POLITECNICO DI TORINO

Corso di Laurea in Ingegneria Elettrica

Tesi di Laurea Magistrale

Previsione dei carichi delle reti di distribuzione AT con tecniche di machine learning

POLITECNICO
DI TORINO

Relatore
Prof. Ettore Bompard

Correlatori:
Prof. Tao Huang
Ing. Paolo Cuccia

Laureando
Enrico Marin
matricola: 240250

Anno accademico 2018 – 2019

Sommario

In questo elaborato sono stati studiati, analizzati e sviluppati diversi metodi di previsione dei carichi industriali connessi alla rete di distribuzione in alta tensione. La porzione di rete in esame corrisponde ad un'isola di carico situata nell'Italia nord-occidentale. I metodi, che si basano su studi di regressione lineare, reti neurali artificiali e studi di clusterizzazione, sono ottenuti tramite l'analisi dei dati di carico relativi ai primi 11 mesi del 2017 e testati durante dicembre dello stesso anno. Considerando sia la bontà dei risultati ottenuti che la complessità e affidabilità dei metodi, è stata scelta la tecnica di previsione più adatta per essere implementata nei sistemi del Transmission System Operator (TSO) Italiano. La tecnica selezionata è stata quindi testata in una macchina di prova funzionante in parallelo ai sistemi di controllo del TSO e i risultati sono stati confrontati con i corrispettivi valori ottenuti dal metodo di stima attualmente in uso nei sistemi del TSO. Il confronto ha permesso di determinare qualora il metodo proposto comporti benefici effettivi.

Indice

1 Introduzione generale 6
- 1.1 Obiettivo dell'elaborato . . . 6
- 1.2 Composizione dell'elaborato . . . 7
- 1.3 Descrizione dell'isola di carico in esame . . . 7

2 Il problema di Power-flow 11
- 2.1 Introduzione teorica . . . 11
- 2.2 Semplificazioni . . . 12
- 2.3 Modello matematico . . . 12

3 Futuri investimenti sulle reti elettriche di potenza 17
- 3.1 Introduzione . . . 17
- 3.2 Smart Grid . . . 18
- 3.3 AMI . . . 20
- 3.4 Smart meter . . . 23
- 3.5 Smart Grid e il problema della stima dei carichi . . . 24

4 Regressione lineare 27
- 4.1 Introduzione . . . 27
- 4.2 Teoria . . . 28
  - 4.2.1 Introduzione alla teoria . . . 28
  - 4.2.2 Assunzioni . . . 30
  - 4.2.3 Algoritmi di stima . . . 34
- 4.3 Applicazione della regressione lineare . . . 39

5 Rete neurale artificiale 49
- 5.1 Teoria . . . 49
  - 5.1.1 Introduzione . . . 49
  - 5.1.2 Componenti delle reti neurali artificiali . . . 50

5.1.3 Paradigmi di apprendimento 55
5.1.4 Algoritmi di apprendimento 56
5.2 Implementazione della rete neurale artificiale 57

6 Altri metodi 63
6.1 Combinazione lineare tra regressione e ANN 63
6.2 Clusterizzazione 67
6.2.1 Introduzione teorica al problema di clusterizzazione . . 67
6.2.2 Implementazione 71

7 Confronto dei risultati 77
7.1 Confronto quantitativo dei metodi 77
7.2 Individuazione del metodo più adatto per l'implementazione nei sistemi del TSO 78

8 Implementazione e collaudo 79

9 Conclusioni 89

A Test sui metodi - Risultati completi 91

B Confronto tra stima proposta e stima attuale - Risultati completi 115

Bibliografia 127

Capitolo 1

Introduzione generale

1.1 Obiettivo dell'elaborato

Gli operatori del sistema elettrico di trasmissione (Transmission System Operators, TSO) per gestire in sicurezza ed efficacia le reti di trasmissione e subtrasmissione, devono conoscere con esattezza i parametri che caratterizzano la rete. In aggiunta ai parametri fisici della rete, devono essere noti il modulo e la fase delle tensioni e i flussi di potenza ad ogni nodo. Per fare questo è necessario svolgere uno studio di power-flow. Considerando noti i parametri fisici della rete, per svolgere uno studio di power-flow, è necessario conoscere almeno due delle quattro variabili che caratterizzano ogni nodo (modulo e fase della tensione e scambi di potenza attiva e reattiva con il sistema esterno). Nelle reti ad altissima tensione AAT il numero di variabili note è sovrabbondante e supera il numero di equazioni disponibili, creando un problema sovradeterminato. Attraverso un metodo computazionale chiamato "stima dello stato", è persino possibile verificare la coerenza delle misure disponibili e nel caso, identificare le eventuali anomalie (bad data). Nelle reti di subtrasmissione, invece, è presente il problema opposto, ovvero il sistema è sotto-determinato, in molti nodi, non sono disponibili almeno due variabili note. In particolare, sono mancati quasi tutte le misure in tempo reale di assorbimento delle potenze attive e reattive dei carichi industriali. Si rende perciò necessario stimare queste misure mancanti, verificarne la coerenza tramite il processo di stima dello stato e in seguito svolgere lo studio di power-flow. È da sottolineare che il processo di stima dello stato è in grado di rendere coerente, ma non necessariamente corretti, i dati disponibili. È chiaro quindi che la bontà di questo processo dipende fortemente dalla qualità della stima effettuata. Il TSO Italiano (ruolo assunto dalla azienda Terna

Rete Italia S.p.A.) stima i carichi mancanti effettuando bilanci di potenza ai nodi, quando possibile, o semplicemente assegnando valori fissi. Il problema sarebbe immediatamente risolto se si intervenisse nella rete con un'opera di adeguamento dei sistemi di misura, ma ciò comporterebbe un elevato costo di investimento. L'elaborato si propone di trovare una tecnica di stima più efficace di quella attualmente in uso, sufficientemente semplice e robusta da poter essere applicata nei sistemi reali dei TSO e in grado di sostituirsi ad un oneroso nuovo sistema di misura.

1.2 Composizione dell'elaborato

Dopo un breve capitolo introduttivo, verrà presentata in modo sommario la teoria riguardante il problema di power-flow (*Capitolo 2*) e chiarito il motivo per cui non sono disponibili misure in tempo reale dei consumi elettrici delle utenze industriali e per quale motivo è preferibile procedere alla loro acquisizione tramite stima piuttosto che tramite un nuovo sistema di misura (*Capitolo 3*). In seguito, verranno presentati, sia dal punto di vista teorico che applicativo, i metodi che questo lavoro propone per affrontare il problema della stima dei carichi (*Capitolo 4, 5 e 6*). Nel *Capitolo 7* i metodi vengono confrontati in modo quantitativo e qualitativo per stabilire quale di essi sia il più idoneo secondo criteri di precisione, semplicità e affidabilità. La tecnica selezionata è stata poi testata su una macchina di prova funzionante in parallelo ai sistemi che governano la rete e i risultati delle due stime sono confrontati e presentati nel *Capitolo 8*. Nello stesso capitolo sono stati poi esposti i risultati di una ulteriore convalida svolta su una differente isola di carico. Segue un breve capitolo conclusivo. In appendice sono infine riportati i risultati in forma grafica dei vari processi di stima.

1.3 Descrizione dell'isola di carico in esame

Lo studio è stato svolto su un'isola di carico della rete in alta tensione AT dell'Italia nord-occidentale. La porzione di rete in esame si estende geograficamente tra Torino sud, Cuneo e Alba (figura 1.1). Essa è alimentata da 4 stazioni elettriche attraverso 10 linee ed è composta da 12 carichi industriali, 32 cabine primarie e 15 unità di generazione collegate a 7 bus. 3 di questi sono bus ai quali sono collegati carichi industriali e le relative unità di generazione sono infatti di proprietà dell'utente. In figura 1.2 è riportato lo schema dell'isola considerata. Non rispetta la disposizione geografica dei

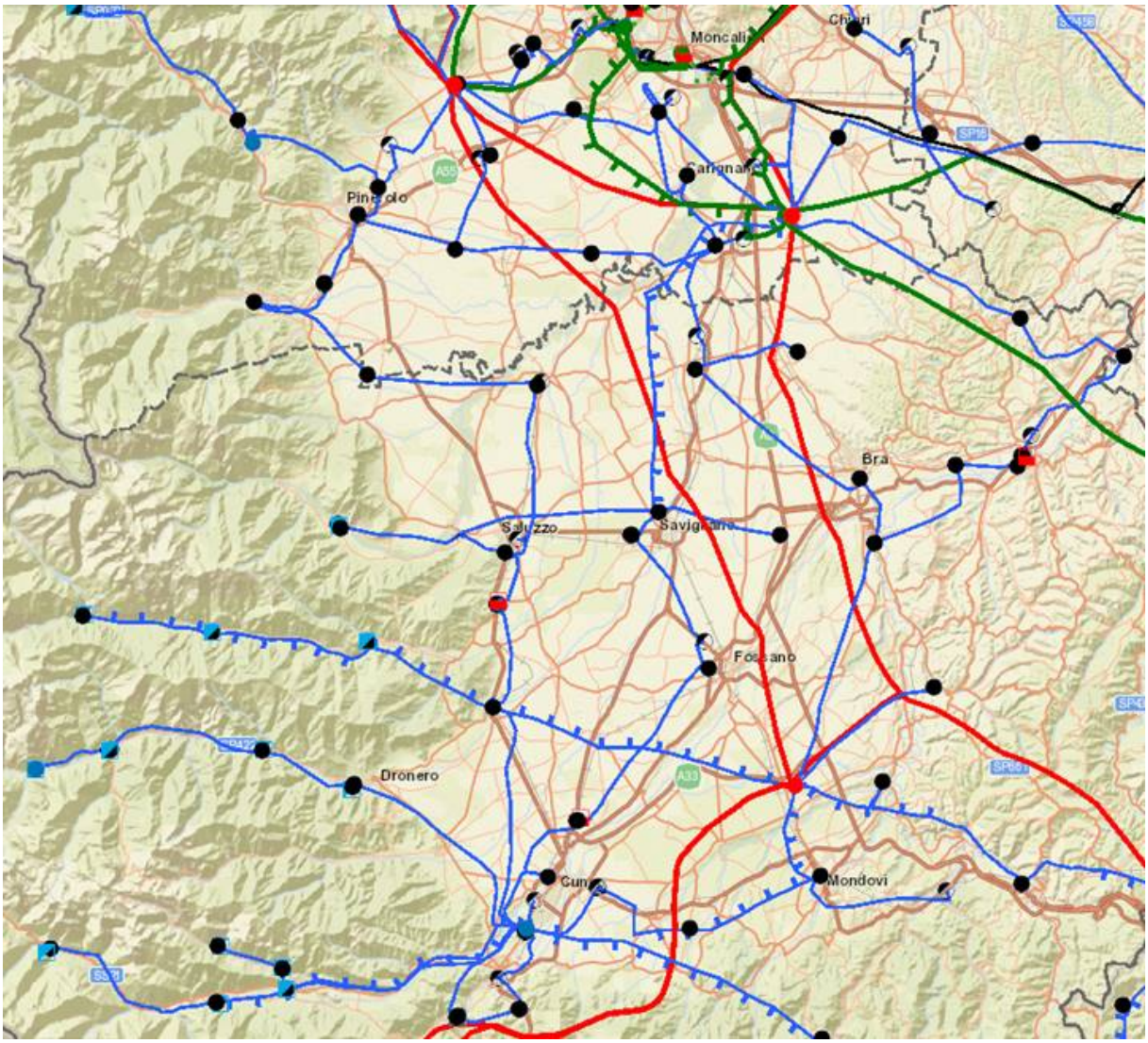


Figura 1.1. Schema geografico della rete a 132 kV (in blu) che compone l'isola di carico in esame.

vari componenti così come la proporzione delle distanze, serve unicamente per dare un'idea della struttura dell'isola, delle connessioni e della complessità del sistema. I nodi in verde (1, 3, 7, 13, 15, 23, 24, 28, 29, 32, 37, 46) sono bus al quale è collegato almeno un carico industriale, possono infatti essere collegate anche unità di generazione di proprietà dell'utente (24, 37 e 46) o cabine primarie. I nodi in rosso (11, 20, 44, 47) rappresentano le stazioni elettriche. I nodi in giallo (40, 41, 42 e 43) sono nodi al quale sono collegate unità di generazione elettrica. Nei nodi restanti, ovvero in blu, sono

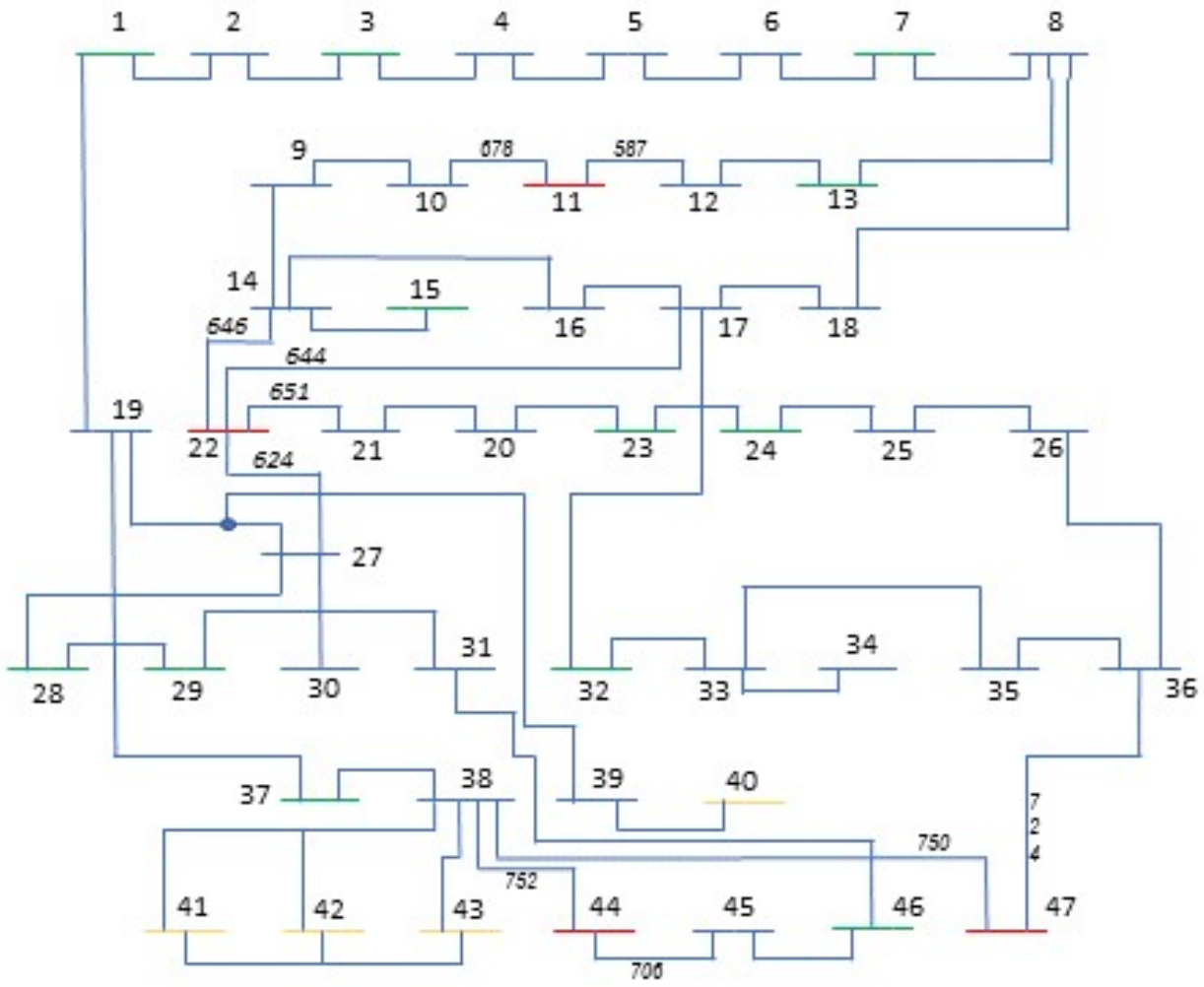


Figura 1.2. Semplice schema elettrico dell'isola di carico di interesse. I nodi in verde (1, 3, 7, 13, 15, 23, 24, 28, 29, 32, 37, 46) sono sbarre al quale è collegato almeno un carico industriale. I nodi in rosso (11, 22, 44, 47) rappresentano le stazioni elettriche. I nodi in giallo (40, 41, 42, 43) sono bus al quale sono collegate unità di generazione. Unità di generazone sono presenti anche nei nodi 24, 37 e 46. I nodi restanti, ovvero in blu, sono unicamente cabine primarie. Sono inoltre indicati i numeri delle linee in uscita dalle stazioni.

collegate unicamente cabine primarie. Solo per un carico, connesso al nodo 37, l'assorbimento è ricavato con precisione da altre misure in tempo reale, effettuando un bilancio al nodo, poiché è disponibile lo scambio netto con

la rete e la potenza generata dal proprio impianto di produzione. Con una semplice sottrazione è quindi ricavato l'assorbimento del carico. Per quanto riguarda gli altri carichi sono solo disponibili lo storico delle misure di assorbimento. In seguito, è riportata la denominazione dei nodi di carico e le stazioni. Per i carichi, per motivi di privacy, non sono riportati i veri nomi degli utenti, ma solo una sigla identificativa:

- Stazioni:
  - 11: Piossasco,
  - 22: Casanova,
  - 44: San Rocco,
  - 47: Magliano.
- Carichi:
  - 1: ABT,
  - 2: LOC,
  - 7: ERB,
  - 13: RVA,
  - 15: ALT,
  - 23: MIR,
  - 24: FEU,
  - 28: USG,
  - 29: MIF,
  - 32: ITA,
  - 37: BGV,
  - 46: MIC.

Capitolo 2

Il problema di Power-flow

2.1 Introduzione teorica

Lo studio di power-flow, anche conosciuto come load-flow, è uno strumento numerico ampliamente utilizzato per calcolare i flussi di potenza sulle linee di un sistema elettrico. È ottenuto a partire dalla conoscenza delle condizioni di lavoro dei carichi e dei generatori connessi alla rete e il suo obiettivo è principalmente di ottenere un completo set di informazioni riguardo ai moduli e alle fasi delle tensioni di ogni bus della rete, oltre che ai flussi di potenza in ogni linea elettrica che compone il sistema. Lo schema logico che sta alla base del processo computazionale si sviluppa a partire dalla definizione delle variabili note e incognite nei vari bus della rete. Per ogni nodo sono presenti quattro variabili: modulo della tensione, fase della tensione, scambio con l'esterno di potenza attiva e di potenza reattiva. Ogni bus deve avere due variabili note e due incognite per rendere possibile l'utilizzo del processo di power flow poiché sono disponibili $2 \cdot N_{nodi}$ equazioni di governo della rete, che sono i bilanci di potenza attiva e reattiva ai nodi stessi. La matrice delle ammettenze, in grado di descrivere il rapporto causale tra tensioni ai nodi e correnti iniettate ai nodi, deve essere fornita o calcolabile dalla conoscenza dei parametri fisici della rete. A questo punto l'analisi può essere effettuata e le variabili incognite ai nodi ricavate. Una volta che modulo e fase delle tensioni sono note per ogni nodo, i flussi di potenza e corrente possono essere calcolate su ogni linea del sistema. Anche altre informazioni possono essere ottenute, come lo scambio con l'esterno di potenza reattiva ad ogni nodo o la potenza attiva assorbita/immessa dai nodi slack (il concetto di nodo slack verrà chiarito in seguito). In altre parole, il power-flow serve principalmente

per determinare lo stato operativo in cui il sistema deve lavorare per soddisfare le richieste di carico e di generazione. Le ulteriori informazioni che si possono ricavare da uno studio di load-flow sono le configurazioni dei trasformatori a rapporto spire variabile e dei PST (Phase Shifter Transformer), le perdite su ogni linea o totali nel sistema, accorgimenti su come gestire la rete per minimizzare il costo operativo o per risolvere congestioni. Il load flow è anche usato nello studio dei guasti di cortocircuito, nella pianificazione di futuri interventi strutturali della rete, in studi di stabilità (sia a regime che in transitorio), nel dispacciamento economico e nello unit commitment.

2.2 Semplificazioni

Molto spesso vengono adottate semplificazioni negli studi di load-flow, a seconda anche del suo utilizzo. Il diagramma monofase viene adoperato poiché il sistema è assunto essere equilibrato. Questo viene utilizzato per costruire il modello matematico della rete, ovvero dei carichi, dei generatori e delle linee di trasmissione. La matrice delle impedenze viene infatti costruita a partire dal modello monofase. Viene inoltre usato un sistema in "per unità" adottando una base opportunamente scelta. Il sistema viene assunto funzionante in condizioni di regime in ogni sua parte: frequenza, moduli delle tensioni e punti di lavoro di carichi e generatori.

2.3 Modello matematico

Le equazioni che regolano il power-flow sono ottenute a partire dalla matrice delle ammettenze. La matrice delle ammettenze è ricavata dai parametri fisici della rete elettrica e deve essere nota o comunque devono essere note le proprietà della rete così da essere ricavabile. Essa descrive la relazione tra tensioni e correnti in una rete elettrica. Per un sistema a n nodi:

$$[Y][V] = [I] \tag{2.1}$$

$$\begin{bmatrix} Y_{11} & \dots & Y_{1n} \\ \vdots & \ddots & \vdots \\ Y_{n1} & \dots & Y_{nn} \end{bmatrix} \begin{bmatrix} V_1 \\ \vdots \\ V_n \end{bmatrix} = \begin{bmatrix} I_1 \\ \vdots \\ I_n \end{bmatrix} \tag{2.2}$$

Dove [Y] è la matrice delle ammettenze, [V] è il vettore contenente le tensioni

ai nodi e [I] il vettore composto dalle correnti iniettate ai nodi.
L'equazione al nodo i-esimo è quindi:

$$I_i = \sum_{j=1}^{n} Y_{ij} V_j \tag{2.3}$$

La relazione tra la potenza attiva e reattiva in per unità iniettata nel sistema al nodo i-esimo e la corrente, sempre in per unità, iniettata nel sistema allo stesso nodo è:

$$S_i = V_i I_i^* = P_i + jQ_i \tag{2.4}$$

Dove I_i^* è il complesso coniugato della corrente iniettata nella rete al nodo i-esimo, P_i e Q_i sono rispettivamente la potenza attiva e reattiva iniettate nella rete al nodo i-esimo.
Per cui,

$$I_i^* = \frac{P_i + jQ_i}{V_i} \tag{2.5}$$

$$I_i = \frac{P_i - jQ_i}{V_i^*} \tag{2.6}$$

$$P_i - jQ_i = V_i^* \sum_{j=1}^{n} Y_{ij} V_j = \sum_{j=1}^{n} Y_{ij} V_j V_i^*. \tag{2.7}$$

Siano

$$Y_{ij} = |Y_{ij}| \angle\theta_{ij} \tag{2.8}$$

e

$$V_i = |V_i| \angle\delta_i \tag{2.9}$$

con θ_{ij} fase del numero complesso che descrive l'ammettenza tra il nodo i e il nodo j e δ_i fase della tensione al nodo i.
L'equazione (2.7) si può quindi riscrivere:

$$P_i - jQ_i = \sum_{j=1}^{n} |Y_{ij}||V_j||V_i| \angle(\theta_{ij} + \delta_j - \delta_i) \tag{2.10}$$

e da qui ricavare:

$$P_i = \sum_{j=1}^{n} |Y_{ij}||V_j||V_i| \cos(\theta_{ij} + \delta_j - \delta_i) \tag{2.11}$$

$$Q_i = -\sum_{j=1}^{n} |Y_{ij}||V_j||V_i| \sin(\theta_{ij} + \delta_j - \delta_i) \tag{2.12}$$

Come già è stato accennato nell'introduzione, ad ogni bus i-esimo sono associate quattro variabili, P_i, Q_i, V_i e δ_i e solo due equazioni, (2.11) e (2.12). Per questo motivo in uno studio di power flow, due delle quattro variabili devono essere definite e le restanti devono essere incognite. Il primo passo per risolvere un problema di load flow è quindi identificare le variabili, ed esse, dipendono dal tipo di nodo a cui si riferiscono. Vengono definiti tre tipi di nodi: PQ, PV e slack o $V\delta$.

- Nodo PQ. Un nodo senza generatori collegati ad esso è considerato un nodo di carico e viene chiamato nodo PQ, poiché sono noti i livelli di potenza attiva P e reattiva Q scambiate. Anche un nodo senza nulla collegato ad esso, o nodo di transito, è considerato un nodo PQ (con P e Q nulli).

- Nodo PV. Un nodo con almeno un generatore connesso viene trattato come un nodo PV, ovvero ad esso è noto lo scambio con l'esterno di potenza attiva e il livello di tensione.

- Nodo slack o $V\delta$. I nodi nel quale è collegato un generatore, come già detto, vengono definiti nodi PV. L'unica eccezione riguarda il nodo al quale è collegato il generatore di maggiore taglia oppure il nodo che si interfaccia con una rete elettrica esterna di grandi dimensioni (ad esempio la rete elettrica di un'altra nazione). Questo viene chiamato nodo slack o $V\delta$ e viene usato come riferimento per la fase della tensione. Il ruolo del nodo slack consiste anche nel prendersi carico di chiudere il bilancio di potenza attiva e reattiva nell'intero sistema. Quest'ultimo ruolo, nei sistemi reali, viene in realtà assunto da più nodi (è il caso dello slack distribuito). Lo sbilancio tra potenze generate e assorbite viene quindi suddiviso tra quest'ultimi.

Una volta definite le $2 \cdot N_{nodi}$ variabili note e avendo a disposizione le $2 \cdot N_{nodi}$ equazioni si può risolvere il sistema per calcolare le restanti variabili incognite nel sistema. È da notare però che il sistema da risolvere non è lineare e non può essere risolto analiticamente. È necessario utilizzare metodi di

tipo iterativo per trovare soluzioni numeriche. La procedura comune consiste nell'avanzare una stima iniziale delle tensioni incognite ad ogni nodo della rete (sia in modulo che in fase), sostituire queste stime nelle equazioni di power-flow e determinare la deviazione dei risultati calcolati dai valori di potenza desiderati, aggiornare le stime delle tensioni attraverso qualche algoritmo numerico (ad esempio Newton-Raphson) e ripetere il processo finché la discrepanza dalle potenze desiderate desiderate sono sufficientemente piccole.

Capitolo 3

Futuri investimenti sulle reti elettriche di potenza

3.1 Introduzione

Una rete elettrica di potenza è un sistema interconnesso composto da generatori, linee di trasmissione, trasformatori e sistemi di distribuzioni concepiti per adempire alla richiesta di carico degli utenti (residenziali, commerciali e industriali). Ad oggi, la maggior parte dell'energia elettrica è generata in poche e grandi unità produttive e quindi deve essere trasmessa, dalla rete di trasmissione, alla rete di distribuzione, fino a raggiungere il cliente finale. Il paradigma del sistema energetico attuale è quindi tendenzialmente centralizzato, con i flussi di potenza unidirezionali, dalle grosse centrali elettriche ai consumatori. È proprio questa impostazione monodirezionale e centralizzata che pone i limiti maggiori di impiego e sviluppo dell'infrastruttura elettrica. I compiti principali di un tradizionale sistema di controllo della rete si possono riassumere in:

- gestire il flusso monodirezionale dell'energia, dalle centrali elettriche agli utenti finali;
- garantire il perfetto bilancio di potenza tra generatori e consumatori in quanto praticamente non è presente alcun sistema di accumulo nella rete;
- mantenere il livello di tensione e di frequenza in tutto il sistema all'interno di una certa banda di valori ben precisa.

Una rete tradizionale è in grado di adempire a questi tre compiti basilari, ma ha una limitata capacità di osservare le proprietà del suo complesso ambiente interno. Quest'ultima caratteristica pone una limitazione considerevole, poiché cela i limiti nei confronti di diverse funzionalità come la previsione dei trend di consumo e la capacità di sopperire con celerità a possibili perdite di generazione.

3.2 Smart Grid

Svariate decine di anni di gestione e studio delle reti elettriche hanno sottolineato che il modo migliore per ottenere un sistema il più possibile efficiente e sicuro è ottenuto interconnettendo quanto più possibile i vari componenti della rete, creando così un sistema molto più robusto, flessibile e resiliente. È questo il motivo che sta spingendo i sistemi elettrici di potenza verso un nuovo paradigma, rappresentato dal concetto di Smart Grid, ovvero un sistema altamente connesso e dotato di un sistema di controllo e gestione intelligente in grado di proteggere e ottimizzare le operazioni dei suoi elementi. Possiede molte caratteristiche innovative come la sua capacità di integrare ogni possibile fonte di generazione ed accumulo in modo più semplice e veloce (concetto di "plug-and-play"). Consente una partecipazione al mercato elettrico più consapevole anche per gli utenti, tramite la possibilità di fornire informazioni sempre aggiornate riguardo il prezzo e il consumo di energia. Queste informazioni sono fondamentali se si vuole creare un mercato più inclusivo e qualora si voglia pianificare le produzioni o i sistemi di riserva secondo criteri economici. Le Smart Grid sono in grado di ottimizzare i costi di manutenzione tramite interventi più mirati e massimizzando le risorse disponibili tramite algoritmi di controllo avanzati. Il nuovo paradigma garantisce una maggiore qualità dell'energia elettrica fornita, soddisfando le sempre più esigenti necessità dei dispositivi elettronici, attraverso un programma di auto intervento sui guasti e continua verifica del sistema. È in grado di identificare, analizzare e agire per ripristinare componenti o sezioni di linee in stato di guasto. Il risultante sistema elettrico sarebbe quindi molto più resiliente nei confronti di fenomeni naturali o attacchi di tipo sia fisico che informatico da parte dell'uomo.

Le numerose nuove funzioni introdotte dalle Smart Grids sono rese possibili grazie all'introduzione di varie tecnologie, impiegate sia nella rete di trasmissione che nella rete di distribuzione. Verranno raggruppate in seguito, a

seconda della loro applicazione principale:

- Capacità di consentire agli utenti una partecipazione attiva sul sistema elettrico:
  - Smart meter: dispositivi di misura in tempo reale, al quale viene integrato un sistema di comunicazione bidirezionale.
  - 'Vehicle-to-grid': i veicoli connessi alla rete elettrica posso partecipare a servizi ancillari, ad esempio la possibilità dei veicoli di svolgere una funzione di accumulo distribuito sta attirando grande interesse.
  - Micro Grids: rete elettriche di dimensioni ridotte composta da sistemi di generazione distribuita, sistemi di accumulo e una connessione con la rete elettrica principale interrompibile a seconda dell'esigenza. Possono essere concepite sia per funzionare in corrente alternata che in corrente continua. La struttura delle microreti permette di funzionare in modalità isolata, in caso di disservizi presenti nella rete principale, senza però compromettere la possibilità di effettuare scambi con il sistema esterno in caso di necessità o convenienza economica.
- Aumento della portata dei componenti fisici della rete:
  - Superconduttori: il loro utilizzo ridurrebbe significativamente le perdite nella rete.
  - HVDC (High Voltage Direct Current): numerosi studi dimostrano i vantaggi di un sistema di trasmissione in corrente continua. Le perdite per effetto Joule sono ridotte, oltre ad evitare problemi di stabilità. Le linee in alta tensione in corrente continua permettono di trasmettere energie su lunghissime distanze e con portate maggiori. Sistemi in corrente continua sono anche utili per interfacciare due sistemi a frequenze diverse.
  - Dynamic Line Rating: adatta il valore dei parametri strutturali delle linee a seconda dei flussi sulle linee stesse e sulle condizioni metereologiche.
  - FACTS (Flaxible Alternating Current Transmission System), VVC (Volt-VAR Control): entrambi i sistemi assistono i flussi di potenza e aiutano a mantenere un'alta qualità dell'energia trasmessa.

  - CVR (Conservation Voltage Reduction): implementato in una rete di distribuzione per controllare la tensione nella sua banda di tolleranza inferiore. Quest'accorgimento permette un risparmio energetico in presenza di utenze i quali consumi sono direttamente proporzionali alla tensione di alimentazione.
  - FCL (Fault Current Limiters): dispositivi in grado di limitare la corrente di guasto, in caso di cortocircuito, senza disconnettere completamente le sezioni incriminate.

- Miglioramento del software di controllo e ottimizzazione:
  - WAMS (Wide Area Monitoring System): è un sistema di monitoraggio che agisce in tempo reale e su larga scala. Ad esempio, il monitoraggio della fase della tensione su un'ampia zona permette di identificare e correggere instabilità, grazie ad una più efficace conoscenza dello status del sistema.
  - PMU (Phase Monitoring Units): è un dispositivo che misura le varie proprietà del segnale elettrico, come il modulo e la fase della tensione, sull'intero sistema, usando un segnale di sincronizzazione comune, solitamente fornito tramite GPS.
  - Algoritmi predittivi: questi tipi di algoritmi sono utilizzati per analizzare l'enorme quantità di dati che le Smart Grids acquisirebbero e stoccherebbero. Da questi database si possono ricavare importanti conclusioni e informazioni, come previsione di guasti, trend di consumo di energia elettrica e programmi ottimizzati di manutenzioni preventive.

Come si può evincere da quanto appena elencato, una delle funzioni sicuramente più importanti e chiave per attuare la transizione dalle tradizionali reti elettriche alle Smart Grids è un sistema in grado di effettuare misure in tempo reale, precise e dettagliate, in grado di connettersi e comunicare con il resto del sistema elettrico. Questo ruolo è assunto da una vera e propria infrastruttura basata su un completo set di componenti hardware e software e viene chiamata AMI (Advanced Metering Infrastructure).

3.3 AMI

Questa nuova infrastruttura di monitoraggio della rete adempie al compito di attuare misure intelligenti sulla rete, di comunicare i dati acquisiti e di ricevere segnali provenienti dal resto del sistema elettrico. È sicuramente uno

dei più importanti e fondamentali componenti di tutta l'infrastruttura intelligente che caratterizzano le Smart Grid. Può essere persino considerato come le fondamenta dei sistemi elettrici emergenti. L'AMI si basa su dispositivi di misura avanzati, un sistema di controllo efficiente, un protocollo di comunicazione ad alta velocità, un sistema di gestione dati avanzato e di tecnologie di storage dei dati acquisiti. La costruzione di questi sistemi permetterebbe la nascita di una nuova relazione di scambi bidirezionali di informazioni, oltre che di energia, tra i sistemi di trasmissione/distribuzione e gli utenti. Queste caratteristiche porterebbero ad una migliore qualità del servizio offerto e della energia fornita. Il centro di controllo sarebbe in grado di acquisire in tempo reale informazioni sul consumo di energia e sull'assorbimento di potenza degli utenti, così come i valori di tensione e corrente, arricchiti con informazioni di tipo temporale. La grande quantità di dati acquisibili permetterebbe anche l'identificazione di eventuali frodi di energia. Non sarebbe solo il centro di controllo a beneficiare di tutte queste nuove funzionalità che introdurrebbe questa infrastruttura, anche gli utenti sarebbero in grado di ottenere informazioni di interesse, come informazioni dettagliate dei propri consumi o sul prezzo dell'energia elettrica.
Le principali funzioni dell'AMI sono quindi:

- Comunicazione bidirezionale: l'infrastruttura rende possibile lo scambio di segnali e di dati tra il centro di controllo, la rete di potenza e gli utenti. Il sistema di comunicazione non verrebbe usato solo per il sistema di metering, infatti può anche essere usato per attivare switch da remoto.

- Supporto all'utente: informazioni quali il prezzo orario dell'energia garantirebbero un controllo automatico del carico elettrico. Gli utenti sarebbero automaticamente incentivati a consumare maggiormente quando il prezzo dell'energia è basso, ovvero quando la richiesta di energia complessiva nel sistema è limitata. Il profilo di consumo totale nell'intero sistema durante l'intera giornata andrebbe a stabilizzarsi su valori più costanti, evitando picchi e portando così a notevoli vantaggi in merito a congestioni e alla pianificazione delle produzioni nelle centrali elettriche.

- Supporto alla generazione distribuita. Fornendo un sistema in grado di supportare una comunicazione bidirezionale delle informazioni, la connessione dei dispositivi da parte degli utenti sarebbe semplificata.

- Funzione di risoluzione automatica di problemi nel sistema di comunicazione. Questa funzione sarebbe in grado di risolvere automaticamente

possibili problemi nel sistema di comunicazione ad esempio riconfigurandola in caso di disconnessioni.

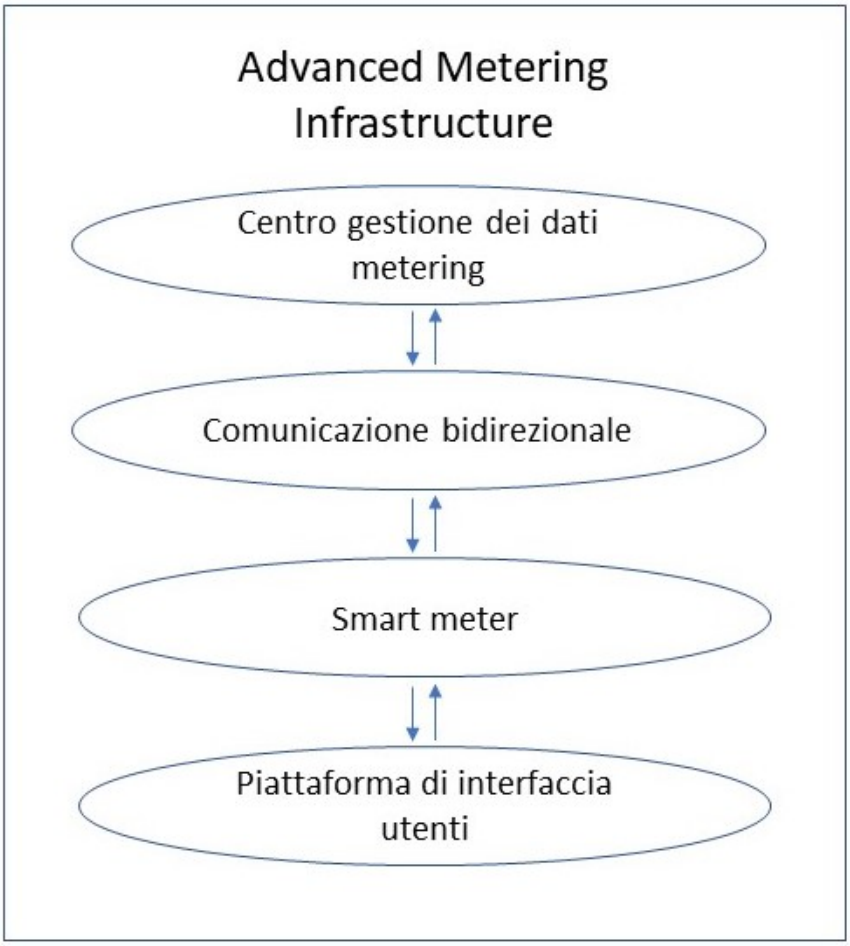


Figura 3.1. Schema logico della struttura di un AMI.

L'Advanced Metering Infrastructure (AMI) può anche essere concepito come suddiviso in quattro livelli logici come nella figura 3.1.

- Il primo livello è costituito dal centro di gestione dei dati, è la stazione informatica principale che acquisisce ed elabora i dati acquisiti dalla rete. Costituisce anche il centro di gestione del sistema di comunicazione.
- Il secondo livello è costituito dal sistema di comunicazione stesso. Ad oggi il sistema di comunicazione si basa su quattro possibili protocolli di

comunicazione: MCWILL, PLC, GPRS e RF.

- Il terzo livello non consiste solo nei dispositivi intelligenti di misura (smart meter), ma anche in un sistema di terminali per accedere o interfacciarsi alla rete, come i generatori di piccola taglia o apparecchiature in grado di interagire con il sistema elettrico (per collezionare dati, per funzioni relative alla vendita di energia o per assistenza clienti da parte del gestore della rete).

- L'ultimo livello svolge la funzione di interfaccia tra gli utenti e il sistema. Consiste in un terminale intelligente ed equipaggiato di uno display interattivo dove sono consultabili informazioni riguardanti consumi, prezzi dell'energia e comunicazioni da parte delle compagnie esercenti del sistema. Il terminale è dotato anche di unità di comunicazione wireless che potrebbero essere utilizzate per monitorare e controllare gli elettrodomestici da remoto. Con quest'ultima funzione l'AMI sarebbe in grado non solo di consentire una interazione bidirezionale tra utenti, sistema elettrico e gestori della rete, ma anche di aprire il sistema verso uno stadio successivo al quale ci si riferisce con il termine IoT (Internet of Things). Ovvero un sistema completamente interconnesso tramite la rete internet.

3.4 Smart meter

Come già rimarcato, le funzionalità delle smart grid si basano largamente sulla osservabilità del proprio sistema. Gli smart meter sono dispositivi introdotti proprio per attuare misure in tempo reale di varie grandezze caratterizzanti del sistema elettrico e di comunicare i dati ottenuti al resto della rete. Sono dispositivi derivanti dai tradizionali dispositivi di misura (contatori) ed infatti mantengono la loro principale caratteristica di misuratori di potenza ed energia elettrica, ma hanno una precisione molto maggiore. Le funzionalità più innovative che introducono sono però, la possibilità di salvare le misure effettuate correlandole con un dato di carattere temporale, salvare e riportare eventi, acquisire e salvare diversi tipi di dati, riportare all'utente il prezzo dell'energia. Questi strumenti di misura vengono equipaggiati con un modulo che permette una comunicazione bidirezionale con il resto del sistema. Il tradizionale paradigma di comunicazione unidirezionale nel quale le informazioni potevano solo essere raccolte dal centro di controllo sarebbe reso obsoleto. Questa nuova possibilità consente agli utenti di partecipare attivamente al mercato elettrico e alla produzione di energia elettrica. Il centro

di gestione del metering è situato nel centro dati della rete di potenza e ha il compito di salvare, analizzare e gestire il sistema di metering e di fornire informazioni basilari per il decision-making della Smart Grid.

3.5 Smart Grid e il problema della stima dei carichi

Anche se le necessità del futuro mercato elettrico stanno portando l'intero sistema di distribuzione e trasmissione verso una dimensione "smart", questa transazione è ancora all'inizio oggi. Un sistema equipaggiato di un AMI non avrebbe più il problema della sotto-determinazione del suo stato. Ogni strumento di misura, così come ogni altra apparecchiatura, sarebbe costantemente connessa al sistema di controllo. In particolare, a riguardo del problema che questo elaborato si propone di risolvere, non sarebbe più necessario tentare di stimare la potenza assorbita dai vari utenti connessi alla rete di alta tensione, poiché ne sarebbe disponibile una loro misura in tempo reale, ad opera degli smart meter. Ad oggi, però, gli utenti non sono equipaggiati con questo tipo di strumentazioni. Le misure, sia di potenza attiva che di potenza reattiva, sono ottenute al quarto d'ora (medie sul quarto d'ora) e sono trasmesse al TSO a fine giornata, spesso con ritardi. Sebbene i permessi di allaccio alla rete di trasmissione sono concessi solo se l'utente si impegna ad equipaggiarsi con un contatore in grado di comunicare in tempo reale con il TSO, gli utenti già connessi non sono obbligati, da contratto, a sostituire i contatori precedentemente installati e, a causa dell'onerosità di una tale operazione, questi non vengono sostituiti (l'istallazione di un nuovo apparecchio di misura comporta un costo di circa 50000€). Il TSO sarebbe quindi costretto a prendersi carico della sostituzione dei dispositivi di misura comportando un investimento molto importante. Sebbene in futuro, come già detto, allo scopo di attuare la transizione dell'attuale sistema verso una dimensione "smart", questi investimenti sarebbero comunque inevitabili, potrebbe essere saggio attendere, poiché in ogni caso le Smart Grid sono un sistema ancora sotto studio e in sviluppo. I dispositivi istallati potrebbero rivelarsi inadeguati.
L'osservabilità del sistema non sarebbe importante solo in ottica Smart Grid, ma anche a causa del diffondersi della generazione distribuita e rinnovabile. Molto spesso questa forma di generazione non è programmabile ed introduce un importante grado di imprevedibilità nel sistema. Più la generazione rinnovabile si diffonde più impegnativa diventa la gestione del sistema. Siccome

le reti in media e bassa tensione stanno quindi diventando sempre meno prevedibili, è necessario rendere completamente controllabili almeno le reti in altissima e alta tensione.
Per queste ragioni, il presente lavoro si propone di trovare un metodo sufficientemente efficace per ottenere pseudo-misure della potenza assorbita dalle utenze industriali, così da ottenere comunque una buona osservabilità della rete in alta tensione e permettere di rimandare l'oneroso adeguamento del sistema di metering.

Capitolo 4

Regressione lineare

4.1 Introduzione

La regressione lineare è un approccio statistico utilizzato per descrivere e modellizzare la relazione causale tra una grandezza (o variabile dipendente) e una o più variabili indipendenti. Se la relazione utilizza solo una variabile dipendente allora viene chiamata regressione lineare semplice, nel caso di più variabili indipendenti viene invece chiamata regressione lineare multipla. Infine, se la risposta è in forma vettoriale, prende il nome di regressione lineare multivariata. A causa della sua semplicità e della sua capacità di descrive una grande varietà di fenomeni naturali, la regressione lineare è un metodo largamente usato. Le sue principali applicazioni oggi possono essere riassunte in:

- Previsione. La regressione lineare può essere utilizzata efficacemente come modello predittivo. Fornito un dataset composto da una popolazione di osservazioni, ovvero coppie di variabili dipendenti e le relative variabili indipendenti, viene stabilita una relazione che in seguito può essere usata per predire risposte in caso di variabili indipendenti non comprese nel dataset iniziale.

- Analisi di correlazione tra variabili. Una regressione lineare è in grado di individuare e descrivere una variazione della variabile dipendente che può essere attribuita a qualche variabile indipendente. Lo strumento sarebbe anche in grado di quantificare il grado della relazione tra due variabili, oppure identificare qualora non ci sia alcuna relazione tra le due o se sono state fornite informazioni ripetitive nel dataset delle osservazioni.

In questo lavoro la regressione lineare verrà utilizzata per cercare e, nell'eventualità, sfruttare una qualche relazione tra le informazioni della rete, disponibili in tempo reale, e le grandezze che si vogliono predire.

4.2 Teoria

4.2.1 Introduzione alla teoria

Per un disponibile dataset composto da n unità statistiche nella seguente forma $\{y_i,\ x_{i1},\ \ldots,\ x_{ip}\}_{i=1}^{n}$, la regressione lineare è un metodo analitico sviluppato per identificare la relazione che meglio descrive il legame tra le variabili dipendenti y_i e le variabili indipendenti x_{ij} del dataset fornito. Nel modello lineare viene anche introdotto un termine di disturbo, chiamato errore statistico ϵ. Quest'ultima variabile è una quantità casuale e non osservabile introdotta per modellizzare il rumore nel sistema e raggruppa tutti gli effetti sulla variabile dipendente che non si possono attribuire alle variabili indipendenti del modello. La relazione tra l'errore e i vari regressori (variabili indipendenti) è un aspetto fondamentale e assolutamente non trascurabile del modello, la loro correlazione è infatti cruciale nella determinazione di quale metodo di stima sia più adatto.
La relazione è quindi nella forma

$$y_i = \beta_0 1 + \beta_1 x_{i1} + \cdots + \beta_p x_{ip} + \epsilon_i = x_i^T \beta + \epsilon_i \qquad (4.1)$$
$$i = 1, \ldots, n,$$

dove la T in apice indica la trasposizione.
Le n equazioni possono essere espresse in modo più efficace tramite la rappresentazione matriciale

$$y = X\beta + \epsilon, \qquad (4.2)$$

con

$$y = \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{pmatrix}, \qquad (4.3)$$

$$X = \begin{pmatrix} x_1^T \\ x_2^T \\ \vdots \\ x_n^T \end{pmatrix} = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & \ldots & x_{np} \end{pmatrix}, \qquad (4.4)$$

$$\beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{pmatrix}, \tag{4.5}$$

$$\epsilon = \begin{pmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_n \end{pmatrix}. \tag{4.6}$$

Sono necessari alcuni commenti a riguardo della notazione e della terminologia:

- y è il vettore composto dalle osservazioni y_i $(i = 1, \ldots, n)$. Queste variabili sono chiamate variabili dipendenti, variabili endogene, variabili predette, variabili risposta o criteri. Nel caso in cui la variabile sia indicata con $\hat{y}$ allora si riferisce ad un valore stimato tramite il modello.
- X è la matrice dei vettori riga x_i^T che rappresentano le osservazioni delle variabili indipendenti x_{ij}. Le variabili x_{ij} sono note come regressori, variabili esogene, variabili esplicative, variabili covariate, variabili predittive o variabili indipendenti. X è chiamata matrice dei regressori.
- β è il vettore dei coefficienti. La sua dimensione può essere sia $p + 1$, se nel suo modello è prevista l'intercetta (coefficiente β_0), che solo p se l'intercetta non è prevista. L'intercetta è il coefficiente relativo al termine costante del vettore x. Gli elementi del vettore β sono detti effetti, coefficienti angolari o semplicemente coefficienti della regressione. L'elemento β_i del vettore β può essere interpretato come la derivata parziale della variabile dipendente rispetto al regressore x_i. La determinazione del modello consiste sostanzialmente nella determinazione dei coefficienti β.
- ϵ è il vettore composto dagli errori ϵ_i, chiamati anche termini di disturbo o termini di rumore.

Sono necessarie anche alcune osservazioni:

- Per un dataset disponibile, è necessario stabilire quale delle variabili modellizzare come variabile dipendente e quale come variabili indipendenti. La decisione può essere semplicemente presa in modo intuitivo qualora

si riesca a presuppore che una variabile dipenda dalle altre, altrimenti è opportuno scegliere la variabile computazionalmente più vantaggiosa.

- La relazione rimane lineare finché rimane lineare nel vettore dei coefficienti β. Per esempio, un regressore può anche essere una funzione non lineare dei dati in ingresso (come nel caso della regressione polinomiale) senza influenzare le proprietà lineari del modello. Un esempio di regressione polinomiale può essere:

$$y = x_0 + \beta_1 x_1 + \beta_2 x_1^2 + \epsilon. \tag{4.7}$$

4.2.2 Assunzioni

I modelli delle regressioni lineari sono ottenuti solitamente a partire da una serie di assunzioni che, se legittime, permettono di semplificare notevolmente i processi computazionali come la stima dei coefficienti. Ad esempio, software come Matlab o Excel, svolgono studi di regressione applicando algoritmi di stima (Ordinary Least Squares) che sono opportuni solo in determinate condizioni. In caso le ipotesi non sono state avanzate in maniera legittima si arriva a risultati non corretti e ingannevoli. Numerose estensioni sono state sviluppate per sopperire all'eventualità di assunzioni non applicabili, in modo completo o parziale, al prezzo di un modello più complicato. In questi casi il problema risulta essere computazionalmente più pesante oppure potrebbe richiedere un dataset più consistente. Le ipotesi più importanti, richieste dagli algoritmi di stima, sono le seguenti: linearità, esogeneità debole, omoschedasticità, indipendenza degli errori e assenza di multicollinearità tra le variabili predittive. Le assunzioni vengono adesso riprese in dettaglio:

- Linearità: significa che la risposta è una combinazione lineare dei coefficienti e dei regressori. Questa ipotesi è molto meno restrittiva di quello che può sembrare ed è semplicemente una restrizione sui coefficienti. I regressori possono infatti essere manipolati in modo arbitrario (ma non lineare). Per esempio, si possono creare varie coppie delle variabili predittive e ognuna di queste trasformata a piacere: x_i può essere aggiunto nel modello anche sotto forma di x_i^2 come nell'equazione 4.7 e quest'ultimo trattato come una nuova variabile x_j (con $j \neq i$) negli algoritmi di stima. Un'altra manipolazione possibile può anche essere una moltiplicazione tra due regressori differenti etc. Questa possibilità di manipolare le variabili indipendenti rende la regressione uno strumento con capacità di fitting talvolta anche troppo elevate. Il problema dell'overfitting

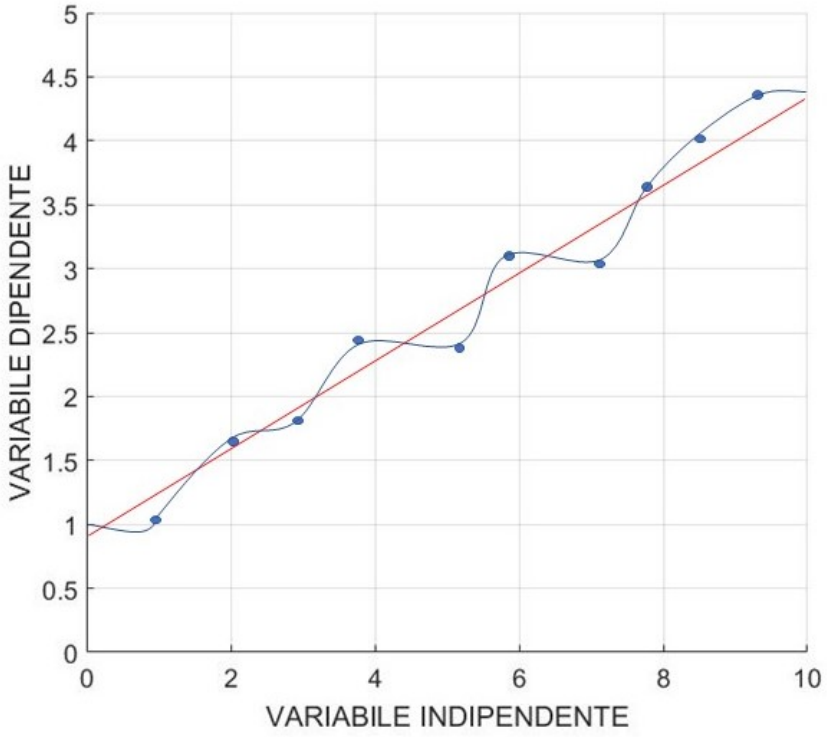


Figura 4.1. Rappresentazione grafica dell'overfitting. La linea rossa rappresenta la relazione corretta, la linea in blu quella ottenuta in caso di overfitting.

(figura 4.1) si può presentare nel caso di una regressione polinomiale. Il modello sarebbe in grado di descrivere con gran precisione le osservazioni contenute nel dataset iniziale, ma perderebbe la capacità di estrapolare la vera relazione tra le variabili, producendo risultati del tutto imprecisi se utilizzato al di fuori dello spazio del dataset iniziale. Nel caso venga utilizzata una regressione polinomiale, qualche tipo di regolarizzazione deve essere utilizzata, come la regressione lineare Bayesiana, per evitare di ottenere risultati svianti.

- Esogeneità debole. Questa proprietà permette di trattare qualunque delle variabili indipendenti come valori fissi. Ad esempio, l'interpretazione che si può attribuire al coefficiente i-esimo, ovvero di derivata parziale rispetto al regressore i-esimo, è possibile solo se si possono considerare fissi tutti gli altri regressori. Questo sostanzialmente significa che le variabili

predittive sono prive di errore (ad esempio prive di errori di misura) e non variabili casuali. Questa assunzione non è realistica per la grande maggior parte dei problemi, ma rilassarla, significherebbe adottare un modello molto più complicato poiché sarebbero da considerare anche gli errori presenti nelle variabili.

- Omoschedasticità o varianza costante. Questa ipotesi afferma che la varianza della variabile dipendente sia la stessa lungo l'intero dominio delle variabili indipendenti x. In altre parole, le risposte y dovrebbero essere ugualmente distribuite intono alla retta o all'iperpiano rappresentato dalla relazione per ogni valore possibile della variabile x (figura 4.2). Nella pratica, questa ipotesi non è quasi mai applicabile nel caso

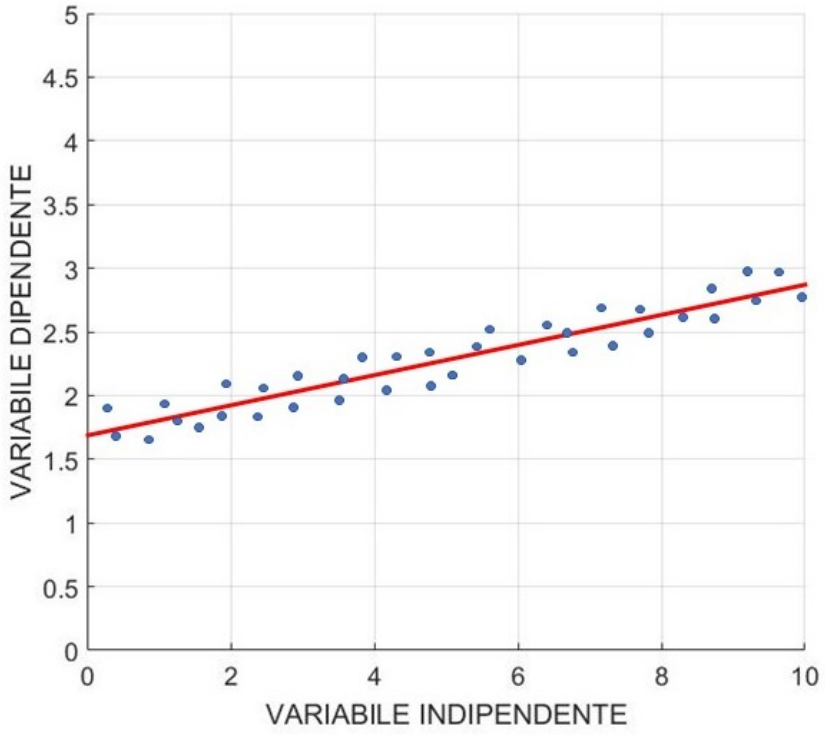


Figura 4.2. Rappresentazione grafica di una distribuzione omoschedastica degli errori.

il dominio delle variabili indipendenti sia ampio. Quando ciò accade, gli

errori sono definiti essere eteroschedastici. Tipicamente varianza e variabile dipendente sono direttamente proporzionali (figura 4.3). È possibile

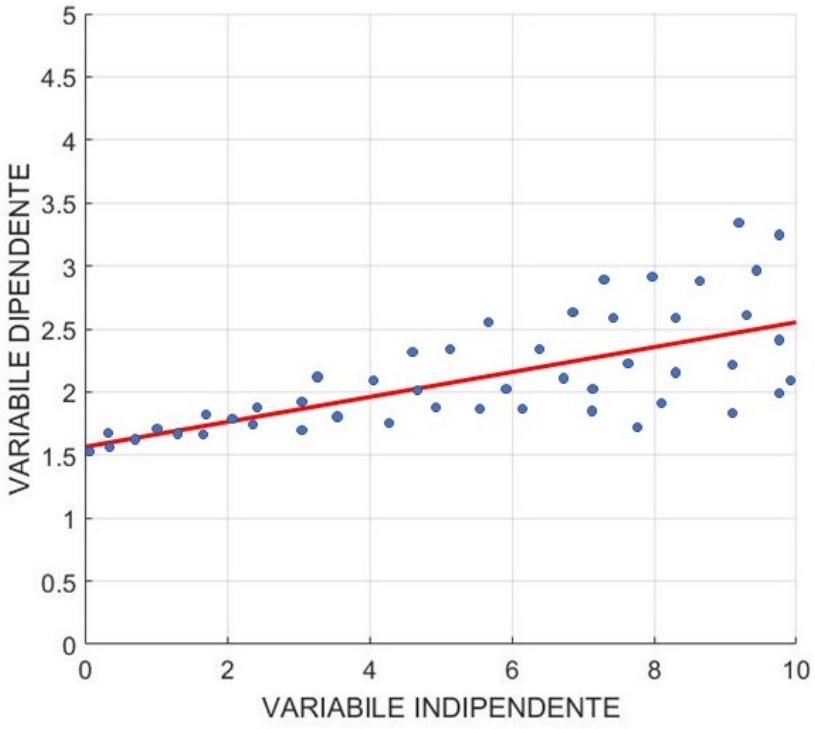


Figura 4.3. Rappresentazione grafica di una distribuzione eteroschedastica degli errori.

individuare questo comportamento semplicemente plottando la varianza in funzione della risposta. Quando l'ipotesi di omoschedasticità non è rispettata, gli algoritmi di stima solitamente usati fornirebbero coefficienti meno precisi e parametri statistici mal interpretabili (ad esempio l'errore quadratico medio, la varianza o la deviazione standard sarebbero scorretti). Comunque, sono stati sviluppati diversi algoritmi di stima dei coefficienti che permettono di rilassare questa assunzione, per esempio, i metodi dei minimi quadrati pesati o l'*heteroscedasticity-consistent standard errors* sanno gestire in modo buono la presenza di eteroschedasticità. Quando la varianza è considerata essere una funzione della risposta si può utilizzare anche la regressione lineare Bayesiana. In alcuni casi è possibile gestire il problema manipolando la risposta del sistema. Ad

esempio, si potrebbe considerare il logaritmo della variabile dipendente, così che avrebbe una distribuzione log-normale.

- Indipendenza degli errori. Questa assunzione è legittima quando gli errori delle variabili indipendenti sono indipendenti tra di loro. L'indipendenza statistica in realtà è un concetto più rigido di una semplice mancanza di correlazione e sebbene spesso non è necessaria, può essere utilizzata se presente. Alcuni metodi, come il metodo dei minimi quadrati generalizzato, è in grado di gestire la dipendenza degli errori anche se questo metodo necessita tipicamente di più dati per il processo di stima dei coefficienti. Alcuni tipi di regolarizzazione possono essere utilizzati in modo da condizionare il modello verso condizioni di indipendenza degli errori ed evitare il bisogno di un dataset maggiore. Infine, anche la regressione bayesiana è un modello generale efficace per gestire la dipendenza degli errori.

- Assenza di multicollinearità perfetta. La proprietà statistica della multicollinearità consiste nella presenza di una relazione lineare tra due colonne della matrice dei regressori X, ovvero si verifica quando due regressori sono linearmente dipendenti. Se il rango della matrice dei regressori non è massimo, questa non è invertibile rendendo inutilizzabile il metodo standard dei minimi quadrati. Questa condizione si incontra quando una variabile predittiva è accidentalmente fornita due volte (trasformare linearmente una delle due non risolve il problema) o quando ci sono più parametri da stimare rispetto alle osservazioni disponibili. In caso di multicollinarità perfetta, il vettore dei coefficienti β non può essere determinato poiché i processi standard di stima dei parametri non avrebbero una soluzione unica. In software come Matlab e Excel il problema è evitato riducendo la dimensione della matrice dei regressori, in caso di multicollinearità, prima di eseguire l'algoritmo di stima (solitamente viene utilizzato l'algoritmo OLS). Comunque, gli algoritmi di stima più generali non sono influenzati da questa condizione.

4.2.3 Algoritmi di stima

Gli algoritmi di stima sono tecniche sviluppate per stimare i coefficienti della regressione e sono sostanzialmente la componente principale di uno studio di regressione lineare. Un gran numero di tecniche sono state sviluppate e differiscono per la complessità dell'algoritmo, ipotesi necessarie, dimensione del dataset necessario, presenza di una soluzione in forma chiusa, immunità

rispetto a valori anomali nel dataset etc.
Alcuni degli algoritmi di stima più comuni sono metodi lineari ai minimi quadrati come il metodo dei minimi quadrati ordinari(Ordinary Least Squares, OLS), il metodo dei minimi quadrati pesati (Weighted Least Squares, WLS), il metodo dei minimi quadrati generalizzato (Generalized Least Squares, GLS) o metodi della massima verosomiglianza (Maximum-likelihood).

Metodi ai minimi quadrati - Minimi quadrati ordinari

Il metodo dei minimi quadrati ordinari (OLS) è il metodo più semplice e comune per la stima dei parametri negli studi di regressione lineare. Come suggerisce il nome, questo metodo è basato sul principio dei minimi quadrati. I parametri del modello lineare sono scelti in modo da minimizzare la somma del quadrato delle differenze tra le variabili dipendenti presenti nel dataset e le corrispondenti variabili dipendenti ottenute dalla relazione. Questo metodo può essere interpretato in modo geometrico, il valore da minimizzare può anche essere visto come la somma dei quadrati delle distanze tra l'iperpiano descritto dalla relazione e i corrispondenti punti nel dataset (figura 4.4). Intuitivamente la bontà del processo di fitting è tanto maggiore quanto minore è questa distanza, in realtà questo metodo non è in grado di identificare la presenza di valori anomali o l'overfitting, di conseguenza, anche se la somma dei quadrati è bassa ciò non significa che il metodo sia stato in grado di estrapolare la vera relazione tra le variabili.
L'OLS è un metodo di stima appropriato solo se le l'ipotesi di esogeneità debole è rispettata, mentre l'algoritmo è ottimale in caso di omoschedasticità e errori non correlati. Inoltre, in caso di distribuzione normale dell'errore, il metodo dei minimi quadrati rappresenta lo stimatore con massima verosomiglianza.
Per regressioni lineari questo metodo si può formalizzare matematicamente con relativa semplicità a partire dal seguente sistema sovradeterminato di n equazioni lineari con p coefficienti, β_0, β_1, β_2, …, β_p (con $p < n$ e uguale al numero di regressori):

$$y_i = \sum_{j=1}^{p} X_{ij}\beta_j, \qquad i = 1,2,\ldots,n \tag{4.8}$$

In forma matriciale diventa:

$$Y = X\beta, \tag{4.9}$$

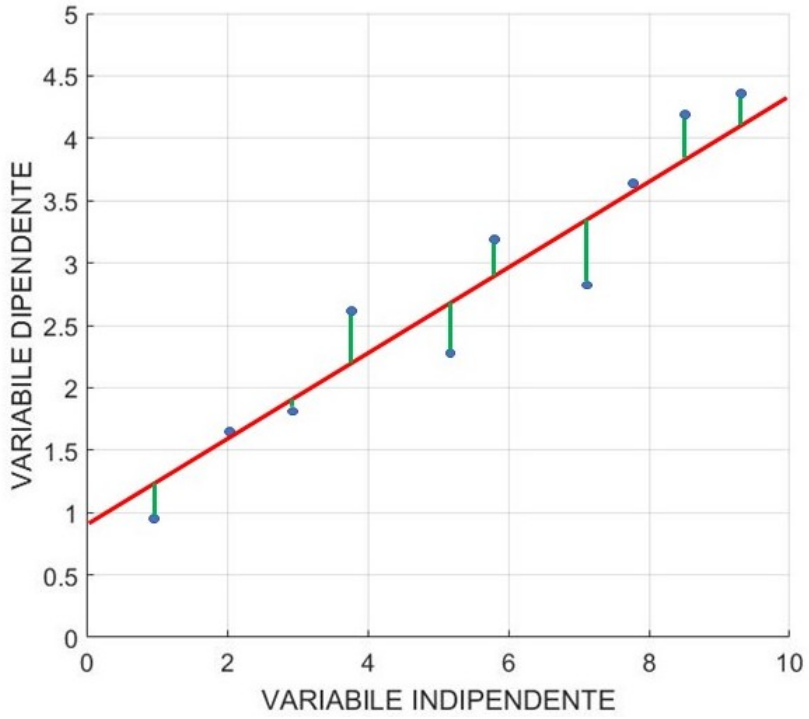


Figura 4.4. Interpretazione grafica OLS. In verde sono riportati gli errori tra osservazioni e modello. L'algoritmo OLS trova i coefficienti della regressione in modo da minimizzare questi errori.

dove i vari componenti sono nella forma come per l'equazione 4.2 . Per una relazione lineare come sopra, la prima colonna di X è spesso popolata interamente da 1 ($X_{i1} = 1$) e solo le altre colonne contengono vere informazioni riguardanti le osservazioni.
Un tale sistema solitamente non ha soluzione, quindi l'obiettivo del metodo diventa trovare i coefficienti β tali per cui minimizzino una funzione obiettivo. Nel caso dello stimatore OLS la funzione obiettivo S è data da

$$S(\beta) = \sum_{i=1}^{n} \left(y_i - \sum_{j=0}^{p} X_{ij}\beta_j \right)^2 = ||y - X\beta||^2 \tag{4.10}$$

che rappresenta la somma dei quadrati delle differenze tra le variabili dipendenti osservate e ottenute tramite il modello. Questo problema ha una sola soluzione solo se la matrice dei regressori X ha le p colonne linearmente indipendenti. La soluzione è ottenuta risolvendo il sistema

$$(X^T X)\hat{\beta} = X^T y. \tag{4.11}$$

La matrice $X^T X$ possiede l'importante proprietà di essere positiva semi-definita. Infine, $\hat{\beta}$ è il vettore dei coefficienti stimati dal metodo OLS e calcolato come segue:

$$\hat{\beta} = (X^T X)^{-1} X^T y. \tag{4.12}$$

Metodi ai minimi quadrati - Minimi quadrati pesati

Il metodo dei minimi quadrati pesati (Weighted Least Squares) è una generalizzazione del metodo dei minimi quadrati ordinari introdotto nella sezione precedente. Questo stimatore deve essere utilizzato nel caso l'ipotesi di omoschedasticità non è applicabile e la matrice di correlazione Ω dei residui risulta essere diversa dalla matrice identità. Questa condizione si incontra quando la varianza delle osservazioni non è costante. La matrice delle varianze Ω deve comunque essere diagonale, infatti questo metodo è considerato come un caso specifico del metodo dei minimi quadrati generalizzato, dove la matrice delle varianze Ω può essere completa. Riassumendo, il metodo dei minimi quadrati pesati deve essere usato quando la varianza delle osservazioni non è costante, però queste non devono essere correlate tra loro.
L'introduzione dei pesi è opportuna quando le osservazioni non sono tutte ugualmente affidabili. Il loro contributo nella sommatoria dei quadrati delle differenze viene così pesato tramite un coefficiente w_i proporzionale al reciproco della varianza della osservazione i-esima. La funzione obiettivo è allora

$$S(\beta) = \sum_{i=1}^{n} w_i \left(y_i - \sum_{j=0}^{p} X_{ij} \beta_j \right)^2 = ||W^{\frac{1}{2}}(y - X\beta)||^2 \tag{4.13}$$

Dove W è la matrice diagonale dei pesi.
Calcolando il minimo di $S(\beta)$ si trova la seguente equazione

$$(X^T W X)\hat{\beta} = X^T W y \tag{4.14}$$

e da questa, i parametri sono trovati con

$$\hat{\beta} = (X^T W X)^{-1} X^T W y \tag{4.15}$$

Metodi ai minimi quadrati - minimi quadrati generalizzati

Lo stimatore ai minimi quadrati generalizzati (Generalized Least Squares, GLS) è utilizzato quando è presente un certo grado di correlazione tra i residui del modello della regressione. Quando questa condizione si verifica, i metodi dei minimi quadrati ordinari e pesati, non dovrebbero essere utilizzati perché possono essere inefficienti o posso dare risultati ingannevoli.

Altre tecniche di stima dei coefficienti dei regressori

Un gran numero di metodi di stima sono stati sviluppati oltre a quelli ai minimi quadrati. Uno di questi è il metodo della massima verosomiglianza (Maximum Likelihood Estimation, MLE). Il metodo MLE può essere utilizzato quando la distribuzione dei residui ha caratteristiche note, come la sua famiglia parametrica $f\theta$ e la sua relativa varianza θ. Quando $f\theta$ è una distribuzione normale con varianza θ, il metodo fornisce stime identiche a quelle ottenibile con il metodo OLS, mentre, se i residui seguono una distribuzione normale multivariata, con matrice di covarianza nota, i risultati corrispondono a quelli ottenibili con il metodo GLS. Se il termine di errore e i regressori sono indipendenti, lo stimatore ottimale è allora il 2-steps MLE, dove il primo passo ha lo scopo di stimare la distribuzione del termine di errore in un modo non parametrico.

La regressione Bayesiana sfrutta i principi della statistica bayesiana, nel quale sono necessarie un certo numero di assunzioni soggettive preliminari allo studio statistico. Nel caso della regressione, i parametri del modello non vengono ottenuti come un solo preciso valore, ma piuttosto come una distribuzione di probabilità al quale i coefficienti appartengono. La classe $f\theta$ che descrive la distribuzione di probabilità deve essere scelta a priori, mentre è compito del metodo statistico bayesiano determinare il valore del parametro θ. È un approccio solitamente più immune al problema dell'overfitting. Quando le distribuzioni di probabilità dei coefficienti assunte a priori sono di un particolare tipo allora i metodi prendono il nome di regressione Ridge e regressione Lasso. Questi due metodi sono infatti interpretabili come casi particolari della regressione Bayesiana.

Le regressioni Ridge e LASSO (Least Absolute Shrinkage and Selection Operator) sono metodi di stima penalizzate. Sono solitamente utilizzate quando è presente il problema della multicollinearità. Quando la multicollinearità è verificata, le varianze ottenute nel modello ai minimi quadrati sono elevate e portano a produrre stime dei coefficienti lontane dai valori affidabili. Aggiungendo un parametro penalizzante, o bias, al modello, la variabilità dei coefficienti è ridotta. I processi di stima dei metodi penalizzati sono piuttosto simili al metodo dei minimi quadrati. I coefficienti della regressione vengono ricavati risolvendo un processo di minimizzazione, dove la funzione obiettivo è uguale a quella usata nel metodo OLS, ma con l'introduzione del termine

penalizzante λ. Per la regressione Ridge la funzione obiettivo è la seguente:

$$S(\beta) = \sum_{i=1}^{n} \left[y_i - \sum_{j=0}^{p} X_{ij}\beta_j \right]^2 + \lambda \sum_{j=1}^{p} \beta_j^2 \tag{4.16}$$

Maggiore è il valore di λ, più vicini a zero sono i coefficienti che si otterrebbero. La regressione Ridge non riesce ad escludere dal modello i regressori i quali coefficienti sono trascurabili rispetto agli altri coefficienti. Se invece si desidera escludere dal modello questi coefficienti è necessario utilizzare la regressione LASSO. Questo processo di stima è in grado di escludere i regressori meno influenti semplicemente tramite l'introduzione nella funzione obiettivo di un valore assoluto:

$$S(\beta) = \sum_{i=1}^{n} \left[y_i - \sum_{j=0}^{p} X_{ij}\beta_j \right]^2 + \lambda \sum_{j=1}^{p} |\beta_j| \tag{4.17}$$

In entrambi i modelli Ridge e LASSO l'intercetta non è soggetta alla penalizzazione.
Sia la regressione Ridge che la regressione LASSO sono spesso usate quando l'obiettivo è la previsione delle variabili dipendenti y per valori delle variabili indipendenti x non compresi nello spazio del dataset usato per il fitting.

4.3 Applicazione della regressione lineare

Il primo metodo considerato per la stima dei carichi è ottenuto attraverso l'utilizzo di una regressione lineare multipla. L'obiettivo di questo metodo è di trovare una relazione, per ogni carico da prevedere, nella seguente forma:

$$y_p = \beta_{p0} + \beta_{p1}x_{p1} + \beta_{p2}x_{p2} + \dots. \tag{4.18}$$

Dove β_p sono i coefficienti della regressione per il carico p-esimo, y_p è la variabile dipendente e x_{pi} sono le variabili indipendenti (come spiegato all'inizio di questo capitolo). In questo studio i regressori sono le misure disponibili in tempo reale (ovvero le misure dei flussi di potenza sulle linee in partenza dalle stazioni e che alimentano l'isola di carico, gli assorbimenti di potenza nelle cabine primarie e la potenza generata negli impianti di produzione) e la variabile dipendente è il carico da prevedere. Intuitivamente, questa relazione dovrebbe mettere in relazione un carico con la potenza in transito all'inizio della linea che lo alimenta, tutt'al più sottraendo, da quel valore, la

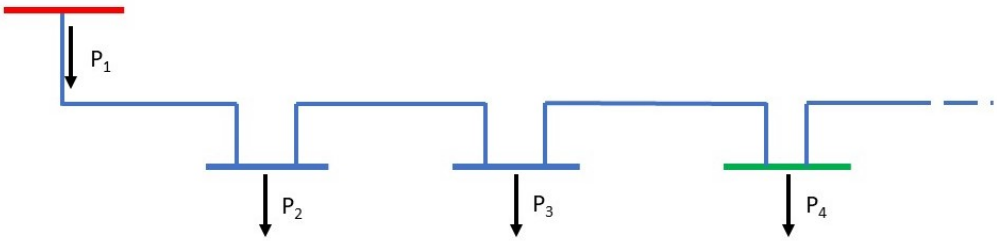


Figura 4.5. Esempio di relazione causale attesa. Il carico P4, è intuitivamente messo in relazione con la misura di potenza P1. In realtà, i coefficienti di correlazione ottenuti, sono molto bassi e non migliorano se a P1 si sottraggono gli assorbimenti delle cabine primarie P2 e P3.

potenza assorbita dalle cabine primarie situate tra il punto dove è effettuata la misura (inizio della linea) e il punto dove è connesso il carico (figura 4.5). Verrebbe infatti da ipotizzare la presenza di una relazione causale tra i valori menzionati. Il primo passo è stato quindi effettuare uno studio di correlazione per verificare se effettivamente è presente una relazione tra un carico ed una potenza in transito all'inizio della linea che lo alimenta (in realtà la rete ha una configurazione magliata, quindi non si può parlare in modo univoco di una linea che alimenta un carico, per questo si potrebbe analizzare la correlazione con la misura sulla linea elettricamente "più vicina"). L'analisi, però, dimostra che c'è una correlazione molto bassa tra di loro, raramente superiore a 0,5 (tabella 4.1). In un'analisi di correlazione un risultato pari a 1 indica una correlazione perfetta, 0 se non esiste alcuna relazione, valori negativi se hanno un rapporto inversamente proporzionale. Questo risultato può essere attribuito al fatto che il profilo della potenza in transito sulle linee in partenza dalle stazioni è influenzata maggiormente dall'assorbimento delle cabine primarie che non dai carichi (figura 4.6). La potenza assorbita delle cabine è infatti considerevolmente maggiore di quella assorbita dagli utenti industriali, oltre al fatto che le cabine sono anche in numero molto maggiore rispetto ai carichi. Se si calcola il coefficiente di correlazione tra il totale della potenza attiva immessa nell'isola di carico dalle stazioni e il totale di potenza attiva assorbita dalle cabine primarie sommato alla potenza attiva generata dagli impianti di produzione, il risultato è pari 0,969 (la forte correlazione si può notare anche dalla figura 4.6), mentre è solamente 0,341 tra la somma delle potenze attive iniettate nella rete dalle stazioni e la somma

	CA_642	CA_644	CA_646	CA_651	MA_724
ALT	0,01	0,42	0,48	0,26	0,54
ERB	0,18	0,49	0,43	0,31	0,54
FEU	0,16	0,26	0,31	0,28	0,45
ITA	0,05	0,47	0,52	0,30	0,60
MIC	0,08	0,34	0,32	0,17	0,34
MIF	-0,19	-0,26	0,15	0,11	-0,02
MIR	-0,24	-0,08	0,41	0,20	0,22
ABT	-0,18	0,09	0,33	0,11	0,15
LOC	0,11	-0,15	-0,21	-0,09	-0,20
USG	0,11	0,41	0,37	0,27	0,45
RVA	0,14	0,45	0,48	0,30	0,56
	MA_750	PI_587	PI_678	SR_706	SR_752
ALT	0,14	0,24	0,45	0,16	-0,15
ERB	0,24	0,44	0,53	0,15	-0,19
FEU	0,15	0,34	0,38	0,17	-0,09
ITA	0,17	0,32	0,54	0,21	-0,17
MIC	0,12	0,26	0,33	0,03	-0,11
MIF	-0,31	-0,08	-0,04	0,36	0,27
MIR	-0,24	0	0,16	0,37	0,26
ABT	-0,13	-0,08	0,08	0,17	0,06
LOC	0,03	-0,02	-0,23	-0,11	-0,02
USG	0,17	0,29	0,42	0,06	-0,17
RVA	0,19	0,40	0,44	0,17	-0,16

Tabella 4.1. CA: Casanova, MA: Magliano, PI: Piossasco, SR: San Rocco. I numeri che seguono nella sigla indicano il numero della linea (figura 1.2).

delle potenze assorbite dai carichi.
Considerando anche gli errori di accuratezza delle misure in tempo reale e la presenza di perdite Joule, risulta quindi chiaro che non è possibile descrivere il comportamento dei carichi tramite una vera relazione causale in una porzione di rete così ampia. Il "rumore" che influenza il modello è talmente importante che si può apprezzare semplicemente effettuando un bilancio, che a livello teorico, dovrebbe fornire le perdite Joule sull'isola. In figura 4.7 è riportato il grafico della seguente operazione: *potenze attive immesse nella rete dalle stazioni - (potenza attiva assorbita dalle cabine primarie + potenza attiva generata dagli impianti di produzione + carico attivo assorbito dalle*

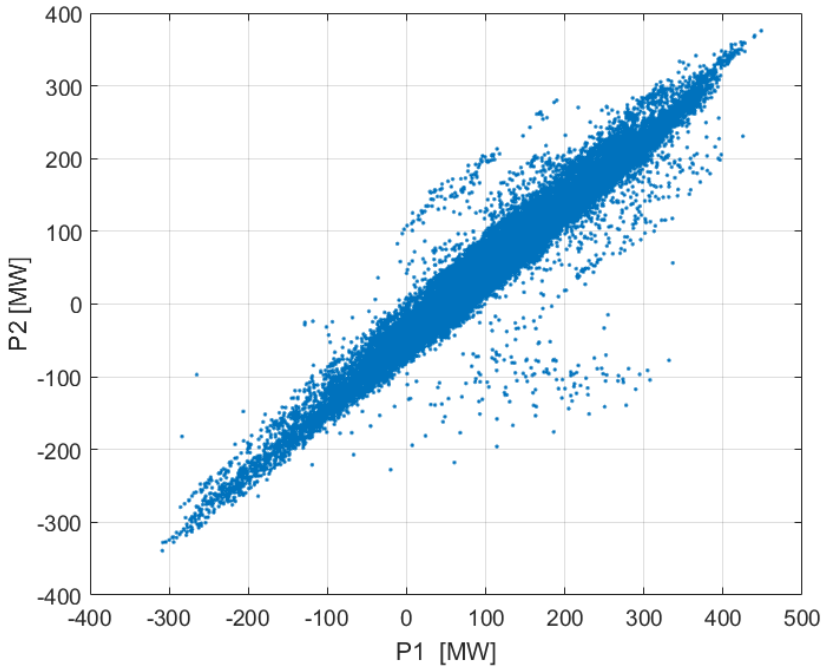


Figura 4.6. Plot tra la potenza iniettata nell'isola di carico dalle stazioni (P1) e la somma delle potenze assorbite dalle cabine primarie e dai generatori(esse saranno quindi negative)(P2)

utenze industriali). La relazione dovrebbe fornire come risultato le perdite Joule sull'isola ed invece i risultati non sono coerenti essendo talvolta pure negative.
L'idea innovativa proposta in questo lavoro è quella di predire i carichi tramite una relazione concepita non per sfruttare una qualche relazione causale, ma per sfruttare le similarità tra i profili d'assorbimento dei carichi e i profili relativi alle misure disponibili in tempo reale. L'analisi di regressione è stata svolta in 3 modi differenti:

- considerando come regressori tutte le misure dei flussi di potenza sulle linee in partenza dalle stazioni e che alimentano l'isola di carico in esame (10 regressori);
- considerando come regressori le misure già indicate al punto precedente più le misure delle potenze degli impianti di produzione (17 regressori);
- considerando come regressori le misure indicate nei due punti precedenti più le misure di assorbimento delle cabine primarie (47 regressori).

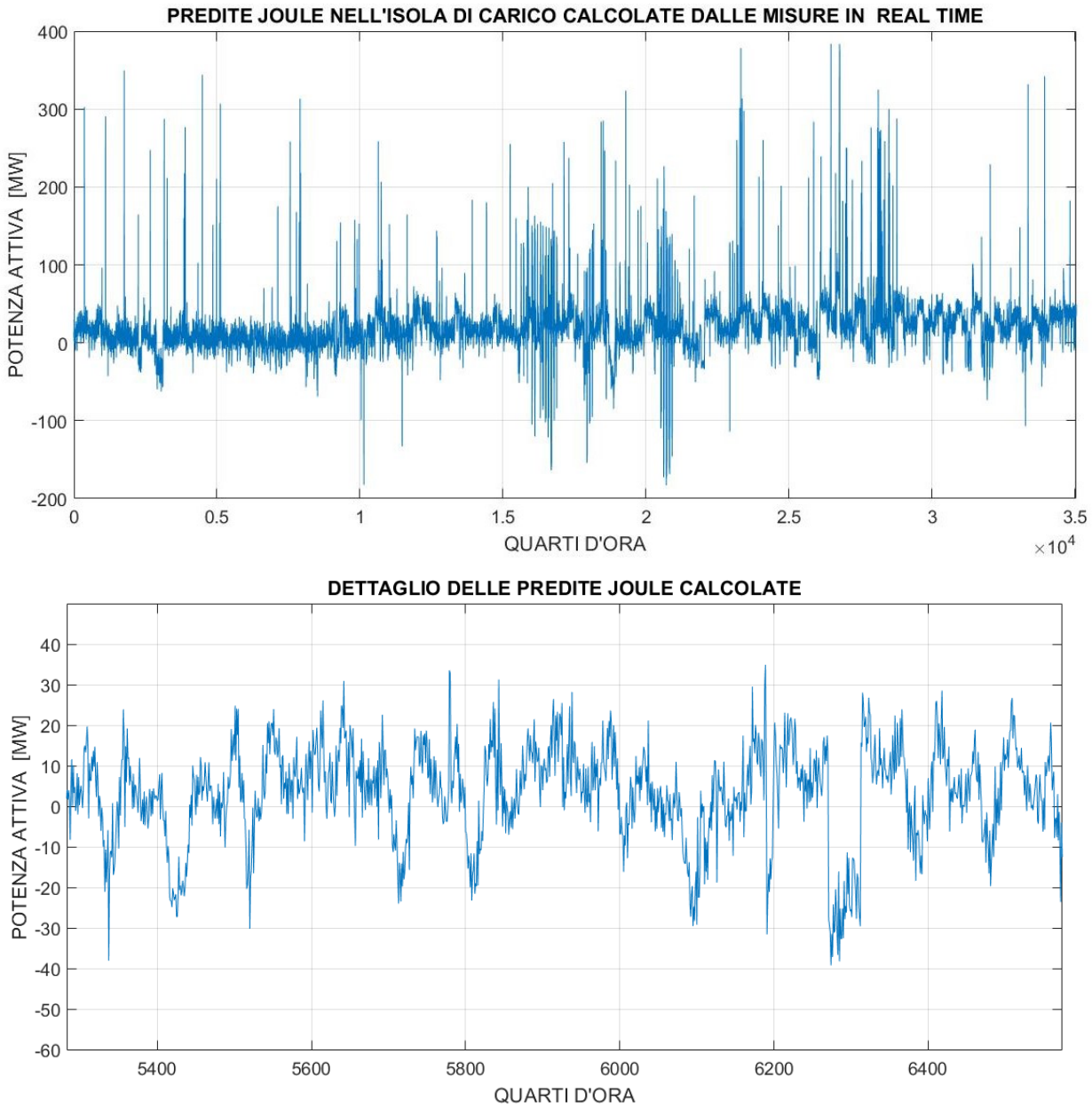


Figura 4.7. Plot del calcolo: *potenze attive immesse nella rete dalle stazioni – (potenza attiva assorbita dalle cabine primarie + potenza generata dagli impianti di produzione + carico attivo assorbito dalle utenze industriali)*. Esso dovrebbe rappresentare le perdite Joule sulle linee, ma da come si può vedere, il risultato non è coerente. Sono infatti talvolta negative e talvolta raggiungono i 300MW.

I risultati ottenuti sono simili nei primi due casi mentre migliorano sensibilmente solo considerando il terzo caso (figura 4.8). Sarebbe possibile,

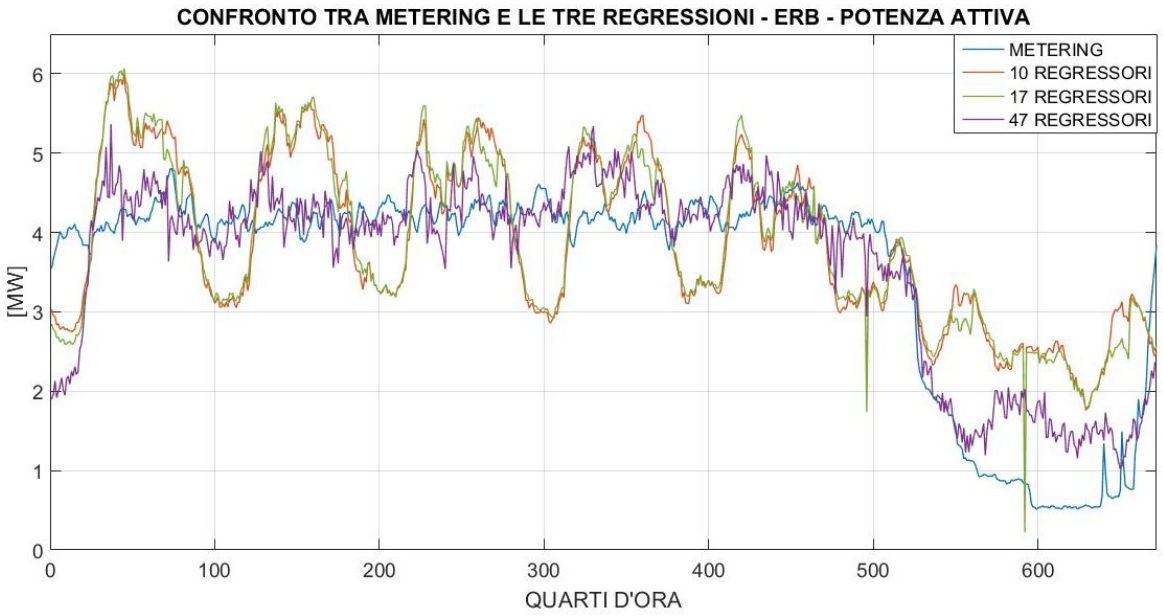


Figura 4.8. Confronto tra i tre metodi di regressione nella settimana tra il l'11 e il 18 dicembre 2017 per il carico ERB, potenza attiva. In blu è riportato il metering, in arancione i risultati stimati dalla regressione con 10 regressori, in verde la regressione con 17 regressori, in viola 47 regressori.

per ogni carico, considerare come regressori solo le misure più correlate (ad esempio le 5 o 10 misure più correlate). Questo semplificherebbe la relazione, ma non l'intero procedimento, poiché sarebbe necessario, per ogni carico, svolgere uno studio di correlazione e identificare le misure più correlate ed in seguito svolgere lo studio di regressione con quest'ultime. I regressori da utilizzare sarebbero quindi differenti volta per volta. È inoltre difficile immaginare che i risultati siano altrettanto precisi. Considerando infine che lo sforzo computazionale necessario per svolgere una regressione con 47 regressori è comunque molto limitato, quest'ultima impostazione è quella presa in considerazione nel prosieguo di questo lavoro.
I coefficienti della regressione sono stati calcolati per mezzo del software Matlab. Il dataset sul quale è stato svolto il processo di fitting consiste nelle misure di potenze al quarto d'ora nell'intero anno 2017 per tutti i 47 regressori e per tutti gli 11 carichi. I dati sono stati caricati nel workspace di Matlab e manipolati per renderli consistenti. Un certo numero di misure sono infatti

mancanti e sono state assunte pari all'ultima misura disponibile. Le unità di misura sono state uniformate a MW e MVar. Un paio di misure relative ai carichi non erano direttamente disponibili dal metering e sono state ottenute indirettamente da altre misure (in caso di un utente equipaggiato anche da qualche unità di generazione di una certa importanza, sono disponibili solo le misure al punto di scambio con la rete e le misure di produzione dei generatori. Il carico, in questi casi, è quindi ottenuto tramite un semplice bilancio). I coefficienti sono quindi ottenuti tramite la funzione "regress" che sfrutta l'algoritmo di stima OLS dopo aver eliminato qualche eventuale colonna linearmente dipendente nella matrice dei regressori.
Nelle figure 4.9, 4.10, 4.11 sono ora riportati alcuni risultati della stima di carico del mese di dicembre 2017 ottenuti con il metodo a 47 regressori, i quali coefficienti sono stati ottenuti sui primi 11 mesi dello stesso anno.

Bisogna sottolineare che dicembre è un mese altamente critico per la stima, poiché sono presenti numerosi giorni festivi infrasettimanali. Il carico durante questi giorni non è previsto con precisione poiché al processo di stima dei coefficienti (OLS) è stato fornito un dataset composto naturalmente da molti più giorni lavorativi che festivi. In questo modo i coefficienti sono stati "pesati" per prevedere con maggior precisione i carichi in giorni lavorati che festivi.

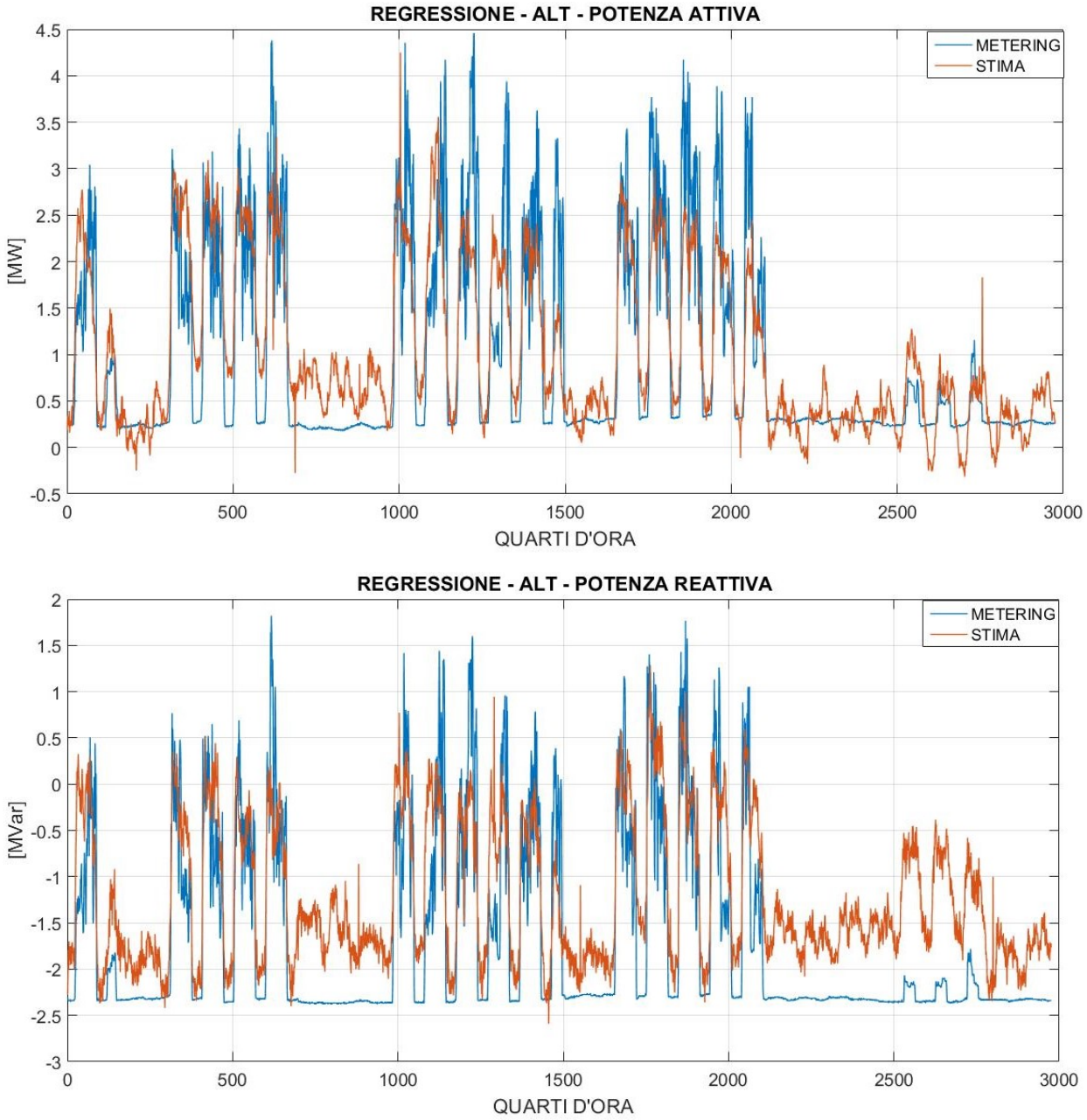


Figura 4.9. Grafico dei risultati della previsione di carico attivo e reattivo dell'utenza ALT ottenuti tramite la regressione. I coefficienti sono ottenuti tramite l'analisi dei dati relativi ai primi 11 mesi del 2017 e i valori plottati sono relativi all'intero mese di dicembre 2017.

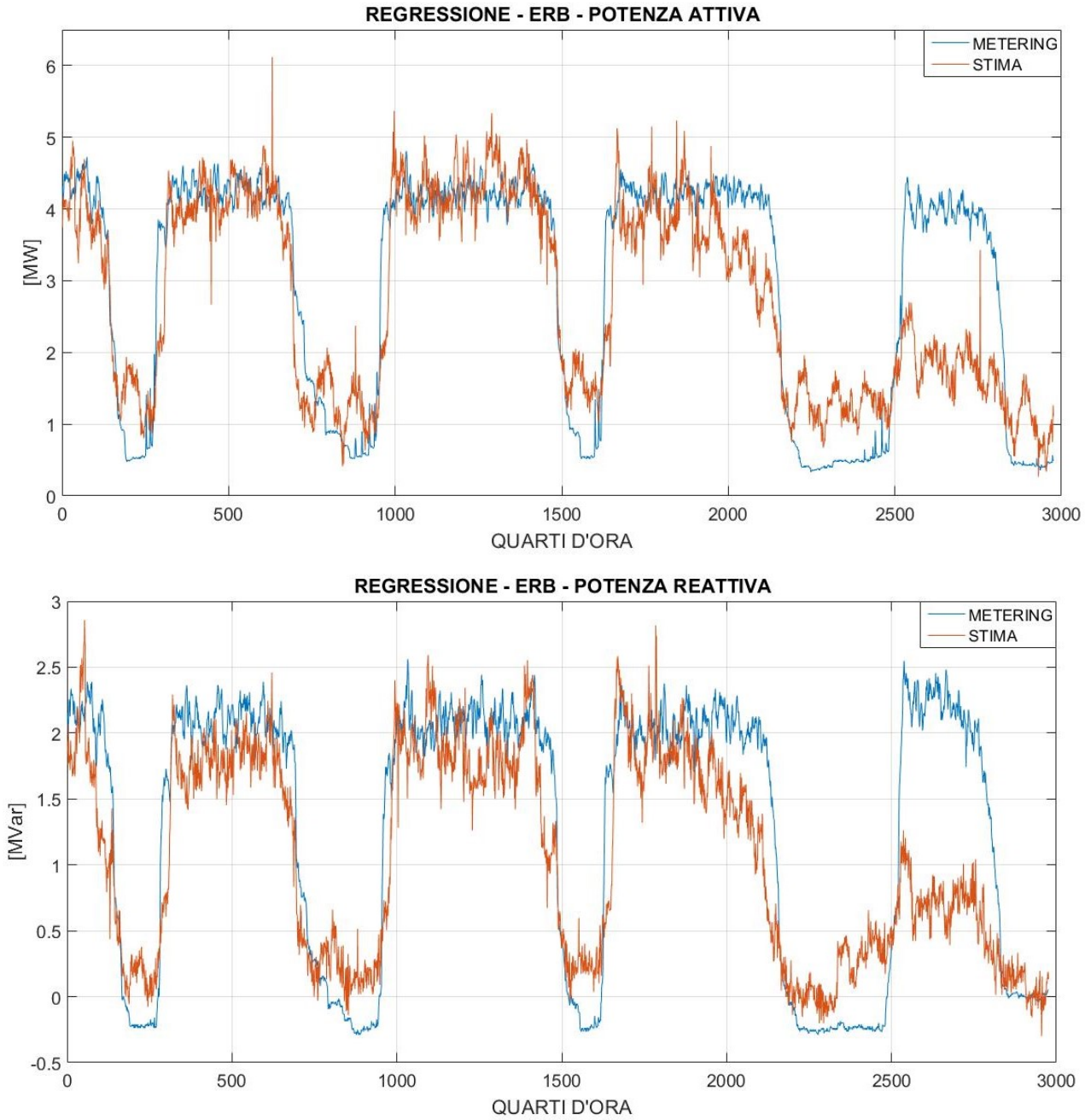


Figura 4.10. Grafico dei risultati della previsione di carico attivo e reattivo dell'utenza ERB ottenuti tramite la regressione. I coefficienti sono ottenuti tramite l'analisi dei dati relativi ai primi 11 mesi del 2017 e i valori plottati sono relativi all'intero mese di dicembre 2017.

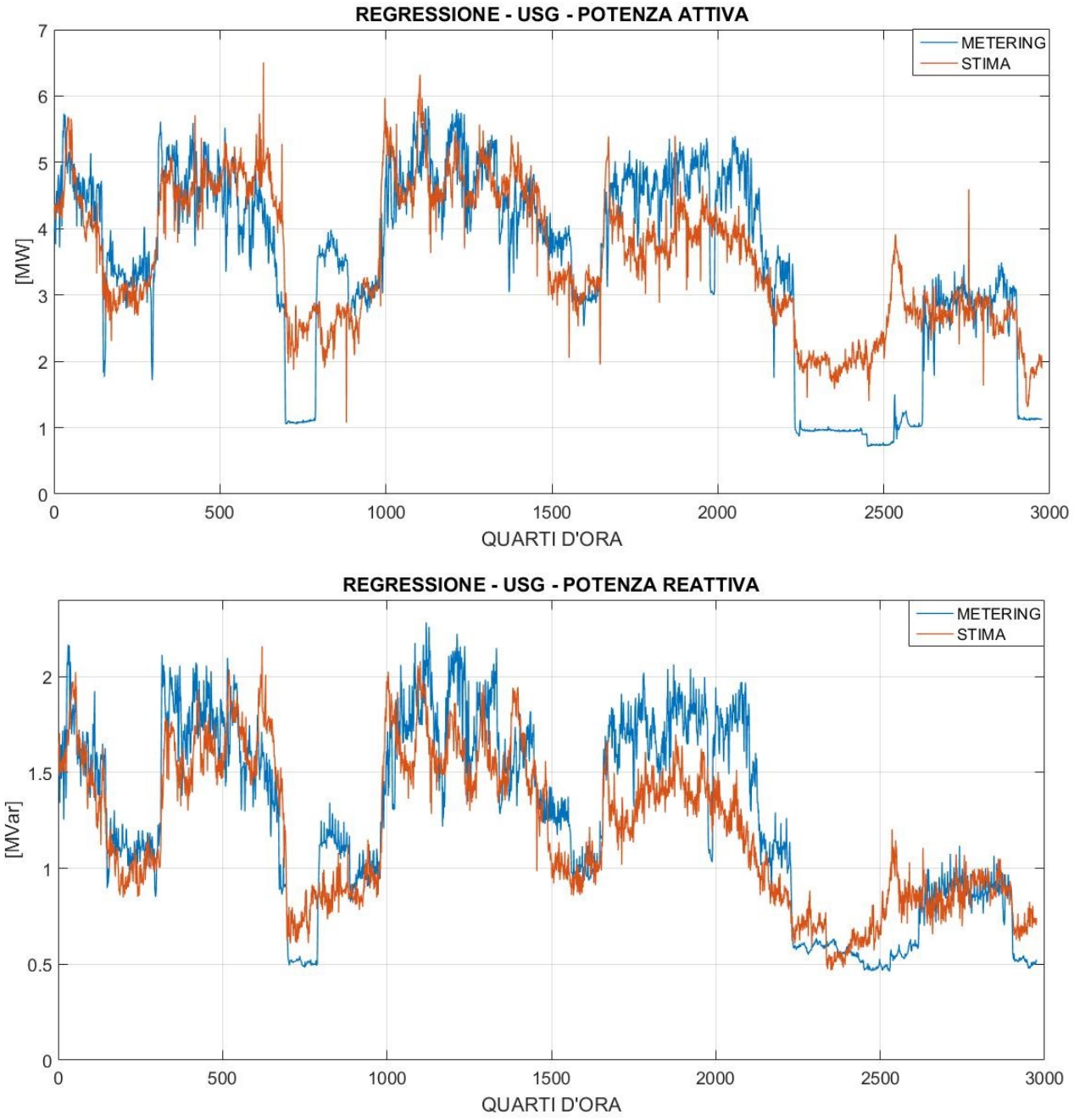


Figura 4.11. Grafico dei risultati della previsione di carico attivo e reattivo dell'utenza USG ottenuti tramite la regressione. I coefficienti sono ottenuti tramite l'analisi dei dati relativi ai primi 11 mesi del 2017 e i valori plottati sono relativi all'intero mese di dicembre 2017.

Capitolo 5

Rete neurale artificiale

Dopo aver analizzato una strategia di previsione che si basa sua un qualche tipo di relazione tra le misure disponibili in tempo reale e le grandezze da stimare, si tenta ora un nuovo approccio al problema. Gli assorbimenti dei carichi industriali sono fenomeni perlopiù stocastici nei quali però è anche possibile riconoscere una componente ricorrente che potrebbe essere sfruttabile. Per esempio, potrebbe essere possibile prevedere l'andamento del consumo di un utente conoscendone l'andamento che esso ha avuto nei giorni precedenti. Strumenti particolarmente adatti per svolgere questo tipo di funzioni sono le reti neurali artificiali.

5.1 Teoria

5.1.1 Introduzione

Le reti neurali artificiali (Artificial Neural Network, ANN) sono sistemi computazionali il quale funzionamento è ispirato dal sistema nervoso del cervello animale. La rete neurale artificiale è una struttura logica concepita per connettere diversi tipi di algoritmi. Una ANN è in grado di processare complesse strutture di dati in input grazie ad un insieme di unità connesse, o nodi, chiamati neuroni artificiali. Così come la struttura della rete, anche i neuroni artificiali sono inspirati dalla controparte biologica. Inoltre, anche le connessioni funzionano come le sinapsi di un cervello e possono trasmettere segnali da un neurone artificiale ad un altro. Un neurone artificiale è in grado di ricevere e processare un segnale ed in seguito inoltrare il segnale ad una unità successiva ad esso connessa.
La caratteristica più importante di una ANN è l'abilità di "imparare" a

svolgere un certo compito senza essere stato precedentemente programmato in alcun modo. Generalmente, il processo di apprendimento consiste semplicemente nel processare un gran numero di campioni esemplificativi. Nel riconoscimento delle immagini, per esempio, le ANN imparano a riconoscere un oggetto semplicemente elaborando una serie di immagini che sono state precedentemente categorizzate. Il sistema è automaticamente in grado di riconoscere delle caratteristiche identificative dal materiale fornito per l'apprendimento e non hanno bisogno di alcuna conoscenza preliminare dell'oggetto da categorizzare.
Durante l'impiego delle ANN, i segnali trasmessi da un neurone artificiale all'altro, attraverso le connessioni, è solitamente un numero reale. Un neurone artificiale somma tutti i segnali ricevuti dagli altri neuroni e compie una qualche funzione non lineare. Il risultato dell'operazione è l'output del neurone e verrà trasmesso ai neuroni successivi. I neuroni e le connessioni tipicamente hanno pesi e bias che si adattano durante il processo di apprendimento. Per esempio, i pesi nelle connessioni rafforzano o indeboliscono la forza dei segnali trasmessi mentre i pesi nei nodi funzionano come una sorta di soglia così che il segnale è trasmesso solo se la somma dei segnali ricevuti supera tale soglia. La struttura tipica di una ANN è organizzata in strati (layer). Il numero dei layer è variabile ed ognuno può compiere diversi tipi di trasformazioni sui loro input. Il primo strato è chiamato input layer e l'ultimo è invece chiamato output layer. Tutti gli altri strati vengono infine chiamati strati nascosti o hidden layer. I segnali possono viaggiare dallo strato di input fino allo strato di output in modo diretto oppure compiendo cicli tra i vari layer. Quando la struttura è concepita per essere aciclica, non ci sono connessioni con gli strati precedenti e vengono chiamate reti *feedfarward* (figura 5.1). ANN *feedfarward* sono tipicamente le Multi-Layer Perceptron (MLP) e le Radial Basis Function (RBF). Reti con una struttura ciclica sono invece definite ricorrenti. Tali reti possono essere interpretate come mostrato in figura 5.2, dove f risulta essere dipendente da sé stessa per le connessioni di feedback. Contrariamente dalle reti feedfarward, le proprietà dinamiche delle reti ricorrenti sono da tenere in considerazione.

5.1.2 Componenti delle reti neurali artificiali

Neuroni

I neuroni artificiali sono il componente principale delle reti neurali. Un arbitrario neurone j-esimo, che riceve un generico input $s_j(t)$, è composto dai

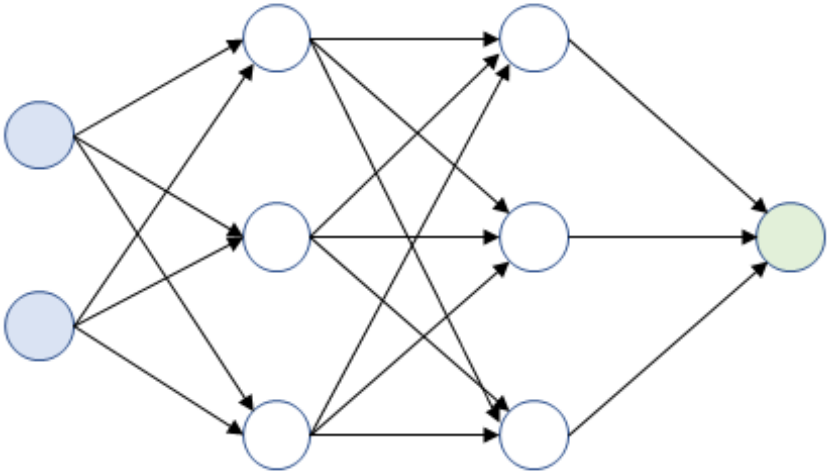


Figura 5.1. Rappresentazione di una rete neurale feedfarward a due input, due strati nascosti a 3 neuroni ciascuno ed un output.

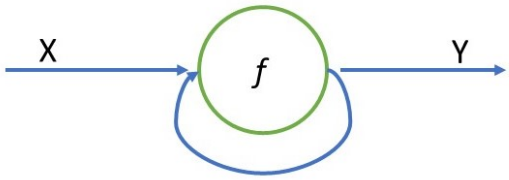


Figura 5.2. Rappresentazione logica di una rete ricorrente.

seguenti elementi:

- Un'attivazione $a_j(t)$. L'attivazione rappresenta lo stato del neurone e dipende da un parametro temporale discreto t.

- Un bias θ_j. I neuroni sono spesso accompagnati da questo parametro che può essere visto come un valore di soglia. I bias non dipendono dal parametro temporale e possono essere modificati solo dall'algoritmo di apprendimento.

- Una funzione di attivazione f. Questa funzione calcola l'attivazione $a_j(t+1)$ dall'attivazione precedente $a_j(t)$, dal bias θ_j e dall'input totale $s_j(t)$ tramite una relazione del tipo:

$$a_j(t+1) = f(a_j(t), \theta_j, s_j(t)). \tag{5.1}$$

  Molto spesso, l'output dei neuroni sono semplicemente le loro attivazioni.

Particolari tipi di neuroni artificiali sono i neuroni di input e di output. Un neurone di input non ha neuroni che lo precedono, mentre il neurone di output non ha neuroni che lo succedono, essi funzionano solo come interfaccia tra il sistema e l'ambiente in cui la rete lavora.

Connessioni e pesi

I neuroni sono collegati tra di loro tramite connessioni che trasmettono l'output di un neurone ai neuroni successivi. Ogni connessione è caratterizzata da un peso, per esempio w_{ij} è il peso assegnato alla connessione agente tra il neurone i e il neurone j. Un peso positivo può essere visto come rafforzamento, mentre un peso negativo può essere visto come un'inibizione del segnale trasmesso. L'input totale di un neurone è quindi dato dalla seguente relazione:

$$s_j(t) = \sum_i w_{ij}(t) a_i(t) + \theta_j(t) \tag{5.2}$$

Le unità con la regola di propagazione come in eq. 5.2 sono chiamate unità sigma. Una regola di propagazione differente è impiegata invece nelle unità sigma-pi:

$$s_j(t) = \sum_i w_{ij}(t) \prod_i a_{ij}(t) + \theta_j(t) \tag{5.3}$$

Nelle unità pi-sigma invece, gli input al neurone vengono moltiplicati invece che sommati.

Funzione di attivazione

La funzione di attivazione, o di propagazione, è una funzione che calcola l'effetto dell'input totale sull'attivazione dell'unità. A partire dall'input totale $s_k(t)$ e l'attivazione corrente $a_k(t)$ produce un nuovo valore di attivazione dell'unità k-esima:

$$a_k(t+1) = \mathcal{F}_k(a_k(t), s_k(t)) \tag{5.4}$$

Spesso, la funzione di attivazione non considera anche l'attivazione attuale, ma è semplicemente una funzione crescente dell'input totale dell'unità:

$$a_k(t+1) = \mathcal{F}_k(s_k(t)) = \mathcal{F}_k(\sum_j w_{jk}(t)a_j(t) + \theta_k(t)) \tag{5.5}$$

In realtà, le funzioni d'attivazione non sono necessariamente funzioni crescenti.
Generalmente, la funzione di attivazione è una funzione che funge da soglia, come la funzione segno, o altre funzioni lineari o semi-lineari. Un esempio può essere la sigmoide (figura 5.3):

$$a_k = \mathcal{F}_k(s_k) = \frac{1}{1 + e^{-s_k}} \tag{5.6}$$

L'output di un'unità potrebbe anche essere una funzione stocastica dell'input totale. In questo caso l'attivazione non è stabilita in modo deterministico, ma l'input totale determina solo la probabilità p che un neurone assuma un valore positivo di attivazione:

$$p(s_k) = \frac{1}{1 + e^{\frac{s_k}{T}}}, \tag{5.7}$$

dove T è un parametro che determina l'inclinazione della funzione di proba-

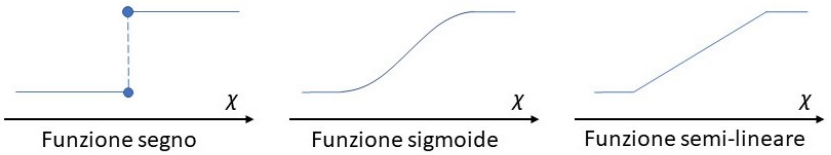


Figura 5.3. Rappresentazione delle tre funzioni di attivazione comunemente più usate.

bilità.

Regola di apprendimento

Una regola di apprendimento è un algoritmo che modifica i pesi e bias della rete neurale, in modo da ottenere, per un dato input alla rete, un output il

più simile possibile a ciò che si desidera. Per uno specifico compito da svolgere e una data rete, il processo di apprendimento consiste quindi nel trovare un set di pesi e bias che risolvono il problema in modo il più possibile ottimale. La capacità di apprendimento è la caratteristica delle ANN che più ha attirato l'attenzione per le sue enormi potenzialità e possibilità applicative.
I procedimenti di apprendimento si basano su un problema di ottimizzazione matematica. Questo significa che deve essere definita una funzione costo $C : \mathcal{F} \to R$ e che deve essere cercata una funzione f^* tale che $C(f^*) \leq C(f)$ $\forall f \in \mathcal{F}$. in altre parole, l'algoritmo di apprendimento ha il compito di trovare una funzione f^* tra quelle possibili nello spazio delle soluzioni $\mathcal{F}$ tale che minimizzi la funzione costo C. Il concetto di funzione costo è fondamentale, ma può essere considerato semplicemente come uno strumento che misuri quanto una soluzione si discosta dalla soluzione ottimale. In un gran numero di applicazioni il costo deve essere una funzione delle osservazioni.
L'algoritmo di ottimizzazione consiste in due passaggi principali: propagazione e aggiornamento dei pesi/bias. Nel primo passaggio, gli input sono presentati alla rete, i segnali si propagano dal layer di input a quello di output, attraversando tutti gli strati intermedi. Il risultato della rete è quindi paragonato all'output desiderato e un termine di errore è calcolato per ogni neurone di output. I termini di errore sono poi propagati dal layer di output indietro ad ogni neurone del sistema, così che un termine di errore è associato ad ogni nodo della rete. Questi valori rispecchiano il contributo di ogni neurone sull'output originale e sono utilizzati dai processi di retro-propagazione per calcolare il gradiente della funzione costo. Nel secondo step, l'algoritmo di ottimizzazione prova a minimizzare la funzione costo utilizzando il gradiente precedentemente menzionato per aggiornare i pesi e i bias. Se nel processo di retro-propagazione, l'aggiornamento dei pesi è attuato per mezzo del metodo della discesa stocastica del gradiente (stochastic gradient descent), allora viene utilizzata la seguente formula:

$$w_{ij}(t+1) = w_{ij} + \eta \frac{\partial C}{\partial w_{ij}} + \epsilon(t) \tag{5.8}$$

Dove, η è il tasso di apprendimento, C è la funzione costo e $\epsilon(t)$ è un termine stocastico.
La funzione costo deve essere opportunamente scelta considerando fattori come il tipo di paradigma di apprendimento (supervisionato, non supervisionato, per rinforzo) e la funzione di attivazione.

5.1.3 Paradigmi di apprendimento

Per ogni compito specifico, sono stati sviluppati tre corrispondenti algoritmi di apprendimento: apprendimento supervisionato, apprendimento non supervisionato e apprendimento per rinforzo.

Apprendimento supervisionato

L'apprendimento supervisionato lavora su un dataset di allenamento costituito da osservazioni. Le osservazioni sono coppie di valori (x, y) con $x \in X$ e $y \in Y$, dove X rappresenta lo spazio degli input e Y lo spazio degli output. L'obiettivo del procedimento di apprendimento consiste nel trovare una funzione $f : X \to Y$ con $f \in \mathcal{F}$ tale che faccia corrispondere nel miglior modo possibile gli output ottenuti dalla rete e le osservazioni. Una conoscenza a priori del problema è necessaria per la definizione della funzione costo. Questa è infatti un qualche tipo di differenza tra il risultato ottenuto dalla rete e il corrispondente output del dataset di allenamento. Poiché il paradigma si basa su una sorta di continui feedback, il processo può essere interpretato come un apprendimento guidato da un "insegnante" o da un "supervisore" il quale ruolo è assunto dalla funzione costo.
La funzione costo più comune è l'errore quadratico medio, che quantifica la discrepanza tra l'output della rete $f(x)$ e il target y nel dataset delle osservazioni. Una tipica rete che utilizza questo tipo di paradigma di apprendimento è il multilayer perceptron (MLP). Per questi tipi di sistemi il costo è minimizzato attraverso il metodo della discesa del gradiente andando così a definire l'algoritmo di training in retropropagazione.
Alcuni dei compiti affrontati con un algoritmo di apprendimento supervisionato sono l'approssimazione di funzioni (regressione) e la classificazione (pattern recognition). Questo tipo di apprendimento si presta anche a risolvere problemi caratterizzati da dati di tipo sequenziale, come riconoscimento gestuale o della calligrafia.

Apprendimento non supervisionato

Nell'apprendimento non supervisionato sono fornite alla rete solo dei dati di input x e la funzione costo da minimizzare. In questi tipi di sistemi è la rete stessa che deve individuare certe proprietà statistiche caratterizzanti nel dataset in ingresso, non è quindi richiesta una conoscenza approfondita di quelle citate proprietà. La funzione costo può essere una qualsiasi funzione degli input x e dell'output della rete, ma dipende dal compito e da eventuali

assunzioni iniziali.
L'apprendimento supervisionato è un paradigma di apprendimento utilizzato solitamente per compiti come il clustering, la compressione e il filtraggio di dati e alcune stime statistiche.

Apprendimento per rinforzo

Nell'apprendimento per rinforzo nessun dataset è direttamente fornito alla rete neurale, ma è generato dall'interazione di un agente con l'ambiente. Un agente è un sistema in grado di interagire con l'ambiente nel quale è immerso tramite sensori e attuatori e di prendere decisioni autonomamente. Per ogni step temporale l'agente intraprende un'azione con l'ambiente e effettua qualche osservazione in modo da generare un input ed un costo istantaneo. L'agente valuta l'input, modifica il suo stato e genera l'azione successiva da intraprendere con lo scopo di minimizzare un costo a lungo termine (ad esempio un costo cumulativo o un costo cumulativo previsto). Lo scopo dell'intera procedura è quello di trovare uno schema intrinseco che il sistema deve seguire per selezionare le azioni da intraprendere.
L'apprendimento per rinforzo deve essere scelto per gestire problemi di decision-making sequenziali come videogiochi o problemi di controllo.

5.1.4 Algoritmi di apprendimento

Come già detto, allenare una rete neurale artificiale significa selezionare una funzione da un set di funzioni disponibili che minimizzi la funzione costo. Negli algoritmi Bayesiani ciò consiste invece nel selezionare una distribuzione tra un set di modelli disponibili, mentre per molti altri algoritmi questo può essere semplicemente visto come una diretta applicazione della teoria di ottimizzazione per retro-propagazione. Infatti, molti usano una qualche variante della teoria della discesa del gradiente combinata con la retro-propagazione. I vari gradienti dei parametri della rete sono calcolati prendendo la derivata della funzione costo rispetto ai relativi parametri. I parametri sono poi aggiornati nella direzione stabilita dal gradiente. Gli algoritmi di retro-propagazione possono essere divisi in:

- Steepest descent;
- Quasi-Newtoniani;
- Levenberg-Marquardt e gradiente coniugato.

5.2 Implementazione della rete neurale artificiale

Gli assorbimenti dei carichi elettrici industriali sono fenomeni fortemente stocastici, ma hanno anche una componente autoregressiva, dovuta ai processi lavorativi ciclici, che può essere sfruttata nei problemi di load-forecasting. Per questo, le reti neurali artificiali sono idonee per questi tipi di compiti. Svariati tipi di ANN, impostazioni, input e algoritmi di training sono stati testati. I risultati migliori sono stati ottenuti per mezzo di una variante inspirata dal problema di Short Term Load-Forecasting proposto nell'articolo [9]. Il lavoro presenta un'applicazione di una rete neurale Radial Basis Function (RBF) per la previsione di carico, delle 24 ore a venire, di un piccolo sistema elettrico. Il modello proposto sfrutta una struttura a più input e ad un solo output, che prevede i carichi orari in modo separato, utilizzando come regressori i carichi storici più correlati. Nell'articolo citato, i carichi sfruttati sono quelli 24, 25 e 48 ore precedenti. Inoltre, sono forniti alla rete anche due indici che discriminano il giorno della settimana e l'ora del giorno del carico da prevedere. Nell'articolo è anche verificato che non c'è una correlazione sfruttabile tra le condizioni metereologiche e il carico elettrico, per cui, pure in questo elaborato, tale relazione non è stata considerata.
In questo lavoro, è stata sfruttata l'app di Matlab "Neural Net Fitting", che adopera una rete feed-forward a due strati, con 10 neuroni nascosti. Alla rete sono state fornite come input i carichi 671, 672, 673, 97 e 96 quarti d'ora precedenti al carico da stimare. Inoltre, sono stati forniti un indice relativo al quarto d'ora (da 1 a 96), un indice relativo al giorno (da 1 a 7) e un indice indicante se il carico da prevedere cade in un giorno festivo nazionale. L'algoritmo di training scelto è il "Bayesian Regularization". Questo algoritmo necessita tipicamente di più tempo, ma ha buone caratteristiche generalizzanti in caso di dataset ridotti o rumorosi. Il dataset fornito in input sia in fase di allenamento che in fase di utilizzo sono da considerate largamente affetti da valori anomali. Numerosi valori sono mancanti o non sono consistenti (ad esempio picchi di potenza non spiegabili). Questo metodo produce risultati molto buoni (i risultati quantitativi sono presentati nel *Capitolo 7*), più precisi che i risultati ottenuti dalla regressione completa di tutti e 47 i regressori.
Nelle figure 5.4, 5.5, 5.6 sono riportati alcuni risultati della stima di carico del mese di dicembre 2017 ottenuti con la rete neurale, allenata sui primi 11 mesi dello stesso anno. Bisogna sottolineare che dicembre è un mese critico

per la stima, poiché sono presenti numerosi giorni festivi infrasettimanali. Il carico durante questi giorni non è previsto con precisione poiché esso non segue il proprio profilo ricorsivo il quale la ANN tende a ricopiare.

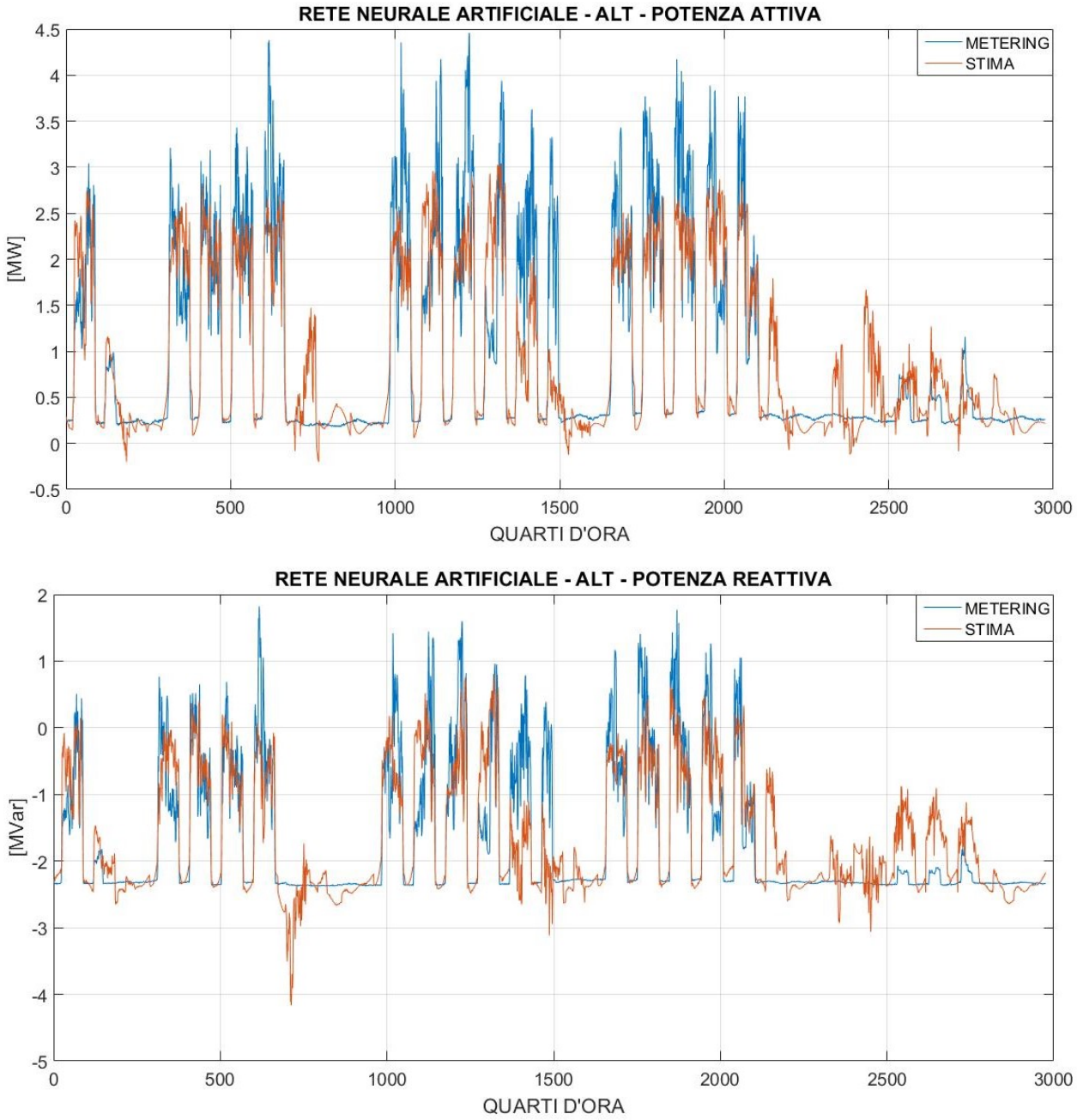


Figura 5.4. Grafico dei risultati della previsione di carico attivo e reattivo dell'utenza ALT ottenuti tramite la ANN. La rete è stata allenata con i dati relativi ai primi 11 mesi del 2017 e i valori plottati sono relativi all'intero mese di dicembre 2017.

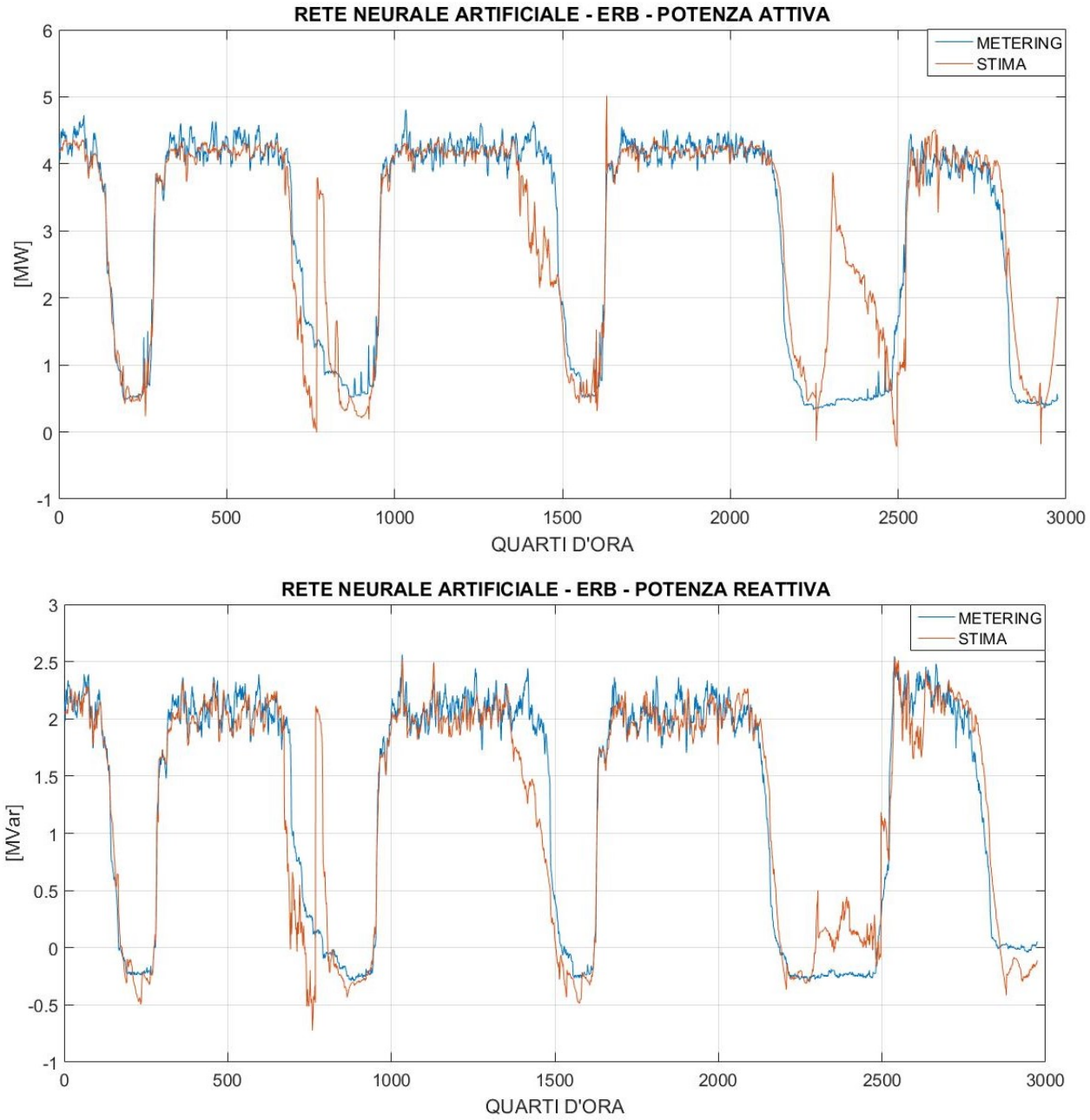


Figura 5.5. Grafico dei risultati della previsione di carico attivo e reattivo dell'utenza ERB ottenuti tramite la ANN. La rete è stata allenata con i dati relativi ai primi 11 mesi del 2017 e i valori plottati sono relativi all'intero mese di dicembre 2017.

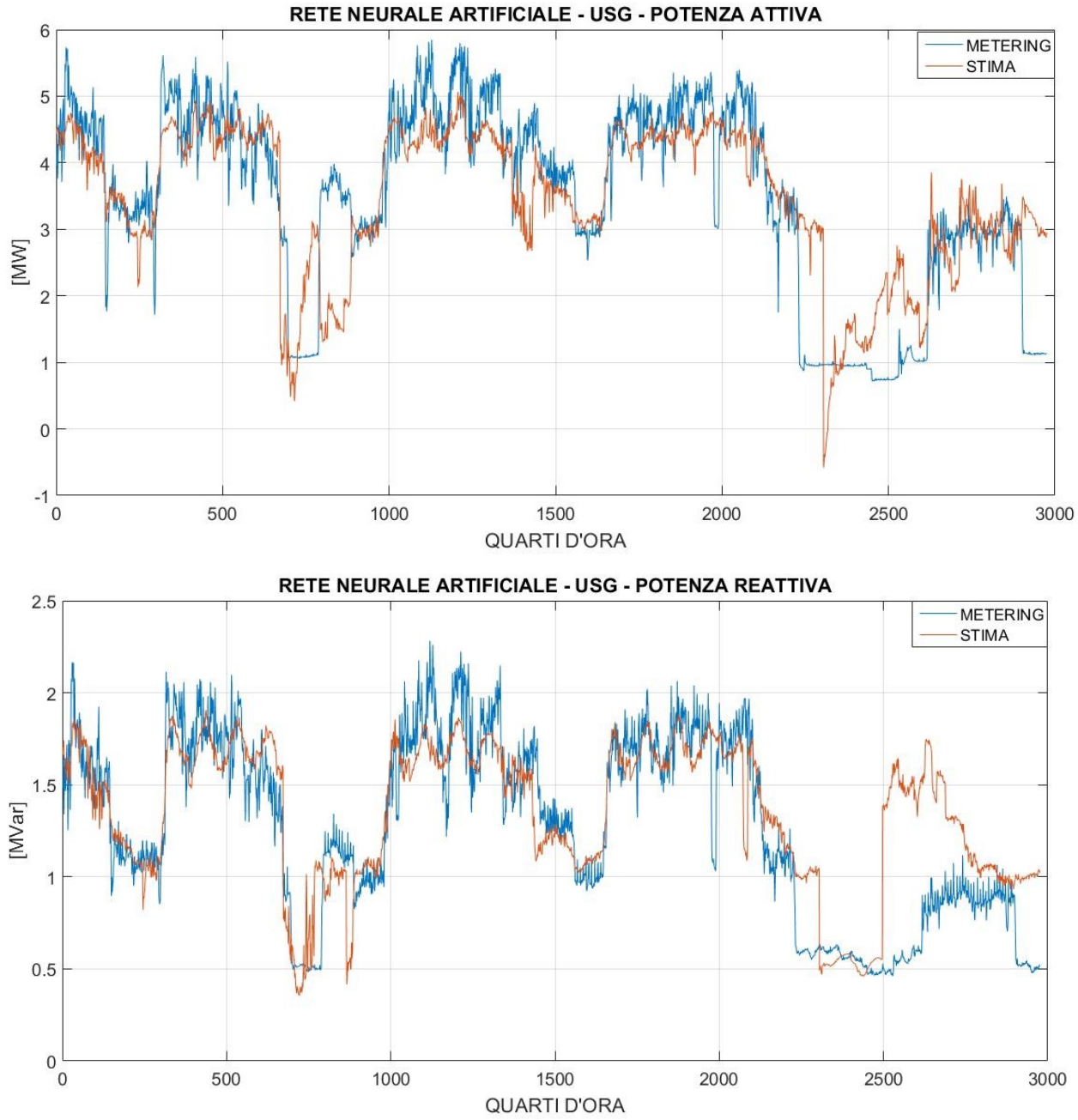


Figura 5.6. Grafico dei risultati della previsione di carico attivo e reattivo dell'utenza USG ottenuti tramite la ANN. La rete è stata allenata con i dati relativi ai primi 11 mesi del 2017 e i valori plottati sono relativi all'intero mese di dicembre 2017.

Capitolo 6

Altri metodi

6.1 Combinazione lineare tra regressione e ANN

Sia la regressione con 47 variabili sia la rete neurale forniscono risultati piuttosto buoni, anche se talvolta producono risultati anomali, come picchi o buchi. Per esempio, la rete neurale, molto accurata in caso di un profilo di assorbimento ricorsivo, ha difficoltà nello stimare il carico durante i giorni festivi infrasettimanali o in caso di assorbimenti anomali, ad esempio per motivi interni all'azienda (guasti, scioperi, etc.). Nelle figure 6.1 e 6.2 sono mostrati le previsioni di potenza attiva per il carico ERB durante una settimana di dicembre 2017. La ANN è solitamente accurata, ma la stima è molto imprecisa al 25 di dicembre proprio perché rappresenta una festività durante un giorno infrasettimanale (lunedì). La regressione, viceversa, è solitamente meno precisa, ma l'errore durante il 25 dicembre è più limitato. Per questo motivo è stato scelto di combinare i due metodi. Con una semplice media aritmetica tra il risultato ottenuto con la ANN e con la regressione, il carico stimato ha ancora una buona precisione, ma è più immune dai valori anomali. Per esempio, in caso di un picco imprevisto da parte di uno o dall'altro metodo, l'errore della previsione è all'incirca dimezzato. Come mostrato in figura 6.3 l'errore al 25 di dicembre è quasi dimezzato, mentre globalmente la precisione non è fortemente corrotta. Questo può essere considerato un metodo semplice, ma molto valido, poiché offre una buona precisione generale e una discreta immunità dagli errori anomali.

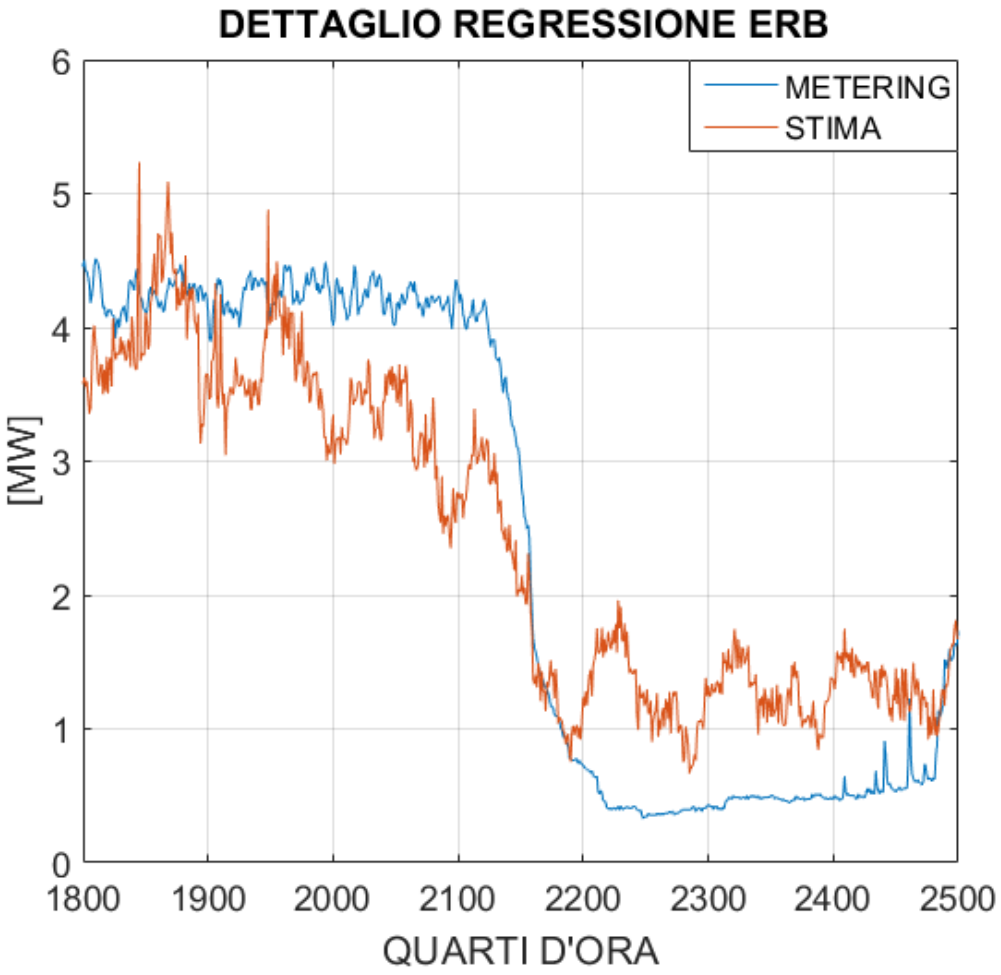


Figura 6.1. Dettaglio della stima per il carico ERB ottenuta tramite la regressione. E' rappresentata circa una settimana. Si può notare come l'errore massimo assoluto sia limitato anche al 25 dicembre (tra i 2304 e 2400 quarti d'ora), ma generalmente la previsione non è molto precisa.

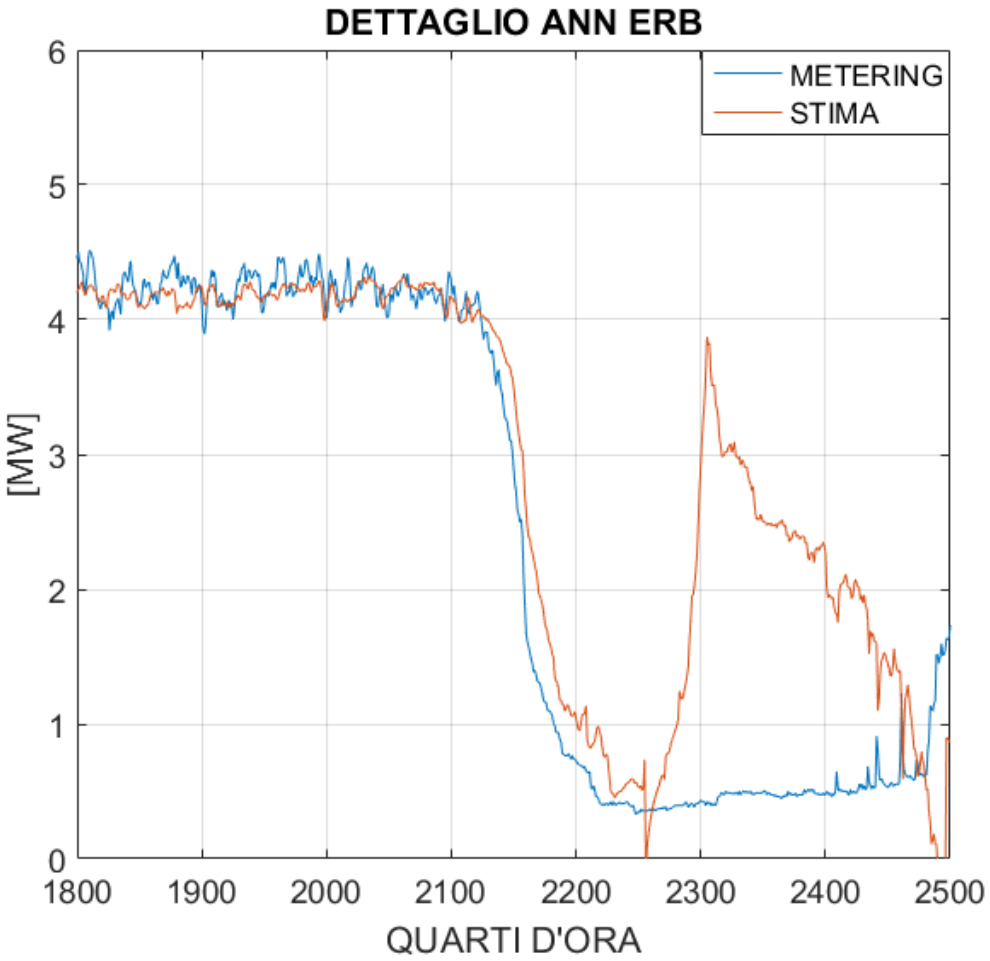


Figura 6.2. Dettaglio della stima per il carico ERB ottenuta tramite rete neurale. E' rappresentata circa una settimana. Si può notare come la previsione sia solitamente molto precisa tranne in alcuni tratti dove l'errore assoluto può essere elevato, come al 25 dicembre (tra i 2304 e i 2400 quarti d'ora).

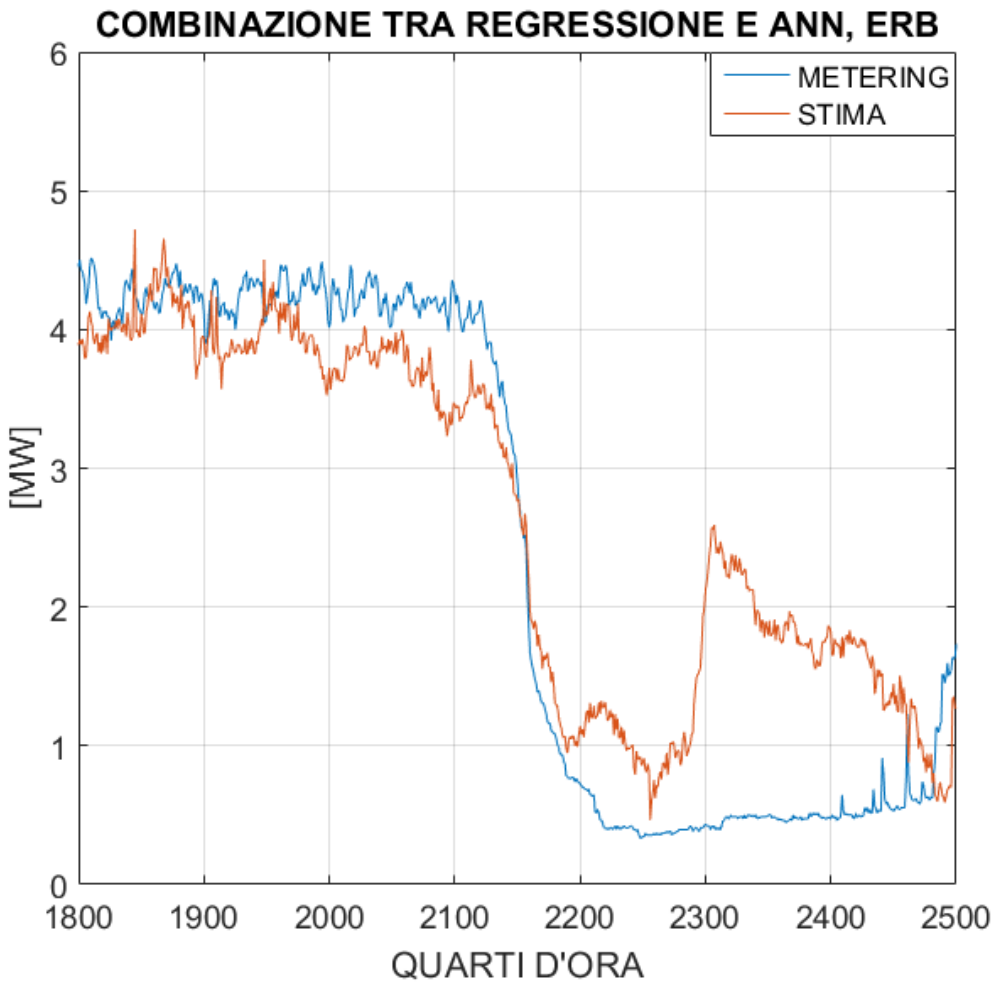


Figura 6.3. Dettaglio della stima per il carico ERB ottenuta tramite una combinazione lineare tra la regressione e la rete neurale. E' rappresentata circa una settimana. Si può notare come si sia ottenuto un compromesso tra precisione generale e errore assoluto massimo.

6.2 Clusterizzazione

In questa sezione verrà introdotto un metodo che si serve di una clusterizzazione preliminare dei dati. Siccome per alcuni carichi i profili di assorbimento sono molto differenti tra giorni festivi e lavorativi, si è pensato che trattare le due stime in modo differente possa fornire risultati interessanti.

6.2.1 Introduzione teorica al problema di clusterizzazione

Il clustering è un processo di raggruppamento di un insieme di oggetti fisici o astratti in classi di oggetti simili in modo che i componenti di una stessa classe siano il più simile possibili tra di loro e il più dissimili possibili con gli oggetti di classi differenti. Le classi vengono chiamate cluster. La formalizzazione del concetto di similarità è l'aspetto più importante del problema di clusterizzazione. Essa dipende dalla natura del problema, spesso viene espressa in termini di distanza euclidea in uno spazio multidimensionale, ma talvolta non è un criterio adeguato o semplicemente non è tecnicamente applicabile, basti pensare al caso in cui gli attributi da classificare non sono di natura numerica. La bontà dell'intero problema di clusterizzazione dipende dall'adeguatezza della misura scelta.
Le tecniche di clustering sono innumerevoli. In seguito vengono presentate, brevemente, le loro principali classificazioni che differiscono per:

- Approccio:
  - Metodi aggregativi o bottom-up: ogni elemento da classificare è inizialmente considerato un cluster a se stante. In seguito, questi vengono raggruppati, secondo certi criteri di somiglianza, fino ad ottenere un numero prefissato di cluster o finchè la distanza tra i diversi clusters supera un certo valore, anch'esso prefissato.
  - Metodi divisivi o top down: inizialmente tutti gli elementi appartengono ad un unico cluster che viene successivamente suddiviso secondo criteri di similarità. Il processo si interrompe al raggiungimento di un certo requisito (solitamente il numero di cluster desiderato).
- Possibilità, per un certo elemento, di essere assegnato a più cluster differenti:

  - Clustering esclusivo o hard clustering: un elemento può essere assegnato ad un solo gruppo. Non esistono quindi elementi in comune tra cluster differenti;
  - Clustering non esclusivo o soft clustering o fuzzy clustering: ogni elemento può appartenere a più di un cluster secondo un certo grado di appartenenza.

- Algoritmo di divisione dello spazio:

  - Clustering partizionale o non gerarchico o k-clustering: i cluster sono rappresentati tramite un vettore centrale, chiamato centroide, che non deve necessariamente appartenere al dataset. Quando il numero di cluster è fissato ad un valore k il problema diventa un problema di ottimizzazione: trovare i k centroidi e assegnare gli oggetti ai relativi cluster in modo da minimizzare la distanza quadratica tra gli oggetti del cluster e i relativi centroidi;
  - Clustering gerarchico: questi algoritmi connettono gli oggetti per formare cluster in base alla loro distanza. Un cluster può essere definito dalla distanza massima necessaria per connettere i componenti al suo interno. A seconda della distanza si formeranno cluster diversi. Questo algoritmo non fornisce una sola classificazione degli elementi, bensì un dendrogramma in grado di rappresentare a che distanza i cluster si dividono/uniscono (figura 6.4).

Algoritmo k-means

Nell'elaborato verrà usato l'algoritmo "k-means", un algoritmo di clustering partizionale che permette di suddividere un insieme di oggetti in k gruppi sulla base delle loro caratteristiche. La suddivisione avviene tramite la minimizzazione della varianza totale intra-cluster. L'algoritmo segue una procedura iterativa. Inizialmente crea k partizioni assegnando ad ognuno di essi elementi in modo casuale o usando alcune informazioni euristiche. Quindi calcola il centroide di ogni gruppo. Costituisce quindi una nuova partizione associando ogni punto d'ingresso al cluster il cui centroide è più vicino ad esso. Quindi vengono ricalcolati i centroidi per i nuovi cluster e così via, finché l'algoritmo non converge.
Dati N oggetti con i attributi, modellizzati come vettori in uno spazio i-dimensionale, è definito l'insieme degli oggetti $X = X_1, X_2, \ldots, X_N$. La

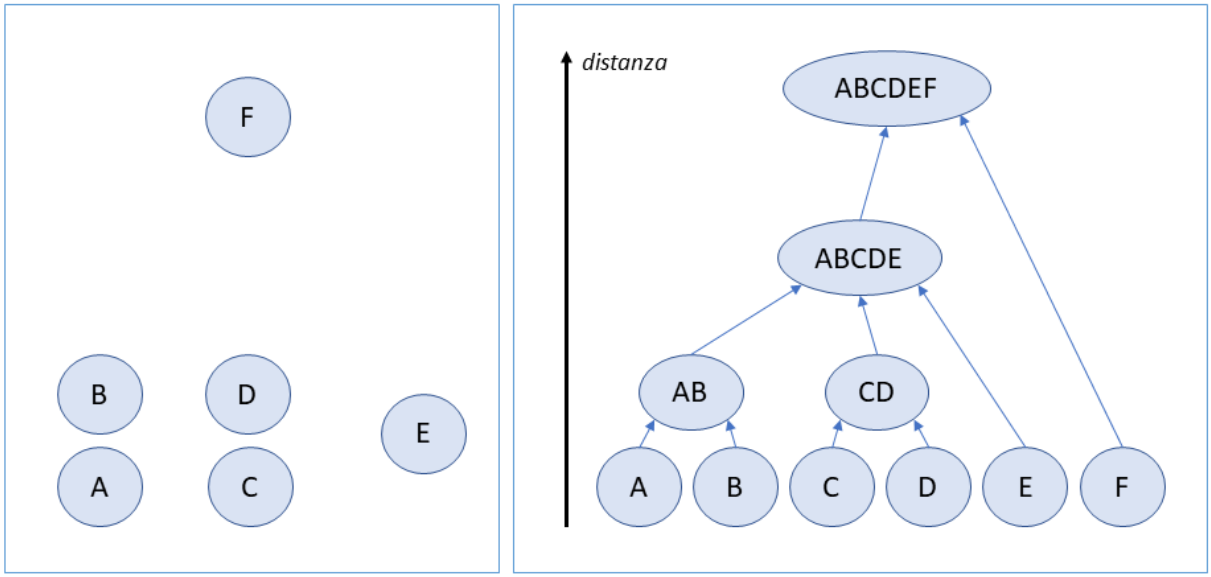


Figura 6.4. Rappresentazione di un dendrogramma e del suo significato, ovvero di indicare a che distanza si uniscono i cluster. Per distanza si intende la distanza massima di un elemento dal centroide del cluster al quale appartiene.

partizione degli oggetti si definisce $P = P_1, P_2, \ldots, P_K$ tale che soddisfano le seguenti proprietà:

- $\cup_1^K P_i = X$, l'unione di tutti i cluster deve contenere tutti gli elementi iniziali;
- $P_i \cap P_j = \emptyset,\ i \neq j$, ogni oggetto può appartenere ad un solo cluster;
- $\emptyset \subset P_i \subset X$, ogni cluster deve contenere almeno un oggetto e nessun cluster può contenere tutti gli oggetti;
- $1 < K < N$, il numero di cluster da individuare deve logicamente essere in numero maggiore di uno e inferiore al numero di oggetti.

Una partizione si formalizza mediante una matrice $U \in \mathbb{N}^{K \times N}$, il cui generico elemento $u_{ij} = \{0,1\}$ indica l'appartenenza, o meno, dell'oggetto j al cluster i. L'insieme dei K centroidi è invece indicata con $C = C_1, C_2, \ldots, C_k$. La funzione obiettivo è invece definita come:

$$S(U, C) = \sum_{i=1}^{K} \sum_{X_j \in P_i} ||X_j - C_i||^2 \tag{6.1}$$

Il minimo della funzione obiettivo viene calcolata con la seguente procedura iterativa:

1. U_v e C_v vengono generati in modo casuale;
2. Si calcola U_n tale che minimizzi $S(U, C_v)$;
3. Si calcola C_n tale che minimizzi $S(U_v, C)$;
4. si verifica la convergenza, se questa non è raggiunta vengono imposti $U_v = U_n, C_v = C_n$ e torna al passo 2.

Dove la convergenza viene verificata tipicamente con i seguenti criteri:

- U_n calcolata è identica alla matrice trovata nell'iterazione precedente;
- la differenza tra la funzione obiettivo calcolata all'ultima e alla penultima iterazione sia inferiore ad un valore prefissato.

L'algoritmo presenta due svantaggi: la casualità di U e C alla prima iterazione fanno sì che la soluzione trovata non rappresenti necessariamente l'ottimo globale ed è richiesto di stabilire a priori il numero di K cluster.

Algoritmo K-means in Matlab

Le clasterizzazioni in questo studio sono state svolte con il software Matlab tramite la funzione *kmeans*, con, in input, il dataset da partizionare e il numero di cluster desiderati. L'algoritmo di calcolo è il medesimo descritto nella sotto-sezione precedente, ma le assegnazione iniziali dei centroidi avvengono per default tramite l'algoritmo *k-means++* in modo euristico.
L'algoritmo sceglie i centroidi iniziali, assumendo il numero di cluster sia pari a k assegnato, nel seguente modo:

1. Seleziona un osservazione in modo casuale dal dataset X. L'osservazione scelta è il primo centroide c_1;
2. Calcola le distanze tra ogni osservazione e c_1. La distanza tra c_j e l'osservazione m è denotata con $d(x_m, c_j)$;
3. Seleziona il centroide successivo, c_2, in modo casuale da X secondo la probabilità
$$\frac{d^2(x_m, c_1)}{\sum_{j=1}^{n} d^2(x_j, c_1)} \tag{6.2}$$
.

4. Per scegliere il centroide j-esimo:

   (a) Calcola le distanze tra ogni osservazione e ogni centroide e assegna ogni osservazione al centroide più vicino;

   (b) Per $m = 1, \ldots, n$ e $p = 1, \ldots, j-1$, seleziona il centroide j in modo casuale da X con probabilità

   $$\frac{d^2(x_m, c_p)}{\sum_{h;x_h \in C_p} d^2(x_h, c_p)} \tag{6.3}$$

   dove C_p è il set di tutte le osservazioni più vicine al centroide c_p e x_m appartiene a C_p.

5. Ripete il passo 4 finché tutti i k centroidi sono scelti.

6.2.2 Implementazione

La clusterizzazione in questo studio è stata utilizzata per poter affrontare la previsione dei carichi in modo differente a seconda del tipo di giorno in esame. L'algoritmo k-means è in grado di categorizzare i giorni in base ai profili di assorbimento giornalieri. I dataset così divisi possono essere studiati, indipendentemente, nel modo che sembra più opportuno per sviluppare le tecniche di stima ritenute più efficaci. Nel momento dell'utilizzo è però necessario riconoscere preventivamente in che tipo di giorno si trova il carico da stimare per selezionare il corrispondente metodo.
In questo studio sono stati testate numerose soluzioni. Ognuna utilizza un modello a due cluster in grado di dividere il dataset in giorni lavorativi e festivi. Le due tecniche di stima considerate sono state: regressione (a 47 regressori come descritto nel *Capitolo 4*) sia per giorni feriali che festivi, rete neurale per i giorni lavorativi e regressione (a 47 regressori come descritto nel *Capitolo 4*) per i giorni festivi e rete neurale sia per giorni lavorativi che festivi. Le reti neurali testate sono state numerose. Per quanto riguarda i giorni festivi, la previsione generalmente è sempre precisa ed affidabile grazie alla forte componente ricorsiva degli assorbimenti. Nei giorni festivi, invece, la componente stocastica è molto importante, rendendo la loro stima un compito particolarmente impegnativo.
I risultati migliori sono stati ottenuti usando reti neurali sia per giorni festivi che lavorativi impostate come segue:

- Giorni lavorativi: 140 input. 120 input corrispondono ad un carico ogni 4 (quindi ogni ora) partendo dal carico esattamente un giorno prima

al carico da stimare e andando a ritroso per 5 giorni. Ovvero, se $L(t)$ è il carico da stimare, con t indicante il quarto d'ora, gli input sono $L(t-96)$, $L(t-100)$, ..., $L(t-572)$. 15 input corrispondono ai carichi relativi ai 3 quarti d'ora più vicini allo stesso quarto d'ora del carico da stimare, ma nei 5 giorni precedenti, ovvero $(L-95)$, $L(t-96)$, $L(t-97)$, $L(t-191)$, ..., $L(t-481)$. Infine, 5 input corrispondono ai carichi corrispondenti allo stesso quarto d'ora del carico da stimare, ma dei 5 giorni precedenti, ovvero: $L(t-96)$, $L(t-192)$, ..., $L(t-480)$. Il dataset dal quale vengono presi gli input è composto rigorosamente solo da giorni lavorativi.

- Giorni festivi: 85 input. Un input indica se il giorno relativo al carico da stimare è una festività nazionale come possono essere il giorno di Natale oppure Capodanno piuttosto che un semplice giorno festivo come un sabato. 72 input corrispondono ad un carico ogni 4 (quindi ogni ora) partendo dal carico esattamente un giorno prima al carico da stimare e andando a ritroso per 3 giorni. Ovvero, se $L(t)$ è il carico da stimare, con t indicante il quarto d'ora, gli input sono $L(t-96)$, $L(t-100)$, ..., $L(t-380)$. 15 input corrispondono ai carichi relativi ai 3 quarti d'ora più vicini allo stesso quarto d'ora del carico da stimare, ma nei 3 giorni precedenti, ovvero $(L-95)$, $L(t-96)$, $L(t-97)$, $L(t-191)$, ..., $L(t-289)$. Infine, 5 input corrispondono ai carichi corrispondenti allo stesso quarto d'ora del carico da stimare, ma dei 3 giorni precedenti, ovvero: $L(t-96), L(t-192), \ldots, L(t-288)$. Il dataset dal quale vengono presi gli input è composto rigorosamente solo da giorni festivi.

Si è scelto di applicare il metodo non per tutti gli utenti dell'isola di carico in esame, ma solo per i carichi con una distinzione più marcata nel profilo di assorbimento tra giorni lavorativi e festivi, ovvero: ALT, ERB, FEU, ITA, USG, RVA. Nelle figure 6.5, 6.6, 6.7 sono riportati alcuni risultati in forma grafica.

Come si può notare, gli errori più importanti si verificano ancora nei giorni festivi. Come già accennato, la componente stocastica è molto importante in quei giorni. Inoltre, il cluster composto dai giorni festivi è ancora molto eterogeneo. Ad esempio, in figura 6.8 viene riportato il plot di tutti i giorni festivi e si può notare come alcuni abbiano un profilo piatto, a valori bassi, ed altri invece arrivano ad avere picchi elevati. Questo succede perché la clusterizzazione riconosce come giorni festivi sia giorni in cui gli assorbimenti sono stati paragonabili a quelli di un giorno lavorativo, ma solo per metà giornata, sia giorni in cui i processi lavorativi sono stati completamente

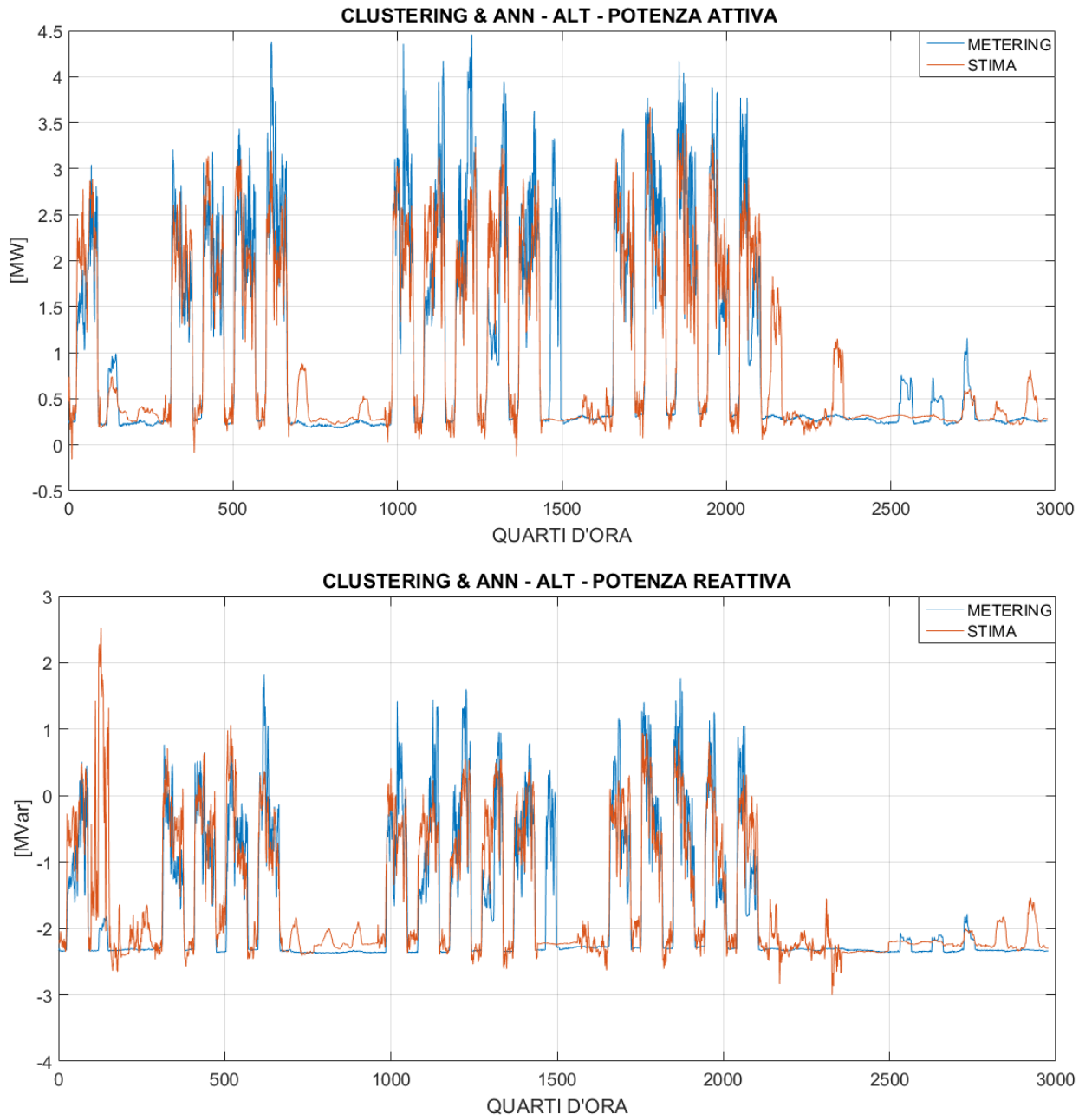


Figura 6.5. Grafico dei risultati della previsione di carico attivo e reattivo dell'utenza ALT ottenuti tramite due ANN scelte a seconda del tipo di giorno in cui si trova il carico da stimare (feriale o festivo). Le reti sono state allenate con i dati relativi ai primi 11 mesi del 2017, divisi in due cluster, e i valori plottati sono relativi all'intero mese di dicembre 2017.

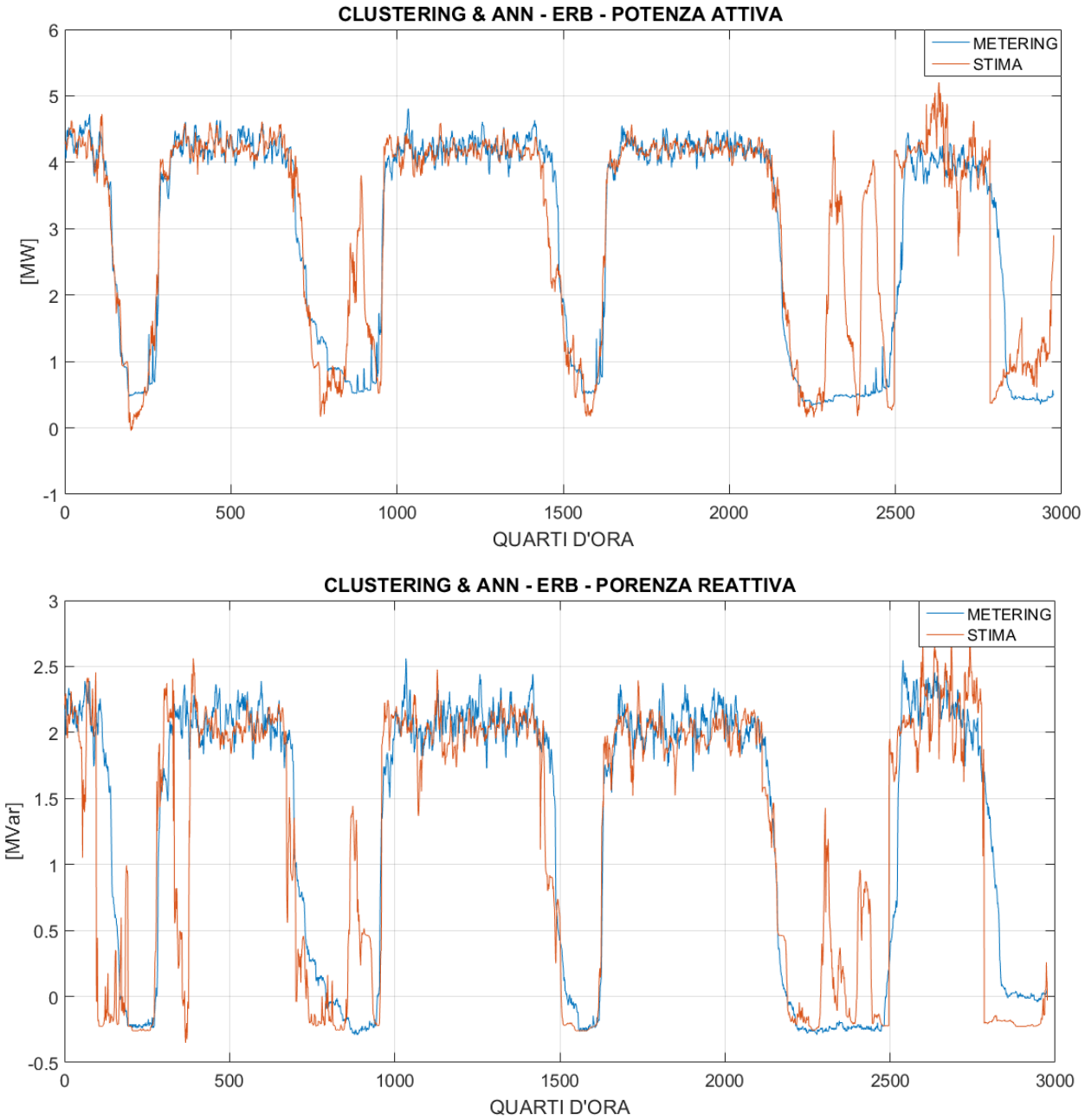


Figura 6.6. Grafico dei risultati della previsione di carico attivo e reattivo dell'utenza ERB ottenuti tramite due ANN scelte a seconda del tipo di giorno in cui si trova il carico da stimare (feriale o festivo). Le reti sono state allenate con i dati relativi ai primi 11 mesi del 2017, divisi in due cluster, e i valori plottati sono relativi all'intero mese di dicembre 2017.

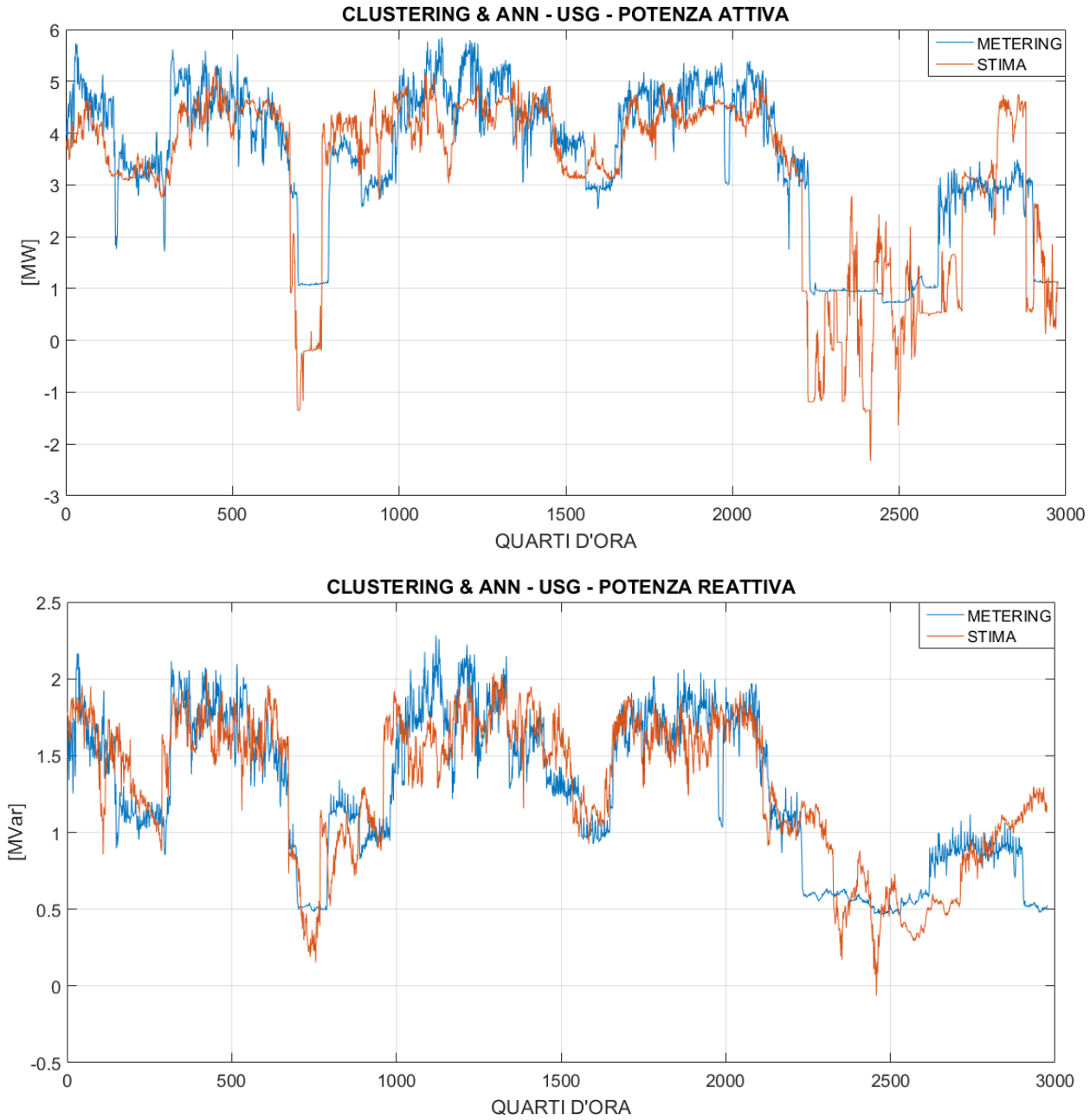


Figura 6.7. Grafico dei risultati della previsione di carico attivo e reattivo dell'utenza USG ottenuti tramite due ANN scelte a seconda del tipo di giorno in cui si trova il carico da stimare (feriale o festivo). Le reti sono state allenate con i dati relativi ai primi 11 mesi del 2017, divisi in due cluster, e i valori plottati sono relativi all'intero mese di dicembre 2017.

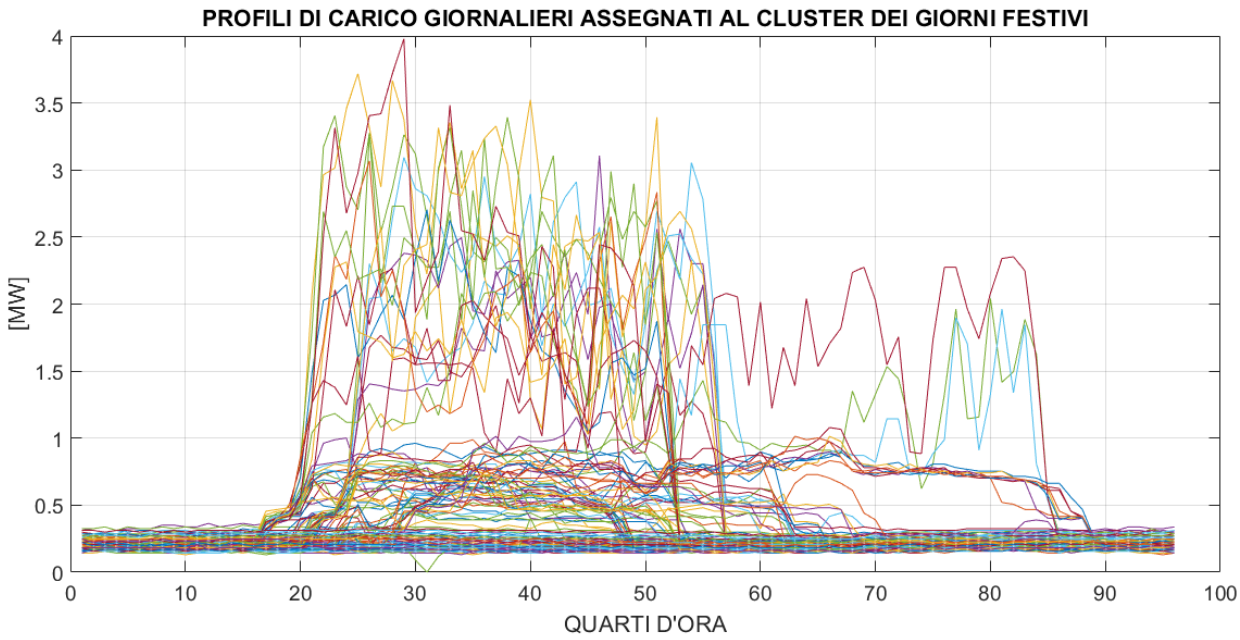


Figura 6.8. Plot di tutti i profili di carico giornalieri del 2017, per il carico ALT, che popolano il cluster dei giorni festivi. Come si può facilmente notare, la classe si potrebbe dividere ulteriormente in almeno altri 3 sottogruppi.

interrotti, come può essere per esempio nel giorno di Natale. I risultati potrebbero migliorare andando a partizionare il dataset in un numero di cluster maggiore, ma si creerebbero due problemi, il primo consiste nella difficoltà di individuare a priori a quale cluster corrisponde il giorno del carico da stimare, il secondo, consiste nella possibilità che alcuni cluster siano popolati da un numero molto ridotto di dati, non sufficiente per allenare adeguatamente una rete neurale o per trovare coefficienti coerenti di una regressione lineare.

Capitolo 7

Confronto dei risultati

7.1 Confronto quantitativo dei metodi

È stato deciso di quantificare la bontà della stima dei differenti metodi confrontando la media del loro errore relativo nel mese di dicembre. Per un equo confronto, entrambi i metodi sono stati "allenati" sui primi 11 mesi del 2017 e testati su dicembre dello stesso anno. I valori di confronto sono stati ottenuti normalizzando l'errore assoluto medio (Mean Absolute Error, MAE) di ogni carico con il relativo picco annuale di potenza assorbita e mediando i risultati così ottenuti tra tutti e 11 i carichi studiati. In altre parole, per un carico viene calcolato il suo MAE, questo valore viene poi diviso con il relativo valore massimo di assorbimento rilevato durante l'intero 2017. Questa operazione viene ripetuta per ognuno degli 11 carichi e, infine, viene fatta una media. Risultati:

- Regressione con 10 regressori: 0.195;
- Regressione con 17 regressori: 0.198;
- Regressione completa di tutti e 47 i regressori: 0.166;
- ANN: 0.103;
- Combinazione lineare tra ANN e regressione: 0.122.

Gli stessi studi sono stati svolti anche per le potenze reattive, ottenendo le stesse conclusioni. Risultati:

- Regressione con 10 regressori: 0.198;

- Regressione con 17 regressori: 0.180;
- Regressione complete di tutti e 47 i regressori: 0.156;
- ANN: 0.118;
- Combinazione lineare tra ANN e regressione: 0.116.

Per quanto riguarda il metodo comprendente di clusterizzazione e due reti neurali differenti, l'errore è stato calcolato solo per i 6 utenti al quale il metodo è stato applicato (ALT, ERB, FEU, ITA, USG, RVA). E' perciò messo in confronto con l'errore ottenuto solo per quei 6 carichi con il metodo ANN (rivelatosi il più preciso).
Potenza attiva:

- ANN: 0.090;
- clusterizzazione e due ANN: 0.084.

Potenza reattiva:

- ANN: 0.111;
- clusterizzazione e due ANN: 0.112.

7.2 Individuazione del metodo più adatto per l'implementazione nei sistemi del TSO

Considerando che lo scopo di questo lavoro è quello di trovare un metodo per stimare l'assorbimento di potenza dei carichi industriali, da applicare nel vero sistema di controllo del TSO Italiano, è stata scelta la regressione lineare (a 47 regressori). Anche se la regressione non ha prodotto i risultati più precisi a livello teorico, esso rappresenta il metodo più adatto per essere implementato in un sistema di controllo reale per via della sua semplicità e stabilità. La regressione è inoltre meno influenzata dalla tangibile possibilità che qualche misura in tempo reale o nel sistema di metering dei carichi sia mancante o siano affette da errore. Una rete neurale è molto più condizionata da questi problemi. Soluzioni possono essere trovate, ma andando a rincarare ulteriormente la complessità del metodo. Per quanto riguarda i metodi concernenti la clusterizzazione sorge l'utleriore problema di individuare preventivamente il tipo di giorno in cui si trova il carico da stimare.

Capitolo 8

Implementazione e collaudo

Il metodo è stato implementato per mezzo di uno script Python nel quale il codice ha la funzione di produrre l'equazione nella stessa forma dell'equazione 4.18, dove le variabili indipendenti sono le misure disponibili in tempo reale come descritto nella *Capitolo 4* e i coefficienti sono stati trovati per mezzo di uno studio di regressione svolto su Matlab utilizzando i dati raccolti nell'intero anno 2017. Il codice è inserito in una macchina di prova che lavora con gli stessi input della macchina correntemente in funzione in Terna. I risultati ottenuti dalle due macchine sono infine confrontati per verificare qualora il metodo proposto porti effettivamente ad un miglioramento delle stime dei carichi. Il confronto è stato effettuato nei giorni dal 18 al 22 gennaio 2019, un periodo di tempo comprensivo sia di giorni feriali che festivi.
Un confronto quantitativo è stato effettuato in modo uguale a quello illustrato nella *Capitolo 7*. Ovvero, i valori di confronto sono stati ottenuti normalizzando l'errore assoluto medio (MAE) di ogni carico con il relativo picco di potenza assorbita e mediando i risultati così ottenuti tra tutti e 11 i carichi studiati. Il valore dei picchi scelto è relativo all'anno 2017, così da rendere possibile un confronto anche con i valori trovati nel *Capitolo 7*. In altre parole, per un carico viene calcolato il suo MAE, questo valore viene poi diviso con il relativo valore massimo di assorbimento rilevato durante l'intero 2017. Questa operazione viene ripetuta per ognuno degli 11 carichi e, infine, viene fatta una media. Risultati:

- stima proposta, potenza attiva: 0.186;
- stima attuale, potenza attiva: 0.609;

- stima proposta, potenza reattiva: 0.259;
- stima attuale, potenza reattiva: 0.772.

I risultati mostrano un miglioramento sostanziale della stima dei carichi. Nelle figure 8.1, 8.2, 8.3 sono riportati alcuni risultati. Dai risultati è anche emerso che la stima attuale dell'assorbimento di potenza attiva del carico FEU, effettuata tramite bilancio è molto affidabile e rappresenta l'unico caso in cui la stima attuale è da preferirsi alla stima proposta in questo lavoro. Tutti i risultati delle stime sono presentate in forma grafica e riportate in appendice B.

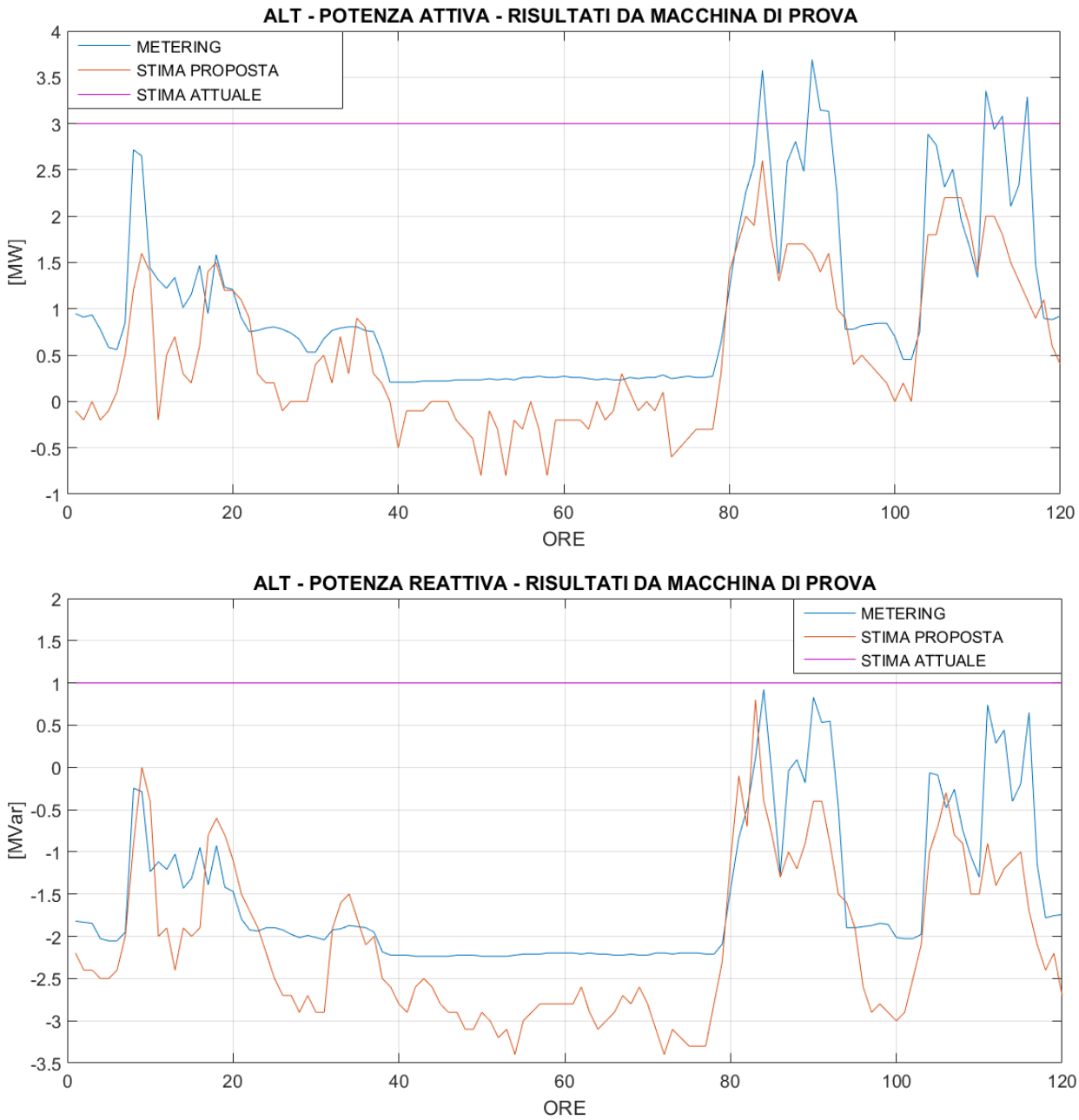


Figura 8.1. Risultati provenienti dalla macchina di prova confrontati con la stima attuale operata da Terna. Il confronto è stato effettuato tra i giorni 18 e 22 gennaio 2019 e si riferisce al carico, sia attivo che reattivo, dell'utente ALT. E' stato valutato un solo quarto d'ora ogni quattro per alleggerire l'onerosa operazione di estrazione dei dati dai database Terna.

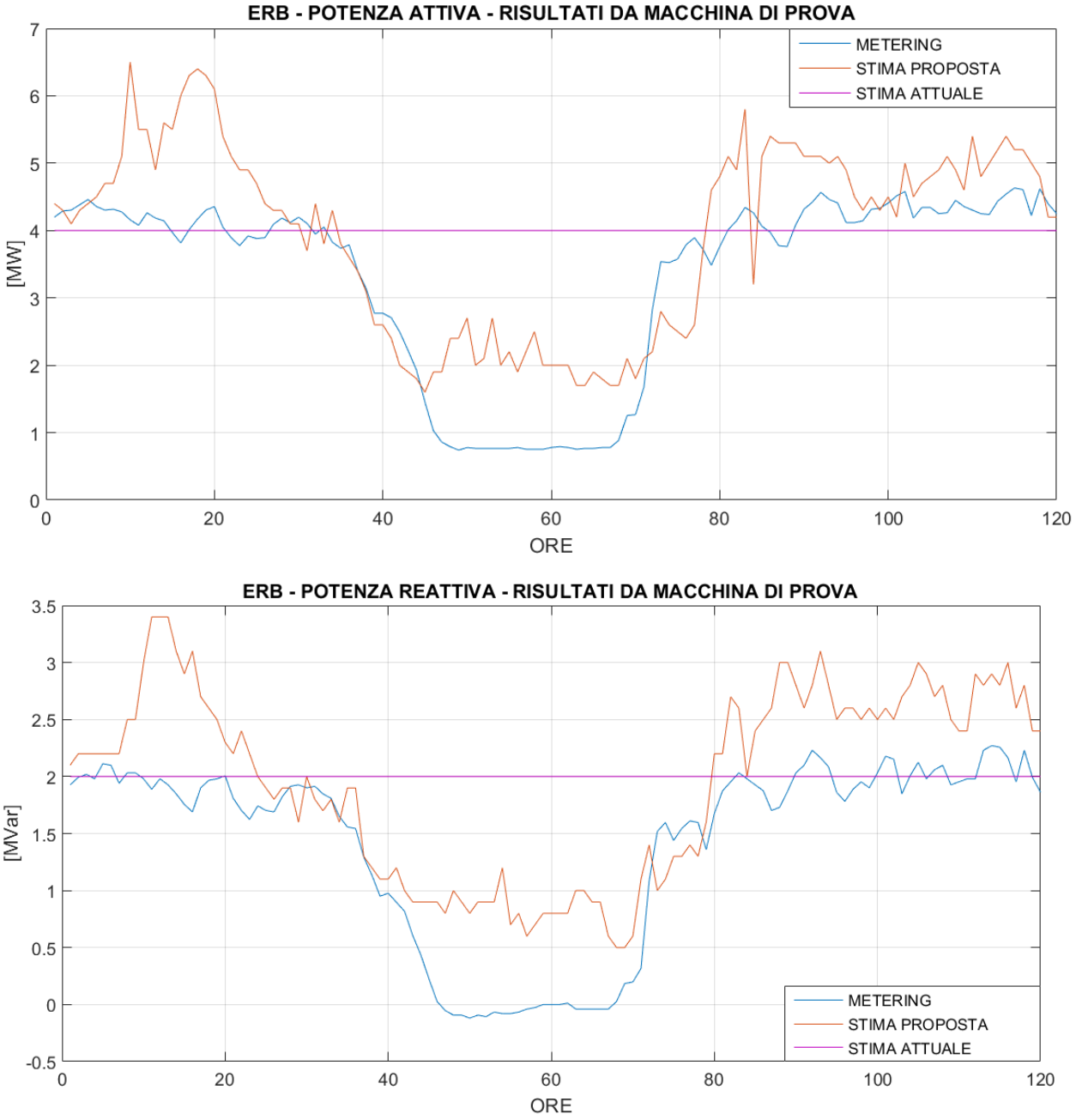


Figura 8.2. Risultati provenienti dalla macchina di prova confrontati con la stima attuale operata da Terna. Il confronto è stato effettuato tra i giorni 18 e 22 gennaio 2019 e si riferisce al carico, sia attivo che reattivo, dell'utente ERB. E' stato valutato un solo quarto d'ora ogni quattro per alleggerire l'onerosa operazione di estrazione dei dati dai database Terna.

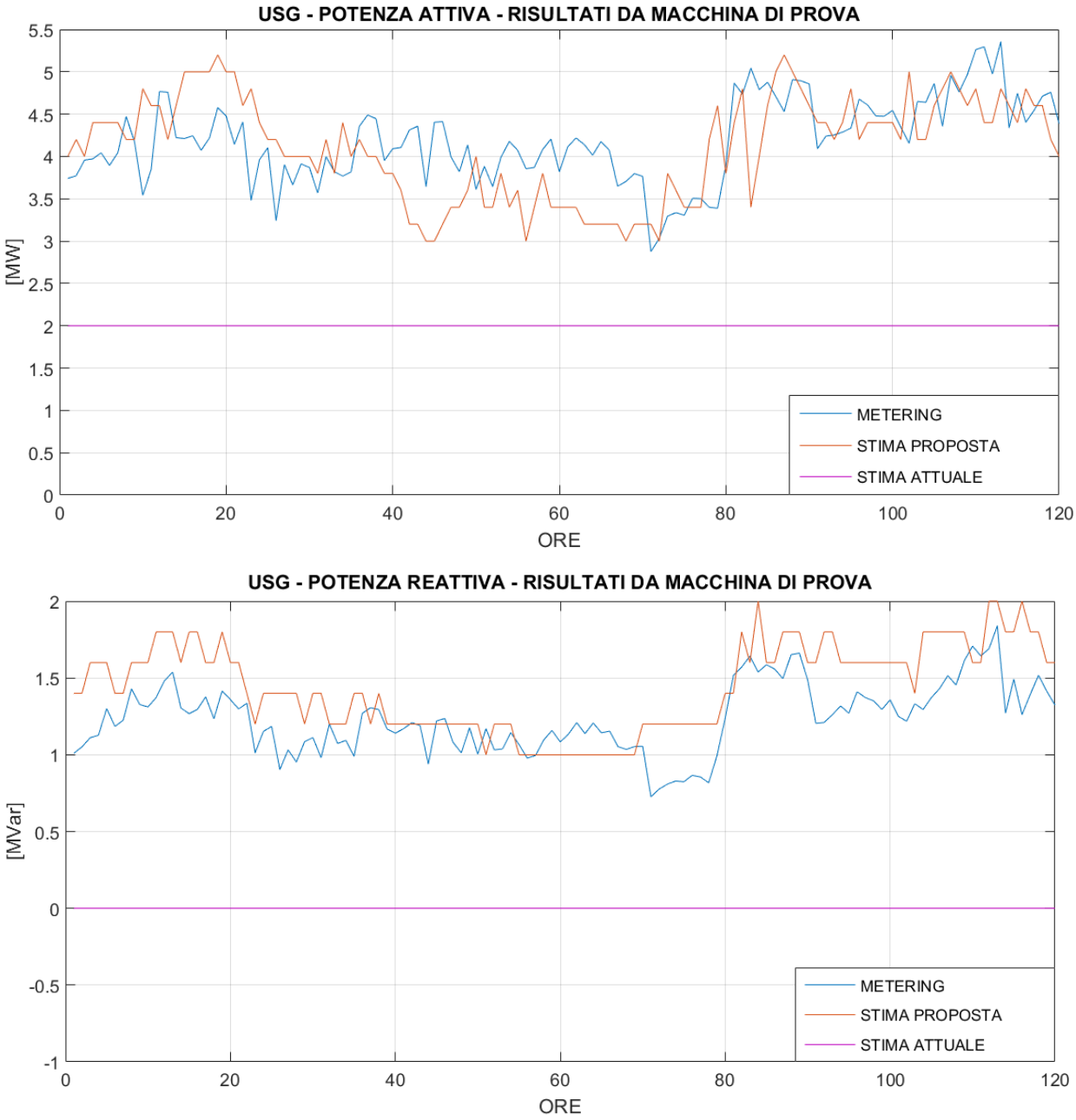


Figura 8.3. Risultati provenienti dalla macchina di prova confrontati con la stima attuale operata da Terna. Il confronto è stato effettuato tra i giorni 18 e 22 gennaio 2019 e si riferisce al carico, sia attivo che reattivo, dell'utente USG. E' stato valutato un solo quarto d'ora ogni quattro per alleggerire l'onerosa operazione di estrazione dei dati dai database Terna.

Per rendere inequivocabile l'idoneità del metodo proposto esso è stato testato anche in una seconda isola di carico, situata nel milanese e di dimensioni più ridotte rispetto all'isola sul quale è stato sviluppato questo metodo. I carichi da stimare sono 4 e i regressori 17. In isole di carico più piccole il minor numero di regressori è compensato da una maggior correlazione tra i carichi e le misure disponibili in tempo reale.

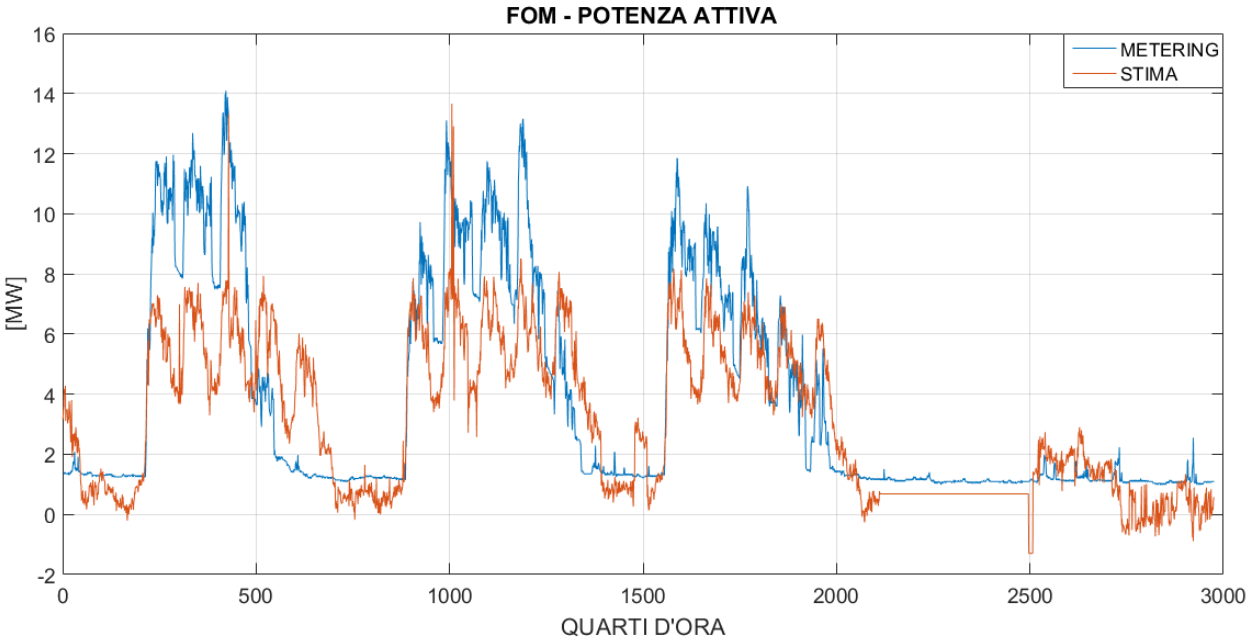


Figura 8.4. Grafico dei risultati della previsione di potenza attiva dell'utente FOM. I coefficienti della regressione sono stati ottenuti dall'analisi dei dati relativi ai primi 11 mesi del 2017 e i valori plottati sono relativi all'intero mese di dicembre 2017.

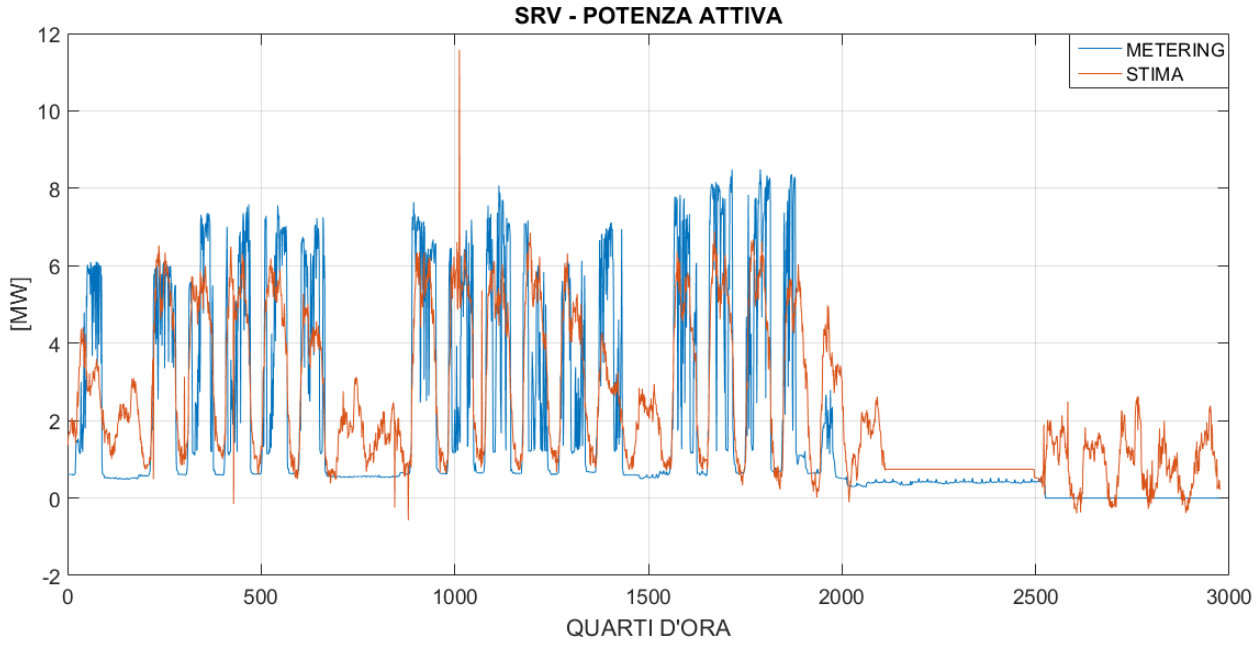


Figura 8.5. Grafico dei risultati della previsione di potenza attiva dell'utente SRV. I coefficienti della regressione sono stati ottenuti dall'analisi dei dati relativi ai primi 11 mesi del 2017 e i valori plottati sono relativi all'intero mese di dicembre 2017.

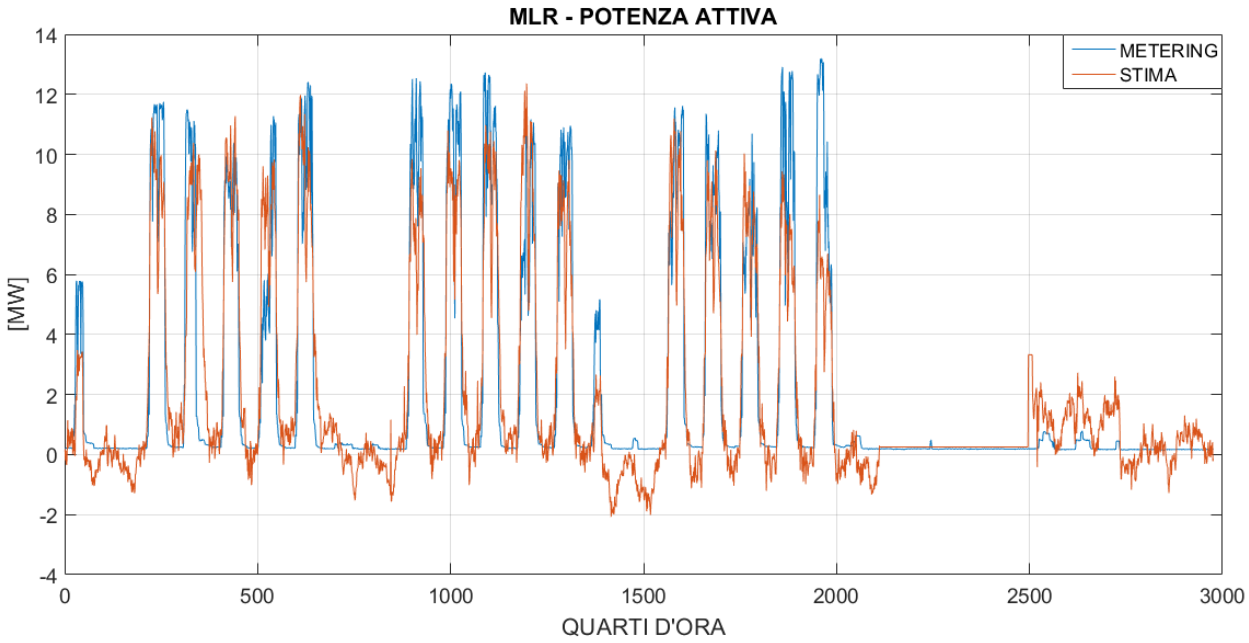


Figura 8.6. Grafico dei risultati della previsione di potenza attiva dell'utente MLR. I coefficienti della regressione sono stati ottenuti dall'analisi dei dati relativi ai primi 11 mesi del 2017 e i valori plottati sono relativi all'intero mese di dicembre 2017.

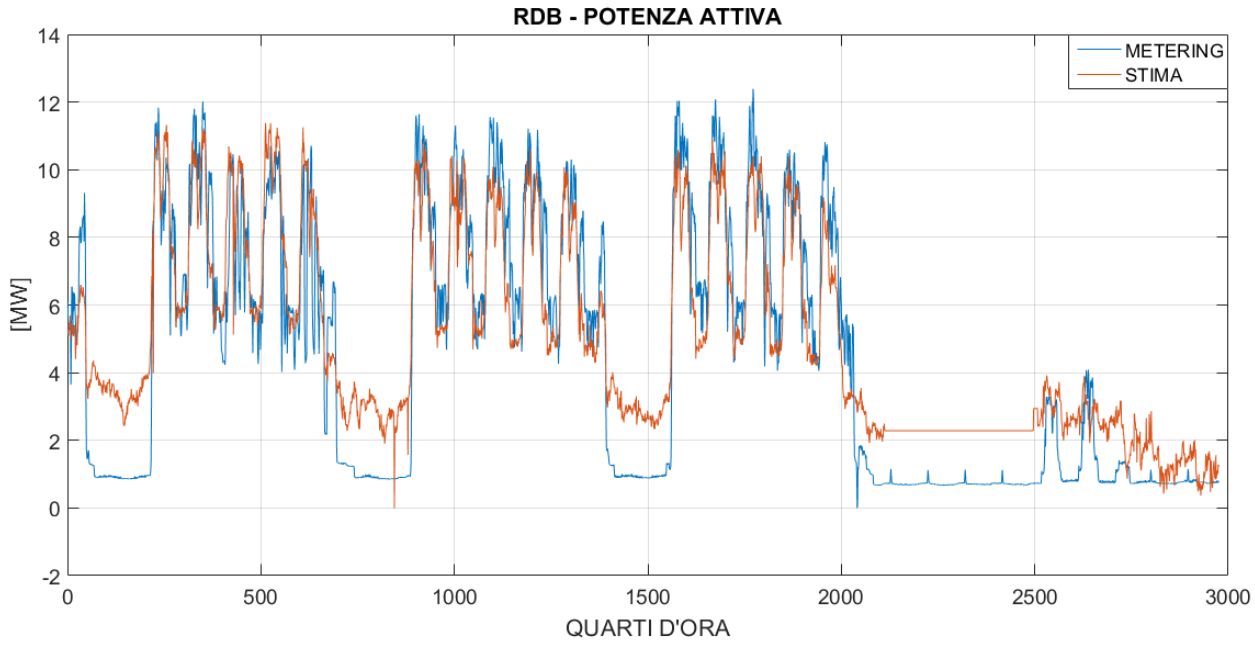


Figura 8.7. Grafico dei risultati della previsione di potenza attiva dell'utente RDB. I coefficienti della regressione sono stati ottenuti dall'analisi dei dati relativi ai primi 11 mesi del 2017 e i valori plottati sono relativi all'intero mese di dicembre 2017.

Capitolo 9

Conclusioni

Sono state presentate diverse tecniche di previsione di carico per utenti industriali connessi alla rete in alta tensione. Tra loro, il miglior risultato, a livello teorico, è considerato essere una combinazione di una regressione lineare multipla e di una rete neurale. La regressione utilizza come variabili dipendenti le misure in tempo reale dei flussi di potenza sulle linee in uscita dalle stazioni, le produzioni degli impianti di generazione e gli assorbimenti delle cabine primarie. Essa è concepita non per sfruttare una relazione causale tra qualche grandezza disponibile in tempo reale e la grandezza da prevedere, ma sfrutta le somiglianze dei profili di carico. La rete neurale, invece, utilizza i dati storici di assorbimento e degli indici indicanti in che giorno della settimana e in che ora del giorno cade il carico da stimare e se questo cade in una giornata festiva nazionale. La rete neurale si è rivelata essere la più precisa, ma talvolta produce valori molto distanti da quelli reali. Mediando i suoi risultati con quelli ottenuti tramite una regressione si ottiene un metodo in grado di fornire risultati con errori assoluti sempre contenuti, al prezzo di una precisione leggermente inferiore.
Per l'implementazione nei sistemi dell'azienda Terna è stata però scelta la sola regressione (la stessa illustrata pocanzi), in quanto comunque sufficientemente precisa e molto più semplice, robusta ed affidabile delle reti neurali. Il metodo è stato implementato in una macchina di prova che ha permesso il confronto tra il metodo di stima proposto e quello attualmente in opera. I risultati mostrano un miglioramento sostanziale delle previsioni, sia di potenza attiva che di potenza reattiva.
Un ulteriore convalida è stata effettuata applicando il metodo in una seconda isola di carico. Anche in questo caso i risultati indicano che il metodo è efficace.

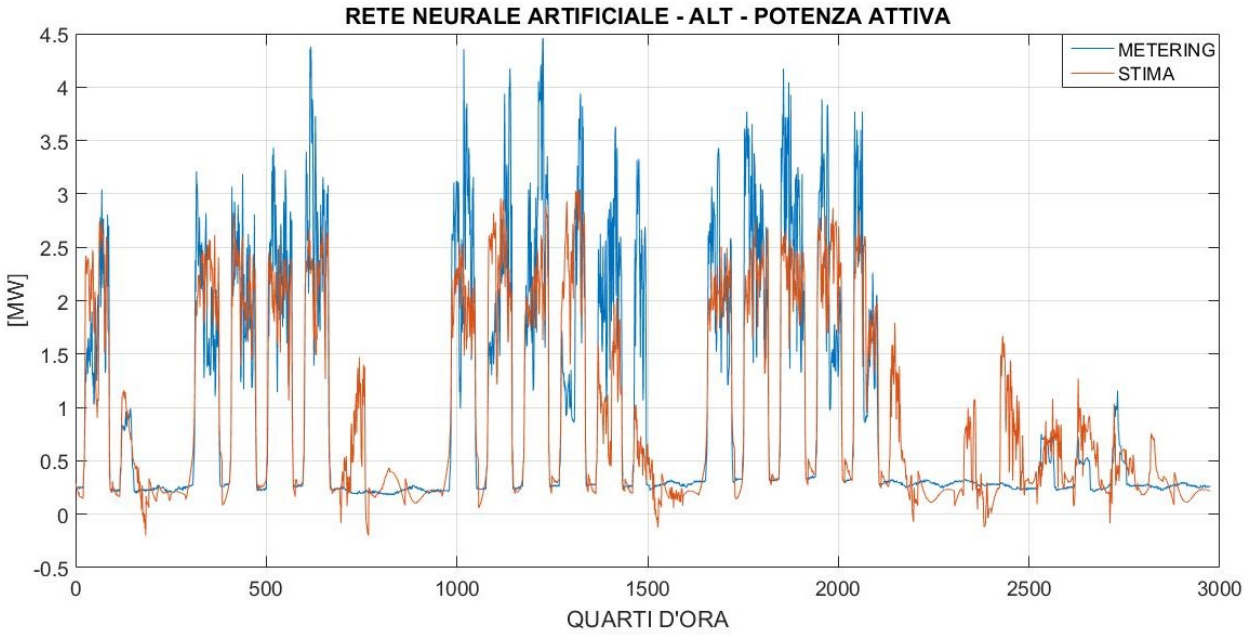
Si ritiene quindi che il metodo proposto sia implementabile nei sistemi dell'azienda Terna e possa permettere un rinvio della sostituzione dei sistemi di misura attuali.
Possibili sviluppi di questa tesi si potrebbero concentrare sulla fase di clusterizzazione. In particolare è da risolvere il problema dell'individuazione, a priori, del cluster al quale il giorno da stimare appartiene. In seguito saranno necessari competenze di programmazione e gestione dati per l'implementazione dei metodi più complessi nei sistemi del TSO.

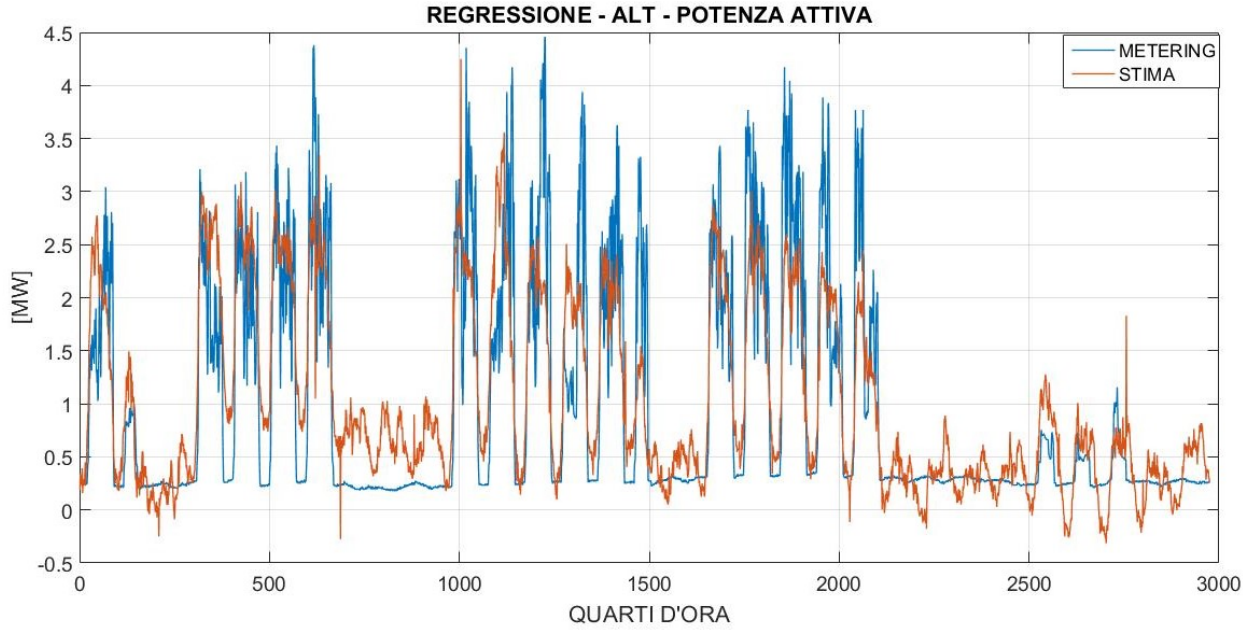
Appendice A

Test sui metodi - Risultati completi

Di seguito sono riportati i risultati della previsione di carico su dicembre 2017 per i metodi principali esposti in questo elaborato, ovvero: regressione lineare multipla e rete neurale artificiale (come descritto nei *capitoli 4* e *5*).

ALT


RETE NEURALE ARTIFICIALE - ALT - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA


REGRESSIONE - ALT - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA

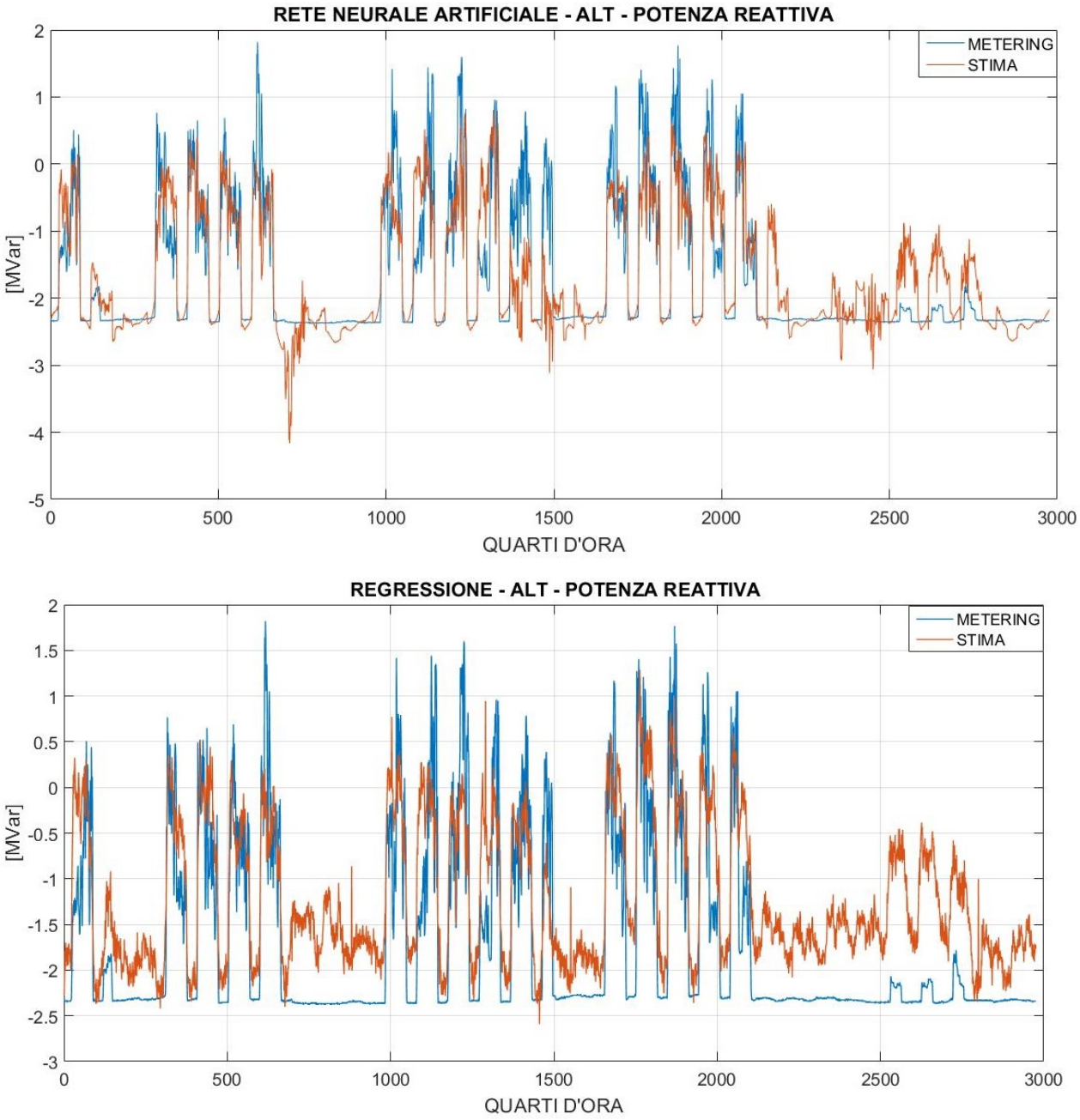
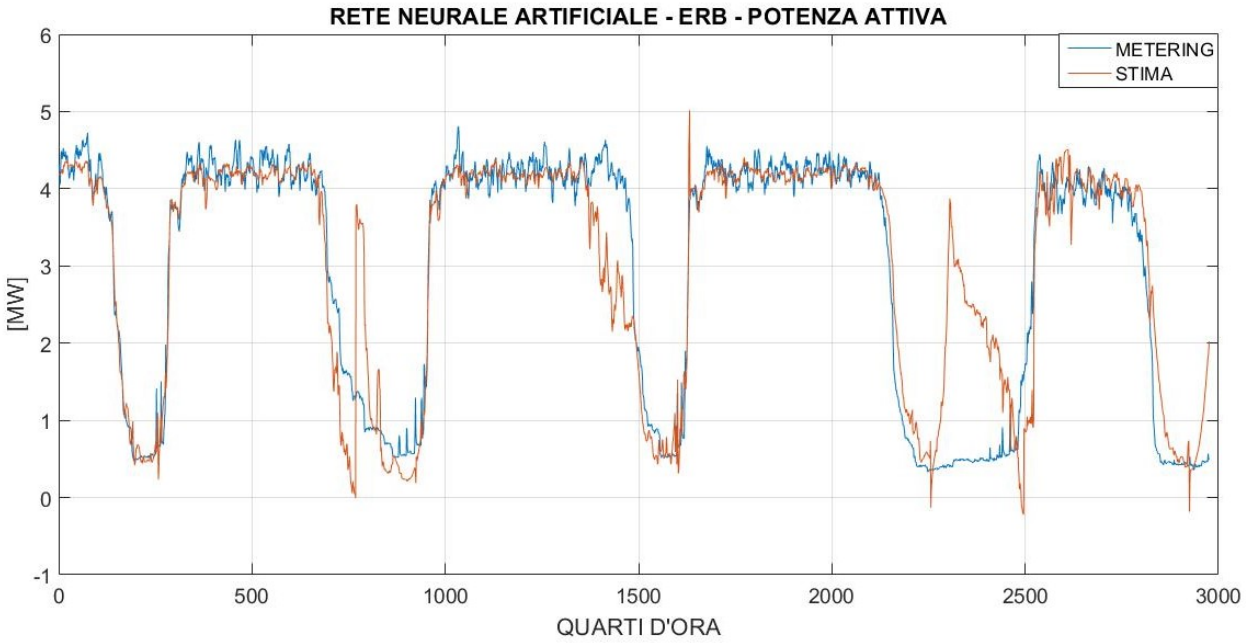
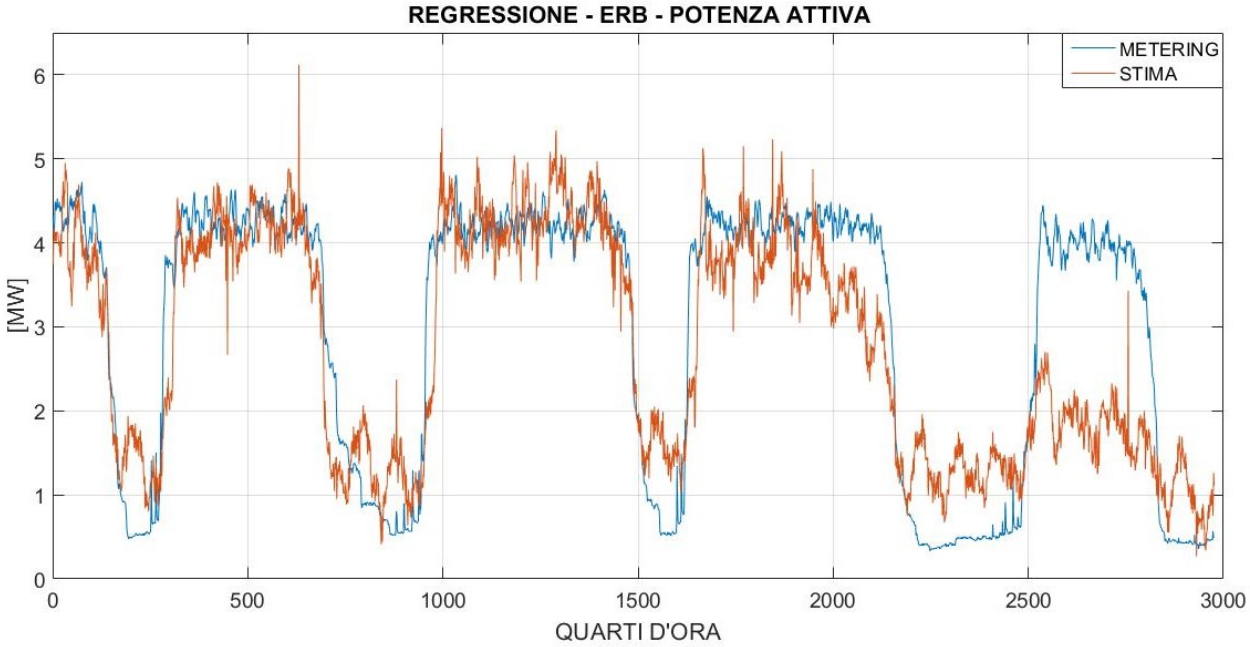


Figura A.1. Risultati della stima di potenza attiva e reattiva per l'utente ALT nel mese di dicembre 2017.

ERB


RETE NEURALE ARTIFICIALE - ERB - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA


REGRESSIONE - ERB - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA

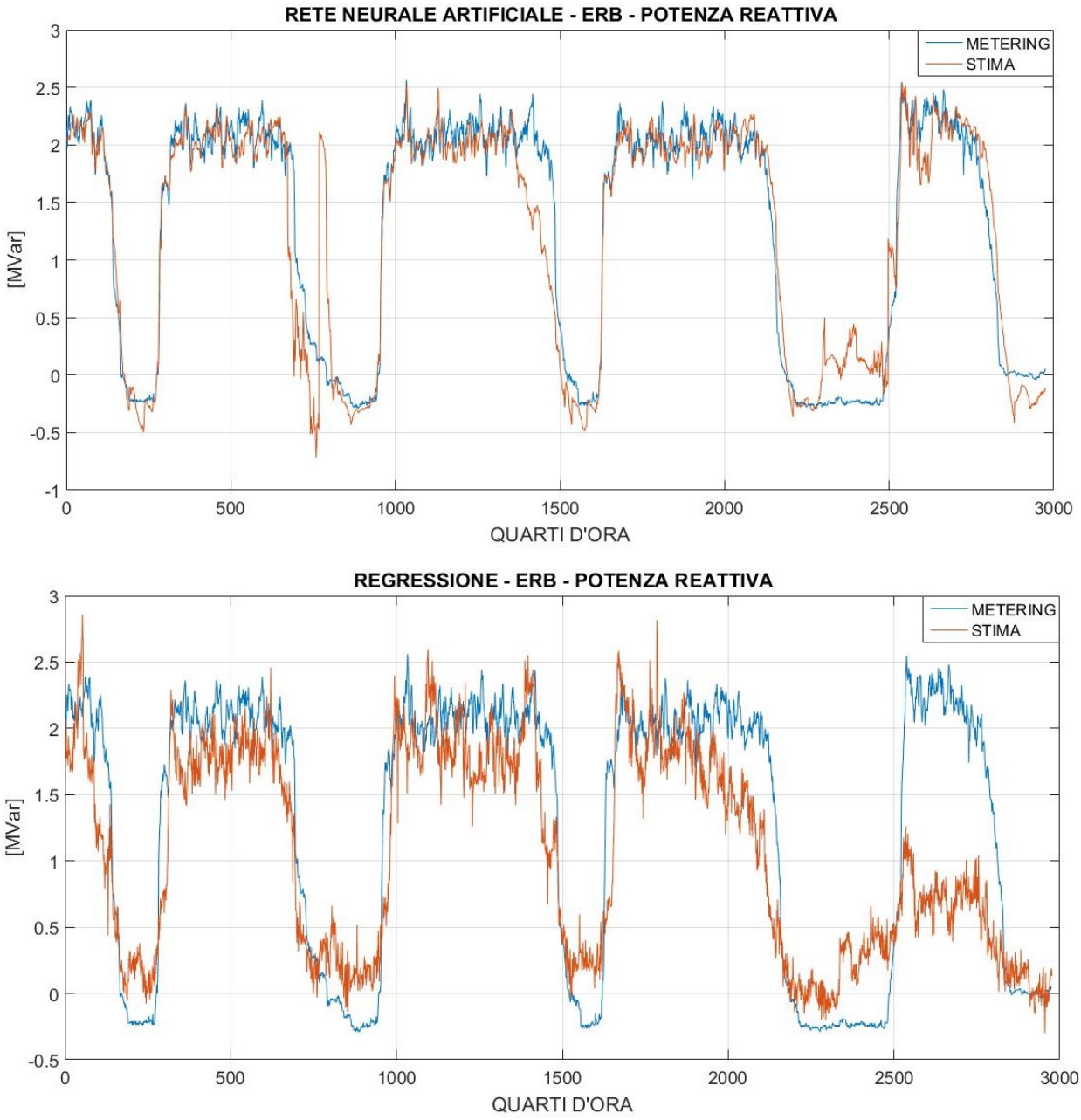
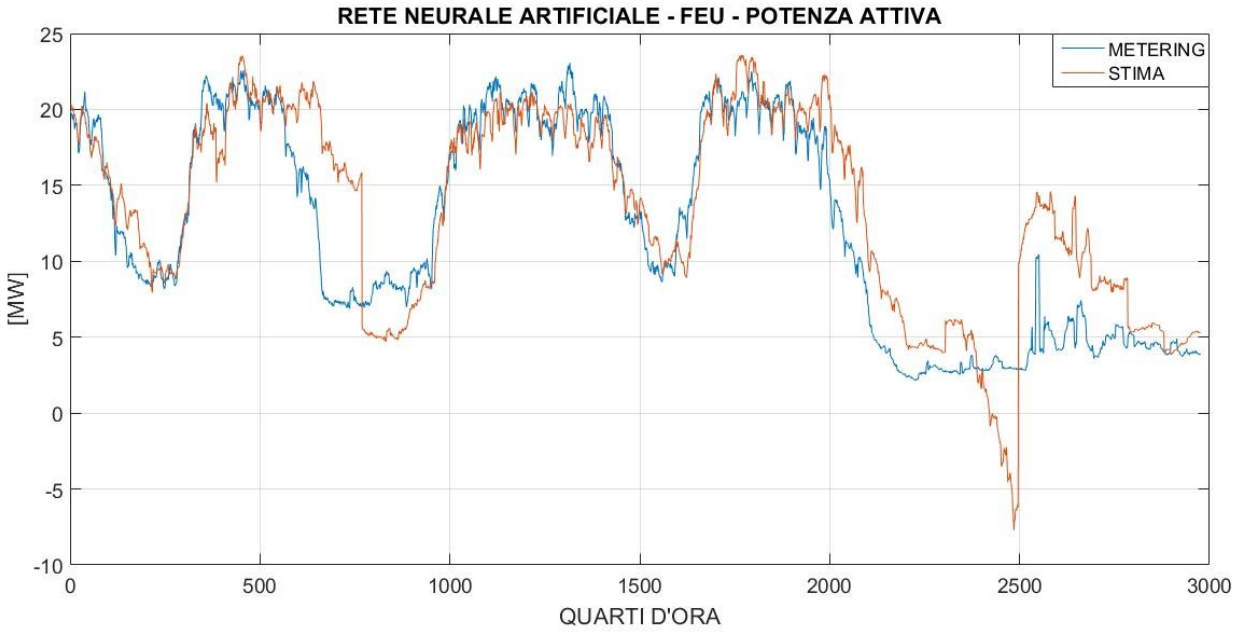
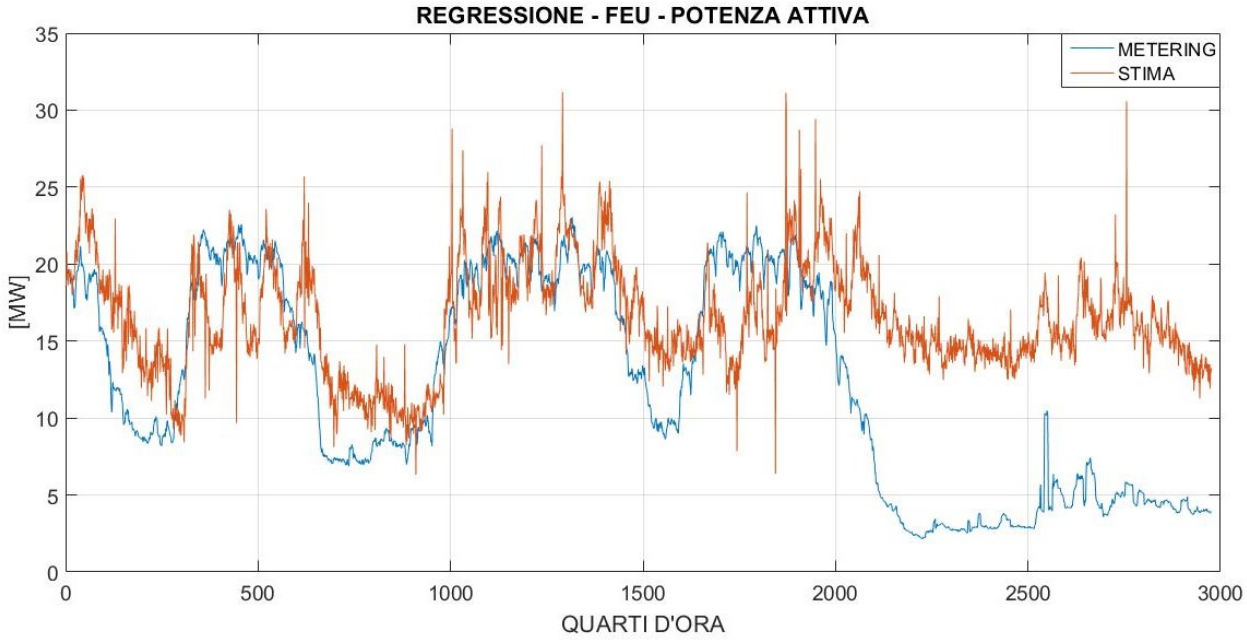


Figura A.2. Risultati della stima di potenza attiva e reattiva per l'utente ERB nel mese di dicembre 2017.

FEU


RETE NEURALE ARTIFICIALE - FEU - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA
25
20
15
10
5
0
-5
-10
0
500
1000
1500
2000
2500
3000


REGRESSIONE - FEU - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA
35
30
25
20
15
10
5
0
0
500
1000
1500
2000
2500
3000

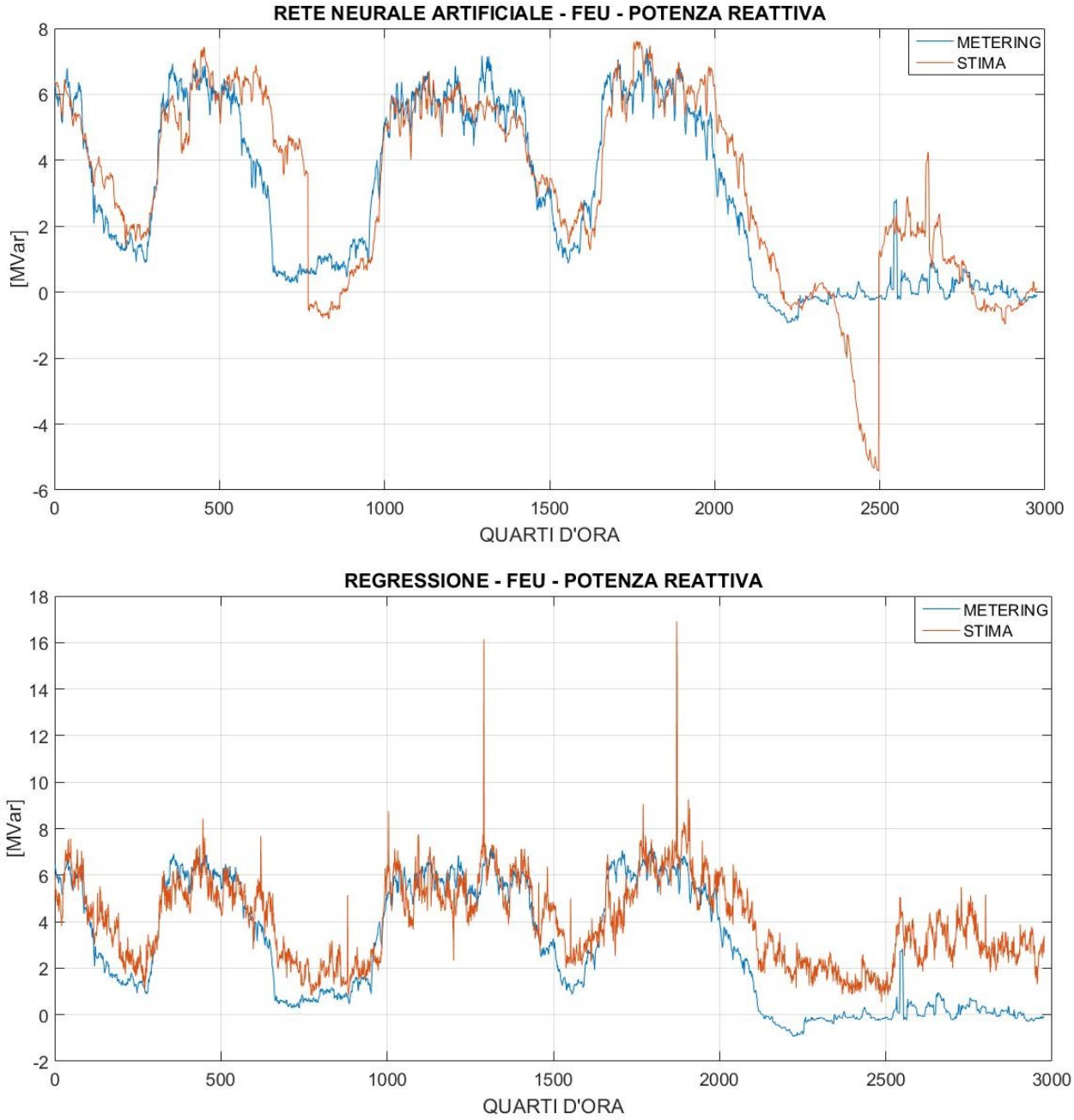
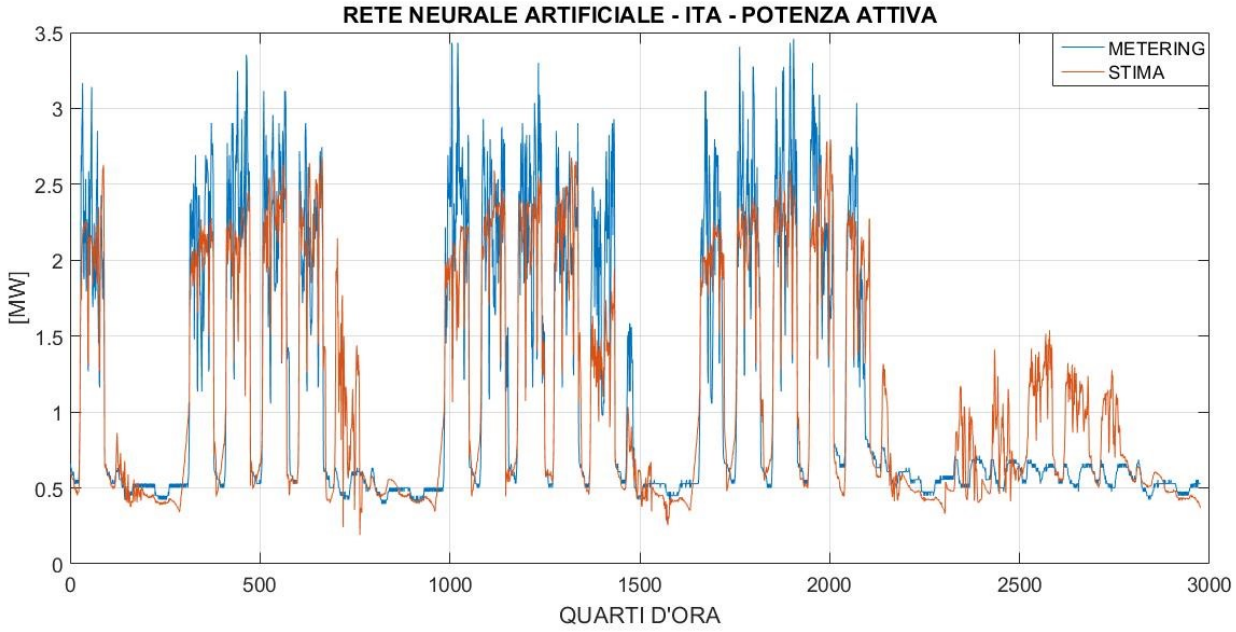
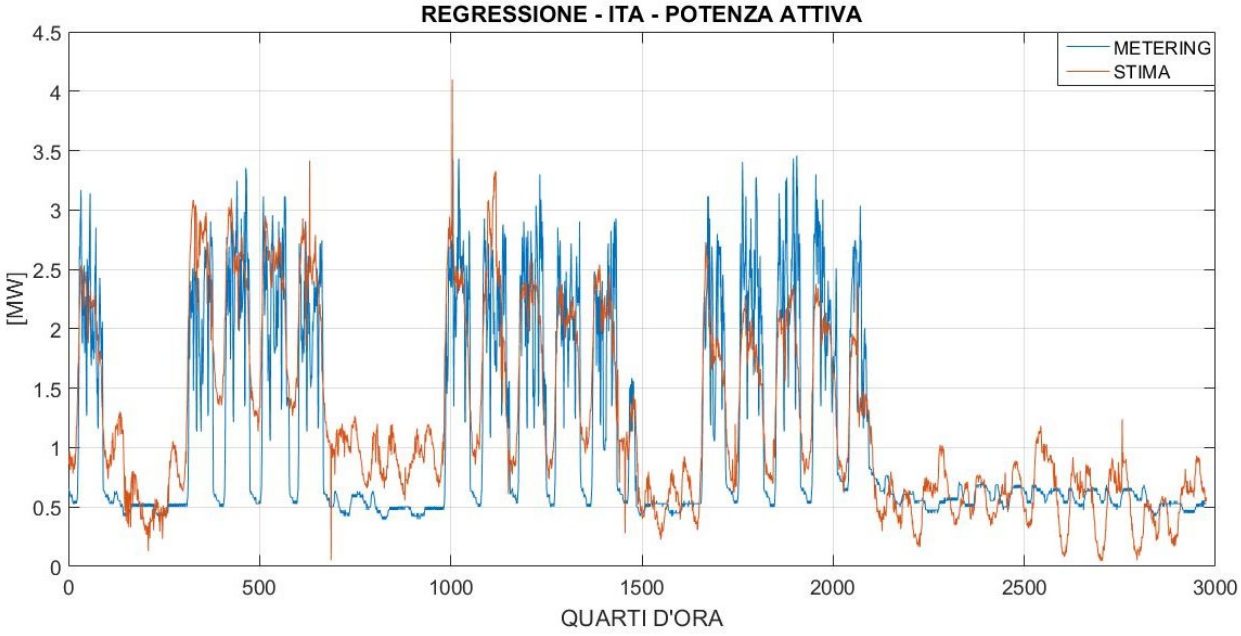


Figura A.3. Risultati della stima di potenza attiva e reattiva per l'utente FEU nel mese di dicembre 2017.

ITA


RETE NEURALE ARTIFICIALE - ITA - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA
0
0.5
1
1.5
2
2.5
3
3.5
0
500
1000
1500
2000
2500
3000


REGRESSIONE - ITA - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
0
500
1000
1500
2000
2500
3000

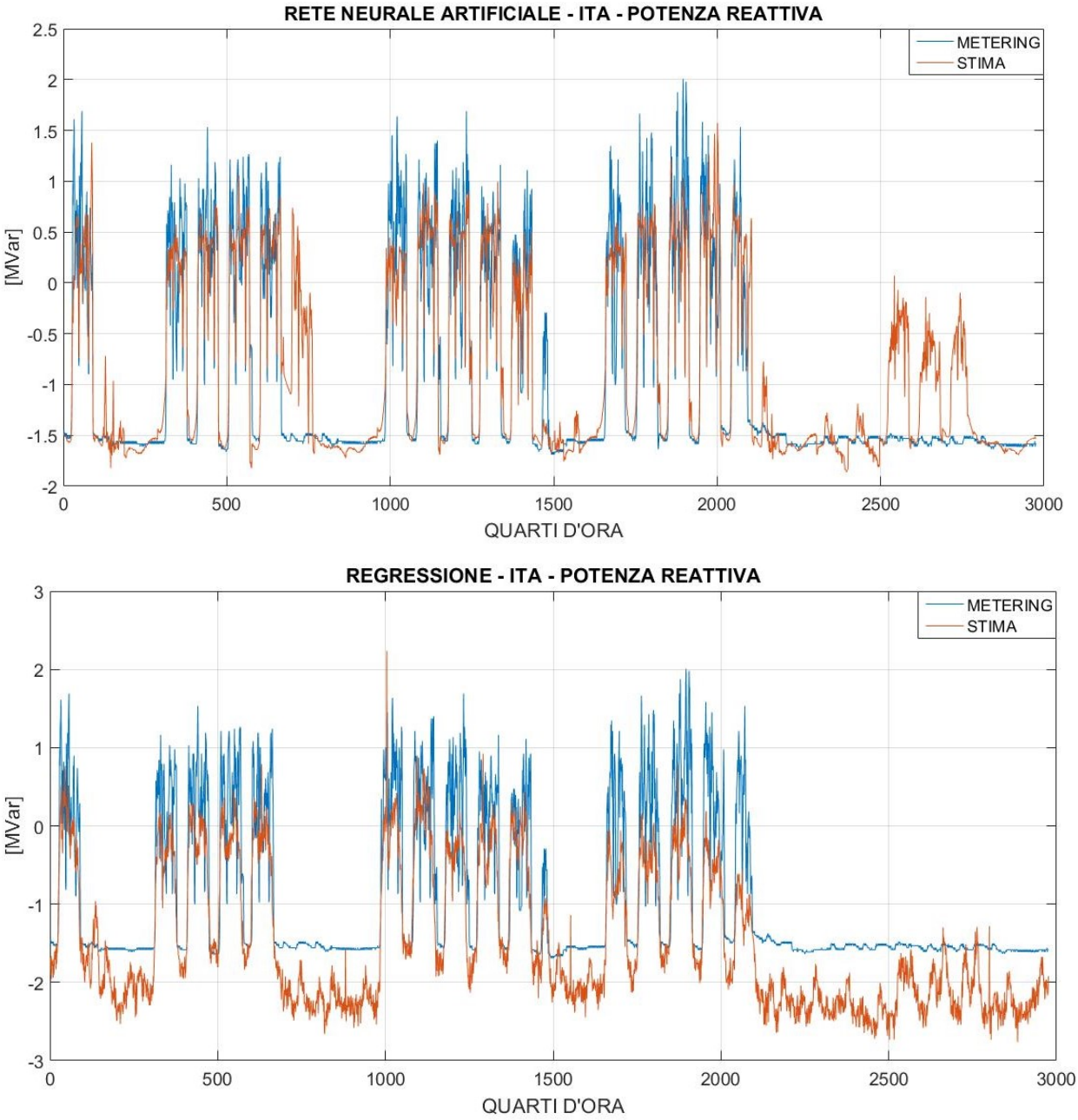
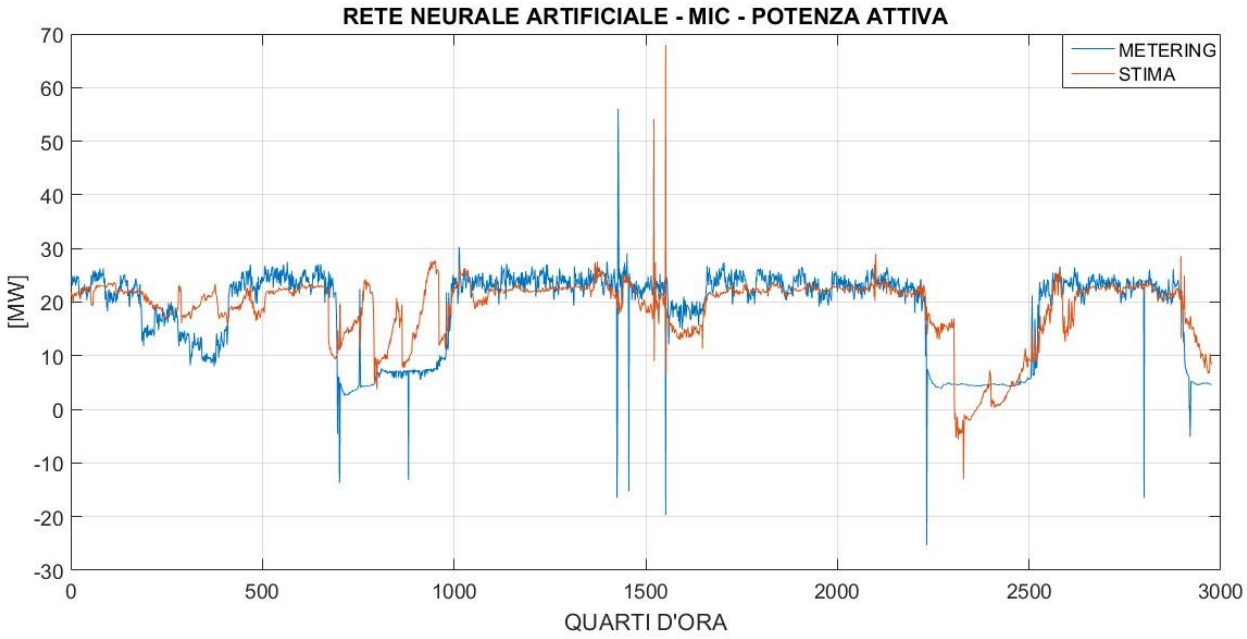
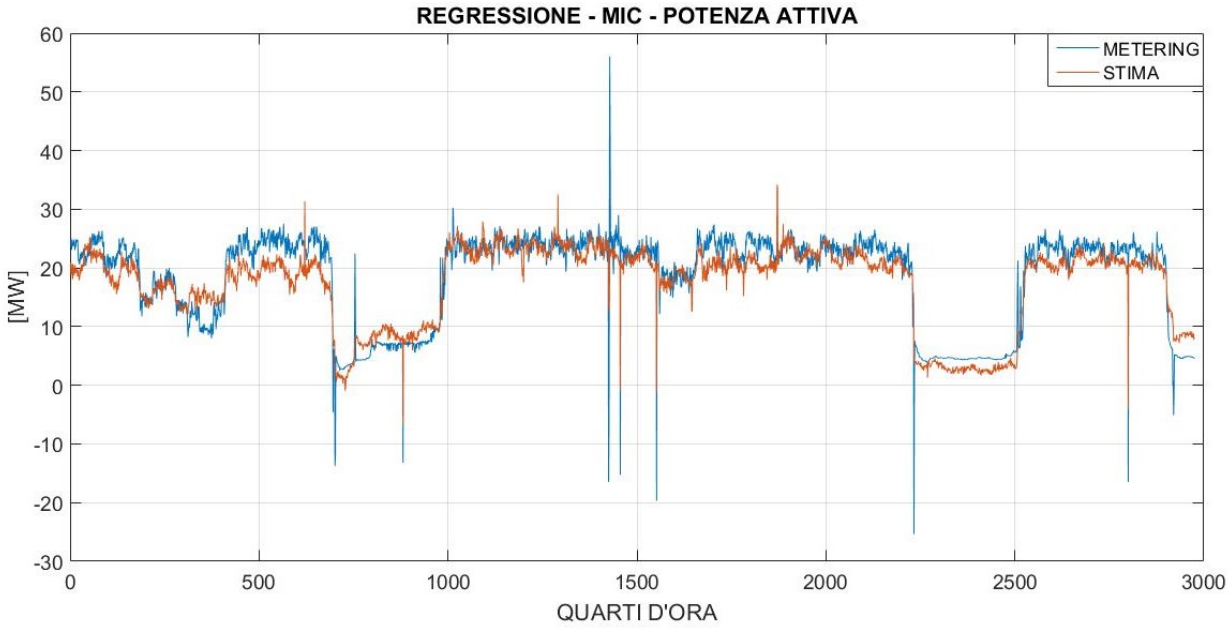


Figura A.4. Risultati della stima di potenza attiva e reattiva per l'utente ITA nel mese di dicembre 2017.

MIC


RETE NEURALE ARTIFICIALE - MIC - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA


REGRESSIONE - MIC - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA

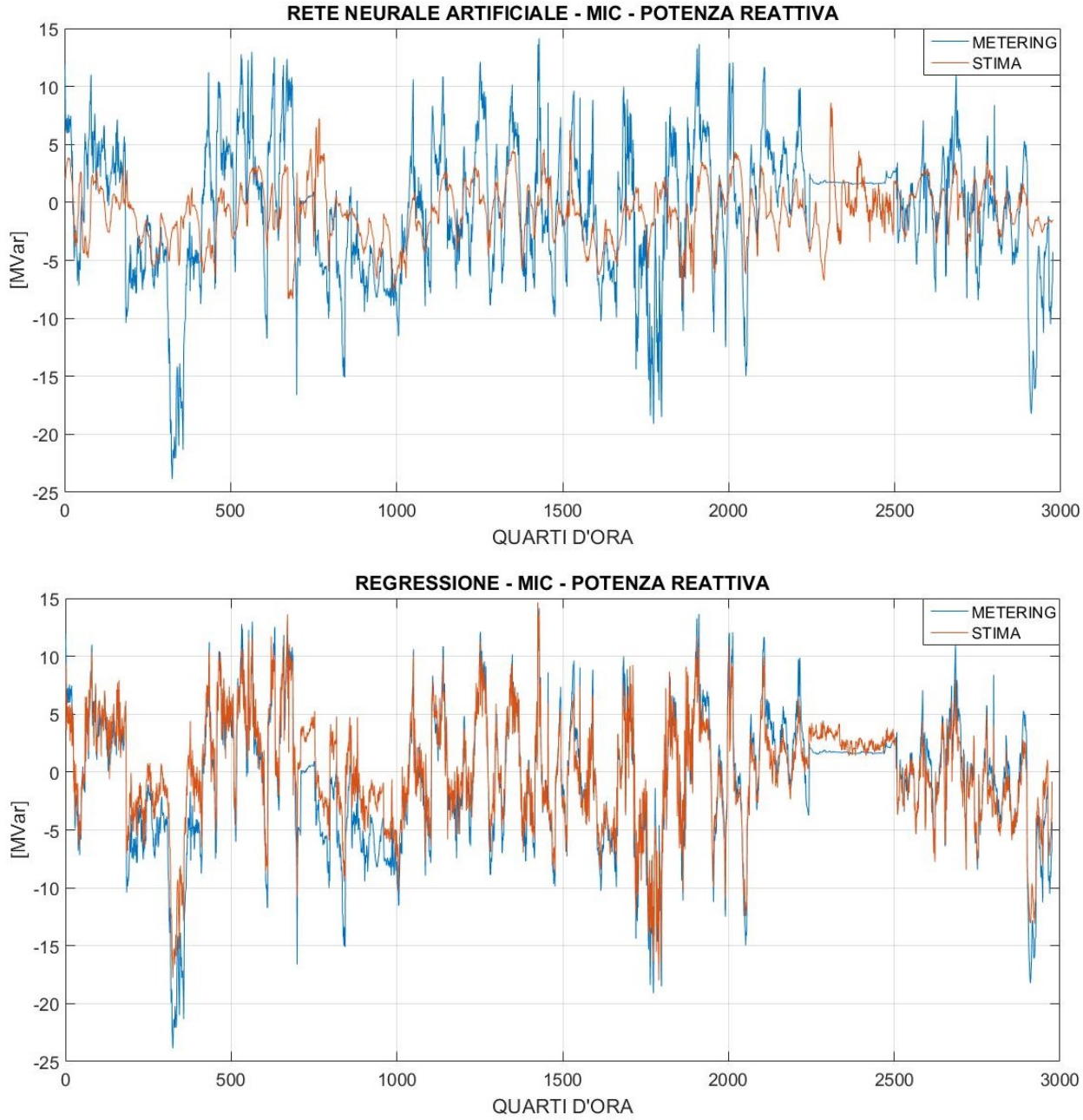
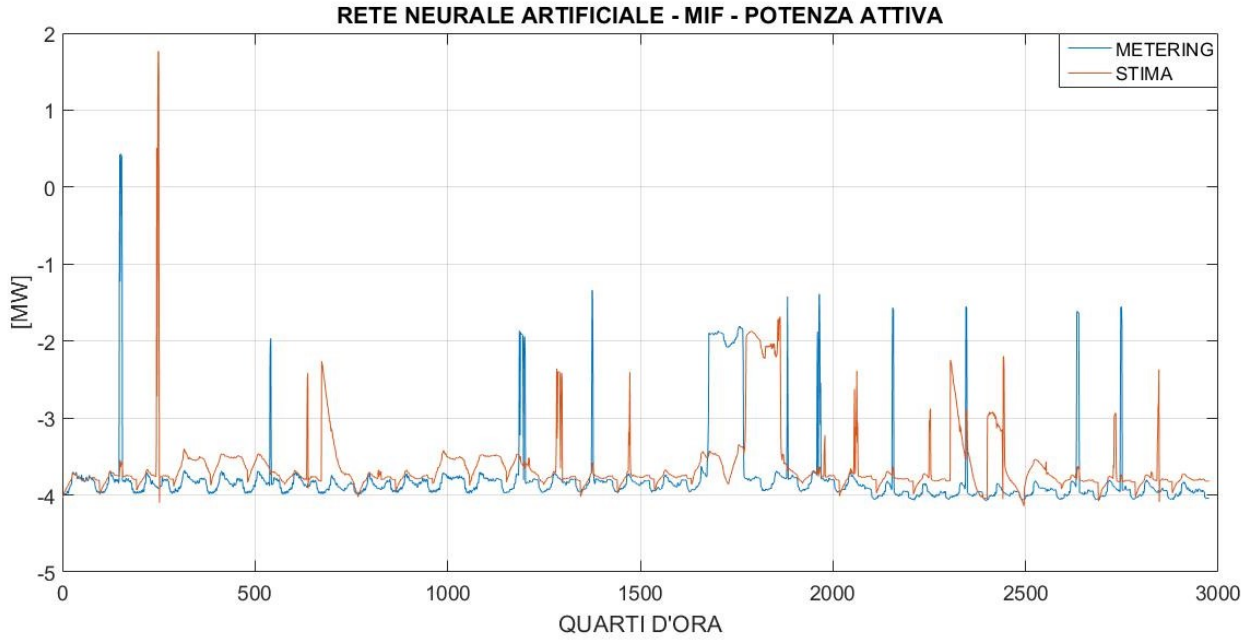
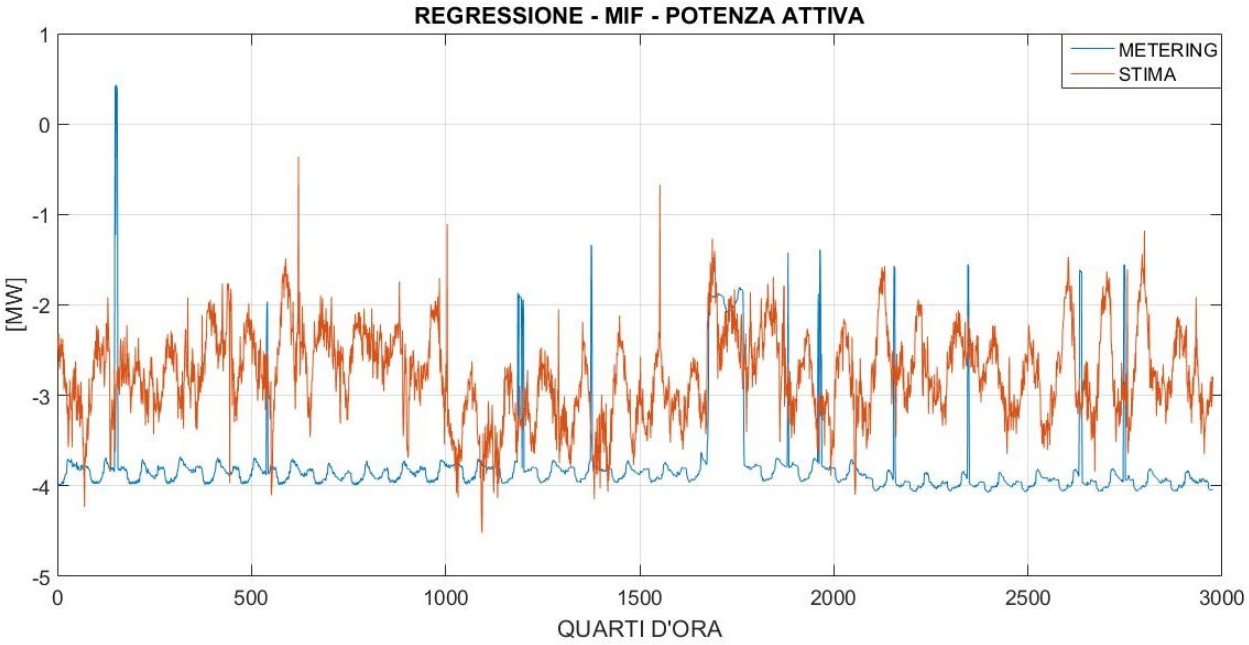


Figura A.5. Risultati della stima di potenza attiva e reattiva per l'utente MIC nel mese di dicembre 2017.

MIF


RETE NEURALE ARTIFICIALE - MIF - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA
2
1
0
-1
-2
-3
-4
-5
0
500
1000
1500
2000
2500
3000


REGRESSIONE - MIF - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA
1
0
-1
-2
-3
-4
-5
0
500
1000
1500
2000
2500
3000

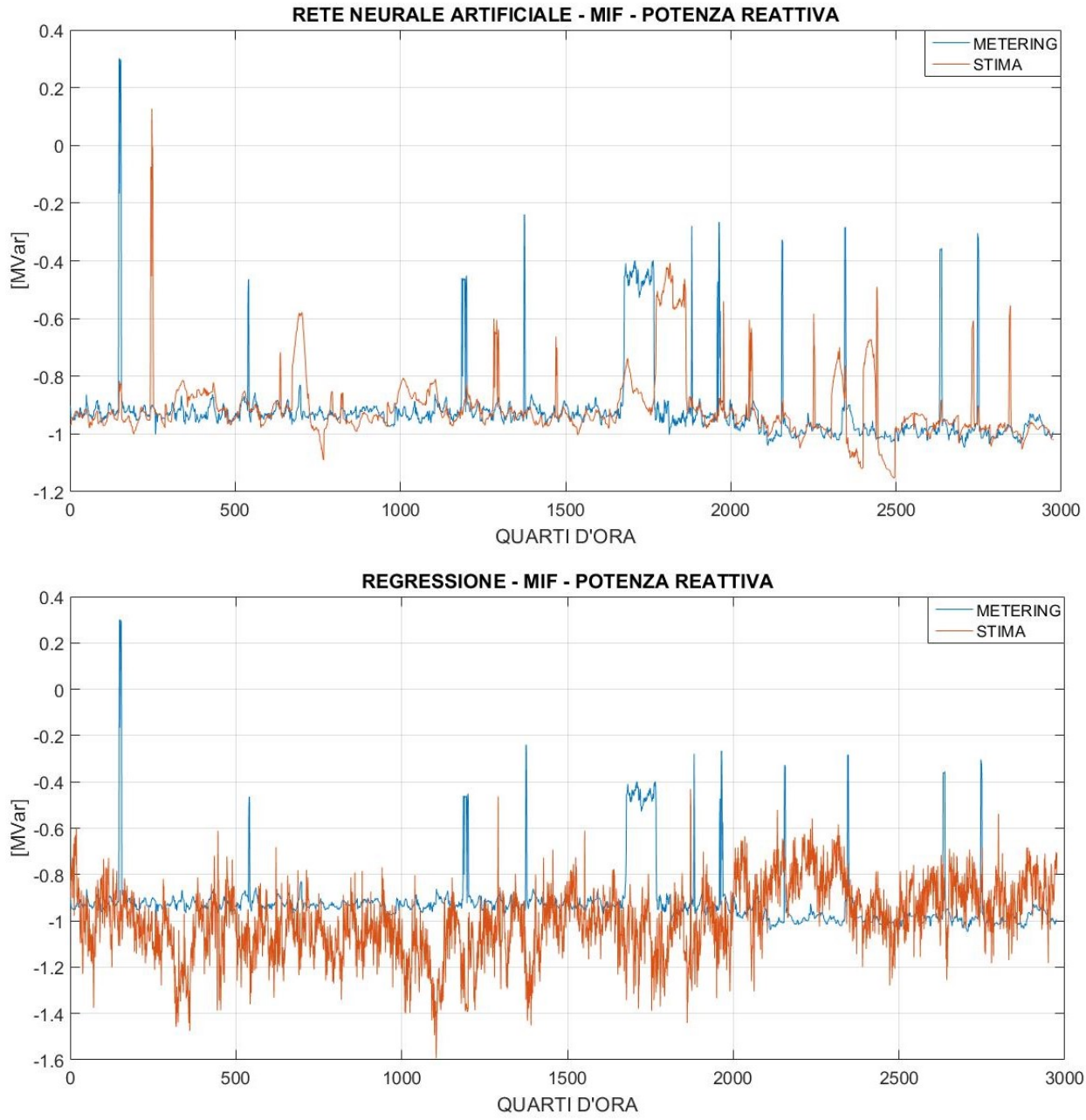
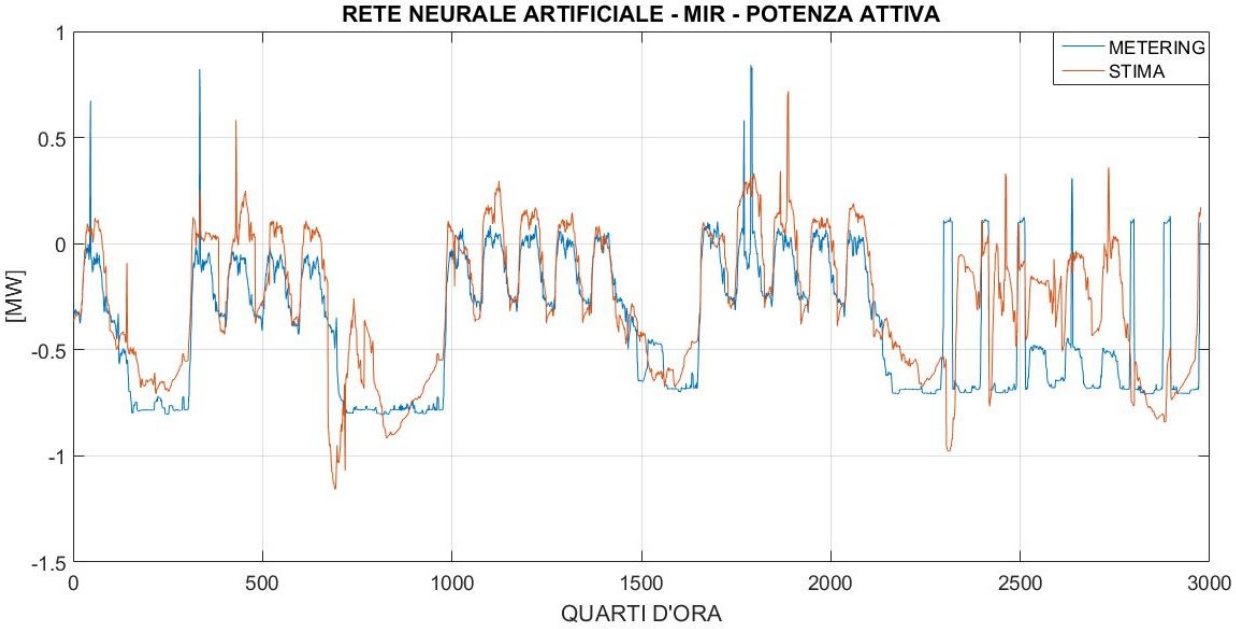
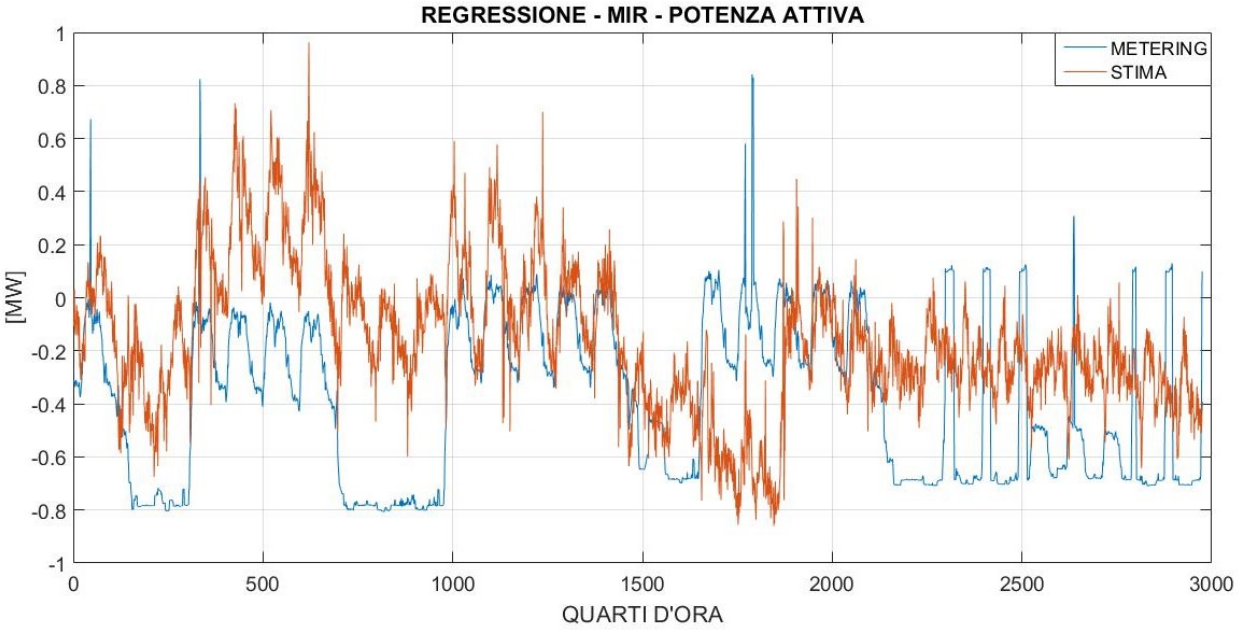


Figura A.6. Risultati della stima di potenza attiva e reattiva per l'utente MIF nel mese di dicembre 2017.

MIR


RETE NEURALE ARTIFICIALE - MIR - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA


REGRESSIONE - MIR - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA

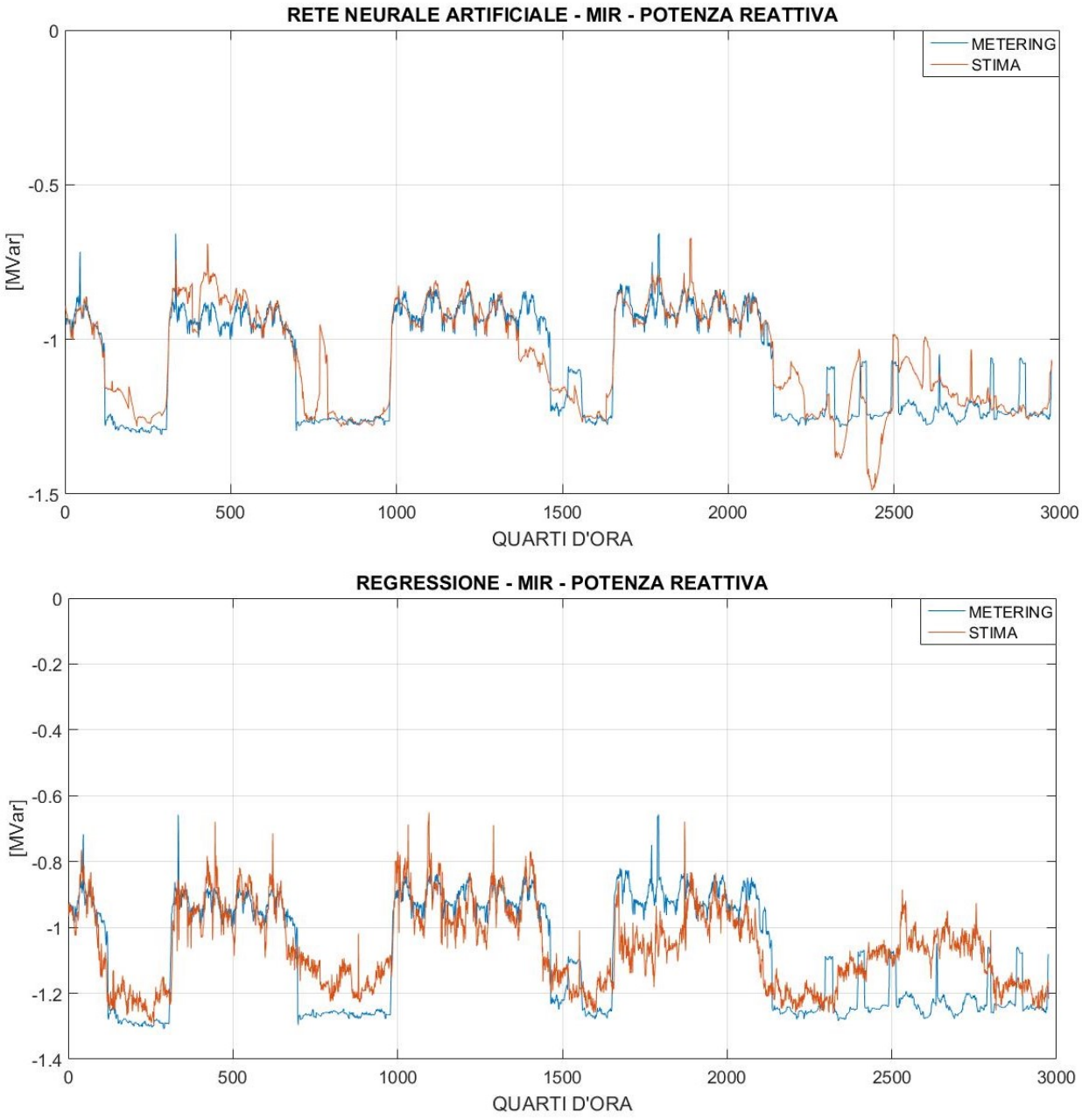
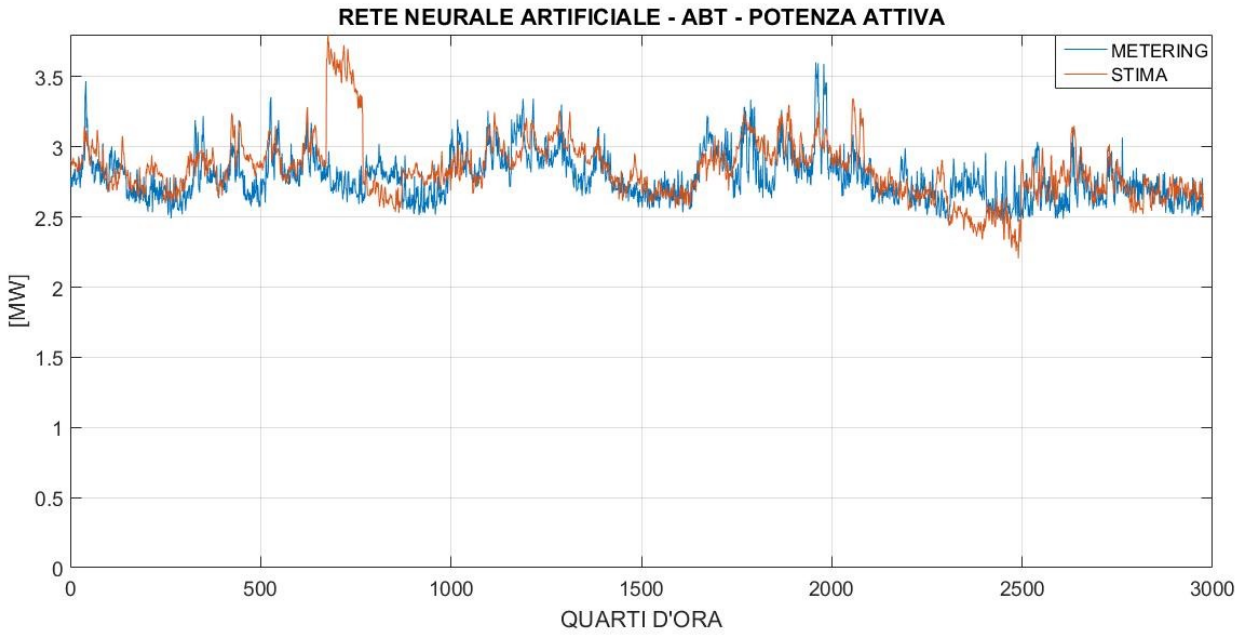
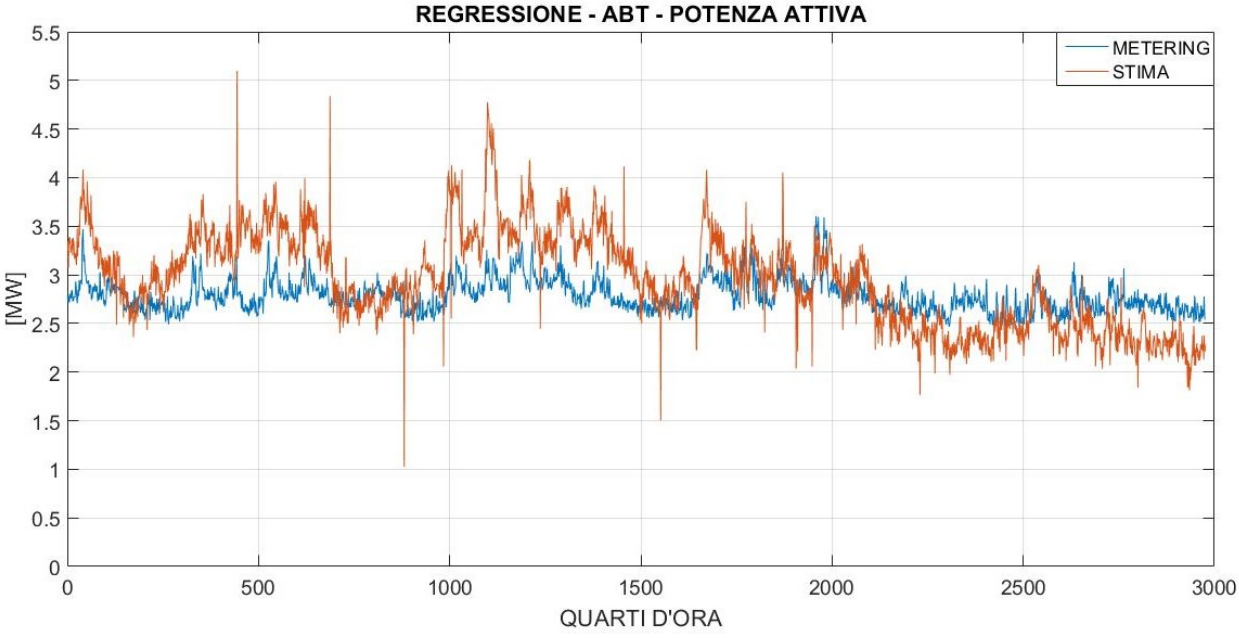


Figura A.7. Risultati della stima di potenza attiva e reattiva per l'utente MIR nel mese di dicembre 2017.

ABT


RETE NEURALE ARTIFICIALE - ABT - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA
0
0.5
1
1.5
2
2.5
3
3.5
0
500
1000
1500
2000
2500
3000


REGRESSIONE - ABT - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
5
5.5
0
500
1000
1500
2000
2500
3000

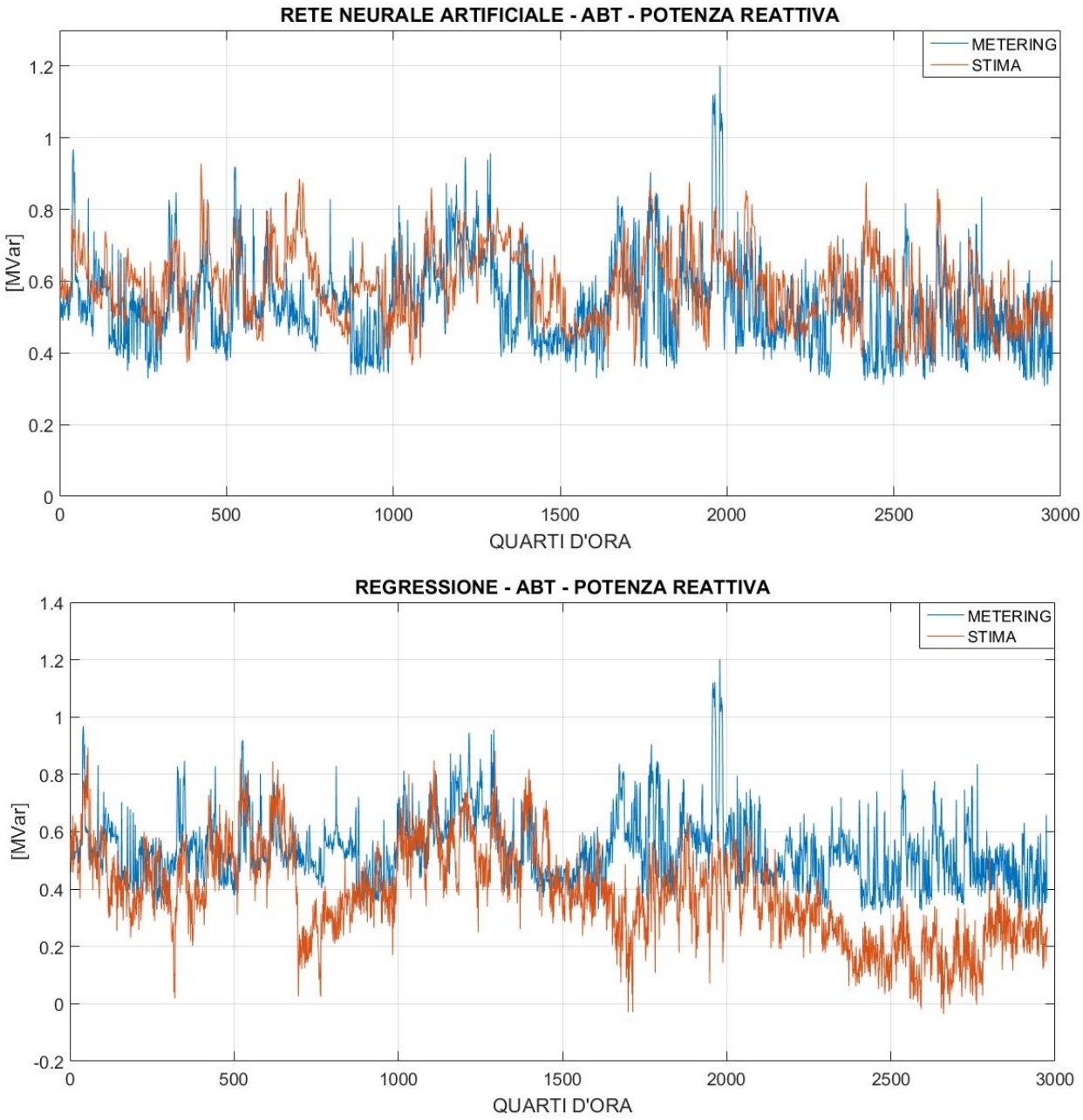
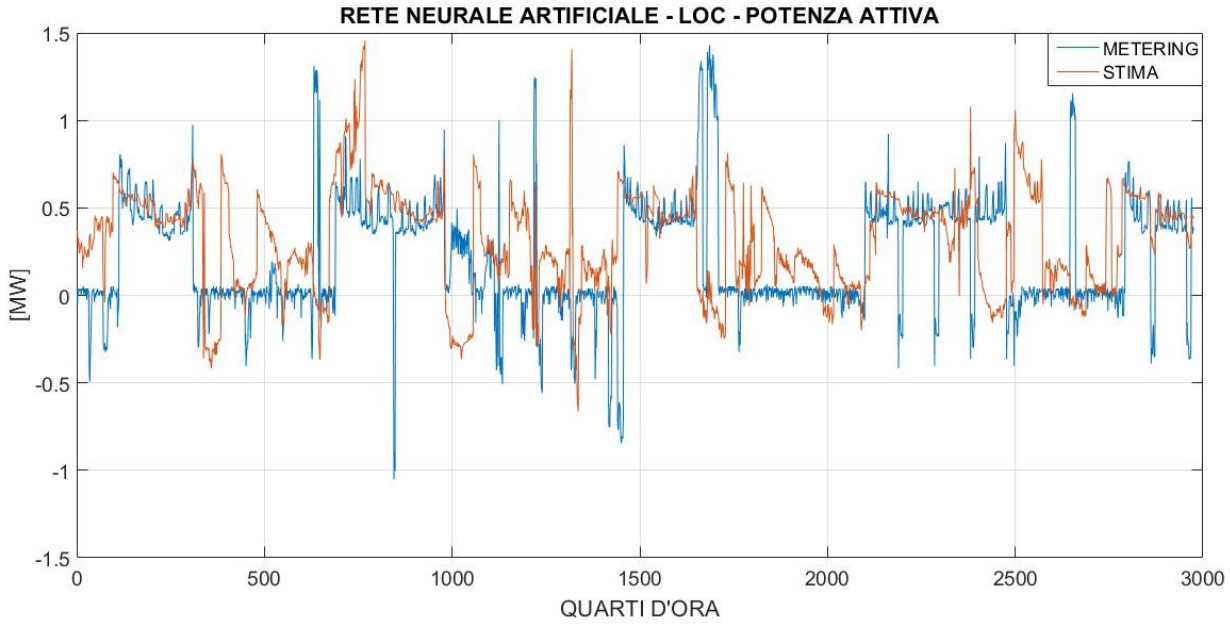
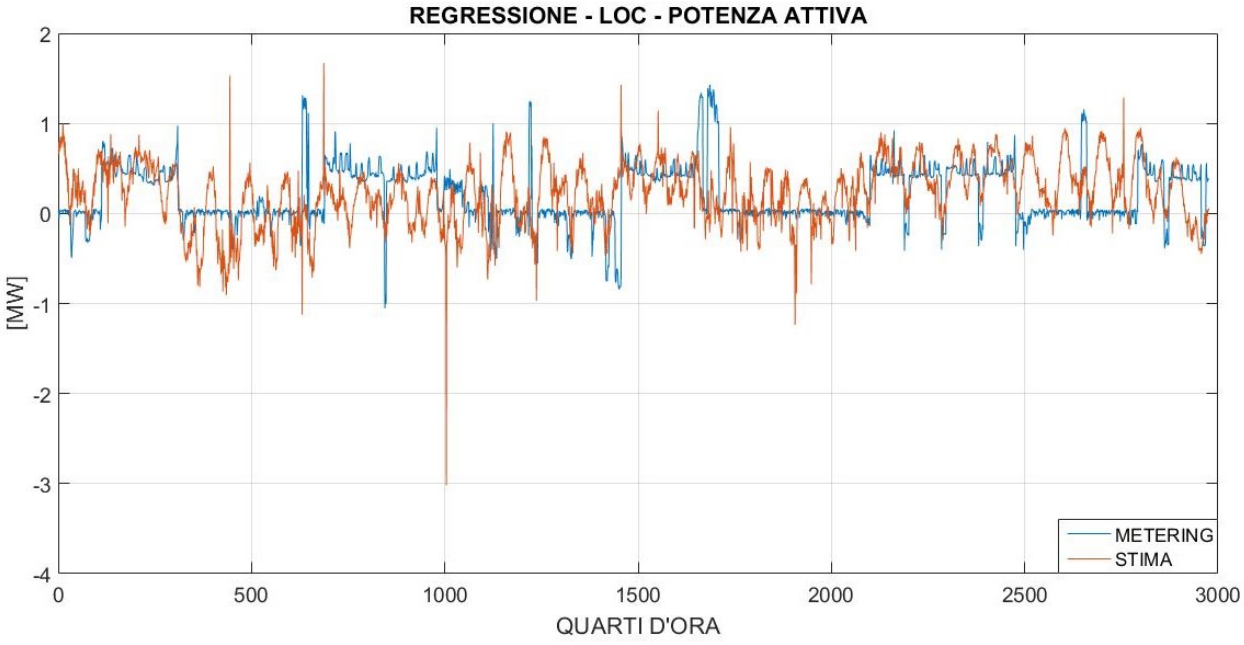


Figura A.8. Risultati della stima di potenza attiva e reattiva per l'utente ABT nel mese di dicembre 2017.

LOC


RETE NEURALE ARTIFICIALE - LOC - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA


REGRESSIONE - LOC - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA

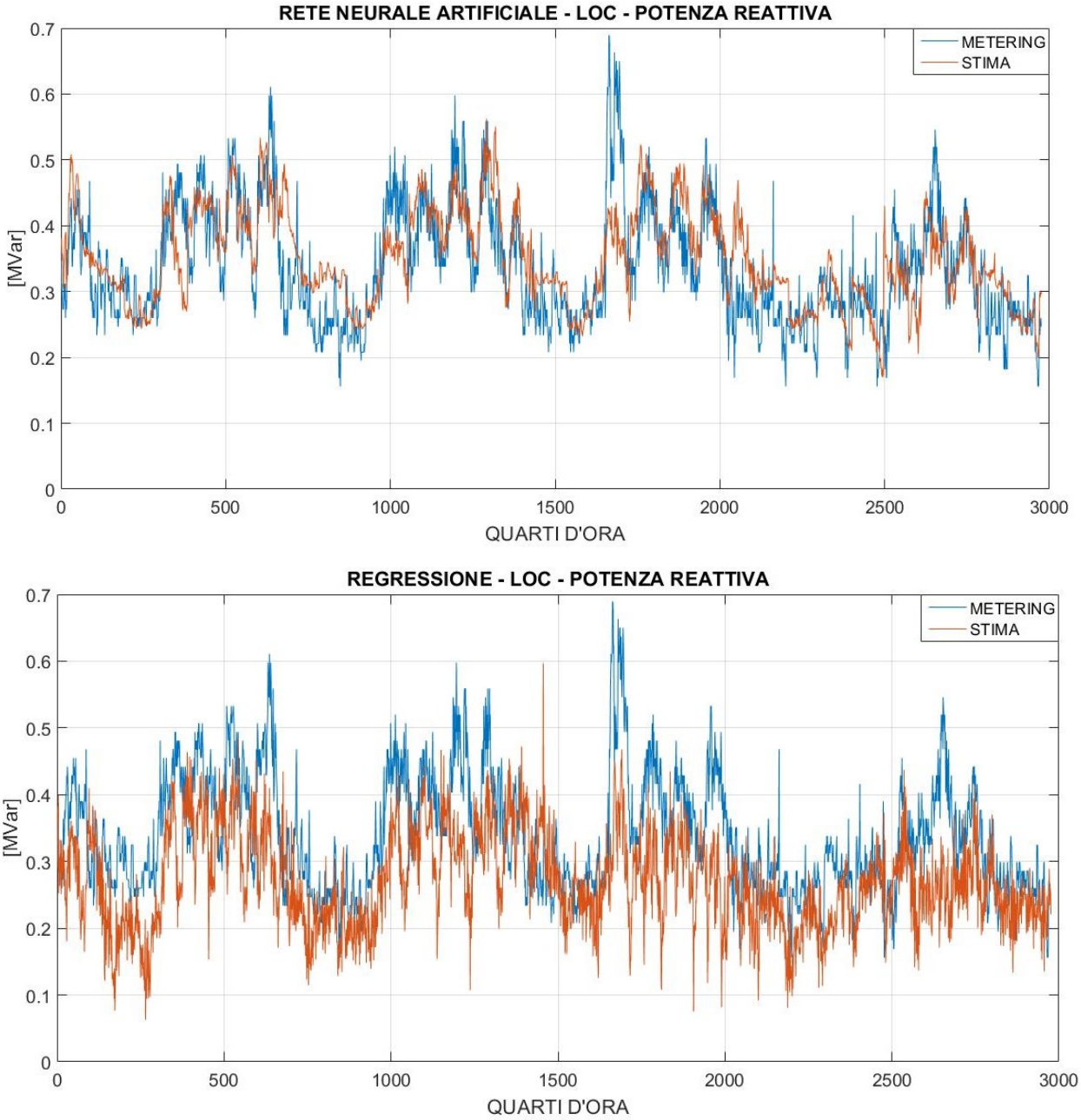
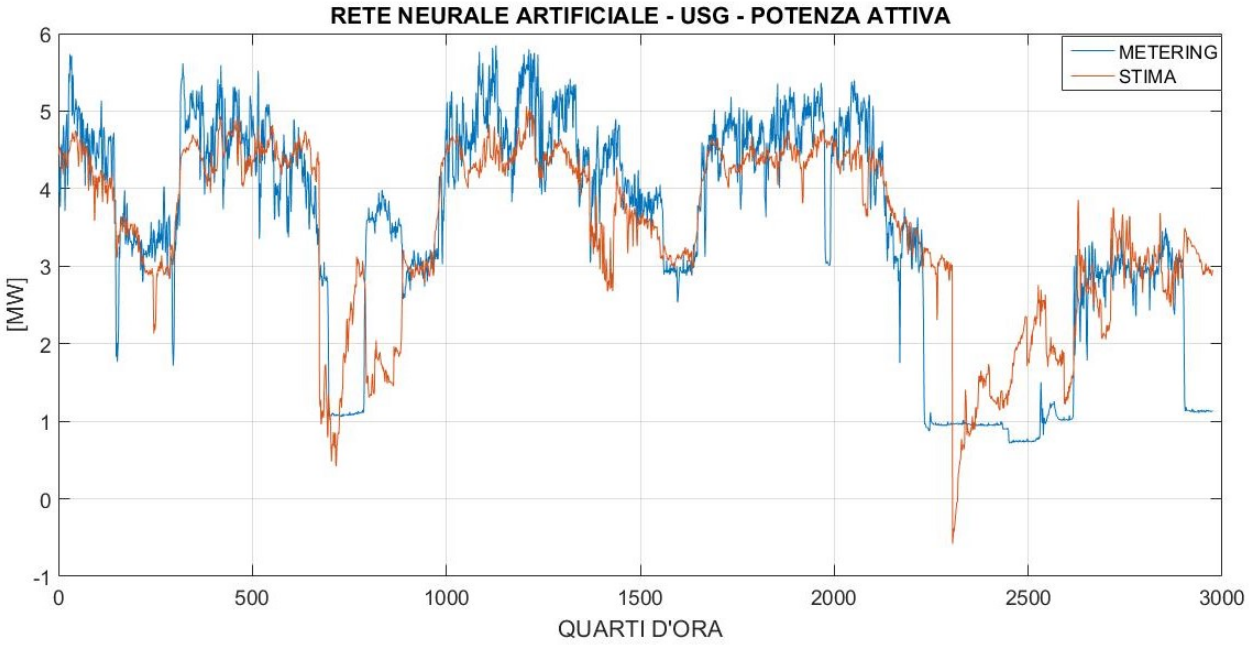
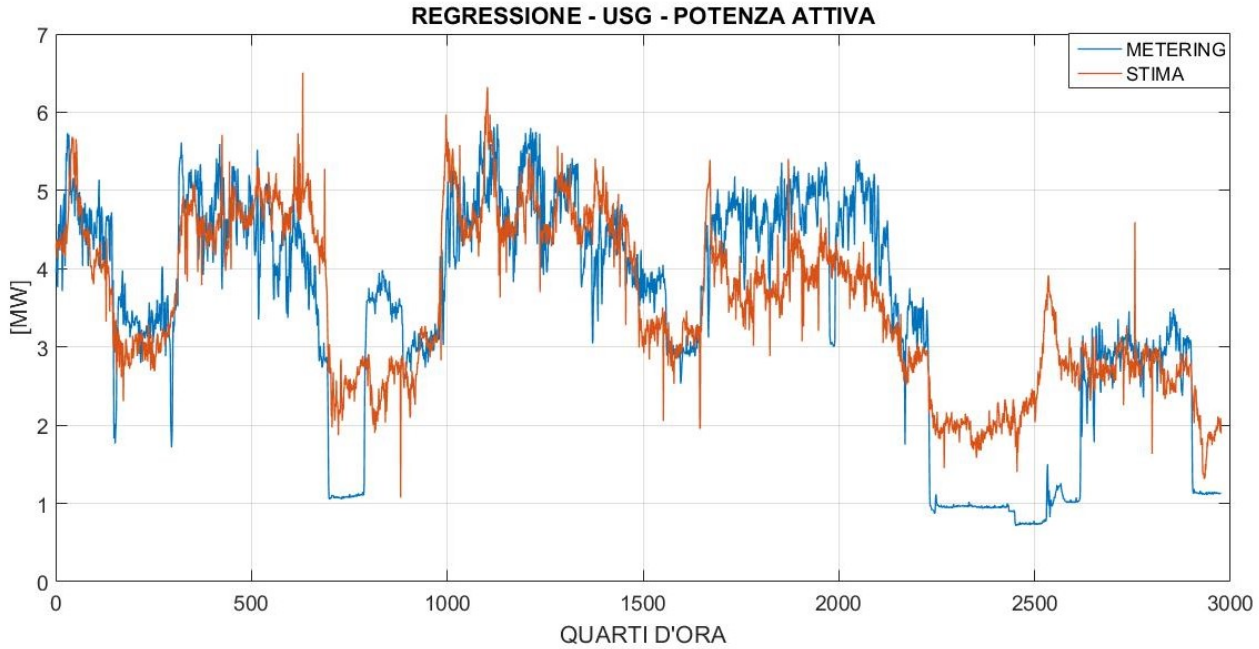


Figura A.9. Risultati della stima di potenza attiva e reattiva per l'utente LOC nel mese di dicembre 2017.

USG


RETE NEURALE ARTIFICIALE - USG - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA


REGRESSIONE - USG - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA

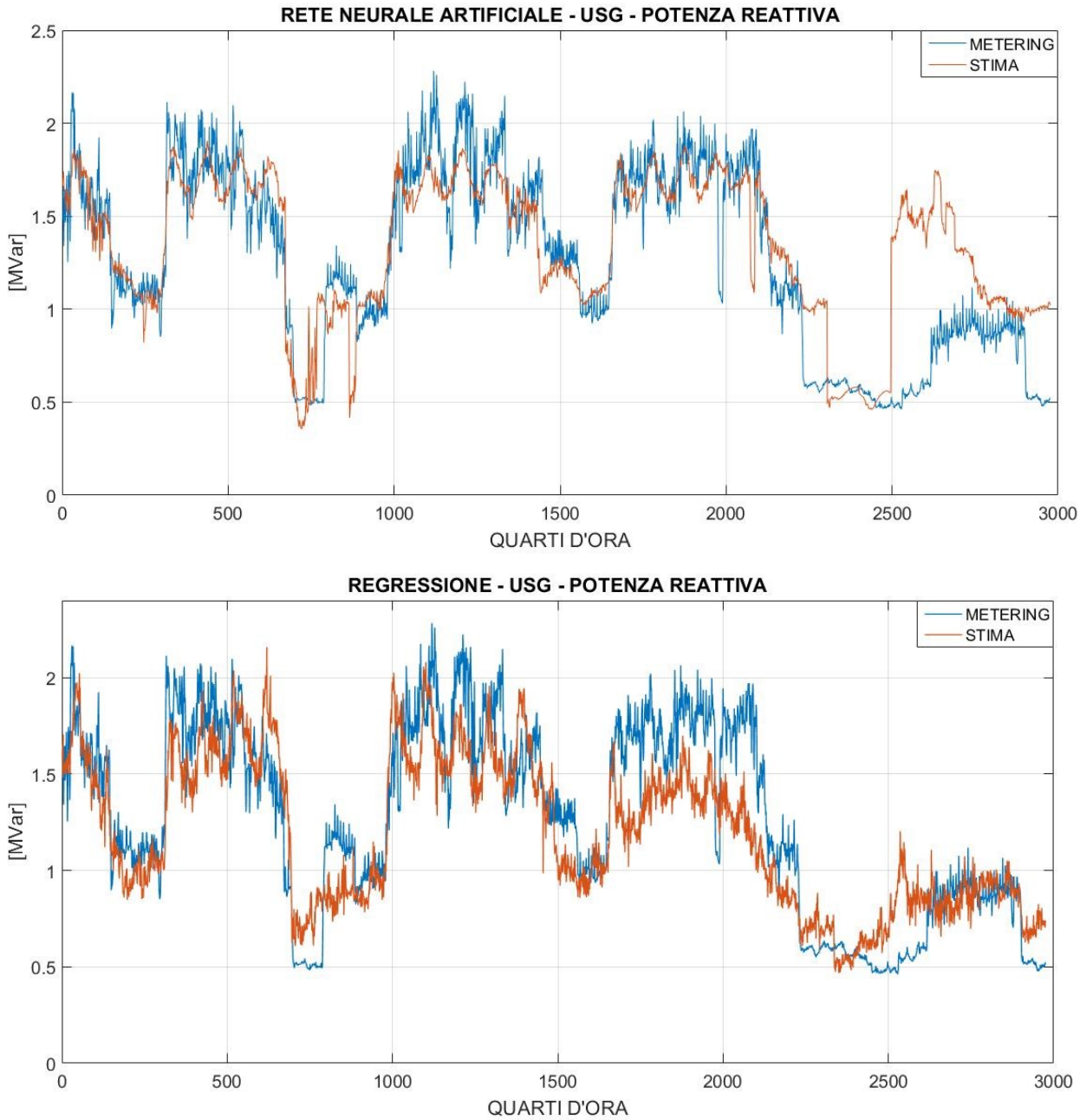
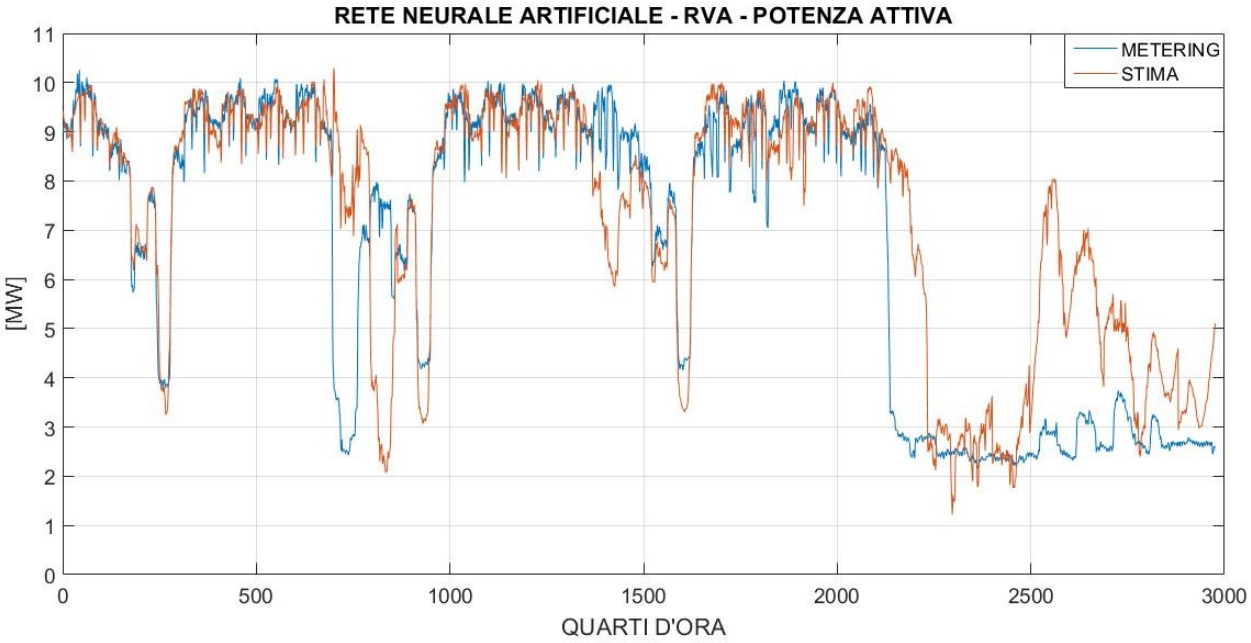
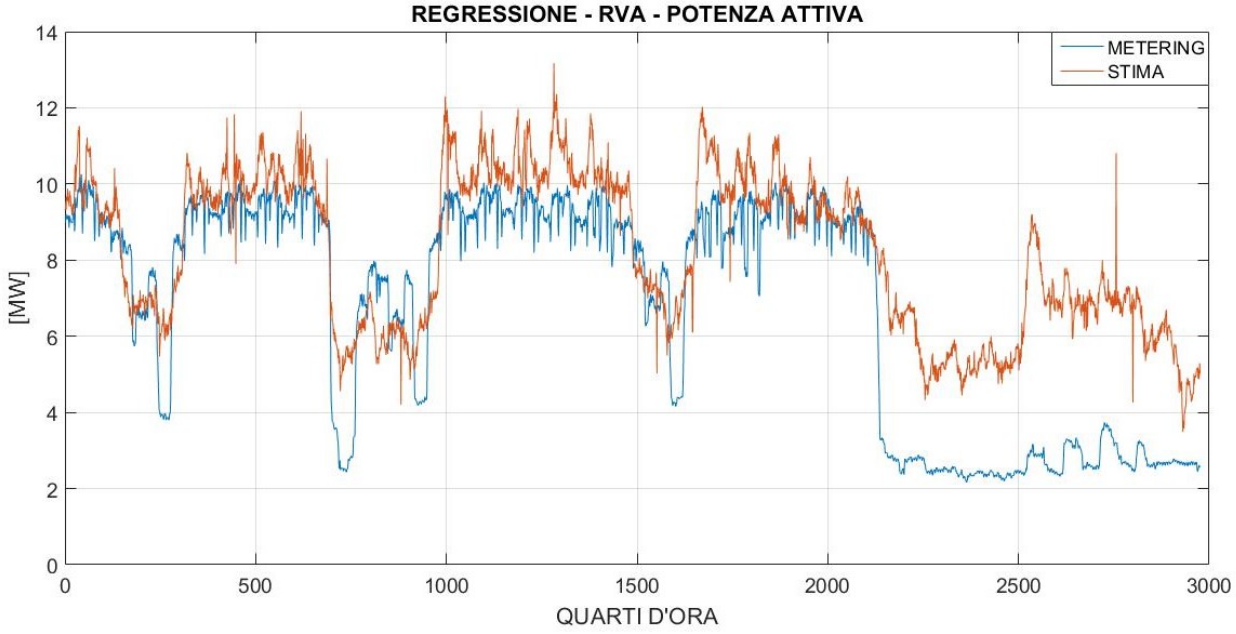


Figura A.10. Risultati della stima di potenza attiva e reattiva per l'utente USG nel mese di dicembre 2017.

RVA


RETE NEURALE ARTIFICIALE - RVA - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA
0
500
1000
1500
2000
2500
3000
0
1
2
3
4
5
6
7
8
9
10
11


REGRESSIONE - RVA - POTENZA ATTIVA
METERING
STIMA
[MW]
QUARTI D'ORA
0
500
1000
1500
2000
2500
3000
0
2
4
6
8
10
12
14

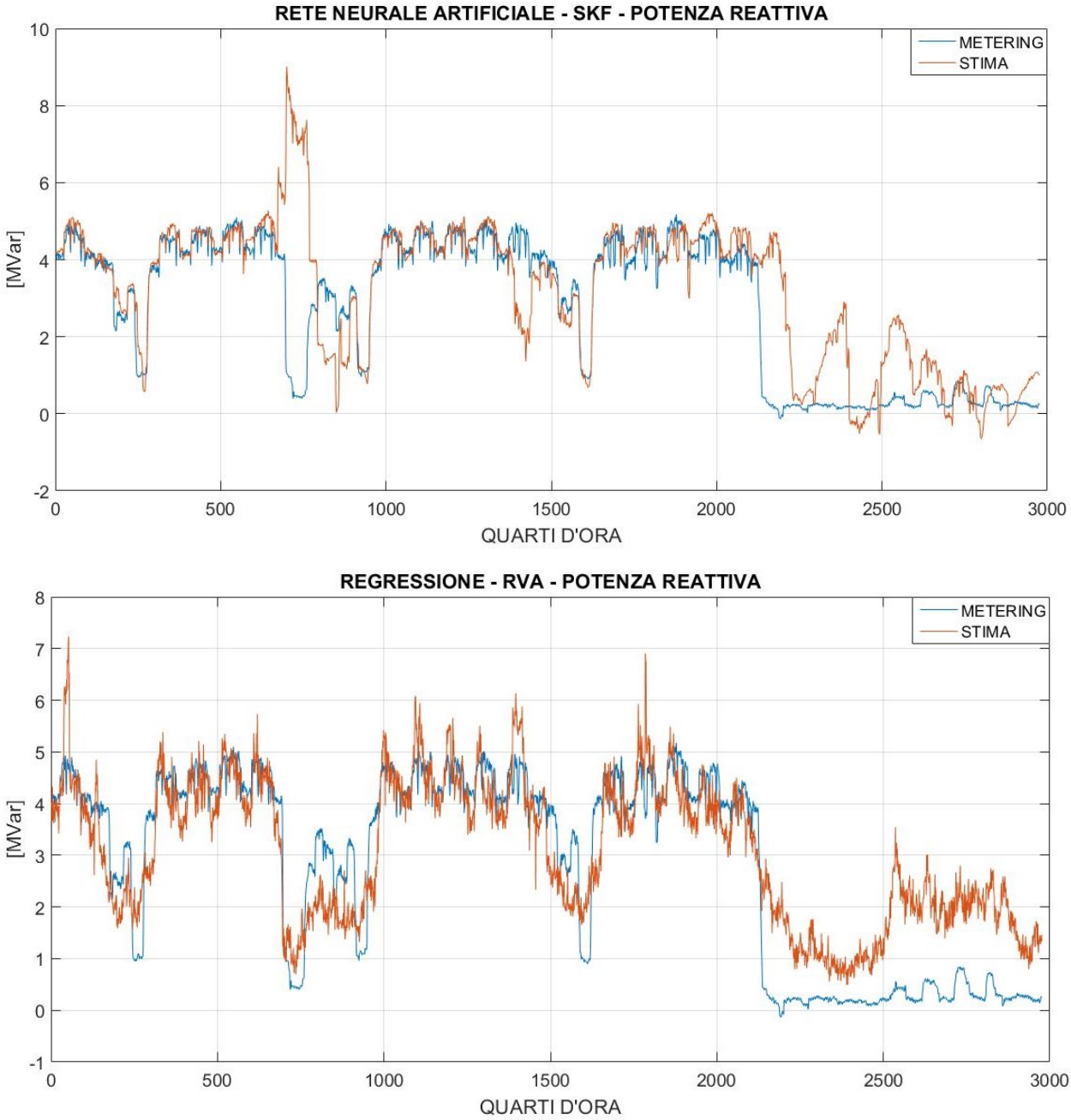


Figura A.11. Risultati della stima di potenza attiva e reattiva per l'utente RVA nel mese di dicembre 2017.

Appendice B

Confronto tra stima proposta e stima attuale - Risultati completi

In seguito, sono riportati i risultati della macchina di prova confrontate con i risultati dei metodi di stima attualmente adottati dall'azienda Terna, nel periodo tra il 18 e il 22 gennaio 2019.

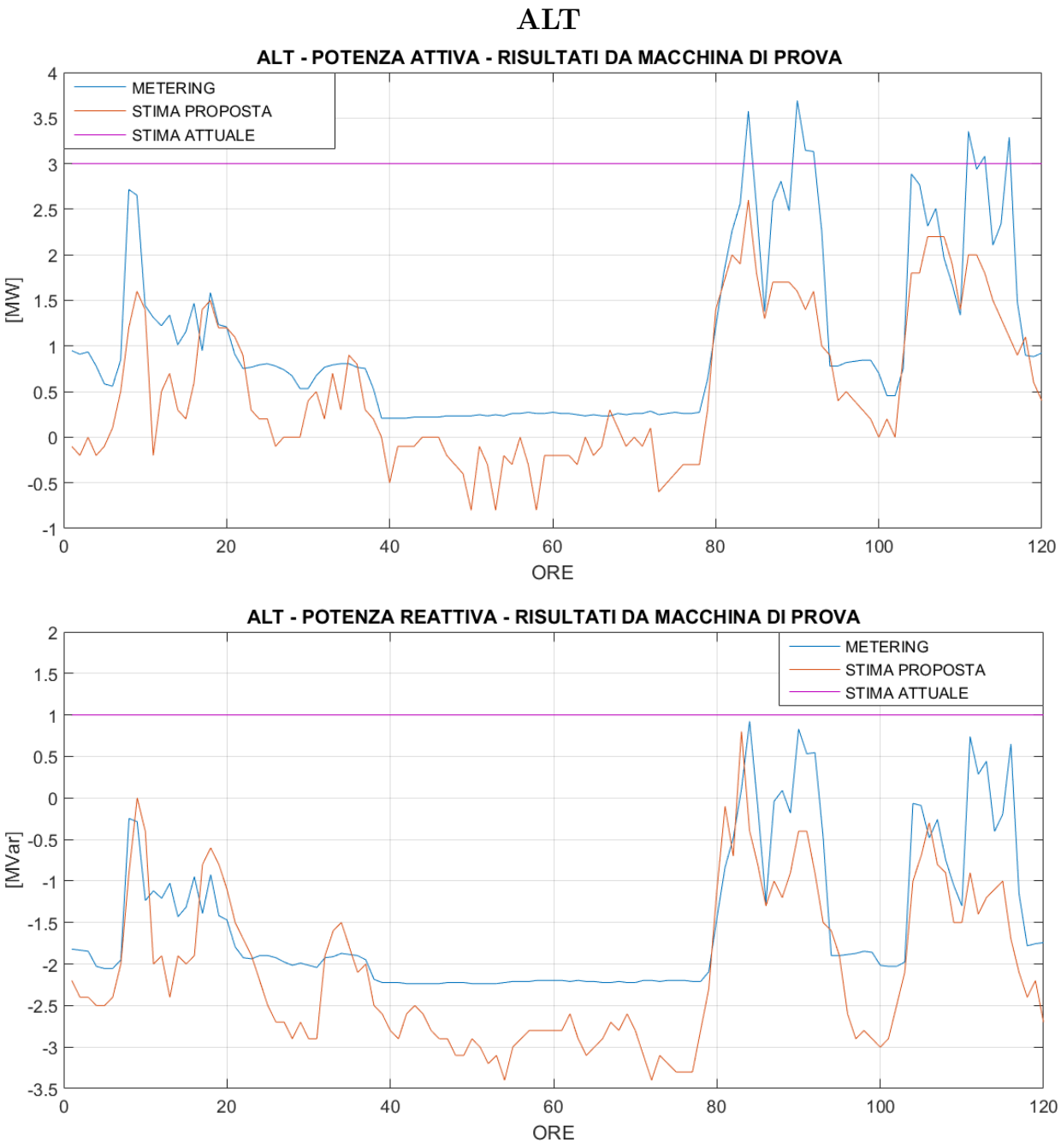


Figura B.1. Risultati per il carico ALT, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

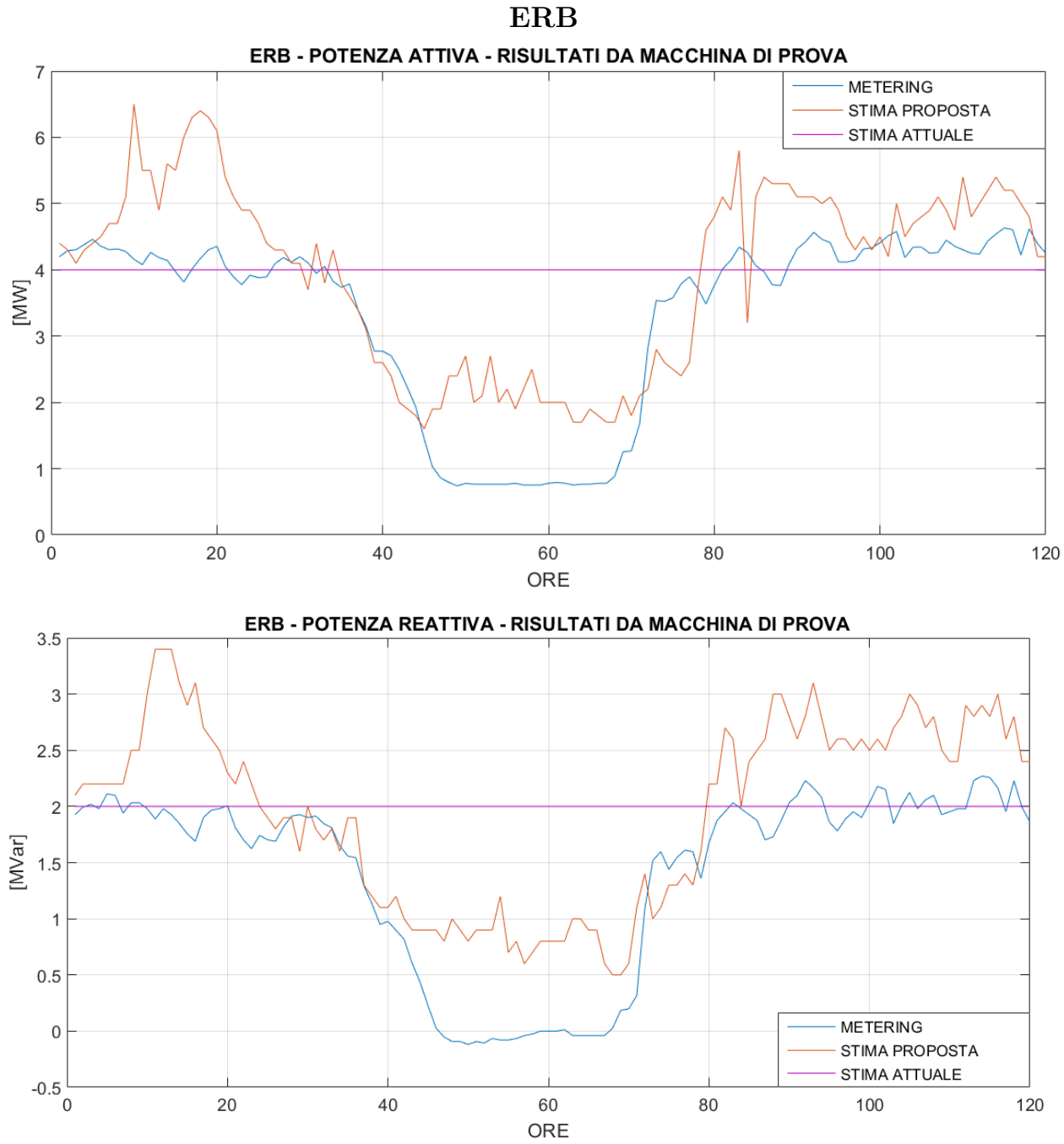


Figura B.2. Risultati per il carico ERB, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

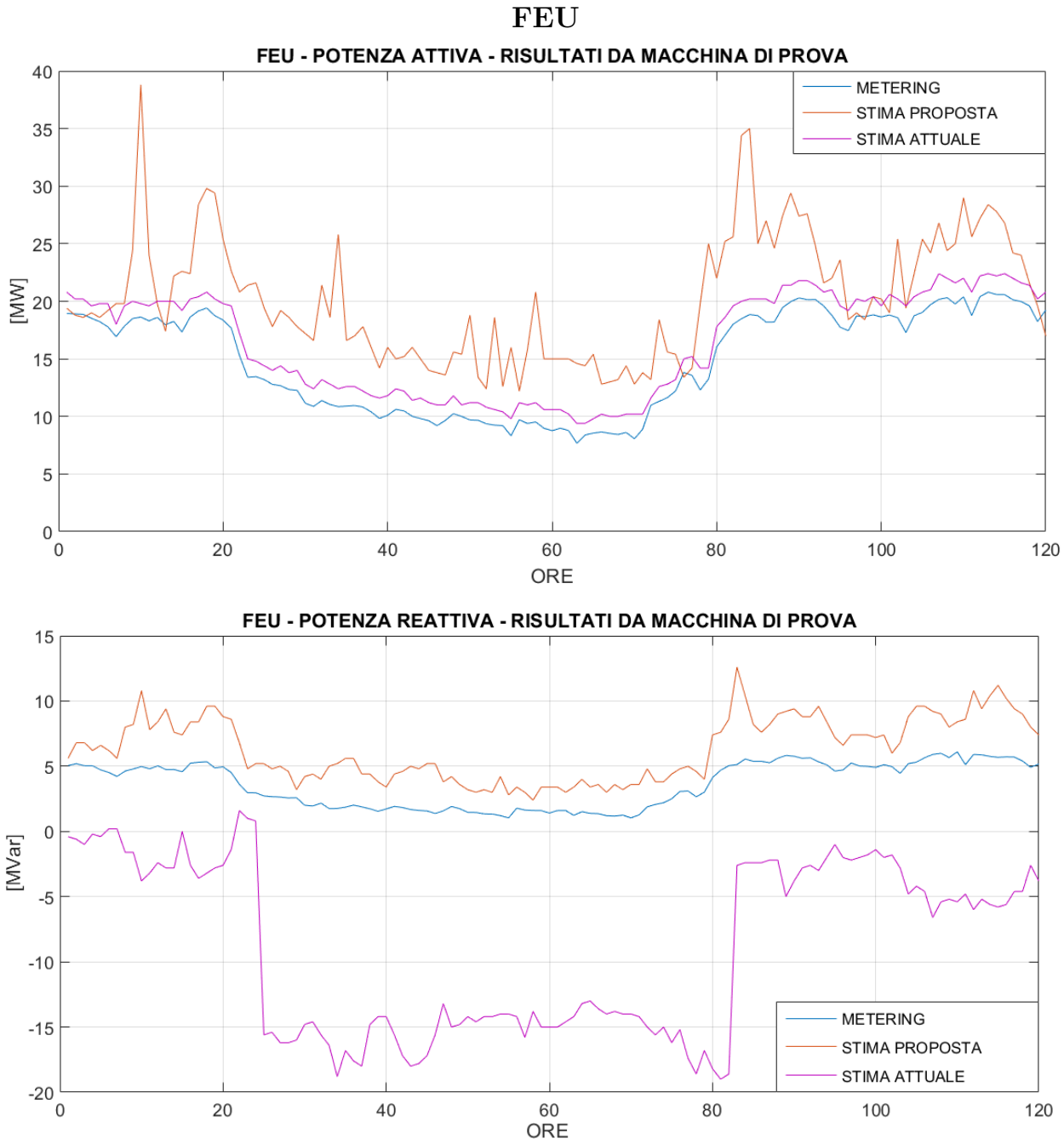


Figura B.3. Risultati per il carico FEU, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

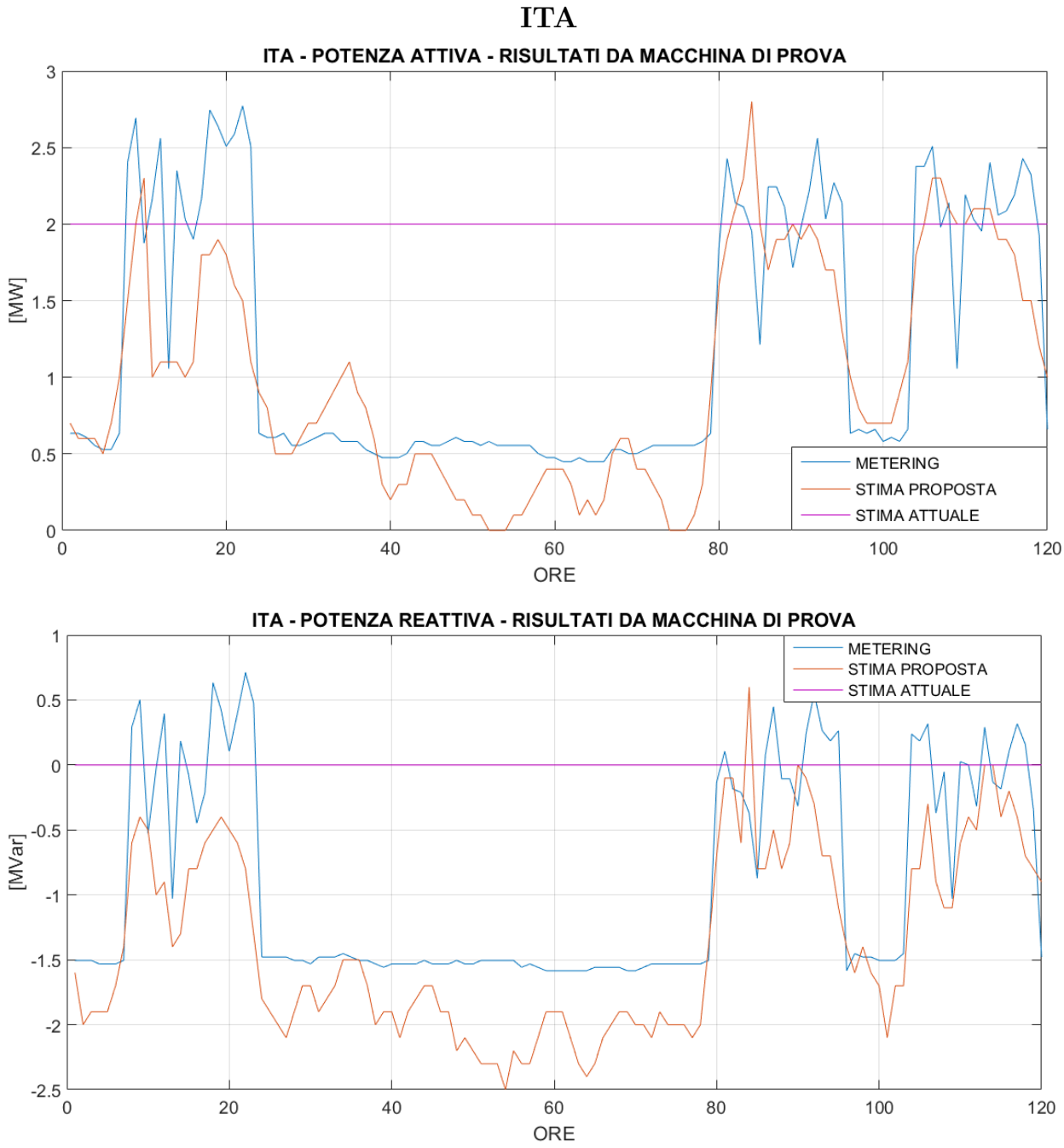


Figura B.4. Risultati per il carico ITA, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

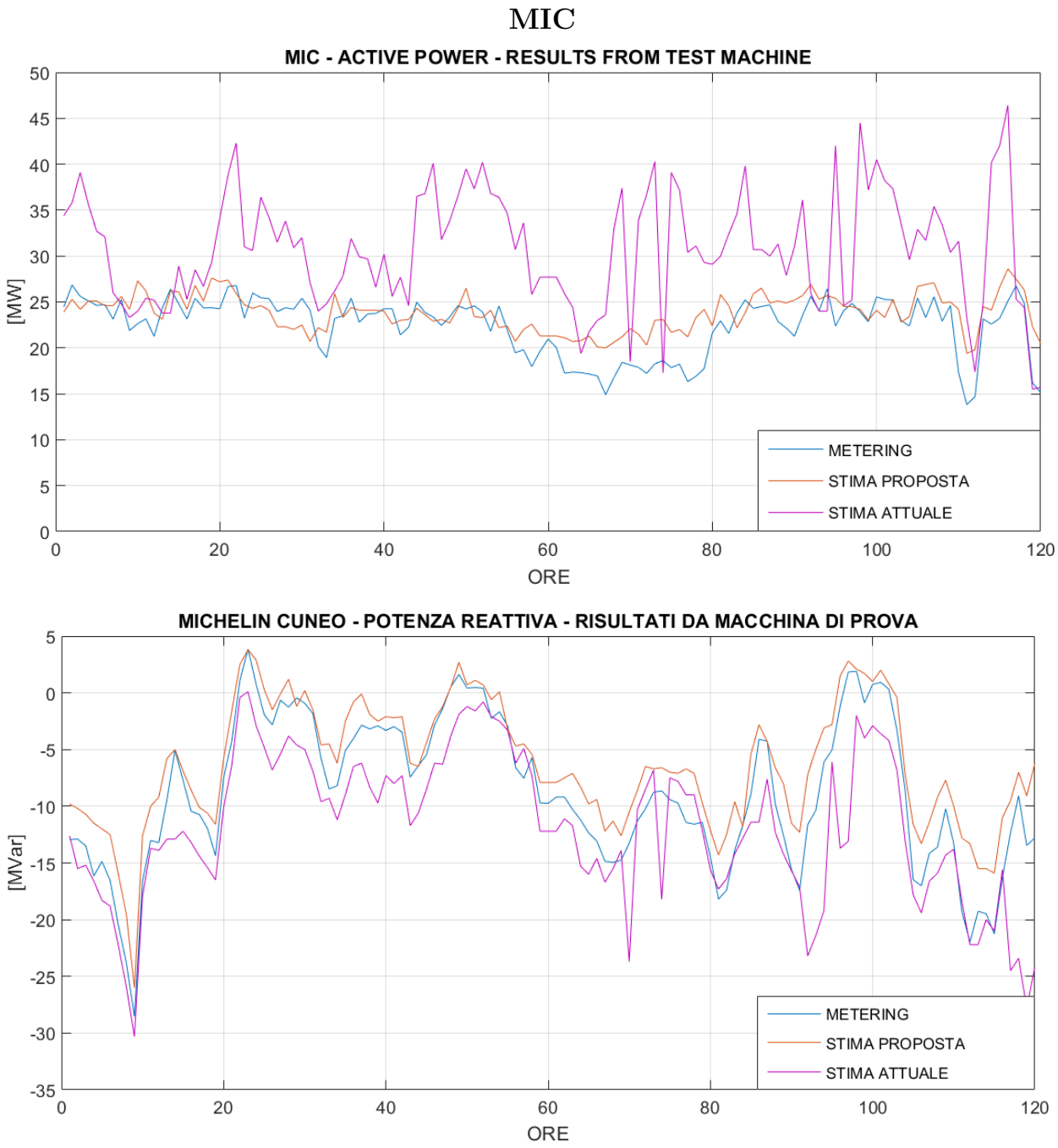


Figura B.5. Risultati per il carico MIC, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

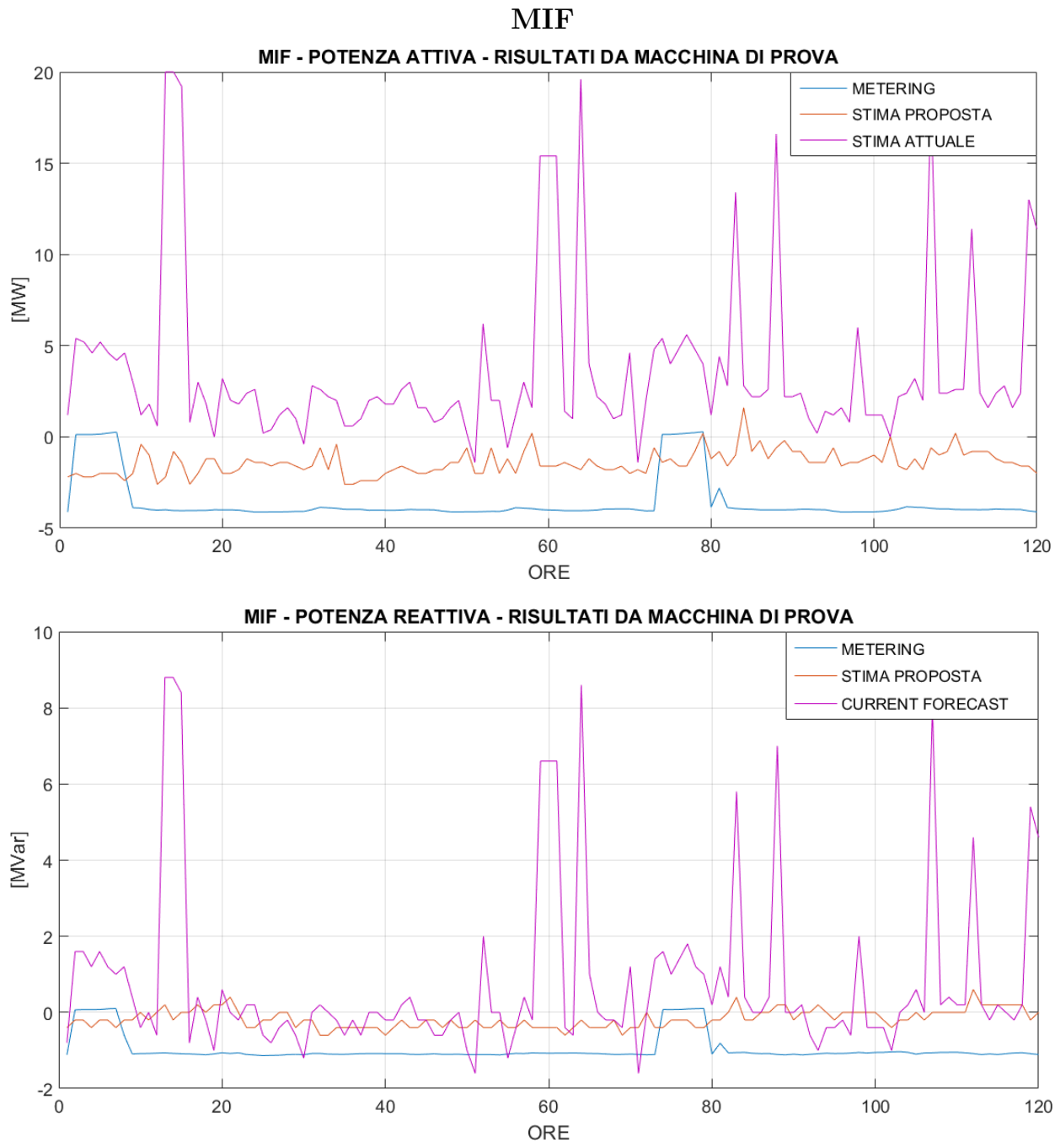


Figura B.6. Risultati per il carico MIF, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

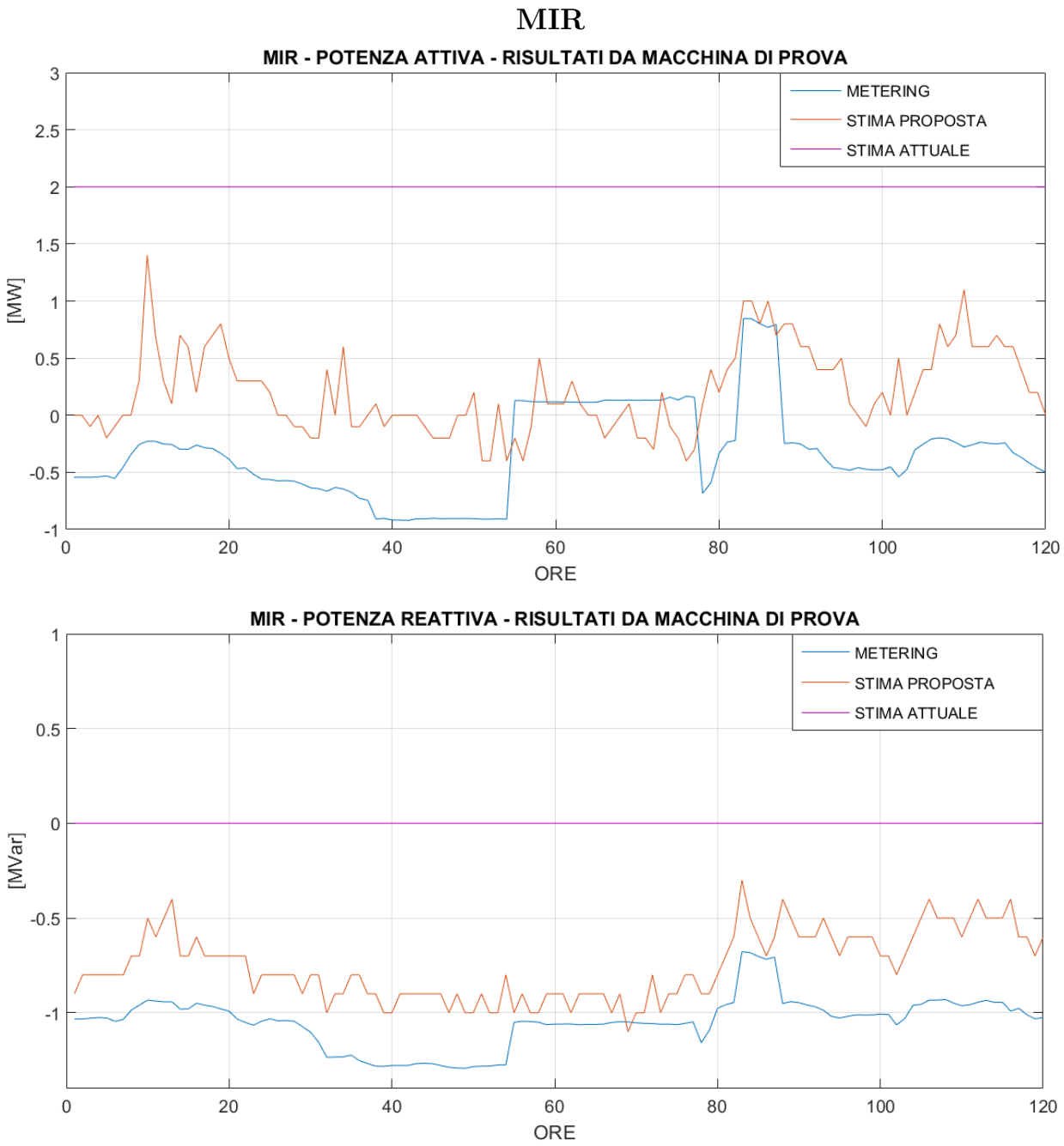


Figura B.7. Risultati per il carico MIR, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

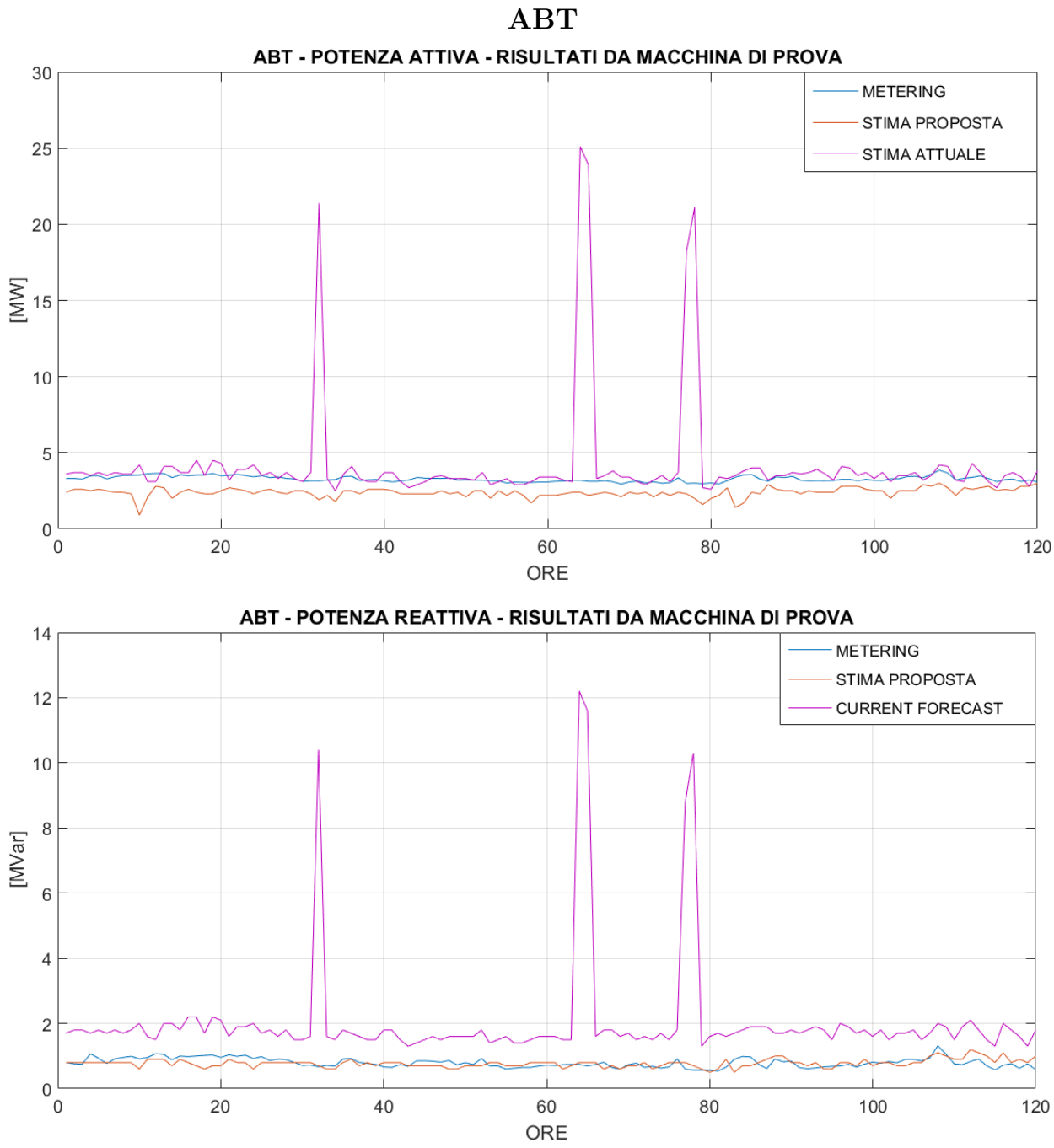


Figura B.8. Risultati per il carico ABT, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

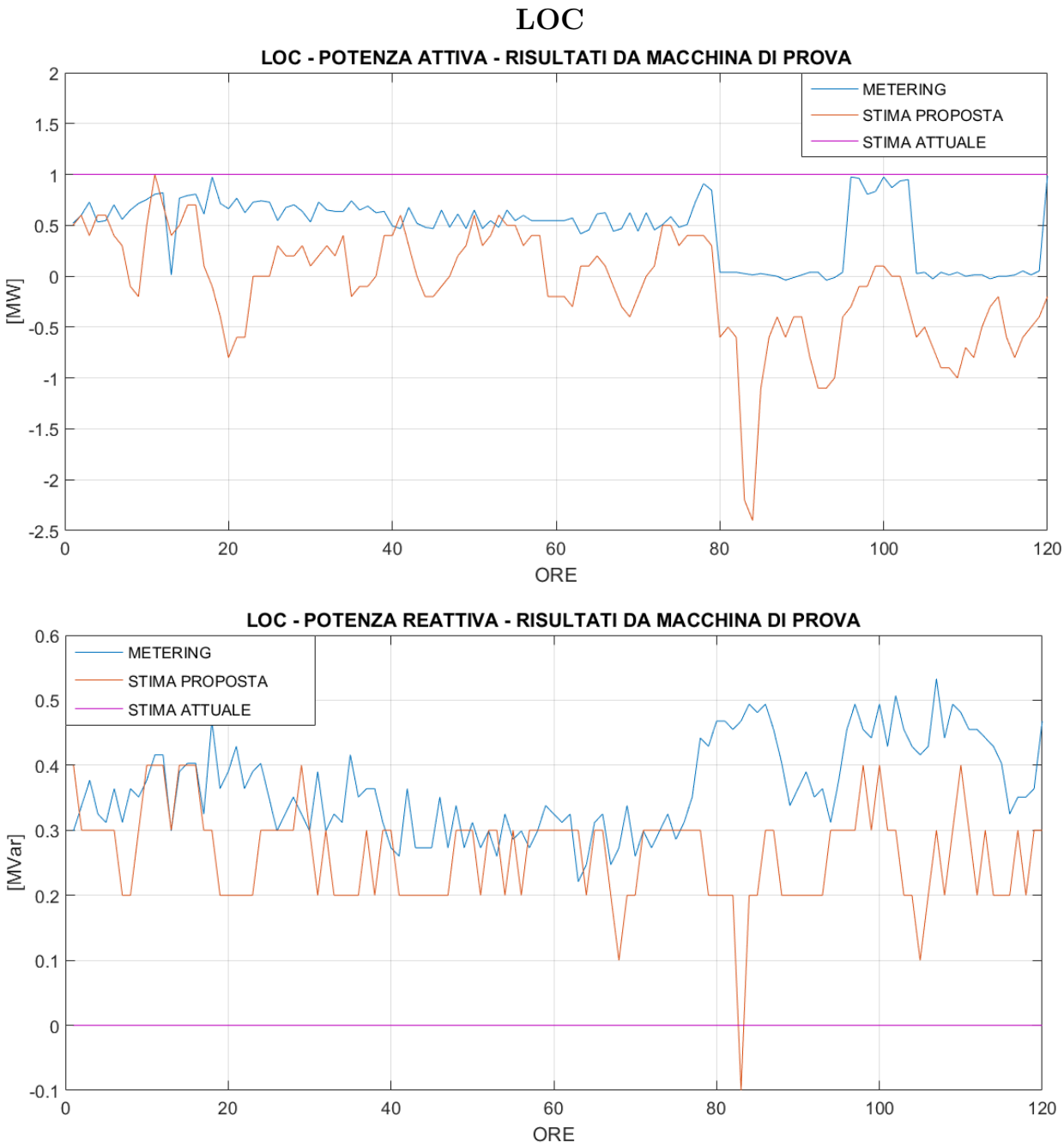


Figura B.9. Risultati per il carico LOC, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

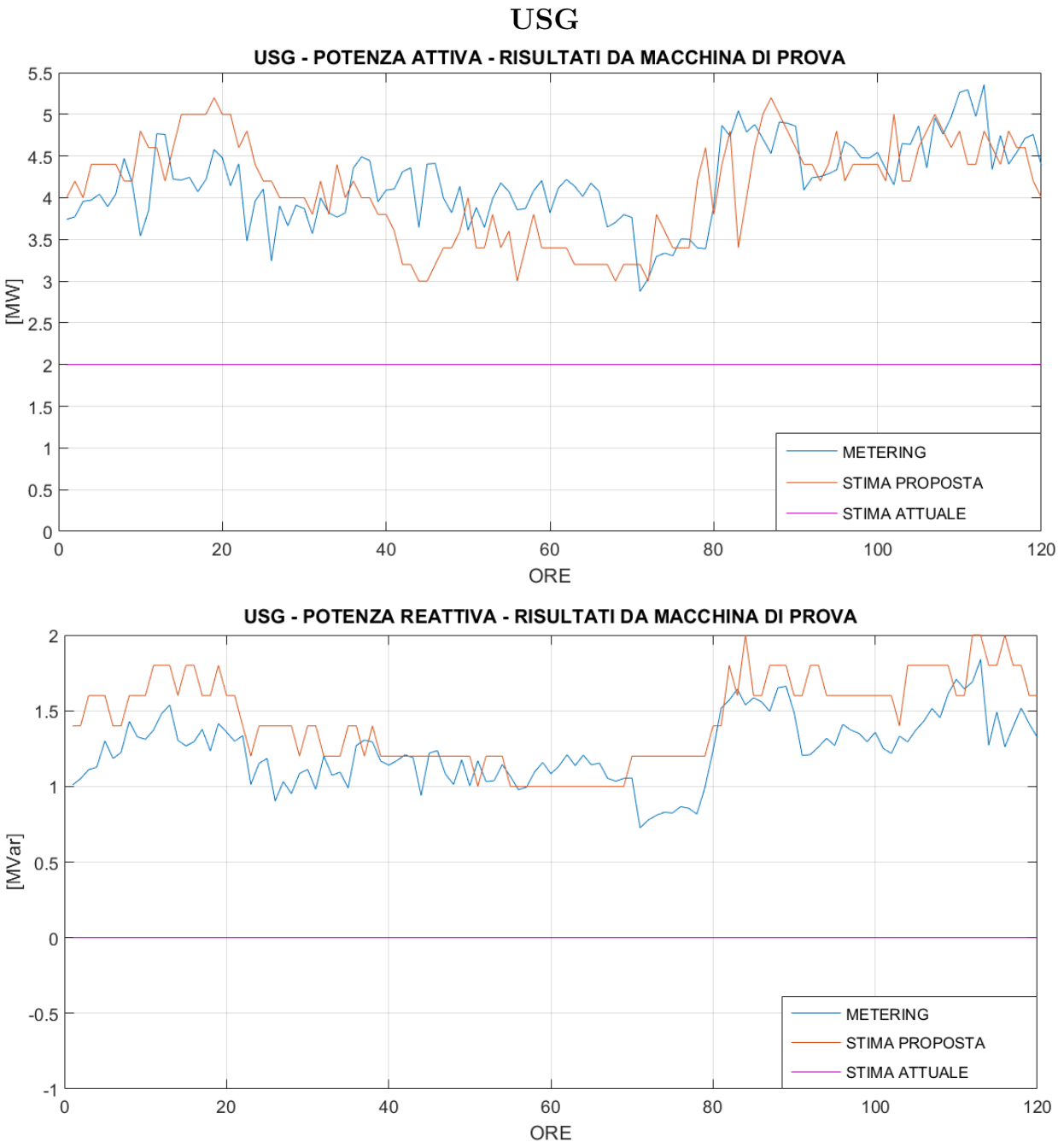


Figura B.10. Risultati per il carico USG, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

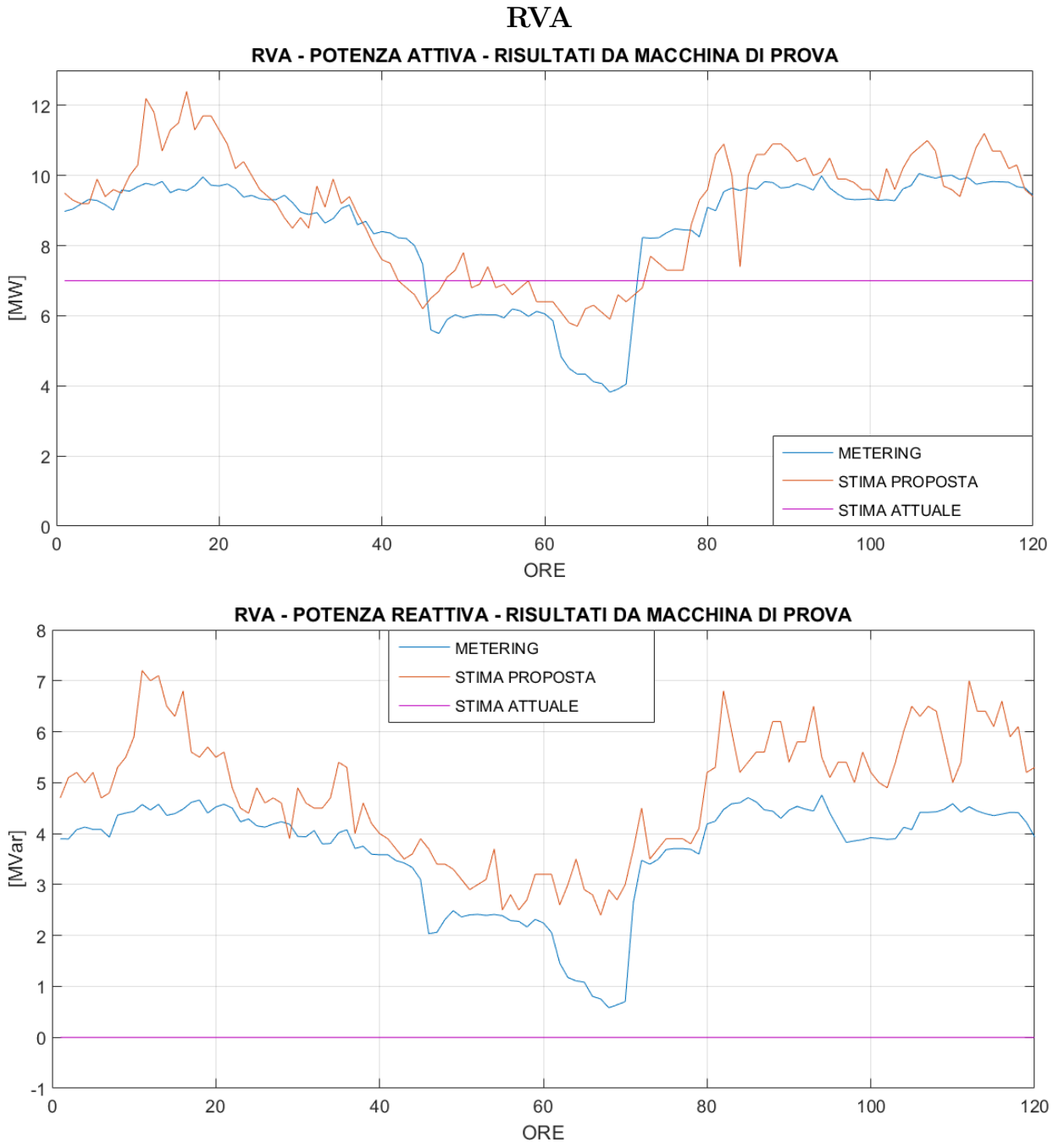


Figura B.11. Risultati per il carico RVA, relativi al periodo tra il 18 e il 22 gennaio 2019. Vengono confrontati i risultati ottenuti dalla macchina di prova, nel quale è implementato il metodo proposto, i risultati della stima attualmente in utilizzo e il valore reale di consumo.

Bibliografia

[1] https://www00.unibg.it/dati/corsi/64023/44086-regressione_ lineare.pdf.

[2] http://statweb.stanford.edu/ tibs/sta305files/ Rudyregularization.pdf.

[3] "*Introduction to Linear Regression Analysis*", Douglas C. Montgomery, Elizabeth A. Peck, G. Geoffrey Vining, (2012, Wiley).

[4] https://areeweb.polito.it/didattica/gcia/Apprendimento_ Mimetico/Lucidi/Lucidi_ Reti_ Neurali1.pdf.

[5] B. Krose, P. van der Smagt, "*An introduction to Neural Networks*", University of Amsterdam, 1996.

[6] N. Babadi, S. Nouri, S. Khalaj, "*Challenges and opportunities of the integration of IoT and smart grid in Iran transmission power system*", 2017 Smart Grid Conference (SGC).

[7] Z. Luhua, Y. Zhonglin, W. Sitong, Y. Ruiming, Z. Hui, Y. Qingduo, "*Effects of Advanced Metering Infrastructure (AMI) on relations of Power Supply and Application in smart grid*", CICED 2010 Proceedings.

[8] J. Zhang, Z. Chen, X. Yang, K. Chen, K. Li, "*Ponder over Advanced Metering Infrastructure and Future Power Grid*", 2010 Asia-Pacific Power and Energy Engineering Conference.

[9] E. Bompard, E. Carpaneto, G. Chicco, R. Napoli, F. Piglione, "*Short-term load forecasting of a small electric utility by a fast learning RBF neural network*", PMAPS 2000.

[10] C. Mosca, "*Metodi e modelli real-time per la stima dello stato nei sistemi di sub-trasmissione AT*", Tesi Magistrale Politecnico di Torino.

[11] MathWotks, *MATLAB documentation*, https://it.mathworks.com/help.

[12] A. Bracale, G. Carpinelli, P. De Falco, and T. Hong, "*Short-term industrial load forecasting: A case study in an Italian factory*", in Proc. IEEE Power Energy Soc. Innovative Smart Grid Technol. Eur., Turin, Italy, 2017, pp. 1–6.