze per scopi predittivi. A questo scopo, si scrive la probabilità di un segno di sbilanciamento macrozonale positivo  $\pi_{t,h}$  come funzione di un vettore di covariate  $\mathbf{Z}_{t,h}$ :

$$\mathbb{E}[Y_{t,h} \mid \mathbf{Z}_{t,h}] = \pi_{t,h} = \pi(\mathbf{Z}_{t,h})$$

dove  $\mathbf{Z}_{t,h} = (Z_{1,t,h}, \dots, Z_{m+k,t,h})$  include  $m$  variabili lagged  $Y_{t-i,h}$  per  $i = 1, \dots, m$  e  $k$  variabili esogene  $X_j, h$  per  $j = 1, \dots, k$ .

Un modello additivo nelle covariate  $\mathbf{Z}$  di dimensione  $p$  [6] ha, in generale, la forma:

$$f(\pi) = \beta_0 + f_1(Z_1) + \dots + f_p(Z_p) = \beta_0 + \sum_{j=1}^p f_j(Z_j) \quad (3.6)$$

dove  $\pi = \mathbb{E}[Y \mid \mathbf{Z}]$  è l'aspettativa condizionale della risposta,  $f$  è la funzione di collegamento e  $f_1, \dots, f_p$  sono funzioni arbitrarie delle covariate. Queste funzioni possono includere forme semplici e lineari così come funzioni smooth e non lineari. Le forme delle funzioni di previsione marginali dipendono dalle scale dei predittori. Per i predittori continui, di solito si testano funzioni smooth (cioè continue e differenziabili)  $f_j(Z_j)$  che possono essere lineari o non lineari in  $Z_j$ . Per i predittori categorici, invece, viene utilizzata la codifica *dummy*.

Per garantire che le funzioni marginali siano identificabili, la somma  $f_1(Z_1) + \dots + f_p(Z_p)$  è vincolata ad avere media zero.

Nel lato sinistro,  $f(\cdot)$  rappresenta una caratteristica della distribuzione condizionale specificata. Nel caso in analisi, per garantire che  $\pi$  appartenga all'intervallo  $[0,1]$ , essendo una probabilità, si assume che la relazione tra  $\pi$  e le covariate  $\mathbf{Z}$  possa essere descritta dalla funzione logistica:

$$\pi(\mathbf{Z}) = \frac{e^{\sum_j f_j(Z_j)}}{1 + e^{\sum_j f_j(Z_j)}}$$

Trattandosi di una variabile risposta binaria, la sua densità appartiene alla famiglia esponenziale e, essendo un caso particolare della distribuzione binomiale, la funzione canonica di link è la funzione logit. Questo significa che la relazione tra le variabili predittive e la probabilità di successo nel modello logistico può essere rappresentata utilizzando la seguente funzione:

$$f(\pi(\mathbf{Z})) = \log \left( \frac{\pi(\mathbf{Z})}{1 - \pi(\mathbf{Z})} \right)$$

Questo modello permette di catturare relazioni non lineari tra le covariate e la probabilità del segno di sbilanciamento macrozonale positivo, sfruttando al meglio le informazioni storiche e le variabili esogene disponibili.

### 3.2.2 *Boosting* del gradiente a componenti: Ottimizzazione dei Modelli Additivi Generalizzati

Nell'analisi condotta è stato impiegato il *boosting* del gradiente per migliorare la capacità predittiva e affrontare due problemi comuni nell'implementazione dei Modelli Additivi Generalizzati (GAM): la selezione delle variabili e la scelta del modello. Questi problemi includono la selezione del sottoinsieme più informativo di covariate e la scelta delle migliori rappresentazioni delle variabili predittive continue, come tra effetto lineare o non lineare.

Il *boosting* del gradiente è un metodo di apprendimento automatico che ottimizza l'accuratezza delle previsioni e ottiene stime dei modelli statistici tramite tecniche di discesa del gradiente. Una caratteristica distintiva di questo metodo è la selezione delle variabili integrata nel processo di adattamento, senza l'uso di tecniche euristiche come la selezione *stepwise*. Inoltre, gli algoritmi di *boosting* del gradiente producono regole di previsione che mantengono l'interpretabilità, a differenza di metodi come le foreste casuali che generano previsioni "black-box".

Nel contesto del *boosting* del gradiente l'obiettivo è stimare la funzione di previsione ottimale  $f^*$ , definita come:

$$f^* := \arg \min_f \mathbb{E}_{Y, \mathbf{Z}}[\rho(Y, f(\mathbf{Z}))]$$

dove  $\rho$  è la funzione di perdita, che in questo caso è la funzione di log-verosimiglianza negativa del GAM, su tutto l'insieme delle possibili funzioni di previsione  $f$  che prendono i predittori  $Z_1, \dots, Z_p$  come variabili.

Poiché le distribuzioni esatte di  $\mathbf{Z}$  e  $Y$  sono tipicamente sconosciute, il *boosting* del gradiente minimizza il rischio empirico  $R$  definito come:

$$R := \frac{1}{n} \sum_{i=1}^n \rho(Y^i, f(\mathbf{Z}^i))$$

dove  $\mathbf{Z}^1, \dots, \mathbf{Z}^p$  e  $Y^1, \dots, Y^n$  sono punti campione indipendenti e identicamente distribuiti di  $\mathbf{Z}$  e  $Y$ , rispettivamente.

$R$  è un'approssimazione del rischio teorico  $\mathbb{E}_{Y, \mathbf{Z}}[\rho(Y, f(\mathbf{Z}))]$ , corrispondendo alla media empirica delle contribuzioni al logaritmo negativo della verosimiglianza  $\rho(Y^i, f(\mathbf{Z}^i))$ ,  $i = 1, \dots, n$ , del campione di dati osservati  $(\mathbf{Z}^1, Y^1) \dots, (\mathbf{Z}^n, Y^n)$  sostituiti in  $\rho$ .

Il *boosting* del gradiente utilizza la discesa del gradiente per stimare l'insieme di valori  $\hat{f}_1 := \hat{f}(Z_1), \dots, \hat{f}_p := \hat{f}(Z_p)$  che minimizzano  $R$ . La procedura inizia con stime grezze arbitrarie (ad esempio, tutte zero) dei valori previsti  $\hat{f}_1, \dots, \hat{f}_p$ , aggiornando poi iterativamente le stime delle funzioni aggiungendo ad ogni passo una piccola frazione della derivata negativa di  $R$  rispetto a  $\hat{f}_1, \dots, \hat{f}_p$ , valutata alle stime correnti, detta anche gradiente negativo. Ad ogni iterazione, si calcola dunque il gradiente negativo  $U$  della funzione di perdita empirica  $R$  rispetto alle previsioni correnti. Questo gradiente rappresenta la direzione di discesa più ripida.

Il gradiente viene stimato ( $\hat{U}$ ) tramite *base-learners*, che sono piccoli modelli di regressione addestrati utilizzando un sottoinsieme delle variabili predittive. Ogni *base-learner* è un modello semplice, come una regressione lineare o una spline, e ciascuno utilizza una o più variabili predittive.

Per ciascuna delle  $p$  variabili predittive continue, sono stati specificati due *base-learner*: uno lineare (regressione lineare) e uno non lineare liscio (P-spline), secondo [Kneib et al.](#). Per le variabili predittive categoriche, è stato utilizzato un *base-learner* lineare con variabili dummy rappresentanti il predittore.

Il *boosting* del gradiente a componenti utilizza esclusivamente il miglior *base-learner*, cioè quello che si adatta più efficacemente al gradiente in ogni iterazione. Una piccola frazione  $\nu$  di  $\hat{U}$  viene aggiunta ai valori correnti  $\hat{f}_1, \dots, \hat{f}_p$ , e successivamente  $\hat{U}$  viene nuovamente stimato nelle iterazioni successive utilizzando il miglior *base-learner*. Comunque, per il fattore di lunghezza del passo si adotta il valore  $\nu = 0.1$ .

Nelle tabelle [3.1](#), [3.2](#) sono elencati alcuni esempi di chiamate ai *base-learner* nel pacchetto **mboost** implementato in R.

Chiamata	Tipo di Effetto
bols(x)	Effetto lineare: $x^\top \beta$ dove $x^\top = (1, x)$
bols(x, intercept = FALSE)	Effetto lineare senza intercetta: $\beta \cdot x$
bols(z)	OLS fit del fattore z (ovvero, effetto lineare dopo la codifica dummy)
bols(z, df = 1)	OLS fit con penalità Ridge con un grado di libertà e fattore z
bols(x1, x2, x3)	Unico <i>base-learner</i> per tre variabili: $x^\top \beta$ con $x^\top = (1, x1, x2, x3)$
bols(x, by = z)	Interazione di più variabili: $x^\top \beta \cdot z$ . Se z è un fattore, viene stimato un effetto separato per ciascun livello del fattore

Tabella 3.1: Alcuni esempi di effetti ottenibili tramite *base-learner* bols() nel pacchetto **mboost**.

Chiamata	Tipo di Effetto
bbs(x, by = z)	Coefficiente variabile: $f(x) \cdot z = \beta(x)z$ (con variabile continua z). Se z è un fattore, viene stimato un effetto liscio separato per ciascun livello del fattore
bbs(x, knots = 10)	Effetto <i>smooth</i> con 10 nodi interni
bbs(x, boundary.knots = c(0, 2 * pi), cyclic = TRUE)	Funzione periodica con periodo $2\pi$
bbs(x, center = TRUE, df = 1)	Decomposizione P-spline

Tabella 3.2: Alcuni esempi di effetti ottenibili tramite *base-learner* bbs() nel pacchetto **mboost**.

La selezione del miglior *base-learner* in ogni passo permette la selezione delle variabili e contribuisce a evitare l'*overfitting*. Alcuni *base-learner* potrebbero non essere mai selezionati prima che l'algoritmo si interrompa, consentendo così una selezione automatica delle variabili. L'algoritmo sceglie tra le variabili e tra i *base-learner* lineari e non lineari, garantendo così la scelta del modello ottimale.

### 3.3 Valutazione e Risultati del Modello

L'analisi condotta finora permette di stimare la funzione di previsione  $\hat{f}$  per  $\pi_{t,h}$  (vengono stimati  $\hat{f}^1 \dots \hat{f}^p$ ), la probabilità di segno di squilibrio macrozonale positivo. La stima  $\hat{f}$  viene quindi utilizzata per stimare la probabilità  $\hat{\pi}_{t,h}$  invertendo la funzione stessa:

$$\hat{\pi}_{t,h} = \hat{f}^{-1}(\mathbf{Z}_{t,h})$$

dove  $\mathbf{Z}_{t,h}$  è il vettore delle covariate al tempo  $t$  e periodo di carico  $h$ .

Una volta ottenuta la stima  $\hat{\pi}_{t,h}$  è necessario tradurre questa probabilità in una previsione del segno di squilibrio  $Y_{t,h}$ . La regola più semplice è quella di prevedere un segno positivo se la probabilità è maggiore di 0,5:

$$Y_{t,h} = \begin{cases} 1 & \text{se } \hat{\pi}_{t,h} \geq 0.5 \\ 0 & \text{se } \hat{\pi}_{t,h} < 0.5 \end{cases} \quad (3.7)$$

Nel contesto analizzato, l'obiettivo è una previsione a due giorni in avanti, poiché le decisioni operative devono essere prese entro le 9 del mattino del giorno per il giorno successivo. Pertanto, le informazioni disponibili fino al giorno  $t - 2$  vengono utilizzate per prevedere il segno di squilibrio al giorno  $t$ . La stima della probabilità di uno squilibrio positivo a due giorni in avanti è data da:

$$\hat{\pi}_{t,h} = \hat{f}^{-1}(\mathbf{Z}_{t-2,h})$$

dove  $\mathbf{Z}_{t-2,h}$  include solo le informazioni disponibili fino al giorno  $t - 2$  per il periodo di carico  $h$ .

La stima della probabilità di uno squilibrio positivo mediante il *boosting* del gradiente e la sua conversione in previsioni del segno di squilibrio sono strumenti potenti per supportare le decisioni operative. La previsione del giorno  $t$ , basata su informazioni disponibili fino al giorno  $t - 2$ , garantisce che le decisioni possano essere prese in tempo utile, mantenendo l'accuratezza e l'affidabilità delle previsioni.

Valutare le prestazioni di un modello GAM logistico è un passo cruciale nel comprendere il suo comportamento nella classificazione. Al fine di ottenere una visione completa delle capacità predittive del modello, è essenziale considerare una serie di metriche di valutazione comuni per i modelli di classificazione.

1. **Matrice di Confusione.** Questo strumento fornisce un'analisi dettagliata delle prestazioni del modello, mostrando il numero di predizioni corrette e errate fatte per ciascuna classe target. Oltre a fornire una panoramica generale delle prestazioni, una matrice di confusione permette di valutare la capacità del modello di distinguere tra le diverse categorie di classificazione.
2. **Accuratezza.** L'accuratezza del modello, espressa come la percentuale di predizioni corrette rispetto al totale delle predizioni effettuate, offre una visione complessiva

delle sue prestazioni. È importante notare che l'accuratezza può essere influenzata dalla distribuzione delle classi nel dataset di test e quindi va interpretata con cautela, specialmente quando le classi non sono bilanciate.

3. **Precisione e Sensitività.** Queste metriche diventano fondamentali quando le classi nel dataset di test sono sbilanciate. La precisione misura la percentuale di istanze predette come positive che sono effettivamente positive, mentre la sensitività misura la percentuale di istanze positive che sono state correttamente identificate dal modello. Queste due metriche forniscono una valutazione dettagliata delle prestazioni del modello, consentendo di valutare la sua capacità di individuare correttamente gli esempi positivi.
4. **F1-score.** F1-score è una media armonica di precisione e sensitività ed è particolarmente utile quando si desidera ottenere una misura complessiva delle prestazioni del modello. Essendo una media armonica delle due metriche, il F1-score bilancia l'importanza di precisione e sensitività, fornendo una valutazione più equilibrata delle capacità del modello.
5. **Receiver Operating Characteristic (ROC) Curve e Area Under the Curve (AUC).** La curva ROC e l'Area Under the Curve (AUC) sono utilizzate per valutare la capacità discriminativa del modello. La curva ROC visualizza il trade-off tra il tasso di vera positività (true positive rate) e il tasso di falsa positività (false positive rate) al variare della soglia di classificazione. L'AUC fornisce una misura complessiva della capacità del modello di distinguere tra le classi, indipendentemente dalla scelta della soglia di classificazione.

Questi strumenti forniscono una valutazione dettagliata e completa delle prestazioni del modello GAM logistico, consentendo di identificare eventuali aree in cui può essere migliorato. Oltre a queste metriche di valutazione standard, è importante considerare anche il contesto specifico dell'applicazione e i requisiti del problema di classificazione in questione.

### Caso di esempio:

Considerando la macrozona Nord e la serie storica dello sbilanciamento macrozonale a partire dal 1° Febbraio 2022, si desidera prevedere la probabilità che il segno dello sbilanciamento macrozonale sia positivo per il mese di Agosto 2023. Si esegue un test retrospettivo utilizzando dati coerenti con quelli disponibili per formulare la previsione. Questo approccio implica l'utilizzo di modelli e tecniche di previsione che si basano esclusivamente sui dati disponibili fino al relativo giorno  $t - 2$ , al fine di stimare la probabilità di uno sbilanciamento positivo per ciascun giorno  $t$  e ora  $h$  del mese di agosto 2023.

Per valutare le prestazioni di previsione del modello 3.6, utilizzando la regola 3.7, sono state effettuate previsioni "out of sample" di due giorni in anticipo per il periodo di Agosto 2023. Ciò ha comportato un totale di 744 ( $31 \times 24$ ) previsioni di segno. I risultati ottenuti sono stati valutati in termini delle metriche precedentemente descritte, sia globalmente che per ogni ora del giorno. Nella Tabella 3.3 sono presenti le relative metriche di valutazione, mentre nella Figura 3.9 sono riportate le matrici di confusione e le curve ROC.

	Accuratezza	Precisione	Sensitività	F1-score
Valutazione Globale	68%	56%	69%	62%
Valutazione fascia oraria 00:00	74%	57%	80%	66%
Valutazione fascia oraria 06:00	58%	56%	60%	58%
Valutazione fascia oraria 12:00	61%	47%	81%	60%
Valutazione fascia oraria 18:00	93%	88%	88%	88%

Tabella 3.3: Metriche di valutazione sia in termini globali che per fascia oraria.

Per avere un'ulteriore valutazione della probabilità predetta dal modello, è stato creato un grafico (Fig. 3.8) che confronta i valori di probabilità predetti per l'intero periodo con i valori di sbilanciamento macrozonale. Questi ultimi sono stati opportunamente riportati nell'intervallo  $[0,1]$  per agevolare il confronto. L'osservazione di un andamento molto omogeneo tra le due serie fornisce ulteriore supporto alle previsioni effettuate dal modello.

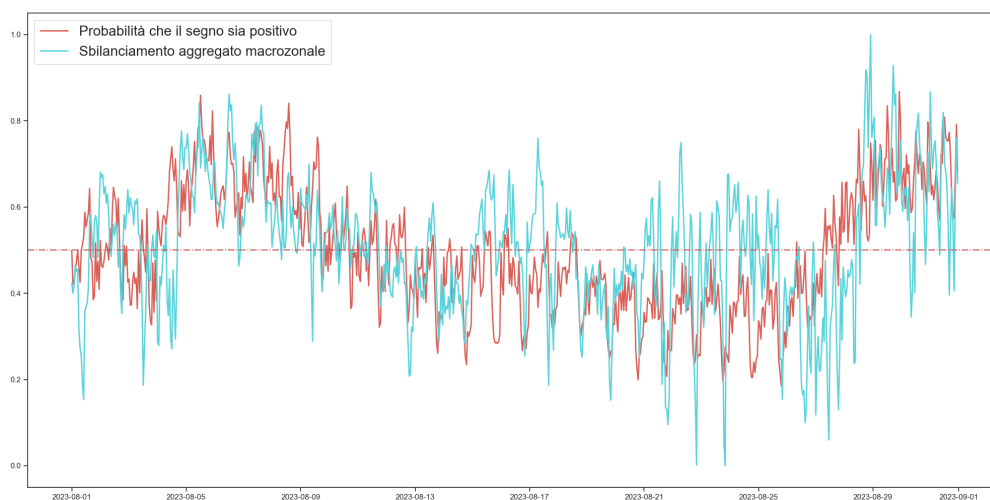
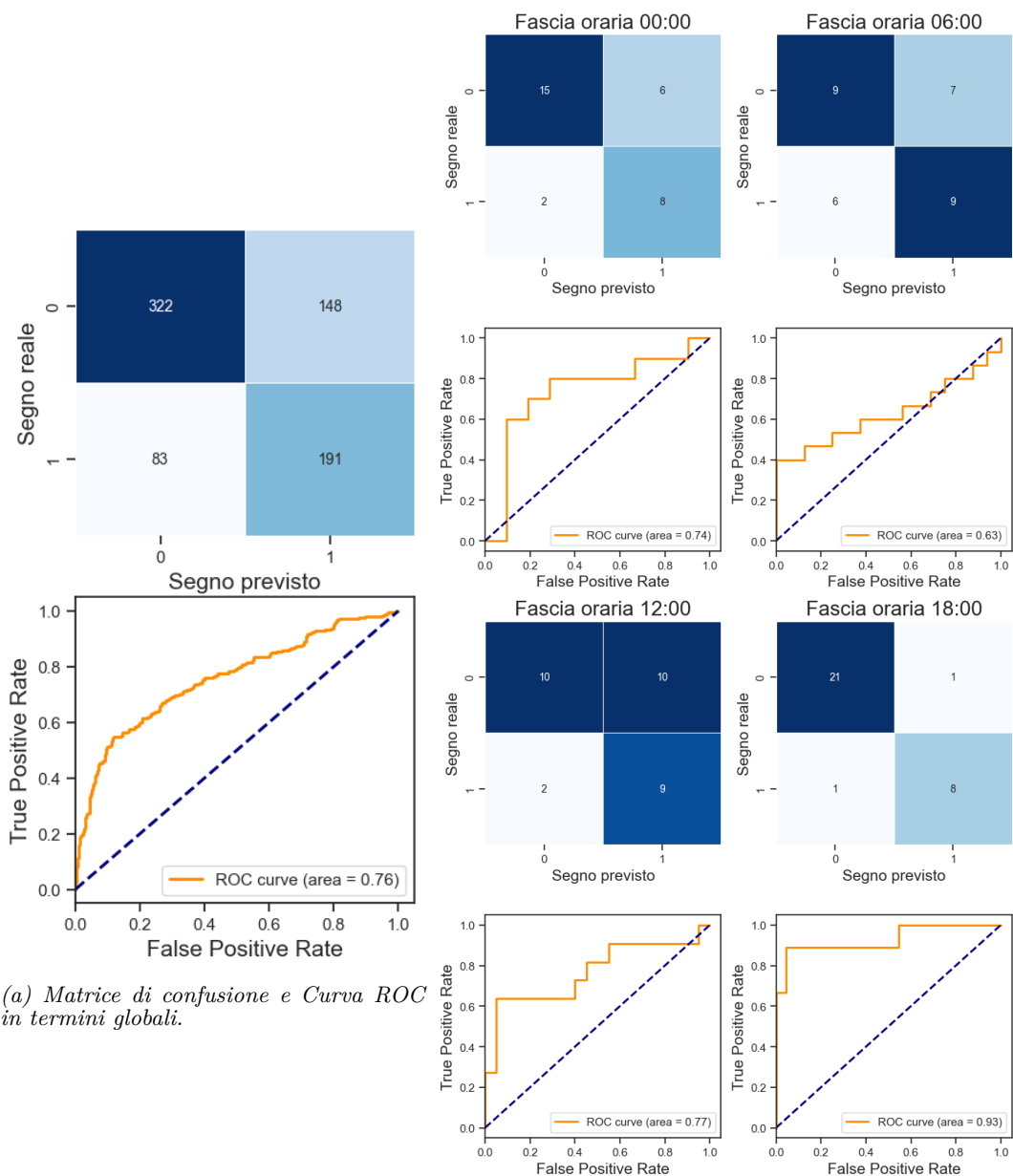


Figura 3.8: Probabilità di segno positivo a confronto con il volume di sbilanciamento macrozonale



(a) Matrice di confusione e Curva ROC in termini globali.

(b) Matrice di confusione e Curva ROC in termini di fascia oraria.

Figura 3.9



A seguito di un confronto dei risultati con quelli ottenuti da un modello randomico e da un modello semplice (media storica), è emersa l'efficacia del modello di previsione. I valori delle metriche di valutazione per ciascun modello sono riportati nella Tabella 3.4.

	Accuratezza	Precisione	Sensitività	F1-score
Predizioni con il modello GAM	68%	56%	69%	62%
Predizioni con modello randomico	50%	37%	48%	41%
Predizioni utilizzando media storica	46%	38%	79%	52%

Tabella 3.4: Metriche di valutazione sia in termini globali che per fascia oraria.

### 3.3.1 Considerazioni sull'Importanza del Livello di Probabilità Previsto

Nel contesto delle previsioni ottenute dal modello GAM logistico è fondamentale considerare il livello di probabilità previsto insieme alla semplice regola di classificazione basata sulla soglia del 50%. Questo perché tale regola non tiene conto del grado di incertezza nelle previsioni, ignorando informazioni preziose.

I valori di  $\hat{\pi}_{t,h}$  che si avvicinano a 0,5 potrebbero non essere particolarmente informativi riguardo a  $\hat{Y}_{t,h}$ , poiché rappresentano casi in cui la previsione è incerta. Una regola più efficiente dovrebbe quindi considerare anche il livello di probabilità previsto, distinguendo tra previsioni più sicure e meno sicure.

Ad esempio, se  $\hat{\pi}_{t,h}$  è molto vicino a 0, il modello predice con una certa sicurezza uno sbilanciamento macrozonale negativo, mentre se  $\hat{\pi}_{t,h}$  è prossimo a 1, il modello predice con una maggiore confidenza uno sbilanciamento macrozonale positivo. In entrambi i casi, le previsioni sono più affidabili rispetto ai casi in cui  $\hat{\pi}_{t,h}$  si avvicina a 0,5.

Considerare il livello di probabilità previsto consente quindi di avere una visione più dettagliata e informativa sulle previsioni del modello. Questo può essere utile soprattutto in contesti in cui la precisione delle previsioni è cruciale, consentendo di prendere decisioni più consapevoli e informate basate sulle probabilità previste.

Si osserva infatti dal grafico 3.10 che all'aumentare della confidenza della probabilità prevista, si registra un aumento del numero di segni correttamente predetti. Questo comporta un miglioramento dell'accuratezza complessiva del modello, in quanto si verifica una migliore discriminazione tra le previsioni.

Nel grafico, è evidente che quando  $\hat{\pi}_{t,h}$  si avvicina a 0, le previsioni dello sbilanciamento macrozonale negativo sono più affidabili, mentre quando  $\hat{\pi}_{t,h}$  si avvicina a 1, le previsioni dello sbilanciamento macrozonale positivo diventano più attendibili. Questo aumento della precisione delle previsioni alle estremità della scala delle probabilità conferma l'importanza di considerare il livello di confidenza previsto per ottenere risultati più accurati.

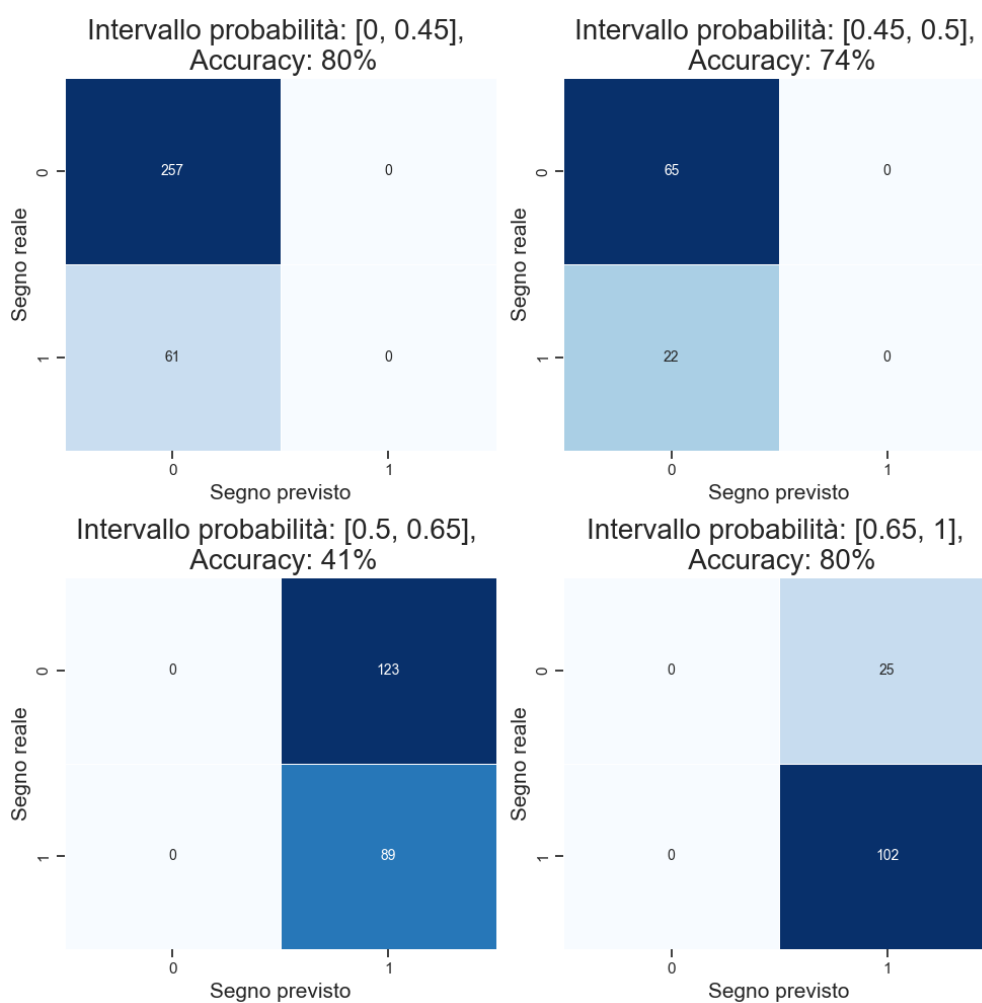


Figura 3.10: Matrice di confusione in termini globali, al variare della probabilità predetta.

## Capitolo 4

# Ottimizzazione del Sistema di Previsione Energetica: Strategie per la Riduzione delle Perdite

Per ottimizzare il sistema di previsione e ridurre l'errore, si utilizza l'informazione dai capitoli precedenti per fornire un "consiglio" di aggiustamento. Questo "consiglio" mira a correggere l'errore di previsione dell'utente, tenendo conto non solo degli intervalli di confidenza sullo sbilanciamento, ma anche della previsione del segno macrozonale. L'obiettivo è fornire raccomandazioni per migliorare la precisione delle previsioni. È fondamentale che gli operatori del sistema energetico dispongano delle informazioni necessarie per prendere decisioni informate e tempestive, mantenendo l'equilibrio tra domanda e offerta di energia e assicurando la stabilità del sistema elettrico.

Considerando l'impatto diretto del sistema di regolazione degli sbilanciamenti sulla gestione economica delle unità di produzione e consumo, le decisioni operative devono tenere conto sia della quantità di energia prodotta o consumata, sia del contesto macrozonale in cui avvengono tali operazioni. Tuttavia, è importante ricordare che eventuali situazioni di guadagno non dovrebbero incentivare uno squilibrio intenzionale. Aumentare intenzionalmente lo sbilanciamento per ottenere profitto sfruttando la differenza di prezzo tra il Mercato del Giorno Prima e il Mercato dei Servizi di Dispacciamento è proibito e va contro gli obiettivi di sostenibilità del sistema.

L'analisi si concentra sull'obiettivo fondamentale di ridurre le perdite derivanti da una disparità di prezzo penalizzante, in particolare quando lo sbilanciamento dell'unità coincide con quello macrozonale della zona in cui opera. Per affrontare efficacemente questa sfida, viene adottato un modello decisionale specifico che si focalizza sulla correzione degli

errori di sbilanciamento quando sono in linea con la zona di riferimento. Questo approccio strategico mira a migliorare l'efficacia complessiva del sistema elettrico, garantendo al contempo un comportamento etico e responsabile da parte delle unità partecipanti al mercato. L'obiettivo è ridurre al minimo le perdite economiche associate agli squilibri, garantendo un funzionamento più stabile e sostenibile del sistema nel suo complesso. Attraverso questo modello decisionale, si mira a ottimizzare l'utilizzo delle risorse energetiche disponibili e a massimizzare i benefici per tutte le parti coinvolte nel mercato energetico.

## 4.1 Problemi di Ottimizzazione: Caratteristiche e Applicazioni

Un problema di ottimizzazione rappresenta una sfida nella ricerca della soluzione migliore tra molte alternative possibili, tenendo conto di criteri specifici o vincoli definiti. Questi problemi sono diffusi in diversi settori, tra cui ingegneria, economia, gestione aziendale, logistica, produzione, finanza e molti altri.

Le caratteristiche chiave di un problema di ottimizzazione includono:

- **Variabili decisionali.** Le variabili decisionali costituiscono gli elementi che possono essere manipolati per influenzare l'esito di un problema di ottimizzazione. Esse rappresentano le decisioni da prendere al fine di conseguire l'obiettivo prefissato. In sintesi, sono i fattori che l'ottimizzatore può regolare o impostare in modo da ottenere il miglior risultato possibile. Modificando queste variabili, si cerca di massimizzare o minimizzare la funzione obiettivo, rispettando contemporaneamente tutti i vincoli e le restrizioni del problema.
- **Dominio delle variabili.** Il dominio delle variabili definisce l'insieme di possibili valori che le variabili decisionali possono assumere durante il processo di ottimizzazione. Questo spazio di ricerca limita le combinazioni di valori per trovare la soluzione ottima. Una corretta definizione del dominio assicura che il processo di ottimizzazione produca risultati coerenti con le condizioni e i vincoli del problema, contribuendo così a individuare la soluzione ottimale in modo efficiente.
- **Funzione obiettivo.** La funzione obiettivo rappresenta la misura del successo desiderato e può essere definita come la quantità da massimizzare o minimizzare. Questa funzione dipende direttamente dalle variabili decisionali, in quanto il suo valore è determinato dalle scelte effettuate riguardo a quest'ultime.
- **Vincoli.** I vincoli rappresentano le restrizioni o condizioni che devono essere soddisfatte durante il processo di ottimizzazione. Essi definiscono i limiti entro cui le variabili decisionali possono muoversi e le relazioni da rispettare tra di esse. I vincoli sono essenziali perché definiscono il contesto operativo e garantiscono che la soluzione ottimale sia realistica e praticabile. Contribuiscono a guidare il processo di ottimizzazione verso soluzioni che rispettano le specifiche esigenze del problema, assicurando che il risultato finale sia valido e applicabile nella pratica.

### Caso di Studio:

Consideriamo il caso specifico di ridurre le perdite derivanti da una disparità di prezzo penalizzante, minimizzando l'errore di sbilanciamento in linea con la zona di riferimento.

Dato il giorno  $t$  e fissata una fascia oraria  $h$ , sia il seguente vettore di **variabili decisionali**:

$$\Delta_{t,h,i} = (\Delta_{t,h,i}^+, \Delta_{t,h,i}^-) \quad (4.1)$$

dove  $\Delta_{t,h,i}^+$ ,  $\Delta_{t,h,i}^-$  rappresentano, rispettivamente, la correzione da apportare al valore di previsione a livello orario, espressa come percentuale in valore assoluto, per aumentare o diminuire il valore previsto.

La definizione del **dominio delle variabili decisionali**, vincolato da  $\Delta_{max,t,h,i}$ , è cruciale per garantire la stabilità del processo decisionale. Limitando le correzioni entro un intervallo massimo consentito, si evitano fluttuazioni eccessive che potrebbero compromettere l'efficacia della previsione. Questo limite massimo assicura che le correzioni rimangano entro un range praticabile e non distorcano eccessivamente il valore previsto. Questo approccio mira a garantire che le correzioni apportate siano pragmatiche e realistiche, consentendo un adeguato controllo sul processo di ottimizzazione e una migliore gestione degli squilibri. In questo elaborato  $\Delta_{max,t,h,i} = 5\%$ . Il dominio delle variabili decisionali risulta quindi essere in  $\mathbb{R}^2$ , in quanto variabili continue, con i seguenti vincoli:

$$0 \leq \Delta_{t,h,i}^+, \Delta_{t,h,i}^- \leq \Delta_{max,t,h,i} \quad (4.2)$$

### Strategie di Correzione

Nel contesto dell'ottimizzazione dell'errore di sbilanciamento, la direzione e l'entità delle correzioni dipendono dal segno macrozonale di riferimento. Questo segno indica se la macrozona è *lunga* (con un eccesso di energia disponibile) o *corta* (con un deficit di energia disponibile).

Dato il giorno  $t$  la fascia oraria  $h$ , siano  $Err_{t,h,i}$  lo sbilanciamento dell'unità in termini di errore percentuale e  $Segno_{t,h}$  il segno della macrozona di riferimento.

- Macrozona Lunga ( $Segno_{t,h} = 1$ ): in questa situazione, la correzione dovrebbe essere al ribasso per ridurre lo sbilanciamento e non contribuire all'eccesso di energia nella macrozona.
- Macrozona Corta ( $Segno_{t,h} = 0$ ): in questa circostanza, la correzione dovrebbe essere al rialzo per ridurre il deficit energetico della macrozona.

Per rappresentare correttamente i casi descritti e garantire che le correzioni siano effettuate nella direzione appropriata in base al segno macrozonale, vengono introdotti **vincoli operativi**:

$$\begin{cases} \Delta_{t,h,i}^+ \cdot Segno_{t,h} = 0 \\ \Delta_{t,h,i}^- \cdot (1 - Segno_{t,h}) = 0 \end{cases} \quad (4.3)$$

Questi vincoli assicurano che:

- Macrozona lunga ( $Segno_{t,h} = 1$ ): la correzione al rialzo  $\Delta_{t,h,i}^+$  deve essere zero. Questo vincolo  $\Delta_{t,h,i}^+ \cdot Segno_{t,h} = 0$  impedisce di aggiungere ulteriori correzioni positive in eventuali situazioni di eccedenza.
- Macrozona corta ( $Segno_{t,h} = 0$ ): la correzione al ribasso  $\Delta_{t,h,i}^-$  deve essere zero. Questo vincolo  $\Delta_{t,h,i}^- \cdot (1 - Segno_{t,h}) = 0$  impedisce di ridurre ulteriormente la produzione in una situazione di deficit energetico.

Per allineare gli obiettivi di sostenibilità del sistema con la strategia di correzione, si considerano ulteriori vincoli operativi.

Nel caso di macrozona lunga ( $Segno_{t,h} = 1$ ), se lo sbilanciamento dell'unità  $Err_{t,h,i}$  è negativo non viene applicata alcuna correzione. Se lo sbilanciamento dell'unità  $Err_{t,h,i}$  è positivo la correzione  $\Delta_{t,h,i}^-$ , allora, deve ridurre lo sbilanciamento al più fino a zero, evitando di cambiare il segno dello sbilanciamento. Nel caso di macrozona corta ( $Segno_{t,h} = 0$ ), se lo sbilanciamento dell'unità  $Err_{t,h,i}$  è positivo non viene applicata alcuna correzione. Se lo sbilanciamento dell'unità  $Err_{t,h,i}$  è negativo, la correzione  $\Delta_{t,h,i}^+$  deve ridurre lo sbilanciamento al più fino a zero, evitando di cambiare il segno dello sbilanciamento.

- $Segno_{t,h} = 1 \wedge Err_{t,h,i} \leq 0 \implies \Delta_{t,h,i}^- = 0$
- $Segno_{t,h} = 1 \wedge Err_{t,h,i} \geq 0 \implies \Delta_{t,h,i}^- \leq Err_{t,h,i}$
- $Segno_{t,h} = 0 \wedge Err_{t,h,i} \geq 0 \implies \Delta_{t,h,i}^+ = 0$
- $Segno_{t,h} = 0 \wedge Err_{t,h,i} \leq 0 \implies \Delta_{t,h,i}^+ \leq |Err_{t,h,i}|$

Per rappresentare queste condizioni, si considerano i seguenti **vincoli operativi**:

$$\begin{cases} (Err_{t,h,i} - \Delta_{t,h,i}^-) \cdot \Delta_{t,h,i}^- \geq 0 \\ -(Err_{t,h,i} + \Delta_{t,h,i}^+) \cdot \Delta_{t,h,i}^+ \geq 0 \end{cases} \quad (4.4)$$

Questi vincoli assicurano che, per la macrozona lunga, non solo la correzione al rialzo sia zero, ma anche che la correzione al ribasso non superi l'errore positivo; per la macrozona corta, assicurano che la correzione al rialzo non superi l'errore negativo. Questo approccio non solo ottimizza il processo decisionale, ma promuove anche un comportamento etico e responsabile da parte delle unità partecipanti al mercato. Nella pratica operativa, le correzioni vengono applicate in modo tale da rispettare rigorosamente le condizioni specifiche del sistema elettrico, con l'obiettivo di minimizzare l'errore di sbilanciamento solo quando è in linea con il segno macrozonale.

Nel caso specifico preso in esame, dato lo scopo principale, la **funzione obiettivo** è definita come segue:

$$\begin{aligned} f(\Delta_{t,h,i}, Segno_{t,h}, Err_{t,h,i}) &= (Err_{t,h,i} - \Delta_{t,h,i}^-) \cdot Segno_{t,h} \\ &\quad - (Err_{t,h,i} + \Delta_{t,h,i}^+) \cdot (1 - Segno_{t,h}) \end{aligned} \quad (4.5)$$

Questa riflette la volontà di correggere gli errori di previsione solo nella direzione opposta al segno macrozonale.

Integrando questa funzione obiettivo con i vincoli operativi precedentemente delineati, si sviluppa un modello decisionale robusto che ottimizza le correzioni di sbilanciamento. Questo modello non solo minimizza le perdite economiche, ma migliora anche la coerenza e l'efficienza dell'intero sistema elettrico nazionale. Grazie a questo approccio, le entità partecipanti al mercato sono in grado di prendere decisioni più informate e responsabili, contribuendo così a un sistema energetico più sostenibile e resiliente.

Il **modello decisionale** ottenuto è il seguente:

$$\begin{aligned}
 & \min_{\Delta_{t,h,i}} f(\Delta_{t,h,i}, Segno_{t,h}, Err_{t,h,i}) \\
 & \text{s.t. } \Delta_{t,h,i}^-(1 - Segno_{t,h}) = 0, \\
 & \quad \Delta_{t,h,i}^+ \cdot Segno_{t,h} = 0, \\
 & (Err_{t,h,i} - \Delta_{t,h,i}^-)\Delta_{t,h,i}^- \geq 0, \\
 & \quad - (Err_{t,h,i} + \Delta_{t,h,i}^+)\Delta_{t,h,i}^+ \geq 0, \\
 & 0 \leq \Delta_{t,h,i}^-, \Delta_{t,h,i}^+ \leq \Delta_{max,t,h,i}.
 \end{aligned} \tag{4.6}$$

Considerando che la funzione obiettivo è additiva (è lineare nelle variabili) e i vincoli sono separati nelle variabili decisionali, è vantaggioso suddividere il problema di ottimizzazione in due parti distinte, una per  $\Delta_{t,h,i}^+$  e una per  $\Delta_{t,h,i}^-$ . Questa separazione semplifica il processo di ottimizzazione e favorisce una ricerca della soluzione più efficiente. Dunque, la funzione obiettivo e i vincoli possono essere suddivisi così:

- Problema di ottimizzazione per  $\Delta^+$

$$\begin{aligned}
 & \min_{\Delta_{t,h,i}^+} - (Err_{t,h,i} + \Delta_{t,h,i}^+) \cdot (1 - Segno_{t,h}) \\
 & \quad \Delta_{t,h,i}^+ \cdot Segno_{t,h} = 0, \\
 & \quad - (Err_{t,h,i} + \Delta_{t,h,i}^+)\Delta_{t,h,i}^+ \geq 0, \\
 & \quad 0 \leq \Delta_{t,h,i}^+ \leq \Delta_{max,t,h,i}.
 \end{aligned} \tag{4.7}$$

- Problema di ottimizzazione per  $\Delta^-$

$$\begin{aligned}
 & \min_{\Delta_{t,h,i}^-} (Err_{t,h,i} - \Delta_{t,h,i}^-) \cdot Segno_{t,h} \\
 & \text{s.t. } \Delta_{t,h,i}^-(1 - Segno_{t,h}) = 0, \\
 & \quad (Err_{t,h,i} - \Delta_{t,h,i}^-)\Delta_{t,h,i}^- \geq 0, \\
 & \quad 0 \leq \Delta_{t,h,i}^- \leq \Delta_{max,t,h,i}.
 \end{aligned} \tag{4.8}$$

### 4.1.1 Modelli di Programmazione Vincolata alla Casualità per Problemi Stocastici: Approcci e Gestione dei Vincoli

Nel contesto di un problema stocastico 4.9, dove le variabili coinvolte, oltre a quelle decisionali, sono soggette a incertezza o rischio, i modelli di Programmazione Vincolata alla Casualità (Chance Constraints Problem) sono utilizzati per gestire tale incertezza nella formulazione dell'ottimizzazione. Questi modelli stabiliscono limitazioni sul grado accettabile di rischio associato alle decisioni prese.

$$\begin{aligned} \min_x \quad & f(x, \boldsymbol{\xi}) \\ \text{s.t.} \quad & g_i(x) \leq 0 \quad i = 1, 2, \dots, p_1 \\ & g_i(x, \boldsymbol{\xi}) \leq 0 \quad i = p_1 + 1, p_1 + 2, \dots, p \end{aligned} \quad (4.9)$$

dove  $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_m)$  è un vettore di variabili stocastiche. Si denotano  $G^d, G^s$  i set di indici dei vincoli. Pertanto  $|G^d| = p_1, |G^s| = p - p_1$ .

La Programmazione Vincolata alla Casualità è un potente strumento per gestire la casualità e l'incertezza all'interno di problemi di ottimizzazione. I tre tipi di funzioni obiettivo proposti da [Charnes and Cooper](#) offrono diverse prospettive per affrontare questo tipo di incertezza. Sia un problema di ottimizzazione stocastico di minimizzazione:

- **Modello-E:** Minimizza il valore atteso della funzione obiettivo, cercando il miglior risultato medio possibile considerando tutte le possibili realizzazioni delle variabili stocastiche coinvolte.

$$\min_x f(x, \boldsymbol{\xi}) \implies \min_x \mathbb{E}_{\boldsymbol{\xi}}[f(x, \boldsymbol{\xi})]$$

- **Modello-V:** Minimizza il quadrato medio generalizzato della funzione obiettivo, dando maggiore peso ai risultati vicini al valore atteso. Si vogliono comunque considerare gli effetti delle deviazioni rispetto a tale valore.

$$\min_x f(x, \boldsymbol{\xi}) \implies \min_x [f(x, \boldsymbol{\xi}) - \mathbb{E}_{\boldsymbol{\xi}}[f(x, \boldsymbol{\xi})]]^2$$

- **Modello-P:** Massimizza la probabilità di soddisfare un certo livello di aspirazione della funzione obiettivo, tenendo conto di un livello di rischio specificato. Questo è utile quando ci sono specifici obiettivi che si desidera raggiungere con una certa affidabilità, anche se ciò potrebbe comportare un aumento del rischio.

$$\min_x f(x, \boldsymbol{\xi}) \implies \max_x \mathbb{P}(f(x, \boldsymbol{\xi}) \left\{ \begin{array}{l} \leq \\ \geq \end{array} \right\} u)$$

dove  $u$  è uno specifico livello di rischio.

L'attenzione è focalizzata sulla capacità del sistema di soddisfare i vincoli (misure di rischio) con una certa affidabilità in un ambiente incerto. Questi vincoli stocastici possono essere rappresentati in tre modi.



1. Soddisfare il valore atteso dei vincoli:

$$\mathbb{E}_{\boldsymbol{\xi}}[g_i(x, \boldsymbol{\xi})] \leq 0, \quad i \in G^s$$

Si indica con  $\bar{G}^s$  l'insieme dei vincoli il cui valore atteso deve essere soddisfatto,

$$\bar{G}^s = \{g_i \mid \mathbb{E}_{\boldsymbol{\xi}}[g_i(x, \boldsymbol{\xi})] \leq 0\}$$

2. **Soddisfare i vincoli in modo probabilistico:** Si definisce un insieme costituito dalla combinazione di  $r$  ( $\leq p_1$ ) vincoli di disuguaglianza stocastica come  $\hat{g}_{i,r}(x, \boldsymbol{\xi})$  ( $|\hat{g}_{i,r}| = r$ ), definito come:

$$\{g_{i_1}, g_{i_2}, \dots, g_{i_r} \mid i_1, i_2, \dots, i_r \in G^s \text{ e } i_1 \neq i_2 \neq \dots \neq i_r\}$$

L'insieme costituito da tutte le combinazioni uniche di  $r$  vincoli di disuguaglianza stocastica è denotato da  $\hat{G}_r^s$ , definito come:  $\{\{\hat{g}_{i,r}\}, \{\hat{g}_{j,r}\}, \dots\}$ .

Il numero di insiemi all'interno dell'insieme  $\hat{G}_r^s$  è uguale a:  $C_r^{p_1} = \frac{p_1!}{(p_1-r)!r!}$

I vincoli probabilistici sono dati da:

$$\mathbb{P}[\hat{g}_{i,r}(x, \boldsymbol{\xi}) \leq 0] \geq \alpha_{i,r}, \quad \hat{g}_{i,r}(x, \boldsymbol{\xi}) \in \hat{G}_r^s \quad (4.10)$$

dove  $\alpha_{i,r}$  è la misura di probabilità minima che i vincoli devono soddisfare. Basandosi sul valore di  $r$ , possiamo interpretare il vincolo 4.10 in due modi diversi:

- (a) Quando il valore  $r = 1$ , il vincolo 4.10 si riduce a vincoli di probabilità individuale nella forma:

$$\mathbb{P}[\hat{g}_{i,1}(x, \boldsymbol{\xi}) \leq 0] \geq \alpha_{i,1}, \quad \hat{g}_{i,1}(x, \boldsymbol{\xi}) \in \hat{G}_1^s$$

che può essere ulteriormente interpretato come:

$$\mathbb{P}[g_i(x, \boldsymbol{\xi}) \leq 0] \geq \alpha_i, \quad i \in G^s \quad (4.11)$$

dove  $\alpha_i$  è la misura di probabilità minima (indice di affidabilità/sicurezza) per ciascun vincolo  $i$ .

- (b) D'altra parte, quando  $r > 1$ , il vincolo 4.10 si riduce a vincoli di probabilità congiunta. Tali vincoli sorgono quando si richiede che un gruppo di vincoli stocastici soddisfi un dato livello di affidabilità.

$$\mathbb{P}[g_i(x, \boldsymbol{\xi}) \leq 0, \quad i \in G^s] \geq \alpha \quad (4.12)$$

È ben noto che  $\mathbb{P}\{\boldsymbol{\xi} = y\} = 0, \forall y \in \mathbb{R}^m$ , quindi i vincoli probabilistici di uguaglianza non hanno alcuna giustificazione matematica. Si indica con  $\Gamma$  l'insieme dei vincoli nel problema CCP;  $\Gamma$  include i vincoli deterministici e i vincoli probabilistici.

### Caso di studio:

Nel caso in analisi, dove le variabili stocastiche sono l'errore percentuale e il segno per il giorno  $t + 1$  e compaiono sia nella funzione obiettivo che nei vincoli, la Programmazione Vincolata alla Casualità può essere particolarmente utile. Si utilizza il modello-E, l'obiettivo diventa minimizzare il valore atteso della funzione obiettivo e si richiede che i vincoli vengano soddisfatti in modo probabilistico, interpretati come 4.12 (vincoli di probabilità congiunti).

Al netto delle precedenti considerazioni il problema di ottimizzazione 4.6, dopo avere riformulato il vincolo di uguaglianza in due vincoli di disuguaglianza in modo da ottenere la medesima restrizione, viene riformulato nel seguente modo, fissato  $\alpha$ :

$$\begin{aligned} \min_{\Delta_{t,h,i}^+} \quad & \mathbb{E}_{(Err_{t,h,i}, Segno_{t,h})} [-(Err_{t,h,i} + \Delta_{t,h,i}^+) \cdot (1 - Segno_{t,h})] \\ & \mathbb{P} \left\{ \begin{array}{l} \Delta_{t,h,i}^+ \cdot Segno_{t,h} \leq 0, \\ -(\Delta_{t,h,i}^+ \cdot Segno_{t,h}) \leq 0, \\ (Err_{t,h,i} + \Delta_{t,h,i}^+) \Delta_{t,h,i}^+ \leq 0, \end{array} \right\} \geq \alpha \\ & 0 \leq \Delta_{t,h,i}^+ \leq \Delta_{max,t,h,i} \end{aligned} \quad (4.13)$$

## 4.2 Struttura Integrata di Ottimizzazione: Algoritmo Genetico e Simulazione Monte Carlo

Per la risoluzione del problema di ottimizzazione si propone una struttura costituita da un Algoritmo Genetico e da una simulazione Monte Carlo. La natura non lineare e non convessa dei problemi di Programmazione Vincolata alla Casualità viene efficacemente trattata utilizzando l'Algoritmo Genetico, che permette di esplorare ampi spazi e di individuare soluzioni ottimali o quasi ottimali. Contestualmente, la natura stocastica dei problemi viene gestita tramite la simulazione Monte Carlo, che consente di modellare e analizzare le incertezze intrinseche nei dati e nelle variabili del problema.

L'Algoritmo Genetico si basa sui principi dell'evoluzione naturale, applicando processi come la selezione, il crossover e la mutazione per generare nuove popolazioni di soluzioni. Questo metodo è particolarmente adatto per problemi complessi dove la funzione obiettivo non è ben definita o presenta numerosi ottimi locali.

D'altra parte, la simulazione Monte Carlo rappresenta un potente strumento per gestire la variabilità e l'incertezza, eseguendo un gran numero di simulazioni per approssimare la distribuzione delle variabili aleatorie. Questa tecnica permette di ottenere stime accurate delle performance delle soluzioni in condizioni di incertezza, fornendo una base solida per la valutazione e l'ottimizzazione.

L'integrazione di questi due metodi nella struttura proposta non solo garantisce una ricerca approfondita nello spazio delle soluzioni, ma assicura anche che quelle trovate siano robuste rispetto alle variazioni e alle incertezze dei dati.

### 4.2.1 Simulazione stocastica per vincoli di probabilità

La gestione dei vincoli mediante metodi Monte Carlo è particolarmente vantaggiosa quando i vincoli non sono regolari o sono complessi da calcolare. In generale, per stimare la probabilità di soddisfacimento dei vincoli, si utilizza un approccio Monte Carlo, generando un gran numero di vettori casuali e calcolando la frequenza con cui le soluzioni soddisfano i vincoli.

In generale, sia il vincolo di probabilità congiunto:

$$G(x) = \mathbb{P}[g_i(x, \boldsymbol{\xi}) \leq 0, \quad i \in G^s] \geq \alpha \quad (4.14)$$

dove  $\boldsymbol{\xi} = (\xi_1, \xi_2, \dots, \xi_m)$  è un vettore stocastico  $m$ -dimensionale e ogni  $\xi_i$  segue una distribuzione specifica. Per ogni valore  $x$ , si utilizza la tecnica Monte Carlo per stimare  $G(x)$ . Utilizzando la tecnica di simulazione Monte Carlo (MC), il valore dei vincoli di probabilità viene calcolato sulla base della frequenza con cui la soluzione attuale soddisfa i vincoli.

Si generano  $N$  vettori casuali indipendenti:

$$\boldsymbol{\xi}^s = (\xi_1^s, \xi_2^s, \dots, \xi_m^s), \quad s = 1, 2, \dots, N$$

dalle rispettive distribuzioni di probabilità, con metodi di generazione ampiamente trattati in letteratura e riassunti da [Rubinstein et al.](#) Nel sottoparagrafo successivo [4.2.1](#) si analizza nel dettaglio i metodi di generazione utilizzati per le variabili stocastiche del caso in esame.

Indicando con  $N'$  il numero di volte in cui:

$$g_i(x, \boldsymbol{\xi}_s) \leq 0, \quad i = 1, 2, \dots, p, \quad s = 1, 2, \dots, N'$$

ovvero il numero di volte che il vettore casuale soddisfa i vincoli, secondo la definizione di base di probabilità,  $G(x)$  può essere stimato come:

$$G(x) = \frac{N'}{N} \quad (6)$$

Questo significa che il vincolo di probabilità [4.14](#) è soddisfatto se e solo se

$$\frac{N'}{N} \geq \alpha$$

Questa stima, pur essendo approssimativa e soggetta a variazioni tra diverse simulazioni, risulta utilizzabile per problemi reali, poiché la determinazione del livello di confidenza  $\alpha$  non richiede una precisione estrema.

Nell'algoritmo implementato, si tiene conto del rischio associato alle variabili casuali utilizzate nel problema di ottimizzazione. La scelta del parametro  $\alpha$  dipende dal livello di confidenza della stima della segno macrozonale. Più alta è la fiducia nella stima (cioè

quanto la probabilità si avvicina a 0 o a 1, come discusso nel Capitolo 3), meno si è avversi al rischio. Al contrario, se la probabilità è più vicina a 0.5, si è meno fiduciosi e quindi più avversi al rischio.

Questo concetto si traduce in una logica di selezione del parametro  $\alpha$  all'interno dell'algoritmo. Ad esempio, se la probabilità stimata è maggiore o uguale a 0.65, il parametro  $\alpha$  viene impostato a un valore più basso  $\alpha = 0.5$ , indicando una minore avversione al rischio. Allo stesso modo, se la probabilità stimata è inferiore o uguale a 0.45, il parametro  $\alpha$  viene impostato a 0.5, riflettendo ancora una volta una minore avversione al rischio. Per tutte le altre probabilità, viene utilizzato un valore più alto di  $\alpha$ , ad esempio  $\alpha = 0.7$ , indicando una maggiore avversione al rischio.

Se il parametro alpha è più piccolo, infatti, significa che i vincoli probabilistici sono meno stringenti. Questo perché tale valore rappresenta un fattore che influenza la generazione delle soluzioni nell'algoritmo genetico, in base al livello di rischio desiderato. Quando alpha è più piccolo, si accetta un maggiore grado di incertezza o rischio nelle soluzioni prodotte, il che si traduce in vincoli meno severi. L'algoritmo sarà più incline a generare soluzioni che possono violare i vincoli probabilistici con una maggiore probabilità, poiché viene data meno enfasi al soddisfacimento rigoroso di tali vincoli. Questo può portare a una maggiore esplorazione dello spazio delle soluzioni, consentendo di considerare un numero più ampio di possibilità, anche se alcune di esse possono non rispettare completamente i vincoli.

## Metodi di generazione

1. Generare valori casuali per variabili con una distribuzione specificata nota, come ad esempio l'esponenziale o la normale, coinvolge due fasi distinte. Inizialmente, si genera una sequenza di numeri casuali distribuiti uniformemente nell'intervallo tra 0 e 1. Questa sequenza successivamente viene trasformata attraverso un processo noto come *generazione di variabili casuali*, al fine di ottenere una sequenza di valori casuali che rispetti la distribuzione desiderata.

Nel **caso analizzato** questo vale per la variabile stocastica segno macrozonale. Come detto nel capitolo 3, si assume che il segno macrozonale segua una distribuzione bernoulliana. Dunque, stimata la probabilità che il segno sia positivo tramite il Modello Generalizzato Additivo è possibile generare i valori casuali desiderati.

La distribuzione di Bernoulli è una distribuzione discreta che può assumere solo due valori: 0 e 1, con probabilità di successo  $p$  per 1 e probabilità di fallimento  $1 - p$  per 0.

Un approccio comune è utilizzare il metodo dell'inversione della funzione cumulativa. Per la distribuzione di Bernoulli, la funzione cumulativa è:

$$F(x) = \begin{cases} 0 & \text{per } x < 0 \\ 1 - p & \text{per } 0 \leq x < 1 \\ 1 & \text{per } x \geq 1 \end{cases}$$

Si possono generare valori casuali per una distribuzione di Bernoulli utilizzando questa funzione cumulativa. Il procedimento è il seguente:

- (a) Si genera un numero casuale  $u$  uniformemente distribuito tra 0 e 1.
- (b) Se  $u < 1 - p$ , viene restituito 0; altrimenti 1.

Questo metodo sfrutta il fatto che la probabilità che  $u$  cada nell'intervallo  $[0, 1 - p)$  è  $1 - p$ , mentre la probabilità che cada nell'intervallo  $[1 - p, 1]$  è  $p$ , rispettivamente corrispondenti a 0 e 1 nella distribuzione.

2. Sia invece una distribuzione empirica non nota. Per ottenere un insieme di campioni da tale distribuzione è possibile eseguire i seguenti passaggi, simili a quelli del punto precedente:
  - (a) Si calcola la funzione di densità cumulativa (CDF) a partire dalla distribuzione empirica.
  - (b) Si genera un numero casuale  $u$  uniformemente distribuito tra 0 e 1.
  - (c) Si applica il metodo della trasformata inversa per mappare il numero casuale uniformemente generato sulla distribuzione empirica. Questo processo individua il valore che corrisponde alla probabilità cumulativa associata al numero casuale uniforme.

Il processo viene ripetuto fino ad ottenere il numero desiderato di valori della variabile casuale. Questo metodo offre un modo efficace per campionare una distribuzione empirica e generare un insieme di campioni che riflettano fedelmente la distribuzione originale.

Nel **caso analizzato** questo è il caso della variabile stocastica errore percentuale. Si ottiene tramite l'algoritmo dei giorni simili un campione di valori di errori percentuali, ciascuno con il proprio peso (vedi Capitolo 2): questo rappresenta la distribuzione empirica pesata della variabile casuale, specificata la fascia oraria e il giorno in esame. Nella Figura 4.2 è riportato un esempio.

Per quanto riguarda una distribuzione empirica pesata, per calcolare la funzione cumulativa empirica, si ordinano i valori in ordine crescente e si calcola la frazione dei valori inferiori o uguali a ciascun valore, considerando le relative frequenze. Successivamente, si inverte tale funzione cumulativa e la si utilizza per generare i valori casuali, e quindi gli scenari di errore percentuale necessari.

È importante sottolineare che nella generazione degli  $N$  vettori casuali, si assume che il segno macrozonale e l'errore percentuale siano variabili casuali indipendenti. Si genera un campione di  $N$  valori del segno macrozonale ( $Segno_{t,h}^1, Segno_{t,h}^2, \dots, Segno_{t,h}^N$ ) e un campione di  $N$  valori per l'errore percentuale di sbilanciamento dell'unità  $i$ , ( $Err_{t,h,i}^1, Err_{t,h,i}^2, \dots, Err_{t,h,i}^N$ ), gli  $N$  vettori casuali indipendenti data la fascia oraria  $h$  sono

$$\xi_{t,h,i}^s = (Err_{t,h,i}^s, Segno_{t,h}^s), \quad s = 1, 2, \dots, N$$

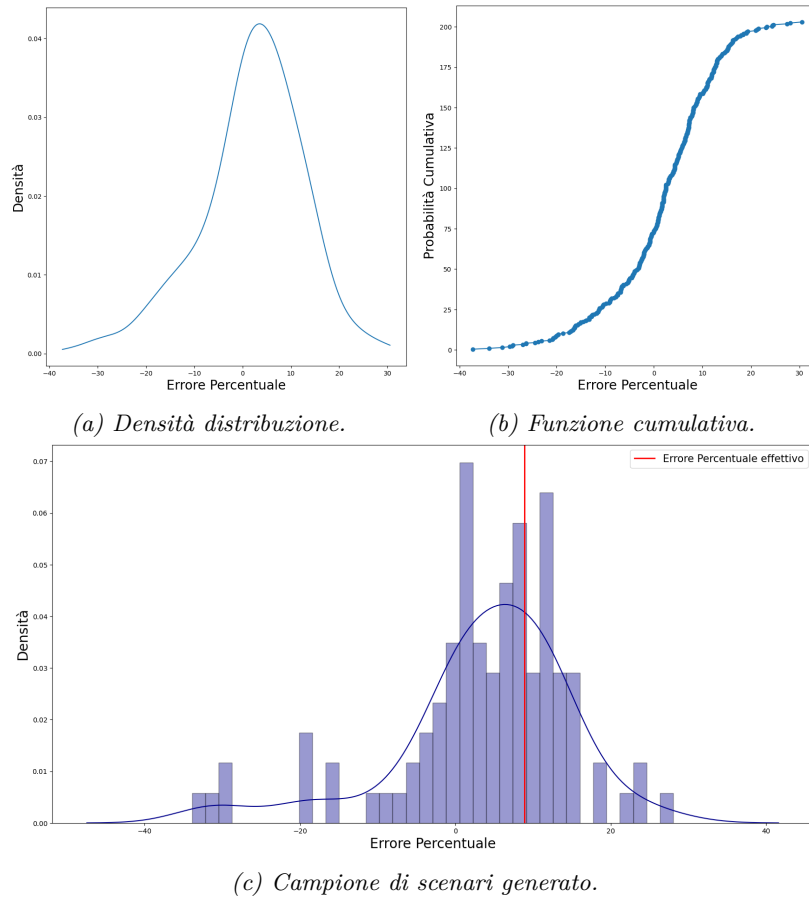


Figura 4.1: Generazione scenari a partire da distribuzione empirica pesata.

## 4.2.2 Algoritmo genetico

Un algoritmo genetico (GA) è una tecnica di ottimizzazione e ricerca ispirata ai principi della selezione naturale e della genetica evolutiva. Questi algoritmi sono utilizzati per trovare soluzioni approssimate a problemi complessi attraverso l'evoluzione iterativa di una popolazione di soluzioni candidate.

Nei GA, le soluzioni candidate sono rappresentate come "cromosomi" o "individui", solitamente codificati come stringhe di caratteri (ad esempio, stringhe binarie) o vettori di numeri reali. Una popolazione iniziale di cromosomi viene generata casualmente o utilizzando una conoscenza a priori, rappresentando un insieme di soluzioni potenziali al problema. Ogni cromosoma viene valutato utilizzando una funzione di fitness che misura quanto è buona la soluzione rappresentata da quel cromosoma rispetto al problema da risolvere. La funzione di fitness guida l'evoluzione della popolazione, che avviene attraverso tre principali operatori genetici: selezione, crossover e mutazione. La selezione è il processo di scelta di cromosomi dalla popolazione attuale per generarne una nuova, con

maggior probabilità di essere selezionati quelli con una fitness maggiore. Il crossover combina due cromosomi genitori per produrre uno o più cromosomi figli, permettendo lo scambio di informazioni tra soluzioni. La mutazione altera uno o più geni di un cromosoma con una piccola probabilità, introducendo variazione nella popolazione e aiutando a evitare di rimanere bloccati in ottimi locali. Ogni nuova generazione di soluzioni è vista come un miglioramento rispetto alla precedente, e questo processo viene ripetuto iterativamente fino a quando nella popolazione non compare una soluzione accettabile al problema.

In questa sezione, viene sviluppato un Algoritmo Genetico (GA) integrato alla simulazione stocastica per la risoluzione dei problemi di Ottimizzazione vincolati alla casualità (CCP). Vengono analizzati in dettaglio la struttura di rappresentazione, il processo di inizializzazione, la verifica dei vincoli mediante simulazione stocastica, la funzione di valutazione (fitness), il processo di selezione, l'operazione di crossover, l'operazione di mutazione e il processo di terminazione. L'efficienza dei GA nel trovare soluzioni ottimali dipende spesso dalla corretta personalizzazione della codifica, degli operatori di riproduzione e delle metriche di fitness per i problemi ingegneristici specifici. Ogni componente del GA deve essere adattato per affrontare le peculiarità del problema CCP, garantendo così un'efficace esplorazione dello spazio delle soluzioni e la convergenza verso soluzioni ottimali che soddisfino i vincoli stocastici.

Si riporta in dettaglio la personalizzazione del problema 4.6, nella rappresentazione 4.13 per  $\Delta_{t,h,i}^+$ . Per il problema riguardante  $\Delta_{t,h,i}^-$  il ragionamento è analogo.

### Struttura di Rappresentazione

La rappresentazione delle soluzioni nel GA è un aspetto cruciale. In questo contesto, le soluzioni sono rappresentate come cromosomi, dove ognuno di questi è la codifica binaria della possibile soluzione del problema di ottimizzazione. Per esempio, definita la dimensione  $C = 10$ , si ha il seguente cromosoma:

$$cromosoma = [0,0,0,1,1,1,0,0,1,1]$$

Dati limiti della soluzione reale del problema

$$0 \leq \Delta_{t,h,i}^+ \leq \Delta_{max,t,h,i} \implies lb = 0, ub = \Delta_{max,t,h,i},$$

la conversione dalla codifica binaria alla soluzione reale è:

$$\Delta_{t,h,i}^+ = \frac{lb + (ub - lb)}{2^C - 1} \sum_i cromosoma[i] \cdot 2^{C-1-i} \quad (4.15)$$

### Verifica dei Vincoli tramite Simulazione Stocastica

I vincoli di probabilità sono verificati utilizzando tecniche di simulazione stocastica, discussa nel paragrafo precedente. Per ogni soluzione candidata, la simulazione stocastica viene utilizzata per calcolare la probabilità che la soluzione soddisfi i vincoli. Questo è

essenziale per assicurare che le soluzioni proposte siano valide rispetto ai vincoli stocastici del problema.

Siano  $\xi_{t,h,i}^s = (Err_{t,h,i}^s, Segno_{t,h}^s)$ ,  $s = 1, 2, \dots, N$  i vettori casuali indipendenti, sia  $\Delta_{t,h,i}^+$  la soluzione reale corrispondente al cromosoma, allora  $N'$  è:

$$N' = \sum_{s=1}^N \delta(\Delta_{t,h,i}^+)$$

dove

$$\delta(\Delta_{t,h,i}^+) = \begin{cases} 1 & \text{se } \begin{cases} \Delta_{t,h,i}^+ \cdot Segno_{t,h}^s \leq 0, \\ -(\Delta_{t,h,i}^+ \cdot Segno_{t,h}^s) \leq 0, \\ (Err_{t,h,i}^s + \Delta_{t,h,i}^+) \Delta_{t,h,i}^+ \leq 0, \end{cases} \\ 0 & \text{altrimenti,} \end{cases} \quad (4.16)$$

### Processo di Inizializzazione

L'inizializzazione della popolazione è il primo passo del GA. Una popolazione iniziale viene generata casualmente, garantendo che le soluzioni siano distribuite uniformemente nello spazio ammissibile. Questo aiuta a esplorare una vasta area fin dalle prime generazioni.

Si genera quindi un cromosoma codificato come stringa di bit in modo casuale: se la soluzione reale corrispondente a tale cromosoma soddisfa, tramite la simulazione stocastica i vincoli di probabilità 4.16 allora il cromosoma viene considerato individuo della popolazione, in caso contrario viene scartato. Si ripete questo processo fino ad ottenere una popolazione della grandezza desiderata  $p_{size}$ , di individui che soddisfano i vincoli.

### Funzione di Valutazione

La funzione di valutazione, o funzione fitness, è utilizzata per misurare la qualità delle soluzioni. Nel contesto dei problemi CCP, la funzione fitness deve incorporare l'ottimizzazione dell'obiettivo. Dopo il soddisfacimento dei vincoli, vengono calcolati i valori delle funzioni obiettivo per tutti i cromosomi ( $cromosoma_1, cromosoma_2, \dots, cromosoma_{p_{size}}$ ). Si utilizza una stima Monte Carlo del valore atteso della funzione obiettivo del problema 4.13:

$$\mathbb{E}_{\xi}[f(\Delta_{t,h,i}^+, \xi_{t,h,i})] = \frac{1}{N} \sum_{s=1}^N f(x, \xi_{t,h,i}^s) \quad (4.17)$$

dove  $f(x, \xi_{t,h,i}^s) = -(Err_{t,h,i}^s + \Delta_{t,h,i}^+) \cdot (1 - Segno_{t,h}^s)$ .



Siccome il problema di ottimizzazione che si sta considerando è una minimizzazione, più il valore atteso è piccolo migliore è la soluzione considerata. Si utilizza dunque come funzione di fitness, dato il *cromosoma* e la relativa soluzione reale  $\Delta_{t,h,i}^+$ :

$$fitness[cromosoma] = \frac{1}{\mathbb{E}_{\xi}[f(\Delta_{t,h,i}^+, \xi_{t,h,i})] + 1} \quad (4.18)$$

## Processo di Selezione

Il processo di selezione determina quali soluzioni vengono scelte per la riproduzione. Tecniche comuni includono la selezione a torneo, la selezione proporzionale alla fitness (*roulette wheel*) e altre strategie che favoriscono le soluzioni con valori di fitness superiori, mantenendo una certa diversità nella popolazione.

Nell'algoritmo genetico implementato viene selezionata la coppia di individui meritevoli tramite la selezione a torneo. Si seleziona casualmente un sottoinsieme della popolazione attuale, la dimensione di questo sottoinsieme è nota come dimensione del torneo e può essere specificata a priori. Si valutano i cromosomi all'interno del sottoinsieme utilizzando la funzione di fitness. I due cromosomi con la fitness più alta vengono selezionati come vincitori del torneo. Si ripete il torneo fino a quando non si seleziona il numero desiderato di coppie di cromosomi per la generare successiva generazione ( $\frac{P_{size}}{2}$ ).

## Operazione di Crossover

Il crossover è un operatore genetico che combina due soluzioni (genitori) per produrre nuove soluzioni (figli). Questo processo aiuta a esplorare nuove aree dello spazio di ricerca combinando caratteristiche di soluzioni diverse. Vengono utilizzati vari tipi di crossover, come il crossover a un punto, a due punti o uniforme, a seconda della natura del problema.

In questo algoritmo genetico, è stato implementato un operatore di crossover che sfrutta una strategia a punto singolo. Viene generato casualmente un valore compreso tra 0 e 1. Si confronta questo valore con la probabilità di crossover  $p_{cross}$ . Se il valore casuale è inferiore a  $p_{cross}$ , non si esegue il crossover e si mantengono intatti i genitori, procedendo direttamente alla fase successiva per la generazione di una nuova popolazione. Se questo valore casuale, invece, supera la probabilità di crossover  $p_{cross}$ , viene selezionata casualmente una posizione intera nell'intervallo  $[1, C]$ , dove  $C$  rappresenta la lunghezza dei cromosomi. I geni dei due genitori a partire dalla posizione selezionata vengono quindi scambiati per generare due figli. Una volta creati, viene effettuata una verifica per garantire che soddisfino i vincoli imposti dal problema. Se entrambi i figli risultano fattibili, ovvero rispettano tutti i vincoli, procedono alla fase successiva dell'algoritmo genetico. Tuttavia, se uno o entrambi i figli non sono fattibili, i genitori originali vengono conservati per il prossimo ciclo evolutivo.

## Operazione di Mutazione

La mutazione introduce variazioni casuali nelle soluzioni, aiutando a mantenere la diversità genetica nella popolazione e a prevenire la convergenza prematura a soluzioni subottimali. La mutazione modifica casualmente uno o più geni in un cromosoma secondo una certa probabilità.

Nel sistema genetico, la probabilità di mutazione è definita come  $p_{mut}$ . Di conseguenza,  $p_{mut} \times p_{size}$  rappresenta il numero atteso di cromosomi che subiranno l'operazione di mutazione. Per determinare gli individui della generazione successiva, dopo aver eseguito l'operazione di crossover il seguente processo viene ripetuto per  $j = 1, 2, \dots, p_{size}$ :

- Selezionare il  $j$ -esimo cromosoma, che a partire dall'operazione precedente può essere un nuovo individuo o un cromosoma che è passato allo step successivo intatto.
- Generare casualmente un numero reale  $r$  dall'intervallo  $[0, 1]$ .
- Se  $r < p_{mut}$ :
  - Generare casualmente una posizione intera  $pos$  nell'intervallo  $[1, C]$ .
  - Mutare il bit nella posizione  $pos$ .

Altrimenti il cromosoma rimane immutato.

Dopo aver effettuato la mutazione, è necessario verificare la fattibilità del nuovo cromosoma. Se questo nuovo individuo risultante dalla mutazione non soddisfa i vincoli, il cromosoma padre popola la generazione successiva.

## Processo di Terminazione

Il processo di terminazione definisce quando il GA deve fermarsi. I criteri di terminazione comuni includono un numero massimo di generazioni, una soglia di miglioramento minimo della fitness o una combinazione di entrambi. È importante scegliere un criterio di terminazione appropriato per bilanciare il tempo di calcolo e la qualità delle soluzioni trovate.

In questo caso viene definito il numero di generazioni. Questo parametro determina il numero massimo di iterazioni che l'algoritmo eseguirà prima di terminare. Questo è cruciale poiché bilancia l'efficienza computazionale con la qualità della soluzione. Un numero troppo basso di generazioni potrebbe non permettere all'algoritmo di esplorare adeguatamente lo spazio delle soluzioni, portando a risultati subottimali. Al contrario, un numero troppo elevato di generazioni può risultare in tempi di esecuzione eccessivamente lunghi senza un significativo miglioramento delle soluzioni.

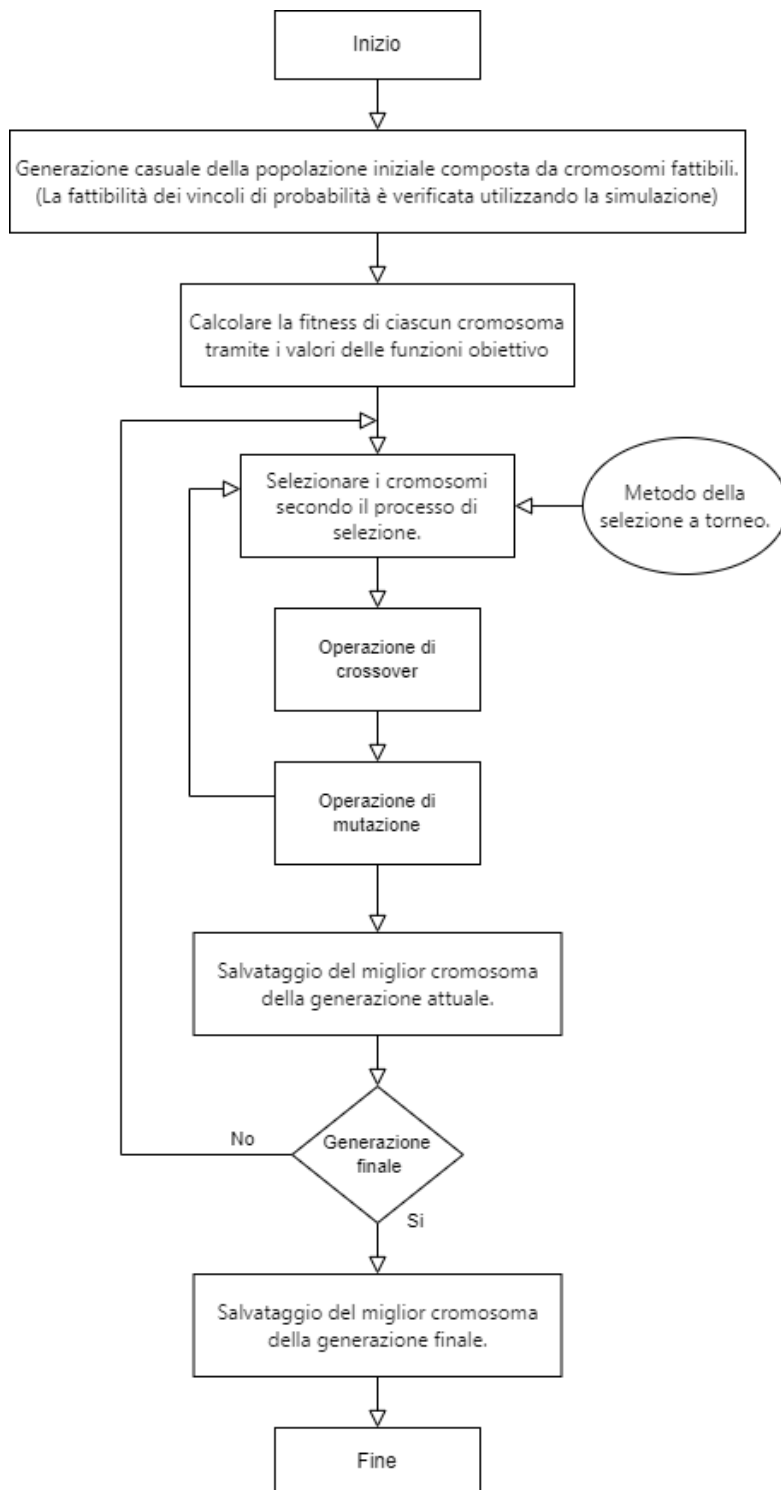


Figura 4.2: Diagramma di flusso algoritmo genetico implementato.

### Parametri di controllo per l'Algoritmo Genetico

Parametro	Valore
Dimensione del vettore delle variabili decisionali ( $n$ )	1
Dimensione della popolazione ( $K$ )	$10 \times n$
Numero di generazioni ( $M$ )	$5 \times K$
Dimensione del cromosoma ( $C$ )	10
Probabilità di crossover ( $p_{\text{cross}}$ )	0.7
Probabilità di mutazione ( $p_{\text{mut}}$ )	0.3
Dimensione dello spazio campione ( $N$ )	100

Tabella 4.1: Parametri di controllo per l'Algoritmo Genetico.

### 4.3 Valutazione dei Risultati e Analisi Economica

Si eseguono più simulazioni dell'algoritmo utilizzando spazi campione differenti. Da ciascuna simulazione si ottiene una soluzione ottimale del problema che può variare da una simulazione all'altra. La soluzione finale viene determinata calcolando la media di queste soluzioni ottimali. Le motivazioni dietro la ripetizione di più simulazioni dell'algoritmo genetico sono varie:

1. **Robustezza dell'algoritmo:** eseguire l'algoritmo genetico più volte su uno stesso problema consente di valutarne la robustezza e la consistenza delle soluzioni. Se l'algoritmo produce risultati simili o identici in diverse esecuzioni, si può avere maggiore fiducia nella validità delle soluzioni ottenute.
2. **Variabilità nelle inizializzazioni:** l'algoritmo genetico dipende fortemente dalla inizializzazione casuale della popolazione. Eseguire più simulazioni con diverse inizializzazioni consente di esplorare una più ampia gamma di spazi di ricerca e di individuare soluzioni migliori.
3. **Valutazione delle prestazioni:** l'esecuzione ripetuta dell'algoritmo consente di confrontare le prestazioni in termini di convergenza, tempo di esecuzione e qualità delle soluzioni ottenute. Questo aiuta a identificare configurazioni migliori dei parametri e strategie di selezione, crossover e mutazione.
4. **Analisi della variabilità dei risultati:** le simulazioni multiple permettono di valutare la variabilità dei risultati ottenuti dall'algoritmo genetico. Questa analisi

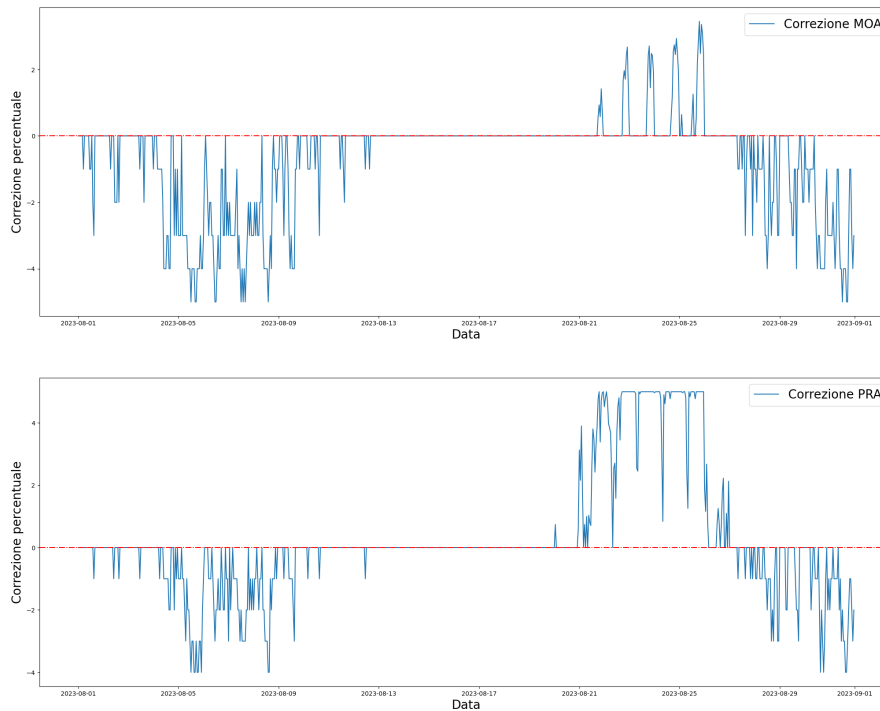
può essere utile per comprendere la consistenza delle soluzioni e per valutare la presenza di ottimi locali o di convergenza prematura.

Complessivamente, eseguire più simulazioni dell'algoritmo genetico è una pratica comune per garantire la validità dei risultati, valutare la robustezza dell'algoritmo e individuare strategie ottimali per la risoluzione del problema considerato.

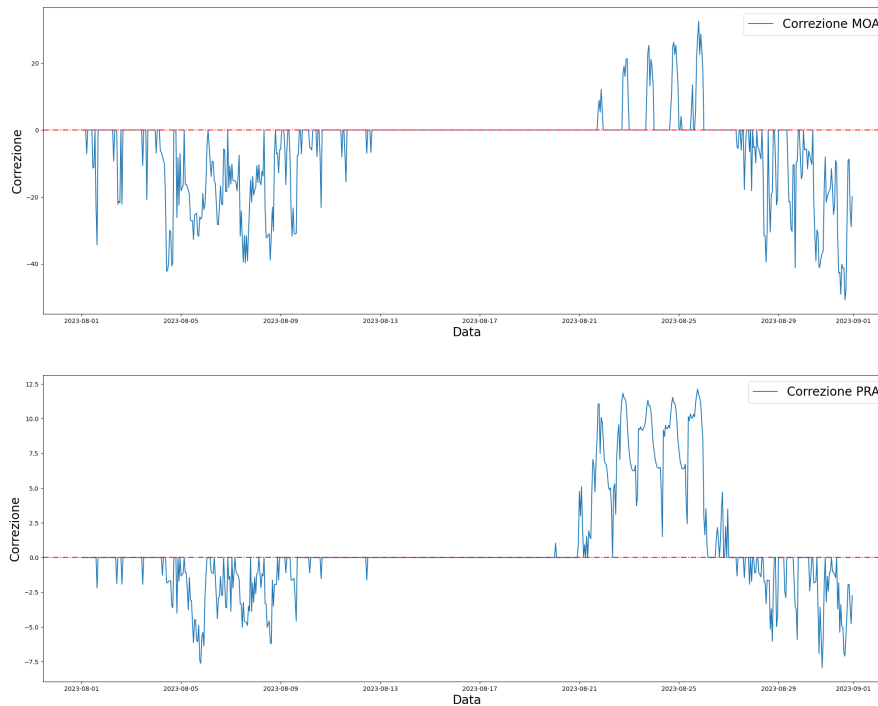
### Caso di studio:

Nel presente studio, si riprende il caso dell'unità  $i$  localizzata nella zona Nord. Si impiega l'algoritmo genetico di ottimizzazione per il mese di Agosto 2023, utilizzando i parametri descritti nella tabella di riferimento 4.1. Per la generazione degli scenari delle variabili stocastiche, l'errore percentuale di sbilanciamento dell'unità e il segno macrozonale relativo, ci si basa sui risultati ottenuti nei capitoli precedenti e sui metodi di generazione sopra citati.

L'intero procedimento viene applicato sia al problema 4.7 che al problema 4.8 al fine di ottenere, per ogni giorno  $t$  e fascia oraria  $h$  del mese in esame, il vettore di correzioni da applicare al risultato della previsione di consumo elettrico. Queste correzioni,  $\Delta_{t,h,i}^+$  e  $\Delta_{t,h,i}^-$ , espresse in termini di valore percentuale, vengono determinate per l'unità  $i$  sia per il MO che per il PRA, separatamente. I risultati sono riportati in 4.3a.



(a) Correzioni in termini percentuali.



(b) Correzioni in termini di volume.

Figura 4.3: Risultati problema ottimizzazione.

Dato che Terna S.p.A determina le penalità basandosi sul volume totale di sbilanciamento effettuato dall'unità  $i$ , ovvero il volume previsto per il MO più il volume previsto per il PRA sottratto del volume effettivo per il MO e del volume effettivo per il PRA, è indispensabile calcolare le correzioni in termini di volume anziché in termini percentuali. Questo è cruciale per valutare l'impatto delle correzioni sia sul piano economico che su quello operativo. I risultati in termini di volume, ottenuti moltiplicando le correzioni percentuali per il valore di previsione, sono riportati in 4.3b.

Viene dunque applicata la seguente correzione riportata nella Figura 4.4.

Nelle tabelle a doppia entrata riportate in 4.3, viene indicata, in termini di volume, la distribuzione dello sbilanciamento rispetto al segno macrozonale prima e dopo l'applicazione della correzione. Questi dati risultano cruciali per comprendere le dinamiche degli sbilanciamenti e le loro implicazioni economiche e operative. La diagonale penalizzante, che rappresenta gli sbilanciamenti con effetti negativi, è evidenziata in rosso per una facile identificazione. Al contrario, gli sbilanciamenti discordi, che possono avere effetti più favorevoli, sono evidenziati in verde. Si nota un miglioramento in termini di sbilanciamento assoluto totale, ma soprattutto una diminuzione dello sbilanciamento sulla diagonale penalizzante, a seguito dell'applicazione delle correzioni.

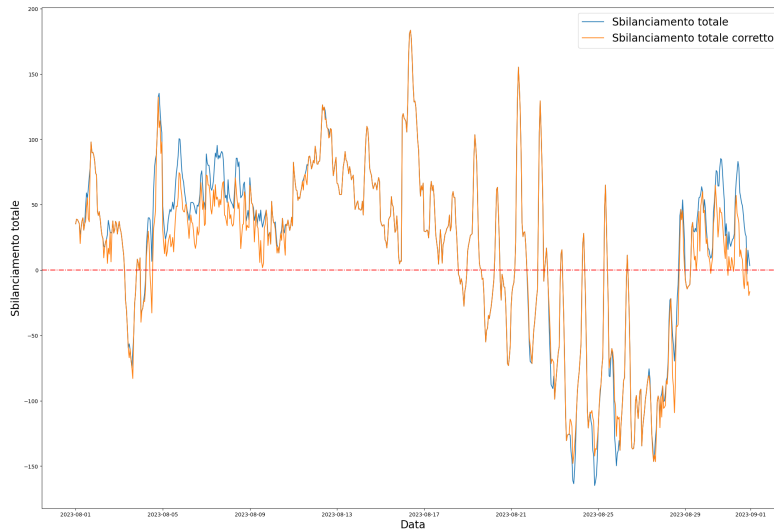


Figura 4.4: Sbilanciamento corretto.

Nella tabella 4.4, invece, viene riportato l'effetto di una correzione casuale compresa tra  $-\Delta_{max,t,h,i}$  e  $\Delta_{max,t,h,i}$ , al fine di valutare l'efficacia dell' algoritmo proposto. Confrontando i risultati ottenuti con correzioni casuali con quelli derivanti dall'algoritmo, è possibile evidenziare i vantaggi specifici dell'approccio ottimizzato, dimostrando così la sua superiorità in termini di riduzione degli sbilanciamenti e delle penalità associate.

		<b>Sbilanciamento macrozonale:</b>	
		Sbilanciamento positivo	Sbilanciamento negativo
<b>Sbilanciamento zonale dell'unità:</b>	Sbilanciamento positivo	13285	15748
	Sbilanciamento negativo	2242	12512

Diagonale penalizzante:	25797
Diagonale non penalizzante:	17990
Diagonale non penalizzante sul totale:	41%
Sbilanciamento totale (assoluto):	43787

Tabella 4.2: Volume totale sbilanciamento rispetto alla macrozona.

		<b>Sbilanciamento macrozonale:</b>	
		Sbilanciamento positivo	Sbilanciamento negativo
<b>Sbilanciamento zonale dell'unità:</b>	Sbilanciamento positivo	10033	14704
	Sbilanciamento negativo	2203	11717
Diagonale penalizzante:		21750	
Diagonale non penalizzante:		16907	
Diagonale non penalizzante sul totale:		43,8%	
Sbilanciamento totale (assoluto):		38675	

Tabella 4.3: Volume totale sbilanciamento rispetto alla macrozona, a seguito della correzione.

		<b>Sbilanciamento macrozonale:</b>	
		Sbilanciamento positivo	Sbilanciamento negativo
<b>Sbilanciamento zonale dell'unità:</b>	Sbilanciamento positivo	13700	16124
	Sbilanciamento negativo	2391	13736
Diagonale penalizzante:		27436	
Diagonale non penalizzante:		18515	
Diagonale non penalizzante sul totale:		40%	
Sbilanciamento totale (assoluto):		45951	

Tabella 4.4: Volume totale sbilanciamento rispetto alla macrozona, a seguito della correzione randomica.



### 4.3.1 Analisi della valorizzazione economica dello sbilanciamento energetico

Come discusso nel capitolo 1, è importante ricordare che il sistema a tariffazione unica per lo sbilanciamento delle Unità Non Qualificate, ossia le unità analizzate nel progetto, rappresenta un approccio semplificato rispetto ai sistemi più complessi che utilizzano prezzi variabili per lo sbilanciamento positivo e negativo. Nel caso di un sistema a tariffazione unica, viene applicata una tariffa uniforme per tutti gli sbilanciamenti, indipendentemente dal fatto che l'operatore stia immettendo più o meno energia di quanto programmato.

Questo approccio semplificato facilita la gestione e il calcolo delle penalità per le Unità Non Qualificate, riducendo la complessità associata alla gestione degli sbilanciamenti energetici. Viene stabilita una tariffa unica per lo sbilanciamento che si applica a tutti gli operatori che si discostano dalle loro previsioni di immissione o prelievo di energia. Tale tariffa è calcolata in base a criteri predefiniti, che possono includere i costi medi del bilanciamento del sistema e altri fattori operativi.

Gli operatori sono soggetti a un addebito in base alla quantità di energia sbilanciata (positiva o negativa) moltiplicata per la tariffa unica:

$$\text{Costo Sbilanciamento} = |S_{t,h,i}| \times P_{t,h}$$

dove  $S_{t,h,i}$  rappresenta lo sbilanciamento dell'unità  $i$  nel tempo  $t$  e nella fascia oraria  $h$ , e  $P_{t,h}$  indica la tariffa unica applicata. L'ente regolatore e il gestore della rete monitorano costantemente le condizioni del sistema e possono aggiustare periodicamente la tariffa unica per riflettere i costi effettivi del bilanciamento.

L'unità si trova però in una posizione di perdita o di guadagno a seconda della sua posizione reciproca con il segno macrozonale relativo. Questa posizione è valorizzata da:

$$\text{Valorizzazione Sbilanciamento} = S_{t,h,i} \times (P_{t,h} - P_{MGP,t,h})$$

Prima di presentare i risultati ottenuti dall'applicazione del progetto, si considera il seguente esempio per chiarire meglio il contesto:

**Esempio:** Si supponga che un'unità abbia programmato di immettere 100 MWh in una certa fascia oraria. Se l'unità immette effettivamente 110 MWh, lo sbilanciamento è di  $S_{t,h,i} = +10$  MWh. Se invece immette solo 90 MWh, lo sbilanciamento è di  $S_{t,h,i} = -10$  MWh. In un sistema a tariffazione unica, la tariffa applicata a questi sbilanciamenti è la stessa, indipendentemente dalla direzione dello sbilanciamento.

Quindi, se la tariffa unica per lo sbilanciamento applicata da Terna S.p.A. è, ad esempio,  $P_{t,h} = 50$  €/MWh, allora:

**Sbilanciamento Positivo:** Per uno sbilanciamento di +10 MWh, l'unità pagherà a Terna S.p.A.:

$$10 \text{ MWh} \times 50 \text{ €/MWh} = 500 \text{ €}$$

**Sbilanciamento Negativo:** Per uno sbilanciamento di -10 MWh, l'unità pagherà a Terna S.p.A.:

$$10 \text{ MWh} \times 50 \text{ €/MWh} = 500 \text{ €}$$

In entrambi i casi, l'importo pagato a Terna S.p.A. sarebbe di 500 €, poiché la tariffa è uniforme per ogni MWh di sbilanciamento, indipendentemente dal fatto che si tratti di un surplus o di un deficit di energia immessa rispetto alla quantità programmata.

Si consideri ora il caso in cui la macrozona dell'unità sia positiva. Questo implica che  $P_{t,MGP} > P_{t,h}$ , ad esempio  $P_{t,MGP} = 55 \text{ €/MWh}$ .

**Sbilanciamento Positivo:** In questo caso, se lo sbilanciamento è positivo,  $S_{t,h,i} = +10 \text{ MWh}$ , l'energia extra è venduta a meno di quanto è stata acquistata nel Mercato del Giorno Prima:

$$10 \text{ MWh} \times (50 \text{ €/MWh} - 55 \text{ €/MWh}) = -50 \text{ €} < 0,$$

dunque, oltre al costo dello sbilanciamento pagato a Terna S.p.A. si è in una posizione di perdita.

**Sbilanciamento Negativo :** Se lo sbilanciamento è negativo,  $S_{t,h,i} = -10 \text{ MWh}$ , l'energia mancante è comprata a meno di quanto sarebbe stata pagata nel Mercato del Giorno Prima:

$$-10 \text{ MWh} \times (50 \text{ €/MWh} - 55 \text{ €/MWh}) = +50 \text{ €} > 0,$$

dunque, oltre al costo dello sbilanciamento pagato a Terna S.p.A. si è in una posizione di guadagno.

Si riportano i risultati in termini di costo dello sbilanciamento totale di Agosto 2023 in 4.5 e di valorizzazione di quest'ultimo in 4.6. Sono anche riportati i risultati ottenuti applicando la correzione randomica.

Le serie dei prezzi del Mercato del Giorno Prima e dei prezzi di sbilanciamento per la macrozona Nord e per ciascun giorno e fascia oraria in esame sono scaricabili dal sito del Gestore dei Mercati Elettrici S.p.A.

<b>Sbilanciamento</b>	<b>Sbilanciamento post correzione</b>	<b>Sbilanciamento post correzione randomica</b>
3.323.327 €	2.873.013 €	3.384.414 €

Tabella 4.5: Costo sbilanciamento, confronto con correzione randomica.

	<b>Sbilanciamento</b>	<b>Sbilanciamento post correzione</b>	<b>Sbilanciamento post correzione randomica</b>
Concorde alla zona	-1.992.250 €	-1.623.499 €	- 1.974.850 €
Discorde alla zona	1.331.077 €	1.249.513 €	1.409.564 €
<b>Totale</b>	<b>-661.173 €</b>	<b>-373.986 €</b>	<b>-565.286 €</b>

Tabella 4.6: Valorizzazione sbilanciamento, confronto con correzione randomica.

I buoni risultati ottenuti evidenziano l'efficacia dell'algoritmo proposto nella gestione dello sbilanciamento energetico. La riduzione dei costi di sbilanciamento totale, insieme ad una valorizzazione migliore, dimostra l'utilità delle correzioni applicate nell'ottimizzazione del processo.

In particolare, l'applicazione delle correzioni ha portato a una significativa diminuzione dei costi di sbilanciamento totale rispetto ai risultati ottenuti senza correzioni o con correzioni casuali. Questo suggerisce che l'algoritmo sia in grado di individuare correzioni efficaci che contribuiscono a mitigare lo sbilanciamento energetico, riducendo al contempo i costi associati.

# Conclusione

Questo elaborato fornisce un'analisi dettagliata del mercato dell'energia elettrica in Italia, concentrando l'attenzione sulle metodologie per la previsione dello sbilanciamento energetico e sull'ottimizzazione del sistema di previsione.

L'analisi del mercato dell'energia elettrica in Italia, descritta nei vari capitoli, offre una visione approfondita di un settore complesso e in costante evoluzione. Partendo dall'istituzione del mercato interno dell'energia nel 1999, sono stati esplorati dettagliatamente il Mercato Elettrico a Pronti (MPE) e i principali sottomercati che contribuiscono all'efficienza e alla stabilità del sistema energetico nazionale.

Il metodo adottato per la previsione dello sbilanciamento energetico zonale ha dimostrato l'importanza di considerare variabili ambientali e temporali per ottenere previsioni precise. L'impiego dell'algoritmo dei giorni simili e l'integrazione dettagliata delle variabili climatiche hanno notevolmente migliorato l'accuratezza delle previsioni. Inoltre, l'utilizzo dei modelli additivi generalizzati (GAM) con l'approccio del *boosting* del gradiente ha fornito una base robusta per predire il segno macrozonale dello sbilanciamento, catturando relazioni non lineari tra le covariate e la probabilità di sbilanciamento.

L'ottimizzazione del sistema di previsione energetica ha mirato alla riduzione delle perdite e alla correzione degli errori di previsione, attraverso l'applicazione combinata di Algoritmi Genetici e simulazione Monte Carlo. Questo approccio ha dimostrato efficacemente come gestire l'incertezza e migliorare l'efficienza delle soluzioni, portando a benefici tangibili sia in termini economici che operativi.

I casi studio specifici hanno confermato l'efficacia delle strategie proposte, evidenziando una significativa riduzione dei costi di sbilanciamento e una migliore valorizzazione del consumo energetico. Questi risultati forniscono agli operatori del sistema energetico strumenti pratici per prendere decisioni informate e migliorare la stabilità del sistema, promuovendo nel contempo comportamenti etici e responsabili.

In conclusione, l'approccio integrato e metodologico delineato in questo studio offre un quadro chiaro e dettagliato per affrontare le sfide del mercato energetico italiano. Attraverso analisi approfondite, previsioni accurate e tecniche di ottimizzazione, si mira a migliorare la gestione dell'energia elettrica, contribuendo così a garantire un approvvigionamento stabile, efficiente ed economicamente vantaggioso.

# Bibliografia

- [1] Patrick Avis and Jeo Lee. Evaluating system imbalance forecasting models for the united kingdom electricity market. 2021.
- [2] Vincent Charles and Udhayakumar Annamalai. Genetic algorithm for chance constrained reliability stochastic optimisation problems. *Int. J. of Operational Research*, 14:417 – 432, 06 2012. doi: 10.1504/IJOR.2012.047513.
- [3] Abraham Charnes and William W. Cooper. Chance-constrained programming. *Management Science*, 6:73–79, 1959. URL <https://api.semanticscholar.org/CorpusID:122269097>.
- [4] A. Géron. *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems*. O’Reilly Media, 2019. ISBN 9781492032618. URL <https://books.google.it/books?id=HHetDwAAQBAJ>.
- [5] GME. Gestore mercati energetici. <https://mercatoelettrico.org/It/Default.aspx>.
- [6] Trevor J. Hastie and Ryan J. Tibshirani. Generalized additive models. 2014. URL <https://api.semanticscholar.org/CorpusID:61047344>.
- [7] Benjamin Hofner, Andreas Mayr, Nikolay Robinzonov, and Matthias Schmid. Model-based boosting in r: a hands-on tutorial using the r package mboost. *Computational Statistics*, 29:3 – 35, 2012. URL <https://api.semanticscholar.org/CorpusID:14182802>.
- [8] Tao Hong and Michael Shahidehpour. Short-term load forecasting, part i: Methodologies and electric load forecasting. *IEEE Transactions on Power Systems*, 30(6): 1235–1242, 2015. doi: 10.1109/TPWRS.2014.2351491. URL <https://doi.org/10.1109/TPWRS.2014.2351491>.
- [9] Kakuzo Iwamura and Baoding Liu. A genetic algorithm for chance constrained programming. *Journal of Information and Optimization Sciences*, 17(2):409–422, 1996. doi: 10.1080/02522667.1996.10699291. URL <https://doi.org/10.1080/02522667.1996.10699291>.

- 
- [10] Thomas Kneib, Torsten Hothorn, and Gerhard Tutz. Variable selection and model choice in geoadditive regression models. *Biometrics*, 65(2):626–634, 2009. ISSN 0006341X, 15410420. URL <http://www.jstor.org/stable/25502326>.
- [11] Francesco Lisi and Enrico Edoli. Analyzing and forecasting zonal imbalance signs in the italian electricity market. *The Energy Journal*, 39(5):1–20, 2018. ISSN 01956574, 19449089. URL <https://www.jstor.org/stable/26534473>.
- [12] Zhenya Liu et al. Load forecasting based on a novel seasonal decomposition method and neural networks. *IEEE Transactions on Smart Grid*, 7(3):1233–1241, 2016. doi: 10.1109/TSG.2016.2543931. URL <https://doi.org/10.1109/TSG.2016.2543931>. Accessed: 2024-06-15.
- [13] Qingqing Mu, Yonggang Wu, Xiaoqiang Pan, Liang Huang, and Xian Li. Short-term load forecasting using improved similar days method. *2010 Asia-Pacific Power and Energy Engineering Conference*, pages 1–4, 2010. URL <https://api.semanticscholar.org/CorpusID:29399821>.
- [14] OECD. *PISA Data Analysis Manual: SAS Second Edition*. 01 2009. ISBN 9789264056268. URL <https://www2.oecd.org/pisa/pisaproducts/pisadataanalysismanualspssandsassecondedition.htm>.
- [15] Giorgia Oggioni and Cristian Lanfranconi. Empirics of intraday and real-time markets in europe: Italy. Technical report, Berlin, 2015. URL <https://hdl.handle.net/10419/111267>.
- [16] Chandra A. Poojari and Bobby Varghese. Genetic algorithm based technique for solving chance constrained problems. *European Journal of Operational Research*, 185(3):1128–1154, 2008. ISSN 0377-2217. doi: <https://doi.org/10.1016/j.ejor.2006.06.045>. URL <https://www.sciencedirect.com/science/article/pii/S0377221706006321>.
- [17] Reuven Y. Rubinstein, Gregory Levitin, Anatoly Lisnianski, and Hanoch Ben-Haim. Redundancy optimization of static series-parallel reliability models under uncertainty. *IEEE Transactions on Reliability*, 46:503–511, 1997. URL <https://api.semanticscholar.org/CorpusID:122473901>.
- [18] Pang-Ning Tan, Anuj Karpatne, Michael Steinbach, and Vipin Kumar. *Introduction to Data Mining Global Edition*. Pearson Deutschland, 2019. ISBN 9780273769224. URL <https://elibrary.pearson.de/book/99.150005/9780273775324>.
- [19] Terna S.p.A. Terna. driving energy. <https://www.terna.it/it>.
- [20] Jianhui Wang et al. Energy demand forecasting: A review of models and methods. *Renewable and Sustainable Energy Reviews*, 113:109232, 2019. doi: 10.1016/j.rser.2019.109232. URL <https://doi.org/10.1016/j.rser.2019.109232>. Accessed: 2024-06-15.