



Politecnico di Torino

Progettazione e gestione di impianti energetici

A.a. 2024/2025

Sessione di laurea Novembre 2025

Monitoraggio e analisi dei consumi della sede EthosEnergy di Torino ed implementazione di un modello predittivo per la loro stima

Relatore:

Prof. Pierluigi Leone

Candidati:

Cataldo Malcangi

Co-relatore:

Prof. Igor Stievano

Abstract

La crescente attenzione verso l'efficienza e la sostenibilità energetica in ambito industriale rende sempre più rilevante l'impiego di modelli previsionali capaci di anticipare l'andamento dei consumi. In questo contesto si inserisce il presente lavoro di tesi, che ha come obiettivo lo sviluppo di un modello di previsione dei consumi elettrici per lo stabilimento EthosEnergy di Torino, finalizzato a migliorare la gestione operativa e strategica dell'energia nel breve periodo (*short-term forecasting*).

Il modello è stato realizzato mediante due differenti approcci: la Regressione Lineare (LR), di tipo statistico e interpretabile, e la Gaussian Process Regression (GPR), basata su tecniche di *machine learning* e in grado di rappresentare relazioni non lineari. Entrambi i metodi sono stati applicati ai dati orari di consumo aziendale, analizzando l'effetto di parametri quali il resampling e l'orizzonte temporale di previsione, e valutandone le prestazioni attraverso le metriche RMSE (Root Mean Square Error) e MAE (Mean Absolute Error).

I risultati hanno evidenziato una maggiore accuratezza e stabilità del modello GPR, capace di cogliere relazioni non lineari e dinamiche complesse del segnale energetico, mentre la LR si distingue per semplicità di implementazione e rapidità di calcolo.

L'applicazione di tecniche di previsione dei consumi si è dimostrata un valido strumento per la pianificazione e l'ottimizzazione dei processi aziendali, offrendo un'opportunità di gestione più consapevole dell'energia e un miglioramento complessivo dell'efficienza operativa. Il lavoro rappresenta inoltre un primo passo verso l'integrazione di strumenti predittivi nei sistemi di monitoraggio aziendale, aprendo la strada a futuri sviluppi orientati all'estensione dell'orizzonte temporale e/o all'impiego di variabili esplicative aggiuntive.

Indice

Elenco delle tabelle	IV
Elenco delle figure	VI
1 Introduzione	1
1.1 Contesto energetico nazionale e introduzione metodi di previsione dei consumi in ambito energetico	1
1.2 Obiettivo e finalità della tesi	4
1.3 Struttura della tesi	5
2 Descrizione dell'azienda oggetto di studio	6
2.1 Presentazione contesto aziendale	6
2.2 Attività produttiva e organizzazione degli spazi	8
2.3 Descrizione impianti esistenti	10
2.4 Profilo energetico attuale e passato dell'azienda	11
2.4.1 Raccolta e analisi dei dati di consumo	11
2.4.2 Indicatori energetici	19
2.5 Utilizzo vettori energetici	23
3 Descrizione modello e metodi utilizzati	24
3.1 Definizione del modello di previsione e obiettivi della previsione . .	24
3.1.1 Linear Regression	26
3.1.2 Guassian Process Regression (GPR)	28
3.2 Struttura e costruzione del modello	30
3.3 Metriche di valutazione utilizzate	34
3.3.1 RMSE	34
3.3.2 MAE	35
4 Previsione dati di consumo dell'azienda EthosEnergy oggetto di studio	36
4.1 Definizione e descrizione del dataset	36

4.2	Risultati modello di previsione	39
4.3	Previsione one-step-ahead	41
4.3.1	Analisi del parametro di resampling	44
4.4	Previsione multi-step ahead	59
4.4.1	Analisi del parametro di resampling	62
4.4.2	Analisi del parametro di orizzonte temporale di previsione .	72
5	Analisi e commento dei risultati	75
5.1	Analisi risultati previsione <i>one-step-ahead</i>	76
5.2	Analisi risultati previsione <i>multi-step-ahead</i>	80
6	Conclusioni	86
6.1	Sintesi e commento finale dei risultati	86
6.2	Considerazioni finali e sviluppi futuri	89
	Bibliografia	91

Elenco delle tabelle

2.1	Spazi stabilimento	8
4.1	Metriche di valutazione LR - one step prediction	42
4.2	Metriche di valutazione GPR - one step prediction	43
4.3	Metriche di valutazione LR – one step prediction ($t_{resampling} = 4$ h)	45
4.4	Metriche di valutazione GPR – one step prediction ($t_{resampling} = 4$ h)	46
4.5	Metriche di valutazione LR – one step prediction ($t_{resampling} = 8$ h)	48
4.6	Metriche di valutazione GPR – one step prediction ($t_{resampling} = 8$ h)	49
4.7	Metriche di valutazione LR – one step prediction ($t_{resampling} = 12$ h)	51
4.8	Metriche di valutazione GPR – one step prediction ($t_{resampling} = 12$ h)	52
4.9	Metriche di valutazione LR – one step prediction ($t_{resampling} = 24$ h)	54
4.10	Metriche di valutazione GPR – one step prediction ($t_{resampling} = 24$ h)	55
4.11	Metriche di valutazione LR – one step prediction ($t_{resampling} = 48$ h)	57
4.12	Metriche di valutazione GPR – one step prediction ($t_{resampling} = 48$ h)	58
4.13	Metriche di valutazione LR – multi step prediction	60
4.14	Metriche di valutazione GPR – multi step prediction	61
4.15	Metriche di valutazione LR – multi step prediction ($t_{resampling} = 4$ h)	64
4.16	Metriche di valutazione GPR – multi step prediction ($t_{resampling} = 4$ h)	65
4.17	Metriche di valutazione LR – multi step prediction ($t_{resampling} = 6$ h)	67
4.18	Metriche di valutazione GPR – multi step prediction ($t_{resampling} = 6$ h)	68
4.19	Metriche di valutazione LR – multi step prediction ($t_{resampling} = 8$ h)	70

4.20 Metriche di valutazione GPR – multi step prediction ($t_{resampling} = 8$ h)	71
4.21 Metriche di valutazione LR – multi step prediction, variazione orizzonte temporale n	73
4.22 Metriche di valutazione GPR – multi step prediction, variazione orizzonte temporale n	74

Elenco delle figure

2.1	Logo azienda EthosEnergy	6
2.2	Sede EthosEnergy Torino	7
2.3	Pianta di stabilimento	8
2.4	Consumo elettrico 2024	11
2.5	Consumo gas naturale 2024	12
2.6	Confronto consumi elettrici 2024vs2023vs2022	13
2.7	Confronto consumi elettrici 2024vs2023vs2022	13
2.8	Confronto consumo gas naturale 2024vs2023vs2024	14
2.9	Confronto consumo gas naturale 2024vs2023vs2024	15
2.10	Consumo acqua potabile 2024	16
2.11	Consumo acqua industriale 2024	16
2.12	Confronto consumo totale acqua potabile 2024vs2023vs2022	17
2.13	Confronto consumo totale acqua industriale 2024vs2023vs2022	18
2.14	Consumi tep elettrici 2024	19
2.15	Consumi tep termici 2024	20
2.16	Consumo totale tep 2024	20
2.17	Produzione tCO ₂ da fonti elettriche 2024	21
2.18	Produzione tCO ₂ da fonti termiche 2024	21
2.19	Produzione totale di tCO ₂ 2024	22
4.1	Consumo stabilimento EthosEnergy periodo di osservazione	37
4.2	Consumo stabilimento EthosEnergy periodo di osservazione	38
4.3	LR - one step prediction [testing period]	41
4.4	LR - one step prediction [zoom 7 days]	42
4.5	GPR - one step prediction [testing period]	43
4.6	GPR - one step prediction [zoom 7 days]	43
4.7	data REAL vs RESAMPLED [zoom 30 days]	44
4.8	LR - one step prediction $t_r = 4$ h [testing period]	45
4.9	LR - one step prediction $t_r = 4$ h [zoom 7 days]	45
4.10	LR - one step prediction $t_r = 4$ h [testing period]	46
4.11	GPR - one step prediction $t_r = 4$ h [zoom 7 days]	46

4.12	data REAL vs RESAMPLED [zoom 30 days]	47
4.13	LR - one step prediction $t_r = 8$ h [testing period]	47
4.14	LR - one step prediction $t_r = 8$ h [zoom 7 days]	48
4.15	LR - one step prediction $t_r = 8$ h [testing period]	48
4.16	GPR - one step prediction $t_r = 8$ h [zoom 7 days]	49
4.17	data REAL vs RESAMPLED [zoom 30 days]	50
4.18	LR - one step prediction $t_r = 12$ h [testing period]	50
4.19	LR - one step prediction $t_r = 12$ h [zoom 15 days]	51
4.20	LR - one step prediction $t_r = 12$ h [testing period]	51
4.21	GPR - one step prediction $t_r = 12$ h [zoom 15 days]	52
4.22	data REAL vs RESAMPLED [zoom 30 days]	53
4.23	LR - one step prediction $t_r = 24$ h [testing period]	53
4.24	LR - one step prediction $t_r = 24$ h [zoom 30 days]	54
4.25	LR - one step prediction $t_r = 24$ h [testing period]	54
4.26	GPR - one step prediction $t_r = 24$ h [zoom 30 days]	55
4.27	data REAL vs RESAMPLED [zoom 30 days]	56
4.28	LR - one step prediction $t_r = 48$ h [testing period]	56
4.29	LR - one step prediction $t_r = 48$ h [zoom 30 days]	57
4.30	LR - one step prediction $t_r = 48$ h [testing period]	57
4.31	GPR - one step prediction $t_r = 48$ h [zoom 30 days]	58
4.32	LR - multi step prediction [testing period]	60
4.33	LR - multi step prediction [zoom 7 days]	60
4.34	GPR - multi step prediction [testing period]	61
4.35	GPR - multi step prediction [zoom 7 days]	61
4.36	data REAL vs RESAMPLED [zoom 30 days]	63
4.37	LR - multi step prediction $t_r = 4$ h [testing period]	63
4.38	LR - multi step prediction $t_r = 4$ h [zoom 7 days]	64
4.39	GPR - multi step prediction $t_r = 4$ h [testing period]	65
4.40	GPR - mutli step prediction $t_r = 4$ h [zoom 7 days]	65
4.41	data REAL vs RESAMPLED [zoom 30 days]	66
4.42	LR - multi step prediction $t_r = 6$ h [testing period]	66
4.43	LR - multi step prediction $t_r = 6$ h [zoom 7 days]	67
4.44	GPR - multi step prediction $t_r = 6$ h [testing period]	68
4.45	GPR - mutli step prediction $t_r = 6$ h [zoom 7 days]	68
4.46	data REAL vs RESAMPLED [zoom 30 days]	69
4.47	LR - multi step prediction $t_r = 8$ h [testing period]	69
4.48	LR - multi step prediction $t_r = 8$ h [zoom 7 days]	70
4.49	GPR - multi step prediction $t_r = 8$ h [testing period]	71
4.50	GPR - mutli step prediction $t_r = 8$ h [zoom 7 days]	71
4.51	LR - multi step prediction $n = 3, 4, 6, 8$ [testing period]	73
4.52	LR - multi step prediction $n = 3, 4, 6, 8$ [zoom 7 days]	73

4.53 GPR - multi step prediction $n = 3, 4, 6, 8$ [testing period]	74
4.54 GPR - multi step prediction $n = 3, 4, 6, 8$ [zoom 7 days]	74

Capitolo 1

Introduzione

1.1 **Contesto energetico nazionale e introduzione metodi di previsione dei consumi in ambito energetico**

La transizione energetica in corso a livello nazionale e internazionale è caratterizzata da una crescente penetrazione delle fonti rinnovabili, dall'evoluzione dei mercati dell'energia e dalla necessità di garantire sicurezza e sostenibilità del sistema. In questo scenario, la capacità di gestire in modo efficiente i flussi energetici, sia dal lato della produzione sia dal lato dei consumi, riveste un ruolo strategico. A tal fine, i modelli di previsione dei consumi energetici rappresentano strumenti indispensabili per anticipare l'andamento della domanda e della produzione, ridurre l'incertezza e supportare decisioni operative e strategiche.

I modelli previsionali sono basati principalmente sull'analisi di serie storiche di consumo o di produzione energetica, che vengono elaborate per identificare andamenti ricorrenti, ciclicità stagionali e tendenze di lungo periodo. A questo nucleo di dati possono essere integrate variabili esogene, come ad esempio le condizioni meteorologiche (fattore decisivo per la produzione da fonti rinnovabili e per la variazione della domanda elettrica stagionale), oppure indicatori socioeconomici, quali crescita del PIL, variazioni demografiche, evoluzione dei prezzi energetici e abitudini di consumo. L'inclusione di queste informazioni consente di arricchire la capacità del modello di catturare la complessità dei fenomeni reali, superando i limiti di una semplice estrapolazione statistica.

Attraverso queste tecniche è possibile stimare in maniera più accurata l'andamento futuro della domanda o della generazione di energia, fornendo un supporto essenziale a diversi livelli decisionali. Le previsioni diventano infatti strumenti

operativi per la pianificazione delle politiche energetiche, per la gestione quotidiana delle reti e dei sistemi di distribuzione, per l'ottimizzazione dell'uso delle risorse e per la definizione di strategie di investimento a medio e lungo termine. In questo senso, la previsione non rappresenta un fine in sé, ma costituisce un mezzo fondamentale per accrescere l'affidabilità, l'efficienza e la sostenibilità della gestione energetica complessiva. Grazie ad esse, operatori, istituzioni e imprese possono ridurre l'incertezza, anticipare criticità e orientare le proprie decisioni verso soluzioni più resilienti e compatibili con gli obiettivi di transizione energetica.

L'impiego dei modelli previsionali in ambito energetico può avere applicazioni eterogenee [1]. A livello nazionale, vengono utilizzati per stimare la domanda elettrica complessiva e supportare decisioni di lungo periodo relative alla pianificazione della capacità produttiva e delle infrastrutture ([2, 3, 4, 5, 6]). In ambito urbano o nelle smart grid, la previsione dei carichi è funzionale alla gestione in tempo reale della rete e alla riduzione degli sbilanciamenti locali ([7, 8, 9]. Nel settore residenziale e degli edifici intelligenti, i modelli sono applicati per ottimizzare i consumi e abilitare programmi di demand response, che favoriscono una partecipazione attiva degli utenti finali ([10, 11, 12, 13]). Anche le fonti rinnovabili non programmabili come il fotovoltaico e l'eolico richiedono strumenti previsionali per ridurre l'incertezza nella produzione, favorire l'integrazione in rete e pianificare le riserve necessarie ([7, 14, 15, 16]). Un ulteriore ambito di applicazione, particolarmente rilevante, è quello industriale, dove la previsione dei consumi energetici consente una migliore programmazione delle attività produttive, l'ottimizzazione dei costi e l'incremento dell'efficienza complessiva dei processi.

I modelli si differenziano anche in funzione dell'orizzonte temporale su cui operano. È possibile distinguere tra breve, medio e lungo termine. Le previsioni di lungo periodo (oltre un anno) sono principalmente utilizzate per la definizione delle politiche energetiche e per gli investimenti infrastrutturali. Quelle di medio termine (settimane o mesi) servono alla pianificazione tattica, ad esempio per la definizione delle tariffe o la programmazione stagionale dei consumi ([17, 18, 19, 20]). Infine, le previsioni di breve termine (da pochi minuti a pochi giorni) assumono un ruolo centrale nelle attività operative: esse sono impiegate per il bilanciamento in tempo reale della rete, per la programmazione giornaliera dei mercati elettrici, per la gestione del dispacciamento di energia da fonti convenzionali e rinnovabili, e per l'attuazione di strategie di demand response. Una previsione accurata a breve termine consente di ridurre i costi legati agli sbilanciamenti, di garantire l'affidabilità del sistema e di migliorare la flessibilità operativa degli attori coinvolti ([21, 14, 15, 22]).

Accanto alle diverse applicazioni e agli orizzonti temporali, i modelli previsionali possono essere realizzati con approcci metodologici differenti. Alcuni sono di tipo

statistico, come i modelli basati su serie storiche, che risultano particolarmente adatti quando si hanno dati regolari e con forte stagionalità. Altri approcci combinano metodi diversi, dando origine a modelli ibridi, che mirano a sfruttare i punti di forza delle singole tecniche. Infine, si sono diffusi modelli basati su tecniche di apprendimento automatico, come le reti neurali, capaci di cogliere relazioni complesse e non lineari nei dati. Ogni approccio presenta specifiche caratteristiche, vantaggi e limiti, e viene scelto in funzione del contesto applicativo, della disponibilità di dati e degli obiettivi della previsione.

1.2 Obiettivo e finalità della tesi

Il presente lavoro di tesi si propone di analizzare e sviluppare un percorso volto al miglioramento della gestione energetica presso l'azienda *EthosEnergy*. In una prima fase è stata condotta un'analisi dei consumi e una raccolta strutturata dei dati aziendali, con l'obiettivo di definire una base informativa solida e affidabile. Successivamente, è stato realizzato un modello di previsione dei consumi energetici, articolato secondo due approcci distinti: uno di tipo statistico e uno che sfrutta tecniche di ML, al fine di valutare e confrontare le diverse potenzialità dei metodi considerati. L'attività si è poi concentrata sullo sviluppo di un modello di *forecasting* dei consumi, finalizzato a stimarne l'andamento nel breve periodo (*short-term forecasting*).

L'utilizzo di tecniche di previsione e l'analisi consapevole delle serie temporali di consumo consentono di supportare la pianificazione e l'ottimizzazione dei processi produttivi dell'azienda, ridurre i costi legati all'energia e migliorare la sostenibilità complessiva delle attività industriali. In questa prospettiva, il forecasting rappresenta non soltanto uno strumento operativo per la gestione quotidiana, ma anche una base di partenza per approfondire le potenzialità applicative di modelli previsionali orientati al miglioramento dell'efficienza energetica e alla valorizzazione dei dati aziendali.

1.3 Struttura della tesi

La tesi si articola in sei capitoli principali, oltre a bibliografia e appendici. Nel primo capitolo viene presentato il quadro energetico di riferimento, con un'introduzione ai metodi di previsione dei consumi e una descrizione degli obiettivi che guidano il lavoro. Il secondo capitolo è invece dedicato all'azienda oggetto di studio, di cui vengono illustrate le caratteristiche produttive e organizzative, la localizzazione e i relativi fattori climatici, gli impianti presenti e l'attuale profilo energetico, corredata da un'analisi dettagliata dei dati di consumo e degli indicatori di efficienza.

Il terzo capitolo introduce il modello di previsione sviluppato e i metodi utilizzati, descrivendo gli approcci scelti, le tecniche implementate e i criteri di valutazione adottati per verificarne l'affidabilità. A seguire, il quarto capitolo presenta l'applicazione del modello ai dati reali dell'azienda EthosEnergy, illustrando i risultati delle previsioni e approfondendo l'effetto delle diverse strategie di elaborazione e confronto delle metriche.

Il quinto capitolo raccoglie l'analisi critica dei risultati ottenuti, mettendo in evidenza i punti di forza e i limiti dei metodi considerati e proponendo possibili interventi volti a migliorare l'efficienza energetica aziendale. Infine, il sesto capitolo sintetizza le conclusioni del lavoro, proponendo una riflessione complessiva e delineando prospettive di sviluppo futuro.

Capitolo 2

Descrizione dell'azienda oggetto di studio

2.1 Presentazione contesto aziendale

EthosEnergy, il cui logo è rappresentato in figura 2.1, è una realtà multinazionale specializzata nella fornitura di soluzioni e servizi nel settore energetico, con un focus particolare sul miglioramento delle performance, sull'efficienza e sulla sostenibilità degli impianti industriali. L'azienda opera a livello globale, offrendo supporto a clienti che spaziano dalla produzione di energia elettrica fino ai settori dell'oil & gas, con un ampio ventaglio di attività che comprendono la manutenzione, la riparazione e l'ottimizzazione di turbine, compressori e altri macchinari critici. Grazie a un know-how consolidato e alla presenza in oltre 100 Paesi, EthosEnergy può contare su numerose sedi operative e controllate distribuite a livello internazionale, che le consentono di integrare tecnologie avanzate con un approccio orientato al cliente, garantendo soluzioni su misura e un'elevata affidabilità operativa.



Figura 2.1: Logo azienda EthosEnergy

All'interno di questo contesto internazionale, un ruolo rilevante è svolto dalla sede di Torino, che rappresenta un polo strategico per il mercato italiano ed europeo. La filiale torinese si concentra in particolare sul monitoraggio, la manutenzione e l'ottimizzazione delle turbine a gas e a vapore, offrendo servizi sia on-site sia presso i propri laboratori specializzati. Le competenze sviluppate in questa sede riguardano non solo la manutenzione ordinaria e straordinaria, ma anche l'implementazione di soluzioni innovative per migliorare l'efficienza energetica e ridurre l'impatto ambientale degli impianti.

La sede EthosEnergy di Torino (figura 2.2) si trova in Corso Romania 661, 10156 Torino (TO), Italia, all'interno di un'area industriale caratterizzata dalla presenza di numerose realtà del settore manifatturiero ed energetico. La posizione strategica, facilmente accessibile sia dal centro cittadino sia dalle principali arterie autostradali, consente all'azienda di mantenere stretti contatti con clienti, fornitori e partner. Le coordinate geografiche del sito sono 45.1037° N, 7.7224° E.



Figura 2.2: Sede EthosEnergy Torino

Dal punto di vista strutturale e logistico, il sito torinese si estende su una superficie complessiva di circa 100.000 m², articolata in più fabbricati con funzioni produttive, amministrative e di servizio. L'edificio principale, che copre un'area di circa 35.000 m², ospita i reparti produttivi, gli uffici e vari locali tecnici, organizzati in campate.

Accanto all'edificio principale, sono presenti due fabbricati adibiti a magazzini, un locale caldaie e compressori, oltre ad aree destinate a mensa, punti di consegna dell'energia elettrica e del gas metano, parcheggi e zone verdi. Gli uffici, distribuiti su tre piani, comprendono anche diverse sale riunioni, a supporto delle attività organizzative e gestionali della sede.

2.2 Attività produttiva e organizzazione degli spazi

Lo stabilimento EthosEnergy di Torino si caratterizza per una chiara suddivisione funzionale degli spazi, che consente di integrare le diverse attività produttive, logistiche e amministrative in modo efficiente. La struttura comprende infatti uffici amministrativi, una vasta area produttiva, diversi magazzini, una mensa aziendale e un ampio cortile interno con parcheggi.

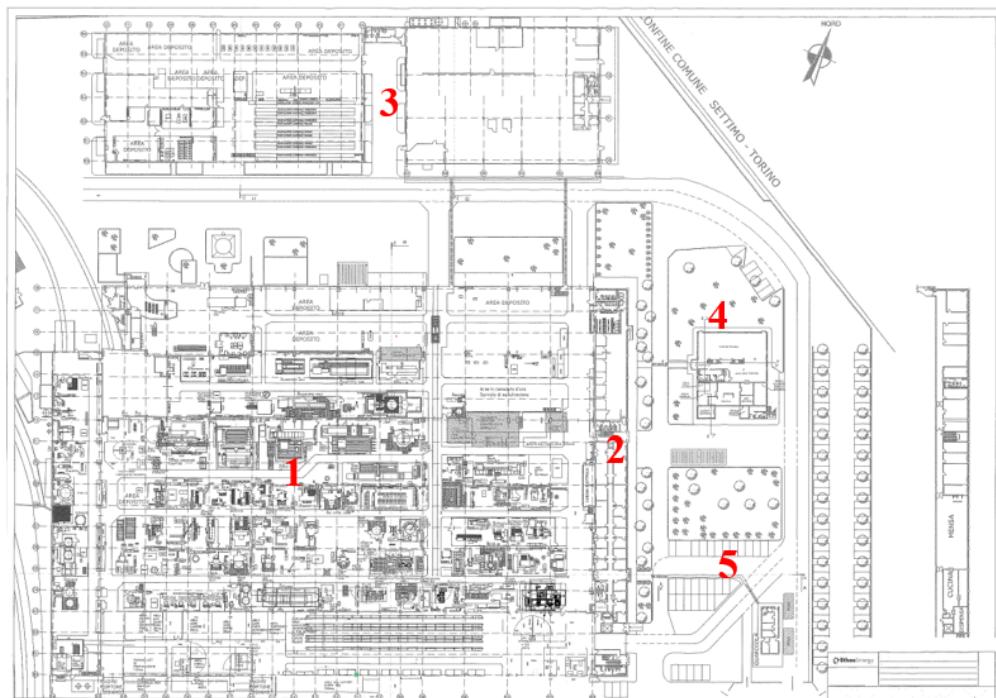


Figura 2.3: Pianta di stabilimento

N.	SPAZI STABILIMENTO
1	Officina
2	Uffici
3	Magazzino
4	Mensa
5	Cortile e parcheggi

Tabella 2.1: Spazi stabilimento

L'area produttiva costituisce il cuore delle attività dello stabilimento ed è organizzata in sette campate, identificate dalle sigle R, 1, 2, 3, 4, 5 e L. Ogni campata è attrezzata con due linee di blindo, ovvero condotti blindati per la distribuzione elettrica che corrono lungo i lati e permettono un'alimentazione sicura e modulare dei macchinari, garantendo ordine e flessibilità nell'organizzazione dell'officina. A ciò si aggiungono un sistema di illuminazione a soffitto e uno o più carriporti che consentono la movimentazione dei carichi pesanti. Al suo interno si svolgono le principali attività produttive della sede torinese, che comprendono la manutenzione, riparazione, revisione e rigenerazione di componenti critici di turbine a gas e a vapore. I magazzini supportano direttamente l'officina e sono destinati allo stoccaggio di ricambi, materie prime e utensili, oltre che alle attività di ricezione e spedizione dei materiali.

Gli uffici, distribuiti su tre piani, ospitano funzioni di ingegneria, pianificazione e gestione delle commesse, mantenendo un legame costante con le attività produttive. La presenza di sale riunioni e locali tecnici consente inoltre un coordinamento immediato tra il personale di officina e quello amministrativo. La mensa aziendale e i servizi connessi – come locali tecnici, caldaie e compressori – completano l'organizzazione degli spazi, contribuendo a rendere lo stabilimento un polo produttivo e ingegneristico pienamente integrato nella rete globale di EthosEnergy.

2.3 Descrizione impianti esistenti

Nello stabilimento sono presenti diversi impianti che garantiscono l'alimentazione energetica e il corretto svolgimento delle attività produttive e di servizio. Di seguito se ne riportano le principali caratteristiche.

Per quanto riguarda l'impianto elettrico, lo stabilimento è servito da un unico punto di allacciamento (POD) trifase in media tensione con la rete elettrica nazionale. L'energia così fornita viene successivamente distribuita attraverso tre cabine di trasformazione interne, che provvedono a convertirla in bassa tensione per renderla disponibile alle diverse utenze. La prima cabina è collegata agli uffici, a una parte dell'area produttiva e alla mensa aziendale, assicurando continuità energetica agli spazi destinati sia al lavoro amministrativo sia ai servizi collettivi; essa copre circa il 65% del consumo totale registrato nel 2025. La seconda cabina alimenta invece la campata "L" dell'area produttiva, i magazzini, il sistema di aria compressa, la guardiola e il parcheggio, integrando dunque sia le esigenze operative della produzione sia le funzioni di supporto logistico e di accesso; i suoi utilizzi corrispondono al restante 35% dei consumi complessivi. La terza cabina, predisposta per l'alimentazione di una parte del magazzino, risulta attualmente disattivata e non contribuisce quindi alla distribuzione interna di energia elettrica.

L'impianto termico dello stabilimento è costituito da generatori di vapore che assolvono un ruolo essenziale, seppur circoscritto. Essi sono infatti impiegati esclusivamente per il riscaldamento degli ambienti e per la produzione di acqua calda sanitaria (ACS), senza interventi diretti nei processi produttivi. L'alimentazione avviene mediante gas naturale, che rappresenta l'unica fonte primaria per questo impianto e garantisce un approvvigionamento affidabile e costante.

Accanto a questi sistemi principali, lo stabilimento dispone di ulteriori impianti complementari che concorrono al corretto funzionamento delle strutture. Tra essi rientrano l'impianto dell'aria compressa, fondamentale per alimentare le macchine e gli strumenti che richiedono questo tipo di servizio, l'impianto di climatizzazione degli ambienti (UTA), destinato a regolare le condizioni termoigometriche interne, e infine l'impianto di illuminazione, che assicura adeguati livelli di comfort e sicurezza in tutte le aree di lavoro e di servizio.

2.4 Profilo energetico attuale e passato dell'azienda

2.4.1 Raccolta e analisi dei dati di consumo

L'analisi dei dati di consumo rappresenta un passaggio imprescindibile per comprendere l'andamento dei fabbisogni di uno stabilimento e per interpretare in modo critico le scelte gestionali e operative che influenzano tali andamenti. I dati raccolti consentono di mettere in evidenza non solo le variazioni stagionali, ma anche le differenze tra gli anni presi in considerazione, offrendo così un quadro completo utile alle fasi successive di modellazione e previsione. In questa sezione vengono presentati i risultati relativi ai consumi elettrici e di gas naturale, con un'attenzione particolare al 2024 e con un confronto diretto rispetto al biennio precedente; a seguire, sarà inoltre riportata la raccolta e l'analisi dei dati di consumo idrico, sia potabile sia industriale, così da fornire una visione integrata e complessiva dei fabbisogni energetici e idrici dello stabilimento.

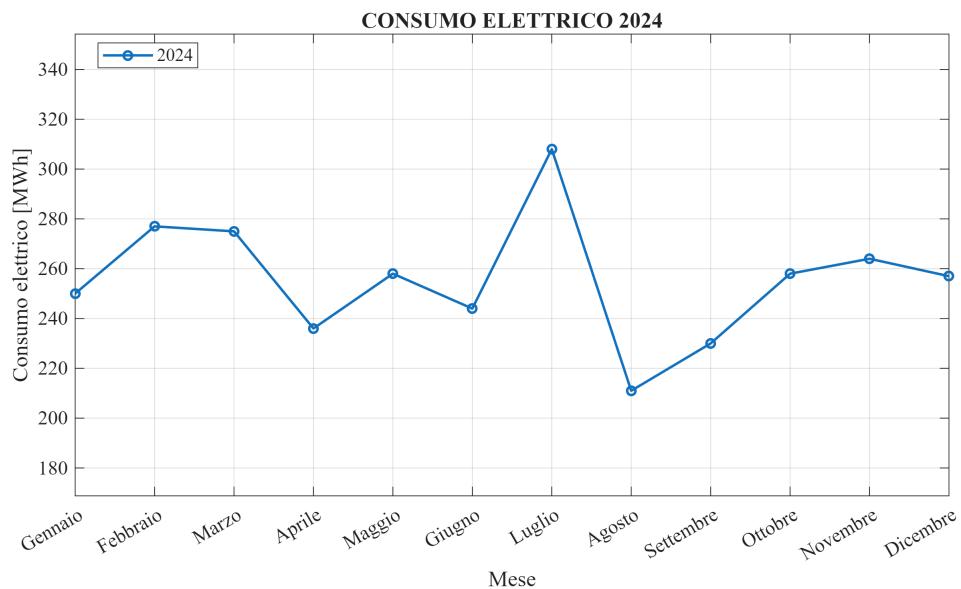


Figura 2.4: Consumo elettrico 2024

Partendo dal consumo elettrico relativo al 2024, riportato in 2.4, si osserva un andamento caratterizzato da una certa variabilità mensile, con valori compresi tra circa 210 e poco più di 300 MWh. Nei primi mesi dell'anno i consumi si collocano su livelli medio-alti, raggiungendo a febbraio un picco intorno ai 278

MWh. Successivamente si registra un calo significativo ad aprile, quando i valori scendono a circa 236 MWh, segno di una riduzione delle attività produttive o di una fase di fermo programmato. Nei mesi estivi l'andamento mostra una ripresa evidente, con un massimo a luglio che supera i 300 MWh, probabilmente legato a un incremento della domanda produttiva e al maggior utilizzo degli impianti di climatizzazione. Agosto rappresenta un'eccezione, con un crollo dei consumi fino a 210 MWh, coerente con la chiusura estiva delle attività. Infine, negli ultimi mesi dell'anno si osserva un progressivo rialzo che porta i consumi a stabilizzarsi tra 258 e 265 MWh. Nel complesso, l'andamento del 2024 riflette un ciclo stagionale abbastanza marcato, con punte estive e cali nei periodi di inattività, suggerendo una forte influenza sia di fattori produttivi sia di elementi climatici.

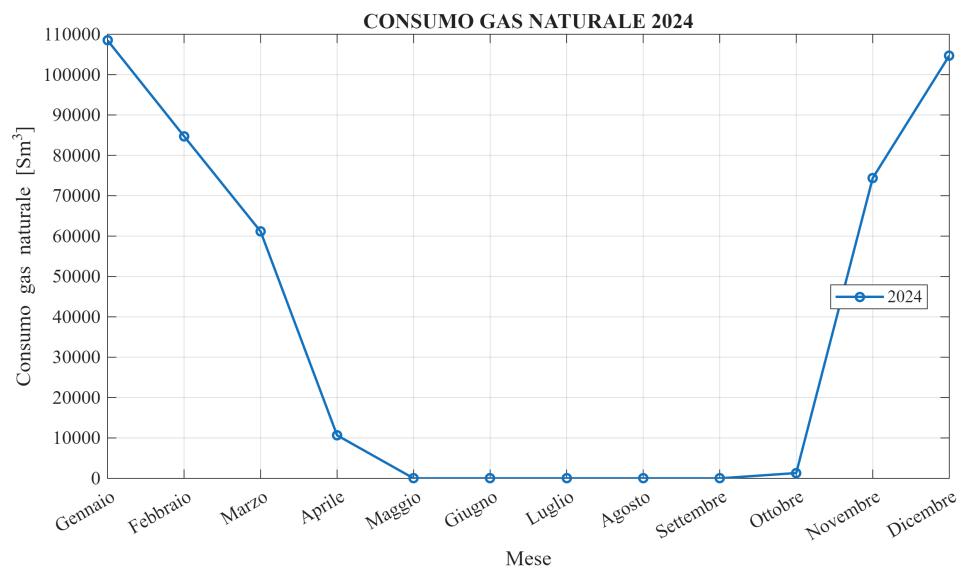


Figura 2.5: Consumo gas naturale 2024

Il consumo di gas naturale nello stesso anno, rappresentato in 2.5, evidenzia invece un comportamento radicalmente diverso, segnato da una netta stagionalità. I consumi sono infatti concentrati nei mesi invernali, con valori che sfiorano i 110.000 Sm³ a gennaio e dicembre, per poi calare bruscamente fino quasi ad annullarsi nel periodo estivo. Da maggio a settembre si osservano infatti valori prossimi allo zero, con una ripresa graduale a partire da ottobre e un incremento deciso tra novembre e dicembre. Questo andamento conferma l'utilizzo del gas principalmente per finalità di riscaldamento ambientale e, in misura minore, per la produzione di calore di processo. L'andamento così regolare rende i consumi di gas relativamente più prevedibili rispetto a quelli elettrici, poiché legati in misura predominante all'andamento climatico e alle esigenze stagionali.

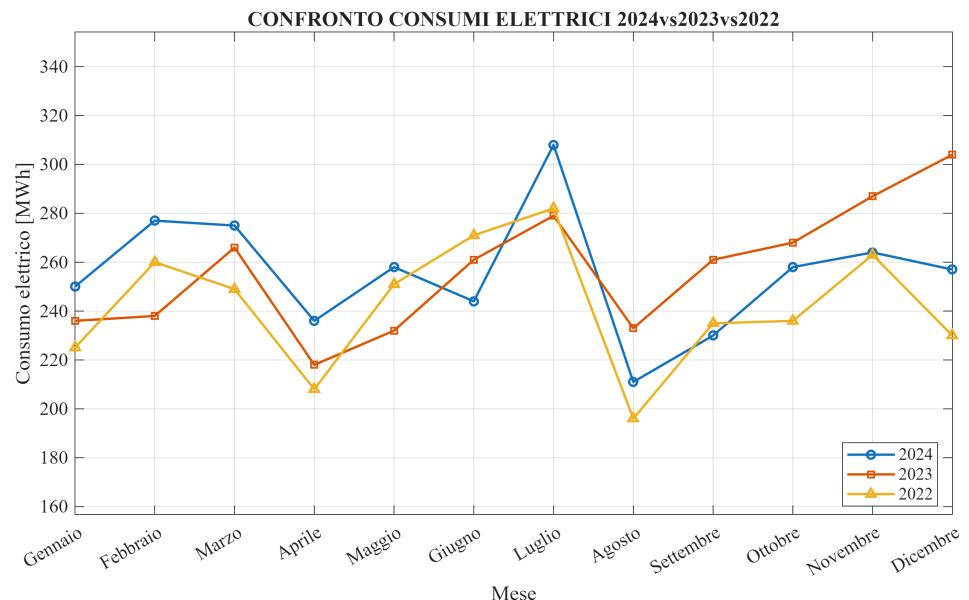


Figura 2.6: Confronto consumi elettrici 2024vs2023vs2022

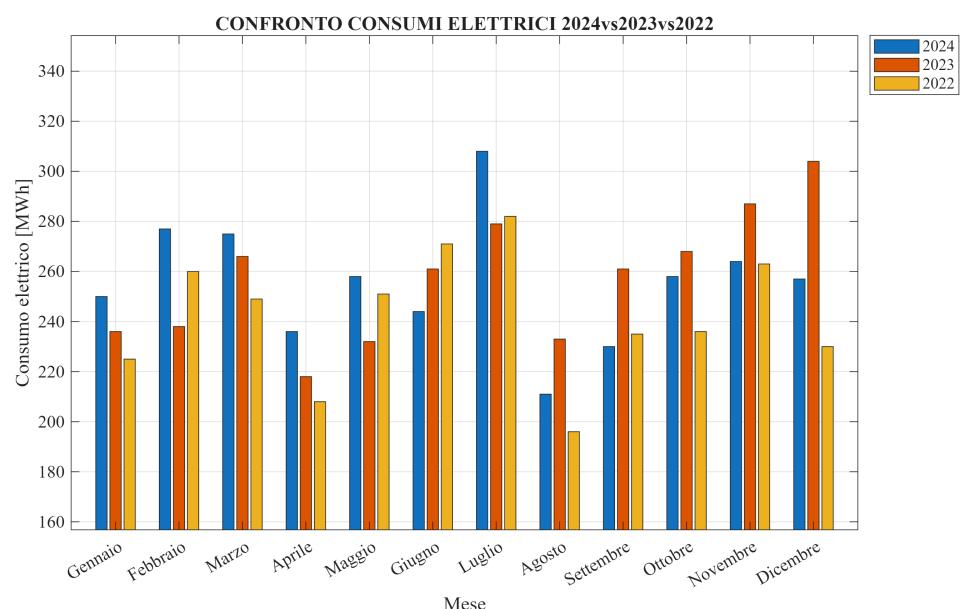


Figura 2.7: Confronto consumi elettrici 2024vs2023vs2022

Per approfondire ulteriormente le dinamiche di consumo, si è condotto un confronto triennale, analizzando il periodo 2022–2024. In 2.6 è riportato l'andamento dei consumi elettrici sotto forma di grafico a linee, che permette di mettere a confronto in modo diretto le curve dei tre anni. Si osserva come la stagionalità sia un elemento comune: tutti gli anni presentano un calo ad agosto e valori più elevati in corrispondenza dei mesi primaverili ed estivi.

Tuttavia, emergono differenze significative nei livelli assoluti. Il 2024 tende a collocarsi su valori leggermente superiori nei primi mesi e in luglio, segnalando una crescita della domanda energetica. Il 2023 mostra invece consumi più elevati nell'ultima parte dell'anno, con un picco che supera i 300 MWh a dicembre, mentre il 2022 si distingue per valori più contenuti, collocandosi quasi sempre al di sotto degli altri due anni. L'osservazione dei grafici a barre, riportati in 2.7, consente di rendere immediatamente visibili tali differenze mese per mese: febbraio e marzo 2024, ad esempio, spiccano per consumi più alti rispetto agli anni precedenti, mentre dicembre 2023 emerge come il mese con il valore massimo assoluto del triennio, intorno ai 305 MWh. Nel complesso, il confronto mette in luce un trend di crescita, pur con oscillazioni legate alla stagionalità e a fattori contingenti.

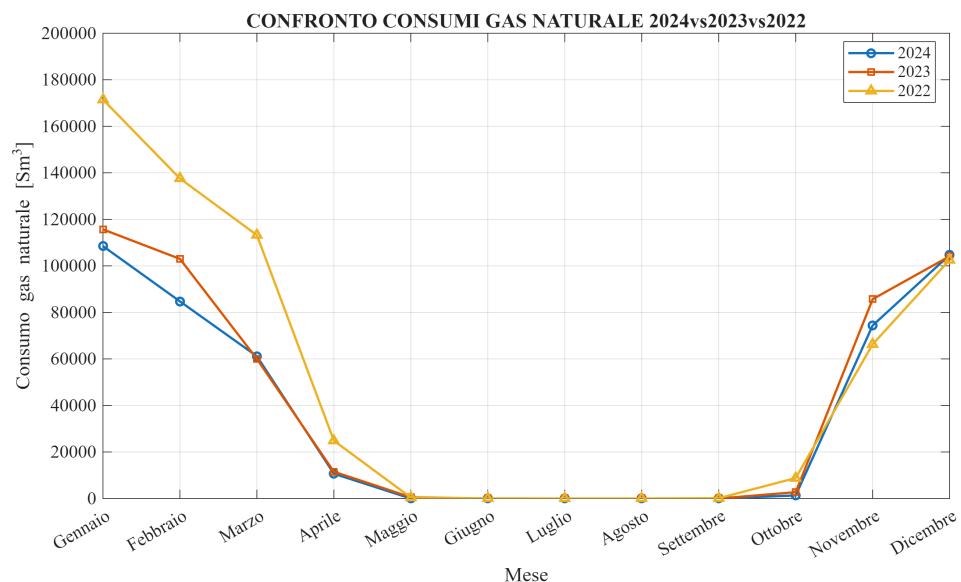


Figura 2.8: Confronto consumo gas naturale 2024vs2023vs2024

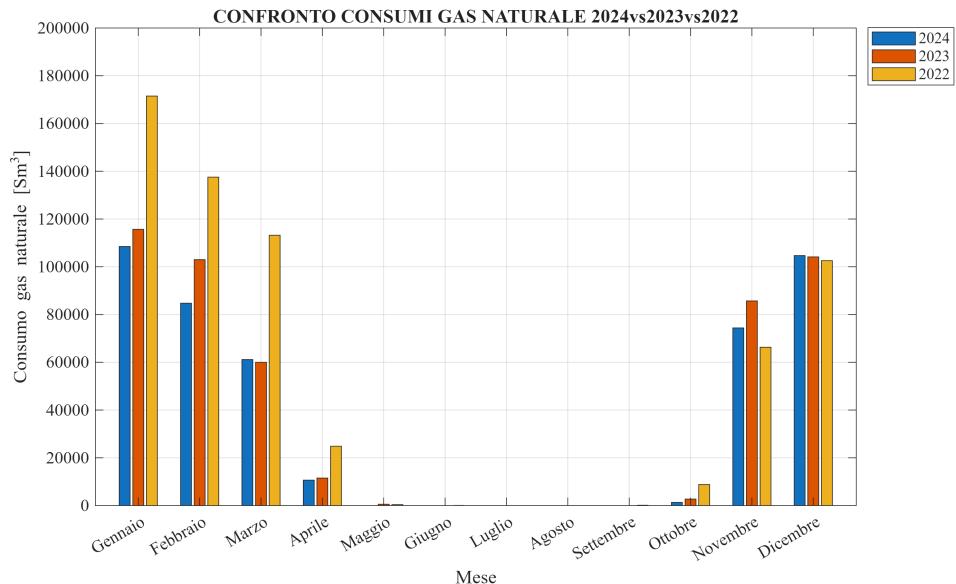


Figura 2.9: Confronto consumo gas naturale 2024vs2023vs2024

Un discorso analogo può essere condotto per i consumi di gas naturale, rappresentati nelle 2.8 e 2.9. Anche in questo caso le curve a linee permettono di evidenziare una stagionalità molto marcata, con consumi concentrati in inverno e pressoché nulli in estate. Ciò che differenzia gli anni è l'intensità dei consumi nei mesi freddi. Il 2022 si distingue come l'anno con i valori più elevati, superando i 170.000 Sm³ a gennaio e mantenendosi su livelli alti fino a marzo. Il 2023 presenta invece valori intermedi, mentre il 2024 appare come l'anno con i consumi più contenuti, soprattutto a inizio anno, quando le differenze con il 2022 risultano particolarmente evidenti. Il grafico a barre conferma questa dinamica, mostrando come i primi mesi del 2022 siano caratterizzati da un consumo nettamente superiore. Questa progressiva riduzione può essere interpretata sia come conseguenza di condizioni climatiche meno rigide negli anni più recenti, sia come risultato di politiche di efficientamento energetico e di un uso più razionale degli impianti. In ogni caso, il trend evidenzia una tendenza positiva verso una maggiore efficienza e una riduzione dei fabbisogni di gas.

L'acqua rappresenta una risorsa strategica per il funzionamento di uno stabilimento industriale, sia come input diretto nei processi produttivi, sia per gli usi accessori che riguardano la vita quotidiana aziendale. In particolare, si distinguono due tipologie di impiego: da un lato l'acqua potabile, utilizzata per usi civili e di servizio, come mense, spogliatoi e uffici; dall'altro l'acqua industriale, proveniente da pozzi o altre fonti dedicate, destinata prevalentemente agli impianti tecnologici, al raffreddamento delle macchine e ad altre attività di supporto alla produzione.

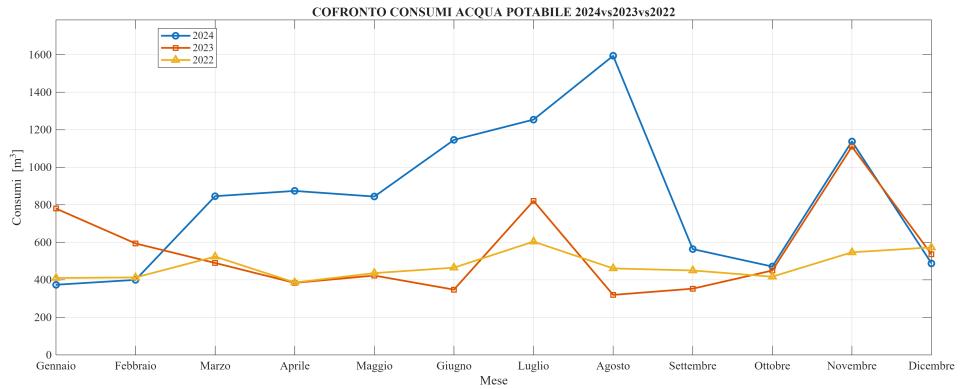


Figura 2.10: Consumo acqua potabile 2024

Il grafico 2.10 mostra l'andamento mensile dei consumi di acqua potabile negli anni 2022, 2023 e 2024. La curva relativa al 2024 si colloca su livelli più elevati rispetto agli altri due anni, con valori che superano in più occasioni i 1.000 m^3 , raggiungendo un massimo ad agosto superiore ai 1.600 m^3 . L'andamento appare caratterizzato da forti oscillazioni, con picchi che si alternano a fasi di riduzione, suggerendo una correlazione con la stagionalità delle attività produttive e con i periodi di maggiore presenza del personale. Al contrario, i consumi del 2022 e del 2023 si mantengono complessivamente più bassi e più regolari, con valori mediamente compresi tra 400 e 800 m^3 al mese. Questo scostamento tra gli anni evidenzia come il 2024 abbia registrato una crescita significativa del fabbisogno di acqua potabile, segnale di un incremento delle attività o di un cambiamento gestionale nella distribuzione delle risorse.

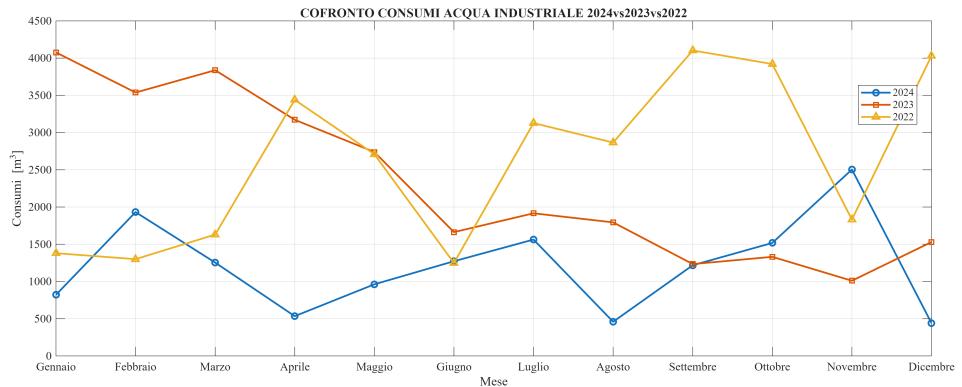


Figura 2.11: Consumo acqua industriale 2024

Il grafico in figura 2.11 riguarda i consumi mensili di acqua industriale. In questo caso, le differenze tra i tre anni risultano ancora più marcate: il 2022 evidenzia

valori molto alti, con punte mensili superiori ai 4.000 m³ e una tendenza generale a mantenersi stabilmente sopra i 2.500 m³ per buona parte dell'anno. Il 2023 presenta invece un andamento intermedio, con consumi compresi tra 1.000 e 4.000 m³, mentre il 2024 mostra i valori più contenuti, scendendo in diversi mesi al di sotto dei 1.000 m³. Tale riduzione suggerisce l'adozione di politiche più restrittive nell'uso dell'acqua industriale o una diminuzione del fabbisogno produttivo, e rappresenta un chiaro indicatore di maggiore efficienza o di un mutato assetto impiantistico. Rispetto all'acqua potabile, il consumo industriale risulta dunque molto più variabile e sensibile a fattori produttivi specifici, con oscillazioni anche significative da un anno all'altro.

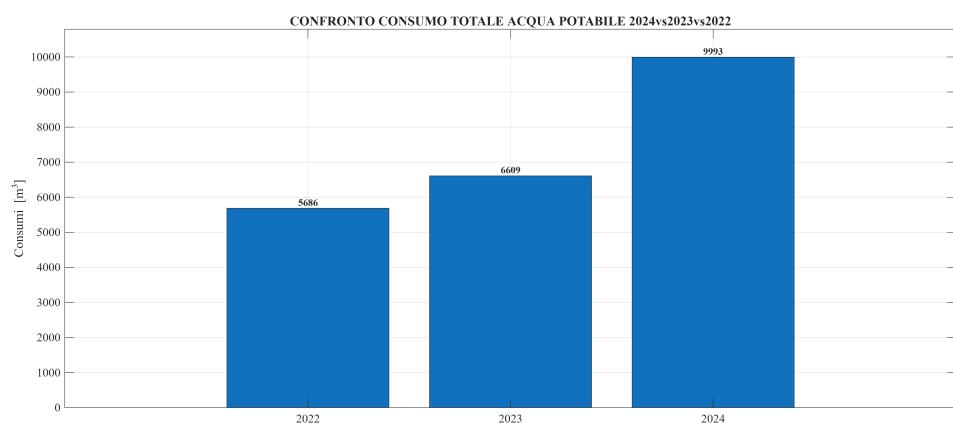


Figura 2.12: Confronto consumo totale acqua potabile 2024vs2023vs2022

Il grafico riportato in figura 2.12 sintetizza i consumi totali annui di acqua potabile nel triennio considerato. I valori mostrano una progressione crescente: si passa dai 5.686 m³ del 2022 ai 6.609 m³ del 2023, fino a raggiungere i 9.993 m³ nel 2024. Questo incremento costante conferma quanto osservato nei dati mensili e segnala un fabbisogno idrico crescente negli usi civili e di servizio dello stabilimento. L'aumento del 2024, in particolare, rappresenta un salto significativo rispetto agli anni precedenti, che merita un approfondimento sulle cause, potenzialmente legate sia a fattori organizzativi interni sia a un maggior numero di addetti presenti nello stabilimento.

Infine, il grafico in 2.13 riporta il confronto dei consumi annui di acqua industriale. In questo caso, il trend è opposto rispetto a quello della potabile: dai 31.588 m³ del 2022 si passa a 27.842 m³ nel 2023, per poi scendere ulteriormente a 14.477 m³ nel 2024. La riduzione è particolarmente evidente nell'ultimo anno, quando i consumi si dimezzano rispetto a quelli iniziali. Tale andamento può essere interpretato come il risultato di una razionalizzazione dei processi produttivi o dell'introduzione di

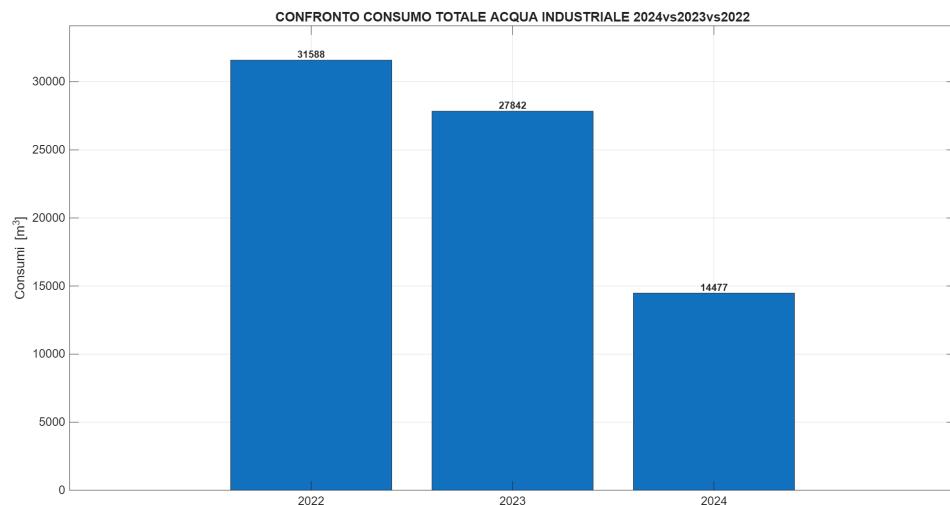


Figura 2.13: Confronto consumo totale acqua industriale 2024vs2023vs2022

pratiche di ricircolo e recupero dell'acqua, con un impatto diretto sul contenimento dei prelievi da pozzo. L'analisi mette quindi in luce un'evoluzione positiva dal punto di vista della sostenibilità ambientale, poiché il calo dei consumi industriali va di pari passo con un utilizzo più consapevole e mirato della risorsa idrica.

In sintesi, i dati analizzati mostrano tre comportamenti distinti: i consumi elettrici risultano distribuiti lungo l'intero anno, con oscillazioni legate sia all'attività produttiva sia alla stagionalità climatica; i consumi di gas sono invece fortemente concentrati nei mesi invernali e mostrano una tendenza alla riduzione nel corso del triennio; infine, i consumi idrici evidenziano dinamiche differenziate tra l'acqua potabile, caratterizzata da un andamento in crescita, e l'acqua industriale, che al contrario registra una progressiva contrazione.

2.4.2 Indicatori energetici

Il TEP (Tonnellata Equivalente di Petrolio) è un'unità di misura convenzionale utilizzata per esprimere i consumi energetici in modo omogeneo, indipendentemente dal tipo di fonte impiegata (elettricità, gas naturale, gasolio, ecc.). Un tep corrisponde alla quantità di energia rilasciata dalla combustione di una tonnellata di petrolio grezzo, pari a circa 41,868 GJ o 11.630 kWh. L'utilizzo dei tep consente di confrontare tra loro differenti vettori energetici, fornendo una visione integrata dei consumi complessivi di uno stabilimento. Ciò è particolarmente utile in contesti industriali complessi, dove coesistono più fonti di energia con caratteristiche fisiche e finalità differenti. Nel caso in esame, i consumi mensili del 2024 sono stati suddivisi in energia elettrica e gas naturale, e successivamente aggregati per ottenere il consumo totale espresso in tep.

Il grafico dei consumi elettrici mensili (figura 2.14) mostra un andamento piuttosto regolare lungo l'anno, con valori mediamente compresi tra 20 e 27 tep. Si osservano lievi incrementi nei mesi estivi, in particolare a luglio, probabilmente correlati a un maggiore utilizzo degli impianti di condizionamento o a un picco produttivo stagionale.

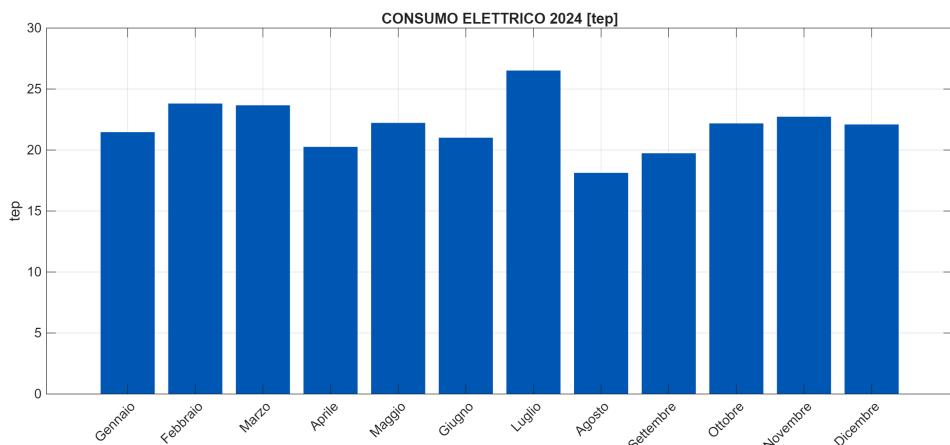


Figura 2.14: Consumi tep elettrici 2024

Per quanto riguarda il gas naturale, l'andamento riportato in 2.15 è invece fortemente stagionale. I consumi risultano elevati nei mesi invernali (gennaio–marzo e novembre–dicembre), raggiungendo punte vicine a 90 tep, mentre nei mesi estivi tendono quasi ad annullarsi. Ciò riflette chiaramente l'utilizzo del gas per la produzione di calore a servizio dei processi e del riscaldamento ambientale.

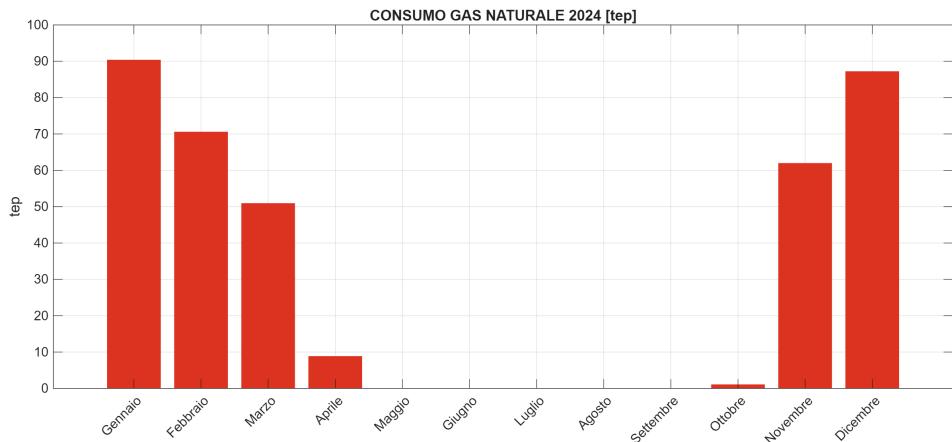


Figura 2.15: Consumi tep termici 2024

Il grafico del consumo totale evidenzia la combinazione dei due vettori. Si nota che l'energia termica (gas) domina nei mesi freddi, mentre in estate la componente elettrica rappresenta quasi l'unico contributo. L'andamento complessivo conferma quindi un profilo stagionale tipico degli stabilimenti industriali con attività termiche concentrate nei mesi invernali.

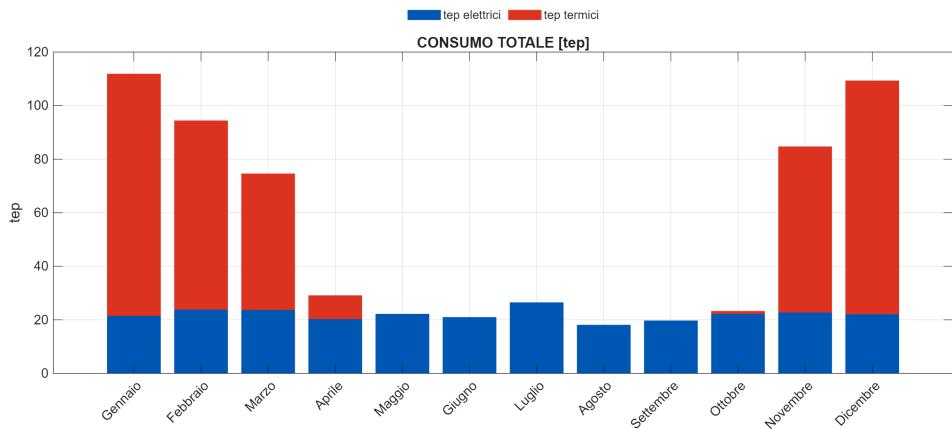


Figura 2.16: Consumo totale tep 2024

Le tonnellate di anidride carbonica equivalente (tCO₂) rappresentano l'indicatore di riferimento per la quantificazione delle emissioni climalteranti associate ai consumi energetici. Questo parametro tiene conto della quantità di CO₂ rilasciata in atmosfera in seguito alla combustione di combustibili fossili o alla produzione elettrica da fonti non rinnovabili. Monitorare le tCO₂ è fondamentale per le aziende, sia in ottica di sostenibilità ambientale che di conformità normativa. La riduzione delle emissioni rappresenta infatti un obiettivo prioritario nel quadro

della transizione energetica e della *Corporate Carbon Accounting*, strumenti chiave per misurare e migliorare le prestazioni ambientali delle organizzazioni.

Il grafico delle emissioni elettriche (figura 2.17) mostra un andamento regolare, simile a quello dei consumi in tep. Le variazioni mensili sono contenute, con valori medi intorno a 60–70 tCO₂. Le fluttuazioni possono essere attribuite sia alla stagionalità della produzione sia a eventuali variazioni nell'intensità emissiva della rete elettrica.

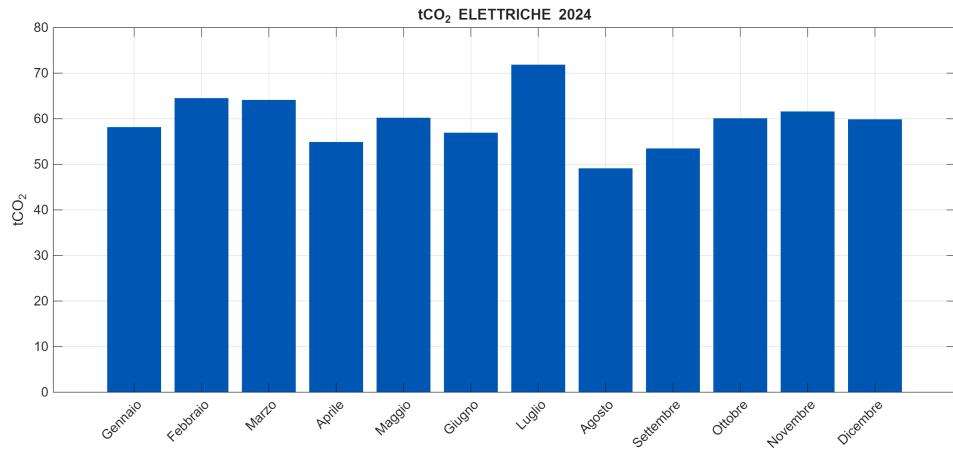


Figura 2.17: Produzione tCO₂ da fonti elettriche 2024

Per le emissioni termiche, invece, si osservano picchi marcati nei mesi invernali, coerenti con i periodi di maggiore utilizzo del gas naturale. I valori massimi superano le 200 tCO₂ a gennaio e dicembre, mentre nei mesi centrali si riducono quasi a zero.

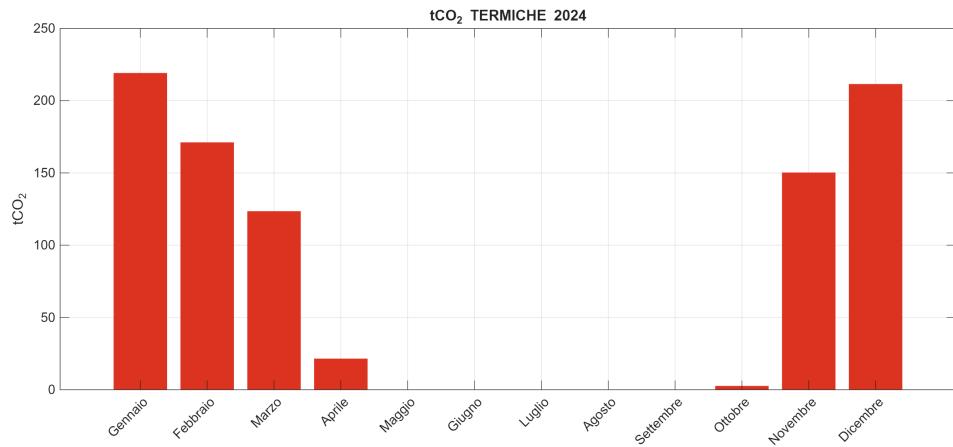


Figura 2.18: Produzione tCO₂ da fonti termiche 2024

Infine, il grafico delle emissioni totali in figura 2.19 mette in evidenza l'andamento complessivo, con una forte incidenza della componente termica durante l'inverno. Questo conferma come gli interventi di efficientamento energetico o di decarbonizzazione debbano prioritariamente concentrarsi sugli usi termici, più impattanti in termini di emissioni.

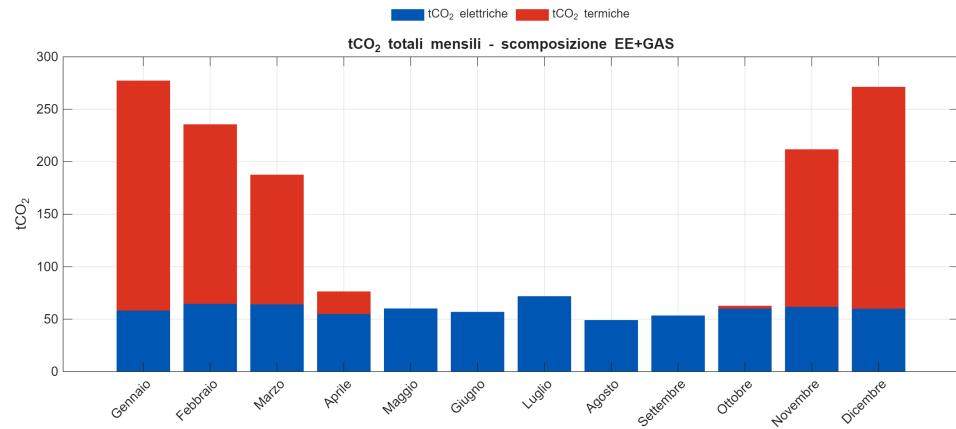


Figura 2.19: Produzione totale di tCO₂ 2024

2.5 Utilizzo vettori energetici

All'interno dello stabilimento i diversi vettori energetici come energia elettrica, gas naturale e acqua rivestono ruoli distinti ma complementari, contribuendo insieme al corretto funzionamento dei processi produttivi e dei servizi di supporto.

L'energia elettrica costituisce il cuore pulsante delle attività industriali, alimentando innanzitutto i macchinari di produzione e i sistemi di pompaggio, indispensabili per garantire continuità e stabilità ai cicli operativi. Accanto a questi utilizzi principali, essa viene impiegata anche per i sistemi di condizionamento e per le pompe di calore, che assicurano comfort climatico sia agli spazi produttivi sia a quelli dedicati ai servizi aziendali. L'energia elettrica supporta inoltre numerose funzioni ausiliarie, come la movimentazione interna delle merci e il funzionamento del sistema di aria compressa, fondamentale per officine e reparti tecnici. Non meno rilevante è il ruolo svolto nell'illuminazione, che copre tanto gli ambienti di lavoro quanto gli uffici, la mensa e gli spazi esterni, garantendo sicurezza e fruibilità in ogni contesto.

Il gas naturale, invece, trova la sua principale destinazione nella produzione di calore. Esso alimenta i forni del reparto produttivo e i sistemi di generazione termica, che comprendono le caldaie principali e quelle di sorveglianza, a servizio degli edifici e della mensa aziendale. Questa risorsa energetica risulta quindi fondamentale per soddisfare sia i fabbisogni termici direttamente connessi alla produzione, sia quelli legati ai servizi generali, contribuendo a mantenere condizioni ambientali adeguate e continuità nei processi.

Infine, l'acqua rappresenta un ulteriore vettore energetico di primaria importanza, con un impiego differenziato a seconda della destinazione. L'acqua potabile viene utilizzata per il fabbisogno umano e i servizi interni, in particolare nella mensa e negli spazi dedicati al personale. L'acqua industriale, invece, risponde a esigenze operative specifiche, come il raffreddamento di impianti, il lavaggio o altre attività produttive, costituendo così un elemento imprescindibile per l'efficienza dei processi.

Nel complesso, lo stabilimento può essere descritto come un sistema energetico integrato, in cui ciascun vettore svolge un ruolo specifico: l'elettricità assicura il funzionamento e l'automazione dei processi, il gas naturale fornisce l'energia termica necessaria a produzione e servizi, mentre l'acqua soddisfa sia i bisogni delle persone sia le esigenze tecniche dell'industria. La gestione coordinata di queste risorse rende possibile non solo il mantenimento della capacità produttiva, ma anche la creazione di un ambiente di lavoro efficiente, sicuro e in linea con le necessità operative quotidiane.

Capitolo 3

Descrizione modello e metodi utilizzati

3.1 Definizione del modello di previsione e obiettivi della previsione

Dopo aver delineato nel capitolo precedente il contesto aziendale e le caratteristiche energetiche dello stabilimento, si procede ora con la definizione e l'implementazione di un modello di previsione dei consumi energetici. L'obiettivo generale di questa sezione è quello di impostare le basi concettuali e metodologiche per la costruzione di modelli predittivi capaci di stimare i consumi futuri dello stabilimento con un orizzonte temporale di breve periodo.

L'intento principale è dunque quello di sviluppare due modelli di previsione distinti, progettati per stimare l'andamento dei consumi orari dell'azienda in un intervallo che varia dall'ora successiva fino a un massimo di 24 ore in avanti. In una prima fase si è optato per una configurazione *one-step-ahead*, che consente di prevedere il valore immediatamente successivo rispetto all'ultimo dato disponibile; successivamente, il modello è stato esteso alla modalità *multi-step-ahead*, per stimare più valori futuri consecutivi. Questa impostazione, orientata al breve termine (*short-term forecasting*), risponde a due esigenze principali: da un lato, alle specifiche richieste dell'azienda, interessata a una previsione utile alla gestione quotidiana dei consumi; dall'altro, alla natura stessa del processo produttivo, caratterizzato da una programmazione non continuativa e da attività organizzate per lotti di lavorazione.

Trattandosi infatti di uno stabilimento che si occupa di servizi di monitoraggio, manutenzione e ottimizzazione di turbine a gas e a vapore, l'andamento dei consumi energetici risulta fortemente dipendente dalle fasi operative e dai carichi di lavoro

del momento. In tale contesto, un modello di previsione a lungo termine risulterebbe poco informativo, in quanto difficilmente correlabile con la reale produzione o con le variazioni operative giornaliere. Di conseguenza, la previsione a breve termine si configura come la scelta più coerente e funzionale, poiché consente una gestione più efficiente dell'energia e un supporto decisionale immediato per chi opera all'interno dello stabilimento.

Per la realizzazione del sistema di previsione si è scelto di confrontare due approcci metodologici differenti, al fine di valutare la loro efficacia e individuare la soluzione più adatta al contesto aziendale. Entrambi i modelli sono stati implementati nel software MATLAB, che ha consentito di gestire l'intero processo di modellazione, addestramento e validazione in modo integrato.

Il primo modello adottato si basa su un approccio statistico tradizionale, e in particolare su una Regressione Lineare (*Linear Regression, LR*), la quale consente di rappresentare la relazione tra le variabili di input e i consumi energetici mediante una combinazione lineare dei predittori. Il secondo modello, invece, sfrutta un approccio più avanzato di apprendimento automatico, noto come *Gaussian Process Regression (GPR)*. Si tratta di un metodo di *machine learning* non parametrico basato sulla teoria bayesiana, che consente non solo di effettuare previsioni, ma anche di quantificare l'incertezza associata ad esse, fornendo così una misura della loro affidabilità.

Il processo complessivo di previsione si articola in più fasi: dopo l'impostazione dei modelli e la definizione delle loro caratteristiche principali (analizzate in dettaglio nel paragrafo 3.2), vengono descritte le tecniche di *resampling* applicate ai dati di consumo (paragrafo 3.3), necessarie per ottenere una serie temporale coerente e adeguata all'analisi. Infine, nel paragrafo 3.4, si introducono le metriche di valutazione utilizzate per confrontare le prestazioni dei due modelli e misurare l'accuratezza delle previsioni ottenute.

3.1.1 Linear Regression

La *Regressione Lineare* (LR) rappresenta uno dei metodi più semplici e diffusi per la modellazione e la previsione di fenomeni continui. Alla base di questa tecnica vi è l'ipotesi che la variabile di risposta y (nel nostro caso, il consumo energetico) possa essere espressa come combinazione lineare di un insieme di variabili esplicative x_1, x_2, \dots, x_n , più un termine di errore che tiene conto delle componenti non spiegate dal modello:

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n + \varepsilon$$

dove:

- β_0 è l'intercetta (o termine costante),
- β_i sono i coefficienti di regressione associati ai predittori x_i ,
- ε è un termine aleatorio che rappresenta il rumore o l'errore di misura, normalmente distribuito con media nulla.

Nel contesto del *time series forecasting*, le variabili esplicative x_i corrispondono in genere ai valori passati della serie temporale, ossia ai consumi energetici registrati nei periodi precedenti. Si parla in questo caso di modello *autoregressivo*, nel quale il valore futuro della variabile dipende linearmente dai suoi valori storici. Il modello più semplice di questo tipo è il modello autoregressivo di ordine p , indicato con $AR(p)$, che può essere espresso come:

$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t$$

In questo caso, il parametro p rappresenta il numero di ritardi (*lags*) considerati nel modello e determina la profondità della memoria temporale. I coefficienti ϕ_i vengono stimati in modo da minimizzare l'errore quadratico medio tra i valori osservati e quelli previsti. Il criterio più comune per la stima è il metodo dei minimi quadrati ordinari (OLS, *Ordinary Least Squares*), che può essere espresso in forma matriciale come:

$$\hat{\beta} = (X^T X)^{-1} X^T \mathbf{y}$$

dove X è la matrice dei predittori (in questo caso i valori passati di y) e \mathbf{y} il vettore dei valori osservati. La soluzione $\hat{\beta}$ rappresenta la stima dei coefficienti che meglio approssimano la relazione lineare nei minimi quadrati.

Nel caso in cui la serie temporale presenti componenti stocastiche non completamente spiegabili dai soli valori passati, è possibile estendere il modello alla forma $ARMA(p, q)$ (*AutoRegressive Moving Average*), che aggiunge una componente di media mobile:

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \varepsilon_t$$

Quando sono presenti anche variabili esogene (cioè input esterni al sistema, come temperatura, produzione o ore lavorate), si parla di modello *ARMAX* (*AutoRegressive Moving Average with eXogenous inputs*):

$$y_t = \sum_{i=1}^p \phi_i y_{t-i} + \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{k=0}^r \beta_k u_{t-k} + \varepsilon_t$$

dove u_{t-k} rappresenta le variabili esterne laggate e r il loro ordine di ritardo.

Applicata al forecasting dei consumi, la regressione lineare (nelle sue varianti AR, ARMA o ARMAX) consente di stimare in modo diretto il valore di consumo previsto, fornendo un modello interpretabile e facilmente aggiornabile. Un aspetto cruciale del suo utilizzo riguarda la scelta dell'ordine del modello, ovvero del numero di lag da considerare: un ordine troppo basso rischia di perdere informazioni temporali rilevanti, mentre un ordine eccessivo può introdurre rumore e fenomeni di sovra-adattamento (*overfitting*).

Un altro elemento da considerare è la possibilità di adottare strategie di previsione differenti: *one-step-ahead forecasting*, in cui il modello prevede solo il passo successivo, oppure *multi-step forecasting*, in cui la previsione è estesa a più orizzonti futuri (ad esempio fino a 24 ore).

Il modello lineare, pur essendo semplice, rimane uno strumento efficace per la previsione a breve termine, specialmente quando la relazione tra i valori passati e futuri del consumo si mantiene relativamente stabile [23, 24, 25, 26, 27, 28, 29, 30]. La sua principale limitazione risiede invece nell'incapacità di rappresentare comportamenti fortemente non lineari o effetti di soglia, che richiedono tecniche più flessibili, come quelle basate sul *machine learning*.

3.1.2 Guassian Process Regression (GPR)

La *Gaussian Process Regression* (GPR) rappresenta un metodo di apprendimento automatico di tipo non parametrico e probabilistico, appartenente alla famiglia dei modelli bayesiani. A differenza della regressione lineare, il GPR non assume una forma funzionale predefinita tra input e output, ma modella direttamente la distribuzione di probabilità delle funzioni che potrebbero descrivere i dati osservati.

Un processo gaussiano è una generalizzazione multivariata della distribuzione normale. In termini semplici, si può pensare a un processo gaussiano come a una distribuzione su funzioni, dove ogni punto della funzione è una variabile casuale e qualunque insieme finito di punti segue una distribuzione normale multivariata. Un processo gaussiano è completamente definito dalla sua media e dalla sua funzione di covarianza (o kernel):

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$

dove:

- $m(x)$ è la funzione di media, che in genere si assume nulla ($m(x) = 0$),
- $k(x, x')$ è la funzione di covarianza che definisce la correlazione tra i valori della funzione in due punti x e x' .

Nel contesto della regressione, si assume che i dati osservati (x_i, y_i) derivino da una funzione latente $f(x)$ perturbata da un rumore gaussiano:

$$y_i = f(x_i) + \varepsilon_i \quad \text{con} \quad \varepsilon_i \sim \mathcal{N}(0, \sigma_n^2)$$

Il compito del modello è quindi stimare, dato un insieme di osservazioni, la distribuzione di probabilità della funzione f in nuovi punti non osservati. Il risultato non è una singola previsione puntuale, ma una distribuzione gaussiana per ciascun punto, descritta da una media predittiva e da una varianza predittiva che rappresenta l'incertezza associata alla stima.

La funzione di covarianza (kernel) è l'elemento chiave del GPR, poiché stabilisce come i punti sono correlati tra loro. Un kernel comunemente utilizzato nel forecasting è quello *squared exponential* (o RBF), definito come:

$$k(x, x') = \sigma_f^2 \exp\left(-\frac{(x - x')^2}{2l^2}\right)$$

dove:

- σ_f^2 controlla la varianza complessiva del processo,
- l (length scale) determina quanto rapidamente decresce la correlazione tra i punti nel tempo.

In pratica, l stabilisce la “memoria” del modello: valori piccoli implicano che solo i punti molto vicini nel tempo si influenzano a vicenda, mentre valori più grandi comportano una dipendenza più lunga. La scelta e la calibrazione del kernel sono quindi fondamentali per adattare il modello ai dati.

Data una serie di osservazioni $X = [x_1, \dots, x_n]$ e $y = [y_1, \dots, y_n]$, la distribuzione congiunta dei valori osservati e di quelli da predire f_* è anch’essa gaussiana:

$$\begin{bmatrix} y \\ f_* \end{bmatrix} \sim \mathcal{N} \left(0, \begin{bmatrix} K(X, X) + \sigma_n^2 I & K(X, X_*) \\ K(X_*, X) & K(X_*, X_*) \end{bmatrix} \right)$$

da cui, attraverso l’inferenza bayesiana, si ottiene la distribuzione condizionata di f_* dato y :

$$p(f_* | X, y, X_*) = \mathcal{N}(\bar{f}_*, \text{cov}(f_*))$$

con media e covarianza predittiva:

$$\begin{aligned} \bar{f}_* &= K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1}y \\ \text{cov}(f_*) &= K(X_*, X_*) - K(X_*, X)[K(X, X) + \sigma_n^2 I]^{-1}K(X, X_*) \end{aligned}$$

Questo approccio consente di ottenere non solo una previsione del valore atteso, ma anche una misura quantitativa dell’incertezza associata a ciascuna previsione.

Nel forecasting dei consumi energetici, il GPR risulta particolarmente utile per catturare le relazioni non lineari e le dinamiche complesse che caratterizzano le serie temporali reali. La previsione a breve termine (ad esempio oraria o giornaliera) può essere formulata modellando i consumi futuri come funzione dei valori precedenti, analogamente a quanto fatto nei modelli autoregressivi, ma con la capacità del kernel di adattarsi a pattern non lineari.

L’aspetto probabilistico del GPR fornisce inoltre un’informazione aggiuntiva di grande utilità operativa: la varianza predittiva permette infatti di quantificare il grado di affidabilità della previsione. Nei contesti industriali, questa informazione consente di gestire in modo più consapevole le strategie di controllo e pianificazione, distinguendo tra previsioni ad alta e bassa confidenza.

Infine, il GPR è un metodo flessibile ma computazionalmente più oneroso: la complessità dell’inversione matriciale cresce con il numero di dati ($O(n^3)$), motivo per cui trova la sua applicazione ideale in problemi di *short-term forecasting*, dove il numero di osservazioni e la finestra temporale sono contenuti ma la precisione richiesta è elevata [31, 32, 33, 34, 35, 36, 37, 38, 39].

3.2 Struttura e costruzione del modello

In questa sezione vengono illustrati la struttura generale e i metodi di previsione adottati per lo sviluppo dei modelli predittivi dei consumi energetici. Come anticipato nel paragrafo precedente, sono stati implementati due approcci differenti: un modello di tipo statistico, basato sulla *Regressione Lineare* (LR), e un modello di apprendimento automatico, fondato sulla *Gaussian Process Regression* (GPR). Entrambi sono stati sviluppati all'interno dell'ambiente *MATLAB*, utilizzando le funzioni e i pacchetti integrati per l'elaborazione di serie temporali e la regressione supervisionata. Gli script di riferimento completi sono riportati in appendice.

La scelta di utilizzare due metodologie differenti nasce dalla volontà di confrontare un approccio classico, fondato su relazioni lineari e facilmente interpretabili, con uno più avanzato, capace di modellare comportamenti non lineari e di fornire stime dell'incertezza associata alle previsioni. Ciò consente di valutare non solo l'accuratezza dei modelli, ma anche la loro flessibilità e robustezza rispetto alle variazioni del segnale di consumo.

Il processo di previsione, comune a entrambi i modelli, segue una struttura modulare che può essere sintetizzata in tre fasi principali. Nella prima fase vengono caricati e preprocessati i dati sperimentali, comprendendo, se necessario, l'operazione di *resampling* dei dati orari, che consente di ottenere serie uniformemente campionate e adatte all'analisi predittiva. Tale operazione viene parametrizzata attraverso la definizione di una finestra temporale di aggregazione, la cui ampiezza potrà variare nei test successivi per valutare l'effetto del *resampling* sui risultati delle previsioni.

Segue la suddivisione del dataset in due sottoinsiemi distinti: uno destinato alla fase di addestramento (circa l'80% dei dati) e l'altro alla fase di test (il restante 20%), necessaria per valutare le prestazioni dei modelli su dati non visti in precedenza. Nella seconda fase viene implementato il modello di previsione vero e proprio, mentre la terza è dedicata alla valutazione dell'accuratezza attraverso specifiche metriche di errore, che saranno descritte nel paragrafo 3.4.

Modello basato su Regressione Lineare (LR)

Il primo approccio si fonda su un modello di tipo autoregressivo, realizzato attraverso gli strumenti della *System Identification Toolbox* di *MATLAB*. In particolare, viene impiegato un modello *ARMAX* (*AutoRegressive Moving Average with eXogenous inputs*), che estende la classica regressione lineare includendo anche componenti che descrivono l’andamento medio mobile degli errori. Questo tipo di modello rappresenta una generalizzazione del modello $AR(p)$, nel quale il valore corrente della serie temporale è espresso come combinazione lineare dei propri valori passati (più un termine di errore).

Nel caso specifico, la serie dei consumi energetici è stata utilizzata come variabile autoregressiva. Il numero di ritardi (*lags*) inizialmente considerato è pari a dieci, valore scelto per catturare l’influenza dei consumi precedenti sull’andamento futuro. Tuttavia, tale parametro non è stato mantenuto fisso: nei capitoli successivi si analizzerà come la variazione del numero di *lag* possa influenzare l’accuratezza del modello e la sua capacità di rappresentare la dinamica temporale dei consumi.

Il modello è stato addestrato sul set di dati di training mediante il metodo dei minimi quadrati (*least squares*), ottenendo la stima dei coefficienti che meglio descrivono la relazione lineare tra i valori passati e futuri della serie.

Dopo la fase di addestramento, il modello è stato impiegato per generare due tipologie di previsioni: una *one-step-ahead*, che stima il valore immediatamente successivo rispetto all’ultimo campione noto, e una *multi-step-ahead*, che prevede più valori futuri consecutivi (fino a quattro nel caso analizzato). Questa doppia configurazione consente di valutare la stabilità del modello e la sua capacità di mantenere buone prestazioni anche in orizzonti temporali più ampi. Le previsioni ottenute sono state quindi confrontate con i valori reali del set di test, e le differenze hanno permesso di calcolare indicatori di accuratezza come l’errore quadratico medio (RMSE).

L’approccio lineare presenta il vantaggio di una notevole semplicità interpretativa, poiché i parametri stimati consentono di identificare in modo diretto la dipendenza del consumo futuro dai valori passati. Tuttavia, esso è meno adatto a rappresentare dinamiche complesse o comportamenti non lineari, che possono caratterizzare i consumi energetici reali in presenza di variazioni operative o stagionali.

Modello basato su Gaussian Process Regression (GPR)

Il secondo metodo adottato si basa sulla *Gaussian Process Regression*, un modello di apprendimento automatico non parametrico che appartiene alla famiglia dei metodi bayesiani. A differenza della regressione lineare, il GPR non assume una forma funzionale predefinita della relazione tra input e output, ma la apprende direttamente dai dati, modellando l'incertezza e la variabilità intrinseca del fenomeno studiato. Questo approccio si dimostra particolarmente efficace quando si lavora con serie temporali rumorose o con relazioni non lineari.

Il modello è stato implementato utilizzando la funzione `fitrgp` di *MATLAB*, che consente di definire e addestrare un processo gaussiano con un determinato tipo di funzione di covarianza. In questo caso, è stato adottato un kernel di tipo *squared exponential* (o *radial basis function*, RBF), che rappresenta una delle scelte più comuni per modellare funzioni lisce e continue. Il modello è stato inoltre standardizzato, in modo da normalizzare i dati e migliorare la stabilità numerica della stima.

Analogamente al caso della regressione lineare, la fase di input è stata costruita generando un insieme di variabili laggate, cioè vettori che contengono i valori di consumo registrati nelle ore precedenti. Anche in questo caso, il numero di *lag* e l'ampiezza dell'orizzonte di previsione (*nahead*) rappresentano parametri iniziali che potranno essere modificati nei test successivi per verificare come la loro variazione influenzi le prestazioni del modello. Tale flessibilità consente di esplorare l'effetto delle configurazioni temporali sulla capacità predittiva del processo gaussiano.

Una delle caratteristiche più rilevanti del GPR è la capacità di fornire, oltre alla previsione del valore medio atteso, anche una stima della deviazione standard associata a ciascuna previsione. Ciò consente di costruire intervalli di confidenza (ad esempio al 95%), che rappresentano l'incertezza statistica legata alla previsione. In termini pratici, questa informazione è preziosa per la gestione energetica dello stabilimento, poiché consente di valutare non solo la previsione in sé, ma anche il grado di affidabilità con cui essa può essere utilizzata nei processi decisionali.

Analogamente al modello lineare, anche il GPR è stato testato sia in configurazione *one-step-ahead* sia in modalità *multi-step-ahead*, per analizzare la capacità del modello di mantenere accuratezza su orizzonti di previsione progressivamente più lunghi.

In conclusione, i due modelli presentano caratteristiche complementari: la Regressione Lineare offre semplicità e rapidità di calcolo, mentre la Gaussian Process Regression introduce una componente probabilistica più raffinata, capace di stimare non solo il valore previsto ma anche la sua incertezza. Inoltre, l'analisi della sensibilità ai parametri — quali numero di *lag* e finestra di *resampling* — fornisce un ulteriore livello di approfondimento utile a comprendere quanto le scelte di configurazione influenzino le prestazioni complessive del sistema di previsione. Questa doppia prospettiva permette di ottenere una visione più completa del comportamento energetico dello stabilimento e fornisce una base solida per le successive analisi comparative sui modelli.

3.3 Metriche di valutazione utilizzate

Per valutare le prestazioni dei modelli di previsione sviluppati, è fondamentale disporre di indicatori quantitativi che consentano di misurare in modo oggettivo la distanza tra i valori previsti e quelli effettivamente osservati. Nel contesto dell'analisi dei consumi energetici, tali metriche assumono un ruolo essenziale, poiché permettono di confrontare modelli diversi, monitorarne l'affidabilità e individuare eventuali limiti o tendenze sistematiche nelle stime [40, 41, 42].

Nel presente lavoro sono state adottate principalmente due metriche: l'Errore Quadratico Medio (*Root Mean Squared Error*, RMSE) e l'Errore Medio Assoluto (*Mean Absolute Error*, MAE). Entrambe misurano la precisione del modello, ma con sensibilità diversa rispetto agli errori di grande ampiezza. Il loro impiego congiunto consente di ottenere una valutazione più completa dell'accuratezza delle previsioni.

3.3.1 RMSE

L'**Errore Quadratico Medio (RMSE)** rappresenta una delle metriche più diffuse per la valutazione di modelli di regressione e previsione. Esso misura la radice quadrata della media dei quadrati delle differenze tra i valori previsti e quelli osservati, ed è definito come:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

dove:

- y_i rappresenta il valore reale osservato al tempo i ,
- \hat{y}_i è il corrispondente valore previsto dal modello,
- n è il numero totale di osservazioni considerate.

L'RMSE fornisce una misura dell'errore medio espresso nella stessa unità di misura della variabile analizzata (ad esempio kWh o tep). Poiché gli errori vengono elevati al quadrato prima di essere mediati, la metrica penalizza in modo più marcato le deviazioni di grande ampiezza: un singolo errore molto elevato può influenzare significativamente il valore complessivo dell'indicatore. Per questo motivo, l'RMSE risulta particolarmente adatto quando è necessario ridurre l'impatto di previsioni fortemente errate o quando si desidera dare maggiore peso alla precisione nei picchi di consumo.

Nel contesto del *short-term forecasting* dei consumi energetici, un RMSE basso indica che il modello è in grado di riprodurre accuratamente l'andamento reale della

serie, mentre valori elevati suggeriscono scostamenti significativi tra previsioni e misure. Essendo una misura assoluta, tuttavia, l'RMSE non consente un confronto diretto tra variabili espresse su scale differenti (ad esempio consumi elettrici e termici), a meno che non venga normalizzato rispetto al valore medio o al range dei dati.

3.3.2 MAE

L'**Errore Medio Assoluto (MAE)** rappresenta un'altra metrica largamente utilizzata per valutare le prestazioni di modelli di previsione. A differenza dell'RMSE, essa calcola la media del valore assoluto delle differenze tra valori reali e stimati, senza elevare al quadrato gli errori. La sua formulazione matematica è la seguente:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

Anche in questo caso, i simboli hanno lo stesso significato precedentemente descritto. L'MAE quantifica, in media, di quanto le previsioni si discostano dai valori effettivi, restituendo una misura diretta e facilmente interpretabile dell'accuratezza del modello. A differenza dell'RMSE, ogni errore contribuisce in modo proporzionale al risultato complessivo: ciò rende la metrica meno sensibile agli outlier e più rappresentativa della deviazione tipica tra previsione e valore reale.

Nel caso della previsione dei consumi energetici, l'MAE consente di stimare l'errore medio commesso dal modello in termini fisici (ad esempio, “in media il modello sbaglia di 15 kWh all'ora”). Un vantaggio importante di questa metrica è la sua robustezza: mentre l'RMSE enfatizza gli errori più grandi, l'MAE fornisce una visione equilibrata dell'errore complessivo, risultando più stabile quando i dati presentano occasionali variazioni improvvise o valori anomali.

In termini di interpretazione pratica, l'RMSE e l'MAE vengono spesso utilizzati congiuntamente: un RMSE significativamente superiore al MAE indica la presenza di alcuni errori di grande ampiezza che influenzano in modo rilevante la prestazione del modello, mentre valori simili suggeriscono una distribuzione degli errori più uniforme. Questa doppia prospettiva consente una valutazione più completa e bilanciata dell'efficacia predittiva dei modelli impiegati.

Capitolo 4

Previsione dati di consumo dell'azienda EthosEnergy oggetto di studio

4.1 Definizione e descrizione del dataset

Per lo sviluppo e la validazione dei modelli previsionali è stato utilizzato un dataset costituito dai dati di consumo elettrico complessivo dello stabilimento EthosEnergy oggetto di analisi. I dati rappresentano la potenza elettrica assorbita istantaneamente dall'impianto, espressa in chilowatt (kW), e sono stati raccolti con una frequenza temporale oraria nell'arco temporale compreso tra giugno 2024 e giugno 2025.

Questo insieme di informazioni fornisce una base solida per lo studio dei comportamenti energetici dell'azienda, permettendo di identificare pattern ricorrenti, stagionalità e variazioni legate all'attività produttiva.

Per una prima analisi esplorativa del dataset, si è rappresentato l'andamento temporale dei consumi sotto forma di serie storica. La Figura 4.1 mostra l'evoluzione della potenza assorbita nel periodo di osservazione: si nota una marcata variabilità giornaliera, con picchi che riflettono le ore di maggiore attività produttiva e cadute di potenza in corrispondenza delle ore notturne o dei periodi di fermo.

Su scala mensile, è possibile individuare un andamento tendenziale che risente sia della stagionalità climatica (temperature più alte o più basse che incidono sui fabbisogni energetici) sia delle dinamiche operative dello stabilimento (turni di lavoro, manutenzioni, festività, ecc.).

La distribuzione irregolare ma ciclica dei valori di potenza suggerisce quindi un

sistema caratterizzato da una componente deterministica e da una componente aleatoria, rendendo il dataset particolarmente adatto a essere trattato con metodi previsionali di tipo statistico e machine learning.

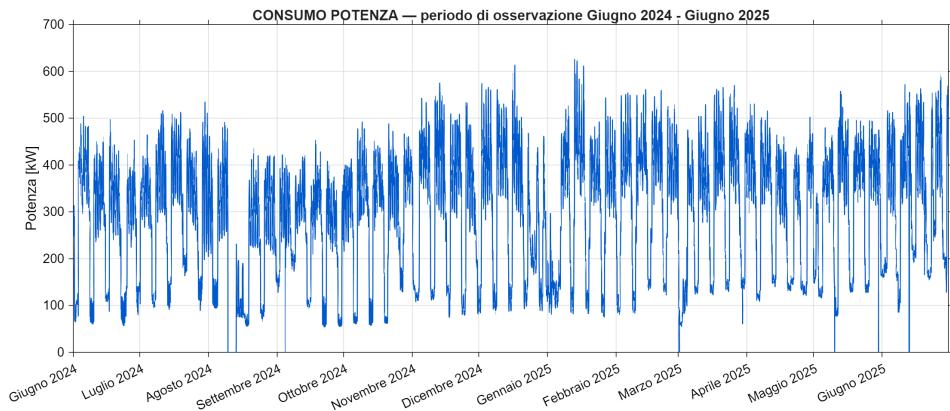


Figura 4.1: Consumo stabilimento EthosEnergy periodo di osservazione

Per approfondire ulteriormente la struttura interna del dataset e visualizzare i comportamenti ricorrenti nell'arco della giornata, si è realizzata una rappresentazione in forma di carpet plot (Figura 4.2).

In questo grafico, sull'asse delle ascisse è riportato il tempo (in mesi), mentre sull'asse delle ordinate sono indicate le ore del giorno (da 0 a 24). Ogni cella è colorata in base al valore medio della potenza assorbita in quell'ora e in quel giorno, permettendo di evidenziare con immediatezza le fasce orarie a maggiore intensità di consumo. Le tonalità più calde (rosso, arancione) corrispondono a livelli di potenza più elevati, mentre quelle fredde (azzurro, blu) indicano minori assorbimenti. Dalla mappa si può osservare come i consumi più elevati si concentrino tipicamente nelle ore centrali della giornata, con una chiara ripetitività giornaliera e differenze stagionali legate al periodo dell'anno.

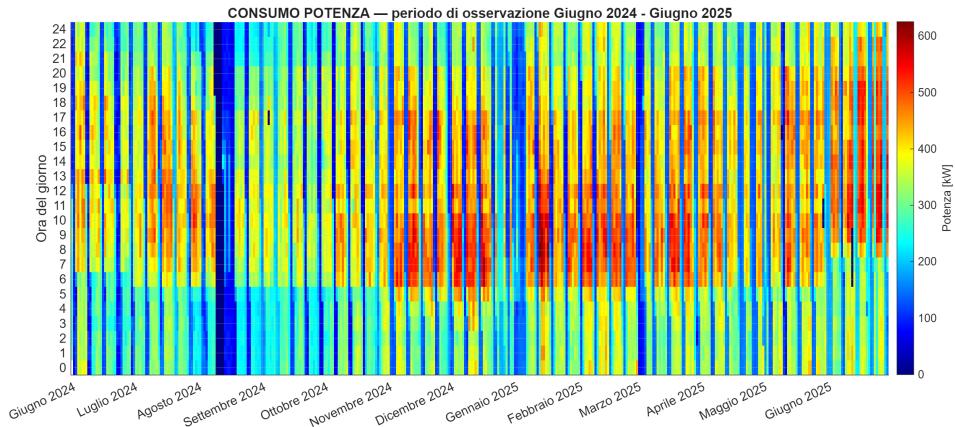


Figura 4.2: Consumo stabilimento EthosEnergy periodo di osservazione

Il dataset fornisce quindi una rappresentazione completa e strutturata del comportamento energetico dello stabilimento, costituendo una solida base di partenza per lo sviluppo dei modelli previsionali. L'integrazione tra analisi temporale e visualizzazione bidimensionale permette di individuare con chiarezza sia le tendenze complessive dei consumi sia le loro ricorrenze giornaliere e stagionali, elementi essenziali per la definizione di modelli predittivi affidabili e coerenti con la realtà operativa.

4.2 Risultati modello di previsione

In questa sezione vengono presentati e discussi i risultati ottenuti dall'applicazione dei modelli di previsione sui dati di consumo energetico dell'azienda oggetto di studio. L'obiettivo di questa fase è valutare la capacità dei modelli di riprodurre e anticipare l'andamento dei consumi, analizzando in che modo la variazione dei principali parametri di configurazione influenzi la precisione e la stabilità delle previsioni.

I parametri impostati per lo sviluppo e la valutazione dei modelli sono tre: il valore di resampling, che stabilisce l'intervallo temporale con cui i dati vengono mediati o ricampionati; il numero di passi temporali in avanti (nahead), che indica l'orizzonte di previsione nel caso delle predizioni multi-step-ahead; e infine il numero di lag temporali (numLags o p), che rappresenta la quantità di osservazioni passate utilizzate come input dal modello per stimare il valore futuro.

I primi due parametri, ossia il resampling e l'orizzonte temporale di previsione, sono stati oggetto di un'analisi sistematica, volta a comprendere come la loro variazione influisca sulla qualità e sulla stabilità delle previsioni. L'obiettivo è stato quello di individuare un compromesso ottimale tra livello di dettaglio temporale e capacità predittiva, valutando al contempo la coerenza dei risultati ottenuti con il comportamento energetico reale dello stabilimento.

Il parametro dei lag temporali, invece, è stato mantenuto costante e pari a dieci in tutte le simulazioni, poiché rappresenta un valore equilibrato sia sotto il profilo tecnico che energetico. Dal punto di vista modellistico, dieci lag consentono di fornire al modello una memoria storica sufficiente a cogliere le relazioni temporali più significative, evitando al contempo un'eccessiva complessità che potrebbe incrementare i tempi di addestramento o introdurre fenomeni di overfitting. Dal punto di vista energetico, tale intervallo risulta coerente con le dinamiche operative tipiche di un impianto industriale, in cui le variazioni di carico sono spesso influenzate da cicli giornalieri e turni produttivi.

Va inoltre osservato che, in assenza di resampling, i dieci lag corrispondono a un orizzonte temporale di dieci ore, mentre nel caso in cui i dati vengano ricampionati, essi rappresentano comunque dieci step temporali indietro, ognuno dei quali copre un periodo più ampio. Ad esempio, con un resampling impostato a quattro ore, i dieci lag equivalgono a una finestra storica di quaranta ore. In questo modo, il modello conserva una profondità temporale adeguata anche al variare della frequenza di campionamento, garantendo robustezza e capacità descrittiva su diversi orizzonti temporali.

Il lavoro di previsione è stato quindi articolato in due fasi principali: nella prima è stata realizzata la previsione one-step-ahead, che stima il valore di consumo

immediatamente successivo a quello disponibile; nella seconda è stata affrontata la previsione multi-step-ahead, in cui vengono stimati più valori futuri in sequenza. Entrambe le fasi hanno previsto l'impiego e il confronto di due differenti approcci modellistici, basati rispettivamente sulla Linear Regression (LR) e sul Gaussian Process Regression (GPR), con l'obiettivo di valutarne le prestazioni e la capacità di adattarsi alle caratteristiche del segnale energetico.

La bontà delle previsioni è stata infine valutata attraverso le metriche RMSE (Root Mean Square Error) e MAE (Mean Absolute Error), che misurano rispettivamente l'errore quadratico medio e l'errore medio assoluto tra i valori reali e quelli stimati. Questi indicatori sintetizzano in modo efficace le prestazioni dei modelli, permettendo di identificare le configurazioni più accurate e stabili nella rappresentazione del comportamento energetico dello stabilimento.

4.3 Previsione one-step-ahead

In questa sezione vengono riportati i risultati ottenuti dalle previsioni *one-step-ahead*, ovvero previsioni del valore di assorbimento energetico a un passo temporale di distanza rispetto al dato corrente. Le simulazioni sono state condotte mantenendo i parametri di riferimento adottati per l'intero studio, fissando il numero di regressori a $p = 10$ e senza applicare alcuna operazione di filtraggio o resampling sui dati originari. L'orizzonte temporale è quindi impostato a $n = 1$, corrispondente a una previsione a un'ora in avanti, in coerenza con il passo temporale dei dati analizzati.

I risultati delle simulazioni sono illustrati nelle figure seguenti, suddivisi per modello di previsione. Per ciascun metodo vengono mostrate due rappresentazioni: la prima riferita all'intero periodo di test, la seconda focalizzata su una finestra temporale di sette giorni consecutivi, al fine di evidenziare in modo più chiaro la coerenza tra valori reali e stime modellate.

Linear Regression (LR)

Le Figure 4.3 e 4.4 mostrano i risultati ottenuti dal modello di regressione lineare.

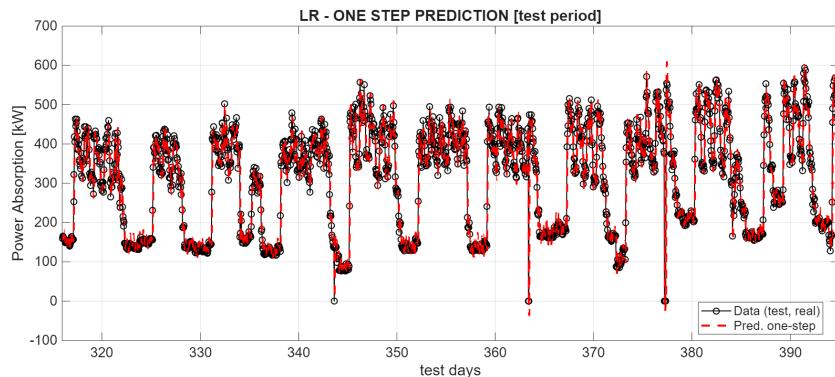


Figura 4.3: LR - one step prediction [testing period]

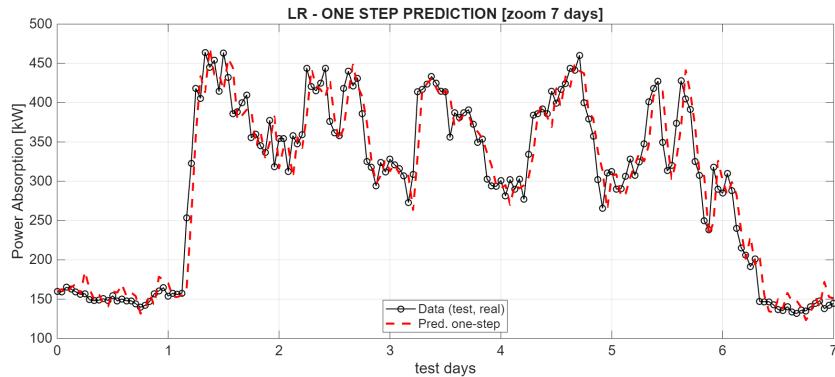


Figura 4.4: LR - one step prediction [zoom 7 days]

La Figura 4.3 rappresenta l'andamento delle previsioni sul periodo di test completo, mettendo a confronto i valori reali con quelli stimati dal modello, mentre la Figura 4.4 riporta un ingrandimento su un intervallo di sette giorni consecutivi, utile a visualizzare con maggiore dettaglio l'andamento delle previsioni.

Metrica	Valore
RMSE	38,44
MAE	22,74

Tabella 4.1: Metriche di valutazione LR - one step prediction

In tabella 4.1 sono riportati i risultati delle metriche di valutazione scelte per lo studio che ha riportato il modello LR

Gaussian Process Regression (GPR)

Le Figure 4.5 e 4.6 riportano i risultati ottenuti tramite il modello di Gaussian Process Regression.

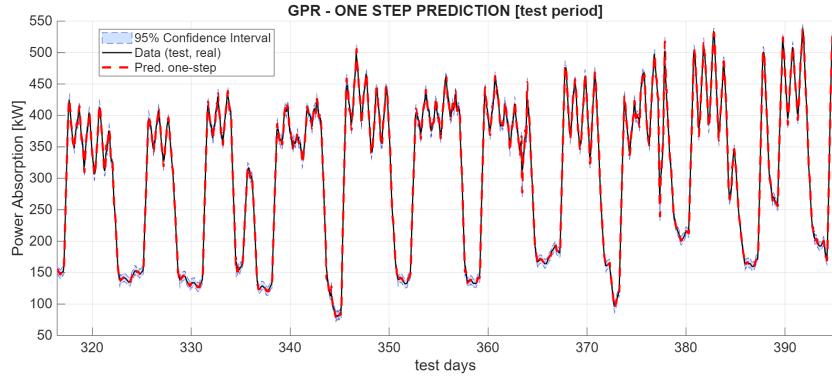


Figura 4.5: GPR - one step prediction [testing period]

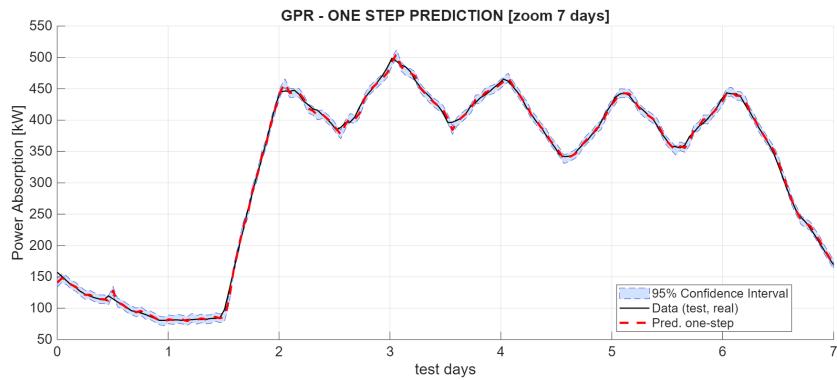


Figura 4.6: GPR - one step prediction [zoom 7 days]

Anche in questo caso, la Figura 4.5 mostra l'andamento sul periodo di test complessivo, mentre la Figura 4.6 propone un dettaglio su un arco temporale di sette giorni. Si osserva inoltre la presenza delle bande di confidenza al 95%, che rappresentano la stima dell'incertezza associata alle previsioni prodotte dal modello.

Metrica	Valore
RMSE	4,43
MAE	2,80

Tabella 4.2: Metriche di valutazione GPR - one step prediction

In tabella 4.2 sono invece riportati i risultati delle metriche di valutazione scelte per lo studio che ha riportato il modello di previsione GPR

4.3.1 Analisi del parametro di resampling

In questa sezione vengono presentati i risultati relativi alla variazione del parametro $t_{resampling}$, che regola l'intervallo temporale con cui i dati di consumo vengono mediati prima dell'elaborazione. Per ciascun valore considerato, viene mostrato innanzitutto un confronto tra il segnale originale e quello preprocessato, al fine di evidenziare come il processo di resampling influenzi la forma e la densità del dataset nel dominio temporale, inoltre saranno anche rappresentati i risultati attraverso grafici con zoom su un intervallo rappresentativo.

Successivamente, sono riportati i risultati delle previsioni *one step ahead* ottenute dai modelli impiegati (prima i risultati del modello LR e poi GPR, quest'ultimo produrrà anche un intervallo di confidenza, anch'esso riportato nei grafici relativi), mantenendo inalterati tutti gli altri parametri di configurazione. Infine, le metriche di valutazione adottate, RMSE e MAE, sono sintetizzate nelle tabelle poste al termine dei grafici di previsione, fornendo un quadro quantitativo dell'accuratezza dei modelli per ciascun valore di $t_{resampling}$.

- **Parametro di resampling $t_{resampling} = 4$ h**

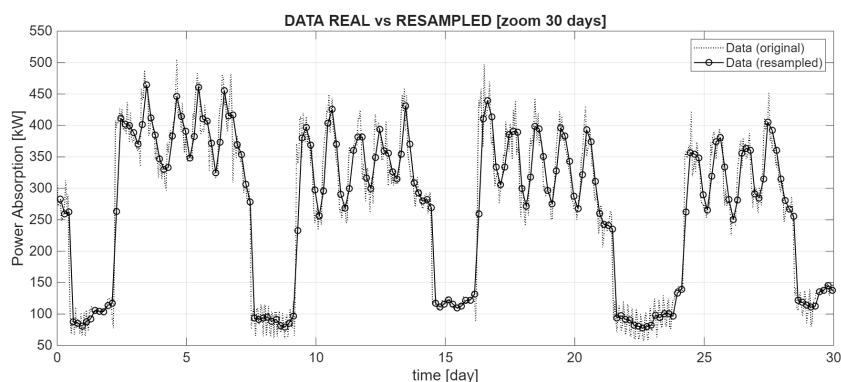


Figura 4.7: data REAL vs RESAMPLED [zoom 30 days]

Linear Regression (LR)

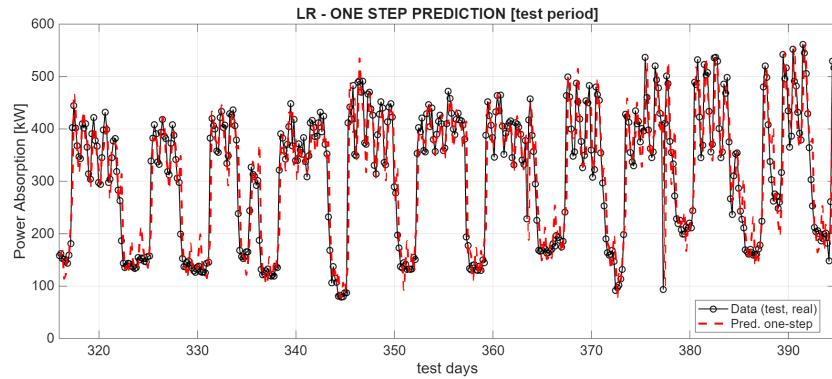


Figura 4.8: LR - one step prediction $t_r = 4$ h [testing period]

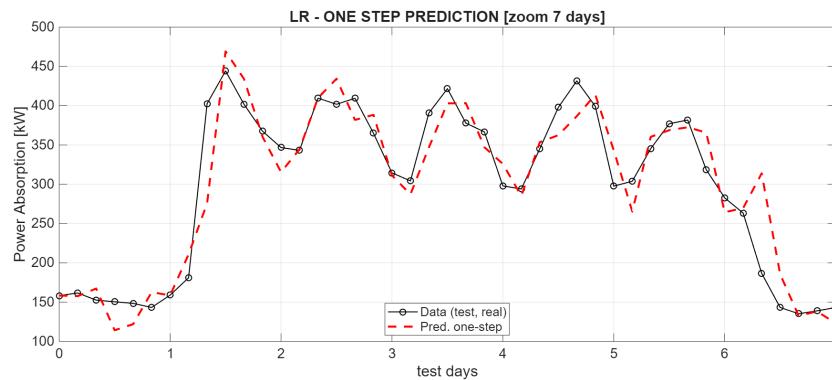


Figura 4.9: LR - one step prediction $t_r = 4$ h [zoom 7 days]

Metrica	Valore
RMSE	55,50
MAE	34,60

Tabella 4.3: Metriche di valutazione LR – one step prediction ($t_{resampling} = 4$ h)

Gaussian Process Regression (GPR)

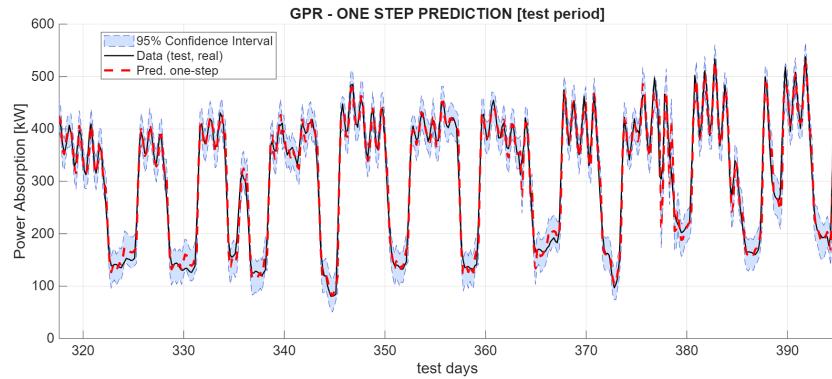


Figura 4.10: LR - one step prediction $t_r = 4$ h [testing period]

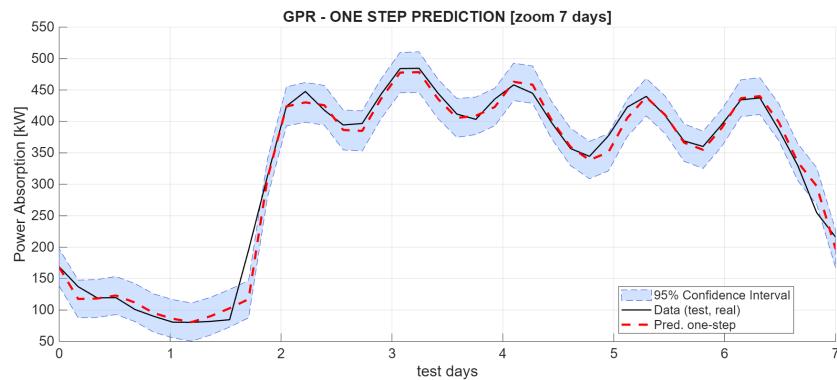


Figura 4.11: GPR - one step prediction $t_r = 4$ h [zoom 7 days]

Metrica	Valore
RMSE	20,41
MAE	12,75

Tabella 4.4: Metriche di valutazione GPR – one step prediction ($t_{resampling} = 4$ h)

- Parametro di resampling $t_{resampling} = 8 \text{ h}$

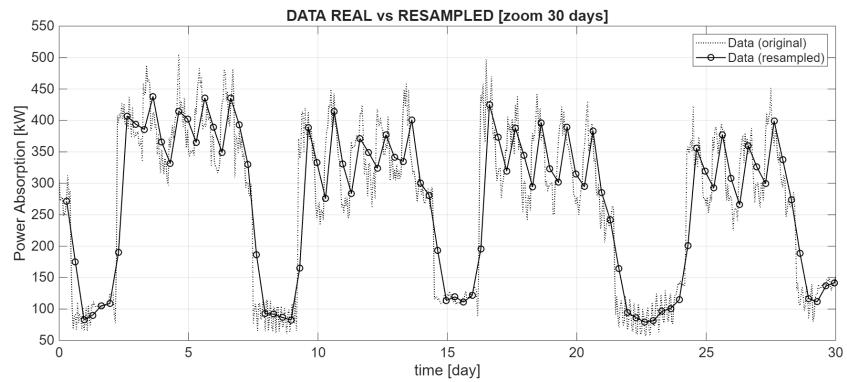


Figura 4.12: data REAL vs RESAMPLED [zoom 30 days]

Linear Regression (LR)

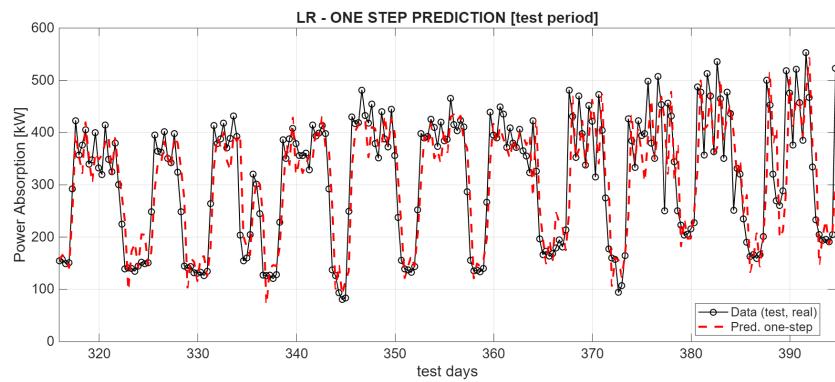


Figura 4.13: LR - one step prediction $t_r = 8 \text{ h}$ [testing period]

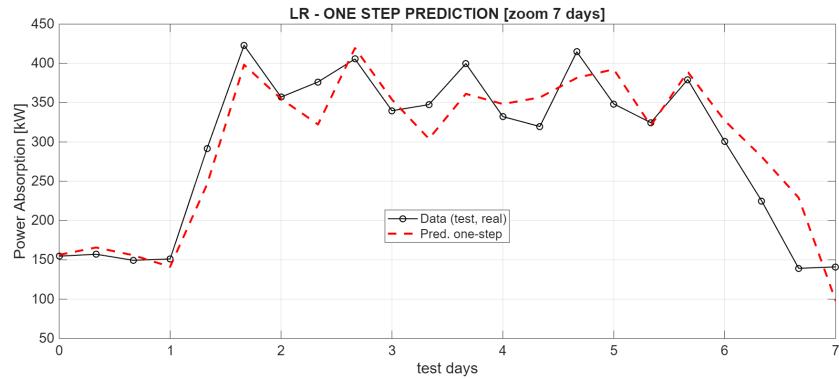


Figura 4.14: LR - one step prediction $t_r = 8$ h [zoom 7 days]

Metrica	Valore
RMSE	64,85
MAE	46,67

Tabella 4.5: Metriche di valutazione LR – one step prediction ($t_{resampling} = 8$ h)

Gaussian Process Regression (GPR)

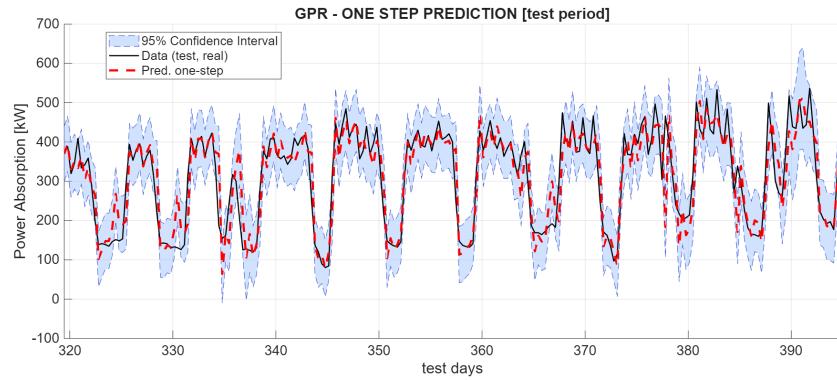


Figura 4.15: LR - one step prediction $t_r = 8$ h [testing period]

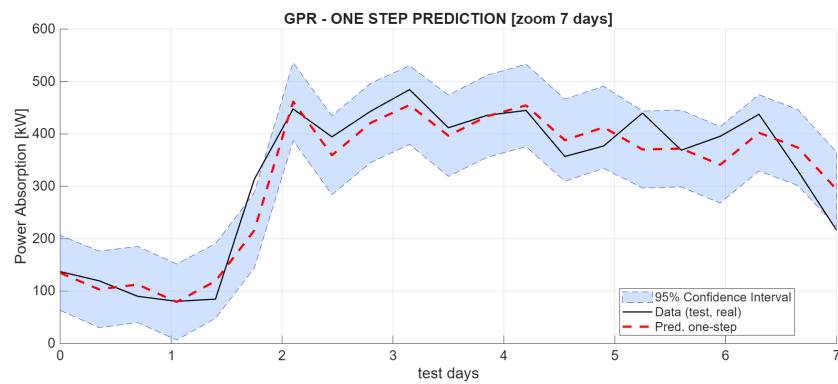


Figura 4.16: GPR - one step prediction $t_r = 8$ h [zoom 7 days]

Metrica	Valore
RMSE	52,95
MAE	36,5

Tabella 4.6: Metriche di valutazione GPR – one step prediction ($t_resampling = 8$ h)

- Parametro di resampling $t_{resampling} = 12 \text{ h}$

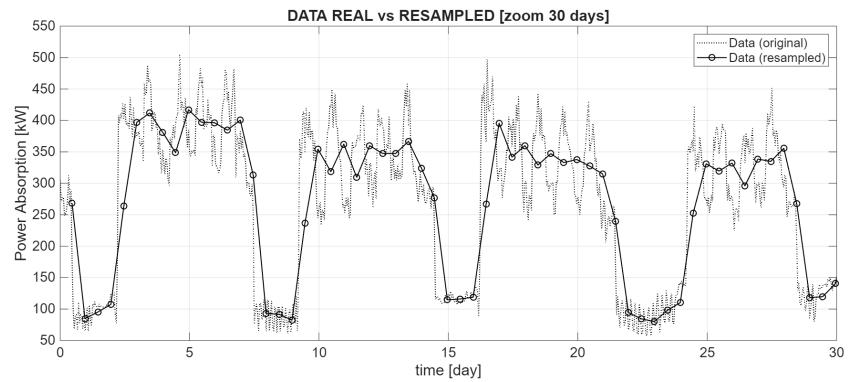


Figura 4.17: data REAL vs RESAMPLED [zoom 30 days]

Linear Regression (LR)

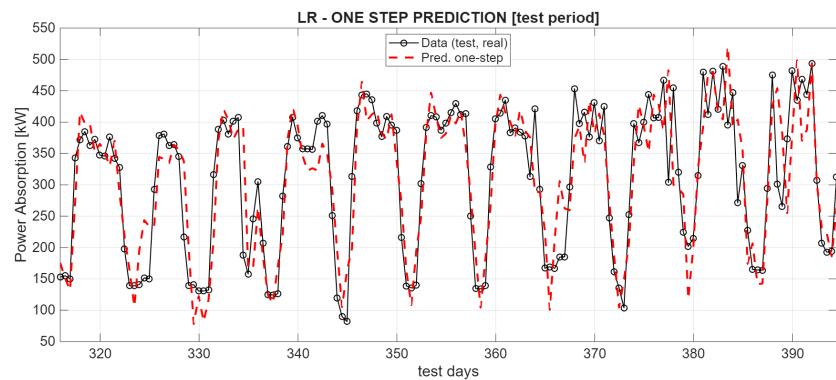


Figura 4.18: LR - one step prediction $t_r = 12 \text{ h}$ [testing period]

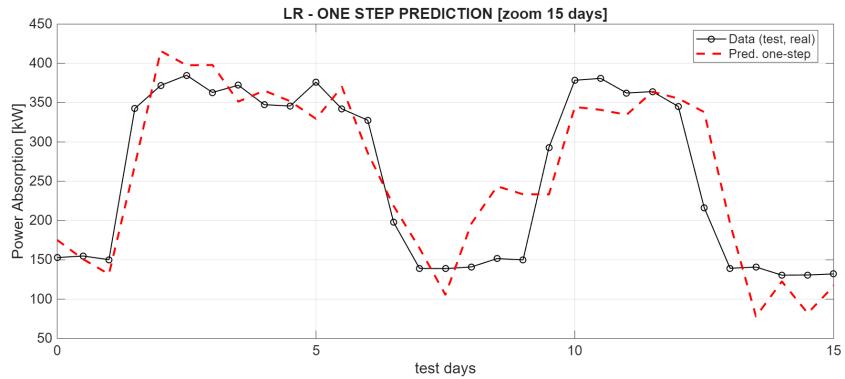


Figura 4.19: LR - one step prediction $t_r = 12$ h [zoom 15 days]

Metrica	Valore
RMSE	58,35
MAE	45,02

Tabella 4.7: Metriche di valutazione LR – one step prediction ($t_{resampling} = 12$ h)

Gaussian Process Regression (GPR)

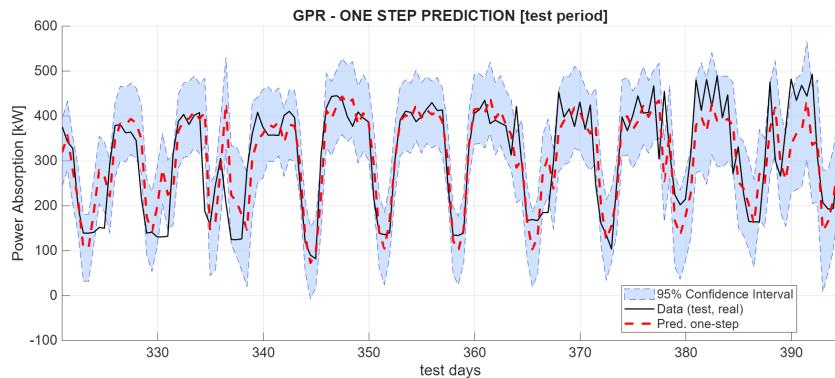


Figura 4.20: LR - one step prediction $t_r = 12$ h [testing period]

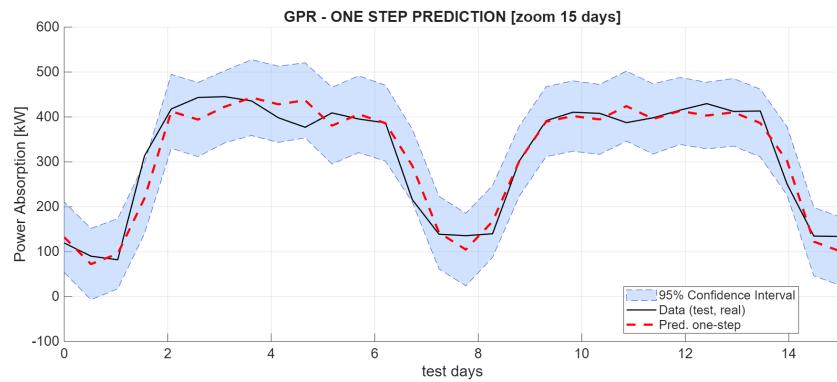


Figura 4.21: GPR - one step prediction $t_r = 12$ h [zoom 15 days]

Metrica	Valore
RMSE	66,63
MAE	48,33

Tabella 4.8: Metriche di valutazione GPR – one step prediction ($t_resampling = 12$ h)

- Parametro di resampling $t_{resampling} = 24$ h

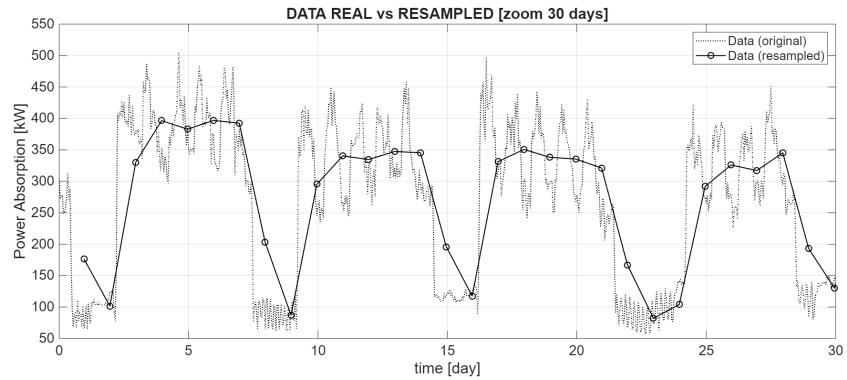


Figura 4.22: data REAL vs RESAMPLED [zoom 30 days]

Linear Regression (LR)

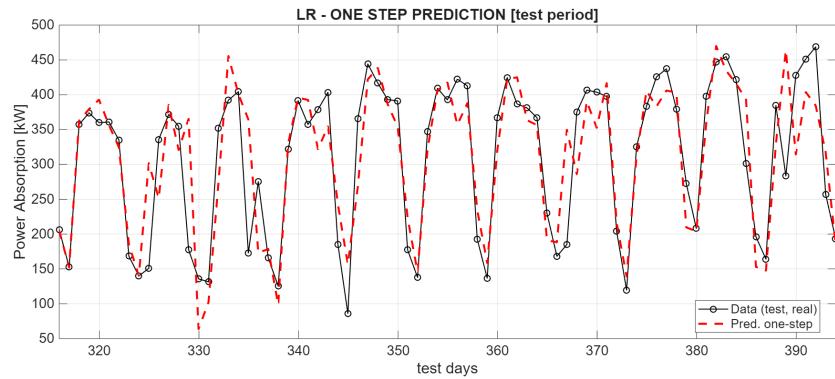


Figura 4.23: LR - one step prediction $t_r = 24$ h [testing period]

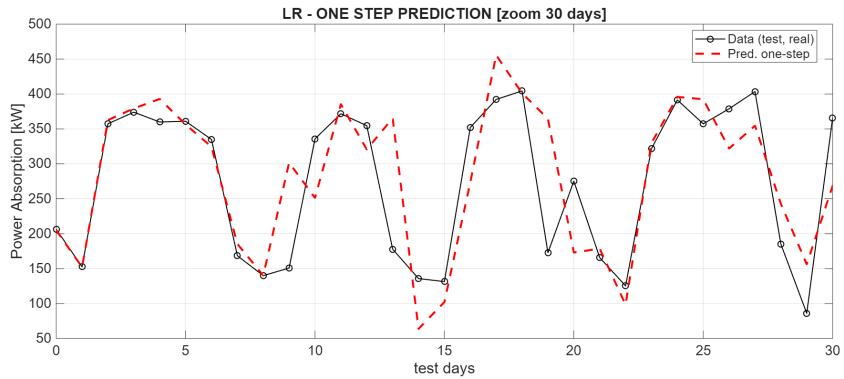


Figura 4.24: LR - one step prediction $t_r = 24$ h [zoom 30 days]

Metrica	Valore
RMSE	61,34
MAE	42,93

Tabella 4.9: Metriche di valutazione LR – one step prediction ($t_{resampling} = 24$ h)

Gaussian Process Regression (GPR)

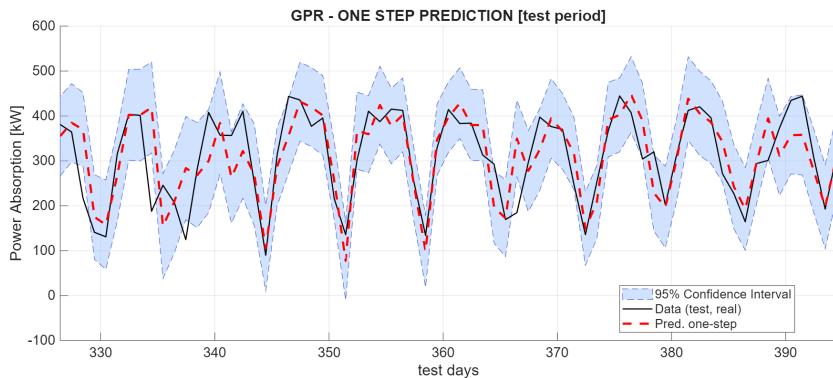


Figura 4.25: LR - one step prediction $t_r = 24$ h [testing period]

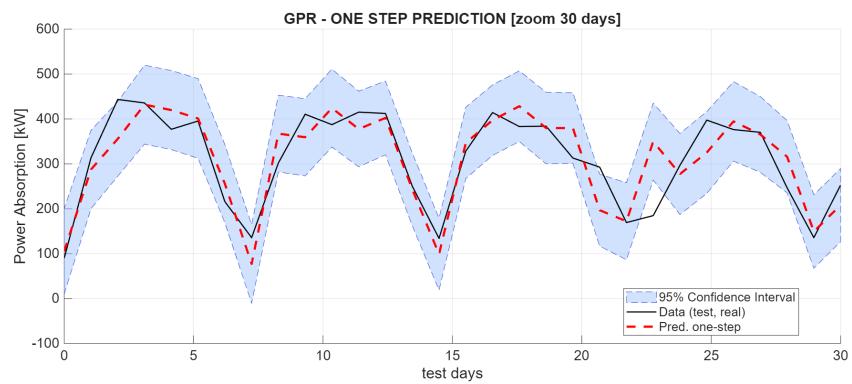


Figura 4.26: GPR - one step prediction $t_r = 24$ h [zoom 30 days]

Metrica	Valore
RMSE	63,97
MAE	45,65

Tabella 4.10: Metriche di valutazione GPR – one step prediction ($t_resampling = 24$ h)

- Parametro di resampling $t_{resampling} = 48$ h

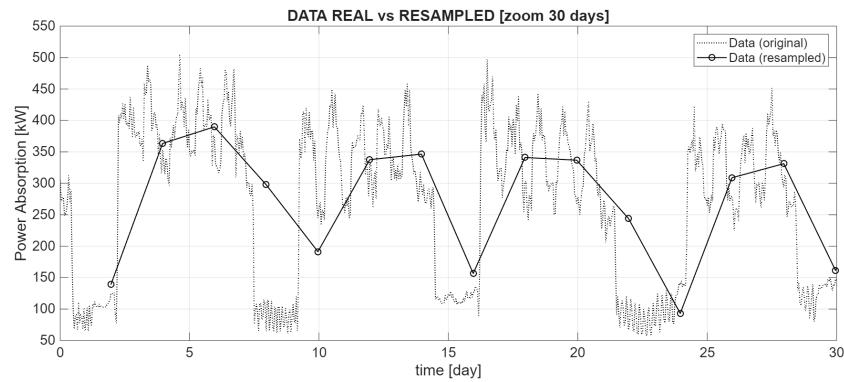


Figura 4.27: data REAL vs RESAMPLED [zoom 30 days]

Linear Regression (LR)

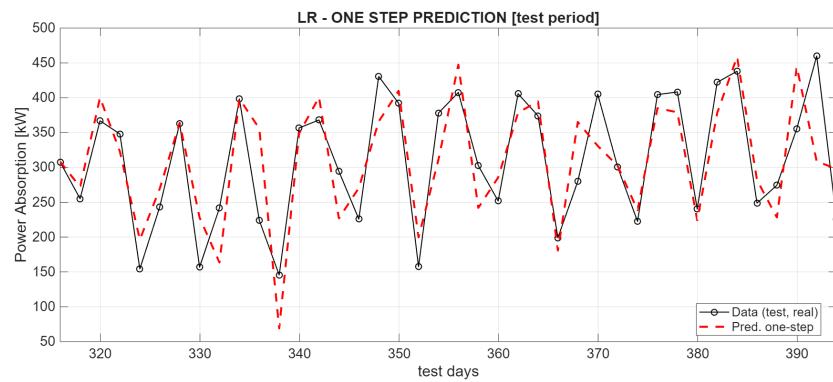


Figura 4.28: LR - one step prediction $t_r = 48$ h [testing period]

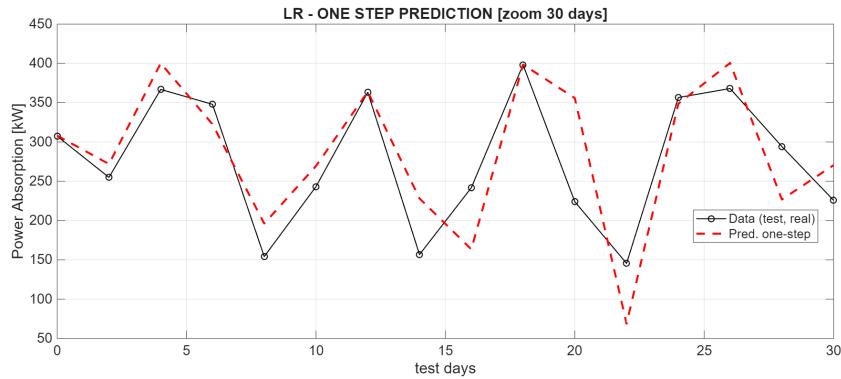


Figura 4.29: LR - one step prediction $t_r = 48$ h [zoom 30 days]

Metrica	Valore
RMSE	55,13
MAE	43,70

Tabella 4.11: Metriche di valutazione LR – one step prediction ($t_{resampling} = 48$ h)

Gaussian Process Regression (GPR)

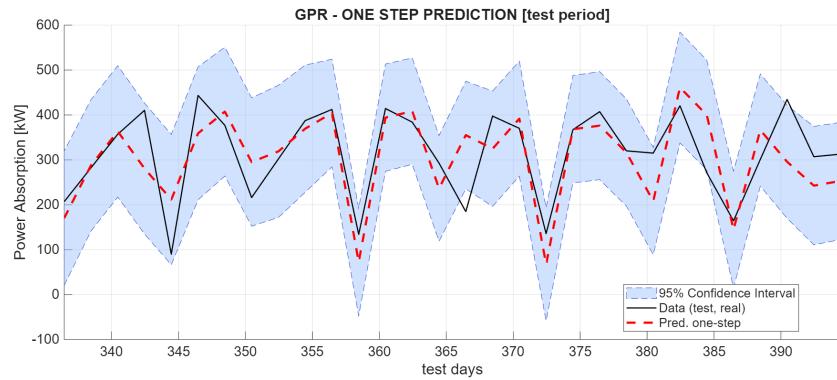


Figura 4.30: LR - one step prediction $t_r = 48$ h [testing period]

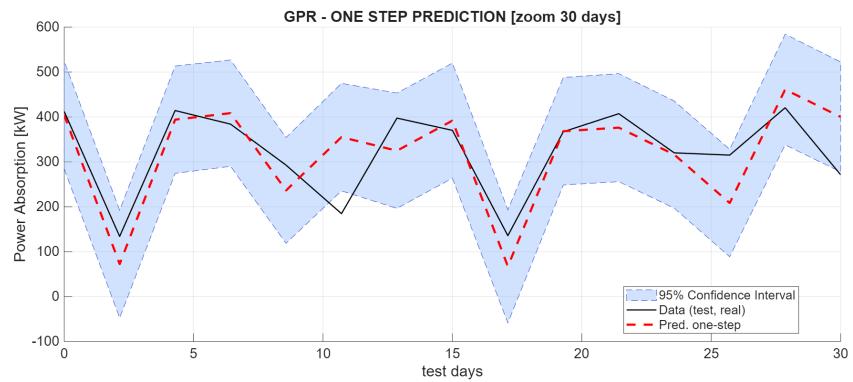


Figura 4.31: GPR - one step prediction $t_r = 48$ h [zoom 30 days]

Metrica	Valore
RMSE	72,48
MAE	56,40

Tabella 4.12: Metriche di valutazione GPR – one step prediction ($t_resampling = 48$ h)

4.4 Previsione multi-step ahead

In questa sezione vengono presentati i risultati relativi alla previsione *multi-step ahead*, ovvero la stima dei valori futuri dell'assorbimento energetico su più passi temporali consecutivi rispetto al dato corrente.

Analogamente a quanto fatto per la previsione *one step ahead*, anche in questo caso il numero di regressori è mantenuto costante e pari a $p = 10$, in conformità con quanto definito nelle sezioni precedenti.

Oltre al parametro dei lag, che rimane invariato, in questa fase viene condotta un'analisi più articolata che coinvolge due aspetti distinti: da un lato, il parametro di *resampling* ($t_{\text{resampling}}$), già esaminato nel Capitolo 4.3.1, il quale consente di valutare come la variazione della frequenza di campionamento influenzi le prestazioni predittive dei modelli; dall'altro, il parametro n , che definisce il numero di passi temporali in avanti da prevedere, e che permette di estendere la capacità di previsione dei modelli su orizzonti temporali più lunghi.

I risultati sono quindi organizzati in modo da mostrare, per ciascuna configurazione dei parametri considerati, le rappresentazioni grafiche delle previsioni ottenute e le relative metriche di valutazione.

Vengono ora riportati i risultati della previsione *multi-step ahead* per i due metodi considerati, corrispondenti al caso di riferimento in cui il parametro di orizzonte temporale è fissato a $n = 4$. Tale configurazione, mantenendo costanti i restanti parametri, consente di valutare la capacità dei modelli di stimare con accuratezza l'andamento del consumo energetico su un intervallo temporale di quattro ore in avanti, rappresentando la base di confronto per le successive analisi di variazione dei parametri.

Linear Regression (LR)

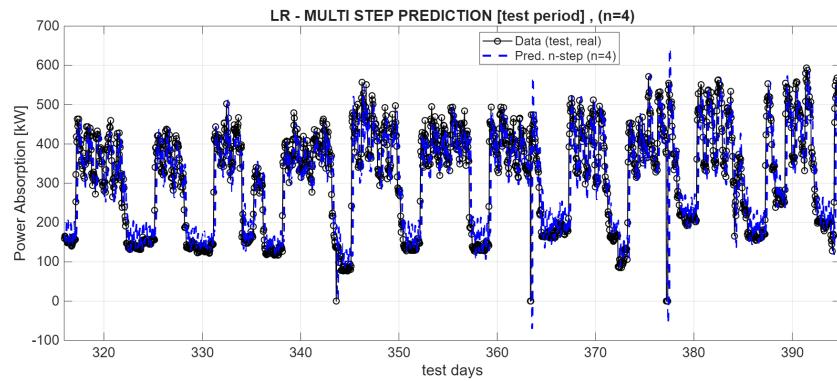


Figura 4.32: LR - multi step prediction [testing period]

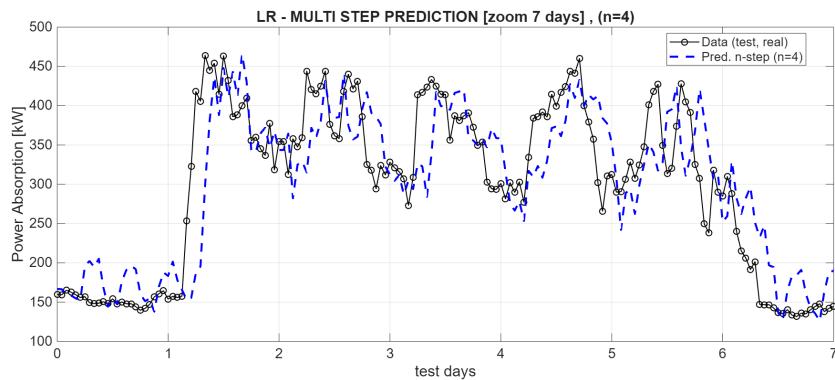


Figura 4.33: LR - multi step prediction [zoom 7 days]

Metrica	Valore
RMSE	75,71
MAE	49,24

Tabella 4.13: Metriche di valutazione LR – multi step prediction

Gaussian Process Regression (GPR)

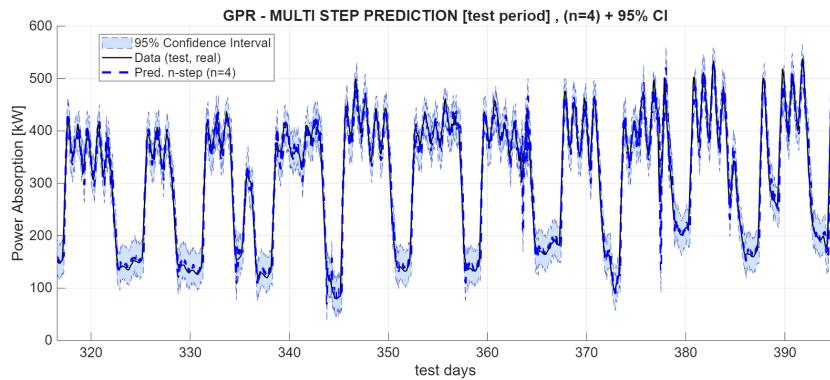


Figura 4.34: GPR - multi step prediction [testing period]

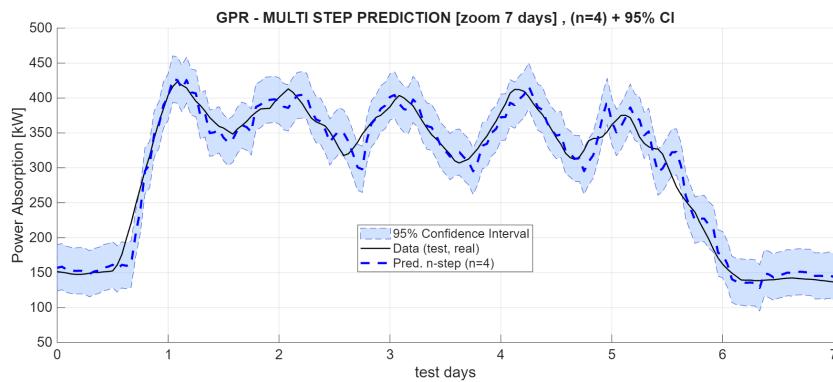


Figura 4.35: GPR - multi step prediction [zoom 7 days]

Metrica	Valore
RMSE	19,55
MAE	12,90

Tabella 4.14: Metriche di valutazione GPR – multi step prediction

4.4.1 Analisi del parametro di resampling

In questa sezione vengono presentati i risultati relativi alla variazione del parametro $t_{resampling}$, che regola l'intervallo temporale con cui i dati di consumo vengono mediati prima dell'elaborazione.

In questo caso, il parametro di orizzonte temporale è mantenuto fisso a $n = 4$, come nel caso base precedentemente analizzato, mentre i valori scelti per il parametro di *resampling* risultano differenti rispetto alla previsione *one step ahead*. In particolare, verranno considerati e analizzati i casi con $t_{resampling}$ pari a 4, 6 e 8 ore, scelte che consentono di esaminare l'effetto del processo di resampling in condizioni che mantengano un livello di accuratezza accettabile, evitando valori troppo elevati per i quali i modelli perderebbero la capacità di rappresentare in modo efficace le dinamiche temporali del segnale.

Per ciascun valore considerato, viene mostrato innanzitutto un confronto tra il segnale originale e quello preprocessato, al fine di evidenziare come il processo di resampling influenzi la forma e la densità del dataset nel dominio temporale. Inoltre, sono riportati grafici con zoom su un intervallo rappresentativo, utili a visualizzare con maggiore chiarezza le differenze introdotte dal preprocessing.

Successivamente, sono presentati i risultati delle previsioni *multi-step ahead* ottenute dai modelli impiegati (prima quelli del modello LR e successivamente quelli del modello GPR, il quale include anche la rappresentazione dell'intervallo di confidenza nei grafici dedicati), mantenendo invariati tutti gli altri parametri di configurazione. Infine, le metriche di valutazione adottate, RMSE e MAE, sono sintetizzate nelle tabelle poste al termine dei grafici di previsione, fornendo un quadro quantitativo dell'accuratezza dei modelli per ciascun valore di $t_{resampling}$.

- Parametro di resampling $t_{resampling} = 4 \text{ h}$

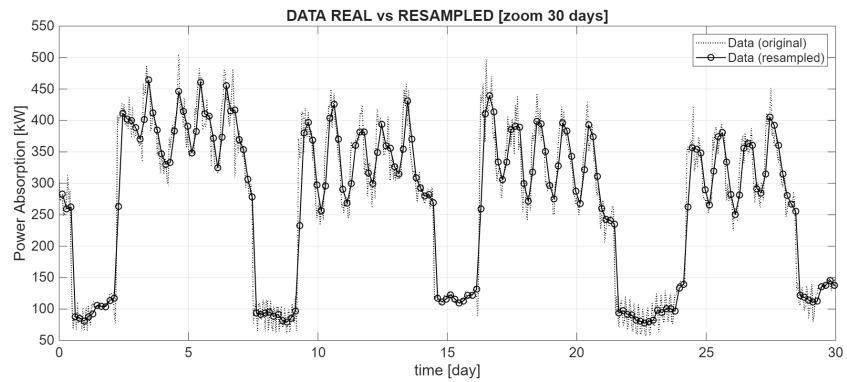


Figura 4.36: data REAL vs RESAMPLED [zoom 30 days]

Linear Regression (LR)

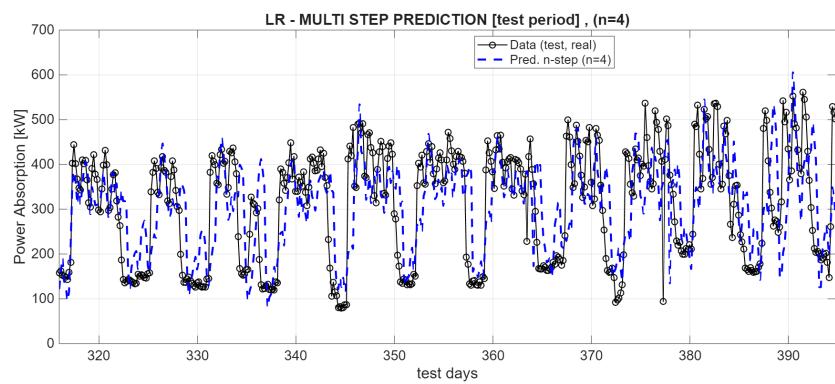


Figura 4.37: LR - multi step prediction $t_r = 4 \text{ h}$ [testing period]

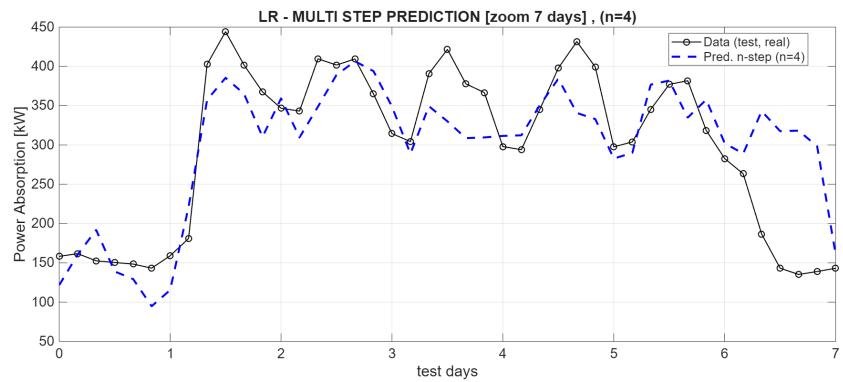


Figura 4.38: LR - multi step prediction $t_r = 4$ h [zoom 7 days]

Metrica	Valore
RMSE	104,91
MAE	79,46

Tabella 4.15: Metriche di valutazione LR – multi step prediction ($t_resampling = 4$ h)

Gaussian Process Regression (GPR)

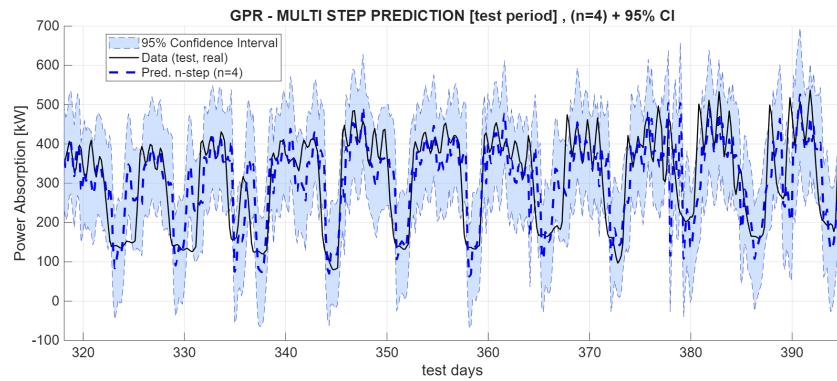


Figura 4.39: GPR - multi step prediction $t_r = 4$ h [testing period]

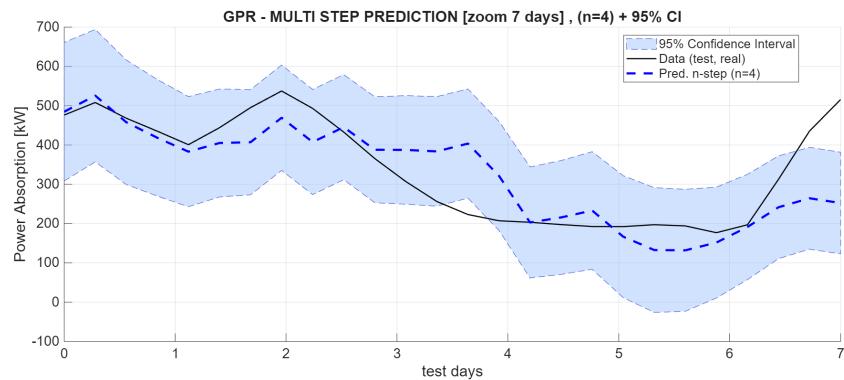


Figura 4.40: GPR - mutli step prediction $t_r = 4$ h [zoom 7 days]

Metrica	Valore
RMSE	84,13
MAE	59,70

Tabella 4.16: Metriche di valutazione GPR – multi step prediction ($t_{resampling} = 4$ h)

- Parametro di resampling $t_{resampling} = 6$ h

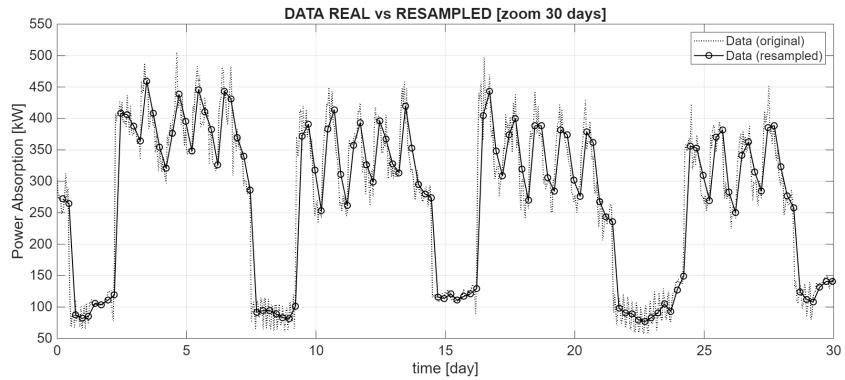


Figura 4.41: data REAL vs RESAMPLED [zoom 30 days]

Linear Regression (LR)

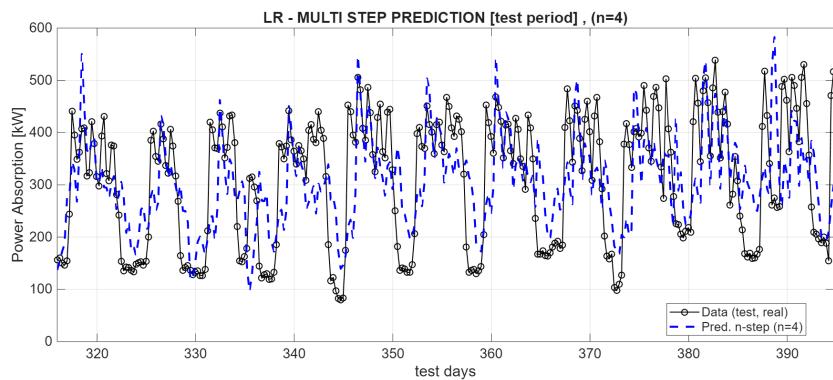


Figura 4.42: LR - multi step prediction $t_r = 6$ h [testing period]

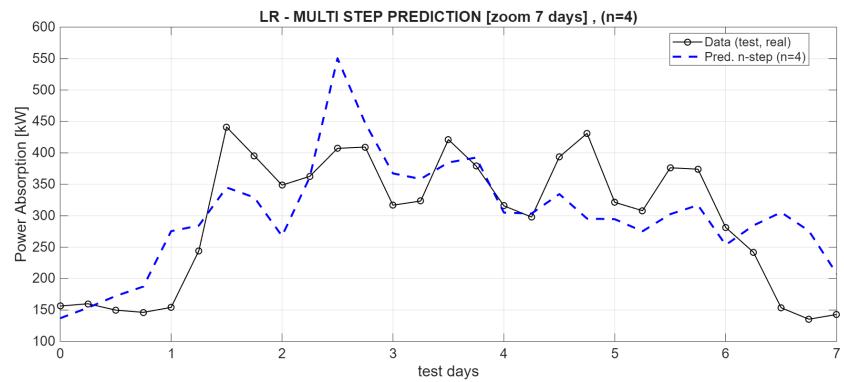


Figura 4.43: LR - multi step prediction $t_r = 6$ h [zoom 7 days]

Metrica	Valore
RMSE	99,44
MAE	80,74

Tabella 4.17: Metriche di valutazione LR – multi step prediction ($t_resampling = 6$ h)

Gaussian Process Regression (GPR)

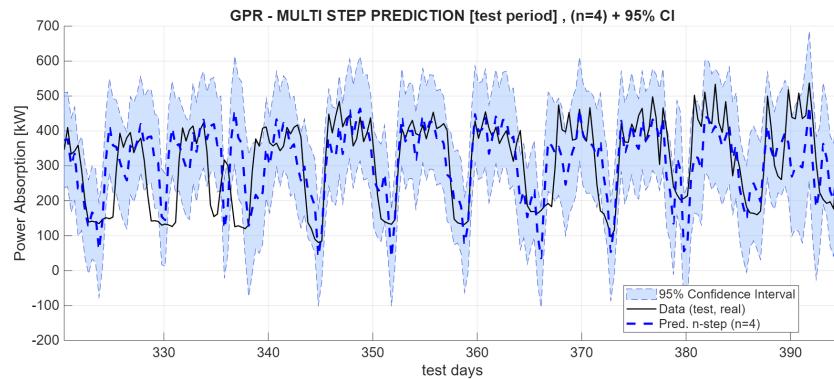


Figura 4.44: GPR - multi step prediction $t_r = 6$ h [testing period]

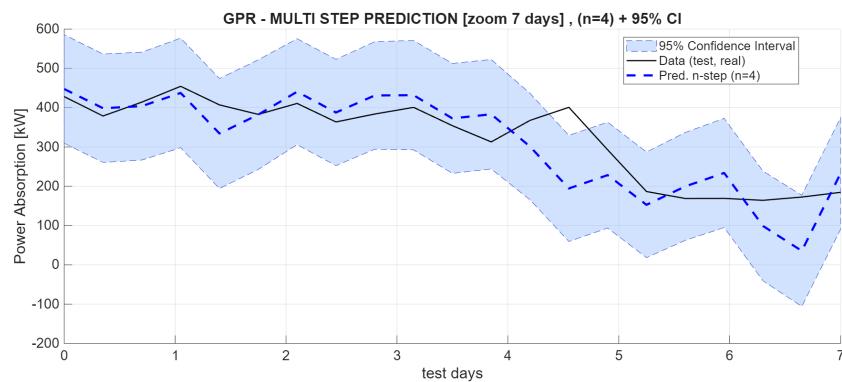


Figura 4.45: GPR - mutli step prediction $t_r = 6$ h [zoom 7 days]

Metrica	Valore
RMSE	93,57
MAE	69,56

Tabella 4.18: Metriche di valutazione GPR – multi step prediction ($t_{resampling} = 6$ h)

- Parametro di resampling $t_{resampling} = 8 \text{ h}$

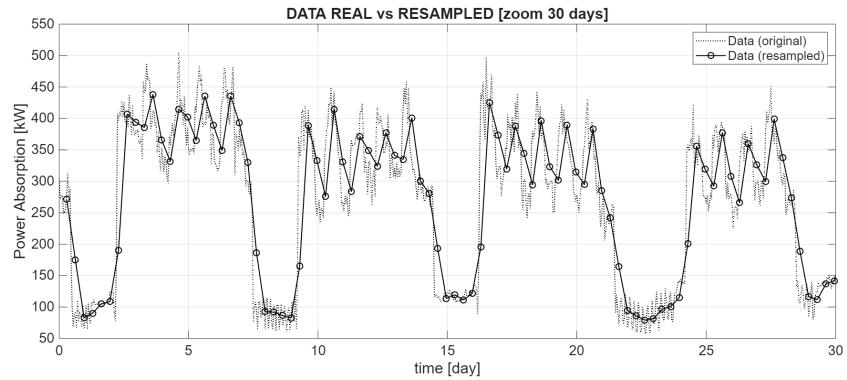


Figura 4.46: data REAL vs RESAMPLED [zoom 30 days]

Linear Regression (LR)

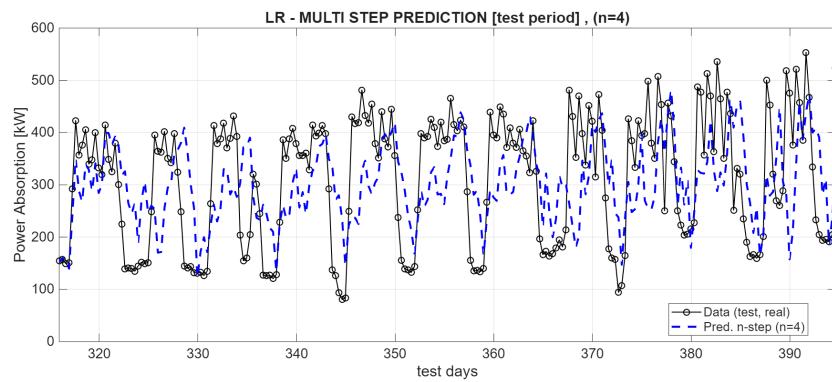


Figura 4.47: LR - multi step prediction $t_r = 8 \text{ h}$ [testing period]

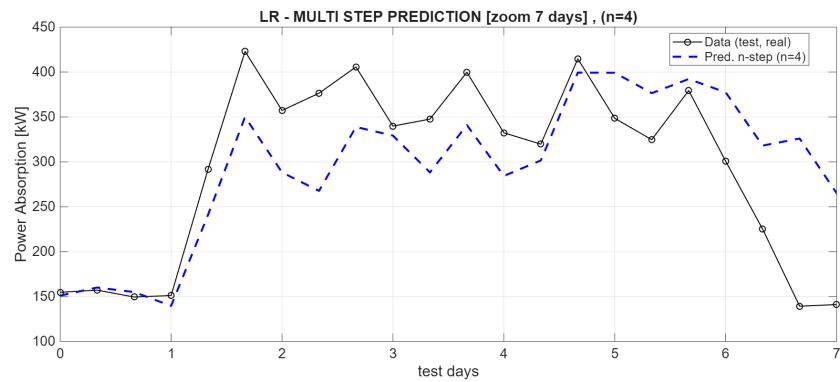


Figura 4.48: LR - multi step prediction $t_r = 8$ h [zoom 7 days]

Metrica	Valore
RMSE	113,29
MAE	94,06

Tabella 4.19: Metriche di valutazione LR – multi step prediction ($t_resampling = 8$ h)

Gaussian Process Regression (GPR)

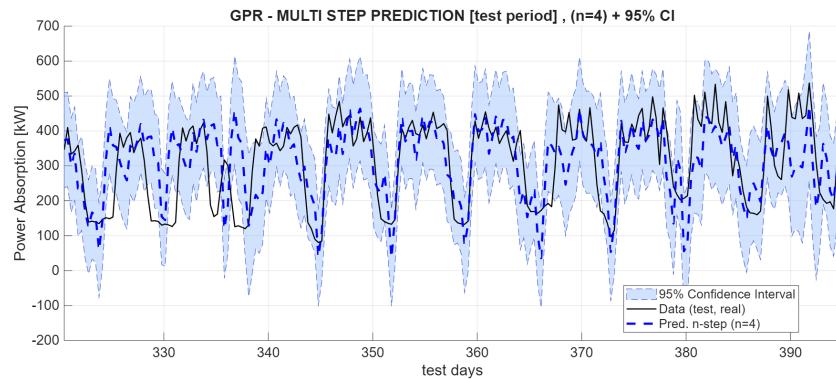


Figura 4.49: GPR - multi step prediction $t_r = 8$ h [testing period]

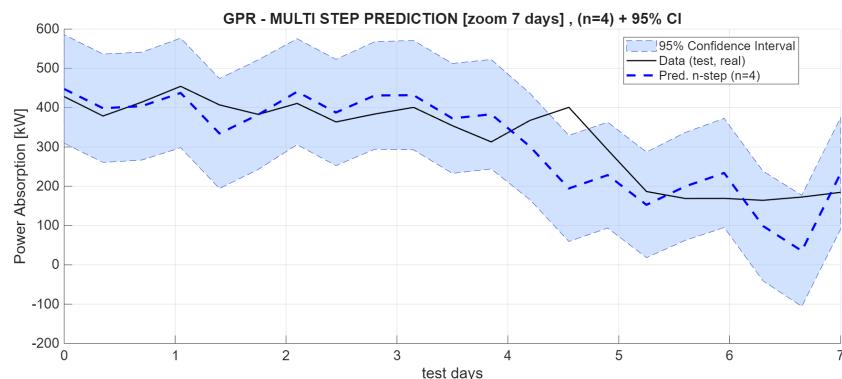


Figura 4.50: GPR - mutli step prediction $t_r = 8$ h [zoom 7 days]

Metrica	Valore
RMSE	98,46
MAE	74,06

Tabella 4.20: Metriche di valutazione GPR – multi step prediction ($t_{resampling} = 8$ h)

4.4.2 Analisi del parametro di orizzonte temporale di previsione

In questa sezione vengono analizzati gli effetti della variazione del parametro di orizzonte temporale di previsione, indicato con n , che rappresenta il numero di passi in avanti per i quali il modello effettua la stima dei valori futuri della variabile di interesse. Come già discusso nei capitoli precedenti, il parametro n controlla quindi la profondità temporale della previsione, definendo la capacità del modello di mantenere accuratezza su orizzonti previsivi più o meno estesi.

Nel presente studio, verranno riportati i risultati ottenuti dai due modelli di previsione al variare del parametro n , considerando i seguenti valori: 3, 4, 6 e 8. Il valore $n = 4$ corrisponde al caso base già presentato nella sezione precedente, e sarà utilizzato come riferimento per il confronto con gli altri scenari.

Per garantire coerenza metodologica e isolare l'effetto del solo parametro di previsione, gli altri parametri di configurazione vengono mantenuti invariati: il numero di regressori è fissato a $p = 10$ e non viene applicato alcun processo di resampling, ossia $t_r = 1$.

Successivamente, per ciascun valore di n considerato, verranno presentati i risultati delle simulazioni ottenute con entrambi i modelli, illustrando attraverso rappresentazioni grafiche il comportamento delle previsioni a più passi in avanti e riportando le corrispondenti metriche di valutazione (RMSE e MAE). Tali risultati consentiranno di osservare come la qualità predittiva dei modelli tenda a variare in funzione dell'aumento dell'orizzonte temporale di previsione.

Linear Regression (LR)

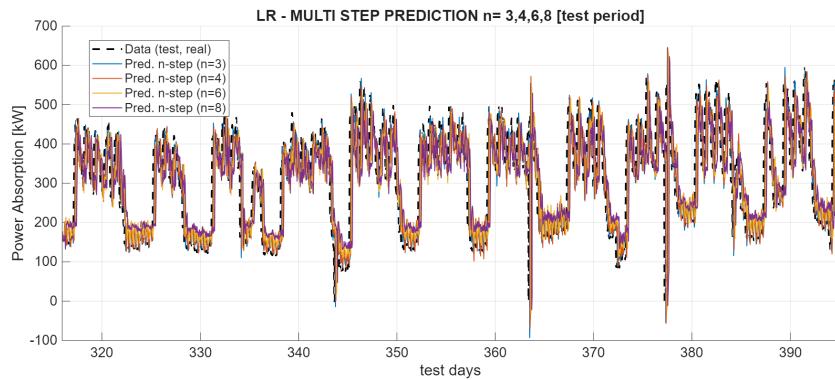


Figura 4.51: LR - multi step prediction $n = 3, 4, 6, 8$ [testing period]

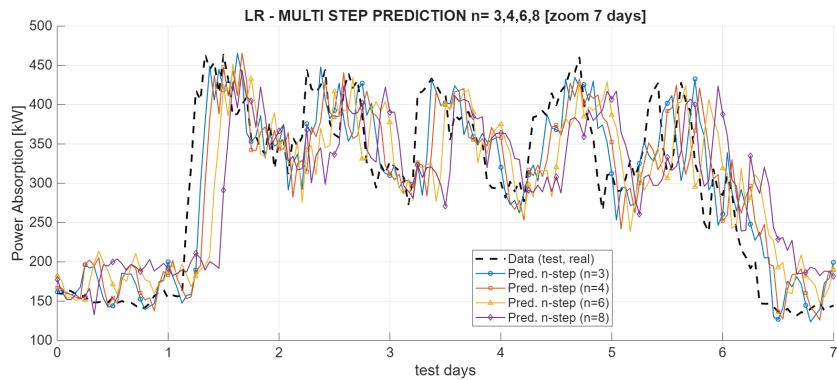


Figura 4.52: LR - multi step prediction $n = 3, 4, 6, 8$ [zoom 7 days]

Metrica	$n=3$	$n=4$	$n=6$	$n=8$
RMSE	66,05	75,71	89,28	98,69
MAE	40,98	49,24	62,39	72,86

Tabella 4.21: Metriche di valutazione LR – multi step prediction, variazione orizzonte temporale n

Gaussian Process Regression (GPR)

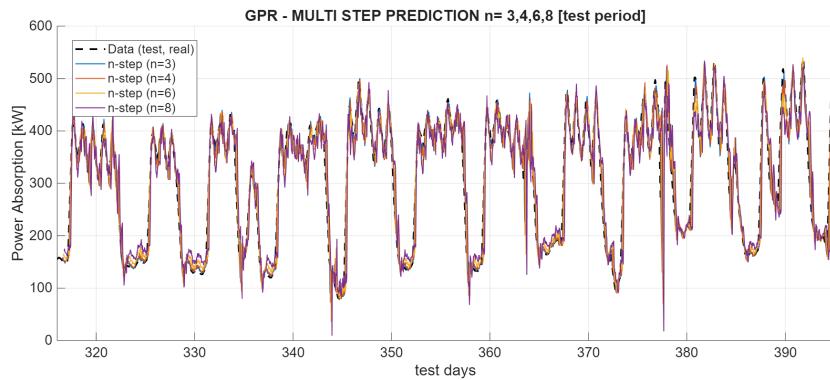


Figura 4.53: GPR - multi step prediction $n = 3, 4, 6, 8$ [testing period]

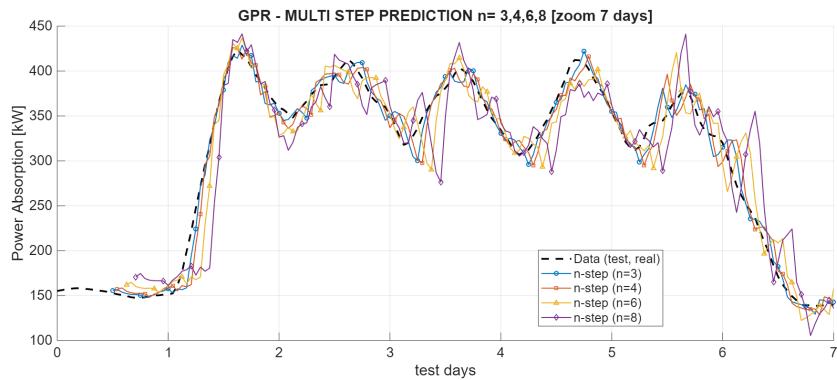


Figura 4.54: GPR - multi step prediction $n = 3, 4, 6, 8$ [zoom 7 days]

Metrica	$n=3$	$n=4$	$n=6$	$n=8$
RMSE	14,45	19,55	30,84	42,93
MAE	9,44	12,90	20,62	29,22

Tabella 4.22: Metriche di valutazione GPR – multi step prediction, variazione orizzonte temporale n

Capitolo 5

Analisi e commento dei risultati

In questo capitolo vengono presentati, analizzati e commentati i risultati ottenuti dalle simulazioni dei modelli di previsione applicati ai dati di consumo energetico dell’azienda oggetto di studio. L’obiettivo è duplice: da un lato valutare le prestazioni dei singoli modelli in differenti condizioni sperimentali, dall’altro effettuare un confronto diretto tra di essi al fine di evidenziare punti di forza, limiti e differenze di comportamento.

L’analisi è articolata in due sezioni principali, dedicate rispettivamente alla previsione *one-step ahead* e alla previsione *multi-step ahead*. Per ciascun caso verranno inizialmente esaminati i risultati relativi al cosiddetto *caso base*, per poi considerare gli effetti dell’introduzione del *resampling* dei dati a diversi intervalli temporali. Nel capitolo dedicato alla previsione *multi-step* sarà inoltre condotta un’analisi specifica sull’influenza del parametro di orizzonte temporale di previsione, al fine di comprendere come la distanza temporale incida sulle prestazioni predittive dei modelli.

Le valutazioni saranno supportate sia da indicatori quantitativi, tramite l’utilizzo delle metriche di errore introdotte nei capitoli precedenti, sia da rappresentazioni grafiche che consentiranno di cogliere in modo visivo l’andamento e la qualità delle previsioni. L’approccio seguito permetterà quindi di delineare un quadro complessivo e comparativo delle capacità predittive dei modelli adottati, fornendo le basi per una discussione critica dei risultati ottenuti.

5.1 Analisi risultati previsione *one-step-ahead*

Analisi del caso base

La prima fase dell'analisi dei risultati riguarda la previsione one-step-ahead, ovvero la stima del valore di potenza assorbita a un'ora di distanza dal dato corrente. Tale configurazione rappresenta il punto di partenza per la valutazione delle prestazioni dei modelli, poiché consente di verificare la loro capacità di catturare la dinamica immediata del segnale di consumo, in condizioni di previsione a brevissimo termine e senza l'introduzione di errori cumulativi.

Linear Regression (LR)

Nel caso standard, riportato nelle figure 4.3 e 4.4, il modello di regressione lineare mostra una buona capacità di seguire l'andamento generale della serie temporale di test, riproducendo con coerenza le variazioni di potenza e mantenendo un trend complessivamente in linea con quello reale. Dall'osservazione dei grafici si nota tuttavia che, pur mantenendo una forma simile al segnale di riferimento, il modello tende a presentare una certa rigidità nella risposta: i picchi massimi e minimi risultano infatti leggermente smussati, e ciò indica una limitata capacità nel rappresentare oscillazioni repentine o variazioni improvvise di carico. Questo comportamento è coerente con la natura lineare del modello, che tende a privilegiare la componente deterministica e a mediare gli scostamenti più accentuati.

Dal punto di vista quantitativo, i valori delle metriche riportati in tabella 4.1 confermano tale osservazione: con RMSE pari a 38,44 e MAE pari a 22,74, il modello evidenzia un errore medio moderato, ma non trascurabile. Il valore dell'RMSE, sensibilmente più elevato rispetto al MAE, indica inoltre la presenza di alcune deviazioni più marcate, che incidono sul calcolo dell'errore quadratico medio. In sintesi, il modello lineare riesce a catturare adeguatamente il comportamento complessivo del consumo, ma mostra difficoltà nel riprodurre fedelmente le fluttuazioni di breve durata e i picchi istantanei, risultando quindi più adatto a previsioni di tendenza che a stime puntuali ad alta precisione.

Gaussian Process Regression (GPR)

Le figure 4.5 e 4.6 mostrano invece i risultati ottenuti mediante il modello di Gaussian Process Regression. In questo caso, il comportamento del modello risulta notevolmente più accurato: la sovrapposizione tra i valori previsti e quelli reali appare pressoché completa per gran parte del periodo di test, segno di una capacità predittiva più raffinata e di una maggiore sensibilità alle variazioni locali del segnale. La presenza delle bande di confidenza al 95%, ben visibili nel grafico, evidenzia inoltre la capacità del modello di fornire una stima dell'incertezza associata a ciascuna previsione, che risulta generalmente contenuta e stabile nel tempo.

La struttura non parametrica del GPR, basata sul kernel di tipo squared exponential, consente di catturare relazioni non lineari tra i campioni temporali, offrendo una modellazione più flessibile rispetto alla LR. Questo si traduce nella capacità di seguire in modo più preciso le variazioni di carico anche su orizzonti temporali brevi, senza incorrere nei fenomeni di “appiattimento” osservati nel modello lineare.

Dal punto di vista numerico, i valori di $RMSE = 4,43$ e $MAE = 2,80$, riportati in tabella 4.2, confermano l'elevata accuratezza del modello. Gli errori sono di un ordine di grandezza inferiore rispetto a quelli ottenuti dalla regressione lineare, e il modesto scarto tra $RMSE$ e MAE suggerisce un comportamento regolare, privo di picchi di errore anomali. L'approccio probabilistico del GPR consente inoltre di gestire in modo efficace il rumore nei dati, migliorando la stabilità del modello anche in presenza di irregolarità o disturbi di misura.

Confronto tra i due modelli

Dal confronto diretto tra i due modelli emerge in modo evidente la superiorità del Gaussian Process Regression rispetto alla Linear Regression. Quest'ultima, pur garantendo una rappresentazione coerente del trend generale, non riesce a catturare la complessità e la variabilità del consumo orario dello stabilimento, limitandosi a una descrizione lineare del fenomeno. Il GPR, al contrario, mostra una maggiore capacità di adattamento, riuscendo a seguire fedelmente la dinamica del segnale e a restituire previsioni di elevata precisione, con un errore medio estremamente ridotto.

In termini di prestazioni, la differenza tra le metriche è netta: il valore di $RMSE$ del GPR è inferiore di quasi un ordine di grandezza rispetto a quello della LR, confermando che il modello basato su kernel risulta molto più adatto a trattare segnali con variazioni non lineari e con una componente stocastica significativa. Anche le bande di confidenza introdotte dal GPR costituiscono un valore aggiunto importante, poiché consentono una valutazione dell'affidabilità della previsione, assente nel modello lineare.

Analisi del parametro di resampling

Successivamente, è stata condotta un'analisi sistematica dell'effetto del resampling temporale ($t_{resampling}$), con l'obiettivo di valutare come la variazione della frequenza di campionamento influenzi la stabilità e la precisione delle previsioni. I casi analizzati corrispondono ai valori di $t_{resampling}$ pari a 4, 8, 12, 24 e 48 ore, mantenendo invariati gli altri parametri del modello ($p = 10$ e $n = 1$).

Linear Regression (LR)

L'osservazione dei grafici mostra come l'aumento del periodo di resampling produca un effetto di progressiva regolarizzazione del segnale, chiaramente visibile nei dataset riportati nelle figure 4.8–4.9, 4.13–4.14, 4.18–4.19, 4.23–4.24 e 4.28–4.29. Nei casi con $t_r = 4$ h (figure 4.8 e 4.9) il segnale appare già più regolare rispetto al caso base: le oscillazioni locali vengono attenuate e i picchi di potenza risultano meno marcati. Tale effetto diventa più evidente con $t_r = 8$ h (figure 4.13 e 4.14) e $t_r = 12$ h (figure 4.18 e 4.19), dove il modello segue ancora l'andamento generale ma mostra una crescente tendenza ad appiattire le variazioni giornaliere. Nei casi estremi con $t_r = 24$ h (figure 4.23 e 4.24) e $t_r = 48$ h (figure 4.28 e 4.29), la previsione risulta quasi perfettamente sovrapposta alla media dei dati, a testimonianza di una forte perdita di dettaglio informativo.

Dal punto di vista numerico, le metriche di tabelle 4.3, 4.5, 4.7, 4.9, 4.11 mostrano un andamento non monotono: per $t_r = 4$ h si osserva un incremento dell'errore rispetto al caso base (RMSE = 55,50; MAE = 34,60), dovuto al fatto che l'aggregazione su finestre più lunghe riduce la sensibilità del modello alle variazioni di breve periodo. Gli errori aumentano ulteriormente per $t_r = 8$ h e $t_r = 12$ h, raggiungendo valori superiori a 60 per l'RMSE, per poi stabilizzarsi attorno a 55–60 nelle configurazioni più spinte ($t_r = 24$ h e $t_r = 48$ h). Ciò suggerisce che, per la LR, il resampling non apporta un beneficio concreto in termini di accuratezza: l'effetto di *smoothing* sui dati attenua la variabilità ma compromette la capacità del modello di adattarsi alla dinamica temporale reale, già limitata dalla struttura lineare.

Gaussian Process Regression (GPR)

Il comportamento del modello GPR al variare di t_r presenta caratteristiche differenti. Nel caso $t_r = 4$ h (figure 4.10–4.11), il segnale mostra un leggero effetto di regolarizzazione, ma il modello riesce comunque a seguire fedelmente le variazioni principali, con bande di confidenza strette e stabili. Con $t_r = 8$ h (figure 4.15–4.16) e $t_r = 12$ h (figure 4.20–4.21), il segnale appare visibilmente più smussato e le bande di confidenza si allargano, segno di un incremento dell'incertezza. Infine, per $t_r = 24$ h (figure 4.25–4.26) e $t_r = 48$ h (figure 4.30–4.31), la previsione mantiene una buona coerenza generale ma tende a sottostimare le variazioni più brusche.

L'analisi delle metriche (4.4, 4.6, 4.8, 4.10, 4.12) confermano tale comportamento:

per $t_r = 4$ h si ottiene un RMSE di 20,41 e MAE di 12,75, valori già più elevati rispetto al caso base ma ancora nettamente migliori di quelli della LR. L'errore cresce progressivamente con l'aumento di t_r , raggiungendo RMSE = 72,48 e MAE = 56,40 per $t_r = 48$ h. Nonostante il peggioramento, il GPR mantiene un comportamento regolare e stabile, con una perdita di accuratezza proporzionale alla riduzione del contenuto informativo del dataset.

Confronto tra i modelli

Il confronto tra i due modelli evidenzia che, anche nei casi con resampling, il Gaussian Process Regression mantiene prestazioni decisamente superiori rispetto alla Linear Regression. In particolare, per intervalli di resampling moderati (4–8 h), il GPR riesce ancora a fornire previsioni con errore medio contenuto, mentre la LR subisce un peggioramento marcato e un'ulteriore perdita di aderenza ai dati.

L'effetto del resampling, dunque, si manifesta in modo diverso per i due modelli: per la LR rappresenta una limitazione aggiuntiva, che riduce ulteriormente la precisione del modello; per il GPR, invece, pur comportando un naturale aumento dell'errore, non compromette la stabilità complessiva del processo predittivo, confermando la maggiore robustezza e flessibilità del modello probabilistico.

5.2 Analisi risultati previsione *multi-step-ahead*

Analisi del caso base

La seconda parte dell'analisi riguarda la previsione multi-step-ahead, ovvero la stima dei valori futuri dell'assorbimento energetico su più passi temporali consecutivi rispetto al dato corrente. Questa configurazione consente di valutare la capacità dei modelli di mantenere accuratezza e stabilità quando l'orizzonte temporale della previsione si estende oltre il passo immediato, condizione di maggiore complessità per entrambi gli approcci analizzati.

Linear Regression (LR)

Nel caso di riferimento, con orizzonte temporale $n = 4$, i risultati del modello di regressione lineare, mostrati nelle figure 4.32 e 4.33, evidenziano un andamento complessivamente coerente con quello della serie reale, ma con una crescente perdita di precisione al crescere del passo di previsione. Il modello riesce a seguire il trend generale del segnale, ma tende a introdurre uno sfasamento temporale tra i valori previsti e quelli osservati, soprattutto nei periodi caratterizzati da variazioni rapide di potenza. Tale comportamento è tipico dei modelli lineari quando la previsione si estende su più step, poiché gli errori generati ai passi precedenti si propagano nei successivi, amplificandosi progressivamente.

Dal punto di vista quantitativo, le metriche riportate in tabella 4.13 mostrano valori di $RMSE = 75,71$ e $MAE = 49,24$, significativamente più elevati rispetto al caso one-step. L'incremento dell'errore è coerente con la complessità maggiore del problema e conferma la tendenza della LR a perdere affidabilità con l'estensione dell'orizzonte di previsione. Il modello conserva tuttavia una buona stabilità, non mostrando oscillazioni spurie o divergenze, ma produce stime eccessivamente mediate, con ridotta capacità di catturare le fluttuazioni di breve periodo.

Gaussian Process Regression (GPR)

Il modello di Gaussian Process Regression, nelle figure 4.34 e 4.35, mostra prestazioni nettamente migliori anche in configurazione multi-step. La curva di previsione segue con maggiore fedeltà l'andamento dei valori reali, mantenendo una buona aderenza anche nelle aree caratterizzate da variazioni repentine. Le bande di confidenza al 95%, presenti nei grafici, risultano più ampie rispetto al caso one-step, segno di un aumento dell'incertezza associata alla previsione, ma restano comunque contenute e coerenti con la variabilità effettiva dei dati.

I valori delle metriche, riportati in tabella 4.14 ($RMSE = 19,55$; $MAE = 12,90$), confermano l'elevata precisione del modello anche su orizzonti temporali multipli. Sebbene gli errori risultino leggermente superiori rispetto alla previsione a un passo, la differenza è contenuta, indice della capacità del GPR di mantenere un buon livello di generalizzazione e stabilità anche al crescere dell'orizzonte di previsione. Il

comportamento del modello mostra inoltre una maggiore regolarità rispetto alla LR, con una distribuzione dell'errore più omogenea e assenza di deviazioni sistematiche.

Confronto tra i due modelli

Dal confronto tra i due approcci emerge nuovamente la superiorità del modello GPR rispetto alla regressione lineare. L'errore medio del GPR risulta di oltre tre volte inferiore rispetto a quello della LR, e la rappresentazione grafica evidenzia una maggiore coerenza con il segnale reale. La Linear Regression mostra limiti strutturali evidenti, legati alla natura deterministica del modello e alla propagazione degli errori tra i passi successivi. Il GPR, grazie alla propria formulazione probabilistica e alla funzione di kernel, è in grado di adattarsi meglio alle relazioni non lineari e di mantenere una maggiore stabilità predittiva anche su orizzonti temporali estesi. Nel complesso, il comportamento del GPR risulta più realistico e continuo, mentre la LR tende a restituire curve più rigide e approssimate, con un progressivo scostamento dai valori effettivi.

Analisi del parametro di resampling

Analogamente a quanto fatto per la previsione one-step-ahead, è stata condotta un'analisi della variazione del parametro di resampling temporale ($t_{resampling}$), mantenendo costante l'orizzonte di previsione a $n = 4$. L'obiettivo è valutare l'effetto della riduzione della frequenza di campionamento sull'accuratezza dei modelli, considerando i casi con $t_{resampling} = 4, 6$ e 8 ore, scelti per garantire un equilibrio tra perdita di dettaglio e stabilità numerica.

Linear Regression (LR)

L'osservazione dei grafici mostra come, anche nel caso di previsione multi-step, l'aumento del periodo di resampling produca un effetto di progressiva regolarizzazione del segnale, chiaramente visibile nei dataset riportati nelle figure 4.36–4.41, 4.46. Nel caso con $t_r = 4$ h (figure 4.37 e 4.38), il segnale risulta già più uniforme rispetto al caso base, con una sensibile riduzione dei picchi e delle fluttuazioni locali. Con $t_r = 6$ h (figure 4.42 e 4.43) e $t_r = 8$ h (figure 4.47 e 4.48), l'effetto di mediazione diventa ancora più evidente: il dataset assume un andamento più liscio e continuo, mentre la previsione del modello lineare segue il profilo medio ma perde progressivamente sensibilità alle variazioni di breve periodo. Di conseguenza, il modello tende a riprodurre correttamente solo il trend generale, mostrando una risposta più rigida e un ritardo nella cattura delle oscillazioni più rapide.

Dal punto di vista numerico, le metriche riportate nelle tabelle 4.15, 4.17 e 4.19 evidenziano un progressivo incremento dell'errore all'aumentare di t_r . Per $t_r = 4$ h si ottiene RMSE = 104,91 e MAE = 79,46, mentre per $t_r = 6$ h i valori crescono ulteriormente, fino a raggiungere RMSE = 113,29 e MAE = 94,06 per $t_r = 8$ h. L'andamento delle prestazioni conferma che l'aumento dell'intervallo di campionamento accentua la perdita di informazione temporale, riducendo la capacità del modello lineare di catturare la variabilità reale del processo. Ne risulta una previsione più stabile ma meno precisa, soprattutto su orizzonti temporali multipli, dove gli errori tendono a propagarsi tra i passi successivi.

Gaussian Process Regression (GPR)

Il comportamento del modello GPR al variare di t_r mostra una maggiore stabilità e capacità di adattamento. Nei casi con $t_r = 4$ h (figure 4.37–4.38), il modello segue accuratamente il profilo dei dati reali, mostrando solo un leggero effetto di regolarizzazione dovuto al ricampionamento, mentre le bande di confidenza restano contenute e indicano una stima dell'incertezza affidabile. Con $t_r = 6$ h (figura 4.44–4.45) e $t_r = 8$ h (figura 4.49–4.50), la regolarizzazione del segnale diventa più evidente: le oscillazioni si riducono e il modello tende a produrre previsioni più lisce, con un allargamento progressivo delle bande di confidenza, che testimonia l'aumento dell'incertezza ma senza perdita di coerenza complessiva. Il GPR riesce comunque a mantenere una buona aderenza ai dati e una notevole capacità di

generalizzazione, anche in presenza di una minore densità informativa.

Dal punto di vista quantitativo, le metriche riportate in tabelle 4.16, 4.18 e 4.20 confermano l'elevata stabilità del modello: per $t_r = 4$ h si ottiene RMSE = 84,13 e MAE = 59,70, che aumentano moderatamente fino a RMSE = 98,46 e MAE = 74,06 per $t_r = 8$ h. L'incremento dell'errore risulta quindi contenuto rispetto alla LR, e la regolarità del comportamento del GPR dimostra la sua maggiore robustezza nei confronti della perdita di risoluzione temporale dovuta al resampling. Il modello probabilistico conserva un buon compromesso tra precisione e stabilità, mostrandosi adatto anche a contesti di previsione multi-step con frequenze di campionamento ridotte.

Confronto tra i modelli

Il confronto complessivo mostra che, anche nella configurazione multi-step, il GPR mantiene prestazioni nettamente superiori rispetto alla LR. Quest'ultima subisce un degrado più marcato con l'aumento dell'intervallo di resampling, dovuto alla somma degli effetti di perdita di informazione e propagazione dell'errore. Il GPR, pur risentendo della riduzione di dettaglio temporale, conserva una buona stabilità predittiva, mantenendo errori relativamente contenuti e un comportamento regolare. Si conferma quindi la maggiore capacità del modello probabilistico di generalizzare e di adattarsi anche a dataset meno densi.

Analisi del parametro di orizzonte temporale di previsione

L'ultima analisi riguarda l'effetto della variazione del parametro di orizzonte temporale (n), mantenendo costanti il numero di regressori $p = 10$ e l'assenza di resampling ($t_r = 1$). I valori considerati sono $n = 3, 4, 6$ e 8 , con l'obiettivo di valutare come l'aumento della distanza temporale della previsione influenzi l'accuratezza e la stabilità dei modelli adottati. L'analisi di tale parametro risulta particolarmente significativa, poiché consente di comprendere fino a che punto ciascun modello riesca a mantenere una capacità predittiva coerente quando la previsione si estende su orizzonti più lunghi, condizione che, come noto, introduce effetti di propagazione dell'errore e perdita di informazione temporale più marcati rispetto alla previsione one-step.

Linear Regression (LR)

Le figure 4.51 e 4.52 mostrano un progressivo peggioramento della qualità della previsione al crescere di n . Il modello, pur mantenendo una coerenza formale con l'andamento generale dei dati, mostra una perdita di aderenza sempre più evidente al segnale reale: i picchi vengono sistematicamente sottostimati, mentre le fasi di minimo tendono a essere sovrastimate, con un conseguente appiattimento del profilo predetto. Tale comportamento si deve alla natura intrinsecamente lineare del modello, che fatica a rappresentare relazioni temporali complesse e non lineari, caratteristiche tipiche del consumo energetico reale. L'aumento dell'orizzonte temporale amplifica inoltre l'effetto di propagazione dell'errore: ogni previsione si basa sulle stime prodotte nei passi precedenti, trasferendo e accumulando progressivamente eventuali deviazioni iniziali, fino a compromettere la stabilità del modello nelle previsioni più lontane.

Le metriche riportate in tabella 4.21 confermano questa tendenza: RMSE passa da 66,05 per $n = 3$ a 98,69 per $n = 8$, mentre MAE cresce da 40,98 a 72,86. L'incremento quasi lineare dell'errore al crescere di n testimonia la difficoltà della LR nel mantenere precisione su orizzonti temporali più ampi, con una perdita di accuratezza proporzionale alla distanza temporale di previsione. In altre parole, il modello riesce a cogliere con sufficiente attendibilità le dinamiche di breve periodo, ma non è in grado di preservare la coerenza della stima quando la previsione si sposta su intervalli temporali più lunghi, rivelando un comportamento progressivamente più rigido e approssimato.

Gaussian Process Regression (GPR)

Nel caso del GPR (figure 4.53 e 4.54), la qualità della previsione si mantiene elevata anche al crescere dell'orizzonte temporale. Il modello mostra una notevole capacità di adattamento, riuscendo a seguire il segnale reale con un degrado graduale e controllato, segno di una maggiore stabilità e flessibilità nella rappresentazione delle relazioni temporali sottostanti. All'aumentare di n , le previsioni mantengono

una buona coerenza con i dati osservati, pur mostrando un leggero smorzamento nelle variazioni più rapide, fenomeno fisiologico dovuto alla crescita dell'incertezza e all'allargamento delle bande di confidenza. Tale comportamento evidenzia come il GPR riesca a bilanciare efficacemente la complessità del modello e la robustezza della stima, preservando un buon livello di generalizzazione anche in condizioni di maggiore difficoltà predittiva.

I valori riportati in tabella 4.22 confermano l'elevata precisione del modello: RMSE cresce da 14,45 a 42,93 e MAE da 9,44 a 29,22 passando da $n = 3$ a $n = 8$. L'aumento dell'errore risulta quindi proporzionale all'ampliamento dell'orizzonte temporale e comunque contenuto rispetto alla LR, dimostrando che il GPR riesce a mantenere un buon equilibrio tra accuratezza e stabilità. La presenza delle bande di confidenza, che si ampliano in modo regolare con l'aumentare di n , fornisce inoltre una misura esplicita dell'incertezza associata a ciascuna previsione, consentendo di valutare non solo la bontà media della stima, ma anche la sua affidabilità statistica. Nel complesso, il GPR conferma la sua capacità di catturare relazioni non lineari e di adattarsi a dinamiche complesse anche in condizioni di informazione temporale ridotta.

Confronto tra i due modelli

Dal confronto tra i due modelli emerge chiaramente la maggiore efficacia della Gaussian Process Regression nel gestire orizzonti di previsione estesi. La Linear Regression mostra un degrado rapido delle prestazioni, con errori che aumentano quasi proporzionalmente alla lunghezza dell'orizzonte temporale, sintomo di una struttura troppo rigida e di una scarsa capacità di modellare le componenti non lineari e stocastiche del fenomeno analizzato. Il GPR, al contrario, mantiene una qualità di previsione elevata anche per n più ampi, grazie alla sua formulazione bayesiana e all'utilizzo della funzione kernel, che consente di apprendere correlazioni più profonde tra i campioni temporali, riducendo l'impatto dell'errore di propagazione.

In termini di accuratezza, il divario tra i due modelli risulta evidente: per $n = 8$, l'errore medio della LR è più che doppio rispetto a quello del GPR, mentre quest'ultimo conserva un andamento dell'errore più regolare e controllato. Dal punto di vista qualitativo, le previsioni del GPR appaiono più continue e realistiche, mentre la LR tende a semplificare eccessivamente la dinamica del segnale, perdendo la capacità di rappresentare le variazioni più significative. Nel complesso, l'analisi dell'orizzonte temporale di previsione conferma che il modello probabilistico non solo garantisce prestazioni più robuste e scalabili, ma rappresenta anche una soluzione più affidabile per applicazioni di previsione energetica su medio-lungo periodo, dove la stabilità e la capacità di gestire l'incertezza diventano elementi determinanti.

Capitolo 6

Conclusioni

6.1 Sintesi e commento finale dei risultati

L’analisi complessiva dei risultati ottenuti nei capitoli precedenti consente di delineare in modo chiaro le prestazioni e le caratteristiche dei due modelli di previsione adottati, ovvero la *Linear Regression* (LR) e la *Gaussian Process Regression* (GPR), applicati ai dati di consumo energetico aziendale nelle diverse configurazioni sperimentali considerate.

Nel caso della previsione *one-step ahead*, entrambi i modelli hanno mostrato buone capacità di adattamento al comportamento reale dei dati, con errori contenuti e una coerenza soddisfacente nella riproduzione delle dinamiche temporali.

La *Linear Regression*, pur essendo un modello lineare e computazionalmente leggero, si è dimostrata efficace nel catturare le tendenze generali, in particolare nei periodi a variazione graduale dei consumi. Tuttavia, nei momenti caratterizzati da oscillazioni più marcate o comportamenti non lineari, la sua accuratezza tende a ridursi.

La *Gaussian Process Regression*, al contrario, ha evidenziato una maggiore flessibilità e capacità di adattamento locale, fornendo previsioni notevolmente più precise e stabili, soprattutto nelle fasi di maggiore variabilità. Tale accuratezza superiore è tuttavia accompagnata da un costo computazionale più elevato, sia in fase di addestramento che di previsione, rendendo il modello meno indicato in contesti in cui la rapidità di elaborazione rappresenta un vincolo operativo.

L’introduzione del *resampling* ha invece mostrato un effetto opposto rispetto a quanto spesso ipotizzato in letteratura: in questo caso, l’aumento dell’intervallo di aggregazione dei dati ha portato a un progressivo peggioramento delle prestazioni di entrambi i modelli. L’incremento di *RMSE* e *MAE* osservato già per intervalli di

resampling di 4–6 ore indica che la riduzione della risoluzione temporale comporta una perdita di informazioni significative per la previsione, probabilmente legate alla variabilità intrinseca del segnale su scale orarie.

La *Linear Regression* mostra un degrado moderato e regolare, mentre la *GPR* — pur mantenendo un livello di accuratezza complessivamente superiore — subisce un incremento più marcato degli errori, segno che la sua capacità di modellare le dipendenze locali risulta penalizzata dalla perdita di dettaglio temporale. Questo comportamento suggerisce che, per il dataset in esame, un'elevata granularità dei dati costituisce un elemento essenziale per il mantenimento di buone prestazioni predittive.

Nel contesto della previsione *multi-step ahead*, le differenze tra i due approcci risultano più evidenti. Con l'aumentare dell'orizzonte temporale di previsione, entrambi i modelli subiscono un fisiologico incremento dell'errore, ma in misura differente.

La *Linear Regression* manifesta una progressiva perdita di coerenza con i dati reali, con errori crescenti e una tendenza a sottostimare o sovrastimare i picchi di consumo. La *Gaussian Process Regression*, pur mostrando anch'essa un degrado delle prestazioni all'aumentare dell'orizzonte, riesce a mantenere una struttura predittiva più stabile e coerente, preservando meglio la forma generale del segnale e contenendo l'aumento dell'errore. L'analisi dell'orizzonte temporale ha dunque confermato che la complessità del fenomeno energetico richiede un modello capace di rappresentare le relazioni non lineari e la propagazione dell'incertezza nel tempo, compito svolto con maggiore efficacia dalla GPR.

Nel confronto complessivo, non emerge un modello universalmente migliore, bensì due strumenti con caratteristiche complementari. La *Linear Regression* si distingue per la semplicità di implementazione, la rapidità di esecuzione e l'efficienza computazionale, che la rendono particolarmente adatta a scenari in cui è richiesta una previsione continua, rapida e a basso costo computazionale, anche a fronte di una minore accuratezza. La *Gaussian Process Regression*, invece, si conferma un modello più sofisticato e accurato, in grado di catturare efficacemente la complessità del fenomeno energetico e di offrire una rappresentazione più realistica dei consumi, sebbene a un costo di calcolo significativamente superiore.

In sintesi, la scelta tra i due approcci dipende dal contesto applicativo: in un sistema di monitoraggio operativo, dove prevalgono i requisiti di rapidità e semplicità, la *Linear Regression* rappresenta la soluzione più efficiente; in un ambito di analisi previsionale o pianificazione energetica, dove l'accuratezza e la capacità di cogliere le dinamiche locali sono prioritarie, la *Gaussian Process Regression* costituisce l'alternativa più appropriata.

Questa complementarità evidenzia come, in prospettiva, l'integrazione dei due approcci — ad esempio mediante una combinazione ibrida o gerarchica — possa rappresentare una via promettente per conciliare prestazioni e costi computazionali in applicazioni reali di previsione energetica.

6.2 Considerazioni finali e sviluppi futuri

I risultati ottenuti nel corso di questo lavoro evidenziano come l'applicazione di modelli di previsione dei consumi energetici possa rappresentare un valido strumento di supporto alle attività aziendali, in particolare nell'ambito dello *short-term forecasting*.

La possibilità di stimare in anticipo l'andamento dei consumi, anche su orizzonti temporali ridotti, consente infatti di migliorare la gestione operativa dell'impianto, ottimizzare l'uso delle risorse e pianificare con maggiore efficienza la produzione e l'approvvigionamento energetico. Un modello predittivo affidabile, capace di cogliere tempestivamente variazioni o anomalie nel profilo di consumo, può inoltre contribuire a identificare situazioni di inefficienza o comportamenti anomali, favorendo interventi correttivi rapidi e mirati.

In questo senso, i modelli sviluppati nel presente lavoro, pur con i loro limiti e differenze prestazionali, hanno mostrato come sia possibile costruire un sistema di previsione efficace a partire da dati storici di consumo, aprendo la strada a un utilizzo concreto di tecniche di *machine learning* a supporto delle strategie di gestione energetica aziendale.

Dal punto di vista metodologico, i risultati ottenuti costituiscono un primo e significativo passo verso l'introduzione, in azienda, di strumenti analitici avanzati per la previsione dei consumi. Si tratta infatti del primo approccio sperimentale interno volto a quantificare e modellare l'andamento energetico dello stabilimento, ponendo le basi per futuri sviluppi più ampi.

Un'evoluzione naturale di questo lavoro potrebbe riguardare, da un lato, il perfezionamento dei modelli esistenti attraverso un affinamento dei parametri, l'utilizzo di tecniche di ottimizzazione automatica o l'introduzione di ulteriori variabili esplicative (quali temperatura, produzione o dati meteorologici); dall'altro, l'estensione dell'orizzonte temporale di previsione, passando da uno *short-term forecasting* come quello oggetto di questo studio a orizzonti temporali più ampi, in grado di fornire una visione predittiva più estesa e strategica del comportamento energetico aziendale.

Questo permetterebbe all'azienda di disporre non solo di strumenti operativi per la gestione quotidiana, ma anche di un supporto decisionale di medio-lungo periodo utile alla pianificazione strategica e alla definizione di politiche di efficienza energetica più consapevoli.

In prospettiva, l'obiettivo potrà essere quello di integrare tali modelli predittivi in un sistema più ampio di monitoraggio e controllo, capace di fornire previsioni in tempo reale e di adattarsi automaticamente ai cambiamenti del comportamento energetico.

Un simile approccio permetterebbe di coniugare l'accuratezza della previsione con la capacità di reazione e adattamento alle condizioni operative, rendendo il processo produttivo più efficiente, sostenibile e in linea con gli obiettivi di transizione energetica aziendale.

Bibliografia

- [1] Anas Thamer Mustafa e Omar Sharaf Al-Deen Al-Yozbaky. «Forecasting energy demand and generation using time series models: A comparative analysis of classical, grey, fuzzy, and intelligent approaches». In: *Franklin Open* 12 (set. 2025), p. 100350. ISSN: 2773-1863. DOI: 10.1016/J.FRAOPE.2025.100350. URL: <https://www.sciencedirect.com/science/article/pii/S2773186325001380> (cit. a p. 2).
- [2] Mehmet Güçyetmez, Sıtkı Akkaya, Murat Uyar e Şekip Esat Hayber. «Forecasting renewable energy consumption with hydrogen integration: A comprehensive regression approach». In: *International Journal of Hydrogen Energy* 142 (giu. 2025), pp. 981–993. ISSN: 0360-3199. DOI: 10.1016/J.IJHYDENE.2025.03.244. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0360319925013655?via%3Dihub> (cit. a p. 2).
- [3] Mehmet Bilgili e Engin Pinar. «Gross electricity consumption forecasting using LSTM and SARIMA approaches: A case study of Türkiye». In: *Energy* 284 (dic. 2023), p. 128575. ISSN: 0360-5442. DOI: 10.1016/J.ENERGY.2023.128575. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0360544223019692?via%3Dihub> (cit. a p. 2).
- [4] André Luiz Marques Serrano, Gabriel Arquelau Pimenta Rodrigues, Patricia Helena dos Santos Martins, Gabriela Mayumi Saiki, Geraldo Pereira Rocha Filho, Vinícius Pereira Gonçalves e Robson de Oliveira Albuquerque. «Statistical Comparison of Time Series Models for Forecasting Brazilian Monthly Energy Demand Using Economic, Industrial, and Climatic Exogenous Variables». In: *Applied Sciences* 14.13 (2024), p. 5846. ISSN: 2076-3417. DOI: 10.3390/app14135846. URL: <https://www.mdpi.com/2076-3417/14/13/5846> (cit. a p. 2).
- [5] Abolore Abdulkareem Nimota. «Time Series Forecast of Nigeria's Electricity Using Auto-Regressive Integrated Moving Average (ARIMA) Model». In: *Journal of Innovative Research in Business, Digitalization & Artificial Intelligence* 1.2 (giu. 2025), p. 14. URL: <https://jirbdai.com.ng/index.php/jirbdai/article/view/50> (cit. a p. 2).

- [6] Maolin Cheng, Jiano Li, Yun Liu e Bin Liu. «Forecasting Clean Energy Consumption in China by 2025: Using Improved Grey Model GM (1, N)». In: *Sustainability* 12.2 (2020), p. 698. DOI: 10.3390/su12020698. URL: <https://www.mdpi.com/2071-1050/12/2/698> (cit. a p. 2).
- [7] Tuo Xie, Gang Zhang, Hongchi Liu, Fuchao Liu e Peidong Du. «A Hybrid Forecasting Method for Solar Output Power Based on Variational Mode Decomposition, Deep Belief Networks and Auto-Regressive Moving Average». In: *Applied Sciences* 8.10 (2018), p. 1901. ISSN: 2076-3417. DOI: 10.3390/app8101901. URL: <https://www.mdpi.com/2076-3417/8/10/1901> (cit. a p. 2).
- [8] Rowell Marquez Hernandez, John Joseph Castillo, Adrian Angel Ragasa, Celinne Mendez e Fatima Agdon. «Energy Consumption Forecasting for Smart Industry Using Auto-Regressive Integrated Moving Average (ARIMA) and Vector Auto-Regression (VAR) Model». In: *Proceedings of the International Conference on Sustainable Information Engineering and Technology (SIET 2023)*. ACM, 2023, pp. 129–136. DOI: 10.1145/3626641.3627607. URL: <https://doi.org/10.1145/3626641.3627607> (cit. a p. 2).
- [9] Idorenyin Markson, Mfonobong Charles Uko e Aneke Chikezie. «Evaluation of Moving Average Model and Autoregressive Moving Average Model (ARMA) for Prediction of Industrial Electricity Consumption in Nigeria». In: *American Journal of Software Engineering and Applications* 6.3 (2017), pp. 67–73. DOI: 10.11648/j.ajsea.20170603.12. URL: <https://www.sciencepublishinggroup.com/article/10.11648/j.ajsea.20170603.12> (cit. a p. 2).
- [10] Aulon Shabani, Adil Paul, Radu Platon e Eyke Hüllermeier. «Predicting the Electricity Consumption of Buildings: An Improved CBR Approach». In: *Case-Based Reasoning Research and Development*. Vol. 9969. Lecture Notes in Artificial Intelligence. Cham: Springer, 2016, pp. 356–369. DOI: 10.1007/978-3-319-47096-2_24. URL: https://link.springer.com/chapter/10.1007/978-3-319-47096-2_24 (cit. a p. 2).
- [11] J. Prabavadhi e S. Kanmani. «Electricity Consumption Monitor and Predictor with Current Sensor Integration». In: *Proceedings of the 2024 International Conference on System, Computation, Automation and Networking (ICSCAN)*. Puducherry, India: IEEE, dic. 2024, pp. –. DOI: 10.1109/ICSCAN62807.2024.10894427. URL: <https://ieeexplore.ieee.org/document/10894427> (cit. a p. 2).
- [12] Ibrahim Ali Kachalla, Christian Ghiaus e Majid Baseer. «Comparative analysis of machine learning models for prediction and forecasting of electric water boilers energy consumption». In: *Applied Thermal Engineering* 267 (mag. 2025), p. 125799. ISSN: 1359-4311. DOI: 10.1016/J.APPLTHERMALENG.2025.

125799. URL: <https://www.sciencedirect.com/science/article/pii/S1359431125003904?via%3Dihub> (cit. a p. 2).
- [13] S. Hadri, Y. Naitmalek, M. Najib, M. Bakhouya, Y. Fakhri e M. Elaroussi. «A Comparative Study of Predictive Approaches for Load Forecasting in Smart Buildings». In: *Procedia Computer Science* 160 (gen. 2019), pp. 173–180. ISSN: 1877-0509. DOI: 10.1016/J.PROCS.2019.09.458. URL: <https://www.sciencedirect.com/science/article/pii/S1877050919316734?via%3Dihub> (cit. a p. 2).
- [14] Mingliang Bai, Yunxiao Chen, Xinyu Zhao, Jinfu Liu e Daren Yu. «Deep attention ConvLSTM-based adaptive fusion of clear-sky physical prior knowledge and multivariable historical information for probabilistic prediction of photovoltaic power». In: *Expert Systems with Applications* 202 (set. 2022), p. 117335. ISSN: 0957-4174. DOI: 10.1016/J.ESWA.2022.117335. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0957417422006935?via%3Dihub> (cit. a p. 2).
- [15] Saifullah Khan, Yasir Muhammad, Ihtesham Jadoon, Saeed Ehsan Awan e Muhammad Asif Zahoor Raja. «Leveraging LSTM-SMI and ARIMA architecture for robust wind power plant forecasting». In: *Applied Soft Computing* 170 (feb. 2025), p. 112765. ISSN: 1568-4946. DOI: 10.1016/J.ASOC.2025.112765. URL: <https://www.sciencedirect.com/science/article/abs/pii/S1568494625000766?via%3Dihub> (cit. a p. 2).
- [16] Abdel Haleem Abdel-Aty, Kottakkaran Sooppy Nisar, Wedad R. Alharbi, Saud Owyed e Mohammed H. Alsharif. «Boosting wind turbine performance with advanced smart power prediction: Employing a hybrid ARMA-LSTM technique». In: *Alexandria Engineering Journal* 96 (giu. 2024), pp. 58–71. ISSN: 1110-0168. DOI: 10.1016/J.AEJ.2024.03.078. URL: <https://www.sciencedirect.com/science/article/pii/S111001682400320X?via%3Dihub> (cit. a p. 2).
- [17] Muamar Mohamed, Farhad E. Mahmood, Mehmood A. Abd, Ambrish Chandra e Bhim Singh. «Dynamic Forecasting of Solar Energy Microgrid Systems Using Feature Engineering». In: *IEEE Transactions on Industry Applications* 58.6 (nov. 2022), pp. 7857–7869. ISSN: 0093-9994. DOI: 10.1109/TIA.2022.3199182. URL: <https://ieeexplore.ieee.org/document/9858626> (cit. a p. 2).
- [18] Gang Zhang, Jingwen Hu e Mingyuan Dong. «Forecasting short-term electricity load of large-scale data center buildings using an improved deep learning model». In: *Energy Informatics* 1.1 (2018), p. 12. DOI: 10.1007/s40565-018-0496-z. URL: <https://link.springer.com/article/10.1007/s40565-018-0496-z> (cit. a p. 2).

- [19] Shobhit Chaturvedi, Elangovan Rajasekar, Sukumar Natarajan e Nick Mc Cullen. «A comparative assessment of SARIMA, LSTM RNN and Fb Prophet models to forecast total and peak monthly energy demand for India». In: *Energy Policy* 168 (set. 2022), p. 113097. ISSN: 0301-4215. DOI: 10.1016/J.ENPOL.2022.113097. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0301421522003226?via%3Dihub> (cit. a p. 2).
- [20] Sujan Ghimire, Ravinesh C. Deo, S. Ali Pourmousavi, David Casillas-Pérez e Sancho Salcedo-Sanz. «Point-based and probabilistic electricity demand prediction with a Neural Facebook Prophet and Kernel Density Estimation model». In: *Engineering Applications of Artificial Intelligence* 135 (set. 2024), p. 108702. ISSN: 0952-1976. DOI: 10.1016/J.ENGAPPAL.2024.108702. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0952197624008601?via%3Dihub> (cit. a p. 2).
- [21] Spyridon Karamolegos e Dimitrios E. Koulouriotis. «Advancing short-term load forecasting with decomposed Fourier ARIMA: A case study on the Greek energy market». In: *Energy* 325 (giu. 2025), p. 135854. ISSN: 0360-5442. DOI: 10.1016/J.ENERGY.2025.135854. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0360544225014963?via%3Dihub> (cit. a p. 2).
- [22] Kamran Hassanpouri Baesmat, Farhad Shokoohi e Zeinab Farrokhi. «SP-RF-ARIMA: A sparse random forest and ARIMA hybrid model for electric load forecasting». In: *Global Energy Interconnection* 8 (3 giu. 2025), pp. 486–496. ISSN: 2096-5117. DOI: 10.1016/J.GLOEI.2025.04.003. URL: <https://www.sciencedirect.com/science/article/abs/pii/S2096511725000490?via%3Dihub> (cit. a p. 2).
- [23] A. Shadab, S. Ahmad e S. Said. «Spatial Forecasting of Solar Radiation using ARIMA model». In: *Remote Sensing Applications: Society and Environment* 20 (2020), p. 100427. DOI: 10.1016/j.rsase.2020.100427. URL: <https://www.sciencedirect.com/science/article/abs/pii/S2352938520302731?via%3Dihub> (cit. a p. 27).
- [24] Bismark Singh e David Pozo. «A Guide to Solar Power Forecasting Using ARMA Models». In: *2019 IEEE PES Innovative Smart Grid Technologies Europe (ISGT-Europe)*. Bucharest, Romania: IEEE, set. 2019, pp. –. DOI: 10.1109/ISGTEurope.2019.8905430. URL: <https://ieeexplore.ieee.org/document/8905430> (cit. a p. 27).
- [25] B. Belmahdi et al. «One month-ahead forecasting of mean daily global solar radiation». In: *Solar Energy* 208 (2020), pp. 425–438. DOI: 10.1016/j.solener.2020.10.043. URL: <https://www.sciencedirect.com/science/article/abs/pii/S0030402620310433?via%3Dihub> (cit. a p. 27).

- [26] Ines Sansa, Zina Boussaada, Mourad Mazigh e Najiba Mrabet Bellaaj. «Solar Radiation Prediction for a Winter Day Using ARMA Model». In: *2020 6th IEEE International Energy Conference (ENERGYCon)*. Gammarth, Tunisia: IEEE, set. 2020, pp. –. DOI: 10.1109/ENERGYCon48941.2020.9236541. URL: <https://ieeexplore.ieee.org/document/9236541> (cit. a p. 27).
- [27] Sharif Atique, Subrina Noureen, Vishwajit Roy, Vinita Subburaj, Stephen Bayne e Joshua Macfie. «Forecasting of Total Daily Solar Energy Generation Using ARIMA: A Case Study». In: *2019 IEEE 9th Annual Computing and Communication Workshop and Conference (CCWC)*. Las Vegas, NV, USA: IEEE, gen. 2019, pp. –. DOI: 10.1109/CCWC.2019.8666481. URL: <https://ieeexplore.ieee.org/document/8666481> (cit. a p. 27).
- [28] Jingmin Wang, Qingwei Zhou e Xueling Zhang. «Wind Power Forecasting Based on Time Series ARMA Model». In: *IOP Conference Series: Earth and Environmental Science*. Vol. 199. 2. IOP Publishing, 2018, p. 022015. DOI: 10.1088/1755-1315/199/2/022015. URL: <https://iopscience.iop.org/article/10.1088/1755-1315/199/2/022015> (cit. a p. 27).
- [29] Yitong Niu, Xiongjie Jia, Chee Keong Lee, Haoran Jiang e Cheu Peng Leh. «Analyzing and Forecasting Laboratory Energy Consumption Patterns Using Autoregressive Integrated Moving Average Models». In: *Laboratories* 2.1 (dic. 2025), p. 2. DOI: 10.3390/laboratories2010002. URL: <https://www.mdpi.com/2813-8856/2/1/2> (cit. a p. 27).
- [30] Putri Gunandra Siregar, Ilham Sahputra e Fidyatun Nisa. «Forecasting of Electrical Energy Consumption Using Autoregressive Integrated Moving Average (Case Study: ULP Meulaboh Kota)». In: *Journal of Geuthée* 4.1 (2023), pp. –. DOI: 10.52626/joge.v4i1.56. URL: <https://joge.geutheeiнституте.com/index.php/jogee/article/view/56> (cit. a p. 27).
- [31] Qizhuan Shao, Rungang Bao, Shuangquan Liu, Kaixiang Fu, Li Mo e Wenjing Xiao. «Short-Term Electric Load Probability Forecasting Based on the BiGRU-GAM-GPR Model». In: *Sustainability* 17.12 (2025), p. 5267. DOI: 10.3390/su17125267. URL: <https://www.mdpi.com/2071-1050/17/12/5267> (cit. a p. 29).
- [32] Mithun Madhukumar, Albino Sebastian, Xiaodong Liang, Mohsin Jamil e Md Nasmus Sakib Khan Shabbir. «Regression Model-Based Short-Term Load Forecasting for University Campus Load». In: *IEEE Access* 10 (gen. 2022), pp. 8891–8905. ISSN: 2169-3536. DOI: 10.1109/ACCESS.2022.3144206. URL: <https://ieeexplore.ieee.org/document/9684449> (cit. a p. 29).

- [33] Anamika Yadav, Rashmi Bareth, Matushree Kochhar, Mohammad Pazoki e Ragab A. El Sehiemy. «Gaussian Process Regression-Based Load Forecasting Model». In: *IET Generation, Transmission & Distribution* 18.5 (2024), pp. 899–910. DOI: 10.1049/gtd2.12926. URL: <https://doaj.org/article/04e2b46381994224b54c00388b262d12> (cit. a p. 29).
- [34] Anwar Jarndal e Saddam Husain. «Forecasting of Electric Peak Load Using ANN-Cascaded, ANN-NARX and GPR Techniques». In: *2020 International Conference on Communications, Computing, Cybersecurity, and Informatics (CCCI)*. Sharjah, United Arab Emirates: IEEE, nov. 2020, pp. –. DOI: 10.1109/CCCI49893.2020.9256555. URL: <https://ieeexplore.ieee.org/document/9256555> (cit. a p. 29).
- [35] Kadir Amasyali e Mohammed Olama. «Gaussian Process Regression for Aggregate Baseline Load Forecasting». In: *2021 Annual Modeling and Simulation Conference (ANNSIM)*. Fairfax, VA, USA: IEEE, lug. 2021, pp. –. DOI: 10.23919/ANNSIM52504.2021.9552156. URL: <https://ieeexplore.ieee.org/document/9552156> (cit. a p. 29).
- [36] Saikat Gochhait e Deepak Sharma. «Regression Model-Based Short-Term Load Forecasting for Load Despatch Centre». In: *Journal of Applied Engineering and Technological Science* 4.2 (2023), pp. 693–710. DOI: 10.37385/jaets.v4i2.1682. URL: <https://doaj.org/article/9178bcc9e7874b0596b23fb5792fd73f> (cit. a p. 29).
- [37] Yaling Zhu, Bo Zhang, Zhenhai Dou, Hao Zou, Shengtao Li, Kai Sun e Qingling Liao. «Short-Term Load Forecasting Based on Gaussian Process Regression with Density Peak Clustering and Information Sharing Antlion Optimizer». In: *IET Generation, Transmission & Distribution* 14.18 (lug. 2020), pp. 3584–3595. DOI: 10.1002/tee.23198. URL: <https://onlinelibrary.wiley.com/doi/10.1002/tee.23198> (cit. a p. 29).
- [38] Manish Kumar e Nitai Pal. «Machine Learning-Based Electric Load Forecasting for Peak Demand Control in Smart Grid». In: *Computers, Materials & Continua* 74.3 (2023), pp. 4785–4799. DOI: 10.32604/cmc.2022.032971. URL: <https://www.proquest.com/docview/3199834947> (cit. a p. 29).
- [39] Anand Krishnan Prakash, Susu Xu, Ram Rajagopal e Hae Young Noh. «Robust Building Energy Load Forecasting Using Physically-Based Kernel Models». In: *Energies* 11.4 (2018), p. 862. DOI: 10.3390/en11040862. URL: <https://doaj.org/article/08d2a2ad932b41baa649eb56bac466c5> (cit. a p. 29).

- [40] T. Chai e R. R. Draxler. «Root Mean Square Error (RMSE) or Mean Absolute Error (MAE)? – Arguments Against Avoiding RMSE in the Literature». In: *Geoscientific Model Development* 7.3 (2014), pp. 1247–1250. DOI: 10.5194/gmd-7-1247-2014. URL: <https://doaj.org/article/6bf18970a5de4497ae44d7dd210fec3b> (cit. a p. 34).
- [41] T. O. Hodson. «Root-Mean-Square Error (RMSE) or Mean Absolute Error (MAE): When to Use Them or Not». In: *Geoscientific Model Development* 15 (2022), pp. 5481–5487. DOI: 10.5194/gmd-15-5481-2022. URL: <https://doaj.org/article/bc6508e8a630422e9731b3d704f6d940> (cit. a p. 34).
- [42] Martin Čalasan, Shady H. E. Abdel Aleem e Ahmed F. Zobaa. «On the Root Mean Square Error (RMSE) Calculation for Parameter Estimation of Photovoltaic Models: A Novel Exact Analytical Solution Based on the Lambert W Function». In: *Energy Conversion and Management* 221 (2020), p. 112716. DOI: 10.1016/j.enconman.2020.112716. URL: <https://www.sciencedirect.com/science/article/pii/S0196890420302545> (cit. a p. 34).