

POLITECNICO DI TORINO

Corso di Laurea Magistrale in Ingegneria Energetica e Nucleare

Tesi di Laurea Magistrale

Analisi stocastica e previsione di serie temporali: il caso del prezzo dell'energia elettrica



Relatori

Prof. Tamea Stefania

Prof. Poggi Davide

Candidato

Tallone Andrea

A.A 2017/2018

INDICE

ABSTRACT	4
INTRODUZIONE	5
Capitolo1: RICERCA IN LETTERATURA.....	6
1.1. <i>Modelli multi agente</i>	<i>8</i>
1.2. <i>Modelli fondamentali</i>	<i>10</i>
1.3. <i>Modelli reduced-form.....</i>	<i>11</i>
1.4. <i>Modelli Statistici.....</i>	<i>12</i>
1.5. <i>Modelli di intelligenza computazionale.....</i>	<i>16</i>
1.6. <i>Modelli Ibridi.....</i>	<i>18</i>
1.7. <i>Risultati in letteratura.....</i>	<i>19</i>
Capitolo 2: ANALISI del MERCATO ELETTRICO ITALIANO.....	26
2.1. <i>STRUTTURA DEL MERCATO ELETTRICO ITALIANO</i>	<i>26</i>
2.2. <i>MGP: MERCATO DEL GIORNO PRIMA</i>	<i>32</i>
2.2.1. <i>Confronti.....</i>	<i>38</i>
2.2.1.2. <i>Confronto settimanale.....</i>	<i>39</i>
2.2.1.3. <i>Confronto mensile</i>	<i>42</i>
2.2.1.4. <i>Statistiche medie giornaliere</i>	<i>43</i>
2.2.1.5. <i>Statistiche medie mensili</i>	<i>52</i>
2.3. <i>MI: MERCATO INFRA-GIORNALIERO.....</i>	<i>56</i>
2.4. <i>CONFRONTI MI e MGP.....</i>	<i>57</i>
2.4.1. <i>Confronto zonale annuale</i>	<i>57</i>
2.4.2. <i>Confronto mensile</i>	<i>62</i>
2.5. <i>VARIABILI ENDOGENE e DATI CORRELATI</i>	<i>65</i>
2.5.1. <i>Potenza rinnovabile installata in Italia</i>	<i>65</i>
2.5.2. <i>Energia prodotta da fonti rinnovabili.....</i>	<i>67</i>
2.5.3. <i>Commodity energetiche per la produzione di energia elettrica.....</i>	<i>70</i>
Capitolo 3: METODO	73
3.1. <i>INTRODUZIONE.....</i>	<i>73</i>
3.2. <i>METODO.....</i>	<i>73</i>
3.2.1. <i>Ricostruzione dello spazio delle fasi.....</i>	<i>75</i>
3.2.2. <i>Ricerca di una relazione funzionale.....</i>	<i>76</i>
3.2.3. <i>Previsione.....</i>	<i>79</i>
3.3. <i>STIMA DELL'ERRORE</i>	<i>81</i>
3.4. <i>SERIE STORICA e AUTOCORRELAZIONE.....</i>	<i>84</i>

CAPITOLO 4: CALIBRAZIONE dei PARAMETRI.....	87
4.1. Set di calibrazione e orizzonte di previsione.....	87
4.2. Scelta dei parametri.....	88
4.2.1 Set estivo.....	92
Orizzonte di previsione a 1 ora.....	92
Orizzonte di previsione a 24 ore.....	97
4.2.2 Set Invernale.....	102
Orizzonte di previsione ad 1 ora.....	102
Orizzonte di previsione a 24 ore.....	106
4.2.3 Confronto a pari orizzonte di previsione e scelta della terna.....	111
Orizzonte di previsione ad 1 ora.....	111
Orizzonte di previsione a 24 ore.....	114
CAPITOLO 5: RISULTATI.....	118
5.1. Set di verifica.....	118
5.2. Orizzonte di previsione ad 1 ora.....	119
5.3. Orizzonte di previsione a 24 ore.....	121
5.4. Riscontro grafico.....	123
5.5. Confronto con lo stato dell'arte.....	125
Capitolo 6 : CONCLUSIONI.....	127
Bibliografia	129
 APPENDICE A – Tabelle capitolo 2	132
APPENDICE B – Diagrammi di calibrazione.....	136

ABSTRACT

In a deregulated market context the electricity price forecasting (EPF) play an essential role. Together with the load prediction there is a strong interest of the companies to forecast them better than others will do, is a matter of profit maximization and planning a proper bid strategy. This thesis aims to provide an overview about the Italian market and proposes a forecasting tool. Firstly an analysis regarding the Italian market is provided, the stock exchange markets price profiles and the influence of endogenous variables. The methodology is able to adapt itself to the complex characteristics and non-linearities involved, it is a model based on the local approximation of dynamic in the phase space constructed starting by the hystorical series based on the “*nearest neighbor*” concept. The results regards the forecasting ability at 1 and 24 hours, isolating the influence of the calibration periods, considering the historical price series as the only input variable. The comparison between obtained result and the state of the art is challenging and the possibilities of improving are good.

La previsione dei prezzi dell'energia elettrica rappresenta un tassello fondamentale per mercati deregolamentati. La competizione fra i soggetti che ne fanno parte rende essenziale prevedere le quantità e il prezzo al fine di massimizzare i profitti e pianificare le strategie di offerta. Questa tesi fornisce una panoramica sul mercato Italiano dell'energia elettrica e propone uno strumento di previsione dei prezzi. In primo luogo si analizzano la struttura e le caratteristiche del mercato Italiano, i profili di prezzo derivati dai mercati di borsa e l'influenza delle variabili endogene. Il metodo di previsione applicato ha la capacità di adattarsi alla complesse caratteristiche e non linearità, è un modello basato sull'approssimazione locale (*nearest neighbor*) della dinamica nello spazio delle fasi costruito a partire dalla serie storica. I risultati riguardano le capacità previsionali a 1 e 24 ore, isolando l'influenza dei periodi di calibrazione, con la sola serie storica in ingresso. Il riscontro dei risultanti con lo stato dell'arte è stimolante ed il margine di miglioramento concreto.

INTRODUZIONE

Questa tesi vuole studiare il mercato elettrico Italiano con l'intenzione di fornire un metodo per previsione del prezzo dell'energia, utilizzando come unico input lo storico del Prezzo Unitario Nazionale (PUN). Visto che l'argomento è ancora poco esplorato nella letteratura scientifica, soprattutto per il mercato elettrico Italiano dove le pubblicazioni sono quasi inesistenti si intende fornire un punto di partenza per sviluppi futuri. In Italia il problema è recente, il passaggio a mercato libero ha seguito un periodo di transizione che si è concluso nel 2004 con l'istituzione del Gestore dei Mercati Energetici S.p.A, ente che attualmente agisce come controparte unica sul mercato (a pronti) Italiano fungendo da acquirente unico. La borsa elettrica Italiana è quindi completamente gestita dal GME e qualunque produttore o consumatore può parteciparvi liberamente. Da qui nasce l'interesse e la necessità di prevedere quantità di energia e il prezzo a cui verranno valorizzate. Si può intuire che nel commerciale c'è un grande interesse economico per questi aspetti. Per un'azienda del settore è fondamentale prevedere il prezzo di equilibrio perché su di esso si basa l'intera strategia di offerte al fine di massimizzare l'utile, quindi, buona parte dei processi decisionali si basano sulle previsioni e possono concludersi con guadagni o perdite ingenti. Dall'altro lato anche il GME e Terna S.p.A devono tutelarsi, siccome la rete elettrica va mantenuta in costante equilibrio è importante che riescano a prevedere la situazione giorno per giorno in modo da approvvigionare risorse o prevedere stoccaggi anche a lungo termine. I dati che vengono utilizzati si sono reperiti dal sito web del gestore dei mercati energetici, disponibili online nella sezione dedicata. Si tratta dei prezzi orari delle sessioni del mercato del giorno prima a partire dal 1 Aprile 2004 fino a 31 Ottobre 2018, da febbraio ad ottobre 2018 servono solo per verificare le previsioni. Per il mercato infra-giornaliero si usano i dati delle prime cinque sessioni del 2016 che si trovano alla stessa sezione del sito GME.

Viene utilizzato un metodo che in passato è stato applicato con successo in ambito idrologico per prevedere le portate fluviali, tal metodo rientra nella categoria dei modelli statistici e utilizza accorgimenti semplici per adattarsi alle forti non linearità del problema. Il metodo prende il nome di "*Previsione non lineare*" e approssima la dinamica locale del sistema considerando solamente alcuni punti vicini e pesandone il contributo per ottenere la previsione. Ci sono comunque riferimenti ad altre classi modelli, ad esempio le reti neurali o Machine Learning che sono in grado di adattarsi all'evoluzione del sistema, nel primo capitolo viene fornita una descrizione dettagliata riguardo i modelli applicati in ambito di previsione dei prezzi dell'energia elettrica con rispettivi punti di forza e debolezza. Brevemente, la tesi comprende una ricerca approfondita sulla la struttura del mercato elettrico e selezione dei dati. Poi viene applicato il metodo e segue la calibrazione dei parametri, quest'ultima rappresenta il punto più ostico dell'analisi dove vengono utilizzati diversi periodi per studiare come incide la stagionalità sulla previsione, quindi si propone uno schema unico per arrivare all'ottimo. Infine si verifica la previsione su un insieme di valori che il metodo non ha mai visto prima e gli ultimi capitoli raccolgono i risultati inserendoli nel contesto dello stato dell'arte e relativi commenti conclusivi.

Capitolo1: RICERCA IN LETTERATURA

In letteratura si trovano pubblicazioni riguardo i diversi approcci per la previsione dei prezzi dell'energia (EPF: Electricity Price Forecasting) e le loro applicazioni; Weron R. in [1] propone un'ottima revisione dove organizza le pubblicazioni sull'argomento: analizza un periodo che va da fine anni 80 al 2013 e propone il grafico in figura in cui si riportano le pubblicazioni divise per anno e tipologia.

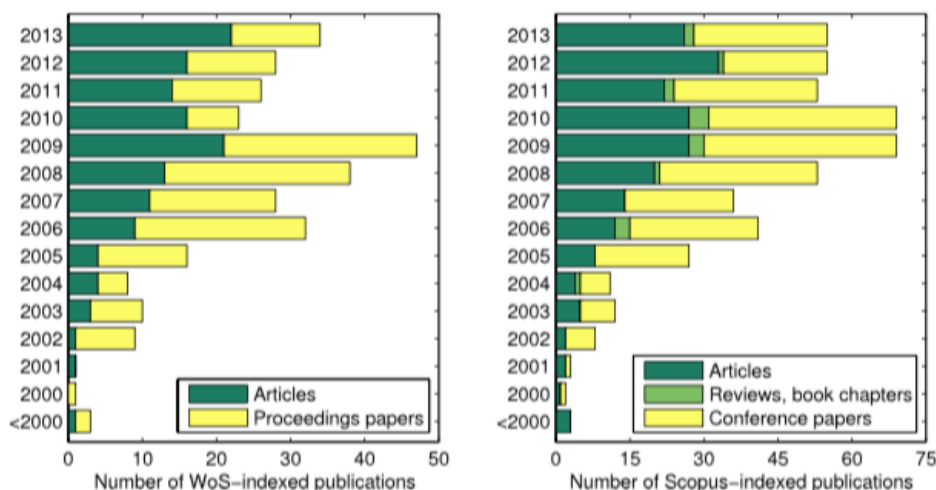


Figura 1 – Pubblicazioni in letteratura riguardo EPF, fonte Weron (2014) [1]

Vi sono poche revisioni che riportano classificazioni esaustive dei modelli utilizzati per EPF, Secondo [2] si identificano tre tendenze di modellizzazione dei mercati studiati: modelli di ottimizzazione, di equilibrio e di simulazione:

- I modelli di **ottimizzazione** si concentrano sul problema della massimizzazione del profitto fra le aziende concorrenti sul mercato, il che rappresenta bene l'aspetto economico del problema ma è ancora un step successivo, prima di tutto serve costruire la previsione in se e perciò non sono modelli utilizzati direttamente in ambito EPF.
- I modelli di **equilibrio** rappresentano il comportamento generale del mercato, prendendo in considerazione la competizione tra tutti i partecipanti, analizzando il tutto in condizione di equilibrio statico del mercato che, come verrà sottolineato, in alcuni casi sarà una semplificazione eccessiva.
- Infine, i modelli di **simulazione** sono un'alternativa ai modelli di equilibrio quando il problema in esame è troppo complesso per essere affrontato all'interno di un quadro di equilibrio formale. Questi rappresentano la soluzione ai problemi accennati per la rottura dell'ipotesi di equilibrio.

In particolare, una prima schematizzazione è da attribuire a [3] pubblicata nel 2009:

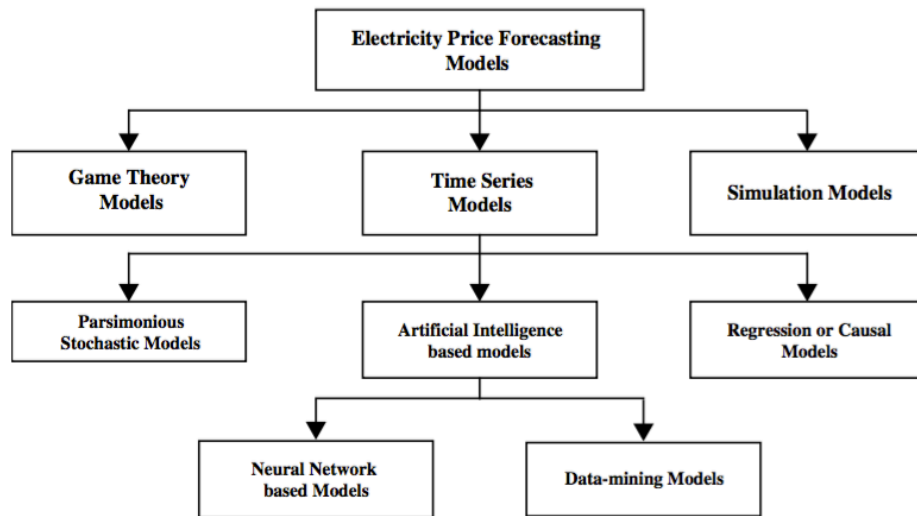


Figura 2 – Classificazione modelli EPF secondo Aggarwal et al. (2009) [3]

Proseguendo poi, la revisione più recente ed esaustiva reperibile letteratura appartiene a [1] che ne suddivide lo spettro in cinque categorie che inglobano le precedenti definite da [3] e [2] ampliandone a sufficienza il livello di dettaglio:

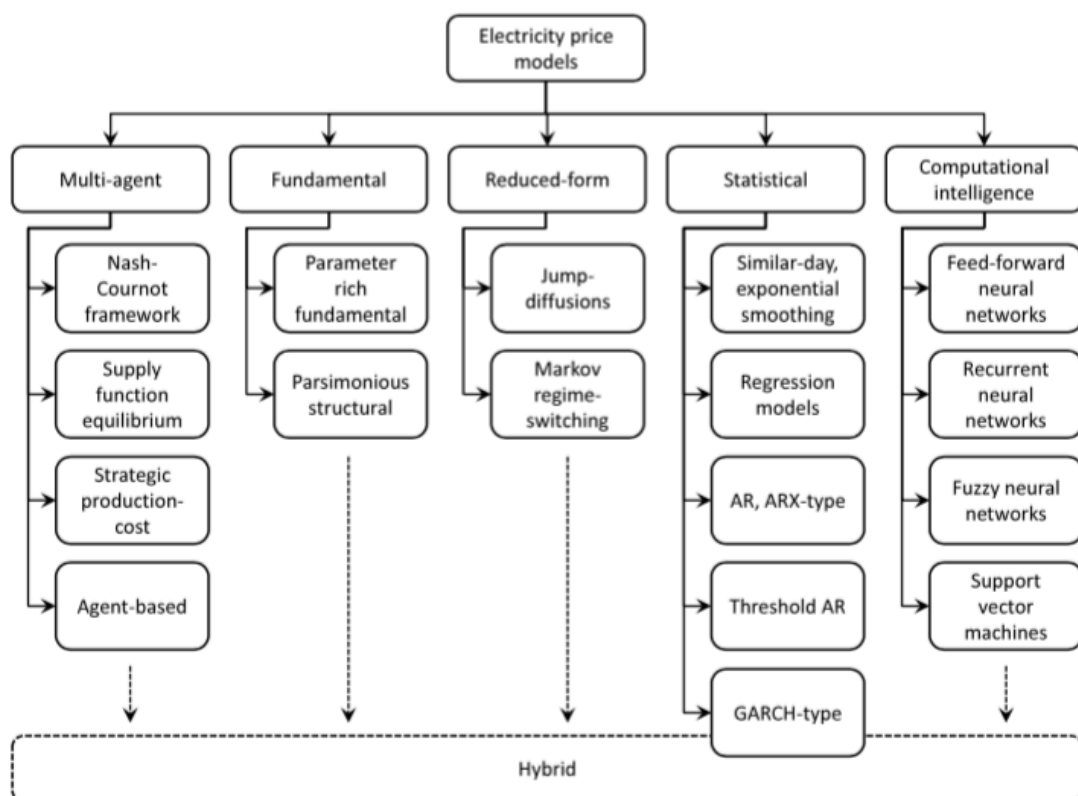


Figura 3 - Classificazione modelli EPF secondo Weron (2014) [1]

Dapprima, quest'ultima classificazione può apparire più caotica rispetto alle precedenti, ma include tutti i metodi citati in letteratura con un livello di dettaglio notevole. Nello specifico la classe dei multi-agent comprende sia i modelli di equilibrio e simulazione [2] sia i "Game theory models" [3], la classe indicata da [3] come "Time series model" è frammentata in "Fundamental", "Statistical" e "Computational intelligence" esplicitate in [1].

A seconda delle caratteristiche di mercato i modelli statistici vengono utilizzati per prevedere in situazioni diverse:

- Modelli di spot price
- Modelli detti "forward"

I primi forniscono una rappresentazione adeguata della dinamica dei prezzi puntuali, che, sulla scia della liberalizzazione dei mercati energetici, diventa uno strumento necessario per scopi di negoziazione. I secondi sono modelli che lavorano bene applicati ai mercati a termine che consentono di fissare i prezzi dei derivati in modo diretto. Per l'analisi di interesse in questa trattazione ci si concentra sulla previsione dello spot-price, quindi si analizzano nello specifico le cinque macro-categorie proposte da [1] specificando dove esse comprendono le altre classificazioni e ampliandole ulteriormente.

1.1. Modelli multi agente

Tale categoria comprende per intero quella definita come "Game Theory models" che hanno come scopo la modellazione delle strategie dei diversi agenti di mercato per poi identificare la soluzione attesa (da qui il nesso con il gaming), in situazioni di mercato oligopolistiche ogni agente sul mercato costruisce la propria curva di offerta in modo da massimizzare i suoi profitti, perciò questi utilizzano modelli matematici per investigarne la soluzione. In più, questi sono modelli basati sui costi (cost-production model, PCM) che hanno la capacità di prevedere i prezzi su scala oraria, tuttavia sono appropriati per i mercati regolamentati con poca incertezza sui prezzi e una struttura stabile, ma non sono adatti per mercati competitivi com'è appunto quello Italiano dopo la totale conversione a mercato libero.

Qui, vi sono di inclusi anche i modelli cosiddetti di equilibrio [3] che sono particolarmente utili per prevedere i livelli di prezzo attesi nei mercati senza una cronologia dei prezzi. D'altra parte, le tecniche di simulazione basate su agenti adattivi sono sempre più popolari e possono affrontare le caratteristiche dei mercati elettrici che i modelli di equilibrio statico ignorano. Dal momento che i modelli di equilibrio e simulazione così definiti condividono molte caratteristiche comuni, si è deciso di considerarle congiuntamente in un'unica ampia classe multi-agente.

In questa enorme categoria rientrano diversi modelli mostrati in Fig.3, i primi tre si basano sull'equilibrio statico del mercato, l'ultimo invece no, nello specifico:

- Nash- Cournot framework: l'energia elettrica viene trattata come un bene omogeneo e l'equilibrio di mercato è determinato sulla base delle decisioni riguardo la capacità dei fornitori. Sfortunatamente, questi modelli tendono a fornire prezzi più alti di quelli osservati nella realtà quindi sovrastimano la variabile prevista.
- Supply function equilibrium (SFE): modella il prezzo come il valore di equilibrio risultante dalla curva di offerta relative alle diverse compagnie e possibilmente la domanda correlata sul mercato all'ingrosso. Per poter calcolare tale funzione è richiesta la soluzione di un sistema di equazioni differenziali, piuttosto che l'insieme tipico di equazioni algebriche che si verifica per Nash-Cournot.
- Strategic-production cost: sicuramente l'approccio meno popolare fra quelli che ipotizzano l'equilibrio statico, è stato proposto da [4] e [5] come modifica dei tradizionali cost-production models (PCM).

I modelli di equilibrio statico discussi sopra si basano su una definizione formale di equilibrio, espressa sotto forma di un sistema di equazioni algebriche o differenziali. Anche se l'insieme di equazioni ha una soluzione, spesso è molto difficile da trovare, perciò si deve ricorrere all'euristica per risolvere il problema [2] [6]. Inoltre, tali approcci di modellizzazione hanno dei limiti nel modo in cui la competizione tra i partecipanti può essere rappresentata.

- Agent-based simulation models: sono modelli che simulano esattamente le interazioni tra gli agenti che operano su quel mercato con metodi che permettono di superare i problemi sopra citati pur non essendo di così difficile soluzione se implementati con una potenza computazionale adeguata. Tale approccio è utilizzato in diversi ambiti ed indicato con l'acronimo ACE (agent-based computational economics). Lo strumento base è una classe di strutture computazionali e regole per simulare le azioni e le interazioni di agenti autonomi (se individui o entità collettive, come organizzazioni o gruppi), con l'obiettivo finale di valutare i loro effetti sul sistema nel suo complesso.

Da un lato, i modelli multi-agente e in particolare i modelli basati su agenti sono una classe di strumenti estremamente flessibili per l'analisi del comportamento strategico nei mercati elettrici. D'altra parte, questa libertà è anche una debolezza, poiché richiede che le supposizioni contenute nella simulazione siano giustificate ed ovviamente, è presente un rischio di modellazione sostanziale. Pertanto, il problema di identificare gli attori rilevanti del mercato e le loro strategie diventa altamente non banale. Inoltre, nonostante le poche applicazioni di previsione discusse sopra, i modelli multi-agente generalmente si focalizzano su questioni qualitative piuttosto che su risultati quantitativi. Possono fornire informazioni sul fatto che i prezzi siano o meno superiori ai costi marginali e su come ciò potrebbe influenzare i risultati di chi ne è agente. Tuttavia, presentano problemi se si devono trarre conclusioni quantitative, in particolare se i prezzi dell'elettricità devono essere previsti con un alto livello di precisione.

1.2. Modelli fondamentali

La prossima classe di modelli è nota come modelli fondamentali o strutturali, cerca di cogliere le relazioni fisiche ed economiche di base che sono presenti nella produzione e nel commercio di elettricità. Vengono postulate le associazioni funzionali tra fattori fondamentali (carichi, condizioni meteorologiche, parametri di sistema, ecc.) e gli input fondamentali sono modellati e previsti in modo indipendente, spesso tramite tecniche statistiche, forma ridotta o di intelligenza computazionale, quindi in qualche modo inglobano l'essenza delle tre classi rimanenti per trattare più fattori endogeni in contemporanea oppure per inglobarne le potenzialità sotto un solo modello ibrido. Per un'ottima introduzione riguardo i fondamenti dei mercati energetici e rispettive modellizzazioni si rimanda al capitolo 4 di [7]. In generale, è possibile identificare due sottoclassi di modelli fondamentali:

- Parameter-rich : come suggerisce il nome vengono tenuti in conto molti parametri, tra cui le variabili endogene quali temperatura, prezzi di combustibili, interruzioni sulla produzione e quant'altro. Iniziano con i processi per i driver primari, quindi costruiscono la curva di offerta, la sua evoluzione ed ottengono i prezzi dell'energia elettrica. I processi simulati mostrano picchi, inversione di tendenza e una struttura di volatilità dei prezzi a termine pressoché corretta. La maggioranza dei modelli di questo tipo sono ideati e sviluppati da utenti privati e perciò è estremamente difficile investigarli nel dettaglio.
- Parsimonious structural models: questa seconda sottoclasse parte da un'analisi empirica delle curve di offerta e domanda del mercato, costruisce il prezzo applicando l'inverso della trasformazione Box-Cox, ovviamente il dettaglio del processo non è così immediato ed è stato introdotto da [8] che oltre ad introdurlo viene calibrato sui mercati di Alberta e California, la previsione ottenuta è continua e presenta ed è capace di esibire i picchi.

L'attuazione pratica di quest'ultimi presenta due problemi principali: Il primo è la disponibilità dei dati. A seconda del mercato, più o meno informazioni sulle capacità e i costi degli impianti, i modelli di domanda e le capacità di trasmissione sono disponibili per il ricercatore o il professionista per la costruzione di tale modello. A causa della natura dei dati fondamentali (che sono spesso raccolti su intervalli di tempo più lunghi, come settimanali o mensili), i modelli fondamentali puri sono più adatti per le previsioni a medio termine che a breve termine ed in genere sono calibrati su dati giornalieri e ignorano le relazioni fini alla risoluzione oraria e perciò non sarebbero adatti al contesto che si analizzerà nelle sezioni successive per il mercato Italiano. La loro applicazione, come quella dei modelli a forma ridotta, è generalmente limitata alla gestione dei rischi e alla determinazione dei prezzi dei derivati.

In effetti, possono essere visti come concorrenti diretti dei primi, consentendo una migliore descrizione dei fondamentali del mercato, anche se a costo di una maggiore complessità dei calcoli analitici e delle procedure di calibrazione. La seconda sfida è l'incorporazione delle fluttuazioni stocastiche dei driver fondamentali. Nella costruzione del modello, facciamo ipotesi specifiche sulle relazioni fisiche ed economiche nel mercato, e quindi le proiezioni dei prezzi generate dai modelli sono molto sensibili alle violazioni di queste ipotesi. Inoltre, più il modello è dettagliato, maggiore è lo sforzo necessario per regolare i parametri. Di conseguenza, esiste un rischio significativo di modellizzazione.

1.3. Modelli reduced-form

Una caratteristica comune di questi modelli è che la loro intenzione principale non è quella di fornire previsioni orarie precise dei prezzi, ma piuttosto di replicare le principali caratteristiche dei prezzi giornalieri dell'elettricità, come le distribuzioni marginali a punti temporali futuri, dinamica dei prezzi e correlazioni tra i prezzi delle materie prime. Tali modelli sono al centro dei prezzi dei derivati e dei sistemi di gestione del rischio, se il processo scelto non è appropriato per cogliere le proprietà principali dei prezzi dell'elettricità, i risultati del modello potrebbero non essere attendibili. Allo stesso tempo, se il modello è troppo complesso, l'onere computazionale ne impedirà l'uso. Da un lato, gli strumenti applicati hanno le loro radici in metodi che sono stati sviluppati per modellare altre materie prime energetiche o tassi di interesse. In un certo senso, le due sottocategorie: i jump-diffusion e i Markov regime switching offrono il meglio dei due mondi: rappresentano un compromesso tra parsimonia e adeguatezza per catturare le caratteristiche uniche dei prezzi.

- Jump-diffusion: sono una classe di modelli introdotti da Robert C. Merton¹ in ambito economico-finanziario ma sono diffusi in più aree di studio, ne sono di esempio le applicazioni per lo studio della diffusione atomica nei reticoli cristallini. In statistica i modelli di diffusione semplici implicano la soluzione di un'equazione differenziale stocastica (SDE) e rappresentano processi Markoviani continui. Tali modelli in ambito economico seguono una serie di casistiche speciali di un'unica SDE così espressa per il prezzo X_t :

$$dX_t = \mu(X_t, t) + \sigma(X_t, t) dW_t + dq(X_t, t) \quad (1)$$

dove dW_t indica un processo di Wiener² standard (moto browniano) e $dq(X_t, t)$ invece un "processo di "salto" puro. Se il primo termine assume un valore costante allora il processo è detto mean-reverting jump-diffusion (MRJD).

- Markov regime switching: l'idea alla base del passaggio al regime di Markov è rappresentata dal comportamento stocastico osservato in un processo come può essere, appunto, l'evoluzione del prezzo dell'energia. Il meccanismo di commutazione tra gli stati è assunto come un processo Markoviano (quali, ad esempio, le discrete random walks) governato dalla matrice di transizione contenente le probabilità P_{ij} di commutazione partendo dallo stato (o valore) i in un certo istante allo stato j all'istante successivo.

Uno dei principali punti deboli dei Jump-diffusion è che incontrano problemi nel prevedere picchi consecutivi alla frequenza rilevata nei dati di mercato. Al contrario, i modelli Markoviani (MRS) sono in grado di prevedere picchi consecutivi in modo molto naturale a patto di costruire un'ottima matrice di transizione. Generalmente non ci si aspetta che i modelli a forma ridotta prevedano con precisione i

¹ https://en.wikipedia.org/wiki/Robert_C._Merton

² https://it.wikipedia.org/wiki/Processo_di_Wiener

prezzi orari, ma si prevede che recuperino le principali caratteristiche dei prezzi (come la prima classe) in genere a scala giornaliera. Tali modelli forniscono un'immagine semplificata, ma ragionevolmente realistica delle dinamiche dei prezzi. È interessante notare che, quando si tratta di previsioni di volatilità o di prezzo, questi modelli sono stati segnalati per funzionare ragionevolmente bene.

1.4. Modelli Statistici

Tale categoria è sicuramente la più ampia ed articolata, raggruppa una grande quantità di metodi e modelli tra cui trova spazio anche quello utilizzato in questo elaborato (si veda capitolo 3) per la previsione di interesse sul “mercato del giorno prima” (day-ahead market) Italiano. Storicamente, il primo afflusso di tecniche EPF consisteva principalmente in metodi statistici di previsione del carico come quantità di energia, da qui è stato sufficiente adattarli sulla previsione del prezzo, per cui i ricercatori sono stati in grado di ottenere modelli EPF e col passare del tempo, sempre più tecniche di elaborazione statistica, econometrica o di segnale sono state introdotte in quest'area.

I metodi statistici prevedono il prezzo corrente utilizzando una combinazione matematica dei prezzi precedenti e / o dei valori precedenti o attuali di fattori esogeni, in genere relativi al consumo e alla produzione o variabili meteorologiche. Le due categorie più importanti sono:

- modelli additivi: la variabile prevista è ottenuta tramite la somma di un certo numero di fattori
- modelli moltiplicativi: la variabile prevista è ottenuta tramite il prodotto di un certo numero di fattori

I primi sono molto più popolari. Si noti, tuttavia, che i due sono strettamente correlati: un modello moltiplicativo può essere trasformato in un modello additivo. I modelli statistici sono interessanti perché alcune loro interpretazioni fisiche possono essere collegate alle componenti, consentendo così di comprendere il loro comportamento.

Sono spesso criticati per la loro limitata capacità di modellare il comportamento non lineare dei prezzi di elettricità e delle relative variabili fondamentali; tuttavia, nelle applicazioni pratiche, le loro prestazioni sono comparabili a quelle delle loro alternative non lineari discusse a seguire nell'ultima classe di metodi.

In seguito si analizzano le principali sottoclassi incluse nei metodi statistici:

- Similar-day and exponential smoothing methods: nello specifico, quelli basati sulle similitudini infra-giornaliere rappresentano un riferimento molto popolare e si basano sulla ricerca di dati storici per giorni con caratteristiche simili al giorno previsto e prendendo i valori storici come previsioni dei prezzi futuri [9]. Tali similitudini possono presentarsi su scala temporali diverse quale settimanale, annuale o mensile, al fine di ottenere previsioni più affidabili è opportuno considerare non solamente un unico giorno ma la previsione può essere una combinazione lineare o una procedura di regressione che può includere diversi giorni simili.

Vi è poi un'altra classe di metodi di più semplice implementazione detti “exponential smoothing” che sono più come nella previsione dei carichi e quindi le curve di domanda piuttosto che in ambito EPF. Essi modellano il prezzo futuro basandosi su un valore pesato sulla base di una media esponenziale dei valori storici della serie di prezzi secondo la seguente relazione:

$$\hat{x}_t = s_t = \alpha x_t + (1 - \alpha) s_{t-1} \quad (2)$$

s_t rappresenta il valore smorzato che viene calcolato come media pesata sul valore assunto all'istante subito prima, il peso decresce esponenzialmente in base al parametro $\alpha \in (0,1)$.

Altri modelli più complessi sono stati sviluppati per processare serie storiche con componenti a tendenza stagionale, l'idea generale è che le previsioni non siano calcolate solamente fondandosi su valori precedenti, ma sia possibile aggiungere una serie indipendente (smoothed) e una componente stagionale.

- Regression models: la regressione è sicuramente una delle tecniche statistiche più diffuse ed utilizzate, in campo EPF si è soliti riferirsi a regressioni multivariate piuttosto che semplici, proprio per la natura ed il numero di variabili che vanno trattate in quest'ambito.

Lo scopo generale di una regressione multivariata è quello di comprendere ed imparare di più riguardo le relazioni tra variabili indipendenti e una variabile dipendente che è quella che vogliamo prevedere. La metodologia di fondo è basata su minimi quadrati: il modello è costruito in maniera tale che la somma dei quadrati delle differenze tra valori osservati e previsti sia ridotta al minimo. Nella sua forma classica, la regressione assume che la relazione tra le variabili sia lineare:

$$P_t = B X_t + \varepsilon_t \text{ dove } B = \begin{pmatrix} b_1 \\ b_2 \\ \dots \\ b_k \end{pmatrix} \quad (3)$$

B è un vettore di coefficienti costanti, X_t è il vettore di regressori, i quali possono essere utilizzati direttamente oppure a seguito di una trasformazione polinomiale³, ε_t indica un termine di errore. I regressori sono selezionati all'interno del campione tra le variabili esplicative considerate, che si presume siano correlate al prezzo, vi è poi una classe di modelli basati sulle regressioni che considerano la variabilità temporale (TVR, time varying regression). Semplicemente il termine di $B(t)$ raggruppa termini non costanti, Nonostante il gran numero di alternative, i modelli di regressione lineare sono ancora tra i più diffusi approcci EPF. Tuttavia, nella maggior parte dei giornali sono combinati con altri metodi, in genere più sofisticati, accade spesso che tali modelli siano difficili da separare dagli approcci auto regressivi, in quanto in letteratura sono spesso chiamati allo stesso modo ma molti utilizzano dei prezzi con ritardo come regressori.

- Autoregression type time series model: basati sul modello lineare auto regressivo classico di ordine p , AR(p), che matematicamente è espresso come segue :

$$P_t = \phi_1 P_{t-1} + \phi_2 P_{t-2} + \dots + \phi_p P_{t-p} + \varepsilon_t \quad (4)$$

³ https://en.wikipedia.org/wiki/Power_transform

combinato con un modello a media mobile MA(q), dove q rappresenta il numero dei valori precedenti del rumore bianco:

$$P_t = \varepsilon_t - w_1 \varepsilon_{t-1} - w_2 \varepsilon_{t-2} - \dots - w_p \varepsilon_{t-p} \quad (5)$$

Il risultato è il modello standard di serie temporali che prende in considerazione la natura casuale e le correlazioni temporali del fenomeno oggetto dello studio è il modello autoregressivo a media mobile il cui acronimo è ARMA (p, q) e la modellazione matematica è espressa attraverso l'uso del backshift operator così definito considerando un caso generale a lag-m:

$$B^m P_t = P_{t-m} \quad (6)$$

Quindi il modello ARMA (p, q) è descritto dalla seguente equazione:

$$\phi B P_t = -w B \varepsilon_t \quad (7)$$

Le serie temporali stocastiche possono essere suddivise in processi stazionari e processi non stazionari. L'assunto di base della stazionarietà include zero media e varianza costante. Nei modelli AR, MA e ARMA sono soddisfatte le condizioni di stazionarietà; quindi sono applicabili solo alle serie stazionarie. Se non accade, allora sarà necessario applicare una trasformazione alla serie per ottenere una forma stazionaria della stessa. Il modello deve la sua fama a Box⁴ e Jenkins⁵ che lo descrivono nel dettaglio in [10, 11] e introducono un terzo parametro d che rappresenta il grado di differenziazione (il numero di volte in cui i dati hanno avuto valori precedenti sottratti), tale azione è giustificata dalla necessità di rendere stazionaria la serie, il modello che ne risulta è detto ARIMA (autoregressive integrated moving average model) o più comunemente detto Box-Jenkins model, se d=0 esso coincide con ARMA.

In alcuni casi la differenziazione semplice a lag-1, anche ripetuta molte volte, non è sufficiente per rendere stazionaria la serie. In particolare, i segnali stagionali di periodo maggiore di uno, come i carichi o i prezzi dell'elettricità, richiedono differenziazioni a ritardi più lunghi. Tali processi sono noti come modelli ARIMA (SARIMA) stagionali.

La notazione generale per l'ordine di un modello ARIMA stagionale con fattori stagionali e non stagionali è $ARIMA(p, d, q) \times (P, D, Q)_s$. Il termine (p, d, q) rappresenta l'ordine della parte non stagionale, mentre $(P, D, Q)_s$ rappresenta l'ordine della parte stagionale. Il valore di s è il numero di osservazioni nel modello stagionale, ad esempio sette per serie giornaliere con periodicità settimanale, 24 per serie orarie con periodicità giornaliera, ecc.

⁴ https://it.wikipedia.org/wiki/George_Edward_Pelham_Box

⁵ https://it.wikipedia.org/wiki/Gwilym_Meirion_Jenkins

I modelli delle serie temporali discussi finora mettono in relazione il segnale in esame con il suo passato e non utilizzano esplicitamente le informazioni contenute in altre serie temporali correlate. Tuttavia, come già accennato, i prezzi dell'energia elettrica sono anche influenzati dai valori presenti e passati di vari fattori esogeni, in particolare la capacità di generazione, i profili di carico e le condizioni ambientali.

Questi modelli non costituiscono una nuova classe; piuttosto, possono essere visti come generalizzazioni di classi esistenti, ad esempio, ARX, ARMAX, ARIMAX e SARIMAX sono controparti generalizzate di AR, ARMA, ARIMA e SARIMA, rispettivamente.

Per cui in un modello ARIMAX, il valore corrente del prezzo spot P_t è espresso linearmente in termini di valori passati della serie stessa, in termini di valori precedenti del rumore e, in aggiunta, in termini di valori presenti e passati delle variabili esogene:

$$\phi B P_t = -w B \varepsilon_t + \sum_{i=1}^k B \psi^i V_i \quad (8)$$

Ove ψ^i sono dei coefficienti polinomiali appropriati e V_i la rispettiva i -esima variabile esogena a cui si riferiscono.

In sostanza, possono essere distinte due classi principali: quelli in cui il regime può essere determinato da una variabile osservabile (e, di conseguenza, i regimi che si sono verificati nel passato e nel presente sono noti con certezza) e quelli in cui il regime è determinato da una variabile latente non osservabile (quali i Markov-regime switching). In quest'ultimo caso, non possiamo mai essere certi che un particolare regime si sia verificato in un determinato momento, ma possiamo solo assegnare o stimare le probabilità delle loro occorrenze.

Il membro più importante della prima classe è il modello AutoRegressive (TAR) Threshold originariamente proposto da [12] e presume che il regime sia specificato dal valore di una variabile osservabile rispetto a un valore di soglia.

- Heteroskedasticity based models (ARC / GARCH): tutti I modelli discussi fino a questo momento si basano sull'ipotesi di omoschedasticità cioè con varianza e covarianza costante. Le serie temporali finanziarie, inclusi i prezzi spot dell'energia elettrica, esibiscono varie forme di dinamica non lineare, che comportano una varianza fortemente dipendente dai valori passati della serie, sono quindi eteroschedastici.

La classe di modelli in cui non è necessario mantenere l'ipotesi di varianza costante è denominata eteroschedastica. Pertanto, questi modelli considerano la varianza come dipendente dal tempo, il primo descritto in letteratura fu l'AutoRegressive Conditional Heteroskedastic (ARCH) di [13] che affrontò con successo il problema dei modelli eteroschedastici seguito da [14] che introdusse il Generalized Autoregressive Conditional Heteroskedastic (GARCH) model.

In conclusione, I metodi statistici prevedono il prezzo corrente utilizzando una combinazione matematica di prezzi precedenti e / o valori precedenti o attuali di fattori esogeni. L'accuratezza della previsione dipende non solo dall'efficienza numerica degli algoritmi impiegati, ma anche dalla qualità

dei dati analizzati e dalla capacità di incorporare importanti fattori fondamentali, come la domanda storica, le previsioni della domanda e dei consumi, le previsioni meteorologiche o i prezzi del carburante. In presenza di picchi, tuttavia, i metodi statistici funzionano piuttosto male, soprattutto per i modelli che si basano sulla sola serie storica del prezzo, ma anche i modelli con variabili esogene non hanno un buon rendimento.

1.5. Modelli di intelligenza computazionale

L'intelligenza computazionale è un concetto difficile definire esaurientemente. E' coerente pensare a CI come a un gruppo molto vario di algoritmi computazionali ispirati dalla natura, tecniche che sono state sviluppate per risolvere problemi che i metodi tradizionali non sono in grado di gestire in modo efficiente. Combina elementi di apprendimento e evoluzione per creare approcci che sono in grado di adattarsi a sistemi dinamici complessi e può essere considerato "intelligente" in questo senso.

I modelli CI sono flessibili e in grado di gestire complessità e non linearità, ciò li rende promettenti per le previsioni a breve termine e numerosi autori hanno riportato la loro eccellente performance in EPF. Sono state applicate anche altre tecniche non parametriche, come la cosiddetta fuzzy-logic, gli algoritmi genetici e la programmazione evolutiva, ma in genere nelle costruzioni ibride. Le reti neurali artificiali (ANN) hanno probabilmente ricevuto la maggiore attenzione nel campo EPF [3], sono modelli matematici che simulano "neuroni" artificiali interconnessi tra loro su più livelli, ispirati da una (pesante) semplificazione delle reti neurali biologiche.

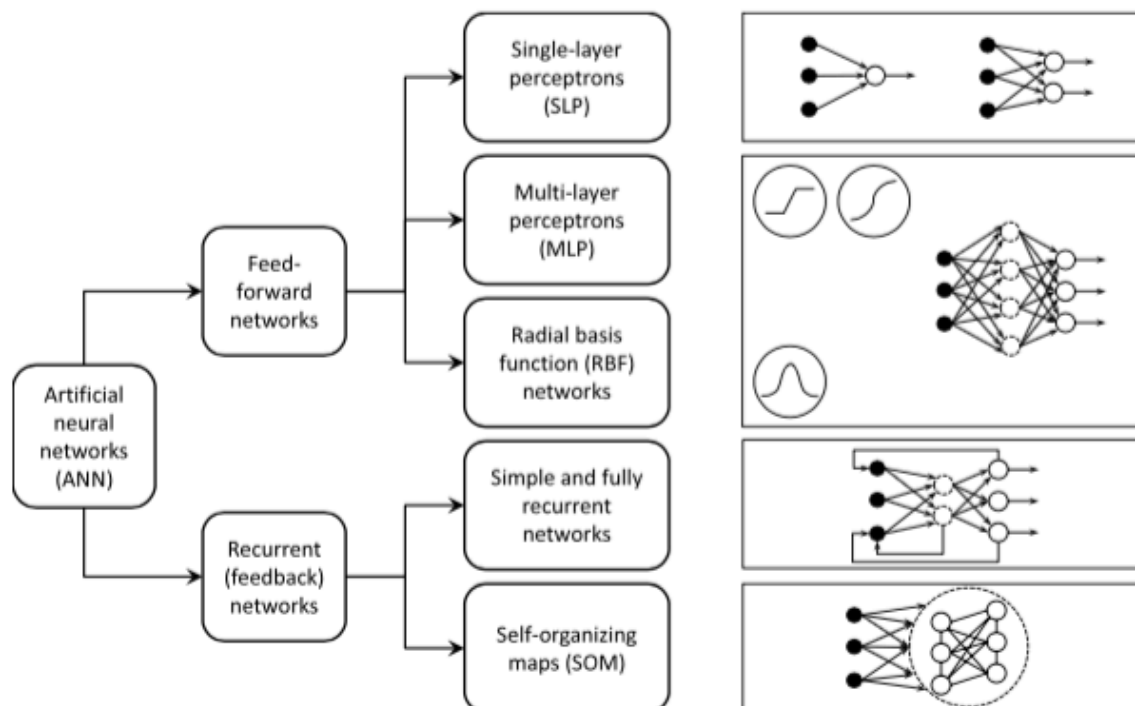


Figura 4 – Gerarchia ANN secondo Weron [1]

- Feed forward neural networks: La rete più semplice, un perceptrone⁶ a strato singolo, non contiene livelli nascosti ed è equivalente a una regressione lineare. Le previsioni sono ottenute da una combinazione lineare degli input. I pesi (corrispondenti ai coefficienti della regressione) sono selezionati usando un "algoritmo di apprendimento" che minimizza alcune funzioni di costo, ad esempio l'errore quadratico medio.

Aggiungendo uno strato intermedio con nodi nascosti, otteniamo il perceptrone multistrato non lineare (MLP). E' opportuno specificare che Le reti di feed-forward sono classificate come statiche, nel senso che producono solo un set di valori di output e non una sequenza di valori da un singolo input. In più non hanno memoria: la loro risposta a un input è indipendente dallo stato precedente della rete, non tengono in considerazione come si sono evolute in precedenza.

- Recurrent neural networks: utilizzano dei circuiti di retroazione (feedback) per risolvere il problema di staticità di cui soffrono le reti feed-forward, per questo motivo le reti ricorrenti sono sistemi dinamici in grado di adattarsi e quando viene presentato un nuovo pattern di input, vengono calcolate le uscite dei neuroni, poi, causa del feedback, gli input per ciascun neurone vengono modificati, il che porta la rete a entrare in un nuovo stato.

In questo caso si riesce ad intuire come tali modelli siano in grado di apprendere durante la loro evoluzione, nello specifico, possono farlo in due modi: supervisionato o non.

I primi evolvono sulla base processi con risposte e comportamenti conosciuti, al contrario per la seconda categoria, alla classe dei "non supervisionati" appartengono le mappe auto organizzanti (SOM) la sequenza di apprendimento usa solo di valori di input, senza alcun segnale di uscita di esempio.

- Fuzzy neural network (FNN): reti basate su una generalizzazione della solita logica booleana, in quanto, invece di un input che assume un valore di 0 o 1, ha determinati intervalli qualitativi ad esso associati. Ad esempio, una temperatura può essere bassa, media o alta.

Secondo [1] ancora un modello rientra in questa categoria, la support vector machine (SVM) introdotta nel 1995 da [15]:

- Support vector machine (SVM): a differenza delle ANN, che tentano di definire funzioni complesse dello spazio di input, SVM esegue una mappatura non lineare dei dati in uno spazio multidimensionale, quindi utilizza semplici funzioni lineari per definire dei confini decisionali nel nuovo spazio. Una caratteristica interessante di SVM è che offre una soluzione unica che è caratterizzata dal minimo globale della funzione ottimizzata. Le applicazioni di SVM nella previsione dei prezzi dell'elettricità sono tipicamente quelle degli elementi nei sistemi ibridi; tuttavia, in uno dei primi articoli su questo argomento [16] confrontano un MLP e un SVM con gli stessi input e concludono che SVM produce previsioni più coerenti e richiede meno tempo per un "allenamento" ottimale.

⁶ <https://it.wikipedia.org/wiki/Perceptrone>

Concludendo anche l'ultima classe di modelli si può affermare che il principale punto di forza degli strumenti di intelligenza computazionale è la loro capacità di gestire la complessità e la non-linearità. In generale, i metodi CI sono migliori nel modellare queste caratteristiche dei prezzi dell'elettricità rispetto alle tecniche statistiche discusse in precedenza. Allo stesso tempo, questa flessibilità è anche la loro principale debolezza, la capacità di adattarsi a comportamenti non lineari e spinosi può non necessariamente portare a previsioni puntuali migliori in certi casi. Un esempio lampante è riportato da [17] che fornisce nuovi spunti per la valutazione delle performance basati sugli sviluppi più recenti nel campo, considera le varie soluzioni implementate dai team coinvolti nella seconda edizione del "Global Energy Forecasting Competition" nel 2014 [18], secondo le metriche proposte in [17] le reti neurali presentano performance mediocri, eppure sono valse al terzo posizione al GEFCom2014. Sempre [17] suggerisce che il metodo implementato fosse lavorasse particolarmente bene per i giorni scelti in competizione e che però non sia un metodo robusto in generale, quindi le reti neurali pretendono grande esperienza di utilizzo al fine di essere effettive.

Inoltre, il pacchetto di strumenti CI disponibili è così vario e ricco che è difficile trovare una soluzione ottimale o peggio ancora, è difficile confrontare a fondo i diversi metodi di CI, anche se l'accuratezza delle previsioni è riportata per lo stesso mercato e lo stesso periodo di test, gli errori dei singoli metodi non sono realmente comparabili a meno che non vengano utilizzati anche periodi identici per la calibrazione e quindi non possono essere usati per formulare affermazioni generali sull'efficienza di un metodo a meno che non sia così. Invece, le conclusioni possono essere tratte solo sulle prestazioni di una determinata implementazione di un metodo, con determinate condizioni iniziali e per un determinato set di dati di calibrazione.

1.6. Modelli Ibridi

Ponendo l'attenzione al futuro, al capitolo 4 di Weron [1] sono raccolti spunti riguardo a possibilità ancora poco esplorate verso le quali potrebbero evolvere modelli sempre più complessi e precisi, comprendono sviluppi della previsione puntuali quindi su intervalli ampi, previsione dei picchi che rappresentano una delle difficoltà maggiori e lo sviluppo di modelli ibridi in cui si sfrutta la natura di diversi modelli per ricavare l'ottimo dai punti forti di ognuno.

Gli ultimi rappresentano la direzione più naturale per il breve termine, sebbene il concetto sia piuttosto semplice e già introdotto verso fine anni sessanta da [19] sembra avere ancora spazio per ulteriori sviluppi, utilizzano modelli diversi per cogliere la natura della serie, da questo punto di vista è opportuno conoscere a fondo i punti di forza delle sottoclassi di modelli e disporre di strumenti utili per la caratterizzazione di una serie.

Il recente lavoro di Yang et al. [20] introduce un modello ibrido che appare molto efficace le cui prestazioni vengono confrontate con altre applicazioni nel paragrafo successivo. Il modello prevede di analizzare separatamente la componenti lineari e non di una serie che le è data in ingresso, si utilizza una wavelet transformation che lavora analogamente ad una trasformata di Fourier e permette di scomporre la serie madre e ottenere un certo numero di sottoserie che vengono poi processate diversamente una volta che l'ipotesi di stazionarietà venga verificata (attraverso l'autocorrelazione e l'ipotesi di radice unitaria). Le serie stazionarie vengono processate secondo un modello ARMA che ne produce la previsione, le restanti vengono processate da un modello KELM (kernel extreme learnig

machine) ottimizzato con un modello SA PSO (Self Adapting Particle Swarm Optimization) che è in grado di adattarsi alle forti non linearità tipiche dei pezzi dell'energia elettrica (i carichi sono molto più facilmente prevedibili invece). I risultati che riporta l'autore sono migliori al confronto dei modelli utilizzati singolarmente, tuttavia non sono presenti informazioni esplicite riguardo l'orizzonte di previsione utilizzato né il tempo di calcolo necessario alla computazione, sicuramente l'implementazione di un metodo composto implica delle difficoltà aggiuntive, ma a rigore di logica anche più tempo, ai fini commerciali una previsione ottima a breve termine non ha alcuna utilità se per ottenere il risultato impiega un tempo maggiore all'orizzonte di previsione.

1.7. Risultati in letteratura

E' importante conoscere l'attuale stato dell'arte in merito alla previsione dei prezzi sui principali mercati elettrici, il mercato libero presenta caratteristiche per le quali è indispensabile sviluppare modelli di previsione sia del carico che dei prezzi, è interesse delle aziende che operano in borsa. Ovviamente le prestazioni raggiunte in ambito commerciale sono particolarmente difficili da ottenere e in quanto tali non vengono pubblicate, né tantomeno i modelli sviluppati dalle aziende. La letteratura scientifica presenta ancora molto margine di sviluppo in questa direzione, e vi sono riferimenti a modelli classici come quelli illustrati in precedenza, la vera innovazione del settore è celata dal segreto aziendale il cui interesse è quello di competere alla migliore previsione e direttamente a massimizzare i propri guadagni.

Ad ogni modo in letteratura si trovano alcune ottime revisioni e articoli in cui si presentano le performance di modelli specifici e confrontate con riferimenti di altri autori o fra modelli meno sofisticati. Prima di tutto, in una pubblicazione del 2009, Aggarwal et al. [3] riporta una tabella interessata riguardo ai mercati più studiati in ambito di ricerca e EPF, prevale quello Californiano seguito da Ontario, New England e PJM ⁷ con pochi riferimenti ad mercati europei dove vi è solo la Spagna e lo UK power pool per poi passare ad Asia e Oceania con Cina, Korea, Queensland e Victoria. Insomma, al 2009 i riferimenti ad applicazioni su mercati reali erano limitati, così tuttora ma alcuni studi più recenti riportati in [21] e [20] sono confrontate le prestazioni reali dei modelli, Bigerna [21] sviluppa un'analisi riguardo l'influenza della temperatura sul mercato Italiano ed è l'unico riferimento disponibile in letteratura mentre Yang et al. [20] sviluppano un nuovo modello ibrido che appare estremamente efficace dal punto di vista delle performance. Per cui a seguire si fa riferimento a queste tre pubblicazioni al fine di fornire una panoramica riguardo lo stato dell'arte delle EPF in ambito di ricerca.

Nel lavoro di [21], l'autrice considera i prezzi zonali nazionali del MGP dal 2005 al 2016, prima le serie sono analizzate singolarmente e sottoposte a test di stazionarietà quali ADF e SURADF [22] e di non linearità introdotto da Kapetanios et al. [23], analizza anche i carichi e le temperature ma prevede solamente i prezzi. Dopodiché vengono confrontati tre modelli: per primo una regressione non parametrica con kernel gaussiano, un modello ARMA con parametri $p=12$ e $q=24$ (orizzonte di previsione a 24h) e come ultimo un modello auto regressivo VAR in diverse varianti. L'effetto della temperatura è introdotto nei modelli VAR2A e VAR2B le cui differenze stanno nei ritardi con cui si considera lo storico di temperatura rispettivamente a breve e lungo termine. I restanti VAR1 e VAR2C considerano rispettivamente variabili legate alla struttura di mercato e condizioni meteorologiche

⁷ <https://www.pjm.com>

estreme (freddo marcato al nord e così via). Un appunto a cui non viene fatto riferimento sta nella metrica, si usa il coefficiente di determinazione r^2 tipicamente usato per valutare le performance di una regressione e il RMSE percentuale (RMSPE), per le regressioni lineari $r^2 = \text{coefficiente di correlazione}^2$ perciò in tabella si riporta quel valore. Invece, il RMSPE è una metrica percentuale che normalizza il quadrato di ogni residuo sul valore reale, ne calcola la media e effettua la radice per arrivare al risultato, comunque verrà discusso nel capitolo 3.

ZONE	KERNEL (R)	ARMA (R)	KERNEL (RMSPE)	ARMA (RMSPE)
NORD	0,57	0,94	6,9	0,9
C. NORD	0,64	0,94	6,2	1,5
C. SUD	0,35	0,94	32	7,3
SUD	0,49	0,94	72,2	2
SICILIA	0,58	0,91	65,2	28,1
SARDEGNA	0,42	0,32	26,7	8,9

Tabella 1 - Coefficiente di correlazione e RMSPE per le se zone nazionali di MGP risultanti da [21]

ZONE	VAR1 (R)	VAR2C (R)	VAR1 (RMSPE)	VAR2C (RMSPE)
NORD	0,95	0,96	0,02	0,01
C. NORD	0,95	0,95	0,038	0,037
C. SUD	0,95	0,95	0,039	0,038
SUD	0,93	0,93	0,055	0,054
SICILIA	0,92	0,93	0,12	0,11
SARDEGNA	0,93	0,93	0,039	0,038

Tabella 2 - Coefficiente di correlazione e RMSPE per le se zone nazionali di MGP risultanti da [21]

L'interesse è confrontare i risultati ottenuti riassunti al capitolo 5 con quelli disponibili in letteratura, a tal fine i valori in tabella 1 e 2 servono come confronto in linea generale, non si considera il PUN ma i prezzi zonali, le metriche non sono del tutto compatibili ma comunque si tratta di prove sullo stesso mercato.

Aggarwal et al. [3] propongono uno studio estremamente esaustivo a proposito, sono raccolte e confrontate un gran numero di pubblicazioni e non solo dal punto di vista delle prestazioni ma includono anche l'estensione dei set di calibrazione dei parametri e valutazione della previsione, più l'orizzonte di previsione che come ribadito è importante nelle applicazioni commerciali. In Tabella 3 e 4 sono raccolte tutte le informazioni estratte da [3], alcune pubblicazioni non mirano al serie di prezzo ma piuttosto a carico, prezzo massimo o medio e alcune alla media oraria del prezzo on-peak, questi vengono tralasciati. Oltre alle prestazioni, gli autori classificano numerosi fattori endogeni influenti e ne riportano

le applicazioni riscontrate, questo è un ottimo spunto per individuare i fattori più influenti in sede di affinamento del modello.

Tuttavia non è specificato su che scala sono raccolti la serie storiche in ingresso, molti considerano variabili aggiuntive ma importa sapere quanti steps temporali sono inclusi nell'orizzonte di previsione, al di fuori dell'ambito accademico è molto importante conoscere questo tipo di informazioni insieme con il tempo computazionale, se il mio codice ha prestazioni eccellenti con orizzonte pari ad un giorno ma i dati sono eccessivi (a causa di uno step dettagliato e necessità di calibrare su set molto ampi), allora la computazione potrebbe produrre risultati in un tempo maggiore azzerandone ogni utilità. Il problema quindi risiede nell'esperienza e abilità per cercare un trade-off tra complessità del modello e potenza computazionale .

PAPER	USCITA	VAL. SET (D)	TEST SET	T	METRICA	PRECISIONE
[24]	P	203	1w	1 step	Daily AvE	2,18 - 11,09
[25]	P	14,28	1d, 30d	1DA	DMAPE	20 – 38 %
[26]	P	180,2	60d, 60d	1DA	DMAPE	8,93 – 12,19 %
[27]	P	77	1w	1DA	DMAPE	11,57 – 12,86 %
[28]	P, Q	363,404,131	1m	1DA	DMAE	1,19 – 1,76
[29]	p	1095,730	2x2d	1,2,3DA	AE	<1c€ 85% casi
[30]	P, C	28	2x1w	1DA	(W)MAPE	8,44 – 15,87 %*
[31]	P, C	90	1w, 1m	1-6h	WMAPE MMAPE	10,69 – 25,77 % 9,75 – 20,03 %
[32]	P, ZP	7,56	1w	Short term	WMAPE	11 – 13 %
[33]	P	-	1w	1DA	DMAPE	10 – 20 %
[34]	p	-	1w	1DA	WMAPE	15,5 %
[35]	p	20	2x3m	1DA	Av. PE	12 – 15 %
[36]	p	-	1w	1DA	DMAPE	3,5 – 5,16 %
[37]	p	182	1w, 92d	1h	WMAPE	15,83 %
[38]	p	30	1m	1h	DMAPE	6,04 %
[39]	p	48	4w stagionale	1DA	WMAPE	7,5 %
[40]	p	-	10d	1 step	APE	8 %
[41]	p	-	2d	1,25,49h	DMPE	2,22 – 8 %
[42]	p	60	1w	1DA	Av. AE	4 – 7,5 %

Tabella 3 - Sommario prestazioni modelli con RETI NEURALI , fonte [3] , * = considerando o meno i valori di picco

La prima colonna indica la grandezze predetta, P significa serie dei prezzi , Q le quantità scambiate, C la curva di carico e ZP i prezzi zionali, sono omessi i modelli specifici, gli algoritmi di apprendimento e le variabili endogene considerate in ingresso al modello che sono comunque reperibili alla tabella 4 in [3]. Dunque, le reti neurali sono molto utilizzate in letteratura e le performance sono abbastanza variabili, non conoscendo l'intervallo del singolo step non si può definire ne quanti punti si usano per calibrare tantomeno per valutare, comunque la metrica più diffusa è il MAPE su diverse scale temporali ,i valori a scala giornaliera (DMAPE) variano dal 3,5 al 38 % che è un range piuttosto ampio. Un

confronto perfetto includerebbe il mercato di applicazione e i periodi di calibrazione e testing perché in base a come vengono scelti questi i risultati cambiano significativamente, basti notare la seconda riga della tabella sopra, si verifica su 30 giorni e per ognuno si valuta il MAPE che varia del 18% in base al giorno che viene predetto.

Non sono riportate le performance per i modelli applicati alla previsione del OPHAP (On-Peak hourly average price) e altre grandezze specifiche come il prezzo medio e massimo, [3] include sei paper a riguardo, i primi due utilizzano solo la serie storica i restanti considerano più variabili, il set di calibrazione varia da quasi 150 giorni a più di 400, il test da una settimana a un anno intero ma comunque ottengono MAPE attorno all'8 – 20% (RMSE = 9,23 in [43] prevedendo i prezzi massimo)

PAPER	MODELLO	USCITA	VAL. SET(D)	TEST SET	T	METRICA	PRECISIONE
[44]	-	P	1	1w	1DA	DMAPE	2,5 – 11,1 %
[45]	AR,ARMA	P	442	45d	1-7 DA	WMAE	3 – 7 %
[46]	AR(X), ARMA(X)	P	272	4w	1DA	WMAPE	3 - 11,1 %
[47]	GARCH	P	147, 105	12m	1DA	WMAPE	9 – 11 %
[48]	DR, TF	P	81, 135, 92	2w, 1w	1DA	DMAPE	3 – 5 %
[49]	ARIMA	p	145, 85, 73, 92	3w & 11w	1DA	WMAPE	8 – 20 %
[50]	-	Av. P	50	2x10d	1DA	MaxAE	1,21 – 4,36 % 36 – 58 %
[51]	-	P	10	2x1d	1h	DPE hPE	1,5% 0,1 – 5,23 %
[52]	ARIMA	P	48	4x1w stagionale	1DA	WMAPE	5- 27 %
[53]	Multivariate arma	P	27	1w	1DA	DAPE max DAPE min	52,2 – 98,7 % 0,1 – 5,3 %
[54]	MLR	P	28	3w	1,2,3DA	-	-
[55]	nonlinear regression	P	540	-	1DA	MSE r²	53,7 – 93,9 0,8 – 0,68

Tabella 4 - Sommario prestazioni modelli STATISTICI (Time serie model), fonte [3]

L'ultima tabella riassume lo stato i risultati per modelli statistici, quindi ARMA(X) /ARIMA(X) / regressioni multivariate / regressioni non lineari, GARCH, MLR e così via. Sebbene si tratti di modelli meno complessi delle reti neurali essi sembrano preformare meglio dei primi, sempre in termini di MAPE i risultati sono più precisi sia a scala giornaliera che settimanale anche senza variabili aggiuntive (vedi AR, ARMA seconda riga).

Come orizzonti di previsione la maggioranza utilizza un giorno intero, che a prescindere dal timestep utilizzato è un valore sensato anche per le applicazioni reali, in particolare il secondo modello analizzato in Tabella 4 è molto performante, usa un intervallo di validazione molto ampio il che implica una buona mole di calcoli e prevede fino ad una settimana prima su 45 giorni diversi rimanendo entro il 7% di WMAPE, questo è senza dubbio uno dei migliori risultati ottenuti considerando la semplicità del

modello e la mancanza di variabili aggiuntive. Fra quest'ultimi molti autori utilizzano un pre-processing delle serie in ingresso al fine di eliminare / separare le componenti caratteristiche, in particolare log e wavelet trasformation. Per concludere il confronto si considera un nuovo metodo ibrido introdotto recentemente da [20] che raccoglie informazioni su tre mercati diversi (PJM, Australie e Spagna) e imposta le previsioni sugli stessi set di calibrazione e test che ritrova in letteratura (per il PJM) e ne effettua un confronto a pari caratteristiche, quindi è un indice veramente effettivo dei miglioramenti introdotti. Brevemente il modello prevede una fase di pre-processo dei dati (wavelet trasformation) seguita da un'analisi di stazionarietà e applicazione di un modello ARMA sulle serie per cui viene rigettata l'ipotesi di radice unitaria e un modello KELM con ottimizzazione SA PSO dei parametri (self-adapting particle swarm optimization). In merito all'interconnessione fra Pennsylvania-New Jersey-Maryland l'autore confronta le performance con [56], [57] e [58] che utilizzano rispettivamente una rete neurale combinata con un metodo a giorni simili, una rete neurale con pre-processo WT e un ELM con pre-processo WT.

La metrica utilizzata è l'errore medio assoluto (MAE) definito come la media dei residui in modulo divisa per la media dei valori reali, riportata in percentuale indica di quanto la previsione si discosta dai valori reali in riferimento alla media dei dati reali su tutto il periodo di test.

2016	DMAE [56] NN-SIMILAR DAY	DMAE [20]	WMAE [56] NN-SIMILAR DAY	WMAE [20] W-SAPSO KELM
20/01	6,93	5,1	-	-
10/02	7,96	4,09	-	-
5/03	7,88	2,84	-	-
7/04	9,02	5,32	-	-
13/05	6,91	3,94	-	-
1-7/02	-	-	7,66	6,35
22-28/05	-	-	8,88	4,1

Tabella 5 – DMAE (%) e WMAE (%): confronto giornaliero performance in letteratura PJM 2016, fonte [20]

2004	WMAE [57] W-NN	WMAE [58] W-ELM	WMAE [20] W-SAPSO KELM
INV : 23-29/02	6,362	6,010	5,278
PRIM : 17-23/05	5,976	4,937	4,805
EST : 23-29/08	5,954	5,843	5,332
AUT: 22-28/10	6,648	6,056	5,492
MEDIA	6,235	5,712	5,227

Tabella 6 - WMAE (%) : confronto settimanale performance in letteratura PJM 2004, fonte [20]

In sostanza il modello [20] aggiunge il processo ARMA e l'ottimizzazione SAPSO al modello di [58], invece i due restanti sono impostati diversamente; in ogni caso si vede che il modello più performante è il W-ARMA-SAPSO KELM sia a scala giornaliera che settimanale, ovviamente era di interesse

dell'autore sottolinearlo ma oggettivamente il confronto è impostato scrupolosamente ed i risultati sono evidenti, la capacità di trattare separatamente le non linearità è un'ottima soluzione per la struttura dei prezzi dell'energia elettrica le cui caratteristiche sono estremamente complesse, molto più dei profili di carico.

Il giorno/settimana influisce sulle prestazioni e in particolare si nota l'influenza delle stagioni, seppure lievi ci sono variazioni in base a che la settimana di test sia in una stagione piuttosto che in un'altra, i mesi primaverili appaiono con l'errore più basso ma non è generalizzabile, in base al modello i pareri sono contrastanti.

GIORNO 2014	KELM	ARMA	W-ARMA- BPNN	W-ARMA- LSSVM	W-SAPSO KELM [20]
GEN	12,1	12,91	11,92	8,7	6,4
FEB	5	5,06	4,6	4,39	2,86
MAR	15,2	18,92	13,28	12,58	2,19
APR	4,78	7,37	3,64	3,46	2,58
MAG	5,63	9,17	4,82	4,57	1,75
GIU	10,54	15,05	7,86	6,67	3,02
LUG	14,8	33	12,94	11,77	7,92
AGO	28,47	74,37	34,3	22,19	9,32
SET	22,63	36,2	15,19	10,45	5,39
OTT	5,4	29,61	4,87	4,68	1,95
NOV	12,44	16,84	11,63	11,05	5,67
DIC	20,11	19,34	18,77	14,94	6,76

Tabella 7 – DMAPE (%): Confronto Victoria , AUSTRALIA 2014 , fonte [20]

SETTIMANA 2013	KELM	ARMA	W-ARMA- BPNN	W-ARMA- LSSVM	W-SAPSO KELM [20]
AUT : 27/10-3/11	4,28	6,36	3,66	3,47	2,18
INV : 28/01-3/02	3,81	3,52	2,95	2,26	1,4
PRIM : 28/04-4/05	9,01	9,23	8,94	7,54	6,26
EST : 28/07-3/08	8,09	9,4	5,59	4,91	3,89

Tabella 8 – RMSE settimanale : Confronto Victoria , AUSTRALIA 2013 , fonte [20]

Il mercato Australiano è separato per il New South Wales (NSW) e Victoria, , i modelli riportati fanno tutti parte dello studio di [20] e non sono confrontati con altri papers. L'autore applica il metodo proposto e lo confronta con i due “semplici” non ibridi (KELM e ARMA) più altri due in cui include il pre-processo WT e rispettivamente una rete neurale “back propagation NN” e un modello SVM combinato con il metodo dei minimi quadrati (LSSVM, least squares support vector machine), usa BPNN e LSSVM proprio perché sono spesso in ambito EPF e sono dei benchmarks. La metrica è il

MAPE in Tabella 9 e RMSE nella successiva ed i valori sono contenuti, il DMAPE minore è a ottobre il massimo ad agosto, la stagione influisce ancora ma il meglio si ottiene in autunno/inverno. Il modello più impreciso è il semplice ARMA seguito dal KELM e poi in cascata tutti quelli ibridi, precisamente BPNN, LSSVM e SASPO-KELM che è il migliore anche per la struttura del mercato australiano, di nuovo emerge l'utilità di trattare separatamente le non linearità presenti su un mercato elettrico deregolato.

GIORNO 2014	KELM	ARMA	W-ARMA- BPNN	W-ARMA- LSSVM	W-SAPSO KELM [20]
GEN	60,84	68,29	41,42	36,93	12,86
FEB	11,93	11,71	9,2	7,45	5,87
MAR	8,03	8,51	9,54	8,03	4,75
APR	5,6	7,03	4,66	4,62	2,46
MAG	5,82	4,63	4,49	4,18	3,47
GIU	5,83	6,77	5,64	5,25	2,28
LUG	3,16	3,62	3,02	2,84	1,86
AGO	4,66	7,47	5,91	5,62	3,67
SET	6,97	8,82	5,25	4,9	2,36
OTT	14,21	14,92	12,83	10,07	4,36
NOV	5,49	5,76	5,27	4,86	3,07
DIC	6,13	8,36	5,72	5,13	3,64

Tabella 9 – DMAPE (%) : Confronto Spagna 2015, fonte [20]

SETTIMANA 2013	KELM	ARMA	W-ARMA- BPNN	W-ARMA- LSSVM	W-SAPSO KELM [20]
AUT : 27/10-3/11	3,71	6,13	4,32	3,33	1,7
INV : 28/01-3/02	6,42	12,07	9,44	7,85	4,76
PRIM : 28/04-4/05	3,6	5,55	4,29	3,58	2,55
EST : 28/07-3/08	2,83	4,52	3,69	3,02	1,75

Tabella 10 – RMSE settimanale : Confronto Spagna 2009, fonte [20]

All'incirca anche per i mercato Spagnolo continua a valere quanto osservato per quello Australiano, il SAPSO-KELM rimane più preciso, il DMAPE massimo a gennaio e minimo luglio e stagionalmente si hanno risultati migliori con la settimana autunnale molto simili all'estate. Al di là dei confronti diretti rintracciabili in letteratura, l'interesse centrale è avere un riferimento recente in merito alle performance ottenute sul mercato Italiano in CAPITOLO 5: RISULTATI, da questo punto di vista bisogna sottolineare che [20] non fa riferimento all'orizzonte di previsione utilizzato, per quanto riguarda il PJM Shrivastava in [58] utilizza orizzonte pari a 24h e riporta anche il tempo medio computazione pari a circa 30 s per script in MATLAB (16 Gb RAM e processore 2,93 GHz, omette il numero di core). Perciò il riferimento al PJM si prevede ad un giorno, verosimilmente sarà così anche per gli Australia e Spagna ma non viene specificato.

Capitolo 2: ANALISI del MERCATO ELETTRICO ITALIANO

In questo capitolo si conduce uno studio relativo al mercato elettrico italiano: la struttura, i mercati, le caratteristiche della borsa e se ne analizza l'evoluzione attraverso lo studio delle serie storiche relative ai prezzi dell'energia elettrica, vengono confrontate le zone di mercato su diversi intervalli temporali e si propongono statistiche fini ad identificare dipendenze utili, variabili fortemente correlate al prezzo dell'energia elettrica e cause di dipendenze più o meno marcate quali il prezzo del petrolio o le condizioni climatiche sul territorio nazionale.

2.1. STRUTTURA DEL MERCATO ELETTRICO ITALIANO

Il mercato elettrico in Italia nasce per effetto del decreto legislativo n°79 del 16 Marzo 1999, noto anche come decreto Bersani, come recepimento della direttiva comunitaria 96/92/CE per la creazione di un mercato interno dell'energia a seguito della liberalizzazione del mercato elettrico, istituita in Italia a partire dal primo Aprile 2004. La fonte principale per le nozioni e i dati storici relativi alla borsa elettrica è il gestore dei mercati energetici [59], fondato a Roma nel 2009 e si occupa di monitorare ed appunto gestire i mercati energetici, a quello elettrico si aggiunge il mercato del gas naturale compreso quello di stoccaggio, il GME si occupa anche di gestire l'organizzazione dei “certificati verdi” e dei titoli di efficienza energetica riguardo ai mercati per l'ambiente, anche conosciuti come “certificati bianchi” relativi alla quota di produzione da fonte rinnovabile obbligatoria per tutti i produttori sul territorio nazionale. Nello specifico, la mappa concettuale in Figura 5 mostra la struttura del mercato elettrico Italiano nella sua completezza.

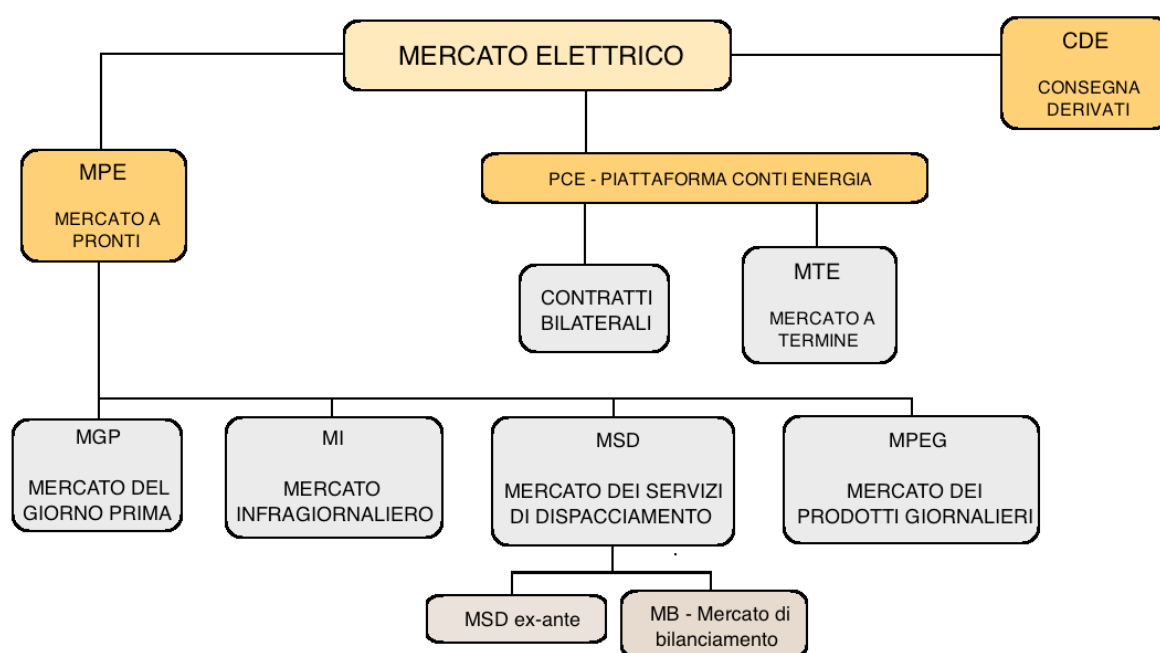


Figura 5 – Struttura del mercato elettrico Italiano

Solitamente ci si riferisce al mercato elettrico a pronti con il nome di “borsa elettrica Italiana” e si articola in diverse fasi, sia per l’ordine cronologico che per gli attori che coinvolge, le trattazioni avvengono su una piattaforma telematica con accesso web per tutti gli operatori accreditati. Essenzialmente il mercato è suddiviso in due filoni principali in base al tipo di contrattazioni coinvolte che posso essere a **termine** o a **pronti**. Ogni sottoclasse presenta le sue caratteristiche e la prima ingloba a sua volta altri mercati articolati su più sedute, vi sono delle caratteristiche interne tipiche di ogni mercato come ad esempio la struttura: i mercati quali MGP ed MI sono mercati d’asta in cui le offerte sono a base oraria e per un dato giorno, al contrario di altri in cui si le negoziazioni si svolgono in modalità continua quali sono MPEG, MSD più il mercato dedicato alle contrattazioni a termine.

A questi due grandi categorie si affianca una piattaforma che si interessa di eseguire i contratti finanziari derivati sull’IDEX ⁸, denominata **Consegna Derivati Energia (CDE)**, questa comprende tutti quei contratti derivati per i quali l’operatore ha richiesto l’operazione di consegna fisica sul mercato, ne sono automaticamente ammessi tutti gli operatori del mercato, bensì la richiesta di consegna fisica sull’ME sia accessibile solo a quelli che dispongono di un conto energia sulla PCE [60] (Piattaforma Conti Energia). La PCE è lo strumento utilizzato per poter valorizzare i contratti negoziati in diverse sessioni dei mercati a termine, sede di contrattazioni bilaterali fra soggetti attivi sulla borsa, gli operatori che concludono contratti al di fuori del sistema delle offerte, tramite la PCE, registrano le obbligazioni e i relativi programmi di immissione e prelievo di energia dalla rete.

Non è da trascurare l’aspetto tecnico riguardo rete di trasmissione (gestita da Terna S.p.a) come anche quello legato alla morfologia del territorio, l’Italia è un penisola stretta e allungata con Sicilia e Sardegna che sono isole a parte, per questo la rete di distribuzione presenta dei vincoli e limiti di transito su dorsali principali piuttosto che linee secondarie. Tali vincoli hanno ripercussioni sul mercato in termini di disgregazione dei prezzi portando a differenze di costo in base alla zona geografica considerata, ciò accade per lo più per le isole dove solitamente si ragiona su valori diversi, invece per le zone confinanti è comune avere un prezzo uniformato con le eccezioni che ne derivano. I prezzi si riferiscono a zone di mercato così organizzate⁹ [61]:

- **Zone geografiche:** in tutto sono sei, raggruppano regioni in base alla posizione geografica sul suolo nazionale:
 - Centro nord (**CNOR**)
 - Centro sud (**CSUD**)
 - Nord (**NORD**)
 - Sardegna (**SARD**)
 - Sicilia (**SICI**)
 - Sud (**SUD**)
- **Zone virtuali nazionali:** che rappresentano poli di produzione limitata, ovvero zone costituite da sole unità di produzione, la cui capacità di interconnessione con la rete è inferiore alla potenza installata delle unità stesse, sono cinque:

⁸ <https://www.borsaitaliana.it/derivati/idx/idx.htm>

⁹ <http://www.mercatoelettrico.org/It/Mercati/MercatoElettrico/Zone.aspx>

- Brindisi (**BRNN**)
 - Foggia (**FOGN**)
 - Montefalcone (**MFTV**)
 - Priolo Gargallo (**PRGP**)
 - Rossano (**ROSN**)
- **Zone virtuali estere:** che rappresentano di interconnessione con l'estero e i market coupling:
- Austria (**AUST**)
 - Slovenia coupling (**BSP**)
 - Corsica (**CORS**)
 - Corsica AC (**COAC**)
 - Francia (**FRAN**)
 - Grecia (**GREC**)
 - Slovenia (**SLOV**)
 - Svizzera (**SVIZ**)
 - Francia coupling (**XFRA**)
 - Austria coupling (**XAUS**)
 - Malta (**MALT**)
- **Zone non attive:** che risultano tali attualmente, in quanto i dati relativi a queste zone sono riportati per il MGP solo per periodi precedenti alla sostituzione del mercato di aggiustamento (MA) con il mercato infra-giornaliero (MI), cioè fino al 2009.
- Calabria (**CALB**)
 - Piombino (**PBNF**)
 - Turbigo-Ronco (**TBRV**)
 - Estero Corsica AC (**E_CO**)
 - Estero Nord-Est (**E_NE**)
 - Estero Nord-Ovest (**E_NW**)
 - Estero Sud (**E_SD**)

In aggiunta alle zone di mercato sui file scaricabili dal sito del Gestore dei Mercati Energetici (GME)¹⁰ è disponibile alla voce

- **NAT** : che indica “Italia senza vicoli” che non è un zona di mercato ma si riferisce al prezzo calcolato in assenza di vincoli tra le zone di rete e tale voce è riportata anche sulle tabelle degli esiti di mercato.

¹⁰ [http://www.mercatoelettrico.org/It/Download/DownloadDati.aspx?val=MGP Prezzi](http://www.mercatoelettrico.org/It/Download/DownloadDati.aspx?val=MGP%20Prezzi)

Di seguito si scende nel dettaglio riguardo la struttura dei due mercati principali, le modalità con cui si effettuano le contrattazioni e le caratteristiche principali dei mercati stessi.

A. Il mercato a pronti (MPE) : comprende la maggior parte delle transizioni di compravendita dell'energia e a sua volta si articola in più mercati che rappresentano in fulcro delle attività di trading sulla borsa elettrica, non prevede contratti bilaterali e l'incontro tra due parti ma la controparte di ogni operatore è il mercato stesso (GME agisce come controparte principale), dove il prezzo è stabilito dall'incontro tra domanda e offerta.

- **Mercato del giorno prima (MGP) :** in questa sede si negoziano quantità di energia in blocchi orari per il giorno successivo attraverso la formulazione di offerte di scambio da parte di ambedue produttori e acquirenti, la seduta si apre alle 8:00 del nono giorno antecedente al giorno della consegna e si chiude alle 12:00 del giorno precedente.

Gli operatori che vi partecipano presentano offerte dove indicano il prezzo massimo o minimo al quale sono disposti ad acquistare o vendere una determinata quantità di energia, a seguito vengono costruite le curve cumulate di domanda e offerta e ne deriva così il prezzo di equilibrio ora per ora. Le offerte vanno formulate entro fine seduta e vengono accettate o respinte alla chiusura della stessa sulla base del merito economico e nel rispetto dei limiti di transito tra le zone, questo rende il MGP un mercato d'asta piuttosto che a contrattazione continua. Quindi, una volta stabilito il prezzo di equilibrio come il quello dell'offerta più costosa che permette di soddisfare la domanda oraria, tutte le offerte vengono valorizzate al prezzo unitario nazionale (PUN), tenendo in conto la congestione sulle linee causata dai limiti di transito, il PUN risulta dalla media dei singoli prezzi zonali ponderati per le quantità acquistate in tali zone. Per quanto riguarda le offerte riferite alle zone virtuali estere esse vengono valorizzate al prezzo marginale di equilibrio della zona a cui appartengono e si trattano allo stesso modo di quelle nazionali.

Mercato infra-giornaliero (MI): da intendersi come appendice al MGP, permette agli operatori di apportare modifiche su quanto definito nel corso della sessione di MGP per il giorno i-esimo, si articola in sette sedute indicate con MI1, MI2, ... , MI7 che hanno luogo a cavallo fra il giorno precedente ed il giorno stesso a cui fanno riferimento le offerte (vedi Figura 32). Il meccanismo di selezione delle offerte, che siano di acquisto o vendita, è lo stesso descritto per il MGP con un'unica differenza, le offerte vengono valorizzate al prezzo zonale di riferimento e non più al PUN come accade nel MGP. Con la nascita del GME S.p.A nel 2009 ha sostituito il Mercato di Aggiustamento (MA) che come descritto in [59] funzionava analogamente al mercato infra-giornaliero a cui potevano partecipare solamente i produttori.

Mercato dei prodotti giornalieri (MPEG): è sede per la negoziazione dei prodotti giornalieri con obbligo sulla consegna dell'energia e si svolge in modalità continua. Vi sono due tipo di prodotti negoziabili: prodotti con “differenziale unitario di prezzo ” e quelli a “prezzo unitario pieno”, la differenza sta nel fatto che il primo si esprime con riferimento al differenziale rispetto al PUN, al quale l'operatore è disposto a negoziare tale prodotto. In aggiunta, per entrambe le tipologie di prodotto sono negoziabili due profili di consegna che sono Baseload e Peak Load, il primo è quotato per tutti i giorni di calendario cosa che non accade per il peak load che è attivo solamente da lunedì

a venerdì. Caratteristiche principali dei prodotti di baseload sta nella disponibilità di consegna in tutti i periodi rilevanti per il giorno di negoziazione, invece nei soli periodi rilevanti al bilanciamento in tempo reale sulla rete elettrica per i prodotti di peak load. Le sessioni di MPEG hanno luogo solamente nei giorni feriali e si articolano in su in due sedute rispettivamente dalle ore 8 alle 17 del secondo giorno precedente e dalle ore 8 alle 9 del giorno precedente, il venerdì si negoziano i prodotti per la consegna per il sabato (h. 8-9), domenica, lunedì e martedì in blocco dalle 8 alle 17, i prodotti del martedì rimangono comunque negoziabili nella sessione del lunedì.

Mercato per il servizio di dispacciamento (MSD) : a sua volta è frammentato in una fase di programmazione che prende il nome di MSD ex-ante ed il mercato di bilanciamento, entrambe si svolgono su più sessioni come definito in Figura 32. È l'unico mercato a pronti in cui la figura che agisce da controparte principale è Terna S.p.A.¹¹ e non il GME, questo è lo strumento che Terna utilizza per stanziare le risorse necessarie alla gestione del sistema elettrico nazionale al fine di risolvere le congestioni, creare una riserva di energia ed effettuare il bilanciamento in tempo reale della rete. Un'altra differenza sostanziale rispetto ai mercati visti fino a questo punto è che le contrattazioni sul dispacciamento vengono concluse al prezzo presentato dall'offerta (pay-as-bid).

Sul MSD ex-ante Terna accetta offerte di acquisto e vendita di energia ai fini della risoluzione delle congestioni residue e della costituzione dei margini di riserva e le offerte relative ad esso vengono presentate tutte in dalle 12:55 del giorno precedente alle 17 del giorno stesso, e gli esiti vengono comunicati secondo sei sessioni a orari diversi, l'ultima alle 18:15 del giorno di consegna.

Sul MB Terna accetta offerte di acquisto e vendita di energia al fine di svolgere il servizio di regolazione secondaria e mantenere il bilanciamento, nel tempo reale, tra immissione e prelievi di energia sulla rete. Anche quest'ultimo è articolato in sei sessioni, per la prima vengono considerate le offerte presentate nell'ultima sessione del MSD ex-ante e per le altre sessioni le offerte si possono presentare dalle 22:30 del giorno precedente e si chiudono un'ora e mezza prima dell'ora in cui possono essere negoziate.

B. Il mercato a termine (MTE) : è la sede di negoziazione dei contratti a termine relativi all'energia elettrica con obbligo di consegna e ritiro, si tratta di accordi bilaterali tra operatori del settore per quantità di energia fissate ad un prezzo fisso e per un intero periodo, ad esempio su scala annuale. Esistono altre piattaforme di trading oltre il MTE messe a disposizione, dietro compenso, da alcuni broker del settore, come GFI, TFS e ICAP, solitamente risulta che in questo ambito la piattaforma MTE sia comunque poco utilizzata a causa delle grandi garanzie richieste dal GME per potervi accedere. Sono ammessi direttamente tutti gli operatori del mercato elettrico e le negoziazioni sono continue come per MSD e MPEG, sul MTE sono negoziabili contratti sia di tipo base-load che peak-load con periodi di consegna che possono essere mensili, trimestrali o annuali, gli operatori vi partecipano presentando proposte nelle quali indicano la tipologia dei contratti e il periodo di consegna, in più dovranno indicare il numero di contratti negoziabili e il prezzo al quale sono disposti ad acquistare o vendere.

¹¹ <https://www.terna.it>

Tutti gli scambi fisici di energia esistono solamente in maniera virtuale, esattamente come avviene nel trasferimento di fondi fra conti differenti, al termine del periodo di negoziazione i contratti sono registrati in corrispondenti transizioni di energia servendosi delle PCE dopo averne confermato la congruenza il regolamento PCE.

Come tempistiche, le sessioni del MTE si svolgono dal lunedì al venerdì (h.9-17) eccetto il penultimo giorno di mercato del mese dove la chiusura viene anticipata alle ore 14.

Concludendo, le tipologie principali di contratti si riferiscono alla borsa elettrica nel caso del mercato dei contratti a pronti, in cui il Gestore dei Mercati Energetici S.p.A svolge la funzione di controparte centrale per tutte le trattazioni registrate su MGP, MI, MPEG invece Terna S.p.A ricopre il medesimo ruolo per il MSD che lo utilizza per svolgere le importanti funzioni di bilanciamento, risoluzione delle congestioni e riserva. Tuttavia, la borsa elettrica non è obbligatoria, l'energia elettrica può anche essere contrattata bilateralmente sulla base di liberi accordi fra gli operatori che partecipano alle sedute del mercato a termine, quindi il GME S.p.A o Terna S.p.A non sono più le controparti centrali con cui avvengono le contrattazioni ma hanno comunque un ruolo fondamentale, il GME ha realizzato e gestisce per conto proprio e di Terna la Piattaforma Conti Energia (PCE) grazie alla quale vengono registrati tali contratti, assicurando così che gli obblighi e i termini di consegna stabiliti siano rispettati secondo le modalità stabilite dalla PCE.

Detto questo, l'analisi riporta in questa trattazione si limita alle operazioni di borsa e quindi al mercato a pronti in cui è disponibile la serie storica di prezzi orari dalla fondazione del GME S.p.A ad oggi senza buchi e su scala nazionale, si analizzano le differenze tra MGP e MI tramite confronti e statistiche temporali e zonali ed infine si effettua la previsione del PUN disponibile solamente per il Mercato del Giorno Prima.

2.2. MGP: MERCATO DEL GIORNO PRIMA

Come già ribadito esso rappresenta il mercato principale su cui vengono costruite la domanda e l'offerta per il giorno successivo in modo tale da organizzarne la distribuzione sulla rete nazionale gestita da Terna attraverso il servizio di dispacciamento, ogni produttore o utente del mercato deve comunicare entro fine giornata le quantità orarie fornite o richieste per il giorno a venire, da qui ne derivano i vari prezzi zonali che insieme con le quantità scambiate contribuiscono al PUN (Prezzo Unico Nazionale). A fine seduta tutte le offerte accettate vengono valorizzate al prezzo dell'ultima offerta necessaria a soddisfare il fabbisogno ora per ora, quindi il più elevato. Da qui appare chiara l'importanza di riuscire a prevedere correttamente il profilo orario dei prezzi, per qualsiasi azienda, che sia pubblica o privata, che opera sulla borsa elettrica è fondamentale sviluppare codici di previsione, eventuali errori sulla previsione di domanda e offerta da parte delle società operatrici si traducono in quantità di energia in eccesso o debito e quindi perdite economiche ingenti su grosse quantità di energia. Considerata la totalità delle offerte avanzata da un'azienda del settore, questa potrà guadagnare solamente su quelle che può fissare sotto il prezzo di equilibrio e che verranno valorizzate a quel prezzo, ovviamente se l'azienda eccelle nella produzione di energia a basso costo tanto maggiore sarà il guadagno, è banale ma queste considerazioni sono cruciali per stabilire piani di produzione e vendita. In più, sono previste more per energia in surplus e guadagno nullo per le offerte che non vengono accettate, quindi in ogni caso l'errore di previsione incide notevolmente sull'utile giornaliero degli operatori.

Riguardo la reperibilità dei dati si fa riferimento alla sezione download sul sito GME ¹² ove sono disponibili gli storici per tutti i mercati in forma più o meno immediata, lo storico del PUN è completo per l'intero periodo mentre le quelli zonali sono confusi e soprattutto non sono continui nel periodo analizzato, non tutte le zone sono presenti per tutti i mercati e per lo stesso periodo di tempo, ciò rende più complicato il confronto fra le diverse serie temporali del MGP e nelle prossime sezioni anche con il mercato infra-giornaliero.

A partire dal D.lgs. 79/1999, che ha effettivamente introdotto l'istituzione della borsa elettrica Italiana, i dati del mercato a pronti relativi alla borsa elettrica sono disponibili dal primo Aprile 2004, perciò il periodo di riferimento da qui in poi comprende valori su scala oraria che vanno dagli **ultimi due trimestri del 2004 al primo trimestre del 2018**, soffermando l'interesse ai soli valori del PUN [€/MWh] per la previsione ed estendendo il confronto anche alle zone di mercato, ove possibile. I valori zonali non sono confrontabili per tutti gli anni, non tanto per la presenza di buchi ma perché dal 2004 ad oggi la struttura della rete si è modificata, la divisione è dettata dai limiti di transito e congestioni sulla rete il che la rende in continua dinamica è imprevedibile, si faccia riferimento al precedente paragrafo per notare le cosiddette "Zone non attive" sul mercato attuale ma che lo erano fino 2008, a seguito nel 2009 è stato fondato il Gestore dei mercati energetici S.p.A. e dal quel punto in poi le zone sono rimaste più o meno invariate eccetto sporadiche eccezioni di anno in anno.

¹² http://www.mercatoelettrico.org/It/Download/DownloadDati.aspx?val=MGP_Prezzi

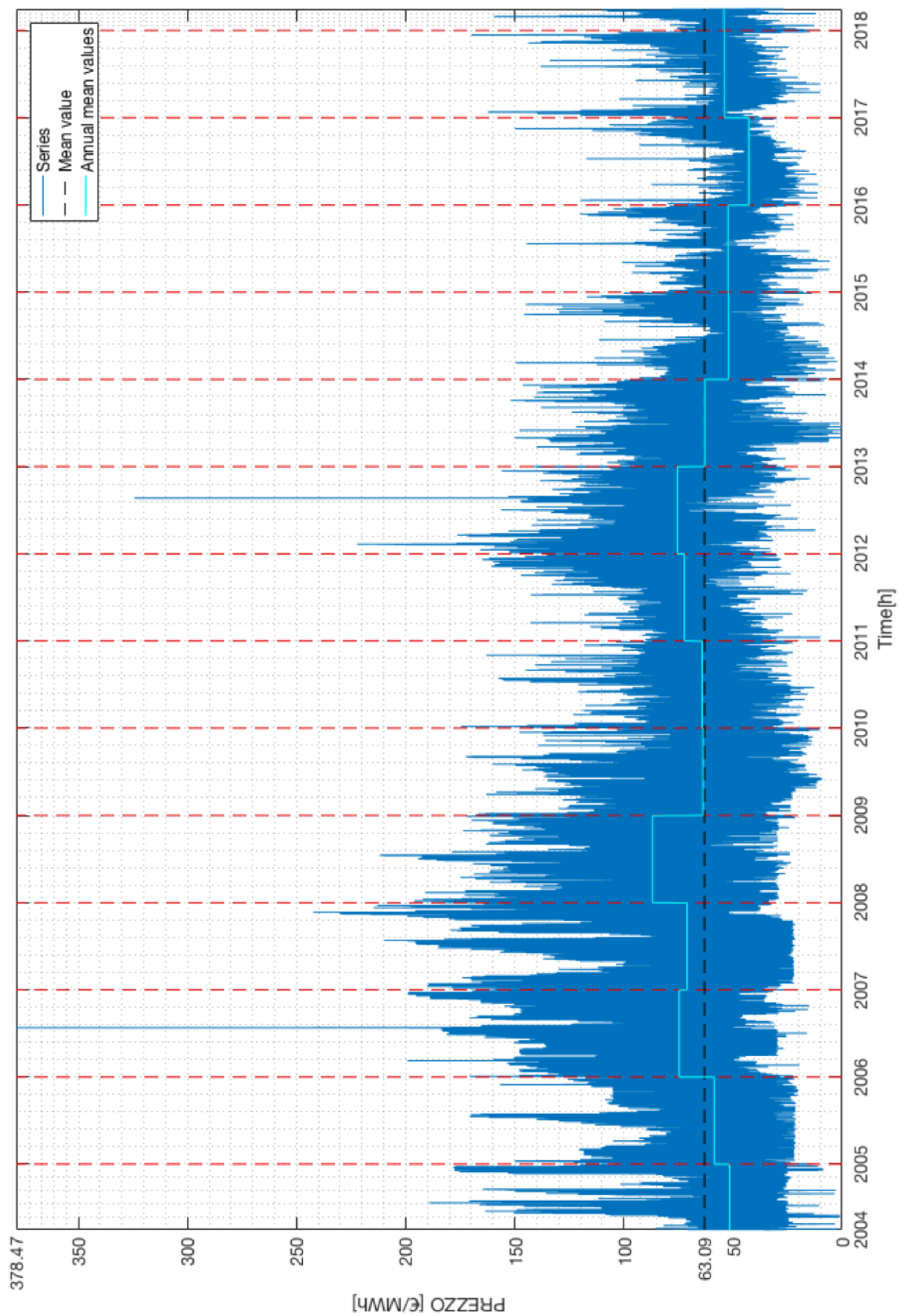


Figura 6 – Serie completa PUN [€/MWh] dal 1 Aprile 2004 al 31 Marzo 2018

In Figura 6 è mostrato l'andamento della serie completa, le linee verticali rosse indicano il primo valore di Gennaio, mentre la linea nera a tratti include l'indicazione sulla media aritmetica dell'intera serie, il grafico viene limitato dai valori massimo e minimo delle ordinate che raggiungono rispettivamente $378,47 \frac{\text{€}}{\text{MWh}}$ nell'estate 2006 (h.11 del 25 Luglio) e $0 \frac{\text{€}}{\text{MWh}}$ al 2014, in più risalta un altro picco anomalo pari a $324,2 \frac{\text{€}}{\text{MWh}}$ nel 2012 (h.20 del 21 Luglio). Il fatto curioso sta nel valore minimo che è registrato come nullo per due ore alle 13 e 14 del 25 Maggio 2013, in termini matematici questo implica che la curva di domanda è traslata per solo valori negativi di prezzo e la curva di offerta presenterebbe anch'essa valori quantità negative, in altri termini si dispone di energia in eccesso che viene valorizzata a prezzo nullo e quindi "regalata". L'orario ed il periodo spingono a pensare ad un surplus dovuto alla produzione da fotovoltaico e se ne trova una giustificazione al paragrafo 2.5.2, dove viene analizzata la produzione da FER in Italia. La linea color ciano indica anno per anno la media aritmetica, questo approccio è funzionale per apprezzare le differenze tra media totale e quella locale, fino al 2008 la media annuale tende a crescere con una leggera depressione al 2007, poi si abbatte e rimane più o meno costante e simile alla media totale per due anni ed infine un leggero incremento fra il 2011 e 12 con successivo decremento fino per arrivare al 2017. Mediamente appaiono prezzi più bassi negli ultimi anni, confrontabili con i valori di fine 2004, nel mezzo invece l'energia elettrica ha assunto valori maggiori, il cui merito ha sicuramente a che fare con la grande crisi che investì prima i mercati Americani e poi anche quelli Europei in quel periodo.

Qualche informazione in più sulla statistica della serie può essere aggiunta in termini di boxplot che illustra la mediana, il primo e terzo quartile e i valori adiacenti superiore inferiore evidenziando i valori fuori scala, a fianco si aggiunge la distribuzione dei valori e la media che risulta superiore alla mediana di poche unità. Dunque, la metà dei dati è contenuta tra i 45 ai $76 \frac{\text{€}}{\text{MWh}}$ ed i valori sono compatti sotto il primo quantile e molto dispersi sopra il terzo, anche se il massimo è elevato ci sono pochissimi valori che se ne avvicinano e comunque la maggioranza dei valori sta sotto i $200 \frac{\text{€}}{\text{MWh}}$.

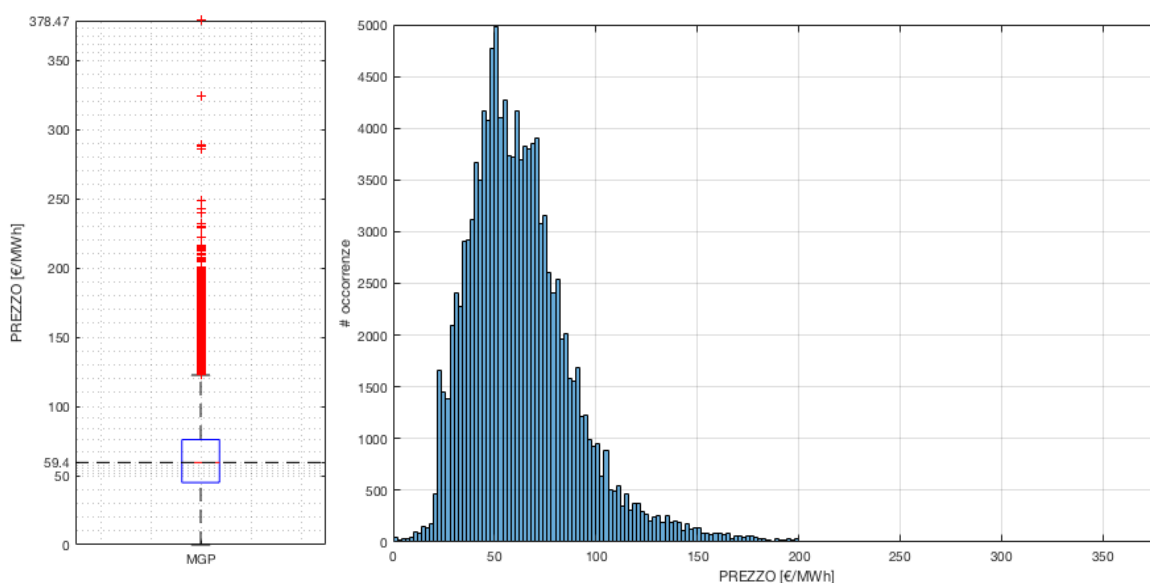


Figura 7 – Sinistra: Boxplot della distribuzione, Destra: distribuzione di frequenze.

Figura 8 e Figura 9 raccolgono le medesima informazioni per i singoli anni, ad eccezione del 2013 non accade mai che la mediana assuma valori maggiori della media, rimane sempre leggermente inferiore o al limite sono praticamente collimate. I valori minimi sono poco variabili e toccano massimo $21,54 \frac{\text{€}}{\text{MWh}}$ nel 2008 mentre i valori massimi vanno smorzandosi nel tempo a partire dal 2006 eccetto il picco nel 2012, ma comunque i valori di picco del 2006 e 2012 sono casi isolati e di poca importanza sulla totalità della serie anche solamente su scala annuale.

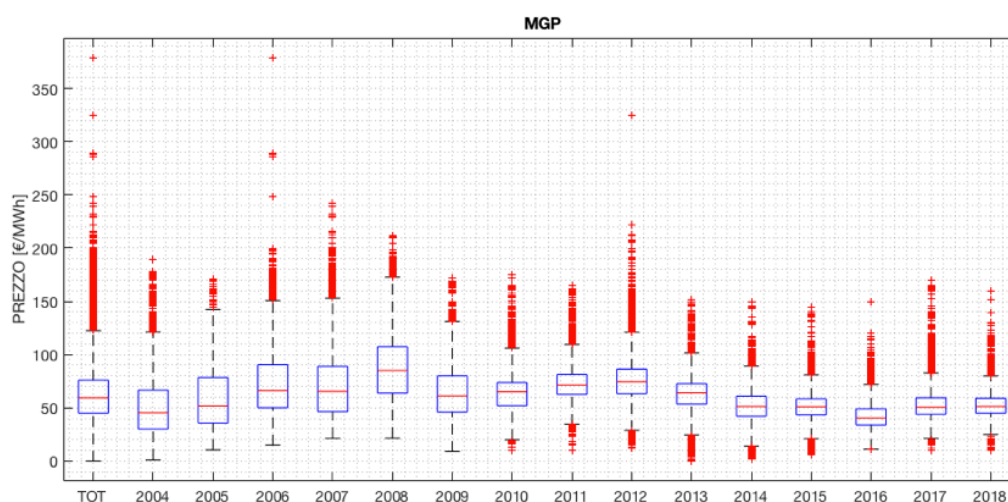


Figura 8 – Boxplot relativi alle singole distribuzioni annuali

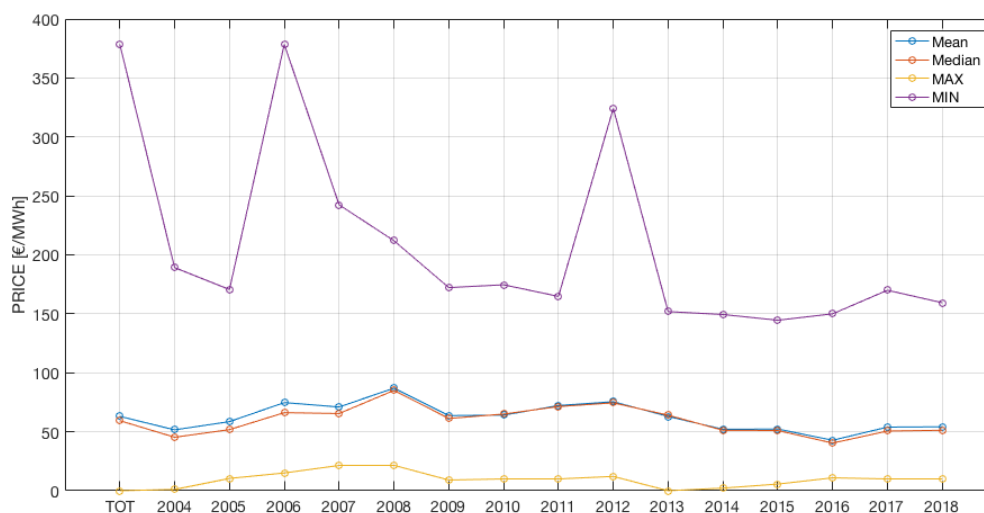


Figura 9 – MGP: Grandezze statistiche di interesse

Si nota come il valore adiacente superiore non comprenda mai il massimo valore della serie, invece non accade lo stesso per il valore adiacente inferiore che risulta comprendere anche il minimo della serie almeno fino al 2009, questo in generale ci fa capire che le serie sono “disperse” benché la metà dei valori sia comunque concentrata nella zona indicata dai box ed un altro 25% sono i valori inferiori al primo quartile, in sostanza i valori sono molto più dispersi nell’intervallo che va dal terzo quartile al massimo, in alcuni anni meno che in altri ma ad esempio nel 2017 i valori sopra al terzo quartile

ricoprono in intervallo più ampio del restante 75% dei dati. Per completezza vengono riassunte le principali grandezze che caratterizzano la serie, raccolti in una tabella riassuntiva in appendice A (pagina 132) che risulta un meno intuitiva dei grafici ma ne esplicita i valori esatti.

Da una nota di carattere matematico si appunta che il VAI deriva dal calcolo $VAI = Q1 - 1,5(Q3 - Q1)$, per cui se la distanza interquartile supera la quota $\frac{Q1}{1,5}$ può anche assumere valori negativi, il che non ha senso per la grandezza di interesse, perciò se il valore minimo è maggiore del VAI allora il esso coincide anche con il minimo di tutta la serie ma non è detto che sia sempre così, come accade dal 2010 in poi.

Per quanto riguarda il VAS valgono le stesse considerazioni, sarà pari al massimo solo se i dati sono compatti abbastanza da assicurare che $MAX < Q3 + 1,5 * (Q3 - Q1)$ e come si nota questo non accade mai, quindi ho sempre dei valori fuori limite, cioè che distano più di una volta e mezza la distanza interquartile dal terzo quartile.

Ponendo l'attenzione alla continuità della serie non ci sono problemi, anche se riferiti ad un periodo così vasto, i valori orari sono tutti disponibili e non presentano buchi di alcun genere, almeno per quanto riguarda la serie del PUN. Al fine della previsione è indispensabile utilizzare una serie storica che sia continua, così non si perdono informazioni utili ed è mantenuto costante lo step temporale fra valori successivi, orario in questo caso.

Quanto osservato prima vale anche per la singole zone di mercato, nel senso che i pochi problemi che si riscontrano non sono valori mancanti nella serie ma bensì valori estremamente elevati che quindi sono di molto fuori scala, tale fenomeno sul MGP si nota ad esempio sulla Corsica (sia AC che DC) in cui i prezzi hanno picchi di 3000 €/MWh isolati oppure anche per lunghi periodi, questo si deve al fatto che l'offerta interna alla zona non è sufficiente a soddisfare la domanda, all'opposto dei casi in cui si registra un surplus di offerta, qui i consumatori si trovano deficit di energia e tutte le offerte avanzate dai produttori vengono valorizzate al prezzo più alto possibile che è limitata da una soglia di sicurezza denominata VENF.

Le tre tabelle Tabella 23, Tabella 24 e Tabella 25 in appendice A (pagina 132) riportano le zone divise per categoria (vedi 2.1) in base al periodo sul quale sono state attive le negoziazioni su ogni zona, sul totale di trenta zone sono state sempre attive solamente quindici, tra cui tutte le zone geografiche più BRNN, PRGP e ROSN e la maggior parte delle zone virtuali estere (AUST, CORS, FRANC, GREC, SLOV e SVIZ). Le zone "non attive" dal 2009 in poi non lo sono comunque state tutte quante in blocco negli anni precedenti, PBNF e E_CO sono da considerare rispettivamente dal 2004 al 2006 e dal 2006 al 2008 ed E_SD a partire dal 2005. Sotto questo aspetto è utile considerare le sole serie complete, evidenziandone le differenze sulla base dello stesso numero di punti analizzati.

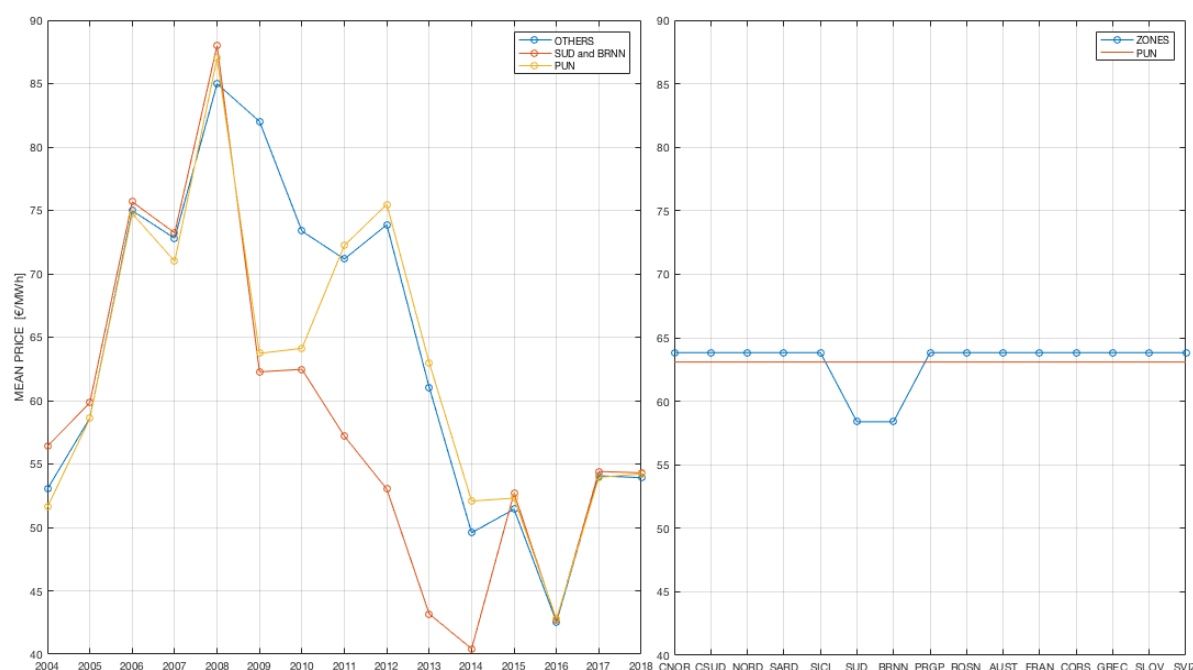


Figura 10 – Sinistra: Profili di prezzo medio annuale per zona, Destra: Valori medi 2004 -2018 zonali

Confrontando le zone di mercato a scala annuale emerge che solamente per il sud Italia e la zona virtuale estera del Brennero (SUD e BRNN) vi sia un'effettiva differenza di prezzi, i profili orari sono del tutto identici e quindi si ragiona con i valori medi annuali. La Figura 10 illustra le differenze tra le curve principali e quella che contribuisce al PUN, è curioso osservare che fino al 2008 le tre curve hanno stato molto simili ma da lì in poi SUD e BRNN hanno registrato prezzi mediamente più contenuti fino al 2015 da cui le curve sono tornate ad essere quasi sovrapposte.

Essendo il PUN costruito sulla base dei prezzi zonali ponderati sulle quantità, la curva gialla indica che fino al 2008 non vi erano grosse differenze e quindi il PUN assumeva valori vicini, invece nel corso del 2009 e 2010 il PUN segue il trend di SUD e BRNN per poi allinearsi con le restanti zone, questo suggerisce che le quantità scambiate sulle zone SUD e BRNN fossero una buona parte del totale e quindi la media pesata converge a quel valore, fenomeno che si inverte dal 2011 dove prevalgono gli scambi di energia sulle restanti tredici zone di mercato.

Il grafico di destra riporta le medie zonali delle serie complete in confronto alla stessa grandezza del PUN, quest'ultima è di poco inferiore ai valori zonali esclusi SUD e BRNN, i quali sono in ribasso di quasi un decina di euro sull'unità prodotta, quindi la depressione 2009-14 di SUD e BRNN incide parecchio sul valore medio, appunto tocca differenze superiori ai venti $\frac{\text{€}}{\text{MWh}}$ come nel 2012. La retta del PUN sul grafico di destra suggerisce che il contributo più influente riguardi le restanti tredici zone, il che è vero globalmente ma non anno per anno, come evidenziato dalla curva gialla sul plot di sinistra.

2.2.1. Confronti

In questo paragrafo si svolge qualche confronto su scale temporali differenti, il fine è quello di raccogliere informazioni utili riguardo la stagionalità della serie quindi la dipendenza da fattori esterni quali la temperatura (approfondito alla sezione 2.2.1.5), i dati di riferimento rimangono quelli del MGP soffermandosi sulla serie del PUN. L'analisi viene svolta su un solo anno, il **2016**, lo stesso che viene considerato nel paragrafo 2.3, che riguarda il mercato infra-giornaliero, viene poi anche accennato qualche riferimento alle altre annate.

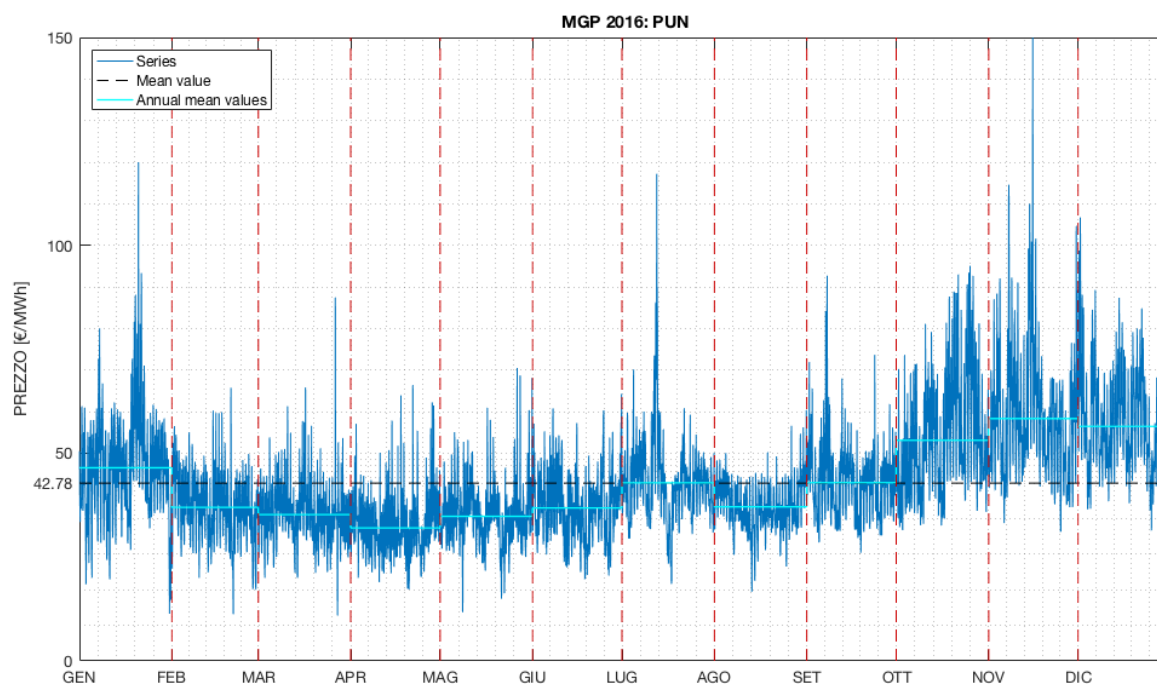


Figura 11 - MGP, Anno 2016 : Prezzo Unitario Nazionale

Il grafico sopra illustra il PUN per l'intero 2016, l'asse delle ascisse riporta i vari mesi secondo la suddivisione delle linee verticali rosse, invece le ordinate indicano il prezzo giornaliero più la stessa quantità a scala mensile (linea spezzata color ciano). Bisogna ricordare che il 2016 registra il valore medio più basso, come evidenziato in Figura 6 in generale gli ultimi anni riportano valori modesti e nello specifico il 2016, per cui si considera proprio quest'ultimo su entrambi i mercati a pronti (MI e MGP), in seconda sede è interessante studiare le fluttuazioni mensili di anno in anno.

Il trend è ben delineato, nei mesi a cavallo fra autunno e inizio inverno il prezzo rimane stabile su prezzi alti con il massimo che cade a su metà Novembre per poi andare in diminuendo in primavera, a Luglio vi è un picco confrontabile con i massimi Gennaio e di conseguenza anche la media presenta un lieve aumento che si sostiene per poco tempo al contrario di quanto accade d'inverno, in più a Luglio e Settembre la media mensile è pari a quella annuale. Per questa si evita il boxplot perché la situazione è la medesima, distribuzione molto più dispersa sopra al terzo quartile e mediana pari a circa $40,4 \frac{\text{€}}{\text{MWh}}$ che rimane di poche unità sotto la media, per come sono definiti VAS e VAI svariati valori risultano fuori scala nella parte alta.

2.2.1.2. Confronto settimanale

Ogni considerazione a seguire continua a valere per serie del PUN e per il 2016 solamente, si confrontano settimane intere comprensive di feriali e festivi per stagioni diverse quali estate e inverno. Comunque il fine ultimo è iniziare a distinguere trend giornaliero e settimanale e l'influenza della stagionalità su di essi. Iniziamo confrontando la prima settimana di Gennaio e Luglio che cadono entro le stesse date essendo l'anno bisestile: dal 4 al 10 del mese, la scelta è data dal fatto che i valori medi mensili sono il massimo estivo ed il minimo del plateau invernale ed entrambi mesi presentano un sfogo anomalo della stessa entità.

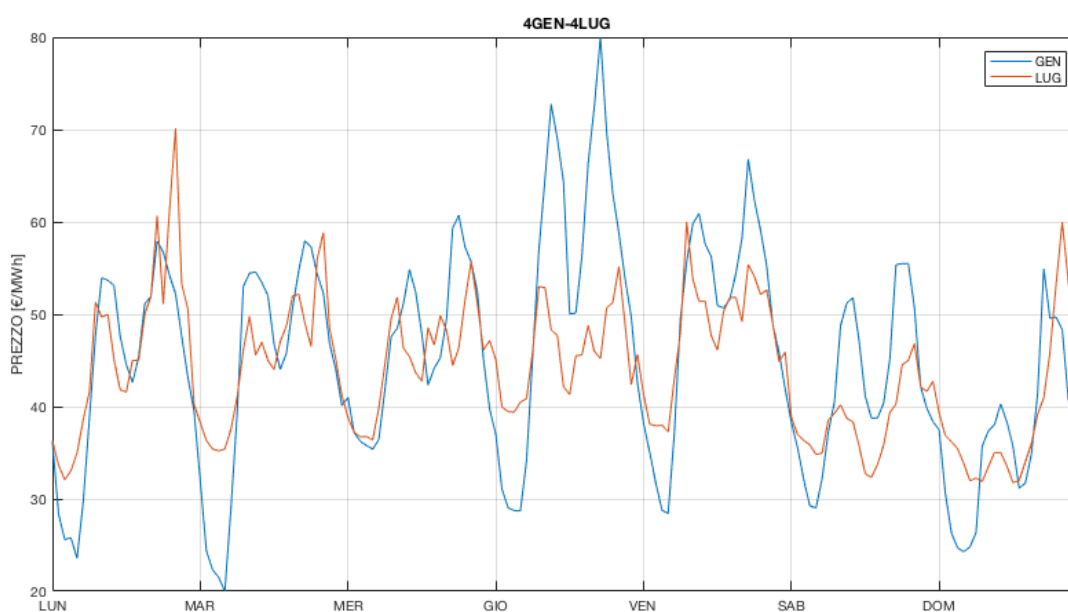


Figura 12 – Confronto prima settimana completa di gennaio-luglio

Un appunto da non tralasciare assolutamente sta nel fatto che il periodo di gennaio è esattamente a cavallo tra della ripresa lavorativa di giovedì 7 gennaio, infatti i picchi in quella giornata sono più ampi e di conseguenza lo sarà anche il prezzo medio, perciò il **festivo infrasettimanale** influisce in questo senso. Comunque la figura sopra evidenzia una caratteristica importante, la curva azzurra segue due picchi ben definiti a fine mattinata e metà pomeriggio con prevalenza del secondo mentre la curva arancio è frastagliata e nei giorni centrali la differenza fra pomeriggio e mattino è quasi nulla, anzi il venerdì è invertita. Notare bene però che la differenza fra giorni feriali e festivi è accentuata su ambedue le curve, soprattutto la domenica vi è un grosso divario fra i massimi locali con prevalenza verso sera, il sabato invece non è un festivo tutti gli effetti e tale fenomeno è poco marcato, ad ogni modo il week end l'energia elettrica è più conveniente in confronto al resto della settimana. E' interessante notare il trend di diminuzione dell'ampiezza delle oscillazioni, mediamente il prezzo nella settimana estiva si mantiene più stabile attorno al valor medio, quest'ultimi non sono indicati a grafico ma contano 45 e 44 $\frac{\text{€}}{\text{MWh}}$ rispettivamente a gennaio e luglio, almeno per la prima settimana non vi sono differenze sostanziali. Come da Figura 17 vi è un valore anomalo (circa 120 $\frac{\text{€}}{\text{MWh}}$) che cade nel pomeriggio del 12

gennaio e 20 luglio, tralasciando le due settimane centrali si cerca conferma nell'ultima settimana, dove gli unici feriali sono i fine settimana.

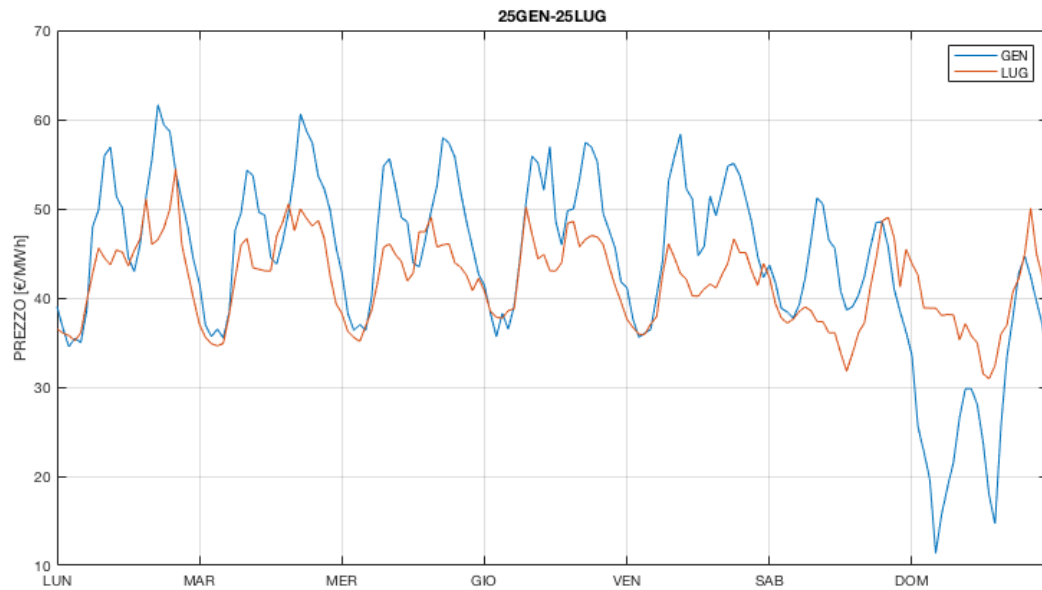


Figura 13 - Confronto ultima settimana completa di gennaio-luglio

In generale quanto scritto per la prima settimana si riporta anche all'ultima, i valori sono più compatti a luglio e i feriali si assentano su profili ripetitivi che però sono meno "lisci" rispetto la prima di gennaio, il sabato e la domenica di luglio praticamente mostrano solo il picco pomeridiano al contrario di gennaio dove rimangono entrambi, conferma anche prezzi medi minori il fine settimana e in particolare la domenica di gennaio. La distribuzione dei dati in ambo i casi è meglio apprezzabile dal boxplot a seguire, si vede bene come anche la mediana praticamente non varia la prima settimana, di poche unità inferiore la seconda.

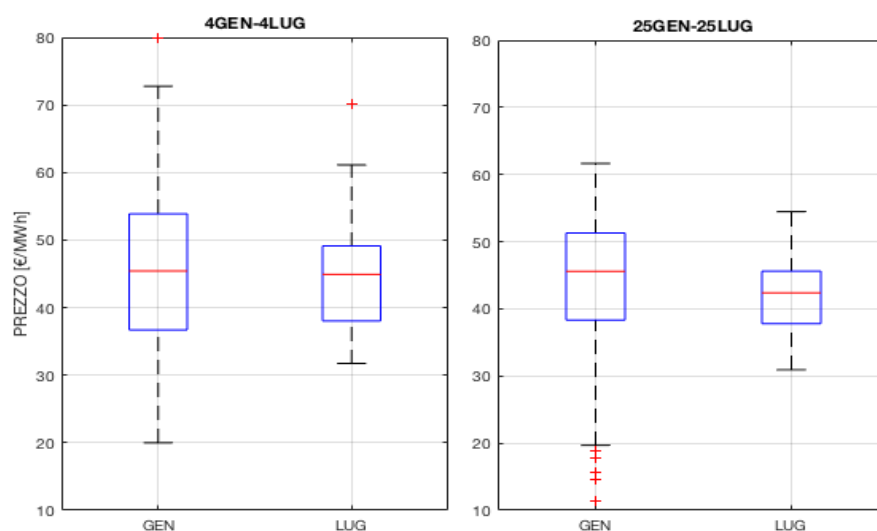


Figura 14 – Boxplot, Sinistra: prima settimana completa del mese, Destra: ultima

Le seconda e la terza settimana includono rispettivamente anomalie su un paio di giorni a luglio e sulla quasi totalità della settimana a gennaio, per questo la media mensile di quest'ultimo è superiore di 4

unità sebbene settimanalmente non sia sempre così, anzi la seconda settimana di luglio è stata più dispendiosa in media.

Espandendo l'analisi delle distribuzioni a tutti gli anni non si trova conferma a Figura 14, in quanto si può trovare un riscontro positivo sul 2011/12/13/14/15 dove si hanno valori più compatti a luglio ma qualche differenza sulla mediana e sui quartili, invece per il 2005/06/07 si ha un trend esattamente opposto ed i valori sono più dispersi a luglio mentre per i restanti anni non si notano grosse differenze. Riguardo ai valori medi mensili dell'intero periodo si faccia riferimento al paragrafo 2.2.1.5 così come al 2.2.1.4 per quelli giornalieri.

Ad ora non si è ancora riscontrata un vero e proprio trend stagionale, è stata analizzata la dispersione dei dati e i profili settimanali in due mesi campione e relativa influenza dei festivi infrasettimanali. A questo punto viene utile un approccio diverso, conviene analizzare le singole stagioni (dal calendario) calcolando il profilo della settimana media per ogni stagione, la Figura 15 riconduce a tale confronto dove si costruisce la settimana media partendo dalle settimane complete per ogni stagione del 2016 che sono 13, 12, e 11 rispettivamente per primavera, estate con autunno e inverno.

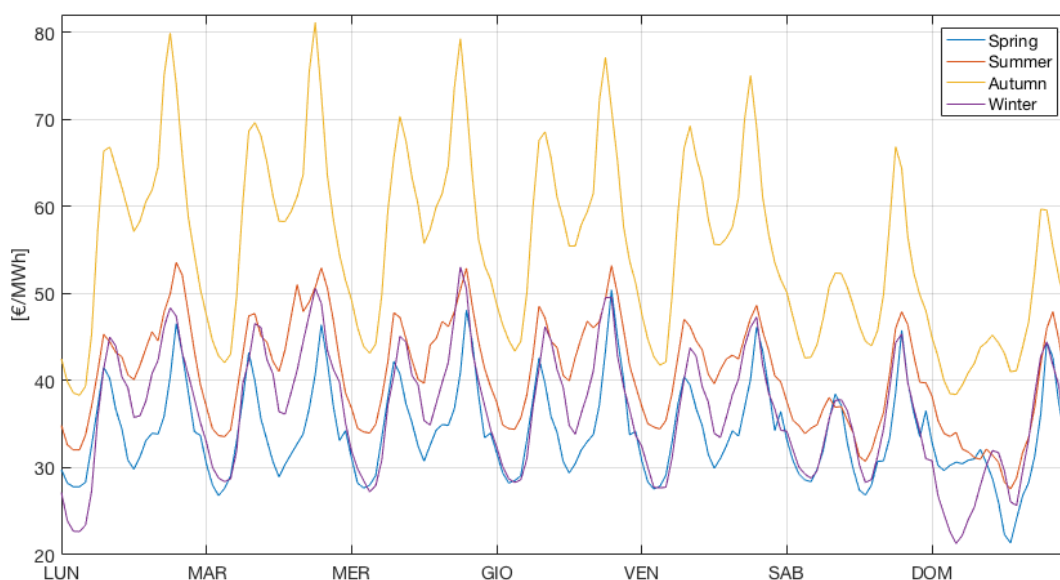


Figura 15 –Settimana media per stagione

Appare molto utile l'ultimo confronto, è veramente apprezzabile solamente la differenza tra autunno e le altre stagioni, a primo impatto appariva che anche i valori invernali fossero mediamente più alti ma il contributo di febbraio e marzo è pesa parecchio sul totale. Escluso l'autunno i profili rimangono molto simili sia come valori che dispersione, prevale il picco pomeridiano e da lunedì venerdì sono le curve sono ripetitive in leggero diminuendo e tutto ciò vale anche per l'autunno, semplicemente tutto traslato a valori superiori.

Quindi, il trend settimanale è univo senza dubbio, a meno dei giorni festivi il feriale medio presenta due massimi locali il cui primo si riduce drasticamente la domenica e poco di meno il sabato. Si investiga ulteriormente la stagionalità in Figura 16, per ogni settimana si mediano i valori 1-24h e poi ancora su tutte le settimane.

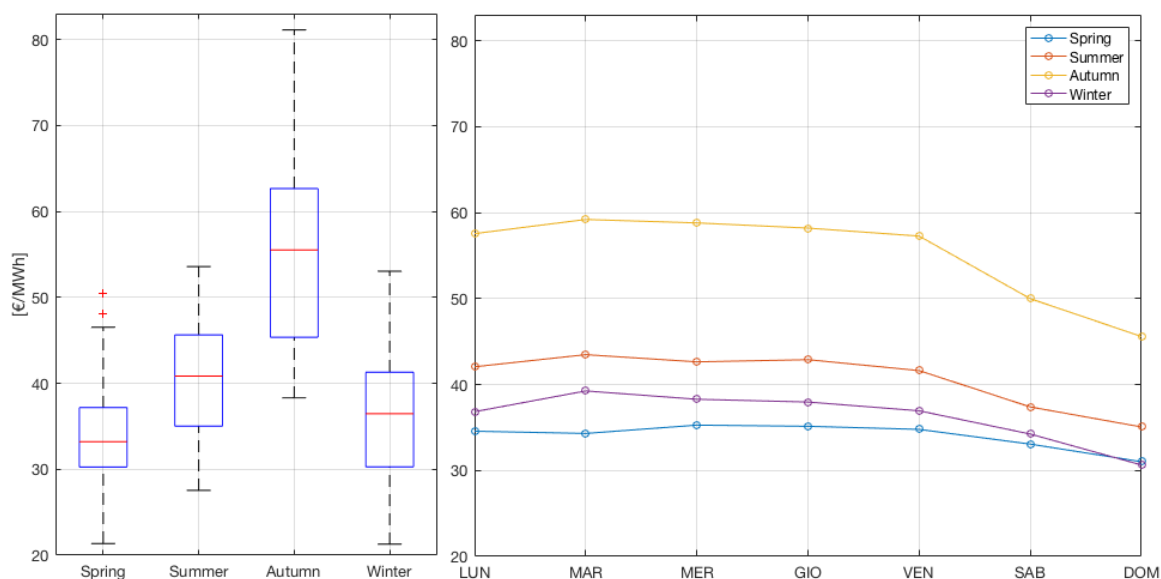


Figura 16 – Sinistra: boxplot delle distribuzioni per la settimana media, Destra: valori medi giornalieri della settimana media

Non solo l'autunno ha mediana alquanto maggiore, ma è anche la distribuzione più dispersa con distanza interquartile maggiore e di conseguenza non presenta valori fuori limite come anche estate e inverno. Le settimane estiva e primaverile sono le più compatte e quindi il prezzo si mantiene attorno la media più di quanto non accada altrove. Secondo il grafico di destra la primavera è la stagione più conveniente, seguita da inverno, estate e autunno con un notevole distacco.

Facendo cenno alla simmetria è interessante come sia più marcata a livello settimanale, salendo su scale temporali più lunghe quale l'anno, esso comprende valori dispersi ma a scala settimanale accade di rado, ciò specifica che nel corso di una settimana è difficile vi siano profili giornalieri molto diversi, però è probabile fra settimane diverse e ancora di più ad intervallo mensile

2.2.1.3. Confronto mensile

Continuando il discorso introdotto con Figura 11, il confronto settimanale è stato redatto sulla base delle anomalie locali presenti nei mesi di Luglio e Gennaio, in particolare i valori di massimo. Ora viene introdotta una visione globale dei trend mensili nella figura a seguire, dove sono rappresentate media, mediana, i valori di massimo e di minimo più i quartili. Troviamo conferma che i valori massimi dei mesi sopra citati siano anomali, benché i valori medi siano comunque in rialzo essi sono molto più tenui, perciò i prezzi sono maggiori di inverno e a Luglio, con valori isolati molto maggiori che derivano da circostanze specifiche relative al giorno di mercato. Le singole sezioni verticali di Figura 17 sono analoghe ad un diagramma a scatola a meno dei valori fuori limite, per cui la fascia compresa tra le curve verde e azzurra è lo scarto interquartile che copre al massimo il 27% dell'intervallo (16% in media), mentre i valori dal terzo quartile in su arrivano al 72% (59% in media) ed essi sono solo la metà di quelli compresi entro lo scarto interquartile

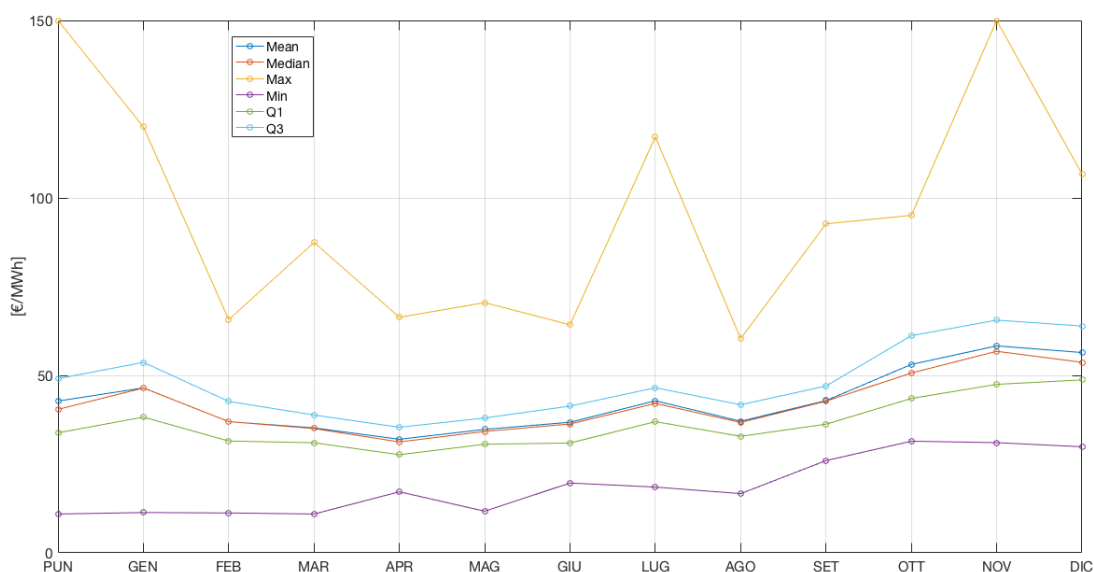


Figura 17 – MGP: PUN 2016, grandezze statistiche di riferimento su scala mensile.

Distribuzioni di anni diversi esibiscono tratti in comune con il 2016 dal punto di vista globale ma è complicato associare una logica riguardo i mesi che mostrano valori anomali, il trend è sempre in aumento in inverno, a prescindere dall'anno, l'entità varia di anno in anno e anche la posizione del plateau che può essere traslato di qualche mese. Invece la situazione nei mesi estivi è casuale, ci sono mesi centrali con piccoli aumenti si prezzo (come Luglio 2016) che nella maggior parte degli anni cade su Luglio-Agosto ma altri, quali 2013 e 2014, anche in primavera e inizio estate.

2.2.1.4. Statistiche medie giornaliere

In questa sezione si va a confrontare l'evoluzione temporale, su scala annuale, delle principali medie giornaliere sull'intero anno; è interessante analizzare la correlazione tra giorni feriali e festivi per capire come ne è influenzato il prezzo dell'energia elettrica. Nei giorni festivi ci si aspettano profili di carico diversi, sia come valore dei picchi che nella posizione degli stessi e di conseguenza anche il prezzo dell'energia elettrica ne sarà direttamente influenzato, per tale analisi si sono considerati solamente gli anni di cui sono disponibili la totalità dei dati orari, cioè dal **2005 al 2017**. Quindi, **da questo punto in poi l'analisi non è più mirata solamente al 2016** come quanto svolto nelle sezioni 2.2.1.2 e 2.2.1.3 ma si espande sull'intero orizzonte temporale. Si è pensato di analizzare singolarmente ogni giorno della settimana mediato sull'anno, a seguito di un confronto fra giorni feriali e non. Quindi, sono stati analizzati singolarmente tutti e sette i giorni della settimana da cui si nota che :

- I giorni **lavorativi** presentano tutti lo stesso trend, indistintamente da giorno della settimana. Nel dettaglio due picchi (metà mattina e pomeriggio) con un minimo la notte ed un plateau di valore intermedio nel periodo che intercorre fra i picchi. Come già ribadito il trend non cambia e neanche i valori nel corso dello stesso anno (non in maniera repentina) , la differenza si apprezza su scala annuale soprattutto sull'entità delle fluttuazioni e in alcuni casi i picchi sono traslati di poco nel tempo. Osservato ciò si è deciso di considerare un'unica statistica con tutti i giorni feriali dell'anno, facendo attenzione alle festività infrasettimanali.

- I giorni **festivi** sono un po' meno standard, in quanto l'unico vero trend stabile è rappresentato dalla domenica in cui il picco mattutino appare lieve e molto “disteso” mentre il secondo rimane più marcato ma comunque contenuto evidenziando che è ancora l'unico vero giorno che si possa considerare festivo, invece per il sabato tutto è più complicato e confuso siccome è difficile includerlo fermamente in una delle categorie. Per quanto riguarda i festivi infrasettimanali (nazionali) invece sono più simili alla domenica ma sono pochi. A conclusione si considera la statistica riferita alle sole domeniche di ogni anno per i 13 considerati.

In generale, per entrambi vale la seguente banale formulazione:

$$\overline{h} = \frac{1}{N} \sum_{i=1}^N \begin{pmatrix} h_1 \\ \dots \\ h_{24} \end{pmatrix}_i = \begin{pmatrix} \overline{h_1} \\ \dots \\ \overline{h_{24}} \end{pmatrix} \text{ con } N = \text{numero di giorni feriali} \quad (9)$$

Siccome i dati sono riferiti ad un periodo di tredici anni, è utile rielaborare i valori delle curve in riferimento al loro valore medio per poter apprezzare le differenze su uno stesso plot:

$$m = \frac{1}{24} \sum_{i=1}^{24} \overline{h_i} \quad (10)$$

Per questo, oltre alle curve medie (fig.8 e 12) si riportano i grafici che riguardano:

- Valori orari **normalizzati** sulla media giornaliera (fig.9 e 13): per sottolineare le distanze in termini relativi al valor medio.

$$\overline{h_N} = \begin{pmatrix} \overline{h_1} \\ \dots \\ \overline{h_{24}} \end{pmatrix} * \frac{1}{m} \quad (11)$$

- Valori orari **scalati** della stessa : per visualizzare gli scarti quantitativi (positivi e negativi) dei singoli valori orari.

$$\overline{h_S} = \begin{pmatrix} \overline{h_1} \\ \dots \\ \overline{h_{24}} \end{pmatrix} - m \quad (12)$$

Come analisi ulteriore si considera la dipendenza del prezzo dell'energia elettrica da quello del **petrolio** (Figura 11) , il che è significativo solo per i giorni feriali in quanto il mercato petrolifero è aperto solamente Lunedì – Venerdì, quindi la domenica non ne dipende affatto. Per fare ciò si utilizzano i prezzi del riferimento europeo : il Brent. I prezzi sono giornalieri ed è bene che la normalizzazione sia

eseguita giorno per giorno e successivamente i valori normalizzati singolarmente vengano mediati per ottenere una curva giornaliera rappresentativa di tutto l'anno.

$$h_p = \begin{pmatrix} h_1 \\ \dots \\ h_{24} \end{pmatrix} * \frac{1}{m_p} \text{ con } m_p = \text{media giornaliera Brent} \left[\frac{\$}{bbl} \right] \rightarrow \overline{h_p} = \frac{1}{N_p} \sum_{i=1}^{N_p} h_{p_i} \quad (13)$$

Non per tutti i feriali italiani è disponibile il prezzo del Brent, perché i due mercati non sono necessariamente sincronizzati, per ciò il procedimento descritto sopra si è applicato ai soli giorni feriali in cui fosse disponibile il prezzo medio giornaliero del Brent.

In Figura 21 i valori sono normalizzati sui prezzi del Brent riportati al valore in euro grazie al tasso di cambio euro-dollaro dello stesso giorno, quindi si fa riferimento alla serie storica¹³ e vengono calcolati i prezzi in euro giorno per giorno e solamente per i giorni che ci interessano. Questo è un passaggio molto importante per tenere conto delle dinamiche di ambedue i mercati Europeo e Americano, il greggio viene commerciato in dollari ed i fattori che contribuiscono alla formazione tasso giornaliero sono molteplici in ambiti molto diversi, perciò differenze minime su prezzi in dollari possono essere percepite in maniera drasticamente diversa sul mercato europeo, in base al periodo e alle circostanze di mercato.

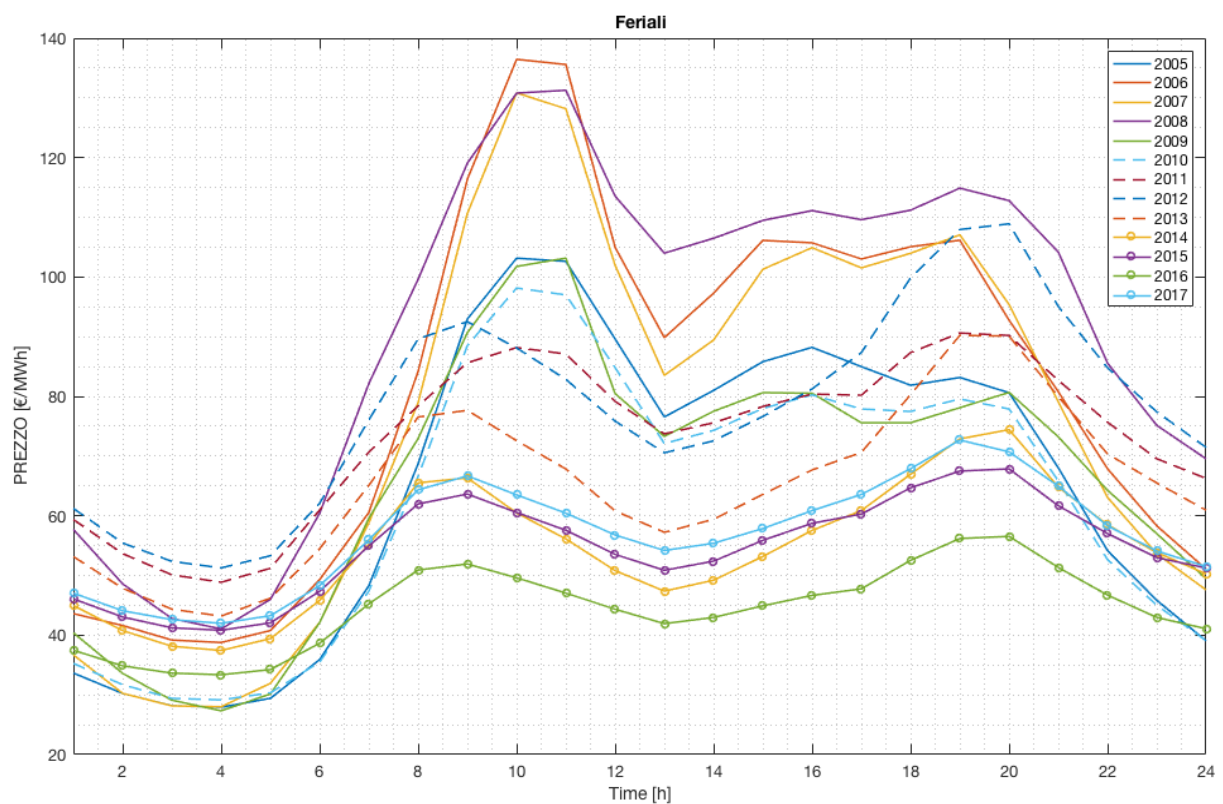


Figura 18 - Feriale medio

¹³ <https://it.investing.com/currencies/eur-usd-historical-data>

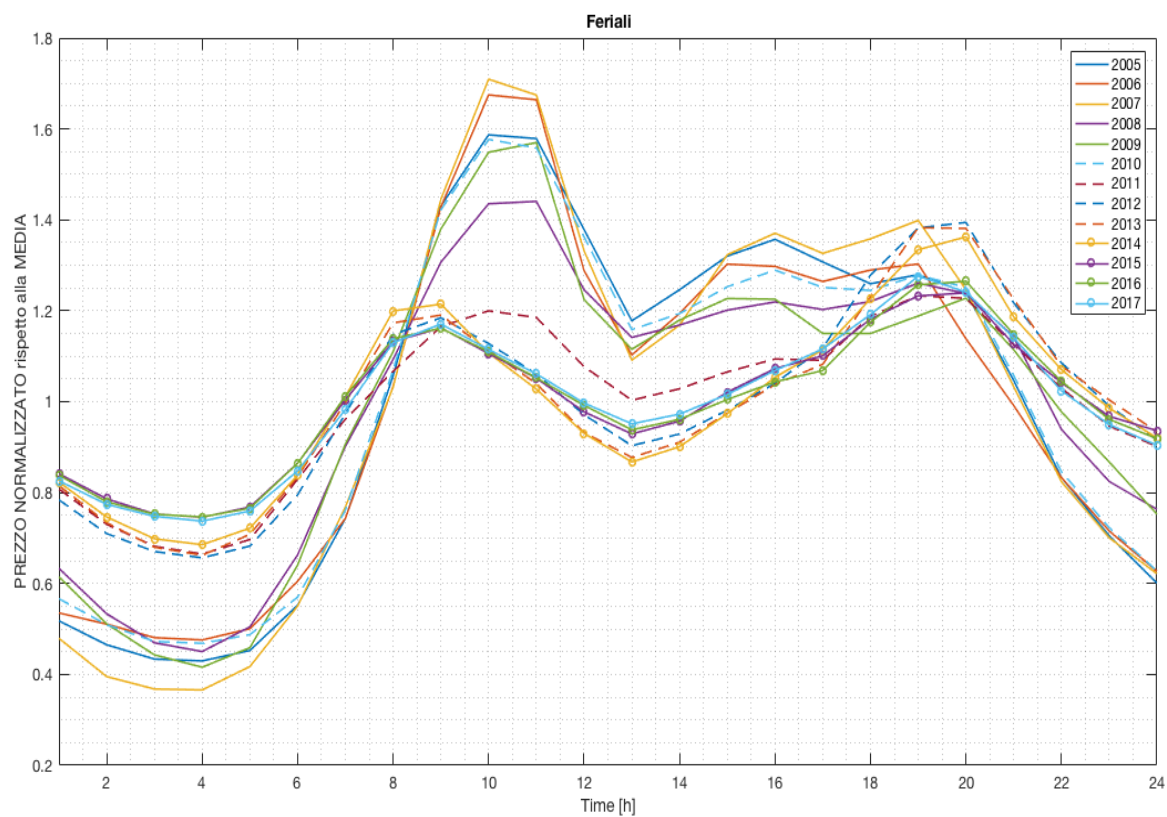


Figura 19 – Feriale medio normalizzato sulla media giornaliera

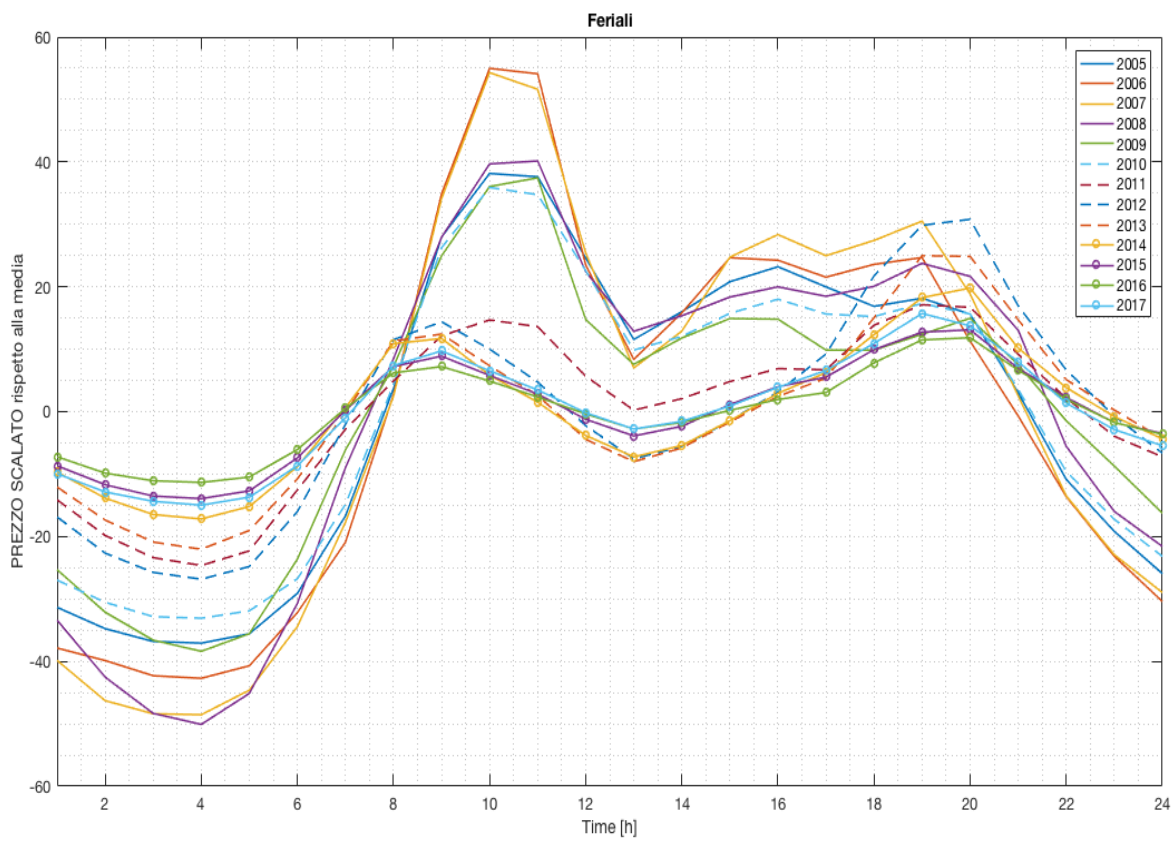


Figura 20 - Feriale medio scalato sulla media giornaliera

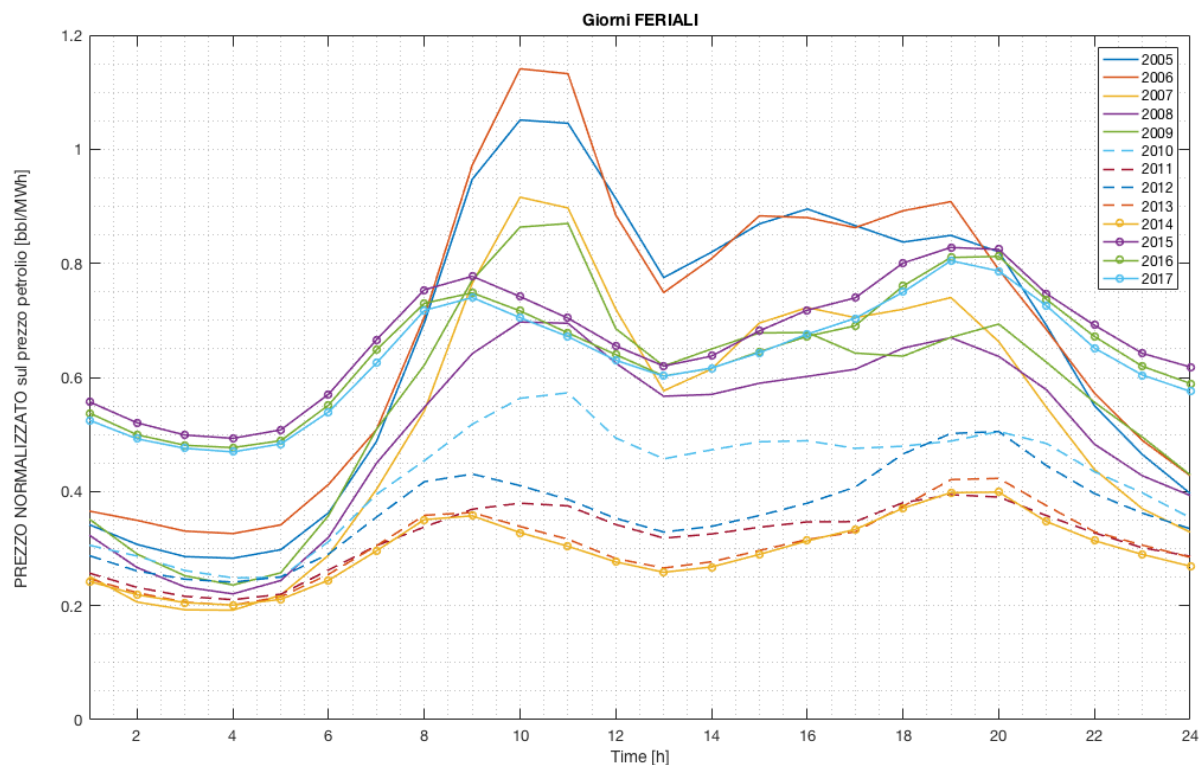


Figura 21 - Feriale medio normalizzato sul prezzo del petrolio riportato in [€/bbl]

Sono individuabili tre fasce temporali in base allo stile della linea :

- **2005 al 2009** : il primo fattore interessante è l'andamento sempre crescente che si nota da Figura 18 con massimo al ed un crollo repentino nel 2009 per ritornare sugli ordini di grandezza del 2005, è interessante notare in Figura 19 come il 2008 presenti un valore medio più alto rendendo la curva normalizzata meno variabile rispetto agli altri anni e questo significa che mediamente il 2008 è stato l'anno più caro dal punto di vista dell'energia, e non solo l'energia elettrica ma tutto ciò è fortemente legato al prezzo del petrolio, in seguito allo stop sulla produzione che si è verificato in quel periodo, ha subito un'impennata per poi precipitare drasticamente. L'influenza del prezzo del petrolio è confermata dalla Figura 21 in quanto la curva è normalizzata su prezzi giornalieri mediamente più elevati (Figura 46), quindi l'influenza di una variabile come il prezzo del combustibile è senza dubbio da considerare ed approfondire.
- **2010 al 2013** : In questo periodo si riscontra un'inversione dell'entità dei picchi, in precedenza il picco più importante era quello mattutino mentre nel pomeriggio non vi era un vero e proprio picco marcato ma piuttosto un leggero aumento. Invece dal 2010 al 2012 si registra che: prima i due picchi divengono confrontabili e poi si invertono, in questo caso la normalizzazione non comporta modifiche ai trend e nemmeno sui valori perciò le curve presentano verosimilmente un valore medio confrontabile. La colpa è da attribuire alla massiccia installazione di potenza da fotovoltaico a seguito degli ingenti incentivi statali al 2010 che negli anni subito a seguire hanno portato ad un surplus sulla produzione (e quindi aumento dell'offerta) nelle ore centrali della giornata bensì la domanda rimanesse invariata. Inevitabilmente, così facendo, il prezzo di equilibrio in quelle ore si

abbassa notevolmente, almeno fin quando vi è sole, una volta passato l'offerta crolla ed il prezzo aumenta portando ad un prezzo più elevato in corrispondenza del tardo pomeriggio.

- **2014 al 2017** : dal 2014 qui in poi il range di variabilità del prezzo risulta molto più contenuto ed il prezzo in se è notevolmente più basso (Figura 18), dalla Figura 19 si vede bene come l'andamento sia perfettamente coerente e le oscillazioni contenute entro i 20% in ambedue le direzioni, il minimo si registra nel 2016. E' utile appuntare una considerazione riguardo il 2014, per tutto il periodo che va dal 2011 al 2014 il prezzo del prezzo del petrolio è sempre rimasto assestato attorno ai 100 / 120 €/bbl mentre il prezzo dell'energia elettrica ha subito un abbattimento non trascurabile.

Soffermandosi nel dettaglio sulla Figura 21 è molto interessante che nel primo periodo (2005-2009) il prezzo dell'energia elettrica era estremamente più variabile se rapportato al prezzo del Brent, per esempio nel 2006 si varia dal 25% al 145% nel corso della sola mattinata ed ancora di più ne 2017. Invece durante gli ultimi due periodi (2010-2017) le curve normalizzate sul petrolio stanno sempre sotto l'unità, nel 2008 si nota ancora una certa variabilità sul 30-100% per poi stabilizzarsi da lì in avanti. Appare difficile generalizzare un legame fra le due variabili, le fluttuazioni dell'una o dell'altra sono difficili da comprendere a fondo ed anche complicati da estrapolare dai grafici precedenti, al fine di recuperare qualche informazione di carattere generale si riporta l'andamento della media delle curve in Figura 18.

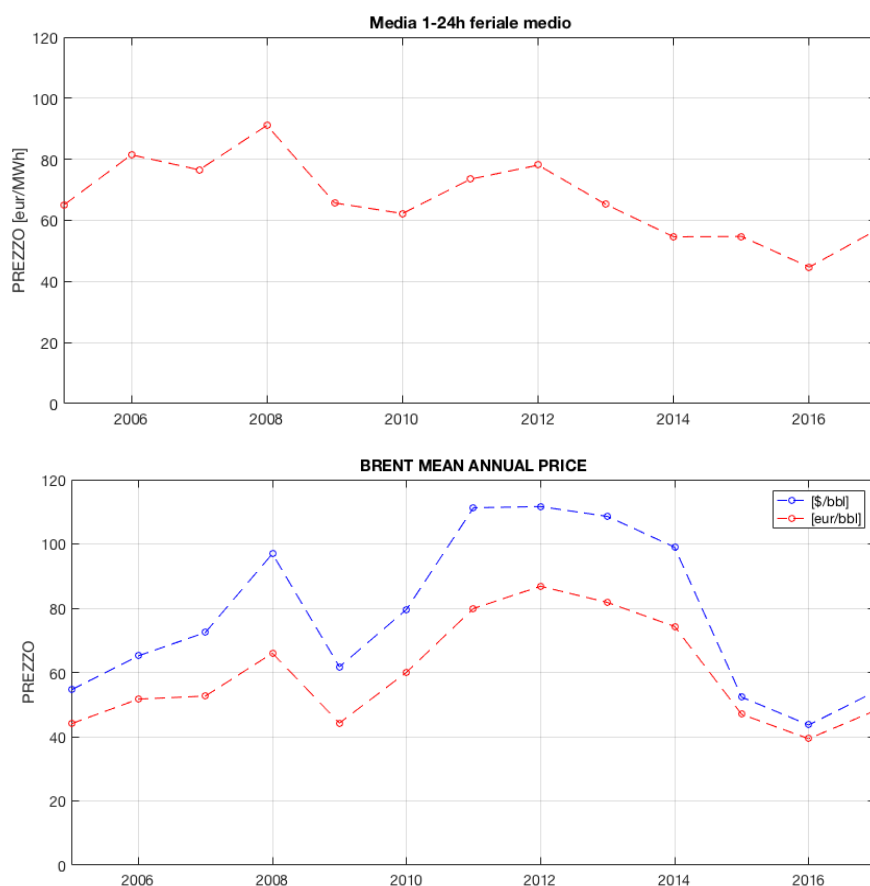


Figura 22 –sopra: media giornaliera del feriale per anno; sotto: prezzo del petrolio Brent in valute diverse

Grazie a Figura 22 si ha una visione più chiara e immediata della situazione riguardo il legame tra prezzo dell'energia elettrica e prezzo del petrolio, ovviamente il tutto in termini di valori medi. In Figura 22 è mostrato l'andamento del prezzo del petrolio riportato ai rispettivi valori in Euro al barile tenendo conto dei tassi di cambio euro/dollaro mediati su tutto l'anno. Il primo periodo è in accordo fra i due trend, dal 2010 in avanti invece la situazione cambia, si ha comunque un andamento di massima che rimane confrontabile ma l'entità degli aumenti/riduzioni sono più contenute, la massima discrepanza corrisponde al 2014 come era evidente già in Figura 21. Quanto analizzato in precedenza mostra un forte legame tra prezzo dell'energia, prezzo del petrolio e energia prodotta dal fonte rinnovabile non controllabile quale il fotovoltaico. Durante il primo periodo si ha una spiccata dipendenza dal prezzo del petrolio, invece nel secondo ne risente di meno, causa la massiccia installazione di impianti fotovoltaici che oltre ad essere i responsabili inversione dei picchi sul trend giornaliero hanno anche avuto un notevole impatto sul mercato elettrico, per non parlare dei problemi relativi alla frequenza di rete dovuti al gran numero di armoniche immesse in rete dagli inverter.

Passiamo adesso all'analisi della giornata festiva media sul mercato del giorno prima Italiano, si fa sempre riferimento al prezzo unitario nazionale (PUN) e ci si sofferma sulla domenica, il sabato non viene considerato a causa del fatto che non è completamente considerabile come festivo, presenta trend confusi e sovrapposti che lo rendono di poco interesse.

Vengono considerati i tre intervalli visti sopra, una caratteristica comune a tutte le curve sta nei picchi, prevale sempre quello pomeridiano e mano a mano che si avanza quello di metà mattinata si attenua sempre più, ad ogni modo il picco mattutino della domenica è traslato in avanti verso tarda mattinata ma soprattutto dura più a lungo ed i valori sono discretamente inferiori.

- **2005 al 2009** : inizia con un valore piuttosto basso (55 €/MWh di massimo) per poi impennarsi al 2006/2007 (80 €/MWh di massimo) e ancora toccando un massimo di 115 €/MWh attorno alle ore 20 nel 2008, quindi anche la domenica l'anno più dispendioso è stato il 2008, il che è curioso in quanto ci si aspetta che ciò sia dovuto una volta ancora al mercato petrolifero che sebbene chiuso la domenica continua a fare sentire il suo pesante effetto.

Ad ogni modo la differenza fra feriali e domenica è marcata nel senso che mediamente la domenica il prezzo scende come evidenziato in Figura 26 ad eccezione del 2013; se si confrontano Figura 24 con Figura 19, la domenica le fluttuazioni hanno toccato il 160 % rispetto alla media contro il 130 % dei feriali, in termini di scarto assoluto al 2014 tali percentuali corrispondono a 30 e 20 €/MWh rispettivamente (Figura 25 e Figura 20) ma da questo punto di vista non corrispondono allo scostamento maggiore.

- **2010 al 2013** : non appare influente la questione del fotovoltaico descritta sopra, almeno non in modo così esplicito come nei feriali, il picco pomeridiano è comunque in aumento con valori simili fra 2012 e 2013, la differenza tra quest'ultimi sta nelle oscillazioni mostrate in Figura 24 dove si vede che mediamente in una domenica mattina del 2013 il prezzo oscillava molto di più rispetto all'anno precedente ed è qui che si trova una spiegazione dell'anomalia mostrata in Figura 26 perché oltre ad essere più volatile il prezzo era comunque maggiore (simile al 2008).

- **2014 al 2017** : situazione stabile per quest'ultimo periodo, trend che presenta un solo e ampio picco nel tardo pomeriggio , valori minimi nel 2016 e massima variabilità al 2014. Il primo picco è ancora distinguibile nel 2014 per poi smorzarsi successivamente , infatti osservando le curve normalizzate (Figura 24) tutte, eccetto il 2014, sono per lo più sovrapposte rendendo la media un buon indice per rappresentare un confronto tra tutte le curve.

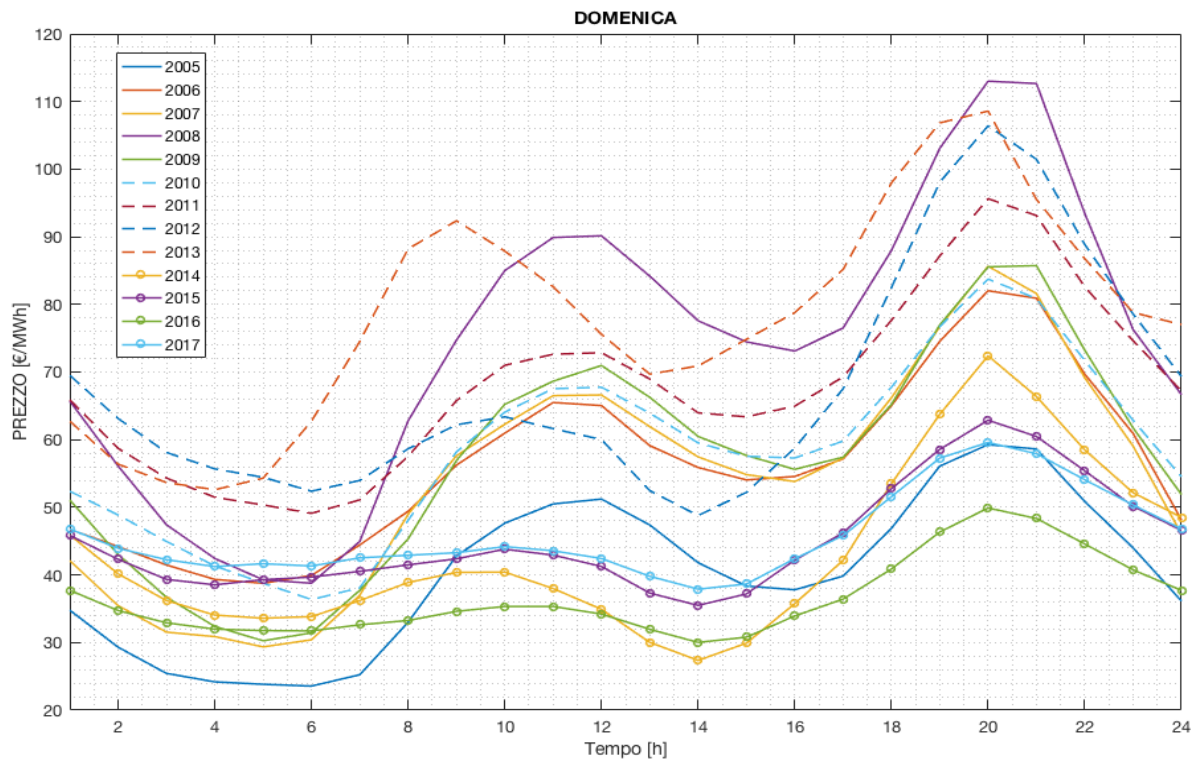


Figura 23 – Domenica media

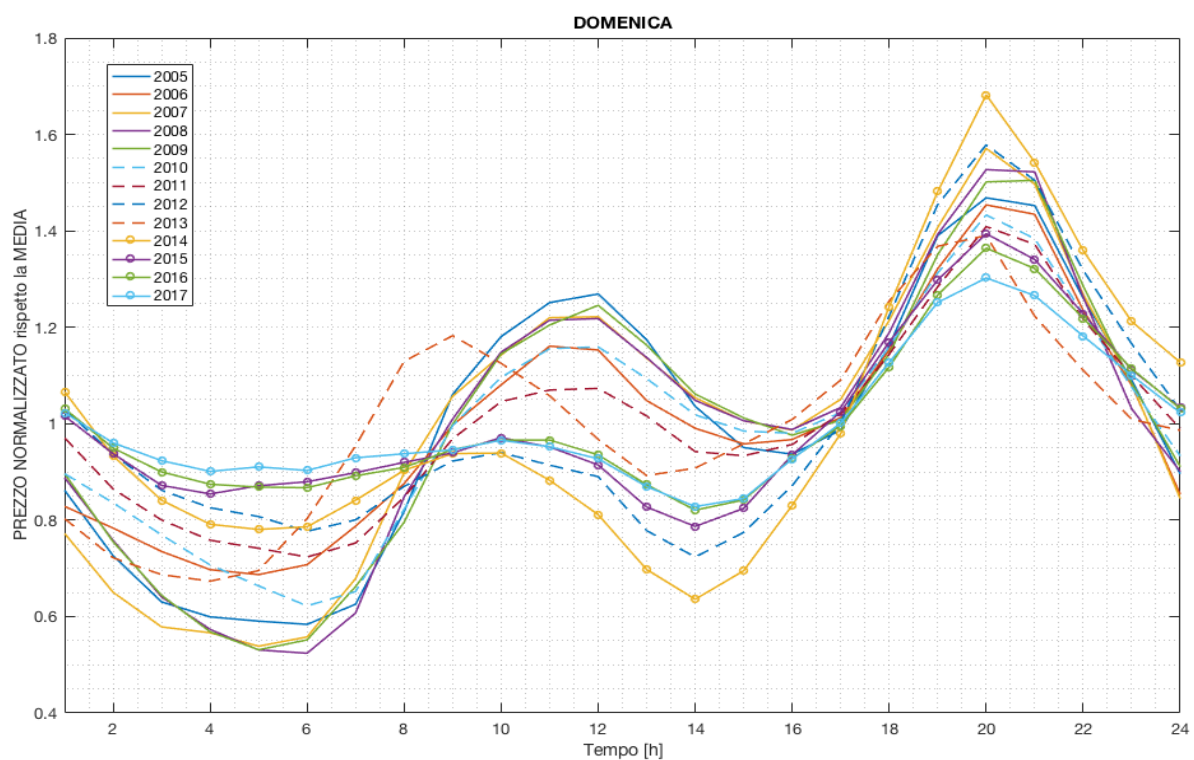


Figura 24 - Domenica media normalizzata rispetto la media giornaliera

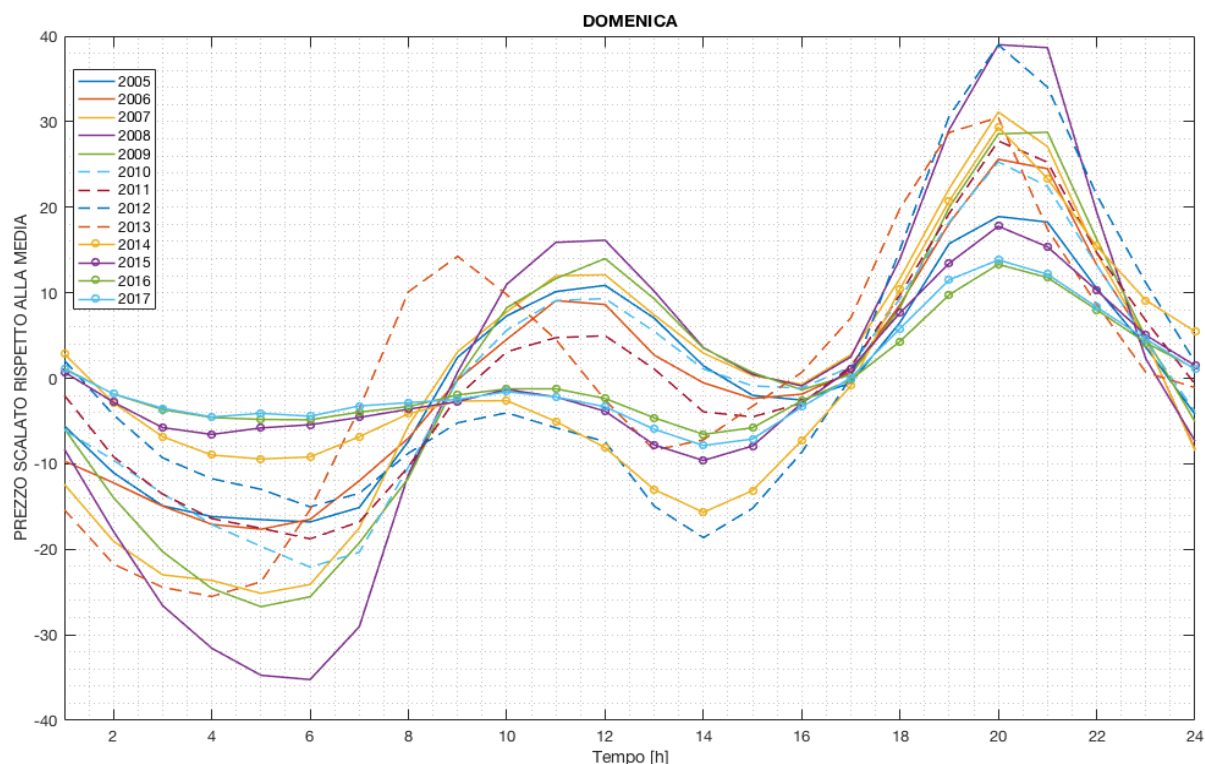


Figura 25 - Domenica media scalata rispetto la media giornaliera

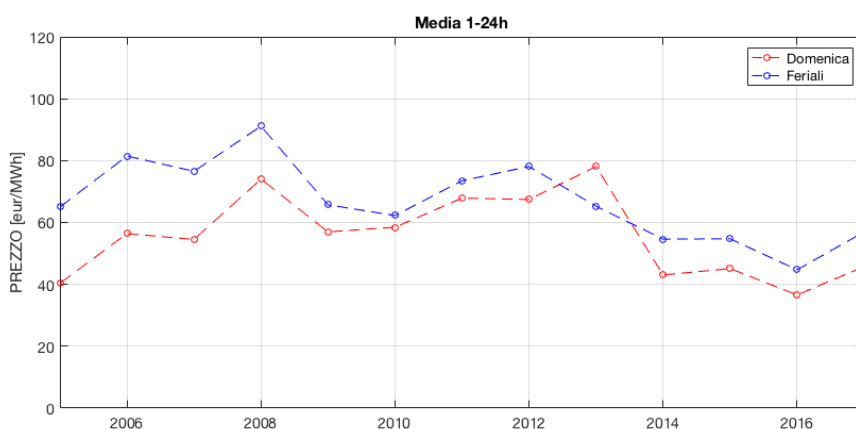


Figura 26 - Valori annuali medi della domenica e feriali a confronto

Ora qualche considerazione di carattere generale, è stato osservato che sono identificabili tre periodi di cui quelli estremi caratterizzati da una certa stabilità sul trend e quello intermedio di transizione fra gli altri due, ben distinguibili nei giorni feriali (Figura 19 o Figura 20) ma più confusi la domenica (Figura 24 o Figura 25). La scelta di confrontare i valori medi sulle 24 ore come in Figura 26 ha senso per i feriali in quanto le curve normalizzate riportate in Figura 19 sono ben distribuite e sovrapposte se si considerano in blocco diversi periodi, quindi la media è un indice di quanto sono traslate verticalmente le curve l'una rispetto all'altra. La questione invece è più complicata la domenica, continua a valere per gli ultimi anni come per i primi (2005/2009 tolto il 2006 e 2015/2017) ma non per il periodo centrale più il 2006 e 2014. In conclusione quanto riportato in Figura 26 è un buon compromesso fra veloce lettura e informazione ma c'è da ricordare che la parte centrale della curva rossa (2010/2013) trascurava gran parte l'informazione sul profilo orario fruibile in Figura 24.

2.2.1.5. Statistiche medie mensili

Si è poi svolta la medesima analisi su scala mensile, anche qui l'intervallo di tempi è stato ristretto ai soli anni si cui si dispone della serie completa dei dati, cioè dal 2005 al 2017 sempre per il valore del prezzo unico nazionale (PUN [€/MWh]) in riferimento al mercato del giorno prima. I valori sono ottenuti mediando i valori orari per ogni mese e per tutti e 13 gli anni considerato, le curve sono plottate in tre periodi su figure differenti, questo complica la lettura ed il confronto ma la natura dei valori è piuttosto caotica su scala mensile perciò un confronto globale delle 13 curve risulta veramente faticoso. Per facilitare la lettura sull'asse delle ordinate è stato imposto un intervallo limite [0-100 €/MWh] .

Suddetta analisi è utile per trarre conclusioni riguardo l'influenza della stagionalità sul prezzo dell'energia elettrica, quindi soprattutto rispetto le principali variabili meteorologiche che incidono sulla produzione da fonti rinnovabili e la **temperatura** che è legata alla domanda di energia elettrica per il condizionamento estivo in quanto esso è attuato esclusivamente da pompe di calore a compressione di vapore (spinte da motori elettrici) oppure ad assorbimento che necessitano calore ma solitamente queste sono presenti in pochi impianti termici con assetto trigenerativo non molto diffusi in Italia. Per quanto riguarda i mesi invernali il fabbisogno energetico richiesto è maggiore rispetto a quello per il raffrescamento ed è fortemente dipendente dalla temperatura esterna ma la maggior parte degli impianti di riscaldamento italiani non sono alimentati elettricamente (come accade per esempio in Francia) ma si tratta di impianti classici a combustione di fonti fossili , per ovvie questioni economiche.

Di seguito si analizzano i trend mensili sui tre periodi sopra citati:

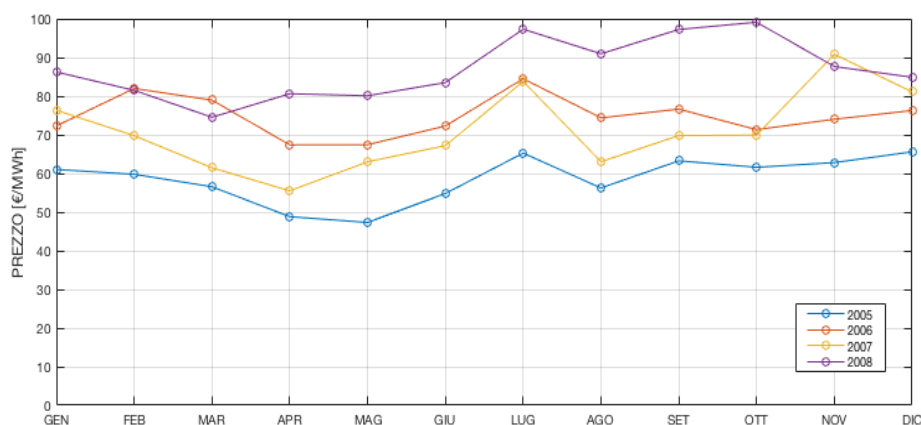


Figura 27 – Primo periodo: valori medi mensili

Iniziando dal primo periodo (Figura 5) si vede per l'ennesima volta che i prezzi massimi dal 2005 ad oggi sono stati raggiunti nei mesi di Luglio e Ottobre 2008 , ma comunque sono stati mediamente più alti rispetto gli anni precedenti, qualche eccezione si nota nei mesi a cavallo del periodo invernale ad esempio novembre e dicembre, in cui 2007 e 2008 sono confrontabili, oppure febbraio e marzo dove accade lo stesso fra 2006 e 2008.

Si apprezza l'aumento dei prezzi nel primo periodo ed è più difficile notare somiglianze confrontando le curve rispetto alla scala giornaliera (Figura 17) dove sono ben evidenti due picchi, anche qui si può generalizzare un trend stagionale in cui è presente un picco in corrispondenza di Luglio-Agosto e un

graduale innalzamento dei prezzi nei mesi invernali, ma non un vero picco (escluso il 2007), il picco estivo è dovuto alla domanda da climatizzazione, qui se ne ha una prima conferma.

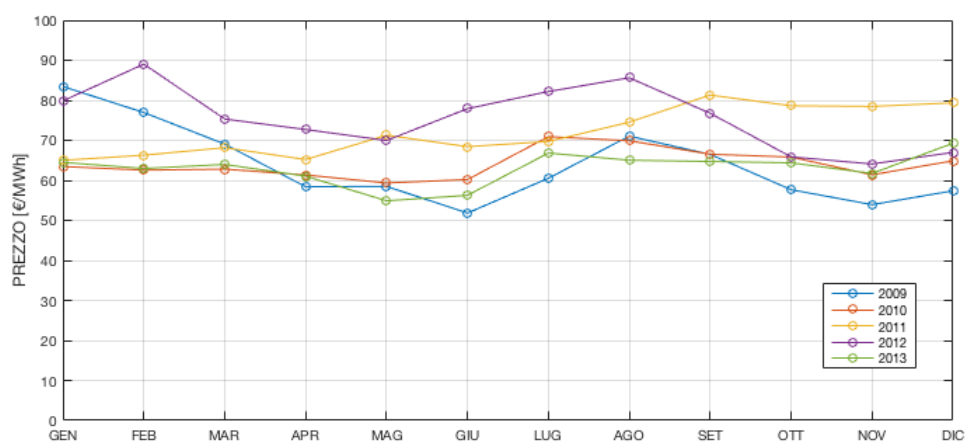


Figura 28 – Secondo periodo: valori medi mensili

Passando al secondo periodo si nota una spiccata somiglianza fra 2010 e 2013 dove la media mensile non subisce grandi variazioni mantenendosi entro i 55-70 €/MWh, al 2009 è ancora marcata la presenza del picco estivo che nei due anni successivi va scomparendo per via della massiccia installazione di solare fotovoltaico già discussa in precedenza.

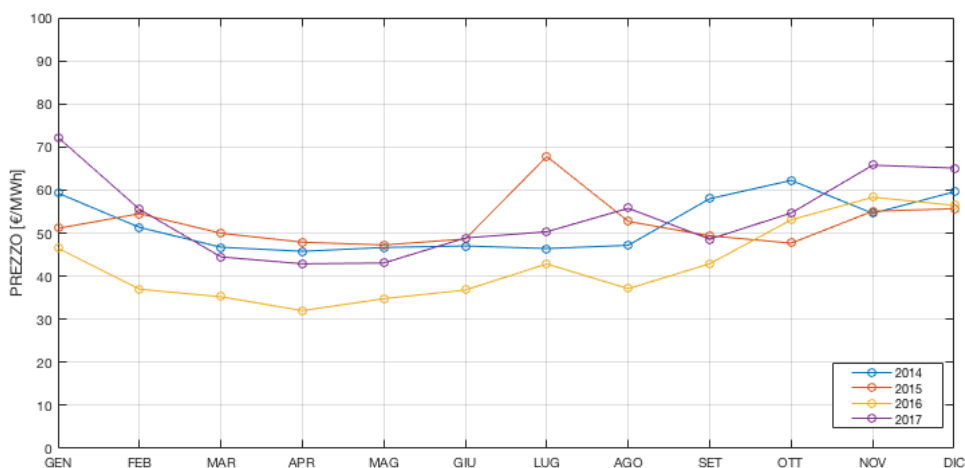


Figura 29 – Terzo periodo: valori medi mensili

Arrivati all'ultimo periodo, è interessante notare che, ad eccezione del 2015, il picco estivo è minore del valore raggiunto nel periodo invernale e nel 2014 non vi è un vero e proprio picco verso Luglio-Agosto bensì per tutta la primavera e estate il prezzo si mantiene costante per poi risalire a settembre e ottobre.

L'analisi condotta da Figura 27 a Figura 29 è poco esplicativa riguardo i trend stagionali, le medie mensili inglobano un elevato numero di giorni sia feriali che festivi per cui le differenze sono meno accentuate e più confuse, in apertura al paragrafo si è fatto cenno alle variabili meteorologiche che

giocano un ruolo importante su scala stagionale fra cui spicca la temperatura come principale fattore influenzante.

Quindi, si sono raccolti i dati relativi alle temperature medie mensili (massime e minime) su scala regionale, i dati sono estrapolati dagli annuari statistici italiani (ASI) [62] [63] [64] [65] [66] [67] [68] [69] disponibili online dal sito ISTAT [70], ogni annuario viene pubblicato entro fine anno le statistiche riportate si all'anno precedente. I dati così ottenuti presentano delle incongruenze in base al periodo di pubblicazione:

- **2009 al 2016:** sono disponibili le temperature medie per ogni regione e la media nazionale in termini di valori assoluti massimi e minimi, nell' ASI 2013 non si riporta la media nazionale ed è quindi calcolata come la media sulle temperature delle singole regioni ponderata sull'estensione di esse rispetto al territorio nazionale.
- **2007 e 2008:** in entrambi, i dati si riferiscono a stazioni di rilevamento e non più all'intera regione, in più è fornita un solo valore medio per il 2007 e non massima e minima come da li in poi.
- **2006 e precedenti:** nessun dato regionale né a scala mensile, sono forniti i valori delle principali stazioni italiane con valori estremi e medie estive ed invernali.

Da questo di vista non ha senso considerare tutto l'arco di tempo dal 2005 in poi (come fatto finora) in quanto i valori di temperatura non si riferiscono alle stesse zone e non sono forniti su scala coerente, per il 2008 si potrebbe pensare di mediare i valori delle stazioni sulle aree a cui si riferiscono ma comunque non si avrebbe la certezza di come sono ottenuti i dati regionali riportati negli ASI a venire.

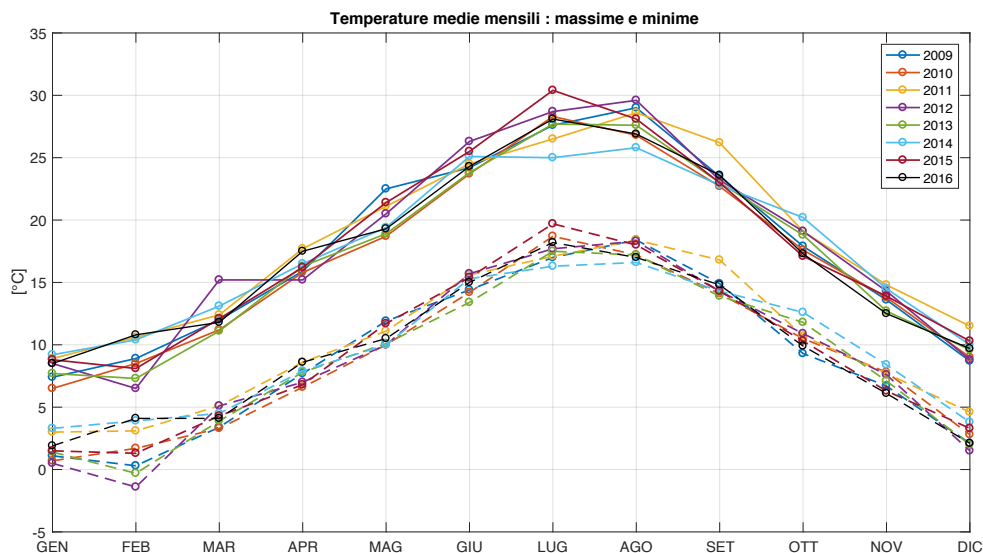


Figura 30 – Temperature medie nazionali : massima e minima mensile

Allora l'influenza delle temperatura viene studiata solamente dal 2009 in avanti per coerenza fra i dati; i profili di temperatura Figura 30 e il risultato della normalizzazione dei prezzi medi mensili con la temperatura media massima sono illustrati in Figura 31, si utilizza il valore massimo a causa della natura della domanda di energia elettrica che deriva dal condizionamento estivo piuttosto che altro e quindi la

domanda di energia sarà tanto maggiore quanto lo è la temperatura, e soprattutto ci aspettiamo una dipendenza mercata nei mesi estivi.

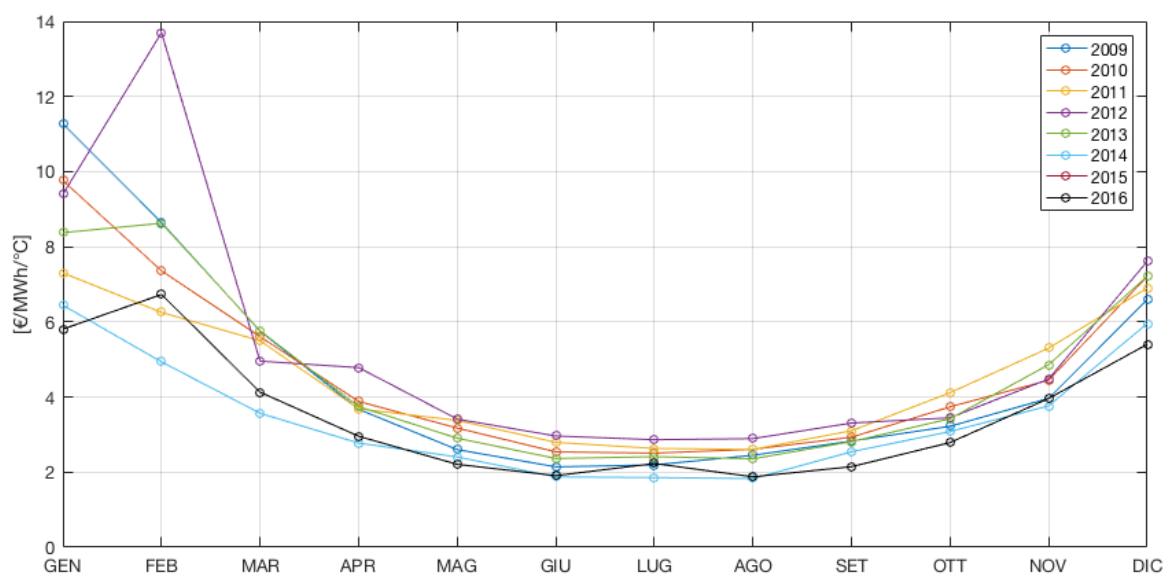


Figura 31 – Profili medi mensili normalizzati rispetto alla temperatura media nazionale massima

A seguito della normalizzazione con la temperatura media massima è evidente un trend stagionale uniforme che presenta prezzi maggiori per temperature minori, quindi in inverno e al contrario l'estate, in un certo senso questo tende ad uniformare il prezzo durante l'anno, se la temperatura aumentasse proporzionalmente alla diminuzione del prezzo esso si manterrebbe pressoché costante a scala mensile, perciò eventuali picchi come quelli presenti nei mesi estivi dipendono direttamente da anomalie sulla temperatura. Infatti vi è un riscontro diretto per 2012 e 2015, in Figura 30 sono gli anni con le estati più calde e proprio per questo anche i prezzi hanno subito incrementi, concentrati a luglio 2015 o distribuiti su giugno-agosto 2012, vale lo stesso per settembre 2011 e febbraio 2012.

Porgendo l'attenzione alla volatilità dei prezzi si sono normalizzate le curve di Figura 27,29 e 30 rispetto alla media annuale, il grafico non è riportato ma per il primo periodo non possiamo avere confronti diretti con i trend di temperatura (causa mancanza di dati) ma comunque si vede che mensilmente la variazione percentuale è limitata $\pm 20 / 30$ % massimo in base all'anno, contro il 70 % riscontrato su scala giornaliera per ambedue feriali e domenica (Figura 19 e Figura 24).

Quindi su scala mensile emerge una forte dipendenza dalla temperatura media massima e le fluttuazioni dei prezzi sono praticamente dimezzate rispetto a quanto emerge dal confronto su scala giornaliera

2.3. MI: MERCATO INFRA-GIORNALIERO

Ha sostituito il mercato di aggiustamento dopo l'entrata in vigore del GME S.p.A nel 2009, su di esso gli operatori idonei possono modificare i programmi di stabiliti su MGP, viene introdotto grazie alla legge 2/09 ed è un appendice fondamentale del MGP che permette di adattare in tempo reale la negoziazione rispetto alle modifiche delle informazioni relative a impianti produttivi e consumi sulla rete. E' lo strumento preliminare con cui Terna S.p.A provvede all'approvvigionamento delle risorse necessarie al dispacciamento attraverso l'omonimo mercato. Le sessioni sono sette articolate su due giorni: quello antecedente al dispacciamento e il giorno stesso, le prime due avvengono subito a valle del MGP e sono attive sin dal 2009. Le ultime cinque avvengono i giorno stesso e sono attive da meno tempo: MI3 e MI4 vengono introdotte il 1 gennaio 2011 e le ultime tre in seguito. Dunque si nota come la struttura dell'MI sia cresciuta nel tempo, secondo il vademecum aggiornato al 2012 [59] (il più recente disponibile sul sito GME) si limitano a quattro sessioni, le prime due costituivano già il mercato di aggiustamento e le ultime lo ampliano. Ad oggi la disposizione dei mercati è schematizzata come segue:

Giorno di riferimento	D-1				D															
	MGP	MI1	MI2	MSD1	MB1	MI3	MSD2	MB2	MI4	MSD3	MB3	MI5	MSD4	MB4	MI6	MSD5	MB5	MI7	MSD6	MB6
Informazioni preliminari	11.30	15.00	16.30	n.d.	n.d.	23.45*	n.d.	n.d.	3.45	n.d.	n.d.	7.45	n.d.	n.d.	11.15	n.d.	n.d.	15.45	n.d.	n.d.
Apertura seduta	08.00**	12.55	12.55	12.55	°	17.30*	°	22.30*	17.30*	°	22.30*	17.30*	°	22.30*	17.30*	°	22.30*	17.30*	°	22.30*
Chiusura seduta	12.00	15.00	16.30	17.30	°	23.45*	°	3.00	3.45	°	7.00	7.45	°	11.00	11.15	°	15.00	15.45	°	19.00
Esiti provvisori	12.42	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.	n.d.
Esiti definitivi	12.55	15.30	17.00	21.45	#	0.15	2.15	#	4.15	6.15	#	8.15	10.15	#	11.45	14.15	#	16.15	18.15	#

** l'ora si riferisce al giorno D-9

* l'ora si riferisce al giorno D-1

° Si utilizzano le offerte presentate sul MSD1

Disciplina del dispacciamento

Figura 32 – Tempistiche delle attività sul MPE

Lo storico degli esiti è disponibile nella sezione download del sito web¹⁴, l'anno di riferimento rimane il 2016 e i dati disponibili on-line non presentano negoziazioni avvenute nelle ultime due sessioni (MI6 e MI7) perciò l'interesse è ristretto alle prime cinque sessioni che sono continue ma la negoziazione 1-24h è prevista solo dalle prime due sessioni:

- MI1 e MI2 → 24 h
- MI3 → 16 h (9-24)
- MI4 → 12 h (13-24)
- MI5 → 8 h (17-24)

¹⁴ http://www.mercatoelettrico.org/It/download/DownloadDati.aspx?val=MI1_Prezzi

I dati sono forniti duplicemente per ogni zona con dicitura “aggiornato”, dopo aver verificato che non vi sono differenze fra valori base e i dati aggiornati vengono tralasciati i quelli ridondanti.

MI rappresenta una proroga indispensabile del MGP e in quanto tale si basa sugli stessi meccanismi di domanda e offerta tenendo conto dei limiti di transito, l'unica differenza è che non si calcola il PUN e le offerte vengono valorizzate al prezzo zonale. Comunque il GME si avvale di un sistema di corrispettivi di non arbitraggio al fine di replicare gli effetti del PUN anche su MI: su ogni **transizione di acquisto** conclusa sul MI che sia valorizzata ad un prezzo maggiore o minore del PUN, l'operatore deve ricevere o pagare suddetto corrispettivo pari alla differenza tra prezzo zonale e PUN; viceversa per una **transizione di vendita**.

Uno dei problemi che salta subito all'occhio è un valore anomalo di 3000/MWh riscontrato verso fine 2016 nelle zone relative alle isole sud quali SICILIA, PRIOLO GARGALLO E MALTA, nel rapporto annuale¹⁵ redatto dal GME risulta che sia su MI1 che su MI4, si sia sfiorato il valore di soglia VENF in quelle zone. Questo problema rende scomodo visualizzare la distribuzione, quindi vengono tralasciati i valori massimi per i mercati di interesse (SICI, MALT e PRGP) in maniera da renderli più leggibili.

Nell'appendice A (pagina 132) vengono raccolte le grandezze di riferimento per la prima sessione solamente, dalla

Tabella 26 si notano i valori spropositati di 3000 €/MWh per Sicilia Malta e Priolo Gargallo.

2.4. CONFRONTI MI e MGP

2.4.1. Confronto zonale annuale

La Figura 33 è ottenuta mediando le serie dei prezzi zonali del 2016, i valori fuori scala appena citati non influiscono sui valori medi perché risulta che Sicilia, Priolo Gargallo e Malta siano comunque le zone in cui l'energia è valorizzata prezzi maggiori su tutte le sessioni. Un valore isolato sebbene di uno/ due ordini di grandezza maggiore non contribuisce a falsare il trend medio di MI1 e MI4. Esclusa la seconda sessione le curve medie seguono un andamento uniforme con piccole variazioni, quasi come se fossero semplicemente traslate lungo le ordinate, in questo senso MI1 è il più economico e MI5 il più caro, i due che restano sono estremamente simili. Invece, il grafico in basso indica l'entità oscillazioni attorno al valor medio, a prescindere da esso e si nota che stanno entro il $\pm 20\%$.

Solo la seconda sessione sembra essere veramente diversa come profilo, mentre le altre sono uniformate di zona in zona, perciò le negoziazioni avvenute durante il MI2 nel 2016 hanno apportato modifiche evidenti agli esiti della prima seduta mentre le altre sedute si sono concluse solamente con a prezzi maggiori.

¹⁵ <http://www.mercatoelettrico.org/It/GME/Biblioteca/RapportiAnnuali.aspx>

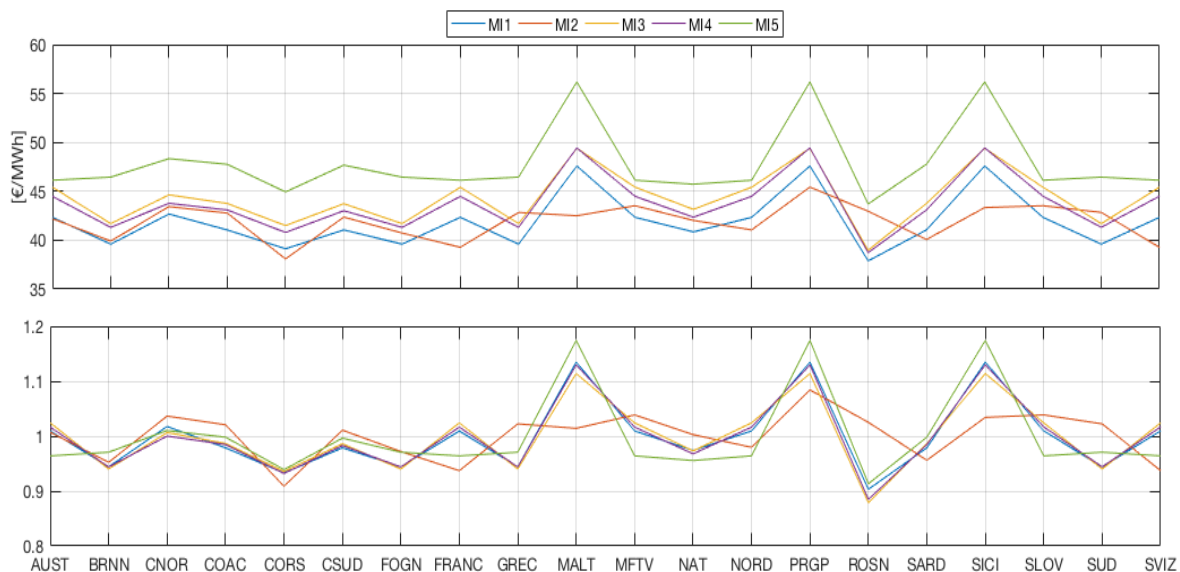


Figura 33 – MI, Sopra: prezzo medio zonale , Sotto: prezzo medio zonale normalizzato

Ancora una volta i valori sotto il primo quartile sono compatti, I valori minimi sono all'incirca gli stessi per tutte le zone mentre i massimi variano con picchi più o meno ampi, in linea di massima si varia dai 160 a 250 €/MWh di picco registrato nella seconda sessione.

Negli altri mercati MI2 e MI5 si aggiunge il market coupling con la Slovenia (BSP) introdotto dal 26 giugno 2016 per ottimizzare la gestione ed il monitoraggio delle dorsali internazionali, non è riportato in Figura 33 perché non solo è presente su due sole sessioni ma comunque sono disponibili poco meno della metà dei valori che si hanno per le altre zone, una media sarebbe poco indicativa.

Ora si introduce il confronto diretto con il MGP, ci si limita ad alcune zone in quanto i due mercati non sempre operano sulle stesse, in base al periodo di cui ho disponibili i dati del MGP non è detto che vi sia già stato introdotto il MI o che le zone attive allora lo siano tuttora. L'anno considerato rimane lo stesso, cioè il 2016, si analizza solamente la media di tutte le zone e la si confronta fra MGP e tutte e cinque le sedute del MI (Figura 33 e Figura 34), in sostanza è un'integrazione del primo grafico del confronto zonale sul solo MI. Vengono tralasciate tre zone:

- **Montefalcone (MFTV) e NAT** : perché riguardano zone che non sono state sede di negoziazione sul MGP al 2016 (si veda Tabella 23).
- **Corsica DC (CORS)**: a seguito di numerosi valori fuori scala ($3000 \frac{\text{€}}{\text{MWh}}$) avvenuti dalle 11 alle 14 del 6 Maggio 2016, dalle 9 del 17 Maggio alle 17 del 13 Giugno 2016 e dalle 9 del 10 Ottobre alle 16 del 13 Ottobre 2016.

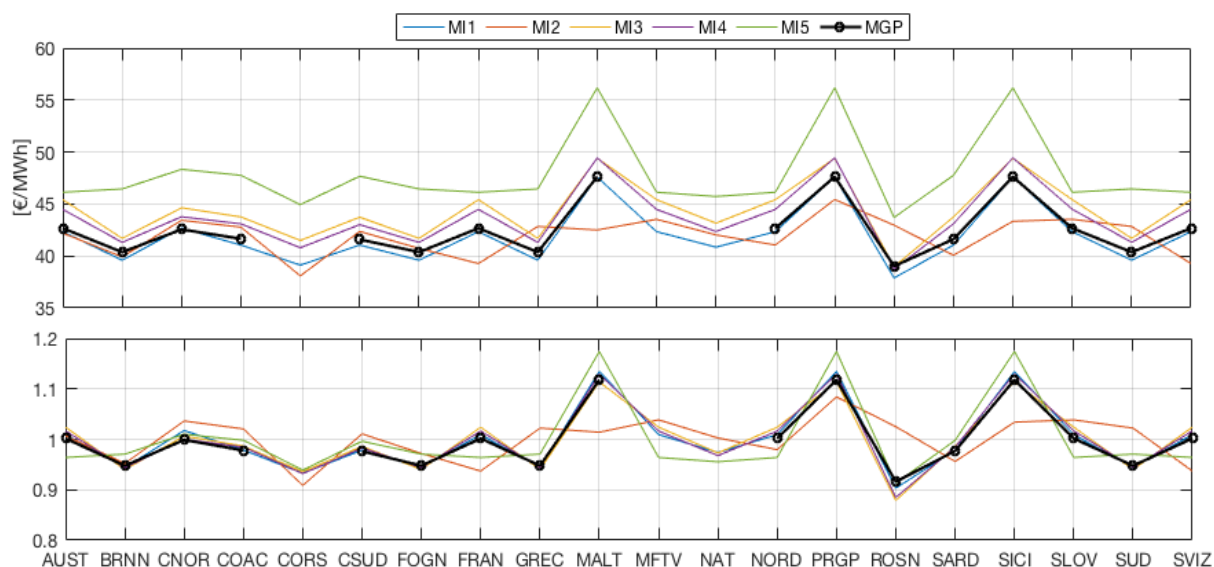


Figura 34 –MGP e MI: prezzo medio zonale

Emerge che i valori medi zonali del MGP siano molto vicini alla prima sessione del MI, in alcune zone il prezzo di equilibrio non viene modificato e nelle restanti la sessione di MI1 si conclude con valori leggermente inferiori. Conferma riguardo i valori massimi raggiunti che rimangono i medesimi e non si modificano in MI1, si ricordare che la Corsica DC (CORS) è stata eliminata perché tocca una media annuale di $292 \frac{\text{€}}{\text{MWh}}$. In sostanza, i programmi stabiliti entro il MGP costituiscono l'ossatura che vien poi adattata in tempo reale alle esigenze delle rete permettendo a Terna di pianificare la distribuzione con atto finale nel MSD. Mano a mano che si avanza nel tempo la modifica dei piani diventa sempre più dispendiosa (vedi MI3,4 e 5) e la seconda sessione si presenta come l'unica che porta modifiche concrete al profilo della curva imposto dal MGP. Partendo da sinistra, i primi punti sono ancora in linea ma dalla Francia in poi il trend subisce modifiche, il massimo di Malta è nettamente smorzato così come la Sicilia, di meno Priolo Gargallo .

A questo punto si prosegue con qualche osservazione riguardo la simmetria, è stato largamente confermato che in ogni caso qualsiasi storico dei prezzi è sempre molto più disperso sopra i terzo quartile, ora si cerca di cogliere questo aspetto e quantificare le differenze tra zone, prima di tutto si analizza la disposizione di media, mediana e moda sulle serie di cui disponiamo su entrambi MI e MGP, sempre per il 2016. Se le tre sono disposte in crescendo o diminuendo la serie è detta rispettivamente asimmetrica positivamente o negativamente, simmetrica se le tre coincidono.

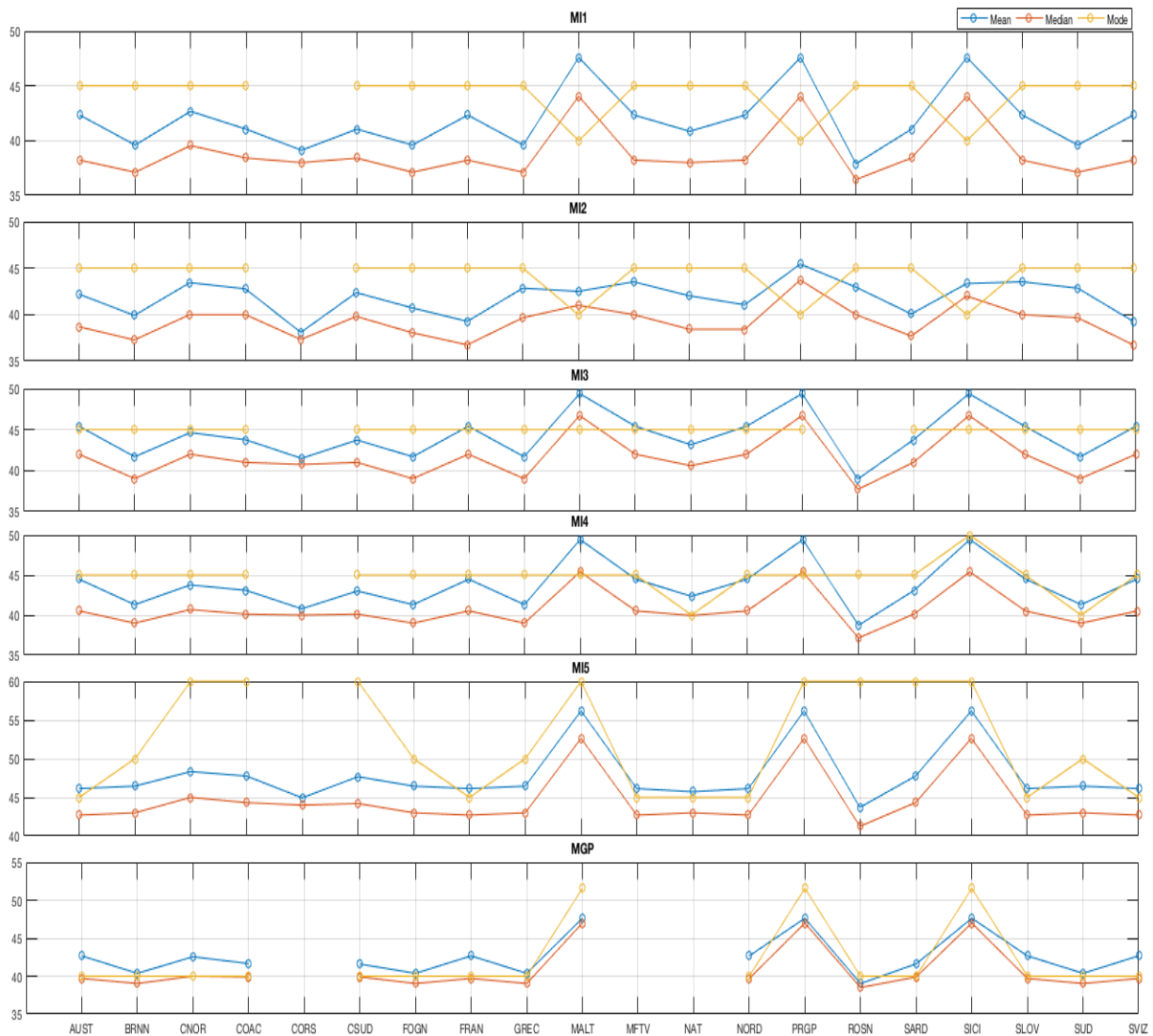


Figura 35 – Media, mediana e moda per tutte le sessioni del MI e MGP

Non è possibile stabilire con certezza la natura della simmetria per le serie di interesse, fino alla terza seduta le serie in cui c'è un massimo locale hanno media, mediana e moda in decrescendo ma da lì in poi non vi è alcun riscontro ulteriore, certamente è molto complicato, se non impossibile, trovare una serie vicina alla simmetria, ci sono delle zone sul MGP in le tre grandezze assumono valor molto simili e che quindi sono quelle con caratteristiche di simmetria più accentuate, questo fenomeno si riscontra bene sul MGP ma non sul MI, dove i valori sono includono su intervalli più ampi. La moda della Corsica è omessa perché in ogni sessione di MI ha molteplici valori orari nulli che ne abbattano il valore cos' come per Rossano in MI3. E' particolare che su MGP la Corsica presenti lunghi periodi di superamento del VENT e invece su MI avvenga in contrario, la faccenda andrebbe analizzata più a fondo tenendo conto dei limiti di transito fissati sulle isole e in più a discapito di zone virtuali estere.

Ultima, ma non meno importate, la Figura 36 illustra i valori di deviazione standard, gli unici valori mancanti riguardano MFTV e NAT dei quali proprio non vi è la serie di partenza e CORS, dove il valore supera gli $800 \frac{\text{€}}{\text{MWh}}$, i tre soliti massimi su MI1 e 4 sono contenuti solo perché sono molto meno frequenti i valori anomali ma i massimi delle quattro serie sono pari a $3000 \frac{\text{€}}{\text{MWh}}$ e sono tutti uguali.

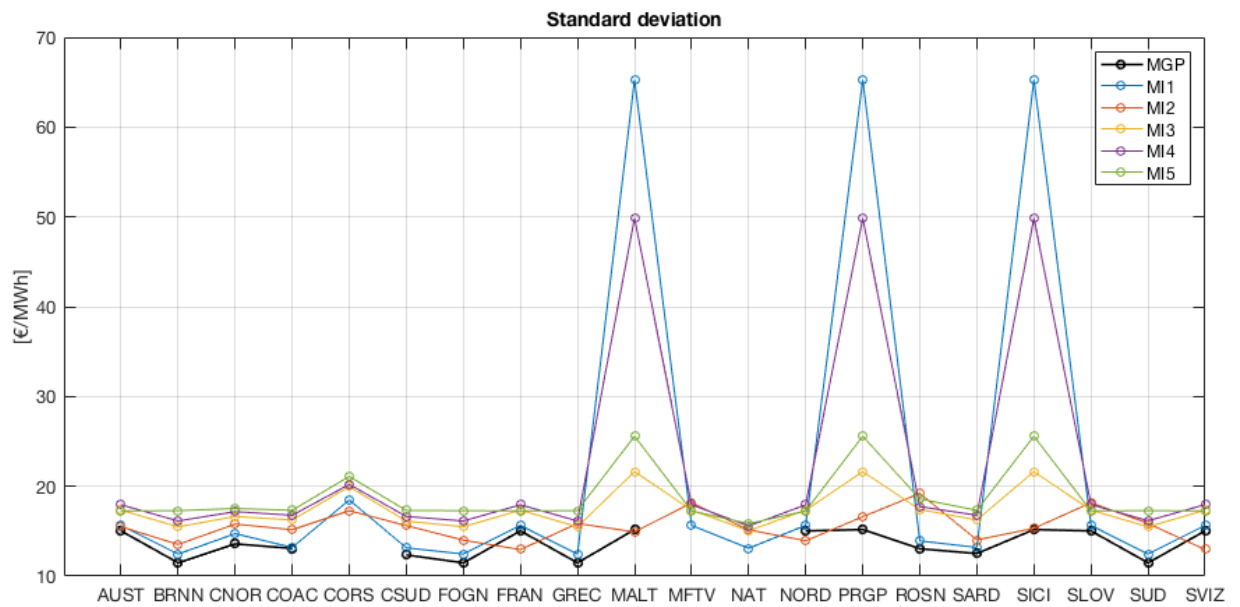


Figura 36 – Zone di mercato: deviazione standard

La deviazione standard insieme alla media (Figura 34) già da sole forniscono buona parte delle informazioni che aiutano a caratterizzare le serie, le medie zonali su MGP appaiono analoghe a MI1 o di poco superiori ma i valori sono mediamente meno dispersi attorno agli stessi (di pochissimo). Le restanti sessioni di MI vanno mano a mano ad esibire deviazioni standard maggiori con il passare del tempo, tolta la seconda che come già si sa segue un profilo diverso.

2.4.2. Confronto mensile

Un ultimo confronto utile è da svolgersi su scala temporale ,nello specifico considera i valori medi mensili e deviazione standard riguardo il 2016 per tutti e sei i mercati, in aggiunta divisi per zona. Non è immediato impostare un confronto efficace secondo queste modalità, venti zone su dodici mesi per sei sedute di mercato sono riassunti nelle figure a seguire tramite diagrammi a barre

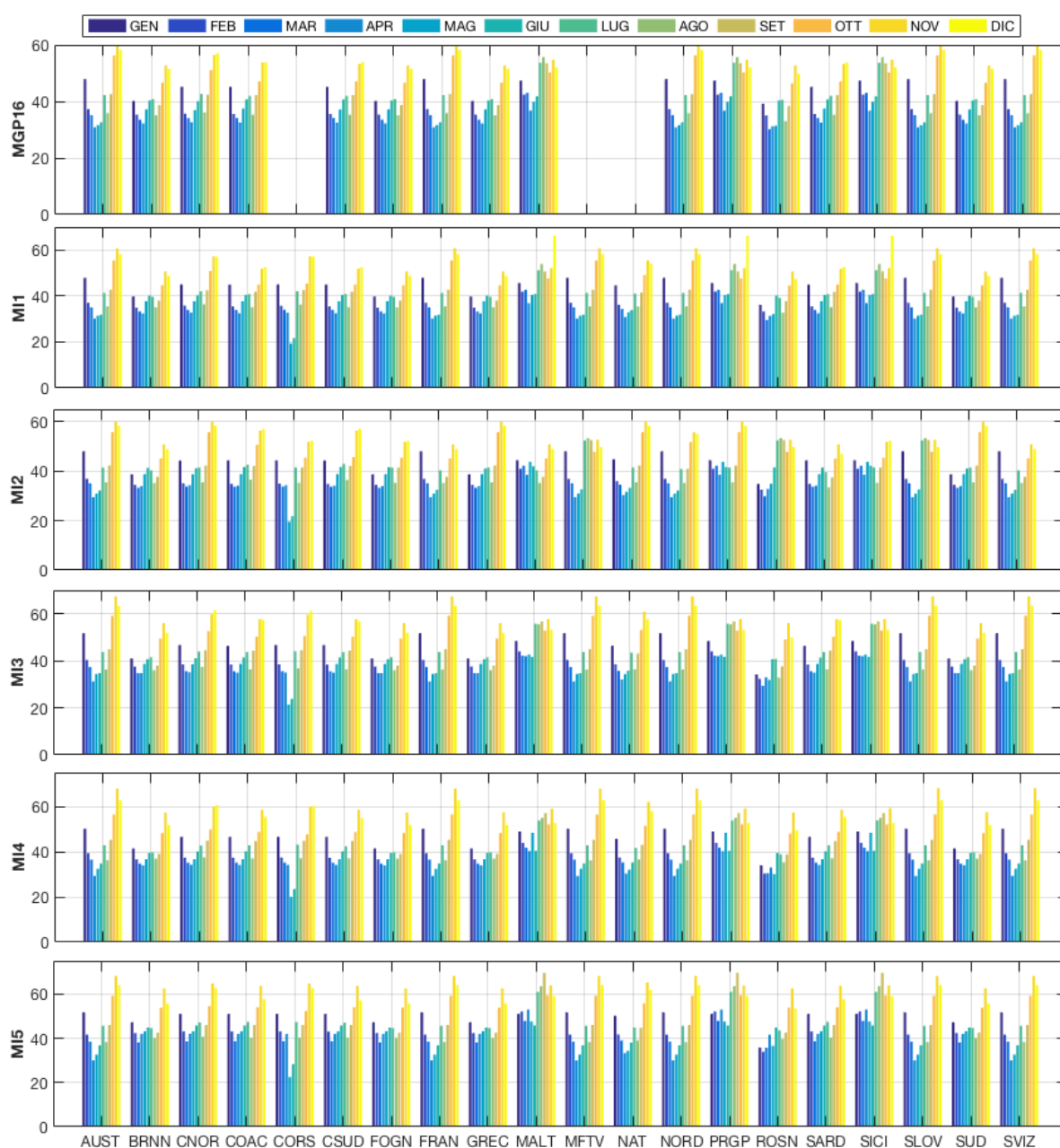


Figura 37 – Media mensile per zona

Ancora una volta si sottolinea la somiglianza fra il mercato del giorno prima e la prima sessione del mercato infra-giornaliero, il trend è lo stesso ed i valori mensili analoghi, si può notare che l'anomalia di MI1 su MALT, SICI e PRGP è avvenuta a dicembre (se si indaga ulteriormente verso l'inizio del mese) ma osservando solo la media non si evince nulla riguardo a quella su MI4.

Includendo nel ragionamento anche la deviazione standard si apprezza che l'anomalia è avvenuta maggio (questa volta verso fine mese) ed evidentemente è stata meno frequente se è tale da non influire sulla media, c'è anche da dire che MI4 si svolge su 12h mentre MI1 su 24h, la media di MI1 è ottenuta sul doppio dei valori quindi il numero di anomalie su MI1 è più del doppio sicuramente (nello specifico 4 su MI1 e 1 su MI4).

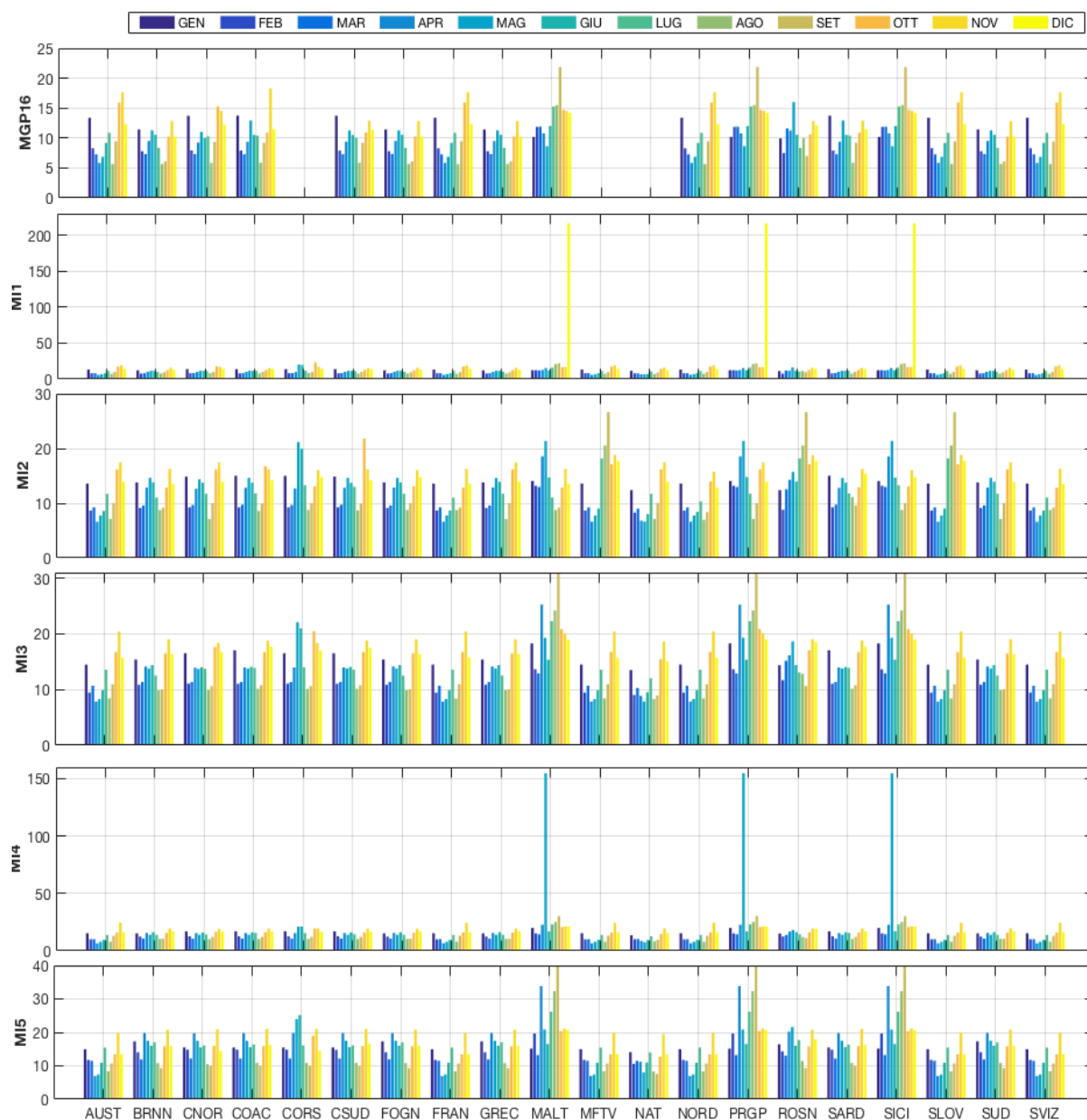


Figura 38 – Deviazione standard mensile per zona

Tolta la quarta sessione, i valori massimi sono sempre MALT , PRGP e SICI ma traslati a settembre, MI2 già in prima sede mostrava differenze di trend e trova conferma con massimi sempre a settembre ma su altre zone: Montefalcone, Rossano e Slovenia.

Le informazioni contenute negli ultimi due grafici sono molte e non immediate, perciò si considera un'unica media comprensiva di tutte le zone nella speranza di generalizzare un trend mensile che si il più possibile ben delineato. La figura che segue conferma le aspettative, considerando indistintamente

tutte le zone vi è un trend uniforme a prescindere dalla sessione, due massimi locali a novembre e luglio
 mano a mano più accentuati con nel susseguirsi delle sessioni, minimi ad aprile ed agosto. Le deviazioni
 standard non presentano differenze sostanziali ma si distingue chiaramente l'occorrenza delle anomalie
 a maggio per MI4 e dicembre per MI1.

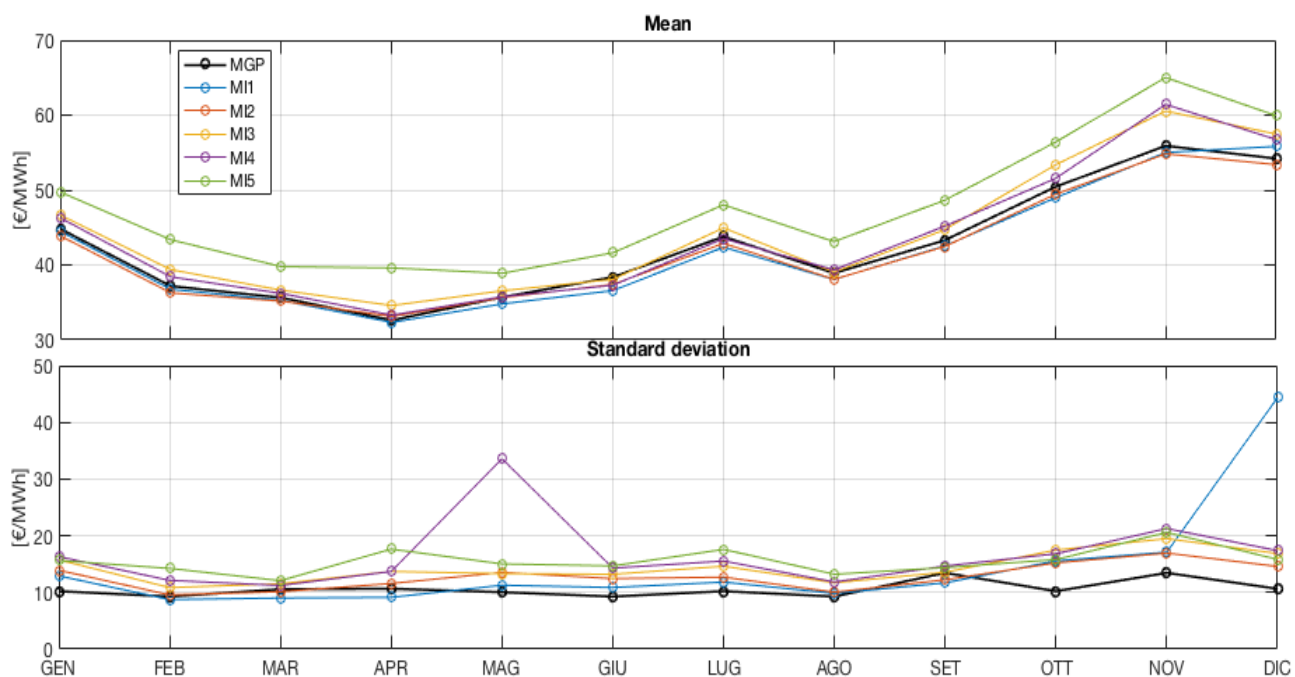


Figura 39 – Valori mensili: media su tutte le zone

2.5. VARIABILI ENDOGENE e DATI CORRELATI

2.5.1. Potenza rinnovabile installata in Italia

Per la ricerca di tali dati si è fatto riferimento al sito istituzionale del Gestore dei Servizi Energetici (GSE) sulla base dei rapporti statistici disponibili dal 2008 al 2016 ¹⁶ [71], su questa base quindi sono riportati i dati annuali totali, in riferimento al periodo che va dal 2002 al 2016 e quelli divisi per fonte dal 2007 al 2016 causa la carenza di dati.

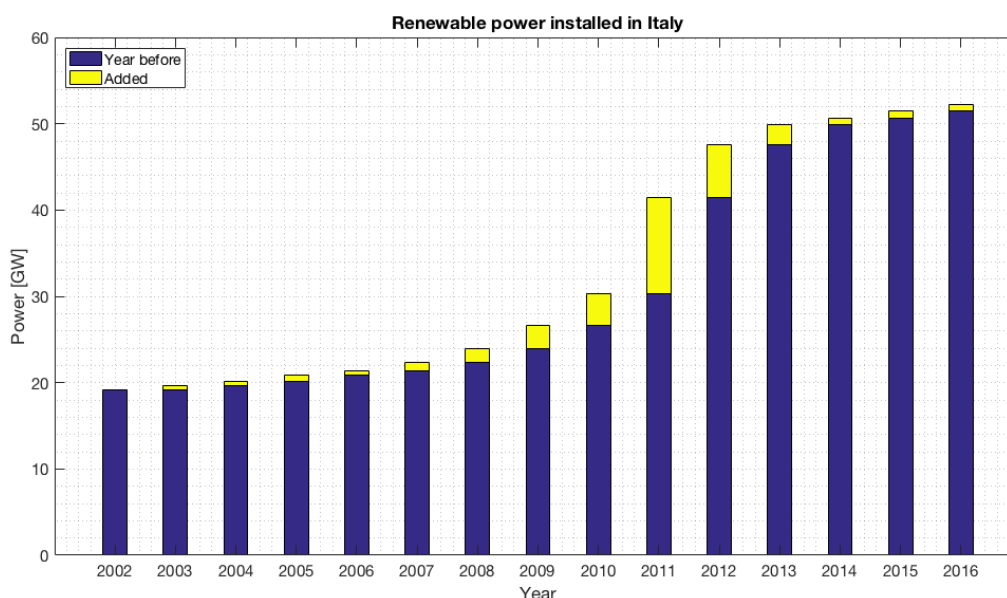


Figura 40 – Potenza da FER installata in Italia dal 2002 al 2016

Nei 13 anni compresi tra il 2003 e il 2016 la potenza efficiente lorda installata in Italia è passata da poco meno di 20 a più di 50 GW, con un incremento esatto di 32.610 MW e un tasso di crescita medio annuo pari al 7,8%; gli anni caratterizzati da incrementi maggiori di potenza sono il 2011 e il 2012.

Fin dagli inizi del '900, il parco elettrico nazionale è stato caratterizzato dall'ampia diffusione di impianti idroelettrici; negli anni più recenti la potenza installata di tali impianti è rimasta pressoché costante (+0,7% medio annuo), mentre tutte le altre fonti rinnovabili sono cresciute in maniera considerevole grazie principalmente ai diversi sistemi pubblici di incentivazione.

Il trend è in netto aumento a partire dal 2009 fino al 2013, mentre nei periodo precedenti o successivi risulta stabile con piccole potenze aggiunte nel corso dell'annata, nei report di GSE si sottolinea che nei sei anni che vanno dal 2005 al 2011 il numero di impianti è sempre raddoppiato rispetto all'anno precedente ma non la potenza installata, la quale ha subito un picco spropositato durante il 2011, questo è dovuto ad impianti di taglia maggiore entrati in esercizio nel 2011, soprattutto fotovoltaici. Quindi nel corso del 2011 l'aumento registrato tocca quasi il 40% (Figura 41) che è più del doppio rispetto al valore appena inferiore ed è venti volte tanto rispetto al 2014, in Figura 42 è evidente che ciò è dovuto quasi totalmente al fotovoltaico con lievi contributi di eolico e bioenergie ma è anche chiaro che la crescita di

¹⁶ https://www.gse.it/Dati-e-Scenari_site/statistiche_site

potenza installata ha avuto notevole incremento nel periodo tra il 2008 e il 2013 e ad oggi siamo nuovamente in un periodo di stazionarietà, i rapporti statistici di un certo anno vengono resi disponibili online generalmente entro verso fine dell'anno successivo, perciò per il 2017 nulla è ancora disponibile.

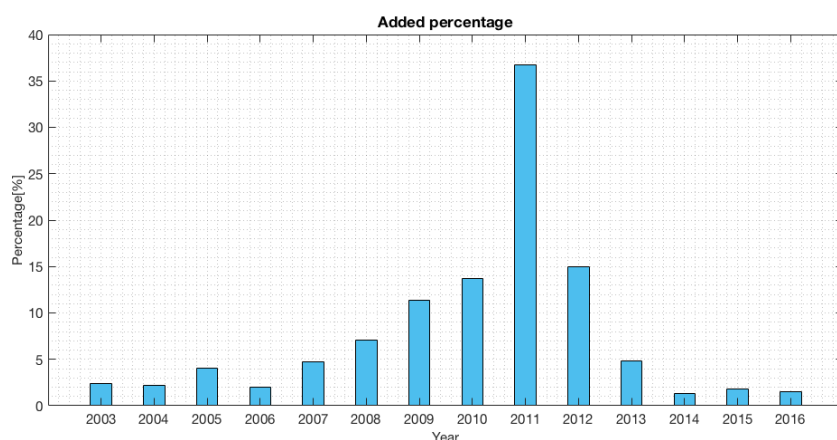


Figura 41 – Potenza da FER: tasso di incremento annuale

Scendendo nel dettaglio si va a suddividere la potenza installata per le singole fonti, prima fra tutte l'idroelettrico, che al 2007 era senza alcun dubbio la più diffusa e che in seguito si è mantenuta pressoché costante fino al 2016 dove, a seguito del marcato incremento del fotovoltaico, quest'ultimo al 2016 è confrontabile con quello idroelettrico.

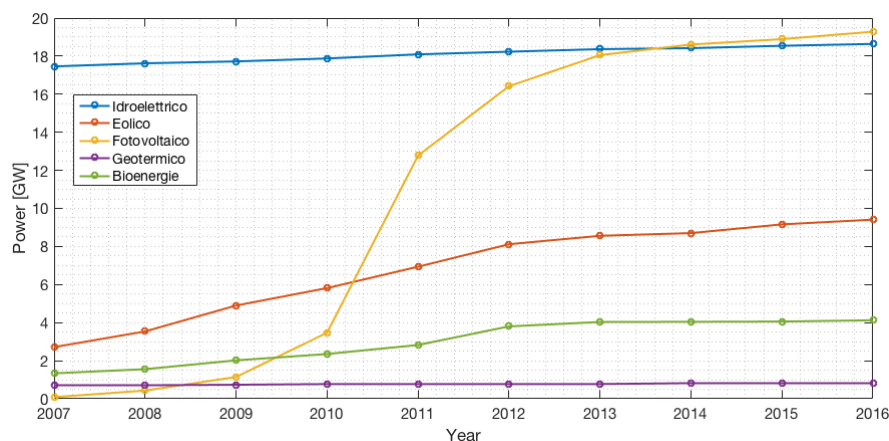


Figura 42 - Potenza da FER installata in Italia: contributo delle singole fonti

La quota parte dovuta al fotovoltaico cresce da circa lo 0,4% del totale a quasi il 37% in soli dieci anni, registrando un incremento notevole, nonché il maggiore; anche la quota di eolico risulta in crescita così come le bioenergie, a prescindere dall'entità dell'incremento esso si sviluppa principalmente fino al 2013 per poi rimanere pressoché invariato.

La quota di geotermico, invece, è stabile fra il 700/800 MW installati per un massimo di pochi punti percentuali sul totale, in Italia sono presenti centrali geotermiche di grossa taglia ma in zone molto specifiche nel sud della Toscana quali Larderello e vi è ancora del potenziale nell'area del monte Amiata.

2.5.2. Energia prodotta da fonti rinnovabili

Analogamente alla potenza installata, i dati necessari sono stati reperiti dal sito del Gestore Servizi Energetici [71] secondo i medesimi rapporti statistici disponibili per il periodo dal 2007 al 2016. Siccome l'energia prodotta è la vera grandezza di interesse quando si parla di quantità immesse sul mercato, anche i rapporti statistici offrono un livello di dettaglio maggiore riportando il contributo delle singole fonti già dal 2002 (Figura 45).

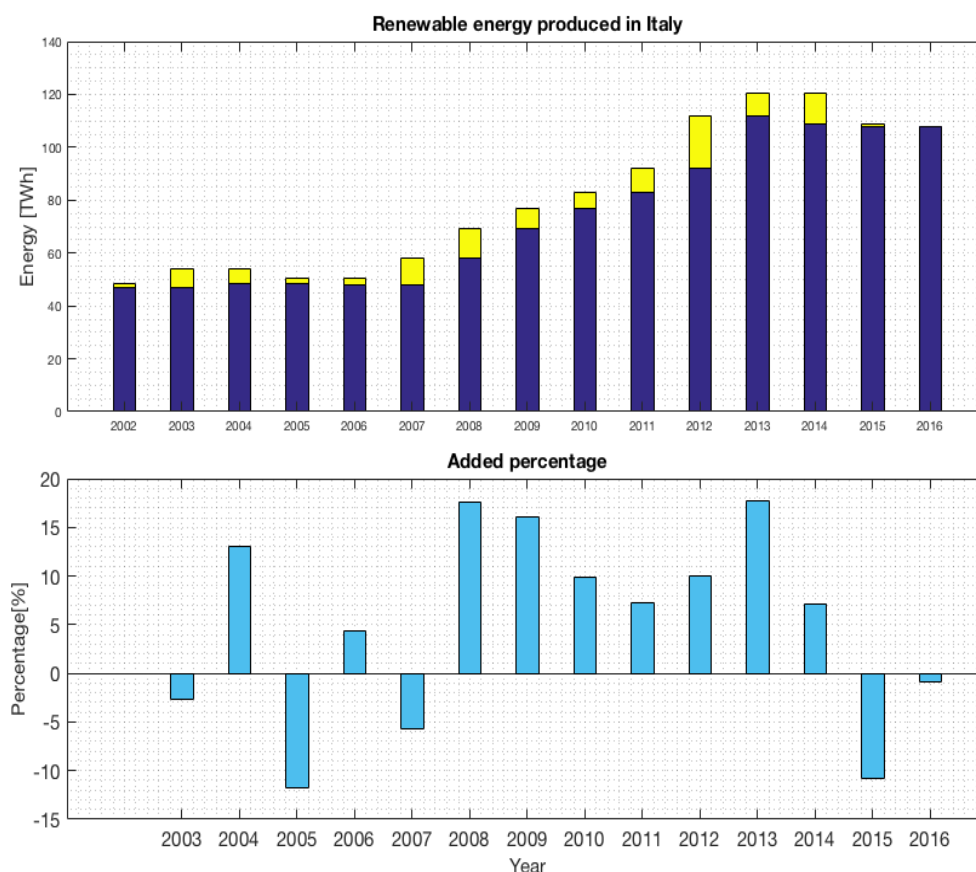


Figura 43 – Sopra: energia prodotta da FER , sotto: percentuale aggiunta/sottratta.

Al 2016 la produzione da fonti rinnovabili sé di poco superiore ai 108 TWh, pressoché costante rispetto all'anno precedente ma ha subito un calo superiore al 10% nel corso degli ultimi due anni benché la potenza sia comunque sempre in aumento. La curva di produzione è più variabile, mentre la potenza installata continua in ogni caso a crescere l'energia prodotta è altalenante fra valori anche in decremento.

La Figura 43 porta con sé un'informazione molto importante, altrimenti tralasciata considerando solamente la potenza, anche se indirettamente essa ci suggerisce il fattore di utilizzazione a piena potenza e quindi la percentuale rispetto al massimo producibile nel corso dell'anno con quella specifica potenza installata. A colpo d'occhio le curve di energia prodotta e potenza installata hanno uno shape simile e ciò indica che tale fattore non sia variato di molto come confermato in **Errore. L'origine r** **iferimento non è stata trovata..** Il valore medio annuale è assestato sui 26,5 % che rimane piuttosto

basso in confronto ai fattori di utilizzazione medi di centrali termoelettriche convenzionali che normalmente stanno sul 80% o anche di più.

Un valore così basso trova spiegazione se si considera la natura intermittente e non programmabile delle FER, le si usa quando sono disponibili e anche se hanno la precedenza sul dispacciamento non sarebbero comunque disponibili 8760 ore annue, ne è un palese esempio il solare fotovoltaico. In breve, è interessante capire quanto effettivamente sono sfruttabili le FER ma non ha senso confrontarle con le fonti fossili classiche in questi termini in quanto sarebbe opportuno calcolare un fattore di utilizzazione mediato sulla reale disponibilità della fonte riportata in Figura 44.

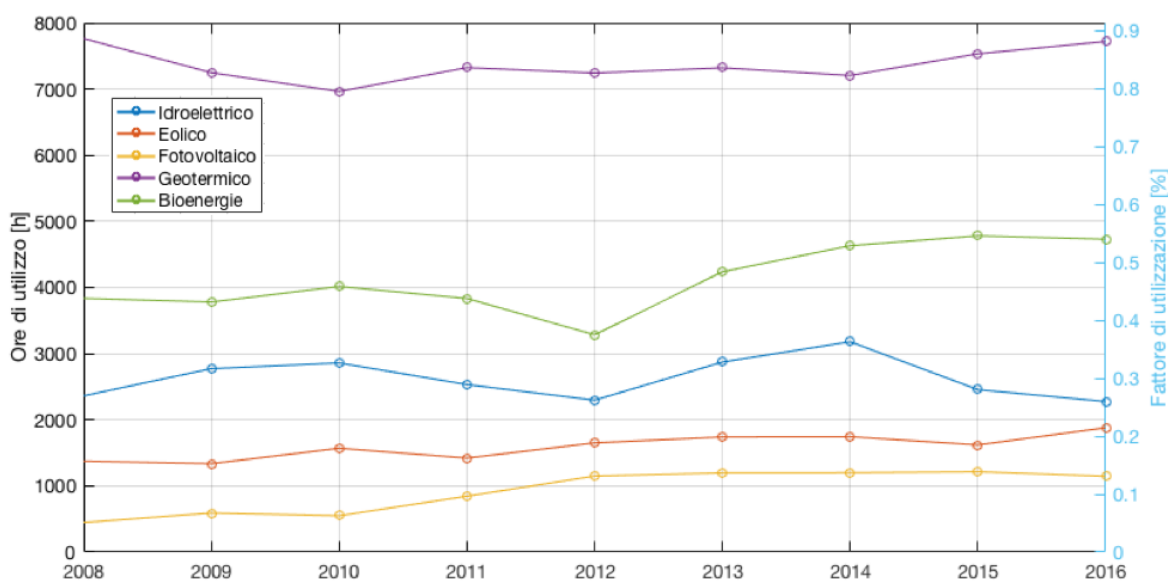


Figura 44 - Ore di utilizzo e fattore di utilizzazione

Considerate tutte le FER, quella geotermica è la più affidabile in termini di continuità di produzione con più di 7000 h/anno (fattore di utilizzazione medio 84%), in quanto è disponibile ininterrottamente ed è analoga ad una centrale termoelettrica convenzionale dal punto di vista della tecnologia applicata (cicli Rankine a vapore dominante, con separazione flash oppure ORC a bassa temperatura). Seconda la categoria che raggruppa le bioenergie, anche qui le tecnologie impiegate sono classici cicli a combustione e la natura di tale fonte è indipendente dall'ora e condizione meteorologiche come le restanti. Come atteso, idroelettrico, eolico e solare fotovoltaico sono le fonti meno sfruttabili proprio per l'intermittenza e non programmabilità che complicano ulteriormente la situazione, peggio fra tutti il fotovoltaico che ad esempio nel 2014 ha lavorato per meno della metà delle ore rispetto l'idroelettrico che quindi a parità di potenza installata quell'anno ha prodotto di conseguenza.

Lo stesso discorso vale anche per l'eolico, le ore di esercizio rimangono limitate ad un massimo di 2000 ore annue (22% di fattore di utilizzazione) comunque più del solare ma con una potenza installata decisamente inferiore.

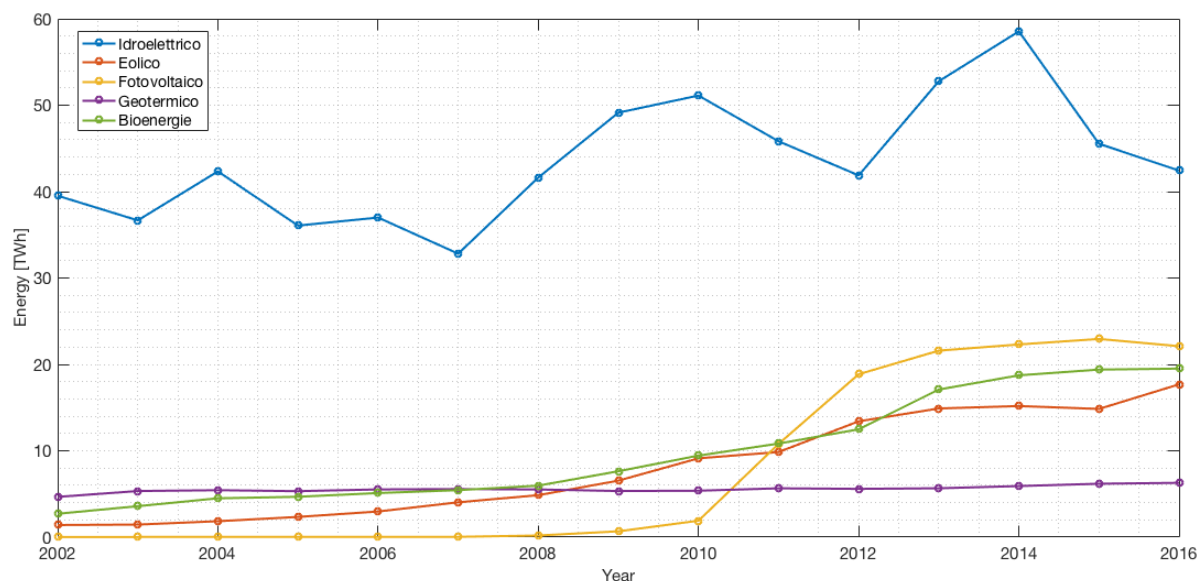


Figura 45 – Energia prodotta da FER : dettaglio per fonte.

In riferimento alla Figura 45, mentre fino al 2008 l'andamento dell'energia generata da FER era legato principalmente alla fonte idraulica, negli ultimi anni è cresciuta progressivamente l'importanza delle "nuove rinnovabili" che sono solare, eolica e bioenergie.

La fonte idraulica, in particolare, ha superato al 2016 un valore di produzione pari a 42 TWh, ossia un valore molto inferiore rispetto ai massimi osservati del 2014 ma comunque non è il valore più basso (che risale al 2007), rispetto alle altre fonti la produzione da idroelettrico è la più variabile, sicuramente.

La fonte solare ha contribuito con un valore di produzione di circa 22 TWh, in calo per il primo anno in assoluto (- 3,7%); tale diminuzione è attribuibile principalmente a un irraggiamento sul territorio nazionale più basso rispetto all'anno precedente piuttosto che ad un effettiva diminuzione delle ore di esercizio. E' da notare che al 2016 bensì la potenza di fotovoltaico abbia raggiunto (e superato) quella di idroelettrico (Figura 42), l'energia prodotta da fotovoltaico pur essendo cresciuta proporzionalmente all'installazione di nuovi impianti sia comunque la metà rispetto alla stessa prodotta con l'idroelettrico.

Stupisce invece la notevole performance della produzione eolica, in crescita di quasi 3 TWh rispetto al dato del 2015 (+19,2%) con quasi 18 TWh raggiunti. La produzione da bioenergie nel 2016 si è attestata invece sui 19 TWh, con un debole incremento, lo 0,6 % in più rispetto l'anno precedente.

Per quanto riguarda la geotermia il livello di energia prodotta è ben assestato su valori che si modificano di pochi punti percentuali di anno in anno, mediamente si ragiona sull'ordine di 5-6 TWh prodotti annualmente.

Al 2016 la prima FER per produzione rimane pur sempre l'idroelettrico ma non per potenza installata dove è superata (di poco) dal fotovoltaico seguito da eolico, bioenergie e geotermico ma dal punto di vista della produzione le bioenergie si impongono sull'eolico causa le limitate ore di esercizio di quest'ultimo. Per consultare i valori dettagliati si rimanda alla Tabella 27 situata al fondo dell'appendice A, pagina 132.

2.5.3. Commodity energetiche per la produzione di energia elettrica

In Italia il parco di produzione è molto vasto e si compone grossi impianti di gestiti da aziende da sempre influenti nel settore come Enel , Eni o Iren nel Torinese, con l'avvento della liberalizzazione del mercato sono nati moltissimi impianti di piccola potenza spartiti su altrettanti micro-produttori. Il mercato delle materie prime sicuramente influisce sul commercio dell'energia prodotta, come e quanto siano legate non è facile da stabilire ma si può valutare come si è modificata la produzione per fonte in Italia. Le fonti fossili tradizionali usate come combustibili per la produzione di energia si classificano per stato di aggregazione:

- **Solidi** : carbone, generalmente lignite che veniva utilizzata in diversi poli di produzione Italiani di grossa taglia per la produzione continua del baseload.
- **Liguidi** : derivati del petrolio che sono maggiormente utilizzati in ambito dei trasposti.
- **Gassosi** : in sostanza gas naturale, negli ultimi tempi ha sostituito grossa parte degli impianti per la produzione continua a seguito della conversione e smantellamento dei vecchi poli a carbone o combustili liquidi.

In ambito di produzione i combustibili fossili vengono utilizzati su impianti che hanno poca flessibilità di utilizzo e quindi forniscono il carico di base, leggerissime fluttuazioni possono essere fronteggiate dalla regolazione di centrali che utilizzano cicli a vapore ma per i carichi di punta si usano cicli a gas o impianti idroelettrici (anche di pompaggio) di grossa taglia .

Serie storica del prezzo del petrolio

Il petrolio è uno dei principali vettori energetici (e non solo) che viene scambiato sul mercato finanziario grazie a contratti detti futures che ne semplificano la negoziazione. Tali contratti di scambio sono stipulati per un certo prezzo e questo tipo di dati sono facilmente reperibili on-line per ampi periodi e su diverse scale temporali, I mercati principali per lo scambio del petrolio sono due : il **NYMEX** di New York e il **Intercontinental Exchange (ICE)** di Atlanta. Attualmente entrambi di proprietà statunitense e su questi mercati sono quotati contratti per la vendita di petrolio di due qualità che rappresentano il benchmark rispettivamente per il continente americano e quello europeo, africano e medio oriente (Russia inclusa) che sono:

- **WTI (West Texas Intermediate)** : anche noto come Texas Light Sweet, è un tipo di petrolio prodotto in Texas e utilizzato come benchmark nel prezzo del petrolio. È una materia prima di elevata qualità dalla cui raffinazione si ottiene un'alta percentuale di benzine e gasolio leggero. Il WTI nasce come petrolio estratto negli Usa.
- **Brent** : l'analogo per il continente Europeo, Brent è un giacimento di petrolio scoperto nel 1971 nel Mare del Nord il cui sfruttamento è iniziato nel 1976. Il termine brent caratterizza oggi il petrolio di riferimento europeo, un prodotto molto leggero, risultato dell'unione della produzione di 19 campi petroliferi situati nel Mare del Nord. Malgrado una produzione limitata, il brent serve

da petrolio grezzo di riferimento a livello mondiale ed il suo prezzo determina quello del 60% del petrolio estratto su scala mondiale.

In passato il Brent era quotato all' international petroleum exchange (IPE) di Londra, ma ad oggi è anch'esso quotato sui mercati americani soprattutto. Il petrolio viene commerciato per consegna immediata (spot) su unità di 1000 barili non frazionabili oppure come futures a Cushing (Oklahoma, USA) e Sullom Voe (Gran Bretagna) e tutte le quotazioni avvengono in dollari. In realtà, le transazioni di petrolio WTI e Brent Blend costituiscono solo una piccola parte del totale degli scambi, ma i prezzi di questi scambi sono utilizzati come prezzo di riferimento per gli altri, vi sono comunque altri benchmarks come il Dubai, Tapis e Isthmus ma sono largamente meno utilizzati.

Nella pratica commerciale, ogni petrolio è quotato rispetto al benchmark di riferimento più una differenza (detta premium), che può essere negativa o positiva. La differenza esistente tra il petrolio in questione ed il benchmark di riferimento è funzione essenzialmente della qualità. Petroli più leggeri o con un contenuto in zolfo minore del loro benchmark di riferimento saranno scambiati con un premium positivo; l'inverso se sono più pesanti o hanno un contenuto in zolfo più elevato.

La ricerca si è svolta in parallelo su più database online per capire quale fosse il più affidabile in termini di continuità temporale dei dati e in fine ho confrontato nel dettaglio i dati provenienti da due fonti ^{17 18} [72] [73] con un riscontro positivo e quindi riporto a seguito Figura 46 il plot completo da Aprile 2004 a Luglio 2018 su scala giornaliera.

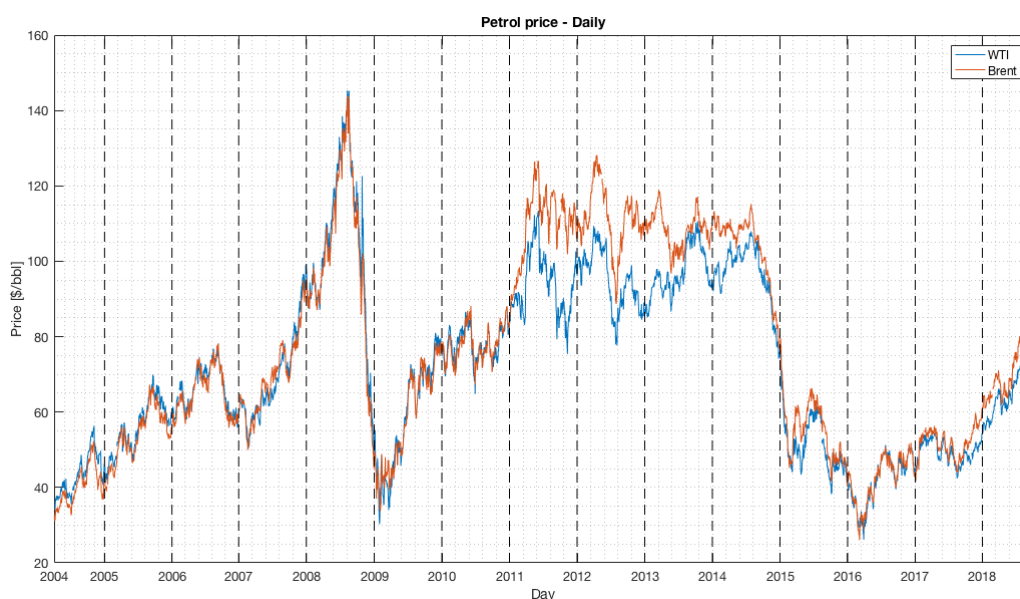


Figura 46 – Storico dei prezzi del petrolio: scala giornaliera.

La serie a scala giornaliera non è completa, su un periodo di 14 anni e quindi più di 5100 giorni si hanno solamente dati relativi a poco più di 3660 giorni che coprono il 70 % del periodo, questo dipende anche dal mercato stesso del petrolio, per cui sono da considerare solo i prezzi per i giorni feriali escludendo

¹⁷ https://www.eia.gov/dnav/pet/pet_pri_spt_s1_a.htm

¹⁸ <https://it.investing.com>

ulteriori giorni festivi infrasettimanali che non coincidono assolutamente con le festività nel calendario italiano.

In Figura 46 risalta il picco raggiunto nel 2008 che separa nettamente in due l'annata dove a cavallo dei mesi estivi ha toccato un massimo assoluto per poi subire un crollo considerevole entro fine anno che ha riportato a valori pari a quelli di inizio 2004. Le cause sono da attribuirsi a molteplici fattori ma prima fra tutte è la grande recessione che ha avuto inizio negli Stati Uniti d'America nel 2007 a seguito di una crisi sul mercato immobiliare che si è poi tradotta in una crisi su tutta l'economia americana per poi assumere gradualmente carattere globale. A seguito i prezzi sono tornati a crescere fino al 2011 (indice di una parziale ripresa economica) e da lì si sono mantenuti verosimilmente stabili fino a buona parte del 2014 per poi crollare nuovamente nel corso di pochi mesi toccando anche valori di poco superiori ai 25 \$/bbl a inizio 2016, dai qui ad oggi il trend è di nuovo in crescita. Le cause del secondo crollo di fine 2014 sono da imputare a fattori diversi rispetto al 2008, l'Arabia Saudita, essendo l'unico paese veramente in grado di scegliere come manipolare il mercato (a causa delle immense scorte a sua disposizione) a seguito dell'aumento considerevole delle trivellazioni da parte delle maggiori compagnie petrolifere statunitensi, ha aumentato l'esportazione di greggio forzando così una forte riduzione dei prezzi producendo un aumento dell'offerta a pari domanda, così che per gli Stati Uniti divenisse svantaggioso continuare a investire sulle nuove e discusse tecnologie che altrimenti gli avrebbero permesso di consolidarsi sul mercato mondiale del greggio.

Tolte le considerazioni di carattere geopolitico legate al prezzo di una commodity energetica così importante come il petrolio è curioso osservare anche la serie su scala annuale, sicuramente non ha molto senso mediare sull'anno il valore di una quantità così volatile in quanto si perdono informazioni utili a riguardo, il 2008 è risultato mediamente più economico del periodo 2011-2015 anche se scendendo nel dettaglio non è così, almeno non per tutto l'anno e sicuramente per l'influenza che ha una variabile del genere su qualsiasi mercato globale il confronto annuale ha senso tenendo in conto certe accortezze necessarie. Ad esempio, il confronto svolto in Figura 22 è giustificato dal fatto che si vuole confrontare un trend annuale di due grandezze fortemente volatili in termini di ordini di grandezza, se dagli stessi si volessero trarre informazioni quantitative per un singolo anno allora l'analisi dovrebbe essere più approfondita. Solamente ragionando su scala mensile la serie rimane comunque ricca di informazioni ma anche più maneggevole:

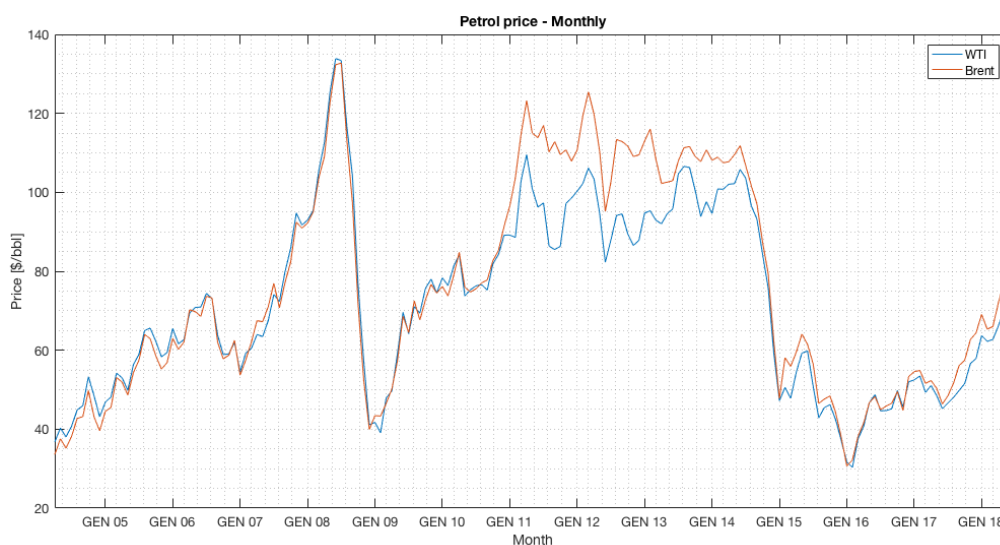


Figura 47 - Storico dei prezzi del petrolio : scala mensile.

3.1. INTRODUZIONE

La previsione di serie temporali ha sempre rappresentato un argomento di notevole importanza in ambito scientifico, in letteratura vi sono molti riferimenti a classi di modelli (Capitolo1: RICERCA IN LETTERATURA) con applicazioni che variano in molti campi di interesse, l'approccio classico ad una previsione consiste nella costruzione di un modello basato su:

- a) Principi di base e formulazioni matematiche
- b) Dati da elaborare

Eppure non è sempre possibile costruire un modello in questi termini, basti pensare alle applicazioni in campo economico e finanziario dove la componente umana rende impossibile la stesura di equazioni esatte ma sono disponibili grosse moli di dati, oppure un fenomeno fisico descritto perfettamente da equazioni differenziali come ad esempio la meccanica dei fluidi, la meccanica classica, l'elettromagnetismo o ancora la trasmissione del calore dove le formulazioni matematiche permettono una previsione accurata ma normalmente acquisire dati a riguardo è difficile se non impossibile, in base alle circostanze una misura può essere accessibile o meno ed è importante considerare il costo della misura.

3.2. METODO

Il metodo utilizzato in questa trattazione trova le sue origini nel 1987 secondo la previsione non lineare proposta da Farmer e Sidorowich [74]; esso basa la previsione sulla sola serie storica della variabile da prevedere come input. Il modello proposto allora sfrutta la natura caotica di alcune serie temporali e la utilizza come punto di forza per la previsione a breve termine (STNLP: short term non linear prediction). Una serie a natura puramente casuale è tale se presenta un certo e irriducibile numero di gradi di libertà tutti indipendenti tra di loro che ne rendono difficile la comprensione e di conseguenza la previsione, anche se fossimo in grado di comprendere a fondo il meccanismo evolutivo della serie, la mole computazionale per elaborazioni di previsione sarebbe consistente e forse insostenibile.

Invece, il caos è allo stesso modo una fonte di casualità nella serie che è presente anche in sistemi deterministici che però permette di rappresentare l'evoluzione della serie con semplici relazioni deterministiche su pochi gradi di libertà, da qui si apprende come una serie che pare casuale e complicata possa invece essere tanto più facilmente descritta quanto più è caotica (quindi con una forte componente deterministica) nella sua evoluzione.

Da questo punto di vista sarebbe utile determinare a priori la natura di una serie, quantificando le componenti stocastica e deterministica, ma chiaramente la questione è molto complessa e tutt'ora risulta difficile stabilire a priori l'accuratezza di un modello predittivo, in questo senso l'esperienza ed il buon senso sono di aiuto.

Il suddetto metodo utilizza il concetto di “**nearest neighbors**” per apprendere informazioni sull’evoluzione dei dati fino all’ultimo valore disponibile, quindi il numero di vicini necessari alla previsione ottima è un indice diretto di quanto sia caotica la serie analizzata, in una serie fortemente deterministica sono necessari pochi vicini per cogliere tutta l’informazione, al contrario se prevale la componente stocastica ma in ogni caso sono considerazioni da farsi a posteriori senza informazioni certe in partenza, l’abilità di chi se ne occupa sta nell’identificare la quantità di informazioni necessarie a descrivere la variabile indagata, che non è solo lo storico ma anche tutte le altre variabili che la influenzano. Analizzare l’autocorrelazione a diversi ritardi può essere un ottimo indice di caos nella serie, essa evidenzia la periodicità di una serie e ne quantifica la misura.

Ulteriori sviluppi hanno portato ad applicazioni a serie caotiche come quelle analizzate in idrologia in merito alle piene dei fiumi le quali sono caratterizzate da grandi non linearità, la trattazione presentata in [75] ammette la possibilità di includere serie di dati relativi a variabili endogene direttamente correlate a quella studiata ma la formulazione matematica è la medesima di [74], in più l’autore propone un metodo per la stima delle fasce di confidenza sulla previsione.

L’applicazione di interesse in questa trattazione è appunto la previsione del prezzo dell’energia elettrica sul mercato italiano, tale variabile è caratterizzata da forte volatilità su scala giornaliera e viene influenzata da un gran numero di variabili esterne quali i prezzi delle commodity energetiche e le condizioni atmosferiche nonché scelte di politica energetica e non solo. Dal punto di vista matematico il modello consiste di tre passi:

- I. Ricostruzione dello spazio delle fasi
- II. Ricerca di una relazione funzionale che legghi i valori passati a quelli futuri
- III. Previsione

Ogni passo richiede accorgimenti illustrati nel dettaglio nei paragrafi a seguire, una questione cardine sta nella valutazione della bontà della previsione che può essere valutata solamente se è disponibile un confronto con i dati reali, per questo motivo si provvede a dividere la serie in ingresso in tre set di dati in ordine cronologico in modo da trarre informazioni quantitative riguardo la precisione della previsione:

- **Fitting set**: utilizzato per costruire l’attrattore¹⁹ del sistema dinamico, sul quale i vettori utilizzati per ricostruire lo spazio delle fasi (14) descrivono le traiettorie di evoluzione temporale proprie del sistema.
- **Validation set** : sul quale si calibrano i parametri ($m_v \tau k_s$) descritti in nel paragrafo successivo, se ne ricava una tripletta di valori ottimali per cui la previsione ottenuta è la migliore in tutto lo spazio delle fasi, per farlo è necessario esplorare ampi intervalli di valori causa la mancanza di metodi certi per la determinare a priori i valori ottimi, sotto questo aspetto la domanda computazionale può essere importante.
- **Testing set**: costituito da pochi punti su cui verificare l’accuratezza della previsione utilizzando i parametri ottimali risultato della calibrazione sul validation set.

¹⁹ <https://it.wikipedia.org/wiki/Attrattore>

3.2.1. Ricostruzione dello spazio delle fasi

Si vuole valutare la previsione di una variabile la cui serie temporale storica è nota, partendo dall'ultimo valore noto $x_1(t_i)$, la serie viene campionata su intervalli regolari per $i = 1, \dots, N$ con passo pari a Δt (orario nel caso dei prezzi dell'energia elettrica), si considerano poi $(n - 1)$ serie di variabili correlate che compongono il sistema. Sulla base dei dati il sistema dinamico viene ricostruito nello spazio delle fasi come proposto da Takens [76] secondo il metodo del tempo di ritardo. Noti i valori delle dimensioni di embedding m_v ($v = 1, \dots, n$) e i tempi di ritardo τ_v ($v = 1, \dots, n$) si procede costruendo i vettori come segue:

$$\begin{aligned} \mathbf{x}(t_i) = \{ & \mathbf{x}_1(t_i), \mathbf{x}_1(t_i - \tau_1), \mathbf{x}_1(t_i - 2\tau_1), \dots, \mathbf{x}_1(t_i - (m_1 - 1)\tau_1), \mathbf{x}_2(t_i - \delta), \\ & \mathbf{x}_2(t_i - \delta - \tau_2), \dots, \mathbf{x}_2(t_i - \delta - (m_2 - 1)\tau_2), \dots, \mathbf{x}_n(t_i), \mathbf{x}_n(t_i - \tau_n), \dots, \mathbf{x}_n(t_i - \\ & (m_n - 1)\tau_n) \} \end{aligned} \quad (14)$$

Dove δ indica un eventuale sfasamento fra diverse porzioni del vettore relative alle diverse variabili e la somma $\sum_v m_v = m$ è la dimensione dello spazio delle fasi così costruito che appare come somma di n sottospazi relativi alle singole variabili considerate nell'analisi, ogni vettore avrà quindi dimensione m . I tempi di ritardo possono essere costanti oppure variare in base alla posizione, in modo da considerare solamente i punti che contengono più informazione ed evitare di introdurre rumore.

Così facendo l'evoluzione del sistema dinamico è schematizzata tramite le traiettorie descritte sull'attrattore dai vettori sopracitati, la rappresentazione del sistema nello spazio delle fasi permette di visualizzare l'evoluzione temporale di quest'ultimo come su volume fisico con parametri m, τ, k dove k indica il numero di vicini da considerare per la previsione, i punti da considerare sono scelti calcolando la distanza euclidea di tutti i punti dal punto di inizio previsione ed ordinandoli per entità.

Farmer e Sidorowich [74] trattano l'argomento secondo un punto di vista matematico e puramente teorico, più qualche applicazione con riscontri relativi all'applicabilità del metodo mostrando gli intervalli di valori dei parametri per cui la previsione risulta migliore di quella pari alla media per tutti i punti previsti. Sulla base di tale approccio risulta possibile stabilire un criterio per la scelta della dimensione dello spazio delle fasi ricostruito (m), definito l'attrattore come l'insieme di stati verso cui evolve il sistema dinamico dopo un tempo lungo abbastanza, quale che sia la sua struttura è definita una dimensione D dell'attrattore per cui esiste un vincolo :

$$m \geq D$$

La difficoltà sta appunto nel calcolo del valore D che per alcune applicazioni, come quelle di [74], è possibile calcolare, ma generalmente esso è difficilmente calcolabile per la maggior parte delle applicazioni di interesse comune alla previsione.

3.2.2. Ricerca di una relazione funzionale

Esclusa la presenza di rumore dinamico, l'evoluzione temporale del sistema è descritta nello spazio delle fasi m - dimensionale da una mappa del tipo:

$$\mathbf{F} : \mathbb{R}^m \rightarrow \mathbb{R}^m$$

$\mathbf{F}$ rappresenta una relazione funzionale che lega il valore all'istante corrente $\mathbf{x}(t)$ ad un valore futuro a cui si è interessati ai fini della previsione, perciò è possibile ottenere una previsione applicando tale relazione al punto di interesse:

$$\mathbf{x}(t + T) = \mathbf{F}(\mathbf{x}(t)) \text{ dove } T = \text{Orizzonte di previsione} \quad (15)$$

In questi termini il problema si riduce a trovare un predittore costituito da una mappa approssimante $\hat{\mathbf{F}}$ con un'espressione semplificata per la previsione che ha come input il medesimo valore temporale:

$$\hat{\mathbf{x}}(t + T) = \hat{\mathbf{F}}(\mathbf{x}(t)) \text{ dove } \hat{\mathbf{F}} : \mathbb{R}^m \rightarrow \mathbb{R} \quad (16)$$

$\hat{\mathbf{F}}$ è una approssimazione di $\mathbf{F}$ nel senso che è definita con dominio m -dimensionale e codominio monodimensionale in quanto il solo elemento di interesse per l'applicazione è appunto il primo: $\hat{\mathbf{x}}(t + T)$. Se la serie di dati in ingresso è caotica, allora, la funzione $\mathbf{F}$ sarà necessariamente non lineare ed è possibile approssimarla seguendo diversi approcci, tre nello specifico:

- **Approssimazione globale** : si assume una forma standard della mappa approssimante, generalmente una forma polinomiale di ordine p e dimensione m definita su tutto lo spazio delle fasi in m dimensioni, i coefficienti vengono fittati sul data set usando una regressione lineare multivariata basata sul metodo dei minimi quadrati. È possibile effettuare previsioni per orizzonti di anche $2T, 3T, \dots$, componendo $\mathbf{F}$ con sé stessa.

Lo svantaggio principale è che ad ogni composizione gli errori di approssimazione crescono esponenzialmente, ovviamente per previsioni a lungo termine è opportuno evitarlo. Inoltre, la ricerca un'unica relazione che sia valida sull'intero dominio presenta delle limitazioni in quanto risente di marcate non linearità e disomogeneità nelle diverse zone dell'attrattore, quindi anche sotto questo punto di vista l'approssimazione globale presenta delle debolezze.

- **Approssimazione locale** : usa comunque una funzione polinomiale di ordine p per approssimare, ma prevede di procedere al fitting degli m coefficienti della regressione per ogni step T . Così facendo è necessario ricalcolare più volte i coefficienti del polinomio ma l'approssimazione sulla porzione di attrattore considerata è molto più precisa rispetto ad un'unica funzione, in un'applicazione di previsione è fondamentale ricostruire il più fedelmente possibile l'attrattore su cui si baseranno le traiettorie descritte nella previsione.

- **Rielaborazione come equazione differenziale:** rappresenta sicuramente la descrizione più dettagliata del problema e dell'attrattore, la si ottiene scrivendo l'equazione (15) in forma differenziale per sostituzione di $\mathbf{x}(\mathbf{t} + \mathbf{T})$ con il suo integrale, come segue:

$$\int \mathbf{x}(\mathbf{t} + \mathbf{T}) \, d\mathbf{t} = \mathbf{F}(\mathbf{x}(\mathbf{t})) \quad (17)$$

Risolvere l'equazione precedente implica conoscere la forma esatta della funzione $\mathbf{x}(\mathbf{t} + \mathbf{T})$ dei valori futuri e la sua dipendenza da altre variabili di cui normalmente non siamo a conoscenza, se lo fossimo la previsione sarebbe ottimale in ogni caso, in quanto si disporrebbe di tutte le informazioni correlate alla variabile predetta, e come già discusso in precedenza, questo è pressoché impossibile da stabilire a priori. Per cui l'equazione (17) descrive esattamente la dinamica del sistema ma è una formulazione matematica fine a sé stessa considerate le applicazioni reali.

Un problema comune a tutti e tre gli approcci è che il numero di parametri da determinare per un generico polinomio di ordine p è: $\frac{(p+m)!}{m!p!} \approx m^p$, che diventa insostenibile per valori di m troppo elevati.

Risultati delle analisi condotte in [74] suggeriscono l'approssimazione locale come approccio più effettivo, quindi non si considera l'intero spazio delle fasi per effettuare la previsione, ma solamente la zona limitrofa al punto di partenza $\mathbf{x}(t_0)$. La scelta riguardo la porzione di spazio da considerare si effettua introducendo una metrica per quantificare la distanza dei punti nello spazio delle fasi:

$$\mathbf{d} = \|\mathbf{x}(\mathbf{t}_0) - \mathbf{x}(\mathbf{t}')\| \text{ con } \mathbf{t}' < \mathbf{t} \quad (18)$$

a seguito si cercano i k punti più vicini al punto di partenza $\mathbf{x}(t_0)$, quindi cerco il minimo valore di $\mathbf{d}$ per cui k stati del sistema descritti da $\mathbf{x}(t')$ hanno tutti distanza minore di $\mathbf{d}$ da $\mathbf{x}(t_0)$. In altri termini si cerca il minimo valore del raggio di una sfera (pari a $\mathbf{d}$) centrata in $\mathbf{x}(t_0)$, entro la quale sono racchiusi k punto antecedenti a $\mathbf{x}(t_0)$. Il numero di vicini è un parametro che va calibrato sul data set in ingresso ed è il parametro che grava sul tempo computazionale per la calibrazione del codice, tanto più è elevato numero di vicini e tanto maggiore è il tempo necessario per la ricerca di questi, soprattutto se la dimensione m dello spazio delle fasi è elevata. In generale, un data set contenente N elementi richiede lo stesso numero di step computazionali per la ricerca di un punto vicino, quindi $k * N$ operazioni solamente per la ricerca dei vicini, la serie utilizzata in questa analisi presenta $N = 122712$ prezzi orari dal 2004 ad oggi, per cui è chiaro che quanto più caotica è la serie è tanto più veloce la calibrazione dei parametri.

Farmer e Sidorowich [74] citano il suddetto problema di onere computazionale proponendo una soluzione riportata da [77] che promette la riduzione da N steps a $\log(N)$, a seguito della partizione del data set tramite un “*decision tree*”.

Successivamente è necessario stabilire quali e quanti punti sono utili e non superflui alla previsione, anche questo passaggio assume caratteristiche non proprio banali perché non si saprà mai con esattezza come sono distribuiti i punti e perciò o li si controlla uno ad uno oppure si introdurrà per forza del rumore, due linee guida per la scelta del numero di vicini sono:

- 1) È bene limitare il valore di k considerando solamente i vicini più significativi, in particolare i punti di picco.
- 2) Dall'altro lato bisogna scegliere k grande abbastanza la fine di ridurre l'effetto del rumore sulla previsione.

Dopo queste considerazioni è chiaro che la scelta di k è tutt'altro che semplice, dipende dalle caratteristiche del set di dati in ingresso e incide molto sul tempo richiesto per il calcolo, perciò è bene considerare caso per caso e non è possibile stabilire delle regole ferree generali, una serie dove prevale il determinismo produrrà risultati migliori con un numero di vicini minore e viceversa se a prevalere è la componente stocastica. Matematicamente bisogna trovare il modo più efficace per pesare il contributo di ogni vicino al valore predetto, lo si fa utilizzando un'approssimazione polinomiale contemplando due possibilità:

- **Approssimazione di ordine zero:** la più semplice ed immediata, si costruisce il predittore locale basandosi solamente sul vicino più prossimo ($k = 1$) e quindi si associa semplicemente quel valore alla previsione:

$$\mathbf{x}_{\text{pred}}(\mathbf{t} + \mathbf{T}) = \mathbf{x}(\mathbf{t}' + \mathbf{T}) \text{ con } \mathbf{t}' < \mathbf{t} \quad (19)$$

- **Approssimazione lineare o di ordine superiore:** approccio di qualità superiore che prevede di fittare le coppie $(\mathbf{x}(\mathbf{t}'), \mathbf{x}(\mathbf{t}' + \mathbf{T}))$ secondo un polinomio di ordine p ($p=1$ lineare) e m variabili i cui coefficienti moltiplicativi vengono calcolati in base al numero di vicini di cui si vuol tenere conto, con l'accortezza di utilizzare:

$$k > m$$

Più precisamente $k = m + 1$ corrisponde ad un'interpolazione lineare ma per assicurare la stabilità della soluzione è vantaggioso utilizzare $k \geq m + 1$, in altri termini il grado minimo di modellazione teorica è pari a $m+1$ ma la presenza di rumore nella serie costringe a utilizzare valori di k maggiori. Così facendo il problema risulta sovradeterminato e i parametri della dinamica locale si stimano con il metodo dei minimi quadrati.

Farmer e Sidorowich [74] analizzano valori differenti di p e osservano che effettivamente, con l'aumentare dell'ordine del polinomio approssimante, aumentano le prestazioni di previsione, ma è comunque un incremento molto scarso e per questo motivo si suggerisce di utilizzare un'approssimazione lineare che rappresenta il compromesso ideale tra accuratezza della previsione e mole computazionale.

Siccome in letteratura non vi sono spunti riguardo applicazioni al mercato elettrico, è opportuno considerare diversi valori di p come ulteriore grado di libertà del problema, al fine di non tralasciare alcuna variabile in gioco.

Concludendo, si rende necessaria la calibrazione di tre parametri fondamentali su cui non abbiamo informazioni precise a priori:

- Dimensione di embedding m
- Fattore di ritardo τ
- Numero di vicini k_s

I primi due contribuiscono alla costruzione dello spazio delle fasi secondo l'equazione (14) attraverso vettori che descrivono l'evoluzione del sistema come traiettorie che compiono sull'attrattore, l'ultimo è fondamentale per poter trattare semplicemente il problema attraverso semplici relazioni locali che permettono di tralasciare la non linearità globale del sistema senza prescindere le prestazioni. A questi si aggiungono due parametri necessari per definire completamente la previsione:

- Orizzonte di previsione T
- Grado del polinomio approssimante p

Influenti sulle performance del codice, in generale la precisione decresce all'aumentare dell'orizzonte di previsione e quindi per ottenere buone previsioni a lungo termine è più efficace eseguire n steps ad orizzonte T piuttosto che un singolo passaggio ad orizzonte $T' = n T$, ovviamente se ne paga il prezzo in termini di tempo di elaborazione. [74] propone il calcolo del massimo orizzonte temporale per tre applicazioni diverse, esso è raggiunto quando la previsione produce uno scarto quadratico medio (RMSE) uguale a quello che si avrebbe approssimando la previsione con il valore medio del dataset, rendendo così inutile la previsione.

Invece il grado p è molto meno influente in termini quantitativi, non è possibile stabilire con certezza come variano le prestazioni con il grado del polinomio ne tantomeno trovare il valore ottimo a prescindere dalla terna m, τ e k_s . Le differenze sono davvero molto lievi e le prestazioni si mantengono al di là che il polinomio sia lineare, quadratico o cubico, nei capitoli a seguire viene verificato e poi si stabilisce $p = 2$ fissato.

3.2.3. Previsione

Partizionando localmente l'attrattore si rende possibile di descrivere una dinamica altrimenti estremamente complessa attraverso una semplice relazione senza che le ripercussioni vadano ad incidere sulla qualità della previsione. L'applicazione di interesse è lo studio del mercato del giorno prima con relativa previsione del prezzo orario, per stabilire quale sia la relazione migliore si fa riferimento a [75] dove lo stesso modello è applicato in ambito idrologico, la mappa approssimante è scelta sulla base delle considerazioni proposte da Porporato e Ridolfi [78] che considerano un mappa a parametri costanti $\mathbf{a}$, quindi la struttura di $\hat{\mathbf{F}}$ non è altro che una regressione lineare:

$$\hat{\mathbf{x}}(\mathbf{t} + \mathbf{T}) = [\mathbf{1}, \mathbf{x}(\mathbf{t})] \mathbf{a} \quad (20)$$

Detto questo, la non linearità che dà anche il nome al metodo non deriva dalla forma che assume la mappa approssimante ma bensì dalla suddivisione del suo dominio che è resa necessaria proprio dalla natura disomogenea e non lineare del sistema. A questo punto i parametri $\mathbf{a}$ si ricavano dalla soluzione di un sistema lineare locale:

$$\mathbf{z} = \mathbf{Y} \mathbf{a} \quad (21)$$

Dove $\mathbf{z} \in \mathbb{R}^{k \times 1}$ è un vettore colonna dei valori predetti, $\mathbf{Y} \in \mathbb{R}^{k \times 2}$ è la matrice dei valori noti costruiti secondo (14) e $\mathbf{a} \in \mathbb{R}^{m \times 1}$ è un vettore colonna dei parametri a cui siamo interessati. Riscrivendo la (21) per esteso si ottiene:

$$\begin{pmatrix} \mathbf{x}(\mathbf{t}_1 + \mathbf{T}) \\ \mathbf{x}(\mathbf{t}_2 + \mathbf{T}) \\ \dots \\ \mathbf{x}(\mathbf{t}_k + \mathbf{T}) \end{pmatrix} = \begin{bmatrix} \mathbf{1}, \mathbf{x}(\mathbf{t}_1) \\ \mathbf{1}, \mathbf{x}(\mathbf{t}_2) \\ \dots \\ \mathbf{1}, \mathbf{x}(\mathbf{t}_k) \end{bmatrix} \begin{pmatrix} \mathbf{a}_0 \\ \mathbf{a}_1 \\ \dots \\ \mathbf{a}_m \end{pmatrix} \quad (22)$$

Ora, si procede applicando il metodo dei minimi quadrati per la stima del vettore di parametri $\hat{\mathbf{a}}$ del modello:

$$\hat{\mathbf{a}} = (\mathbf{Y}^T \mathbf{Y})^{-1} \mathbf{Y}^T \mathbf{z} \quad (23)$$

E poi si minimizzano i residui $\mathbf{e}$ tra i valori reali e quelli calcolati

$$\mathbf{e} = \mathbf{z} - \hat{\mathbf{z}} = \mathbf{z} - \mathbf{Y} \hat{\mathbf{a}} \quad (24)$$

Quindi, a questo punto la previsione viene effettuata applicando i parametri stimati sulla base dei dati reali $\hat{\mathbf{a}}$ all'equazione (20), partendo dal punto iniziale $\mathbf{x}(t_0)$.

3.3. STIMA DELL'ERRORE

A prescindere dal fatto che il codice sia in fase di calibrazione sul validation set oppure in fase di previsione effettiva sul testing set, è sempre necessario poter quantificare l'efficacia della calibrazione in termini di precisione dei dati predetti, per farlo è necessario avere sempre il riferimento ai valori reali ed è per questo che si suddivide il data set nei tre sottoinsiemi. La suddivisione permette di lavorare su serie distinte che siano separate da un gap temporale tale da minimizzare l'autocorrelazione che intercorre fra i tre, nel caso analizzato si ha autocorrelazione massima a scala settimanale e minima alle dodici ore (vedi 3.4).

Stabilire l'efficienza di una previsione non è univoco, vi sono diverse metriche possibili e confrontabili, alcune più robuste ed altre meno, si può ragionare in termini assoluti dell'errore o relativi a un valore di soglia fissato con minimo di prestazioni sotto il quale non ha senso scendere. Innanzitutto si deve partire calcolando il vettore dei residui tra valori reali e predetti, anche detto errore assoluto:

$$\mathbf{AE} = |\mathbf{x}(\mathbf{t}_i + \mathbf{T})_{\text{reale}} - \mathbf{x}(\mathbf{t}_i + \mathbf{T})_{\text{prev}}| = \begin{pmatrix} \mathbf{AE}_1 \\ \mathbf{AE}_2 \\ \dots \\ \mathbf{AE}_{N_{\text{set}}} \end{pmatrix} \quad (25)$$

una rappresentazione grafica di questi in parallelo alla serie stessa fornisce informazioni riguardo ai punti più difficili da prevedere, solitamente i punti di picco. Lo scarto è un'informazione molto utile a scala locale, fornisce una misura della distanza dall'ottimo punto per punto, in prima sede se ne possono trarre informazioni globali calcolandone i valori massimo e medio:

$$\mathbf{AE}_{\text{MAX}} = \max(\mathbf{AE}) \quad (26)$$

$$\overline{\mathbf{AE}} = \frac{1}{N_{\text{set}}} \sum_{i=1}^{N_{\text{set}}} \mathbf{AE}_i \quad (27)$$

I valori di scarto hanno la stessa unità di misura della variabile considerata, nel caso del prezzo dell'energia elettrica si parla di $\left[\frac{\text{€}}{\text{MWh}}\right]$, il numero di punti su cui vengono mediati i valori è indicato da N_{set} che dipende dal set considerato, sarà quello di validazione se è in corso la calibrazione dei parametri oppure il testing se si applicano i parametri ottimi alla previsione effettiva. Per mantenere lo stesso tipo di informazione, un parametro di errore molto robusto indicativo di tutto l'intervallo è lo scarto quadratico medio (Root Mean Square Error), esso è utilizzato frequentemente in letteratura EPF come metrica di errore nella maggior parte delle pubblicazioni di natura econometrica (vedi sezione 3.3 di [1]), solitamente è calcolato su scala giornaliera o settimanale.

$$\mathbf{RMSE} = \sqrt[2]{\sum_{i=1}^{N_{\text{set}}} \frac{(\mathbf{x}_{\text{REALE}} - \mathbf{x}_{\text{PREV}})_i^2}{N_{\text{set}}}} = \sqrt[2]{\frac{1}{N_{\text{set}}} \sum_{i=1}^{N_{\text{set}}} (\mathbf{x}_{\text{REALE}} - \mathbf{x}_{\text{PREV}})_i^2} \quad (28)$$

Sempre [1] include l'errore medio percentuale (Mean Average Percentage Error) e sottolinea che con buona probabilità è la metrica più diffusa in letteratura insieme al RMSE, il MAPE aggiunge l'informazione relativa all'entità del punto predetto, in parole semplici a parità di scarto un punto di picco pesa di meno rispetto ad uno "normale", perciò in alcune situazioni può essere fuorviante se si parla di prezzi.

$$\mathbf{MAPE} = \left[\frac{1}{N_{\text{set}}} \sum_{i=1}^{N_{\text{set}}} \frac{\mathbf{AE}_i}{\mathbf{x}(t_i + T)_{\text{reale}}} \right] 100 \quad (29)$$

Secondo le note riguardo questi primi indicatori di carattere assoluto risulta che il più robusto per l'applicazione di interesse sia il RMSE, molto diffuso in applicazioni simili in letteratura. Per completezza vengono calcolati entrambi e confrontati in sede di calibrazioni dei parametri nel capitolo successivo. Esiste una estensione del RMSE in percentuale (RMSPE) e non più assoluto che tiene conto del rapporto residui e valore reale, come fa anche il MAPE:

$$\mathbf{RMSPE} = \sqrt[2]{\frac{1}{N_{\text{set}}} \sum_{i=1}^{N_{\text{set}}} \left(\frac{\mathbf{x}_{\text{REALE}} - \mathbf{x}_{\text{PREV}}}{\mathbf{x}_{\text{REALE}}} \right)_i^2} 100 = \quad (30)$$

Consideriamo ora un nuovo parametro suggerito da Farmer e Sidorowich [74] con qualche leggera modifica. L'approccio che si vuole proporre utilizza sempre la misura del RMSE ma questa volta lo rapporta un valore standard, fornendo un'informazione riguardo le prestazioni del metodo rispetto ad una previsione in cui i valori sono tutti presi pari alla media dei valori reali:

$$\bar{\mathbf{x}}_{i,\text{reale}} = \frac{1}{N_{\text{set}}} \sum_{i=1}^{N_s} \mathbf{x}(t_i + T)_{\text{reale}} \quad (31)$$

Ne deriva un valore positivo calcolato come il rapporto fra il RMSE della previsione ed il medesimo valore riferito alla previsione costante pari a $\bar{\mathbf{x}}_{i,\text{reale}}$:

$$\mathbf{E} = \sum_{i=1}^{N_{\text{set}}} \frac{\mathbf{AE}_i^2}{[\mathbf{x}(t_i + T)_{\text{reale}} - \bar{\mathbf{x}}_{i,\text{reale}}]^2} \quad (32)$$

È importante sottolineare che nel caso in cui la previsione a valori costanti risulti più efficiente, allora l'errore $E > 1$, e non ha alcun senso implementare un codice e occupare potenza computazionale per una previsione che risulta globalmente peggiore di quella sempre costante pari al valor medio. In questo modo si aggiunge un dettaglio non da poco, riferire tutto a quello che si ottiene con la previsione più banale ci permette di escludere categoricamente quelle combinazioni di parametri che comportano uno sforzo inutile, o per lo meno ci permettono di rendercene conto. Perciò $E \in \mathbb{R}^+$ la previsione è tanto più accurata quanto più il suo valore è prossimo allo zero, il valore unitario indica prestazioni eguali alla previsione costante pari a $\bar{x}_{i, reale}$. Dopodiché l'equazione (32) è stata banalmente riscritta in modo da esprimerne il complemento in percentuale, quindi assume la forma di efficienza percentuale che però può assumere valori negativi se le prestazioni inferiori alla previsione costante.

$$\varepsilon = (1 - E) 100 = \left[1 - \sum_{i=1}^{N_{set}} \frac{AE_i^2}{[x(t_i + T)_{reale} - \bar{x}_{i, reale}]^2} \right] 100 \quad (33)$$

Il valore di efficienza che si ottiene dall'equazione (33) sfrutta l'affidabilità del RMSE e in più fornisce un'informazione aggiuntiva, perciò abbinata al RMSE assoluto permette di descrivere esaustivamente le performance del set di parametri imposti.

Vi sono poi altre grandezze statistiche importanti, non strettamente come errore per misurare la distanza dalla previsione perfetta ma come indici di ripetitività nell'evoluzione della serie, per esempio viene calcolata la persistenza come indice di ergodicità della serie storica oppure un altro indice interessante è il coefficiente di correlazione lineare perché è un indice che esprime l'eventuale linearità che intercorre fra due variabili. Rappresenta il valore incluso nella matrice dei coefficienti di correlazione in cui la è diagonale unitaria perché riporta coefficienti relativi alla combinazione delle variabili con loro stesse, l'espressione applicabile ad un set di dati è la seguente:

$$\rho (\text{reale, prev}) = \frac{\sigma_{reale, prev}}{\sigma_{reale} \sigma_{prev}} \quad (34)$$

Nella recente trattazione di Nowotarsky e Weron [17] vengono analizzate le problematiche riguardo lo stato dell'arte dei nuovi modelli presentati nella letteratura, con l'intento di implementare [1]. Un intero capitolo è dedicato alla valutazione della metrica in cui si propone un'analisi empirica basata su modelli presentati durante la GEFCom2014 [18], tratta di “*Probabilistic forecasting*” in cui vengono assegnate probabilità specifiche per ogni evoluzione del sistema (come il modelli basati sui regime Markoviano presentati in 1.3). Per questi modelli in particolare non si osserva mai la vera distribuzione del processo perciò vengono introdotti i concetti di reliability (affidabilità) e sharpness i quali si utilizzano come metriche alternative anche per le altre classi di modelli. Per valutare queste caratteristiche sono suggeriti diversi indicatori e test applicabili in diversi contesti che però non rientrano nell'interesse di questa applicazione, si tratta di procedimenti più o meno complicati e comunque poco diffusi nella valutazione delle performance il letteratura per cui dal punto di vista di un confronto efficace con la letteratura sono di poco aiuto ma rappresentano gli sviluppi più innovativi in questo campo.

3.4. SERIE STORICA e AUTOCORRELAZIONE

Come già ribadito l'interesse è volto alla previsione della serie del Prezzo Unitario Nazionale (PUN) i cui dati sono disponibili dagli esiti del Mercato del Giorno Prima (MGP) come media ponderata sulle quantità dei prezzi zonali. I dati sono a scala oraria e non presentano discontinuità, perciò la previsione può essere effettuata ad orizzonte minimo pari ad un ora ed esso rappresenterà anche il caso a migliori prestazioni.

In campo di EPF la letteratura cita frequentemente il problema della caratterizzazione delle serie dal punto di vista delle non linearità e stazionarietà di queste, il problema è attuale e molto influente sulla qualità delle previsioni, ad esempio i modelli autoregressivi classici (AR) oppure integrati con metodi a media mobile (ARMA) lavorano bene se trattano serie stazionarie in cui le non linearità sono deboli, invece, serie complesse con marcata non linearità sono difficili da trattare, alcuni metodi che utilizzano reti neurali sono in grado di adattarsi bene ma non è scontato, in generale i metodi più calzanti sono quelli che imparano durante l'evoluzione quindi Machine Learning (ML, ELM). Siccome esiste un gran numero di variabili che contribuiscono alla formazione dei prezzi la serie che ne deriva è molto complessa, ad esempio è più difficile prevedere un profilo di prezzo rispetto ad un profilo di carico, proprio a causa della natura della variabile. I riferimenti di [20] e [21] includono uno studio preliminare della serie utilizzando test specifici: Dickey-fuller (DF e ADF) e autocorrelation function (ACF) specifici per valutare la stazionarietà a cui sovente si aggiunge il Engle Granger (EG) per la cointegrazione e test di non linearità con il Kapetanios test (KAP).

Sebbene servano come conferma non si è utilizzato alcun test, a prescindere le serie di prezzo sono fortemente non lineari per i motivi appena citati e lo stesso vale per la stazionarietà, secondo l'analisi al capitolo 2 è chiaro che la serie non è stazionaria. Comunque si studia l'autocorrelazione, costituisce una conferma da questo punto di vista ma l'interesse principale è identificare le periodicità e le scale su cui sono più forti.

L'autocorrelazione stabilisce il grado di correlazione della serie con se stessa attraverso il campionamento della funzione sul suo dominio. L'algoritmo prevede il calcolo della correlazione tra serie y_t e la stessa con un ritardo k imposto y_{t-k} , si costruisce poi un correlogramma per diversi valori di ritardo $k = 1, \dots, K$ facendo attenzione a che non vengano tralasciati troppi valori del vettore di partenza, per un dato ritardo si considerano gli elementi del vettore dal y_{t-k+1} in avere due vettori di pari lunghezza, in questo senso è bene limitare il lag in base al numero di valori T contenuti nella serie ($K < T/4$ include sempre almeno il 75% dei valori). In particolare, il coefficiente di correlazione è calcolato come rapporto fra covarianza e prodotto delle deviazioni standard [11], ma se si tratta della stessa serie allora si riduce a:

$$r_k = \frac{\sigma_{yy}}{\sigma_y^2} = \frac{\sum_{t=1}^{T-k} (y_t - \mu)(y_{t-k} - \mu)}{\sum_{t=1}^T (y_t - \mu)^2} \quad (35)$$

Il valore di tal coefficiente non supera mai l'unità in modulo ($-1 \leq r_k \leq 1$) è quindi può assumere valori negativi che indicano una correlazione indiretta fra le due serie, se il valore è positivo le serie sono direttamente correlate, per cui se ad un certo ritardo l'autocorrelazione è positiva implica una

periodicità la serie a lag = k . Al contrario un valore negativo di r_k esprime variazioni in senso opposto fra la serie neutra e quella ritardata, un aumento del prezzo all'istante t implica una diminuzione all'istante $t - k$.

Di seguito vengono riportati i correlogrammi costruiti per valori di ritardo massimo fissati su scala giornaliera, mensile e settimanale, che sono verosimilmente i tre orizzonti temporali a breve termine su quali ci aspettiamo che la periodicità sia marcata.

a) **Giornaliera:** con ritardo di 30 h

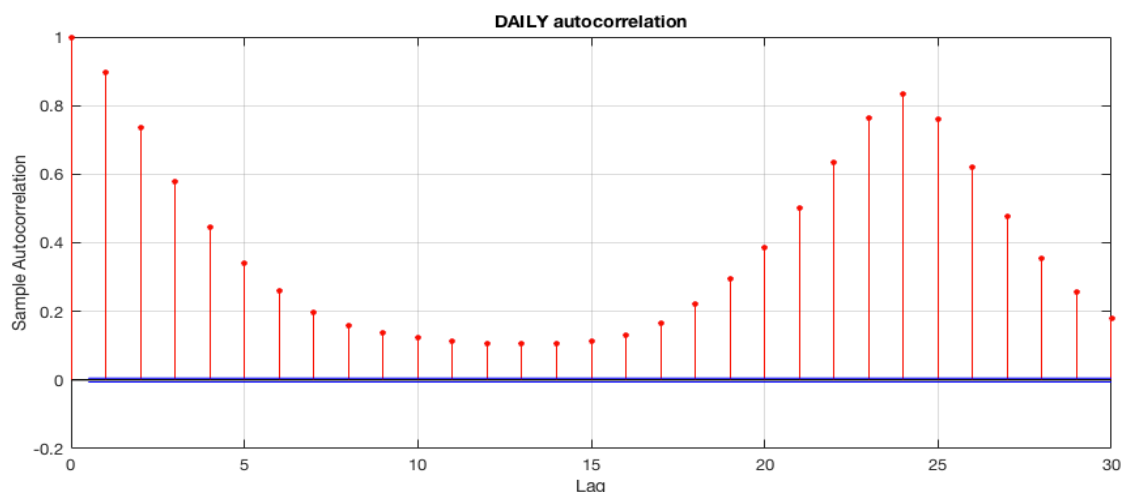


Figura 48 – Coefficiente di autocorrelazione a scala giornaliera ($k = 1,2, \dots, 30$)

Il valore di autocorrelazione si mantiene sempre positivo ed è molto variabile nel corso della giornata, per applicare il metodo di previsione sarà necessario dividere la serie in fitting, valibration e testing set con dei gap fra ognuno in modo da ottenere la minima correlazione fra i set, da qui emerge che dovrò probabilmente scegliere gap ad intervalli di mezza giornata per minimizzare completamente la correlazione fra i set.

b) **Settimanale:** con ritardo di 8 giorni (192 h)

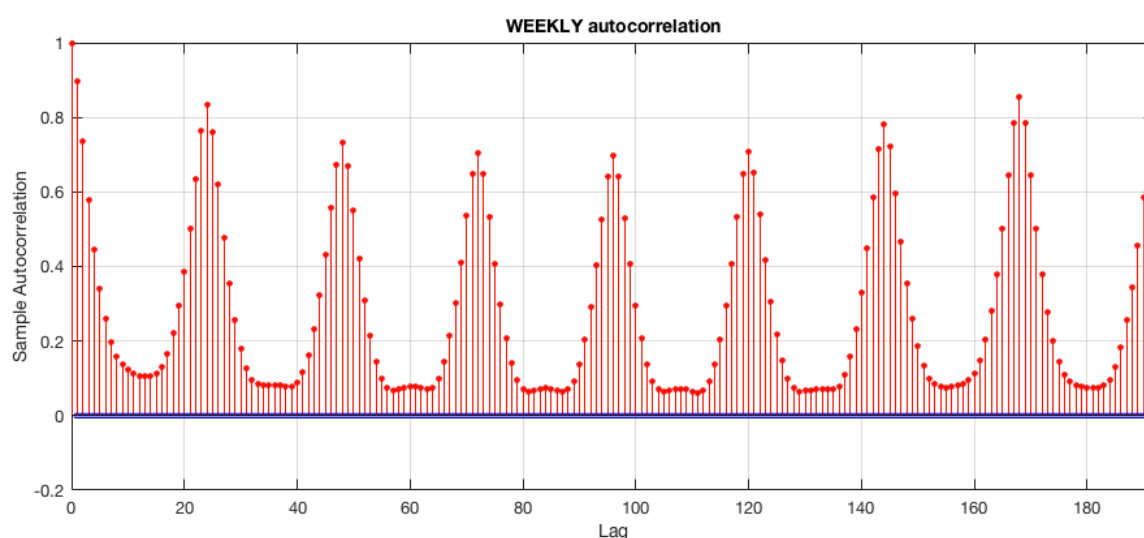


Figura 49 - Coefficiente di settimanale a scala settimanale ($k = 1,2, \dots, 24 * 8$)

Anche su scala settimanale il coefficiente di autocorrelazione è elevato, sempre positivo con un valore a sette giorni che è addirittura maggiore di quello giornaliero, per le giornate centrali della settimana i valori vanno smorzandosi lievemente e la correlazione a 12h non è più minima come a scala giornaliera. Ancora nessun valore negativo, tardano a presentarsi anche su scala mensile ed il primo si verifica a lag = 1600 (pari a 66 giorni e 16 h). Il minimo è raggiunto con ritardo pari a 65778h che è poco più di $T/2$ con valore -0,2027.

c) **Annuale:** con ritardo di 400 giorni (9600 h)

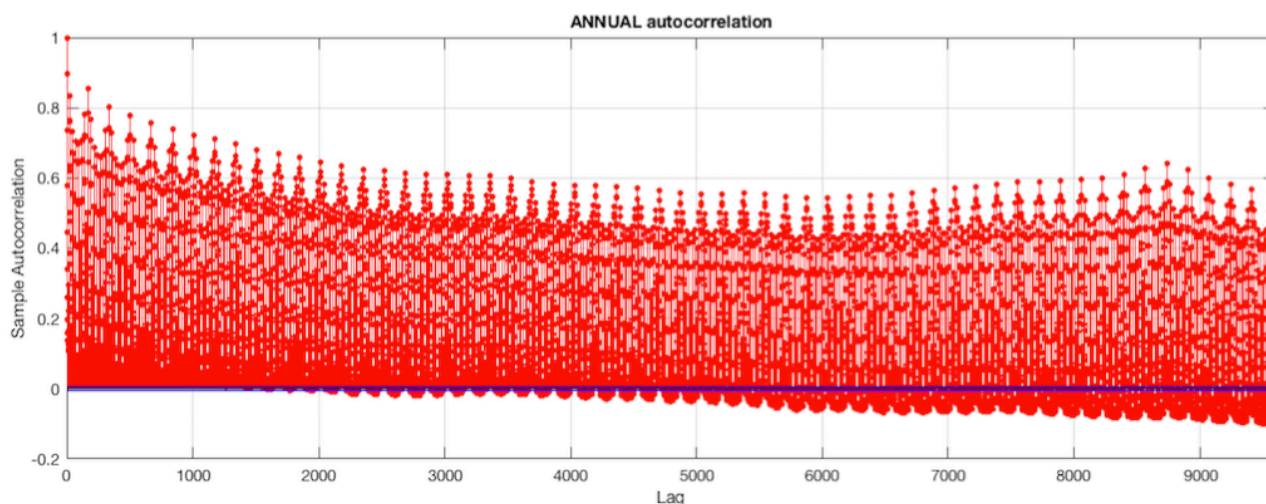


Figura 50 - Coefficiente di autocorrelazione a scala annuale ($k = 1, 2, \dots, 24 \cdot 400$)

Invece su scala annuale si nota comunque una periodicità che è ancora elevata ma non quanto le precedenti, su scala mensile non si apprezza un trend ben delineato e per questo non viene investigato ulteriormente. La serie considerata conta $T = 122712$ valori corrispondenti a 13 completi più gli ultimi tre trimestri del 2004 e il primo del 2018, il massimo valore di ritardo utilizzato è 9600 il che significa che i coefficienti sono ottenuti escludendo un punto soltanto fino a quasi 10000 valori, al massimo si trascura poco più del 7,8% dei dati al fine di calcolare correlazioni a scala annuale.

	MAX	LAG	POSITION
DAILY	0,835	24	1d
WEEKLY	0,855	168	7d
ANNUAL	0,639	8736	364d

Tabella 11 – Autocorrelazione massima, ritardo corrispondente e posizione

Il valore maggiore è quello su scala settimanale invece il minore quello annuale, mentre fra i valori giornalieri e settimanali la differenza è minima il divario con l'autocorrelazione annuale è un po' più ampio e soprattutto l'autocorrelazione annuale presenta dei valori negativi al contrario delle altre.

CAPITOLO 4: CALIBRAZIONE dei PARAMETRI

4.1. Set di calibrazione e orizzonte di previsione

L'obiettivo è ottenere una terna di parametri ottimi da applicare in sede di previsione effettiva sul testing set, quindi occorre considerare diversi periodi dell'anno per capire come questo influisca sui risultati ed è utile scegliere come calibrare al meglio in base al periodo che si vuole predire. I set considerati sono due:

- **Set invernale:** l'intero mese di gennaio 2018 su scala oraria per un totale di 744 punti
- **Set estivo:** 720 valori orari di giugno 2016. Non si considera il 2018 perché è previsto di valutare le previsioni in questo anno completo.

Nel capitolo 3 è spiegato nel dettaglio ogni parametro coinvolto, oltre alla dimensione di embedding m_v , il fattore di ritardo τ ed il numero di vicini k_s si citano anche il grado del polinomio approssimante e l'orizzonte di previsione. Quest'ultimo non è propriamente un parametro da calibrare ma piuttosto da definire in base alle necessità, infatti prevedere ad orizzonte unitario è un buon modo per esplorare il codice ma la applicazioni reali trovano interesse nella previsione a 24h se non di più, ovviamente a discapito delle prestazioni rispetto ad 1h. Invece, il polinomio approssimante non incide parecchio sulle performance, in generale una cubica sembra essere la peggiore ma quello lineare e quadratico sono più o meno alla pari.

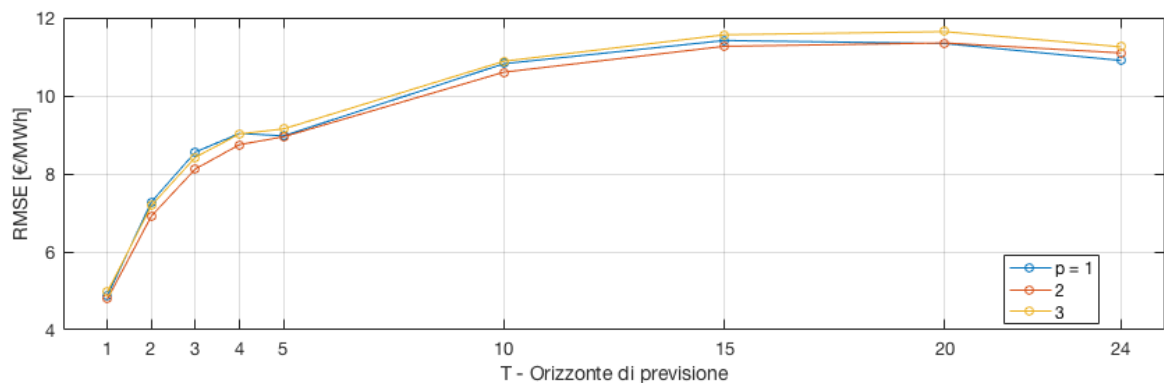


Figura 51 - Curve medie per la scelta di p

La Figura 51 si riferisce ad un set di calibrazione di prova e va analizzata con un occhio critico, è ottenuta mediando i valori di RMSE per tutte le terne con $T=\text{costante}$ e $p=\text{costante}$, indistintamente da m_v , τ e k_s . La calibrazione da cui è costruita utilizza $m_v = [10, 45, 50, 60]$, $\tau = [1, 3, 6, 18]$ e $k_s = [500, 2000, 1000]$ i cui valori di RMSE variano da 4,12 a 15 quindi tengono conto anche delle terne peggiori, non si parla di errori esorbitanti ma è comunque una media, l'informazione che interessa è il trend in aumento mano a mano che aumenta l'orizzonte ($T=24$ migliora però) e l'influenza del grado del polinomio, praticamente sempre è più precisa la quadratica, altrimenti la lineare. Ciò non vuole a

dimostrare che $p=2$ è il valore ottimo, è semplicemente un riscontro che in base alla set di parametri varia leggermente, ad ogni modo da questo punto in avanti si utilizza una quadratica per pesare il contributo di ognuno dei k_s vicini.

Invece, la previsione si effettua su due orizzonti: 1 ora e 24 ore, per ogni set si individuano le terne e con quelle si studia l'influenza delle stagioni e dell'orizzonte di previsione.

4.2. Scelta dei parametri

Il metodo per arrivare alla terna migliore si compone di tre passaggi principali:

A. Calibrazione su tutto lo spazio dei parametri → a priori, non vi è modo di ottenere alcuna indicazione su dove cercare il minimo globale, perciò è necessario esplorare il più possibile un ampio intervallo per ogni parametro, le terne che ne derivano saranno molteplici ed è la fase più dispendiosa in termini di tempo. I parametri variano su intervalli fissi stabiliti:

$$m_v = [5 \ 10 \ 15 \ 20 \ 25 \ 30 \ 35 \ 40 \ 45 \ 50 \ 100 \ 150 \ 200 \ 250 \ 300 \ 350 \ 400 \ 450 \ 500]$$

$$\tau = [1 \ 2 \ 3 \ 4 \ 5 \ 6 \ 7 \ 10 \ 15 \ 20 \ 25 \ 30]$$

$$k_s = [10 \ 20 \ 30 \ 40 \ 50 \ 100 \ 300 \ 500 \ 1000 \ 2000 \ 3000 \ 5000 \ 10000]$$

In base a quanto è ampio il set di calibrazione i tempi di computazione crescono notevolmente per coppie di k_s e m_v elevati mentre il ritardo non incide, i valori di ritardo sono scelti basandosi su alcune calibrazioni preliminari in con $\tau = 1:24$ proprio per capire quali valori escludere a priori.

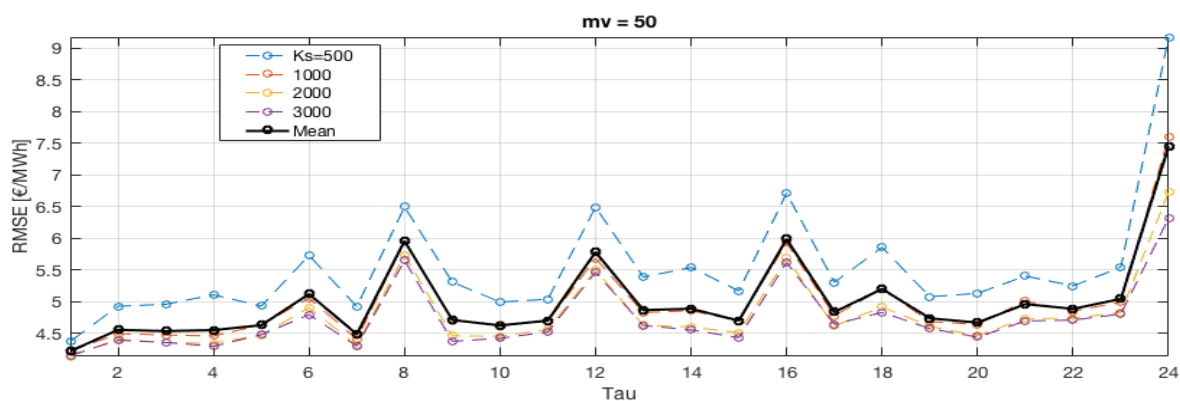


Figura 52 - Sezione dello spazio dei parametri

La figura sopra mostra una sezione dello spazio a numero di embedding fisso e pari a 50 in cui è ben evidente che le terne con $\tau = 8, 12, 16$ e 24 sono da scartare a prescindere dagli altri parametri, non accade solo a $m_v = 50$ ma è un fenomeno che si sostiene, sia sull'orizzonte unitario che a 24 ore.

B. Limite errore e raffinazione degli intervalli → i risultati della prima calibrazione conducono a valori di errore sparsi su 2 o 3 ordini di grandezza, parecchie terne riscontrano valori contenuti ma alcune toccano picchi esorbitanti rispetto al minimo. Queste ultime vanno eliminate per condurre un ragionamento efficace in termini di prestazioni medie, considerare anche le terne disastrose rende l'analisi falsata.

Per cui si impone un limite al valore di errore e si scartano tutte le terne che non lo rispettano, viene calcolata un percentuale di utilizzo dei dati di partenza per capire quanto dati si stanno scartando, se il limite è rispettato allora mi sto muovendo verso il minimo altrimenti occorre infittire la ricerca per muovermi nella direzione esatta, chiaro che il limite va scelto in base al valore minimo assoluto e osservando la distribuzione delle occorrenze dei valori limitrofi.

Una volta imposta la restrizione non è più possibile ragionare con le curve medie perché non tutte le terne sono complete per i valori stabiliti in precedenza (pagina 108), perciò occorre ricondurre la ricerca ad intervalli completi con bassi RMSE che individuano il minimo globale. Si riesce a farlo contando il numero di occorrenze di un dato valore di ogni parametro, per esempio nell'intervallo di $m_v = [5 \ 10 \ 20 \dots]$ si conta quante volte appare ogni valore nel sottoinsieme delle terne che rispettano il vincolo, lo stesso per τ e k_s . Il risultato sono tre vettori del tipo $cont_m = [42 \ 105 \ 78 \dots]$ da cui risalgo al valore più frequente per ognuno dei tre parametri (nell'esempio sarebbe $m_v = 10$ con 105 occorrenze), così facendo si costruisce la terna dei parametri più frequenti con il relativo errore

Seguendo questo schema è possibile studiare come varia il risultato a variare del limite imposto, mano a mano che il limite aumenta considero più terne e i risultati convergono sulle terne più frequenti, quanto più velocemente cresce l'errore spostandosi su altre terne tanto più quella terna costituisce un minimo locale, in altre parole il minimo globale è da cercare nell'intorno delle combinazioni che si sostengono più a lungo al variare del limite fissato. Impostando correttamente il ragionamento è possibile escludere terne o aggiungerne a seconda del caso, prima di tutto limitare l'errore serve per capire quali terne escludere a priori ottenendo un nuovo sottoinsieme, su questo si ragiona per capire se le informazioni sono sufficienti o altrimenti si procede infittendo la calibrazione attorno al minimo.

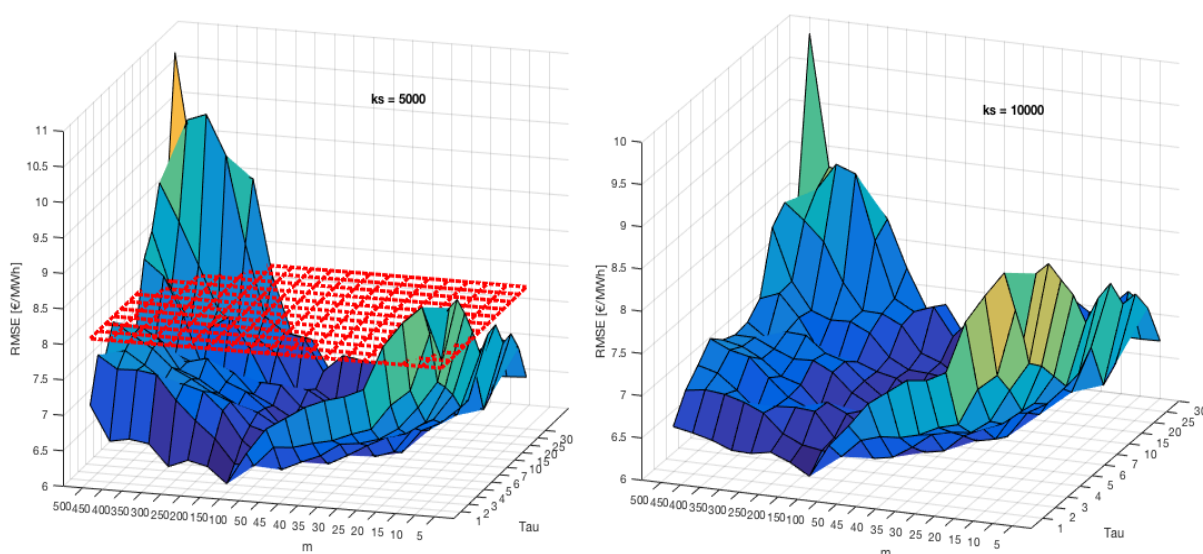


Figura 53 - Rappresentazione grafica del processo di imposizione del limite, $T = 24$, calibrazione globale

La Figura 53 fornisce una sezione tridimensionale a $k_s = cost$ per ottenere un riscontro grafico ed efficace del procedimento descritto prima e ripreso dai grafici in appendice a pagina 132, il piano rosso nella figura di destra rappresenta il limite imposto e la parte del conoide che sta sotto a quel piano è il sottoinsieme delle terne su cui si cercano i parametri più frequenti. Il limite viene applicato ad ogni valore di k_s e gli errori maggiori cadono sotto il 2000 infatti, da destra, le figure sono a $k_s = 5000$ e 10000.

C. Sezioni dello spazio dei parametri → a questo punto la calibrazione conduce verso terne con errori contenuti, quindi inizia a delinearsi la posizione del minimo ma gli intervalli sono ancora troppo ampi per concludere. Occorre una visione d'insieme del problema che include l'errore in funzione di tre parametri, viene da pensare ad un plot 3D con color map, ma non è molto utile su cartaceo ed è più efficace plottarne le sezioni mantenendo fisso un parametro e variando i due che rimangono, come in Figura 52. Se un parametro è fissato si può rappresentare il secondo in ascisse ed il terzo come più curve sullo stesso plot o viceversa, in tutto due sezioni (equivalenti) per ogni parametro quindi sei in tutto.

- $m_v = cost$ → errore in funzione di τ o k_s → m_v figure con k_s o τ curve
- $\tau = cost$ → errore in funzione di k_s o m_v → τ figure con m_v o k_s curve
- $k_s = cost$ → errore in funzione di m_v o τ → k_s figure con τ o k_s curve

L'imposizione del limite serve proprio a eliminare le curve fuori scala facendo sì che una curva media diventi una buona rappresentazione del comportamento generale nel minimo (la curva nera in Figura 52). Inserendo su un solo plot tutte le curve medie così costruite si elimina un grado di libertà (quello mantenuto fisso, m_v nel caso della Figura 52), se tutte le curve di ogni sezione sono ben distribuite su un unico fascio allora ha senso raggrupparle tutte altrimenti è più corretto analizzare le sezioni una alla volta. Si ha la panoramica completa fornendo diagrammi di due parametri mantenuti costanti in quanto il terzo è incluso in entrambi, di solito quelli a $k_s = cost$ sono piuttosto confusi ed è più facile trarre delle conclusioni dagli altri due. A seconda delle circostanze si includono sezioni più elaborate, ad esempio per orizzonte a 24h si studia anche l'errore in funzione di $\tau * m$.

Riguardo le metriche di errore si utilizza il RMSE ed in seconda sede bisogna irrobustire l'analisi basandosi su una seconda metrica, il RMSE valuta solamente l'entità degli scarti senza tenere conto che il valore da prevedere sia "standard" o di picco per esempio, si usa il MAPE come metrica d'appoggio per verificare che tutto sia confermato anche sotto questo aspetto, informazioni più dettagliate sono disponibili al paragrafo 2 del terzo capitolo (a pagina 81).

Dopo si studia l'impatto della previsione a 1 e 24 ore, il processo viene ripetuto per ogni set e poi si confrontano l'invernale e l'estivo sullo stesso orizzonte in modo da giungere ad una sola terna ottima a prescindere dal periodo di calibrazione. Il passo finale produce la terna scelta oppure una nuova calibrazione che può essere un sottoinsieme della prima o una nuova attorno al minimo, in base alle circostanze si valuta se è sufficiente escludere dei valori oppure se ne servono altri per concludere.

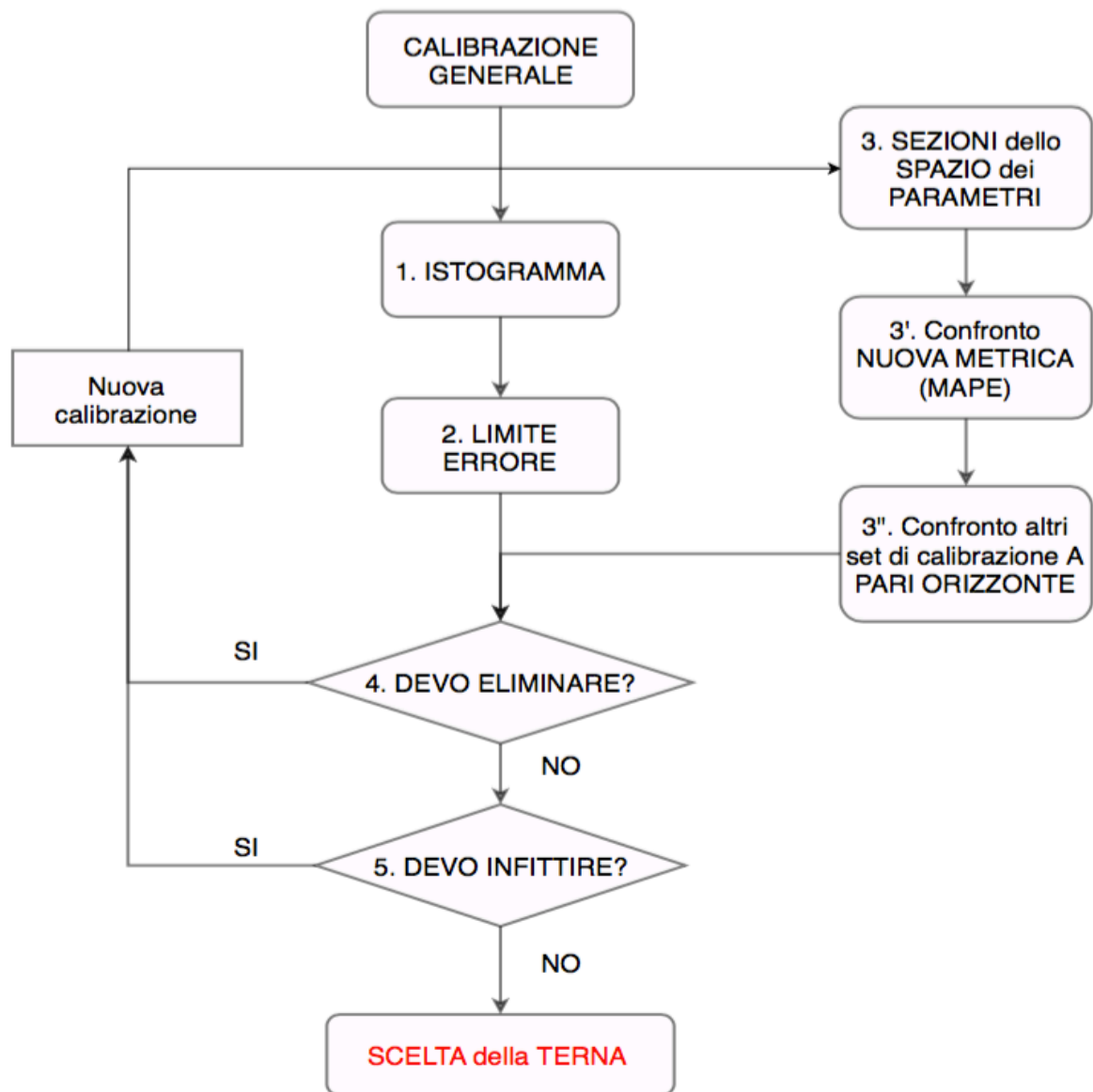


Figura 54 – Diagramma di flusso per la scelta dei parametri

L'ultimo diagramma di flusso schematizza il processo decisionale, riepilogando, la calibrazione generale va subito depurata imponendo un limite di RMSE che porta ad eliminare svariate terne, poi si studiano le sezioni dello spazio dei parametri per la nuova calibrazione considerando una metrica aggiuntiva che valuta l'entità dei residui rispetto il valore reale, infine, dopo aver valutato l'incidenza del periodo su cui si calibra, si decide il da farsi fra tre opzioni:

- Se ci sono ancora terne superflue si escludo per ripartire da capo.
- Se il minimo è ancora confuso si procede calibrando su intervalli specifici.
- Altrimenti la terna di m, τ, k_s è determinata.

4.2.1 Set estivo

Orizzonte di previsione a 1 ora

Si tratta di giugno 2016, trenta giorni e 720 valori orari su cui si effettua la calibrazione indicata a pagina 88, si procede secondo lo schema e si individuano gli estremi:

Minimo assoluto $\rightarrow \text{RMSE} = 2,889 \text{ €/MWh}$ con $m = 350$, $\tau = 1$ e $k_s = 10000$

Massimo assoluto $\rightarrow \text{RMSE} = 6404 \text{ €/MWh}$ con $m = 250$, $\tau = 7$ e $k_s = 500$

L'errore si estende su tre ordini di grandezza e per capire come eliminare le terne peggiori si utilizza un istogramma, la gran parte di queste stanno entro i 25 €/MWh di RMSE quindi si prende quel valore come limite massimo. Attenzione però, questo non equivale a cercare terne ottime con RMSE tollerato fino a 30 30 €/MWh ma cercarle basandosi su un sottoinsieme che rispetta questo vincolo, tal valore massimo è certamente troppo alto per la previsione ottima.

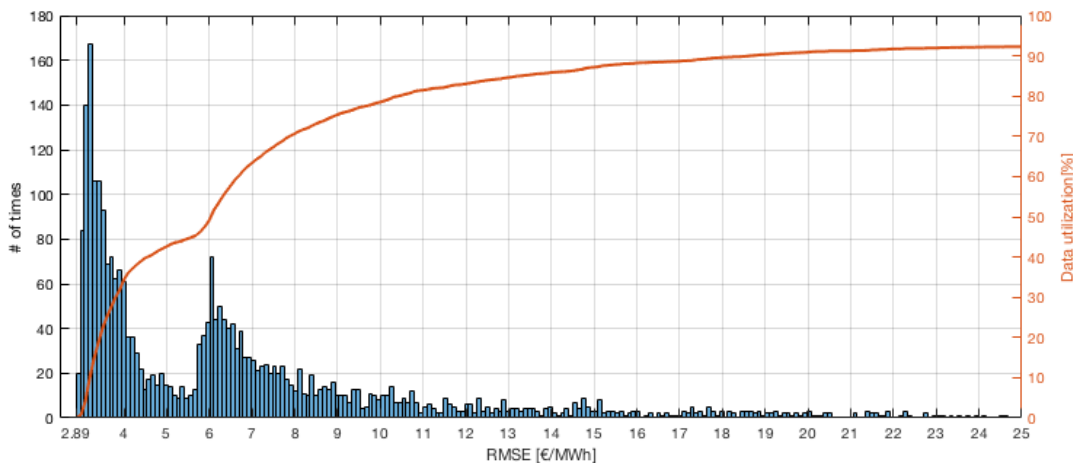


Figura 55 – Istogramma : prima calibrazione, set estivo , $T=1$

È stato anche verificato che i valori esorbitanti fossero casi isolati, e lo sono, quindi si procede con l'esclusione delle terne che non rispettano il limite, poi su questo sottoinsieme si contano le occorrenze per ciascun valore di un parametro alla volta e così si ottiene il valore di m più frequente per quel preciso limite e lo stesso per τ e k_s . Aumentando il limite dal minimo assoluto al massimo scelto (fissato a 25 €/MWh) le terne dei parametri più frequenti cambiano perché includono mano a mano più valori, questo aiuta a restringere l'analisi alle terne migliori.

Ora, gli estremi di riportati sulle ordinate di Figura 80 (Appendice B pagina 136) indicano come restringe la ricerca: $5 \leq m \leq 200$, τ presenta un caso isolato a 10 ma non viene considerato quindi $\tau = [1 \ 2 \ 3]$ e $k_s \geq 2000$.

La calibrazione generale include 2964 terne e questa scrematura ne conserva 156, esclude ritardi notevoli per via dell'orizzonte unitario e necessita di molti vicini, tutte le terne con k_s contenuto sono superflue, gli estremi diventano:

Minimo assoluto $\rightarrow$ **RMSE** = 2,89 €/MWh con $m = 200$, $\tau = 1$ e $k_s = 10000$

Massimo assoluto $\rightarrow$ **RMSE** = 4,961 con €/MWh con $m = 5$, $\tau = 1$ e $k_s = 10000$

Ora, l'errore è limitato con successo e il minimo inizia a manifestarsi, a questo punto si osservano le sezioni dello spazio dei parametri:

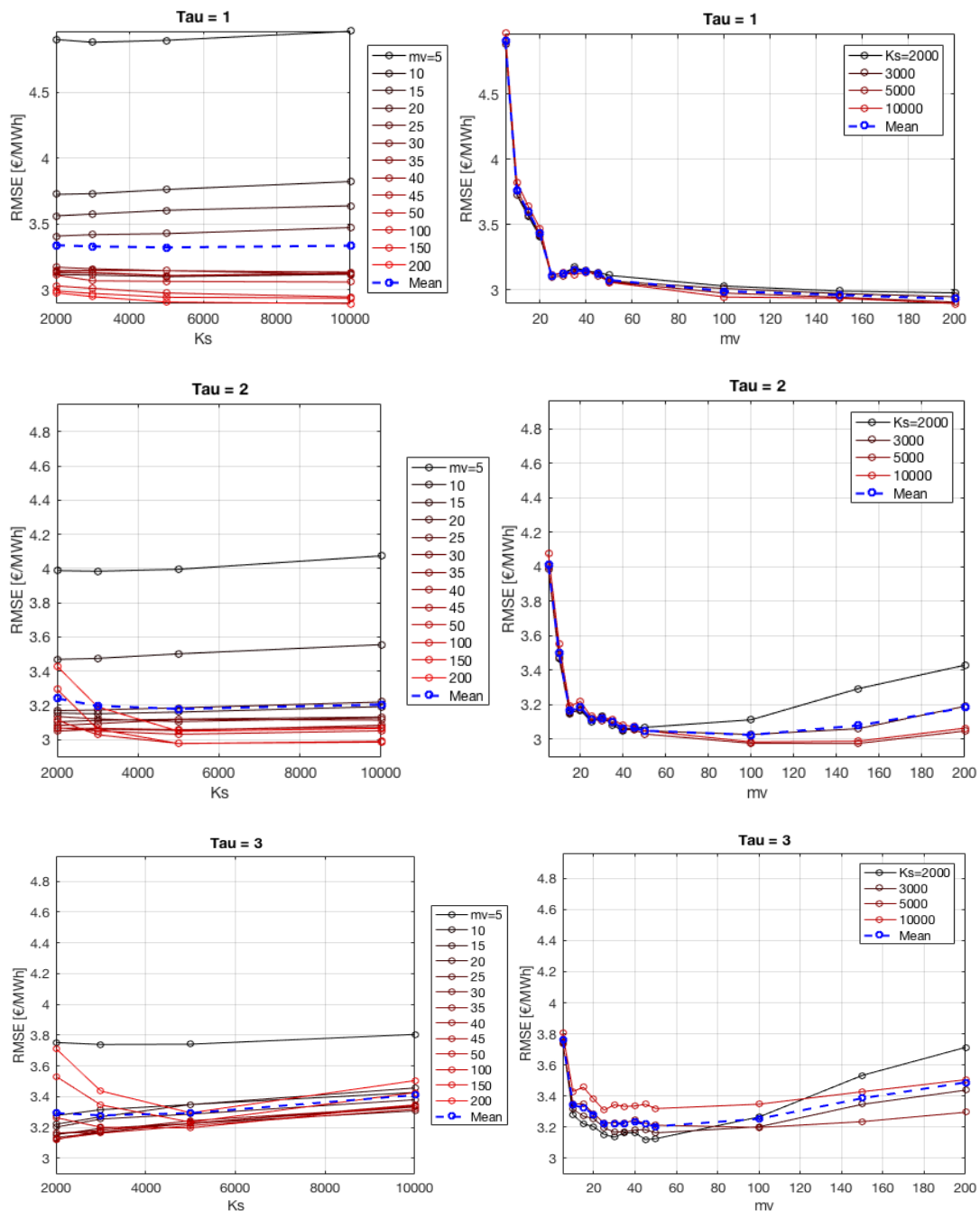


Figura 56 - Sezioni dello spazio dei parametri : $\tau = 1, 2, 3$

L'orizzonte di previsione ad un ora predilige il ritardo unitario poiché è quello da cui trae più informazioni utili, mano a mano che aumenta il ritardo la previsione migliora se il vettore include meno elementi ma è comunque povera di prestazioni se confrontata con le curve a ritardo unitario. Il minimo va cercato a $\tau = 1$ e per alti valori di embedding che vanno infittiti attorno al 150 - 200 e oltre.

Giunti sin qui manca solo verificare i risultati con il MAPE perché lo scarto quadratico medio prende in considerazione solamente l'entità degli scarti senza alcun riferimento al valore reale, se il MAPE conferma, allora la calibrazione sarà robusta abbastanza. l'errore percentuale varia nella finestra dal 6 scarso all'11%.

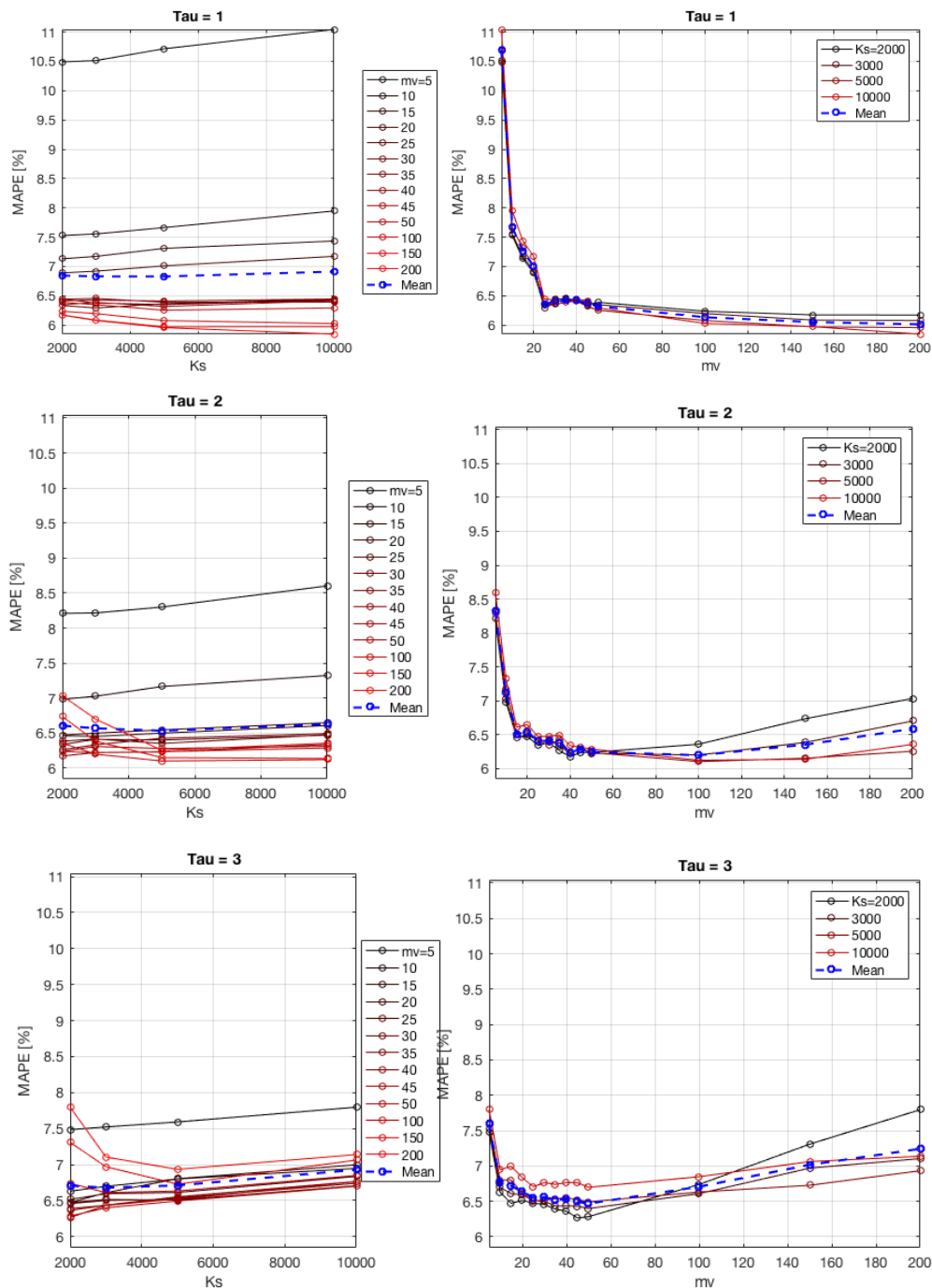


Figura 57 - Sezioni dello spazio dei parametri : $\tau = 1, 2, 3$, MAPE

A prescindere dalla metrica continuano a valere le stesse osservazioni: ritardo unitario con m e k_s da spingere ancora su valori più alti, per il resto c'è un ottimo riscontro con il passaggio alla nuova metrica. Questo set di calibrazione ha bisogno di essere espanso e confrontato con la calibrazione invernale a pari orizzonte, dopo si deve ragionare per giungere ad una conclusione univoca, una sola terna che vada bene a prescindere dal periodo su cui si valuterà il codice.

Riguardo al fattore di ritardo ci sono pochi dubbi, unitario e l'ottimo, per gli altri parametri serve concentrare le ricerche:

$$m = 140 : 10 : 230$$

$$\tau = 1$$

$$k_s = 5000 : 500 : 12000$$

L'idea iniziale prevedeva di cercare $k_s > 10000$ ma dopo qualche analisi preliminare il minimo era già stato raggiunto e così ci si è concentrati anche fra 5000 e 10000, per questa finestra di dimensione di embedding il minimo non è raggiunto nemmeno a $k_s = 7000$. Relativamente a m i valori 140 al 160 sono poco utili ma sono considerati per confrontare un trend curioso evidenziato anche nel set invernale: $m = 160$ corrisponde a poco meno di sette giorni (con $\tau = 1$) quindi il periodo a massima correlazione (vedi SERIE STORICA e AUTOCORRELAZIONE), appunto la Figura 68 riporta un calo delle prestazioni per $m = 160$ indistintamente dal numero di vicini.

Lo stesso vale per il set estivo, le prestazioni risentono della forte correlazione dei dati a scala settimanale senza distinguere il periodo dell'anno, il RMSE lo evidenzia particolarmente il che vuol implica residui maggiori, con il MAPE è simile ma non identico, la zona $m = 150 - 170$ ha sempre una depressione nel mezzo ma in questo caso il minimo è traslato poco in avanti sia per dimensione di embedding che per numero di vicini.

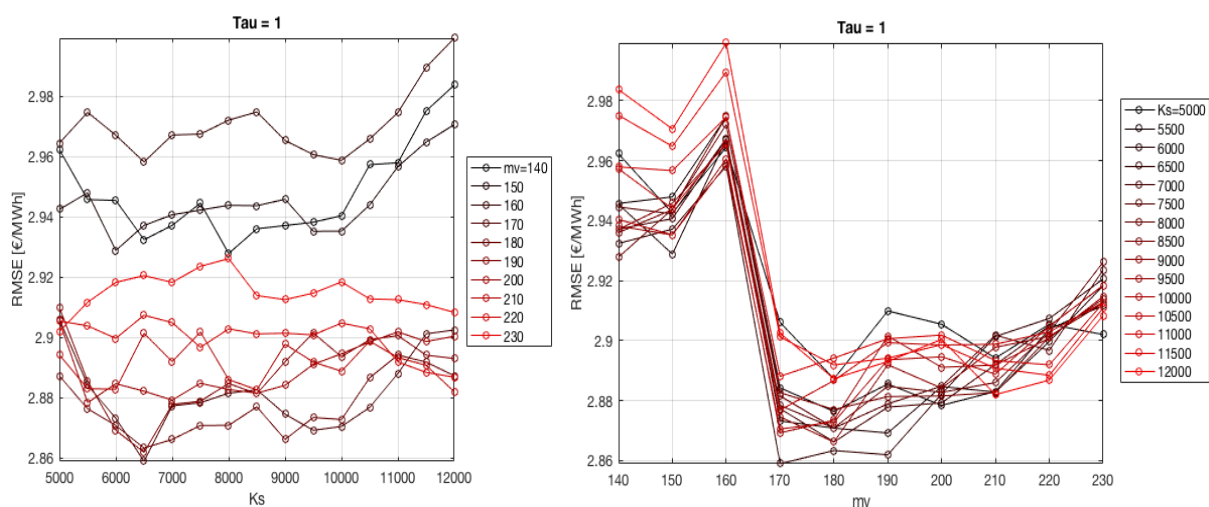


Figura 58 - Sezioni dello spazio dei parametri, $\tau = 1$, RMSE [€/MWh]

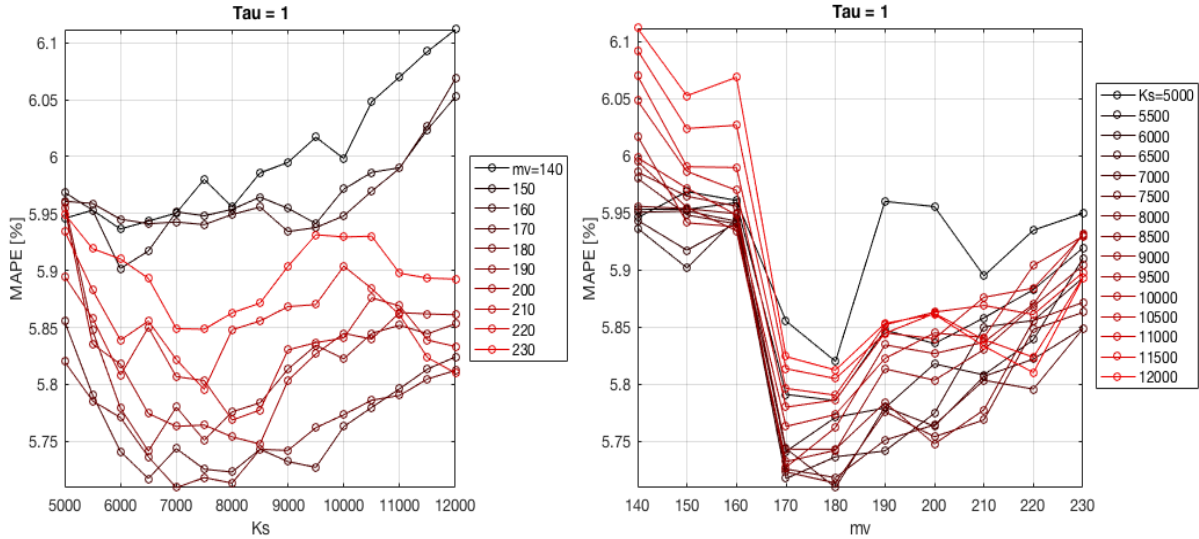


Figura 59 – Sezioni dello spazio dei parametri, $\tau = 1$, sopra : MAPE [%]

Concludendo la calibrazione, per il set estivo ad orizzonte unitario l'ottimo si trova per dimensione m fasi piuttosto alta che sta poco al di sotto del 200, il fattore di ritardo $\tau = 1$ si impone da subito mentre il numero di vicini è un parametro fondamentale che varia parecchio, le prime osservazioni lasciano spazio a valori sopra al 10000 ma affinando non è così. Le ultime due figure mostrano i RMSE e MAPE specifici vicino al minimo e le due metriche portano conclusioni diverse, però si parla di variazioni anche millesimali che stanno sotto la precisione della moneta e che per approssimazione portano a differenze al centesimo di euro sul RMSE e lo stesso per il MAPE anche se in percentuale.

Volendo scegliere, le due terne che eccellono per precisione sono:

$$m = 170 \quad \tau = 1 \quad k_S = 6500 \rightarrow RMSE = 2,86 \frac{\text{€}}{MWh} \quad MAPE = 5,72 \% \rightarrow \text{tempo}^{20} = 307 \text{ s}$$

$$m = 180 \quad \tau = 1 \quad k_S = 7000 \rightarrow RMSE = 2,87 \frac{\text{€}}{MWh} \quad MAPE = 5,70 \% \rightarrow \text{tempo}^{24} = 362 \text{ s}$$

Differenze che sono veramente minime sia come errore che come parametri adottati, un confronto preliminare a pari orizzonte può aiutare, non si è ancora discusso della calibrazione invernale ma si anticipa che utilizza 170 elementi per ogni vettore e accade lo stesso se i due set vengono considerati insieme nella calibrazione finale, per questo motivo il processo è concluso:

$$m = 170 \quad \tau = 1 \quad k_S = 6500 \rightarrow RMSE = 2,86 \frac{\text{€}}{MWh} \quad MAPE = 5,72 \% \rightarrow \text{tempo}^{24} = 307 \text{ s}$$

²⁰ Run singola MATLAB R2016b su MacBook Pro mid 2012, intel dual core 2,5 (3,1) GHz, ram 4Gb DDR3 1600 MHz.

Orizzonte di previsione a 24 ore

Stesso procedimento, calibrazione generale (pg.88) con estremi:

Minimo assoluto $\rightarrow$ **RMSE** = 5,812 €/MWh con $m = 45$, $\tau = 15$ e $k_s = 10000$

Massimo assoluto $\rightarrow$ **RMSE** = 2553 con €/MWh con $m = 500$, $\tau = 20$ e $k_s = 1000$

La maggior parte delle terne stanno entro i 25 €/MWh di RMSE e quindi si imposta quel valore come limite massimo.

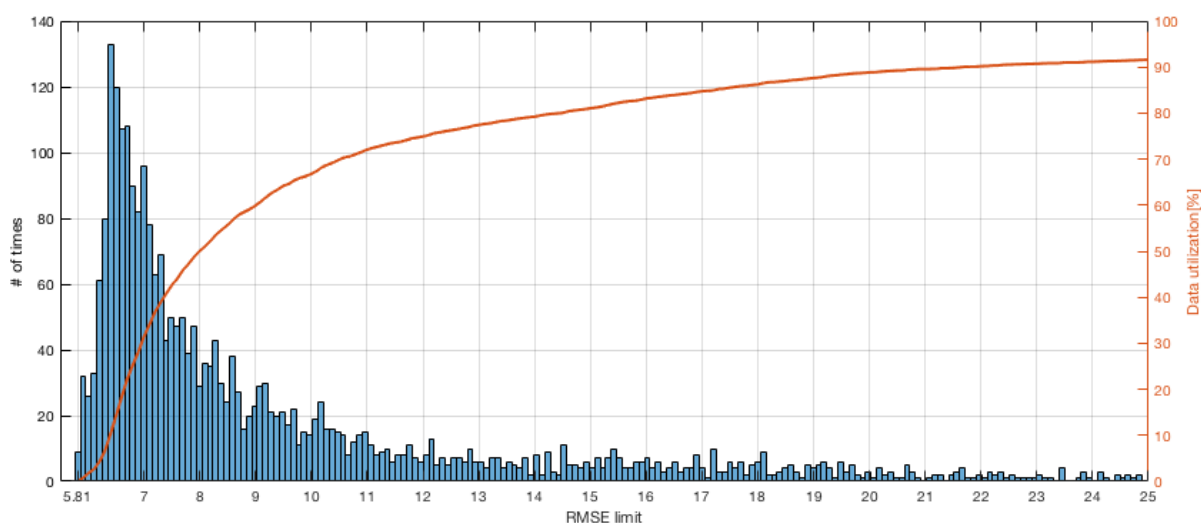


Figura 60 – Istogramma : prima calibrazione, set estivo, $T=24$

I valori esorbitanti sono casi isolati, si procede con l'esclusione delle terne. A priori, le aspettative sull'orizzonte a 24 ore includono previsioni meno accurate ma parametri diversi, probabilmente ritardi maggiori ma per m e k_s è difficile sbilanciarsi, a orizzonte unitario servono molti vicini allora è probabile che sarà così in ogni caso, k_s dipende dal determinismo della serie se questa è prevalentemente casuale allora l'orizzonte di previsione è poco influente.

La situazione non è immediata, tutti e tre i parametri variano su intervalli ridotti ma nessuno è stabilito univocamente, si seguono le indicazioni raccolte in Figura 86 (Appendice B pagina 136) si restringe alle sole terne con m compreso tra 30 e 250, τ fra 2 e 30 e $k_s \geq 2000$. Il sottoinsieme conta 396 terne sulle 2964 di partenza con estremi molto più compatti:

Minimo assoluto $\rightarrow$ **RMSE** = 5,812 €/MWh con $m = 45$, $\tau = 15$ e $k = 10000$

Massimo assoluto $\rightarrow$ **RMSE** = 10,3 con €/MWh con $m = 250$, $\tau = 30$ e $k = 2000$

Proseguendo, possiamo ragionare con le sezioni, i diagrammi sono molti e se ne riportano alcuni a seguito, mostrano differenze riguardo la dimensione di embedding: per m dal 150 in poi il trend ha minimo per ritardo attorno 5, mentre per m attorno al 45 trasla a ritardi maggiori.

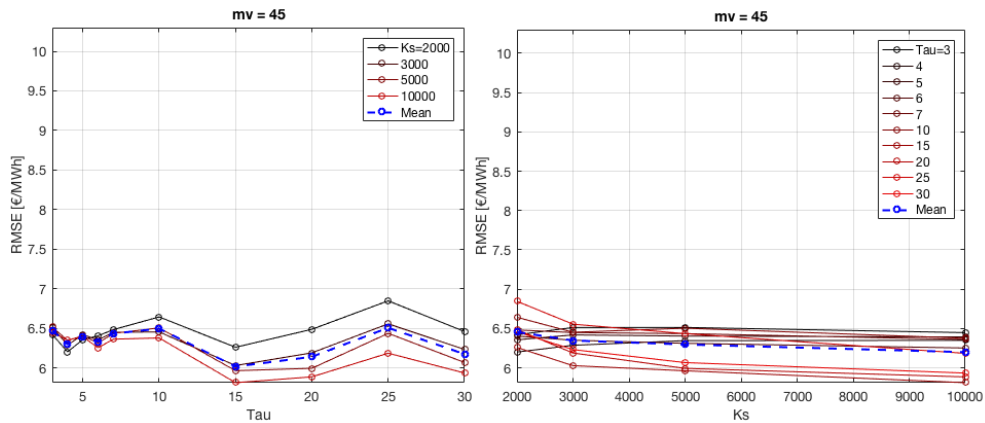


Figura 61 - Sezioni dello spazio dei parametri : $m = 45$

Il minimo più accentuato presenta fasci di curve molto omogenei e la curva media rappresenta bene il fenomeno, esso è verificato per $\tau=15$ ed m circa 45-50. C'è possibilità di miglioramento aumentando il numero di vicini.

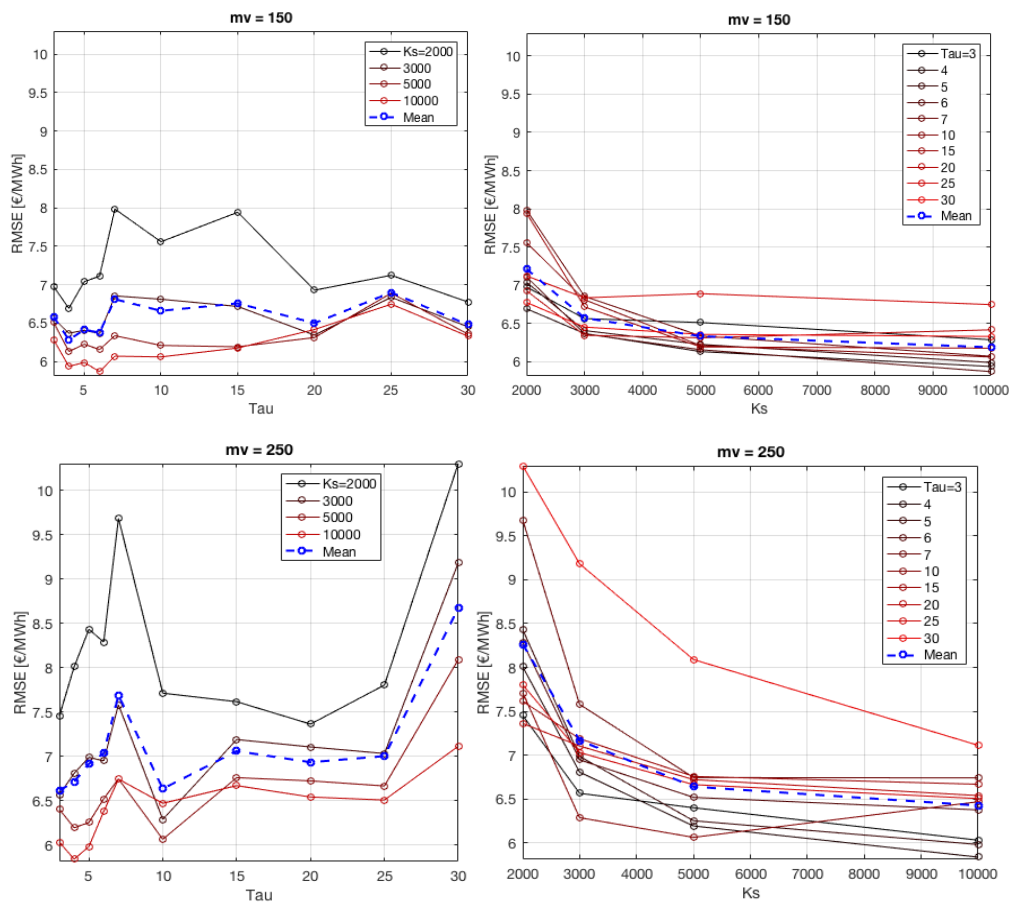


Figura 62 - Sezioni dello spazio dei parametri : $m = 150, 250$

Le ultime due figure interessano valori che portano errori molto vicini al minimo assoluto da $m = 150$ e $\tau = 6$ a $m = 250$ e $\tau = 4$, Il minimo persiste su valori di RMSE simili, a bassi m con ritardo maggiore ($m = 45$ e $\tau = 15$) ed a alti m con ritardi contenuti, a questo punto pare che vi sia un relazione di proporzionalità inversa tra dimensione di embedding e ritardo, perciò si studia l'errore in funzione del prodotto $\tau * m$:

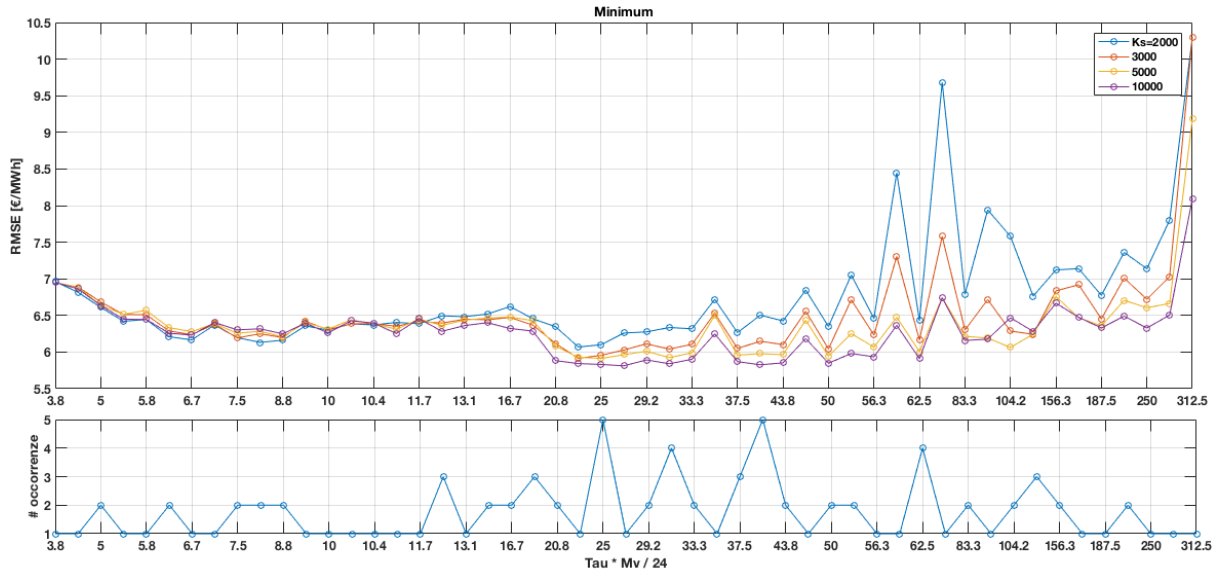


Figura 63 – Sezioni dello spazio dei parametri : $\tau * m$

l'ultimo grafico si costruisce partendo da un valore di k_s alla volta e si plotta la curva dei valori minimi rispetto al prodotto $\tau * m$, con gli intervalli scelti ($m = 30 - 250$, $\tau = 2 - 30$ e $k_s = 2000 - 10000$) ci sono più terne a pari $\tau * m$ (il plot in basso indica quante).

Tale prodotto è un tempo che indica fino a che punto mi muovo a ritroso nella serie per costruire i vettori e per questo viene riportato al numero di giorni dividendo per 24, ad esempio le coppie viste prima sono comprese tra 20,8 e 41,7 e si vede bene che l'errore si stabilizza all'aumentare dei vicini.

Prima di proseguire cerchiamo conferma soffermandoci sul MAPE, che tiene conto del rapporto rispetto il valore reale. Da subito quando si impone il limite di errore i risultati vanno a scremare meglio, $k_s \geq 5000$ (trascura il 2000 a priori), $\tau \geq 5$ e ad m aggiunge il 25 ma esclude il 45.

Ad ogni modo i risultati sono confermati, le osservazioni fatte per le sezioni ad $m = 250$, 150 e 45 continuano a valere, anche il MAPE presenta la dipendenza dal prodotto di $\tau * m$ senza prediligere una coppia precisa e come sempre necessita di moltissimi vicini. L'ultima analisi identifica una zona di minimo che si estende tra 20,8 e 50 giorni ($\tau * m$ tra 500 e 1200) tolte due anomalie, per capire come fissare i due parametri ci si concentra dove le prestazioni sono migliori con un termine di paragone: la curva $k_s = 10000$ eccelle e le combinazioni più numerose sono 25 e 41,7 giorni ($\tau * m = 600$ e 1000).

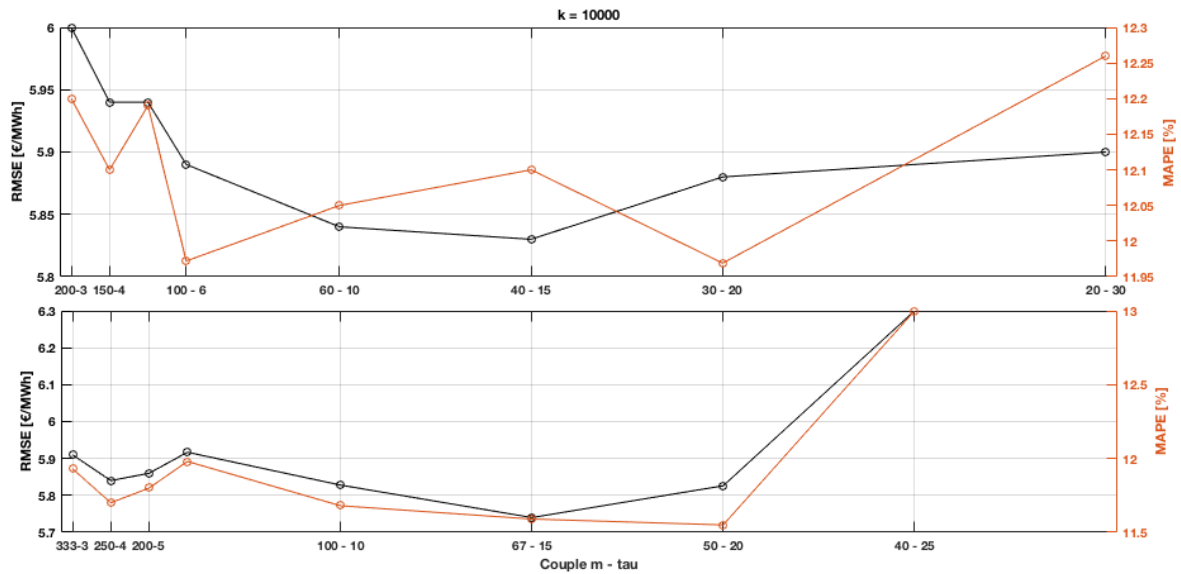


Figura 64 – Dettaglio dei minimi $k_s = 10000$, sopra: $\tau * m = 600 / 25$ giorni, sotto: $\tau * m = 1000 / 41,7$ giorni

La Figura 64 cerca di fornire una visione di insieme del problema con i due valori di $\tau * m$ che appaiono più spesso nella calibrazione, sono state aggiunte alcune coppie specifiche per completare il confronto ($m = 60, \tau = 10$; $m = 120, \tau = 5$; $m = 333, \tau = 3$; $m = 167, \tau = 6$; $m = 67, \tau = 15$).

I valori migliori si ottengono a $\tau * m = 1000$ e per con ritardo circa 15 – 20 per cui si cerca un'ultima volta in quella zona, il grafico che considera $\tau * m = 600$ (Figura 64 in alto) trova conferma solo con il RMSE e inoltre gli errori sono più alti quindi ci aspettiamo un ottimo a $\tau * m$ tendente a 1000 piuttosto che 600.

Si decide di raffinare come segue:

$$m = 40 : 2 : 60$$

$$\tau = 15 \text{ } 20$$

$$k_s = 9000 : 1000 : 15000$$

Risulta che per ritardo pari a 20 si raggiunge il minimo per $m = 50$ e vicini ancora insufficienti però la previsione a ritardo 15 da sempre risultati migliori con l'ottimo oltre $m = 60$ e lo stesso per k_s . Le sezioni $\tau * m$ coprono l'intervallo da 25 a 50 giorni (600 - 1200) ed eccellono a 37,5 giorni (900) quindi a poco più di un mese e una settimana con metriche perfettamente concordi. Ormai il margine di miglioramento è veramente irrisorio quindi si cerca velocemente un'ultima volta:

$$m = 55 \text{ } 60 \text{ } 62 \text{ } 65 \text{ } 70 \text{ } 75 \text{ } 80$$

$$\tau = 15$$

$$k_s = 13000 : 1000 : 16000$$

Si potrebbe indagare il numero di vicini a step di cinquecento ma la precisione è arrivata ad un punto in cui non si migliora sostanzialmente, al massimo si qualche centesimo, per quanto si voglia essere

meticolosi non ha più senso proseguire ma è bene sottolineare che si potrebbe fare. Alla fine, calibrando la previsione giornaliera su luglio 2017 si ottengono ottimi risultati con :

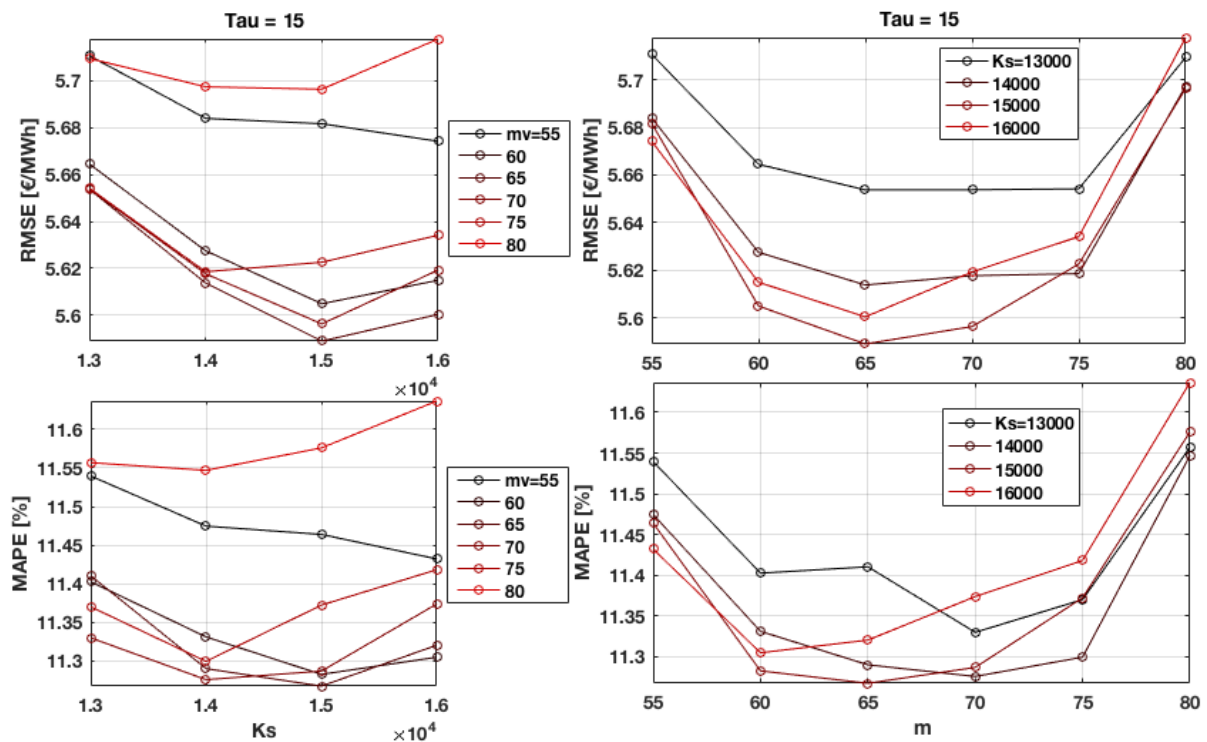


Figura 65 – Sezione dello spazio dei parametri , $\tau = 1$

La soluzione è stata identificata, si presenta subito sostanzialmente diversa da quella ad orizzonte temporale benché si riferisca sempre allo stesso periodo di calibrazione, no resta che vedere cosa succede durante l'inverno, per il momento la terna è:

$$m = 65 \quad \tau = 15 \quad k_s = 15000 \rightarrow RMSE = 5,59 \frac{\text{€}}{MWh} \quad MAPE = 11,27 \% \rightarrow \text{tempo}^{21} = 144 \text{ s}$$

²¹ Run singola MATLAB R2016b su MacBook Pro mid 2012 , intel dual core 2,5 (3,1) GHz, ram 4Gb DDR3 1600 MHz.

4.2.2 Set Invernale

Orizzonte di previsione ad 1 ora

Si tratta di gennaio 2018, trentuno giorni e 744 valori orari su cui si effettua la calibrazione, si procede secondo lo schema individuando gli estremi:

Minimo assoluto $\rightarrow$ **RMSE** = 2,9686 €/MWh con $m = 150$, $\tau = 1$ e $k_s = 10000$

Massimo assoluto $\rightarrow$ **RMSE** = 2974 con €/MWh con $m = 300$, $\tau = 30$ e $k_s = 1000$

Di nuovo, l'errore è esteso su tre ordini di grandezza si utilizza un istogramma in modo da tralasciare le terne peggiori, la maggior parte stanno entro i 25 €/MWh di RMSE che si imposta come limite massimo.

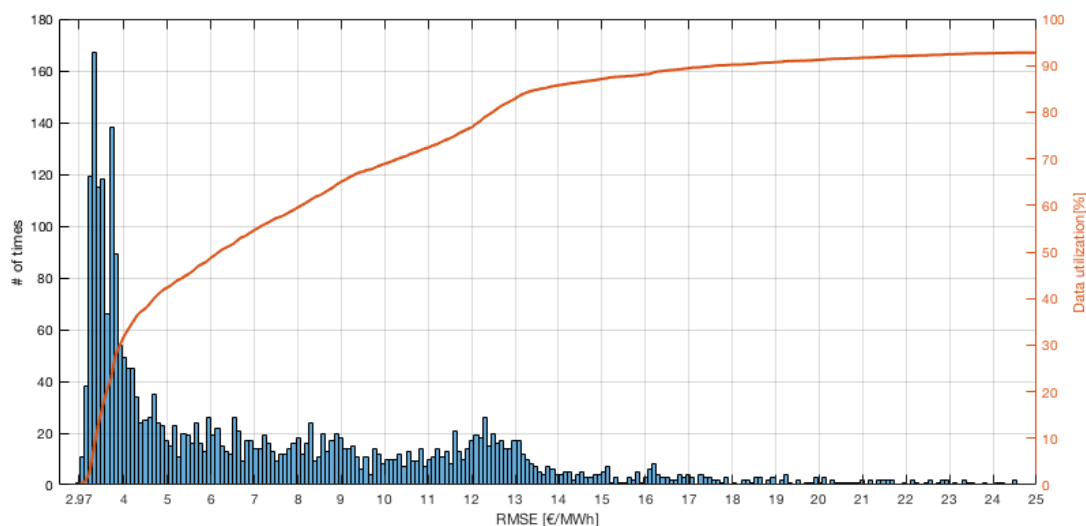


Figura 66 – Istogramma prima calibrazione, set invernale, $T=1$

È stato verificato che i valori esorbitanti fossero casi isolati e poi si procede all'esclusione delle terne peggiori. Dopodiché al solito: si contano le occorrenze ottenendo i parametri più frequenti e si studiano gli intervalli dei parametri più frequenti, procedendo ad eliminare tutte le altre.

Nel migliore dei casi, e anche il più fortunato, sarebbe perfetto riscontrare passo per passo la calibrazione del set estivo, cerchiamo una sola terna che funzioni per $T = 1$ indistintamente dal periodo dell'anno. Certamente sarà difficile però almeno in generale i valori attesi includono bassi ritardi ($\tau = 1$ più probabile) e molti vicini come accadeva con l'estivo sia per orizzonte unitario che 24 ore, invece la dimensione di embedding si presume attorno al 200 per $T = 1$.

Preso il **MAPE** come metrica si giunge alle stesse conclusioni, il minimo è 4,57% per $m = 200$ ed il resto è uguale, m rimane sotto il 200 e τ fisso ad 1 (con qualche picco isolato a 2) e k_s minimo sempre 2000.

Sulla base di Figura 85 (Appendice B a pagina 136) si è in una situazione abbastanza facile perché $\tau = 1$ appare sempre (anche estendendo il limite a 50 e più) in più il minimo globale è da cercare per $m \leq 200$ e $k_s \geq 2000$ perciò una volta eliminate le terne superflue si riparte con il ragionamento. Delle 2964 terne ne rimangono solo 52, tutte sotto i 4 €/MWh di RMSE:

Minimo assoluto $\rightarrow$ RMSE = 2,969 €/MWh con $m = 150$, $\tau = 1$ e $k_s = 10000$

Massimo assoluto $\rightarrow$ RMSE = 3,92 con €/MWh con $m = 5$, $\tau = 1$ e $k_s = 10000$

Con τ già fissato le sezioni sono semplici da analizzare, in un solo grafico vediamo la più importante di tutte includendo anche il numero di vicini e dimensione di embedding, è immediato confrontare i risultati per MAPE e RMSE.

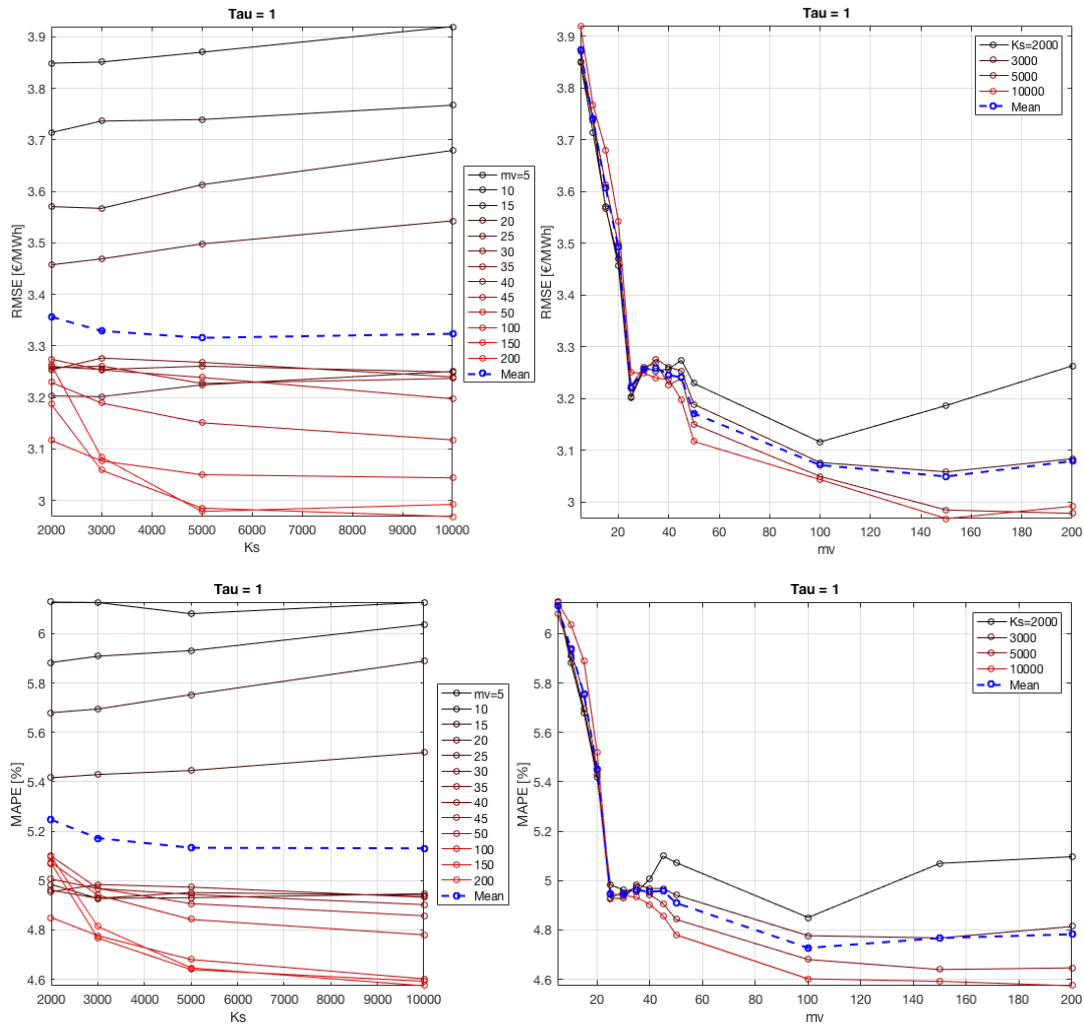


Figura 67 - Sezioni dello spazio dei parametri : $\tau=1$, sopra: RMSE [€/MWh], sotto: MAPE

MAPE ed RMSE sono leggermente diversi localmente, però sul complesso non stravolge i risultati, il MAPE colloca il valori molto vicini al minimo per m già poco sopra a 100 e a 200 ha ancora margine di miglioramento, invece il RMSE lo raggiunge per $m = 150$.

Comunque sia è vero che:

- Con più vicini migliorano le prestazioni
- La dimensione di embedding ha minimo da cercare fra 120-150 fino a 200 – 220

A questo punto si propone questa calibrazione aggiuntiva che dovrebbe chiarire la situazione:

$$m = 120 : 10 : 220$$

$$\tau = 1$$

$$k_s = 5000 \ 6000 \ 7000 \ 8000 : 500 : 15000$$

Da subito si nota che $m = 120, 200$ e 220 sono i “peggiori”, il range di RMSE oramai è irrisorio (2,95 – 3,02 €/MWh) e gli altri valori di m hanno errori di qualche centesimo sotto perciò non ha neanche senso andare a cercare meglio. L’intervallo di k_s è stato ampliato verso il basso perché è ciò che accade nel set estivo e lo stesso per i valori di m al fine di poter fornire un buon confronto fra estate e inverso con orizzonte unitario.

Il confronto con il set estivo conferma il minimo per $\tau = 1$ e m alti anche a 170 come accadeva in quel caso ma diminuendo $k_s < 9000$ le prestazioni peggiorano constatando che per l’inverno servono circa $k_s = 13000 - 13500$ vicini.

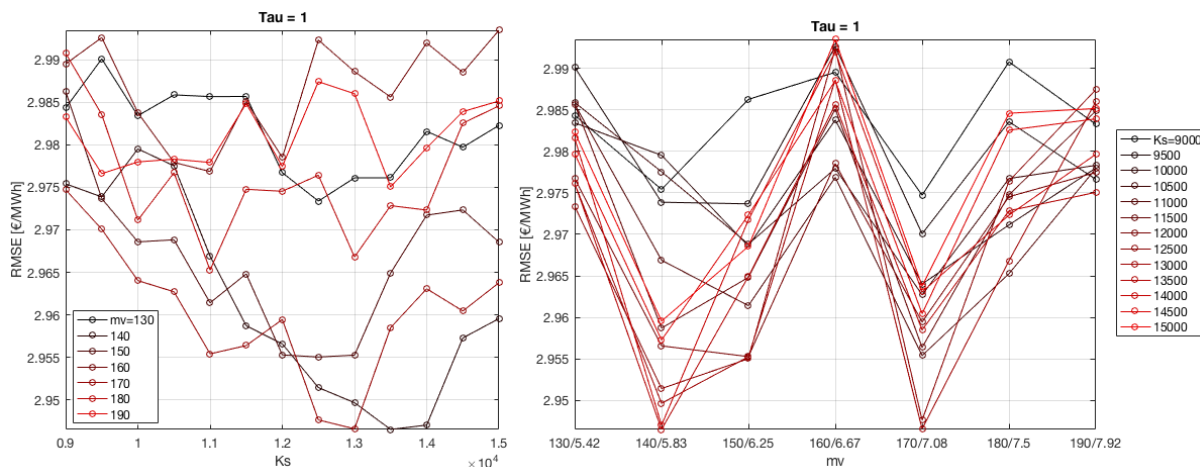


Figura 68 - Sezione dello spazio dei parametri , $\tau = 1$, RMSE

Ormai si parla di variazioni veramente piccole e abbiamo identificato il minimo anche per k_s , che a seconda di m passa da 13000 a 13500, in particolare due terne eccellono:

$$m = 140 \quad \tau = 1 \quad k_S = 13500 \rightarrow RMSE = 2,947 (65) \frac{\epsilon}{MWh} \quad MAPE = 4,46 \% \rightarrow \text{tempo}^{22} = 430 \text{ s}$$

$$m = 170 \quad \tau = 1 \quad k_S = 13000 \rightarrow RMSE = 2,947 (66) \frac{\epsilon}{MWh} \quad MAPE = 4,487 \% \rightarrow \text{tempo} = 589 \text{ s}$$

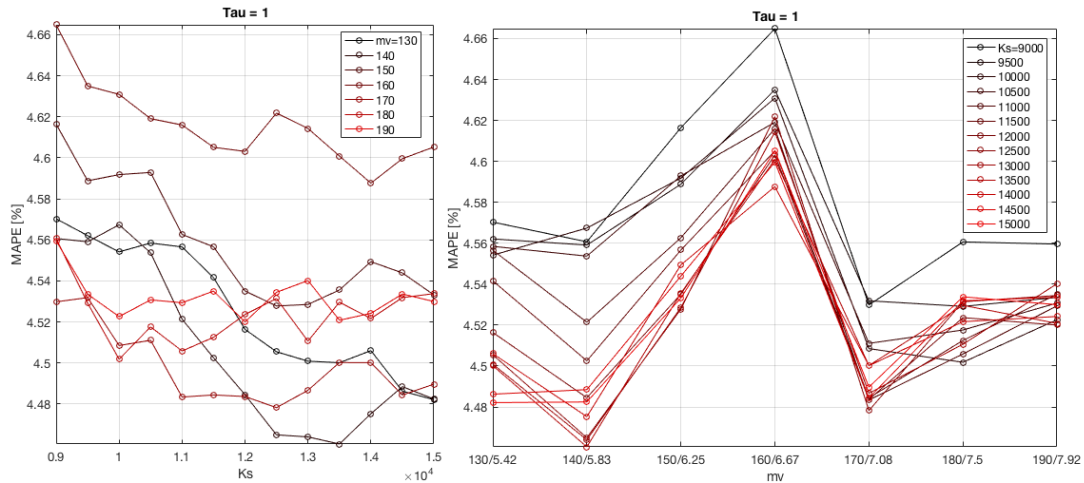


Figura 69 - Sezione dello spazio dei parametri , $\tau = 1$, MAPE

Il MAPE smentisce $m = 170$ ma sottolinea 140, in più l'incremento della dimensione di embedding incide sul tempo di computazione (anche il numero di vicini, ma di meno) e la prima terna è più veloce. Se il set di calibrazione fosse soltanto questo la scelta sarebbe:

$$m = 140 \quad \tau = 1 \quad k_S = 13500 \rightarrow RMSE = 2,947 (65) \frac{\epsilon}{MWh} \quad MAPE = 4,46 \% \rightarrow \text{tempo} = 430 \text{ s}$$

Il MAPE per $m = 170$ è di pochissimo inferiore al 4,8 % per $k_S = 12500$ e non diversamente, come per il set estivo l'errore con questo grado di precisione alla terza cifra decimale è di poco conto, è vero che i risultati per inverno e estate sono leggermente diversi però si discute per valori irrilevanti, il problema principale è il valore di k_S : l'ottimo estivo trova scarse prestazioni in inverno e viceversa.

²² Run singola MATLAB R2016b su MacBook Pro mid 2012 , intel dual core 2,5 (3,1) GHz, ram 4Gb DDR3 1600 MHz.

Orizzonte di previsione a 24 ore

Ennesima ripetizione, estremi di calibrazione:

Minimo assoluto $\rightarrow$ **RMSE** = 5,85 €/MWh con $m = 150$, $\tau = 1$ e $k_s = 5000$

Massimo assoluto $\rightarrow$ **RMSE** = 3220 con €/MWh con $m = 250$, $\tau = 10$ e $k_s = 500$

Innanzitutto le prestazioni calano e tutto sommato non di molto, ci aspettiamo valori di ottimo attorno al **RMSE** = 6 €/MWh e una corrispondenza a **MAPE** circa 11%, la distribuzione è asimmetrica con la coda verso errori alti che sono concentrati sotto i 25 €/MWh (50% il **MAPE**) .

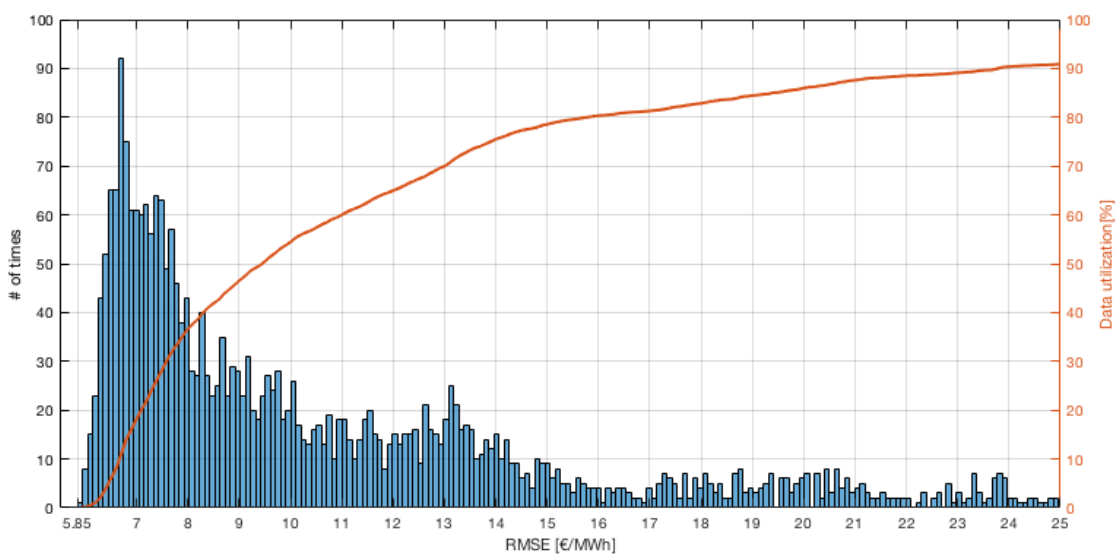


Figura 70 - Istogramma : prima calibrazione, set invernale , $T=24$

Con il limite variabile fra il minimo assoluto e 25 €/MWh di **RMSE** viene costruito il solito diagramma, sia il set estivo che l'invernale a $T = 1$ necessitano di moltissimi vicini ed a questo punto sarà così anche in questo caso, nell'estivo m sta sotto il 100 con una forte dipendenza con il fattore di ritardo che complica la scelta.

Il riferimento per questo caso è la Figura 87 (Appendice B a pagina 136) e poco sotto al limite di 25 €/MWh c'è un valore isolato di **RMSE** = 10 €/MWh a cui corrisponde $m = 400$, è stato tagliato dal grafico e non viene considerato. Procedendo egualmente si eliminano le terne con $m \leq 200$, $k_s \geq 2000$ e $\tau \leq 15$, sempre sulla consiglio di Figura 87 , il risultato include terne comprese tra:

Minimo assoluto $\rightarrow$ **RMSE** = 5,85 €/MWh con $m = 150$, $\tau = 1$ e $k_s = 5000$

Massimo assoluto $\rightarrow$ **RMSE** = 8,95 con €/MWh con $m = 200$, $\tau = 3$ e $k_s = 2000$

La situazione è gemella del set estivo a pari orizzonte, ci sono valori di minimo che derivano da valori di τ ed m diversi ma che conservano il valore del loro prodotto, si riportano le evidenze per qualche valore di m e il diagramma che mostra il minimo dell'errore in funzione di $\tau * m$.

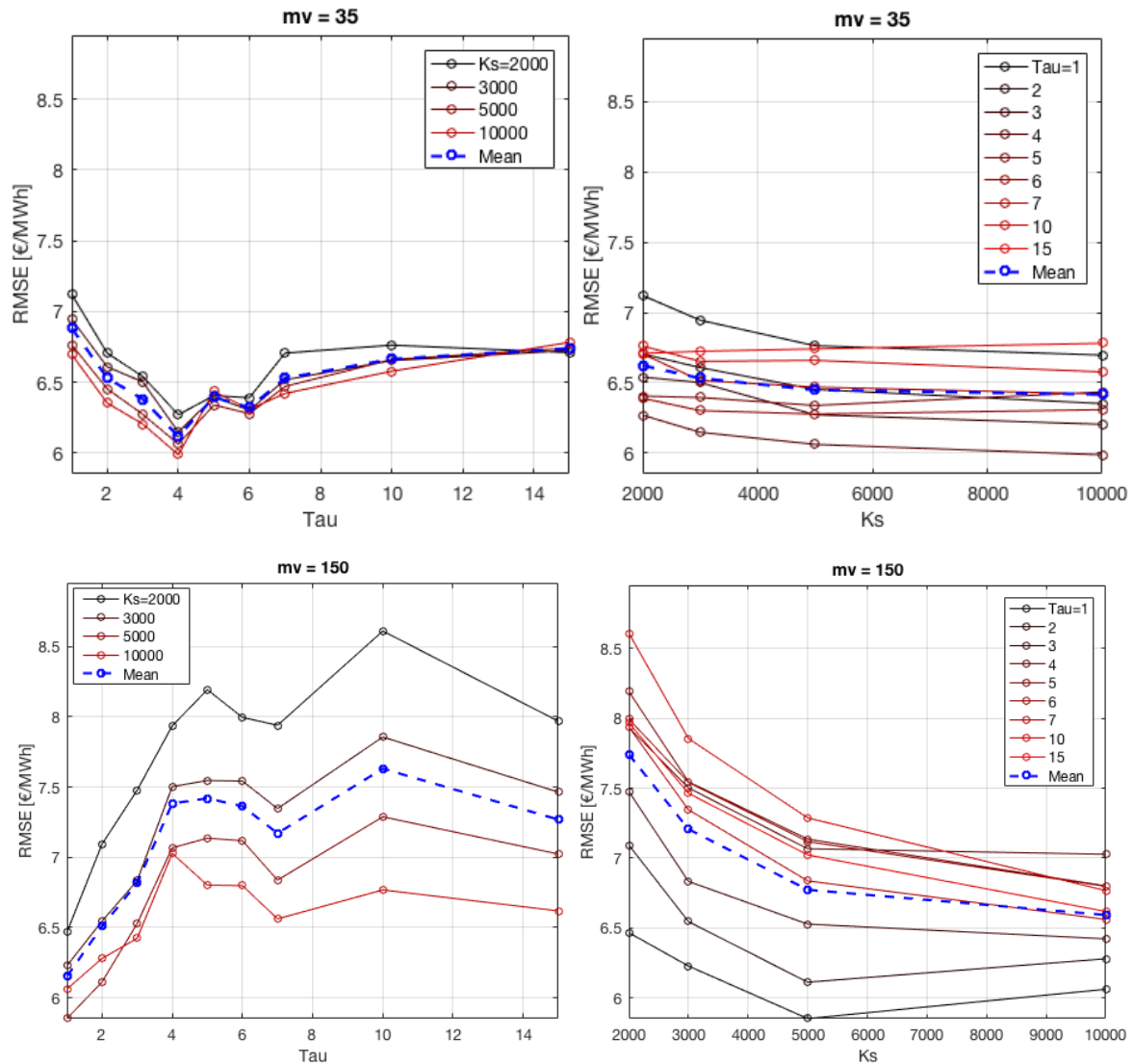


Figura 71 – Sezioni dello spazio dei parametri : $m=35, 150$

Per $m = 150$ è strano che il minimo si sposti a $\tau = 1$ per la previsione ad un giorno, verosimilmente ci si aspetta che un ritardo maggiore sia migliore. Altrettanto anomalo notare che per $\tau=1$ l'ottimo raggiunto prima di $K_s = 10000$, si potrebbe pensare ad una terna "fortunata" che sta fuori dal minimo globale ma è bene non escludere questa possibilità, anzi, m è indagato molto più finemente da 5 a 50 che non altrove perciò cinque delle sei terne che contribuiscono al minimo sono in quella zona.

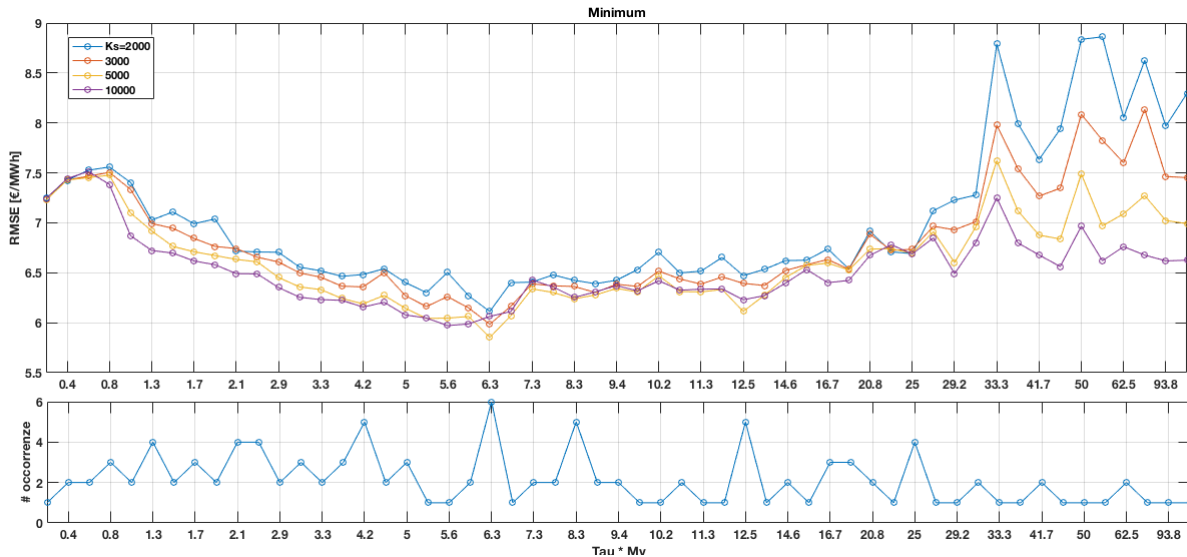


Figura 72 – Sezioni dello spazio dei parametri : $\tau * m$, RMSE

Il minimo assoluto di $\tau * m$ corrisponde a 6,3 giorni e rappresenta esattamente la coppia $m = 150$ e $\tau = 1$ perché dei sei valori presenti è plottato il minore, come ribadito prima da $m = 100$ in poi proseguo a passi di cinquanta mentre fino a $m = 50$ posseggo valori a passi di cinque. E' difficile stabilire come muoversi nella zona vicino alla settimana, ma bisogna cercare lì, il fatto che la terna migliore stia in una zona poco esplorata suggerisce di approfondire la zona limitrofa a $m = 150$. In generale questo minimo suggerisce di soffermarsi su k_s da 5000 (o meno) a 10000 ma non oltre di sicuro, con i dati attuali è meglio 5000 e la figura a valle riporta le sei coppie che incluse dal minimo.

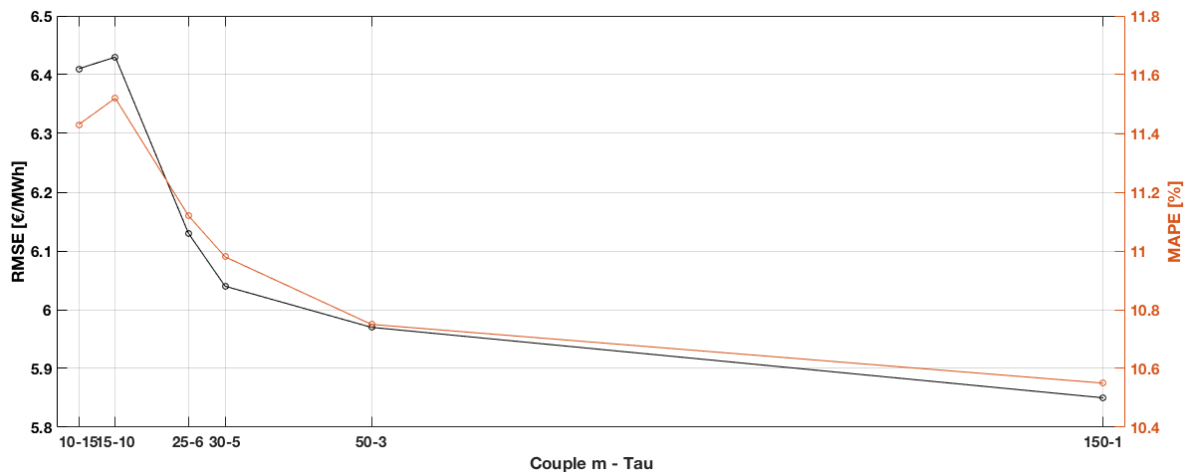


Figura 73 – Dettaglio del minimo, $k_s = 5000$: terne con $\tau * m = 150 / 6,3$ giorni

I primi cinque punti sono concentrati per $m < 50$ ed il minimo è traslato cento unità più avanti, l'ultimo plot indica che il minimo progredisce verso vettori costruiti con molti elementi poco separati fra loro fino a considerare valori in successione. Brevemente la situazione è questa: la prima calibrazione e successiva scrematura indicano una struttura articolata che richiede lo studio della dipendenza tra dimensione del vettore e fattore di ritardo, una volta identificato il minimo al variare del prodotto serve

capire qual è la proporzione ottima fra i due. A questo punto si raffina, cerchiamo terne che costruiscono vettori con molti dati e poco ritardo includendo dai 4000 a 10000 vicini:

$$\mathbf{m} = 100 : 10 : 180$$

$$\tau = 1 \text{ 2}$$

$$\mathbf{k}_s = 4000 : 500 : 7000 \text{ e } 8000 : 1000 : 10000$$

Le figure più indicative sono quelle per $\tau = 1$ poiché gli errori in tal caso sono sempre i più bassi, l'intervallo di $\mathbf{k}_s$ che è stato esplorato indica l'ottimo a 5000 ma vi è qualche indecisione per quanto riguarda la dimensione di embedding.

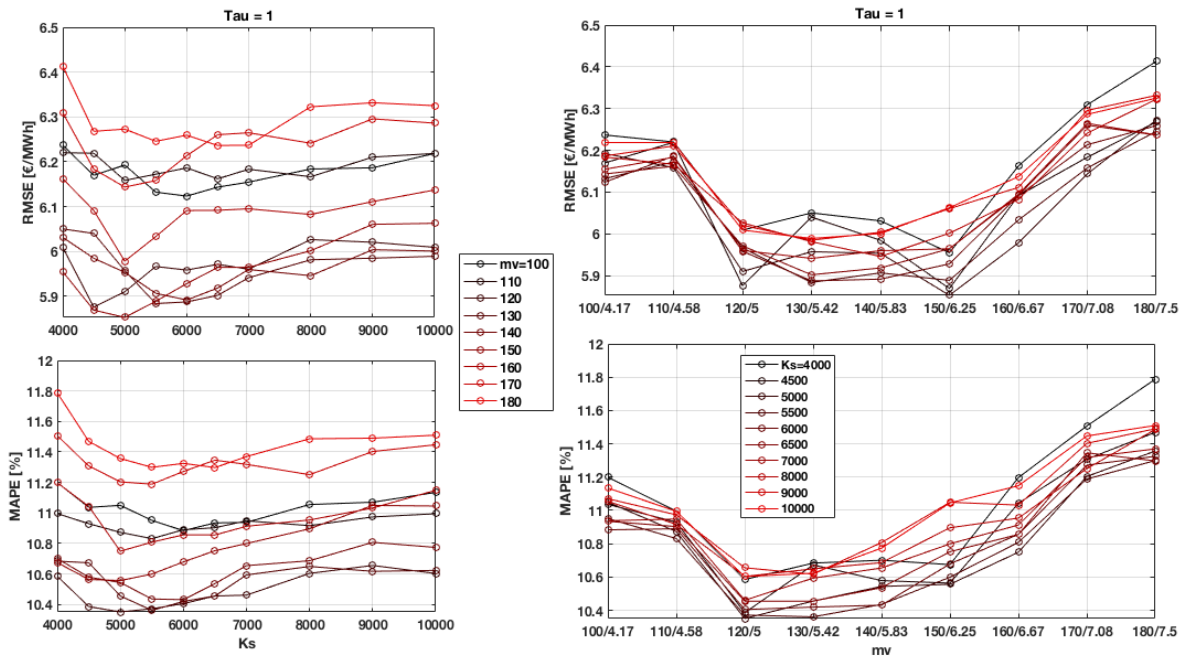


Figura 74 - Sezioni dello spazio dei parametri : $\tau = 1$

C'è incertezza sulla finestra di $\mathbf{m} = 120 / 150$ perché le curve a $\mathbf{k}_s$ costante da 4000 a 5000 sono variabili su entrambe le metriche in quella zona, tolte quelle il fascio per $\mathbf{k}_s \geq 5500$ è uniforme ma esclude il minimo assoluto. Le terne che conducono ai risultati migliori sono le due scritte di seguito e utilizzano 5000 vicini:

$$\mathbf{m} = 150 \quad \tau = 1 \quad \mathbf{k}_s = 5000 \rightarrow \mathbf{RMSE} = 5,85 \frac{\text{€}}{\text{MWh}} \quad \mathbf{MAPE} = 10,56 \% \rightarrow \mathbf{tempo}^{23} = 258 \text{ s}$$

$$\mathbf{m} = 120 \quad \tau = 1 \quad \mathbf{k}_s = 5000 \rightarrow \mathbf{RMSE} = 5,91 \frac{\text{€}}{\text{MWh}} \quad \mathbf{MAPE} = 10,35 \% \rightarrow \mathbf{tempo} = 174 \text{ s}$$

Situazioni di questo genere necessitano di molta attenzione, una calibrazione robusta deve includere un certo livello di affidabilità, applicando la terna ad un periodo che il modello non ha mai visto prima

²³ Run singola MATLAB R2016b su MacBook Pro mid 2012, intel dual core 2,5 (3,1) GHz, ram 4Gb DDR3 1600 MHz.

bisogna tenere conto di un'incertezza vicino al minimo, in altre parole è meglio se la curva presenta un minimo ampio che non vari bruscamente in altre circostanze, chiaramente non è sempre possibile (si veda il caso a $T = 1$ di Figura 68 e Figura 69) ma in questo caso è possibile scegliere di sacrificare un minimo di performance guadagnandone in stabilità. Per cui le terne suggerite rimangono le stesse solo con k_S aumentato a 5500:

$$m = 150 \quad \tau = 1 \quad k_S = 5500 \rightarrow RMSE = 5,89 \frac{\text{€}}{MWh} \quad MAPE = 10,6 \% \rightarrow \text{tempo} = 253 \text{ s}$$

$$m = 120 \quad \tau = 1 \quad k_S = 5500 \rightarrow RMSE = 5,97 \frac{\text{€}}{MWh} \quad MAPE = 10,37 \% \rightarrow \text{tempo} = 184 \text{ s}$$

Osservando nel dettaglio la curva a $k_S = 5500$ si nota un ottimo compromesso fra affidabilità e precisione che include entrambe le metriche, la scelta definitiva:

$$m = 130 \quad \tau = 1 \quad k_S = 5500 \rightarrow RMSE = 5,88 \frac{\text{€}}{MWh} \quad MAPE = 10,36 \% \rightarrow \text{tempo} = 187 \text{ s}$$

4.2.3 Confronto a pari orizzonte di previsione e scelta della terna

Nell'ultimo paragrafo si sono studiati individualmente i due set di calibrazione per apprezzare le complesse dinamiche che portano a soluzioni anche sostanzialmente diverse, la scelta della terna ottimale è un passo estremamente delicato ancor più se vogliamo prevedere serie con spiccate non linearità, per cui, è lecito pensare che la soluzione non sia immediata, infatti non lo è. A questo punto i due set vengono analizzati in parallelo e si studia come costruire la terna ottima elaborando le informazioni acquisite, la previsione ad un ora è la più semplice e precisa ma anche inutile, invece quella giornaliera è interessante ma molto complicata a cui si sommano risultati diversi in estate o in inverno.

Orizzonte di previsione ad 1 ora

Rappresenta le condizioni in cui si riscontrano i risultati migliori e costituisce ‘l'allenamento’ perfetto per la previsione, è il punto di partenza sul quale basare confronti ed osservazioni perché mano a mano che si aumenta l'orizzonte di previsione progressivamente calano le prestazioni del modello.

Riassumendo, i set invernale ed estivo sono concordi riguardo il fattore di ritardo che da subito si palesa essere $\tau = 1$ per entrambi, lo stesso per la dimensione m che include incertezze a cavallo di 160 dove l'errore aumenta causa la forte correlazione settimanale ma comunque è stabile poco sotto al 200. Il problema è il numero di vicini k_s che raddoppia dall'estate all'inverno complicando il tutto, la scelta deve essere unica e l'idea prevede di ricostruire l'errore dei due set congiunti una volta calibrati sul medesimo intervallo. Si utilizzano parte dei dati ottenuti dalle ultime calibrazioni, ampliandoli per ottenere un confronto diretto su intervalli che comprendo l'interna fascia di minimo:

$$m = 140 : 10 : 190$$

$$\tau = 1$$

$$k_s = 6000 : 500 : 14000$$

Dopodiché anche i valori di k_s sono ben distribuiti fra i due minimi che erano 6500/7000 per l'estivo e 13000/13500 per l'invernale, ora si dispone di tutte le terne e relativi errori in duplice copia, una per ogni set. Il problema è stabilire come pesare i contributi stagionali sul totale, allora si procede a ritroso per comporre un nuovo set costituito dalla somma dei due, calibrati sugli intervalli scritti sopra, brevemente il procedimento è:

$$RMSE^2 N_{set} = \sum_i^{N_{set}} (x_{REALE} - x_{PREV})_i^2 \quad (36)$$

Nel caso dello scarto in media quadratica si inizia calcolando la somma dei residui elevati al quadrato per entrambi i set e poi si ottiene il valore globale di errore:

$$RMSE = \sqrt{\frac{RMSE_{est}^2 * N_{est} + RMSE_{inv}^2 * N_{inv}}{N_{est} + N_{inv}}} \quad (37)$$

Lo stesso si applica al MAPE muovendosi a ritroso attraverso le deviazioni percentuali :

$$\frac{MAPE \ N_{set}}{100} = \sum_{i=1}^{N_{set}} \frac{AE_i}{x_{REALE}} \quad (38)$$

e riferendole a entrambi i set :

$$MAPE = \frac{MAPE_{est} \ N_{est} + MAPE_{inv} \ N_{inv}}{100 (N_{est} + N_{inv})} \quad (39)$$

Ad ogni modo ora è possibile avere una visione complessiva del problema, grazie questo passaggio otteniamo un diagramma che racchiude tutte le informazioni per estate ed inverno con un minimo distinto che è unico.

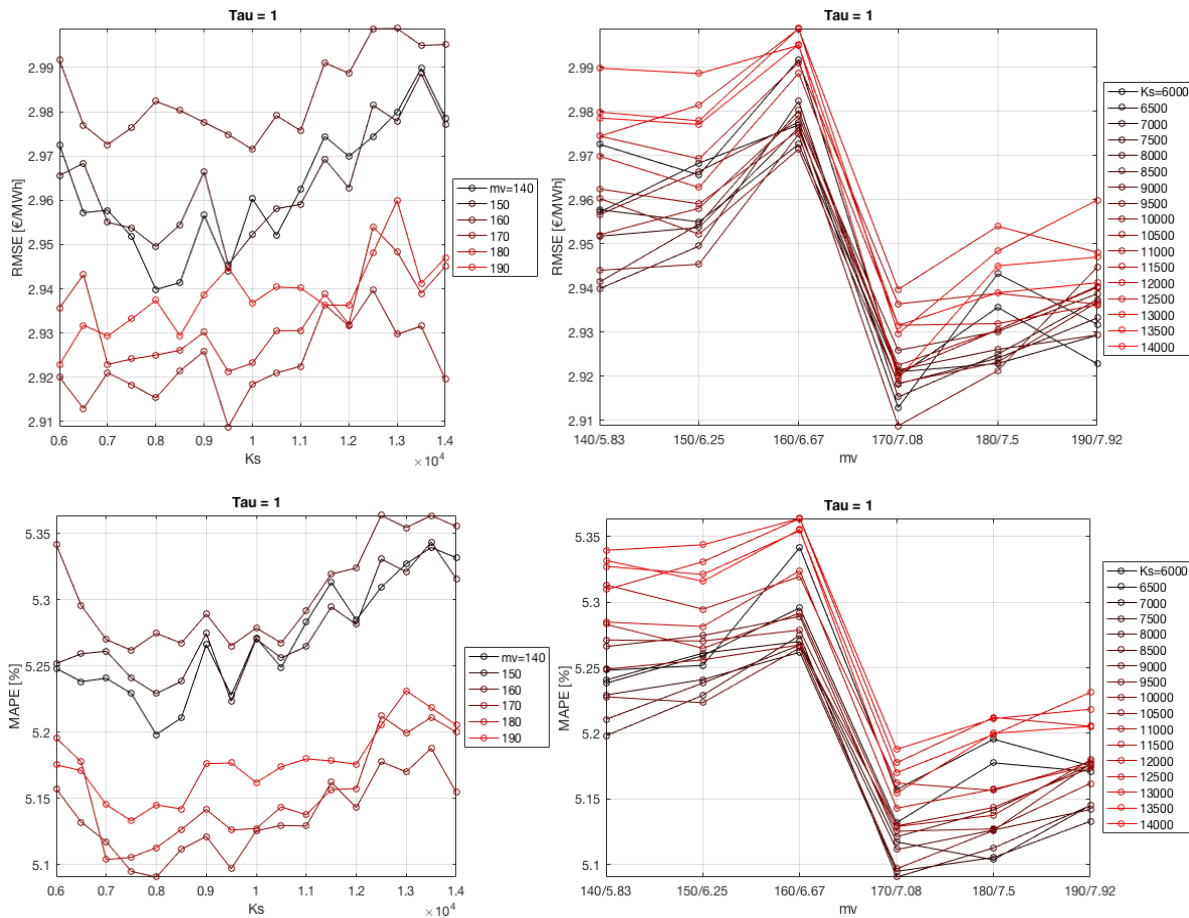


Figura 75 – Sezione dello spazio dei parametri , $\tau = 1$, sopra : RMSE [€/MWh] , sotto : MAPE [%]

Differenze davvero minime fra MAPE e RMSE, piccolo dubbio riguardo il numero di vicini che per l'ennesima volta si conferma come la variabile con maggiori incertezza, le scelte sono due:

$$\begin{aligned} \mathbf{m} = 170 \quad \tau = 1 \quad k_S = 9500 &\rightarrow \mathbf{RMSE} = 2,91 \frac{\text{€}}{\text{MWh}} \quad \mathbf{MAPE} = 5,1 \% \\ \mathbf{m} = 170 \quad \tau = 1 \quad k_S = 8000 &\rightarrow \mathbf{RMSE} = 2,915 \frac{\text{€}}{\text{MWh}} \quad \mathbf{MAPE} = 5,09 \% \end{aligned}$$

Prima di proseguire vi è una questione che nasce spontanea, il fatto che il minimo sia così legato al grado di autocorrelazione settimanale fa pensare che a scala giornaliera accada lo stesso. In questo caso utilizziamo sempre ritardo unitario quindi valori in successione ed $\mathbf{m}/24$ rappresenta quanti giorni passati sono inclusi in ogni vettore. Nel capitolo 3 si sottolinea che la correlazione è molto simile fra scala settimanale e giornaliera, il massimo assoluto si ha per sette giorni ma è comunque utile approfondire cosa succede a cavallo delle ventiquattro ore o multipli. Si riscontra che in corrispondenza di $\mathbf{m} = 25$ l'errore aumenta analogamente al 170 (si veda Figura 57) e da lì in poi diminuisce fino a sette giorni con il brusco “gradino” a $\mathbf{m} = 160$ (6,3 giorni), quindi la correlazione incide sulla previsione in questo senso e con orizzonte unitario le prestazioni decadono puntualmente nel punto a massima correlazione ed il minimo è subito dopo. Detto questo abbiamo la sicurezza che $\mathbf{m} = 170$ sia il valore ottimo e la terna che utilizza più vicini è la migliore perché è forte a prescindere dalla metrica anche se non include il minimo di entrambe.

$$\mathbf{m} = 170 \quad \tau = 1 \quad k_S = 9500 \rightarrow \mathbf{RMSE} = 2,91 \frac{\text{€}}{\text{MWh}} \quad \mathbf{MAPE} = 5,1 \%$$

Per quanto riguarda il tempo di calcolo è impossibile da valutare perché le misure sono costruite a partire da due prove diverse che utilizzano trenta e trentuno giorni, le singole run rimangono attorno a 400s ma includendo il doppio dei dati richiedono più tempo ed in più non riguarda solo le tempistiche ma anche la memoria RAM della macchina che deve far fronte all'incremento.

Orizzonte di previsione a 24 ore

La previsione ad un giorno è una situazione complessa perché le indicazioni portano a considerare un ottimo per il prodotto fra due parametri $\tau * m$ e non una coppia che sia migliore di tutte le altre. Il problema non esiste per la previsione ad un ora perché fissando subito uno dei due parametri il prodotto dipendeva solo più dall'altro e in ogni caso vi è solo un grado di libertà, qui invece la prima scrematura include più valori del fattore di ritardo rendendo possibile comporre coppie diverse che ne conservano il prodotto. A seguito si riepiloga brevemente quanto riscontrato per i due periodi:

- **Calibrazione estiva** : è quella più anomala perché analizzata singolarmente sostiene un minimo che dipende comunque dal prodotto $\tau * m$ ma da subito per valori alti di circa 600 – 1000 che corrispondono a 25 – 42 giorni con il sentore che il minimo tenda all'estremo maggiore. Infatti, proseguendo il minimo si stabilizza per un mese ed una settimana, precisamente 37,5 giorni e $\tau * m = 900$. Sebbene le conclusioni siano sostanzialmente diverse in ogni situazione è molto presente l'influenza della correlazione, qui la correlazione mensile non era neppure accennata ma semplicemente includendo una settimana in più si ricade in un minimo evidente che per questa calibrazione in particolare è il migliore ma potrebbe non esserlo in altre, basti vedere cosa succede inverno. Riguardo ai vicini siamo all'estremo superiore, 15000 vettori la cui dimensione è meno della metà rispetto ad altri casi, questo minimo riceve informazioni da vettori che coprono sempre lo stesso intervallo di tempo ma sono più “*diluiti*” e per questo ha bisogno di molti vicini, anche qui più del doppio della calibrazione invernale per esempio.
- **Calibrazione invernale** : presenta caratteristiche in comune fra i due orizzonti di previsione e mantiene la spiccata dipendenza dal prodotto $\tau * m$, il fattore di ritardo contribuisce ad ottimi risultati anche sopra l'unità purché l'informazione contenuta nei vettori includa dai 5,5 a 6,7 giorni corrispondenti a $\tau * m = 120 - 150$, quindi leggermente meno della settimana a differenza dell'orizzonte unitario. In assoluto questo è il caso che utilizza il minor numero di vicini, però ne servono comunque diverse migliaia per raggiungere il minimo,

Conclusioni sostanzialmente diverse che non vedono molto in comune fra estate ed inverno a parte il bisogno di pesare la previsione sul contributo di moltissimi vicini come accade in ogni singolo caso analizzato. Dal confronto diretto dei due set non si arriva ad una conclusione ma era prevedibile, perciò ricostruiamo l'errore complessivo grazie alle equazioni (37) e (39) però partendo dalla calibrazione generale indicata a pagina 88. Tetto di RMSE fissato a 25 €/MWh (Figura 88, Appendice B a pagina 136) su un massimo di 2785 €/MWh e la ricerca delle terne più frequenti conclude che si mantengano solo quelle con $25 \leq m \leq 200$, $1 \leq \tau \leq 15$ e $k_s \geq 2000$ con minimo a 6,11 €/MWh e massimo che cala a 8,66 €/MWh.

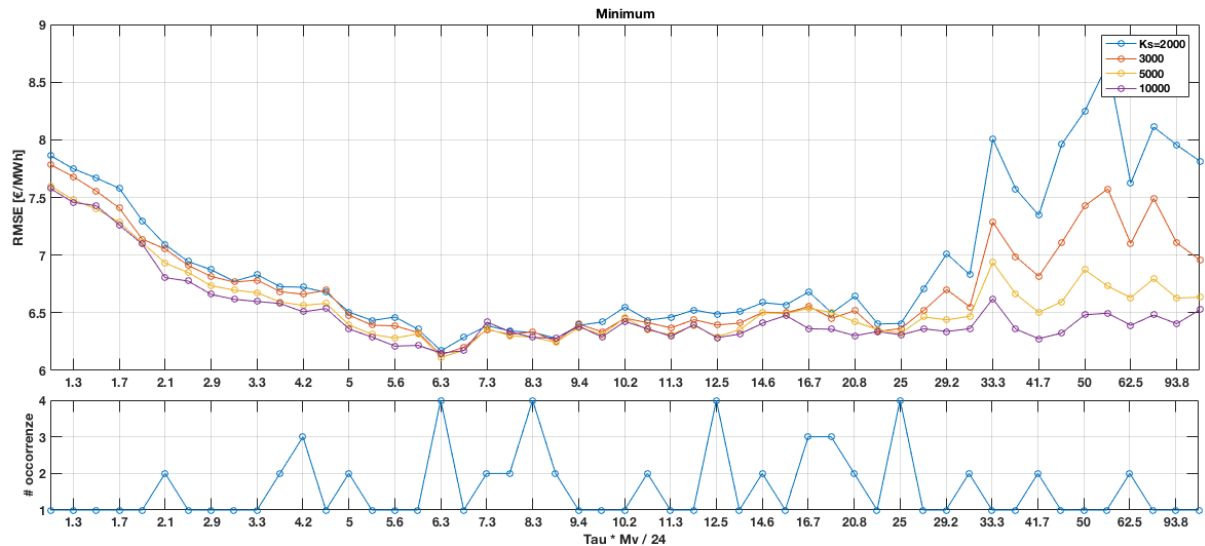


Figura 76 -Sezione dello spazio dei parametri : $\tau * m$, RMSE

Questo approccio considera le informazioni di entrambi i set e le disparità che presentano analizzandoli singolarmente sono chiarite: nella parte di grafico che precede i sette giorni l'errore diminuisce fino a toccare il minimo a $150 / 6,3$ poi aumenta lentamente e rimane poco variabile per sostenere un plateau a valori poco superiori, almeno per $k_s = 10000$. Il MAPE conferma e sottolinea che probabilmente sono sufficienti meno di 10000 vicini in questo caso congiunto dato che la curva migliore è sempre a $k_s = 5000$.

Le circostanze sono complicate solamente perché volendo prevedere con 24 ore di anticipo il modello non converge verso un solo valore di τ ma dipende dalla dimensione del vettore, la correlazione settimanale continua ad incidere sulle prestazioni ed è chiaro che a prescindere dall'orizzonte la scelta ottima preveda di costruire vettori che includano informazioni fino a sette giorni prima, non esattamente sette ma nemmeno troppo oltre o prima, l'esperienza di $T = 1$ suggerisce che sono migliori poco sopra la settimana. A questo punto interessa capire qual è il modo migliore per costruire i vettori includendo abbastanza informazioni e principalmente ci sono due modi per farlo:

- Vettore con pochi valori ma con ampio ritardo fra di essi, ad esempio 25/30 elementi escludendone 6/7 fra ognuno.
- Vettore con molti valori ravvicinati o successivi, simile alla previsione ad orizzonte unitario che utilizza 170 elementi successivi.

Più tutti i casi intermedi come nella calibrazione generale dove le coppie con $\tau * m = 150$ sono quattro : 25/6, 30/5, 50/3 e 150/1 dove l'ultima è la migliore in assoluto e mano a mano che m diminuisce le prestazioni fanno lo stesso (si parla di passare da 6,15 a 6,23 €/MWh). La calibrazione generale indaga bene la zona di m dal 5 al 50 ma non altrettanto il resto tanto che 3 su 4 coppie ottime sono in quella zona ma sembra che la più promettente sia altrove.

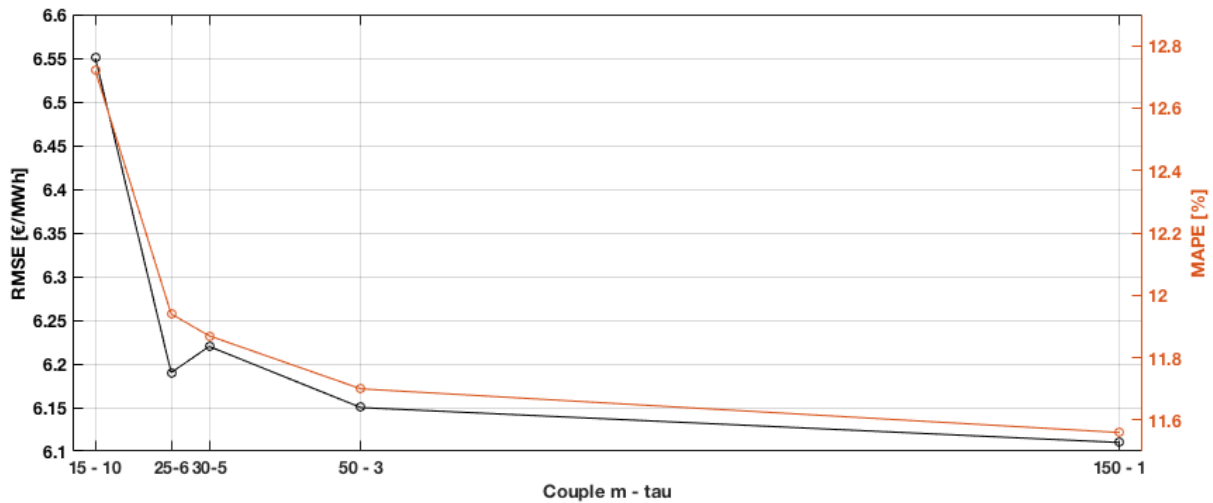


Figura 77 - Dettaglio del minimo, $k_s = 5000$: terne con $\tau * m = 150$

Quindi la previsione funziona meglio con vettori corposi e svariate migliaia di vicini, ovviamente questo incide sul tempo di computazione, sperando di concludere si continua a cercare sempre più nel dettaglio :

$$m = 130 : 10 : 180$$

$$\tau = 1, 2$$

$$k_s = 4000 : 500 : 7000 \text{ e } 8000 : 1000 : 10000$$

Importante notare che ci stiamo muovendo nella stessa direzione della terna ad orizzonte unitario, sebbene nulla sia sicuro se questa calibrazione fornisse risultati migliori già a passi di 10 su alti m allora è probabile che le difficoltà di interpretazione si traducano nello stesso risultato. Si calibra sempre separatamente ricostruendo i risultati in seconda sede.

Si studia un periodo ristretto da 5,4 a 15 giorni corrispondenti a $\tau * m = 130$ e 360 ed il minimo non si muove, sempre fisso a 6,3 con scarto quadratico medio sotto il 6,2 €/MWh e MAPE circa 11,6 % quindi si parla di piccolissimi scostamenti rispetto alla calibrazione generale. Conviene utilizzare elementi in sequenza perché $\tau = 1$ da errori più bassi in generale e per quando riguarda la dimensione del vettore entrambe le metriche convergono a 150 mentre il numero di vicini è stabilito pari a 6000.

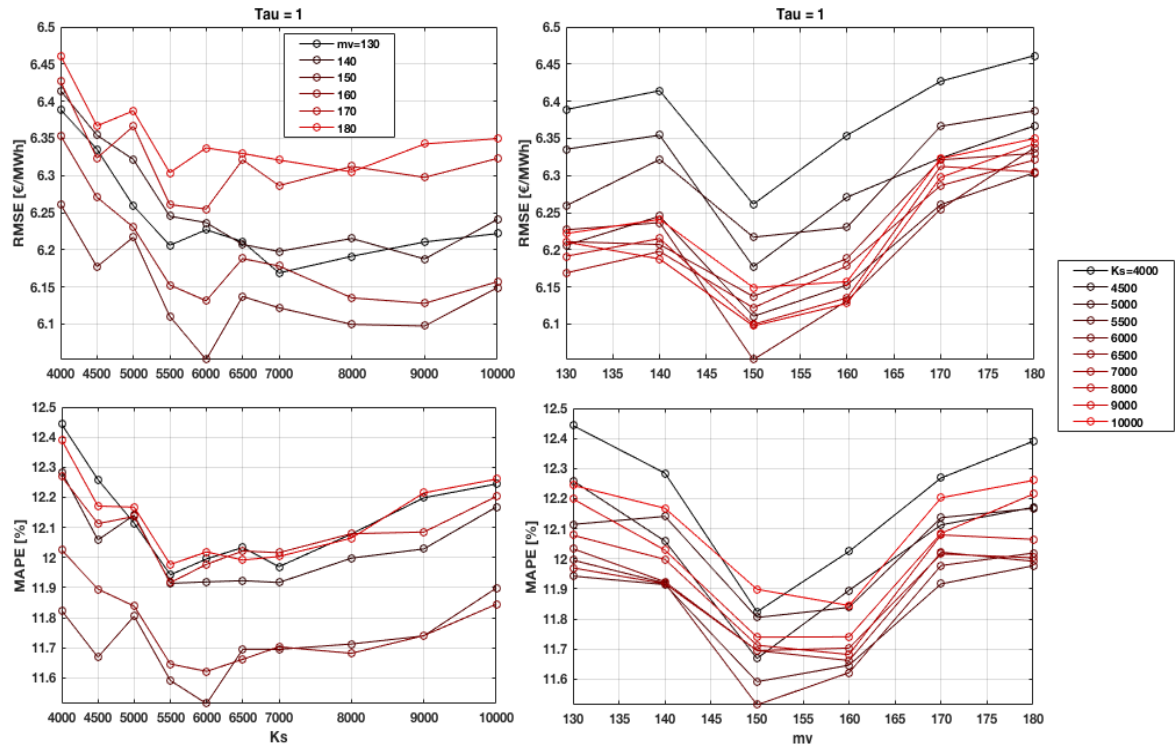


Figura 78 – Sezione dello spazio dei parametri , $\tau = 1$

Come ribadito, le curve con ritardo $\tau = 2$ offrono buoni risultati che però sono peggiori in confronto al ritardo unitario, scelta effettuata anche in questo caso, molto vicina al caso della previsione ad un ora:

$$m = 150 \quad \tau = 1 \quad k_s = 6000 \rightarrow RMSE = 6,09 \frac{\text{€}}{MWh} \quad MAPE = 11,58 \%$$

CAPITOLO 5: RISULTATI

5.1. Set di verifica

Una volta definiti i parametri migliori in ogni set di calibrazione si passa alla previsione vera e propria, non ci sono accortezze ulteriori né modifiche nel modello, semplicemente si impostano i parametri per prevedere un periodo che il modello non ha mai visto prima. Disponiamo comunque del profilo reale ma viene usato solamente per verificare la bontà della previsione, serve per quantificare l'efficacia della calibrazione su un preciso set ed è per questo che si calibra e valuta su stagioni diverse. E' molto importante che il periodo di test sia escluso dai dati di partenza, nello specifico dal fitting set (si veda : Capitolo 3: METODO). Esattamente sono stati utilizzati poco più di dodici anni completi come fitting set e si è calibrato su intervalli mensili e bimestrali fra estate 2017 e inverno 2018. Il testing set si costruisce lungo il 2018, vengono utilizzati quattro mesi, uno per stagione, precisamente:

- **Inverno** : febbraio 2018
- **Primavera** : Aprile 2018
- **Estate** : luglio 2018
- **Autunno** : Novembre 2018

Per ogni mese si considera la seconda settimana completa da cui si risale a giorni feriali e festivi, quindi si applica la previsione su diverse scale temporali che sono mensile, settimanale e giornaliera. A seguito vengono raccolte le terne ottime risultanti dalla calibrazione nel capitolo antecedente, due tabelle riassumono le scelte per la previsione ad orizzonte unitario e giornaliero.

SET CALIBRAZIONE	m_p	τ	k_s	RMSE	MAPE ²⁴
ESTIVO	170	1	6500	2,76 €/MWh	5,72 %
INVERNALE	140	1	13500	2,95 €/MWh	4,46 %
GLOBALE	170	1	9500	2,91 €/MWh	5,1%

Tabella 12 - $T = 1$, Parametri ottimi

SET CALIBRAZIONE	m_p	τ	k_s	RMSE	MAPE ²⁹
ESTIVO	65	15	15000	5,59 €/MWh	11,27 %
INVERNALE	130	1	5500	5,88 €/MWh	10,36 %
GLOBALE	150	1	6000	6,09 €/MWh	11,58 %

Tabella 13 - $T = 24$, Parametri ottimi

²⁴ Calcolato su tutto il periodo di calibrazione : Monthly MAPE per estate e inverno, bimestrale altrimenti.

5.2. Orizzonte di previsione ad 1 ora

SET CALIBRAZIONE →	ESTATE		INVERNO		GLOBALE	
SET TEST ↓	RMSE	MMAPE[%]	RMSE	MMAPE[%]	RMSE	MMAPE[%]
INVERNO : FEBBRAIO 18	4,27 (1 4)	4,29 (2 3)	4,43 (3 4)	4,44 (3 3)	4,30 (2 4)	4,28 (1 3)
PRIMAVERA : APRILE 18	3,97 (1 3)	6,54 (2 4)	4,06 (3 3)	6,57 (3 4)	4,00 (2 3)	6,53 (1 4)
ESTATE : LUGLIO 18	2,18 (1 1)	2,59 (1 1)	2,41 (3 1)	2,87 (3 1)	2,19 (2 1)	2,60 (2 1)
AUTUNNO : OTTOBRE 18	3,56 (2 2)	3,18 (1 2)	3,70 (3 2)	3,35 (3 2)	3,54 (1 2)	3,14 (2 2)

Tabella 14 – Mensile

SET CALIBRAZIONE →	ESTATE		INVERNO		GLOBALE	
SET TEST ↓	RMSE	WMAPE[%]	RMSE	WMAPE[%]	RMSE	WMAPE[%]
INVERNO : 12/02 – 18/02	3,81 (2 2)	4,75 (2 3)	3,89 (3 3)	4,94 (3 3)	3,79 (1 2)	4,71 (1 3)
PRIMAVERA : 9/04 – 15/04	3,90 (2 3)	5,41 (3 4)	3,75 (1 2)	5,15 (1 4)	3,93 (3 3)	5,33 (2 4)
ESTATE : 7/07 – 15/07	1,97 (2 1)	2,40 (1 1)	2,17 (3 1)	2,71 (3 1)	1,95 (1 1)	2,41 (2 1)
AUTUNNO : 15/10 – 21/10	4,88 (2 4)	3,93 (2 2)	5,11 (3 4)	4,23 (3 2)	4,86 (1 4)	3,86 (1 2)

Tabella 15 – Seconda settimana completa del mese

Le due tabelle appena sopra sono dense di informazioni, riportano i risultati dei test mensili e settimanali il cui riferimento temporale esatto è scritto nella prima colonna. La coppia di numeri in parentesi rende immediato confrontare le performance, il numero colorato indica la bontà a parità di testing set (rosso RMSE e blu MAPE) quindi “per riga”, invece, il secondo li classifica a pari calibrazione, cioè “per colonne”. In breve: se vogliamo capire quale terna funziona meglio per prevedere un dato periodo ci riferiremo al numero colorato, invece, se volessimo usare un terna fissa, dal secondo numero capiamo in che periodo otteniamo le previsioni migliori.

Viene naturale pensare che si ottengano risultati migliori prevedendo un periodo compatibile con il set su cui di calibrazione, ma non accade sempre. Ad esempio, se l’interesse è prevedere il **mese** di febbraio la terna migliore non è in nessun caso quella da calibrazione invernale ma bensì l’estiva osservando il RMSE oppure la globale se diamo più peso al MAPE. la calibrazione invernale oltre ad essere la peggiore per prevedere se stessa lo è anche a prescindere dal periodo previsto, cioè, utilizzando quella terna, l’errore è sempre più alto di quello ottenuto con le altre due, quale che sia la stagione indagata o la metrica adottata. Invece, sempre a scala mensile, le terne estiva e globale sono più efficaci e a seconda delle stagioni previste (e metriche differenti) si contendono il titolo di miglior calibrazione senza la prevalenza di una specifica. Gli appunti riguardo i testing set mensili continuano a valere anche su scala **settimanale** con valori confrontabili come entità, non ci sono grosse differenze o cali di prestazioni

improvvisi anche se mediamo su circa un quarto dei dati, non si nota la presenza di valori anomali che pesano di più sulla media. Si sottolinea un fatto curioso: la terna invernale, che prima era la peggiore in ogni caso, ora eccelle se si applica per la previsione della settimana primaverile, per i restanti però rimane la peggiore confermando quanto scritto sopra. Valutando il RMSE la calibrazione globale produce le previsioni più accurate (tolta la primavera ovviamente). Il MAPE, invece, è più altalenante e non predilige un terna ad un'altra.

SET CALIBRAZIONE →	ESTATE		INVERNO		GLOBALE	
SET TEST ↓	RMSE	DMAPE[%]	RMSE	DMAPE[%]	RMSE	DMAPE[%]
INVERNO : 14/02	3,35 (1 2)	4,31 (1 2)	3,51 (3 2)	4,48 (2 2)	3,39 (2 2)	4,50 (3 2)
PRIMAVERA : 11/04	4,50 (2 3)	5,06 (3 3)	4,45 (1 3)	4,74 (2 3)	4,52 (3 3)	4,73 (1 3)
ESTATE : 11/07	1,48 (3 1)	1,71 (1 1)	1,45 (1 1)	1,85 (3 1)	1,46 (2 1)	1,79 (2 1)
AUTUNNO : 17/10	9,23 (2 4)	7,14 (2 4)	9,25 (3 4)	7,25 (3 4)	9,10 (1 4)	6,86 (1 4)

Tabella 16 – DMAPE , Mercoledì

SET CALIBRAZIONE →	ESTATE		INVERNO		GLOBALE	
SET TEST ↓	RMSE	DMAPE[%]	RMSE	DMAPE[%]	RMSE	DMAPE[%]
INVERNO : 18/02	2,88 (2 3)	4,33 (2 3)	3,20 (3 2)	5,13 (3 3)	2,71 (1 3)	3,99 (1 3)
PRIMAVERA : 15/04	4,57 (1 4)	7,31 (1 4)	4,89 (3 4)	7,96 (3 4)	4,67 (2 4)	7,61 (2 4)
ESTATE : 15/07	2,55 (1 2)	3,49 (1 2)	3,21 (2 3)	4,44 (3 2)	2,55 (1 1)	3,58 (2 2)
AUTUNNO : 21/10	2,47 (1 1)	3,08 (1 1)	3,14 (3 1)	3,79 (3 1)	2,56 (2 2)	3,14 (2 1)

Tabella 17 -DMAPE , Domenica

Ponendo l'attenzione a **scala giornaliera** l'errore è riferito al mercoledì come feriale e la domenica per il festivo, come visto nel capitolo 2 la domenica presenta un profilo di prezzo diverso dove il picco mattutino tende a sparire favorendo quello pomeridiano. Ci si aspetta che la domenica sia più difficile da prevedere ed particolarmente sbagliato se parliamo dell'autunno, dove, il mercoledì registra più del doppio dell'errore, invece durante l'estate è vero, il mercoledì estivo da errori minori della domenica a prescindere dalla calibrazione. Discutendo in linea generale, è chiaro che l'estate ha caratteristiche che la rendono la più facile da prevedere ovunque si calibrino i parametri, si capisce bene osservando i numeri in nero tra parentesi: domenica esclusa, la terza riga offre sempre l'errore minimo mentre l'autunno , inverno e primavera sono meno prestanti. La calibrazione invernale funziona peggio delle altre nel più dei casi e in poche situazioni eccelle, in Tabella 12 si vede che la terna invernale prevede vettori di 140 elementi contro i 170 delle altre due e a pagina 104 la scelta includeva proprio quei due casi, non avendo paragoni al momento la scelta si è mossa verso un buon minimo ma che non è affidabile, molto forte localmente ma mediocre altrove. Qui emerge una questione cardine del problema: ai fini della robustezza è indispensabile analizzare a fondo più calibrazioni per confermare la posizione del minimo, la situazione di pagina 104 suggerisce due minimi equivalenti ma impossibili da distinguere senza ulteriori informazioni.

5.3. Orizzonte di previsione a 24 ore

Si ripete l'analisi dei risultati studiando come incide il passaggio ad orizzonte giornaliero. Le aspettative sono praticamente certezze nel senso che senza dubbio l'errore aumenta con l'orizzonte, per il resto è plausibile che si confermi la struttura di mercato che vede la serie estiva come la più semplice da prevedere. Con buone probabilità il fatto che prevedendo ad un ora si eccelle d'estate implica che il motivo risieda nella natura dei dati più che nei parametri o processi coinvolti, quindi, indipendente dall'orizzonte di previsione.

SET CALIBRAZIONE →	ESTATE		INVERNO		GLOBALE	
SET TEST ↓	RMSE	MMAPE[%]	RMSE	MMAPE[%]	RMSE	MMAPE[%]
INVERNO : FEBBRAIO 18	11,29 (1 4)	10,09 (1 3)	12,43 (3 4)	11,31 (3 3)	11,98 (2 4)	10,32 (2 3)
PRIMAVERA : APRILE 18	9,16 (3 3)	17,54 (3 4)	8,78 (2 3)	17,22 (2 4)	8,48 (1 3)	16,18 (1 4)
ESTATE : LUGLIO 18	5,65 (2 1)	7,29 (2 2)	6,51 (3 1)	8,10 (3 2)	5,08 (1 1)	6,30 (1 1)
AUTUNNO : OTTOBRE 18	6,47 (1 2)	6,38 (1 1)	7,10 (3 2)	7,08 (3 1)	6,73 (2 2)	6,79 (2 2)

Tabella 18 – Mensile

SET CALIBRAZIONE →	ESTATE		INVERNO		GLOBALE	
SET TEST ↓	RMSE	WMAPE[%]	RMSE	WMAPE[%]	RMSE	WMAPE[%]
INVERNO : 12/02 – 18/02	5,94 (2 2)	8,52 (2 3)	6,17 (3 2)	8,95 (3 3)	5,67 (1 2)	7,59 (1 3)
PRIMAVERA : 9/04 – 15/04	8,21 (2 4)	12,21 (3 4)	7,62 (1 3)	10,46 (1 4)	8,46 (3 4)	11,41 (2 4)
ESTATE : 7/07 – 15/07	4,47 (2 1)	5,76 (2 1)	5,38 (3 1)	6,73 (2 1)	4,34 (1 1)	5,55 (1 1)
AUTUNNO : 15/10 – 21/10	7,77 (1 3)	6,69 (1 2)	8,79 (3 4)	8,20 (3 2)	8,22 (2 3)	7,30 (2 2)

Tabella 19 – Seconda settimana completa del mese

Il primo focus riguarda il test **mensile** di Tabella 18, la calibrazione invernale continua ad essere piuttosto misera ma a differenza dell'orizzonte unitario non è la peggiore: lo è la primavera. Fra tutte eccelle la terna globale se applicata a mesi primaverili o estivi mentre la calibrazione estiva è ottima per prevedere inverno e autunno, affermare che una terna funzioni meglio se applicata per l'inverno non vuol dire che l'errore invernale sia il più basso fra tutti i periodi di test ma solamente che le altre calibrazioni portano errori maggiori. A scala mensile rimane vero che i periodi più ostici sono inverno e primavera quale che sia la calibrazione adottata, estate ed autunno sono più semplici invece, con prevalenza della prima ma leggero disaccordo fra metriche.

Ora ci si sofferma sulla **Tabella 19** e subito si apprezza qualche differenza, qui si hanno difficoltà nel prevedere in primavera, poi inverno e autunno ma l'estate è sempre il periodo in cui eccellono tutte e tre le calibrazioni a prescindere dalla metrica. Inverno ed estate vengono riprodotti meglio con la calibrazione estiva, l'autunno ha minimo con la calibrazione estiva e sorprendentemente la primavera

con quella invernale. Un commento a riguardo la primavera: la previsione migliore deriva dalla terna invernale che al momento sembra essere la calibrazione meno affidabile, è strano che le altre calibrazioni non riescano a fare di meglio.

SET CALIBRAZIONE →	ESTATE		INVERNO		GLOBALE	
SET TEST ↓	RMSE	DMAPE[%]	RMSE	DMAPE[%]	RMSE	DMAPE[%]
INVERNO : 14/02	3,77 (3 2)	6,16 (3 2)	3,64 (2 2)	5,9 (2 2)	3,28 (1 2)	5,33 (1 2)
PRIMAVERA : 11/04	11,03 (3 3)	15,0 (3 4)	8,33 (2 3)	10,57 (2 4)	8,06 (1 3)	9,63 (1 4)
ESTATE : 11/07	3,42 (3 1)	4,07 (3 1)	2,34 (1 1)	3,0 (1 1)	2,71 (2 1)	3,40 (2 1)
AUTUNNO : 17/10	14,9 (3 4)	9,37 (3 3)	13,9 (2 4)	9,23 (1 3)	13,84 (1 4)	9,33 (2 3)

Tabella 20 – DMAPE , Mercoledì

SET CALIBRAZIONE →	ESTATE		INVERNO		GLOBALE	
SET TEST ↓	RMSE	DMAPE[%]	RMSE	DMAPE[%]	RMSE	DMAPE[%]
INVERNO : 18/02	5,25 (2 1)	6,52 (2 1)	6,67 (3 2)	11,44 (3 3)	4,70 (1 3)	6,02 (1 3)
PRIMAVERA : 15/04	10,28 (3 4)	18,24 (2 4)	8,03 (1 4)	14,44 (1 4)	9,78 (2 4)	19,85 (3 4)
ESTATE : 15/07	6,58 (3 3)	11,22 (3 3)	6,36 (2 1)	10,71 (2 2)	2,89 (1 1)	4,26 (1 1)
AUTUNNO : 21/10	5,93 (2 2)	8,02 (2 2)	6,42 (3 2)	8,70 (3 1)	4,44 (1 2)	5,85 (1 2)

Tabella 21 -DMAPE , Domenica

Adesso si ragiona ad scala giornaliera, per quanto riguarda il **feriale** la calibrazione estiva è sempre la peggiore delle tre e, tolto in autunno, in generale la migliore è quella globale. Ancora una volta l'estate si conferma essere la stagione più prevedibile seguita dall'inverno, primavera e autunno (il MAPE inverte le ultime due). Dal punto di vista della **domenica** la calibrazione globale è di nuovo la più efficace in accordo con il mercoledì, le due che restano sono confuse a tal punto da non distinguerne la peggiore, questo è l'unico caso in cui prevedere l'estate con la stessa calibrazione non dà i risultati migliori benché accada per le altre due terne.

Qualche riga va dedicata per confrontare i risultati mensili e settimanali a diverso orizzonte: quindi fra Tabella 14 / Tabella 18 e Tabella 15 / Tabella 19. E' oramai chiaro che l'estate è più predisposta ad essere predetta a prescindere da altri fattori, invece le caratteristiche invernali del prezzo lo rendono più imprevedibile a scala mensile mentre per primavera e autunno la situazione è meno intuitiva, entrano in gioco le due metriche e per ogni sfumatura si traggono conclusioni diverse. Ad orizzonte unitario la calibrazione invernale è da scartare sia mensilmente che settimanalmente e le altre due si equivalgono in buona misura, per la previsione a 24 ore la situazione è simile per cui la calibrazione più efficace è quella globale.

5.4. Riscontro grafico

E' interessante vedere i profili reale e previsto sullo stesso diagramma con relativo scatter plot, per concentrarsi sulla previsione con applicazioni più concrete si tralasciano i risultati ad orizzonte unitario e ci si limita alla calibrazione globale. Detto ciò, la previsione migliore ha luogo il mercoledì estivo, quella peggiore la domenica primaverile (con il MAPE) o il mercoledì autunnale.

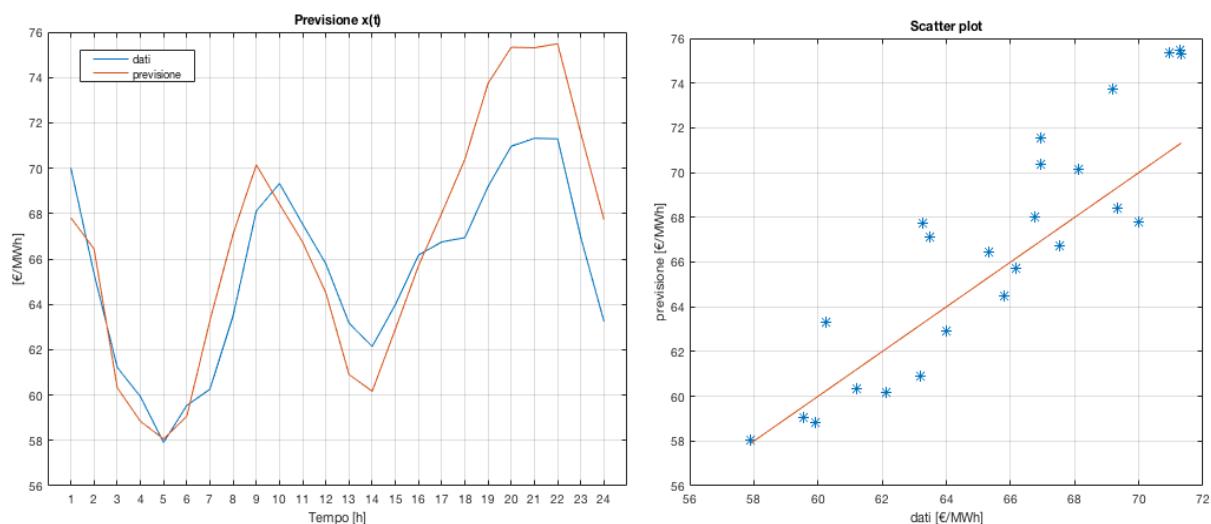


Figura 79 - Calibrazione globale $T=24$, mercoledì estivo, $RMSE = 2,71 \text{ €/MWh}$, $MAPE = 3,4\%$

La Figura 79 è molto accurata secondo entrambe le metriche (RMSE e MAPE), scarti contenuti implicano un basso RMSE e la buona riproduzione dei picchi comporta un MAPE eccellente. Poco meno della metà dei valori è sottostimato e sono tutti vicini alla bisettrice, lo scarto massimo è circa 5 €/MWh e altrove si rimane entro 1 o 2 €/MWh.

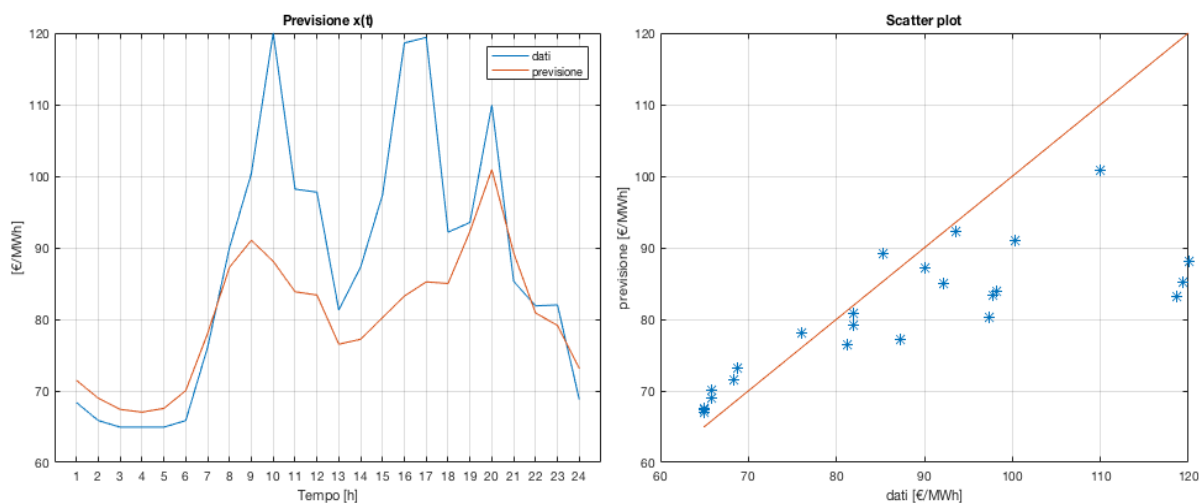


Figura 80 - Calibrazione globale $T=24$, mercoledì autunnale, $RMSE = 13,84 \text{ €/MWh}$

La previsione in Figura 80 penalizza il RMSE perché la maggior parte dei dati è parecchio sottostimata, si tratta di 30 – 40 €/MWh in deficit per valori di picco ed è proprio per questo che il MAPE non gli attribuisce il giusto peso. A parte i primi due picchi la previsione è precisa ma con una variabilità così forte probabilmente nessun modello sarebbe in grado di seguire bene i picchi.

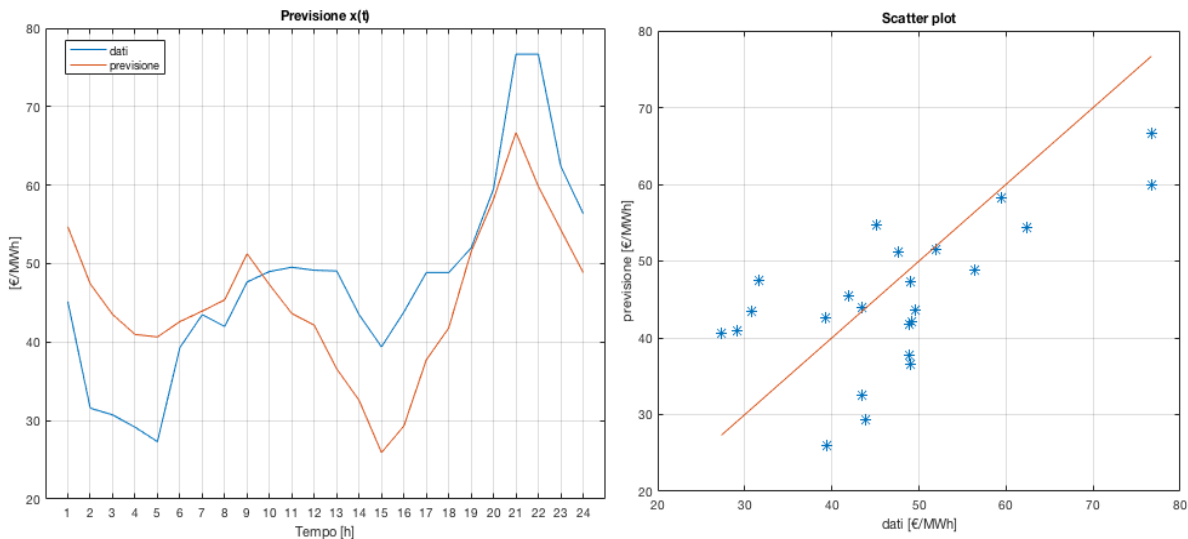


Figura 81 - Calibrazione globale $T=24$, domenica primaverile, $MAPE = 19,85 \text{ €/MWh}$

L'ultima figura rappresenta un caso meno drammatico di quanto si pensasse, il problema sta nei valori attorno alle h.15 che registrano comunque 10 – 15 €/MWh di scarto in negativo sebbene non sia un valore alto, così come le prime ore della giornata mentre lo scarto sul picco delle 21 conta poco oramai. Come si può fare per capire dove dare peso ad una metrica piuttosto che all'altra? Per quanto visto, i profili "comuni" come i primi due sono tanto più prevedibili quanto meno variabili sono i picchi, la situazione è sempre la stessa ed in quel caso è importante fornire il RMSE per capire quanto si lontani dai valori di picco mentre al di fuori la previsione sarà affidabile. Per profili anomali come la domenica o peggio feriali infrasettimanali non possiamo aspettarci che il codice si adatti perfettamente a qualcosa che non si aspetta e in quel caso un'analista esperto saprà come muoversi, da questo punto di vista il MAPE è una metrica più adatta per identificare tali anomalie sul trend. Quindi, un'azienda che pianifica le offerte in base alla previsione deve capire dove sono i picchi nelle giornate "standard" e sapere come correggerli mentre per giornate anomale è tutto più complesso perché l'intero profilo potrebbe essere incerto.

5.5. Confronto con lo stato dell'arte

Nel primo capitolo è stata descritta la situazione attuale in campo EPF (Electricity price forecasting), interessa qui collocare questo lavoro all'interno dello stato dell'arte. Si considerano solo le pubblicazioni in ambito accademico perché nel commerciale le aziende non hanno motivo di pubblicare dati in merito alla precisione che raggiungono. Le **Figura 82 e Figura 83** schematizzano il confronto reso possibile dalla revisione di Aggarwal et al. pubblicata nel 2009 che compara pubblicazioni che utilizzano modelli statistici (Figura 82) e reti neurali. Dal momento che questo metodo rientra nella prima categoria costituisce un confronto efficace. Le suddette figure indicano i range di errore per ogni pubblicazione tramite segmenti verticali e le **linee orizzontali rappresentano i risultati del capitolo 5, le linee a tratto interrotto sono gli estremi di massimo e minimo, invece quella continua rappresenta l'errore medio**. I colori rappresentano la scala temporale: le curve rosse si riferiscono al MAPE giornaliero, le blu al settimanale e le nere al mensile. Per i valori esatti si rimanda alle Tabella 3 e Tabella 4 più tutte quelle della sezione 5.2 del capitolo 5.

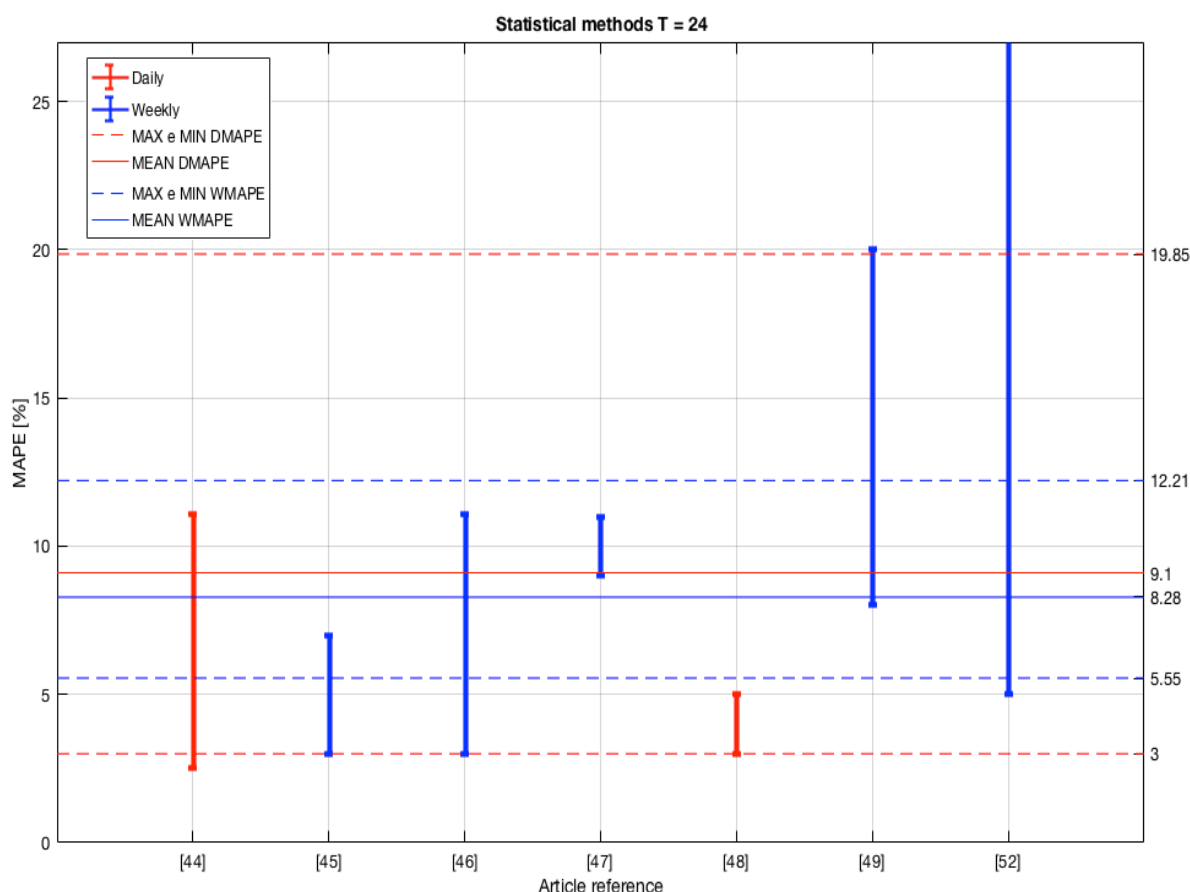


Figura 82 – Confronto con lo stato dell'arte [3] : Metodi statistici

In [3] l'autore censisce sette pubblicazioni che usano metodi statistici e metrica compatibile (MAPE), due con errore giornaliero. I MAPE giornalieri ottenuti in questa tesi variano da 3 a 19,85% che è ampio rispetto a quelli ottenuti da [44] e [48] ma è in linea con il limite inferiore, a scala giornaliera si hanno previsioni meno affidabili che però nel migliore dei casi raggiungono la stessa precisione dei risultati pubblicati. Ora un cenno ai risultati settimanali dei metodi statistici, la Figura 82 mostra che i risultati di

questa tesi sono più compatti perché variano da 5,55 a 12,21%, [45] e [46] riescono a scendere sotto il 5% e rimanere sotto il nostro massimo invece [49] e [52] lo superano di parecchio. La previsione settimanale di [45] è migliore e più affidabile ma in termini di affidabilità [47] non ha rivali con MAPE stabile circa 10%, [49] e [52] arrivano al 25% con un minimo confrontabile a questa tesi il che la rende più affidabile a scala settimanale che giornaliera.

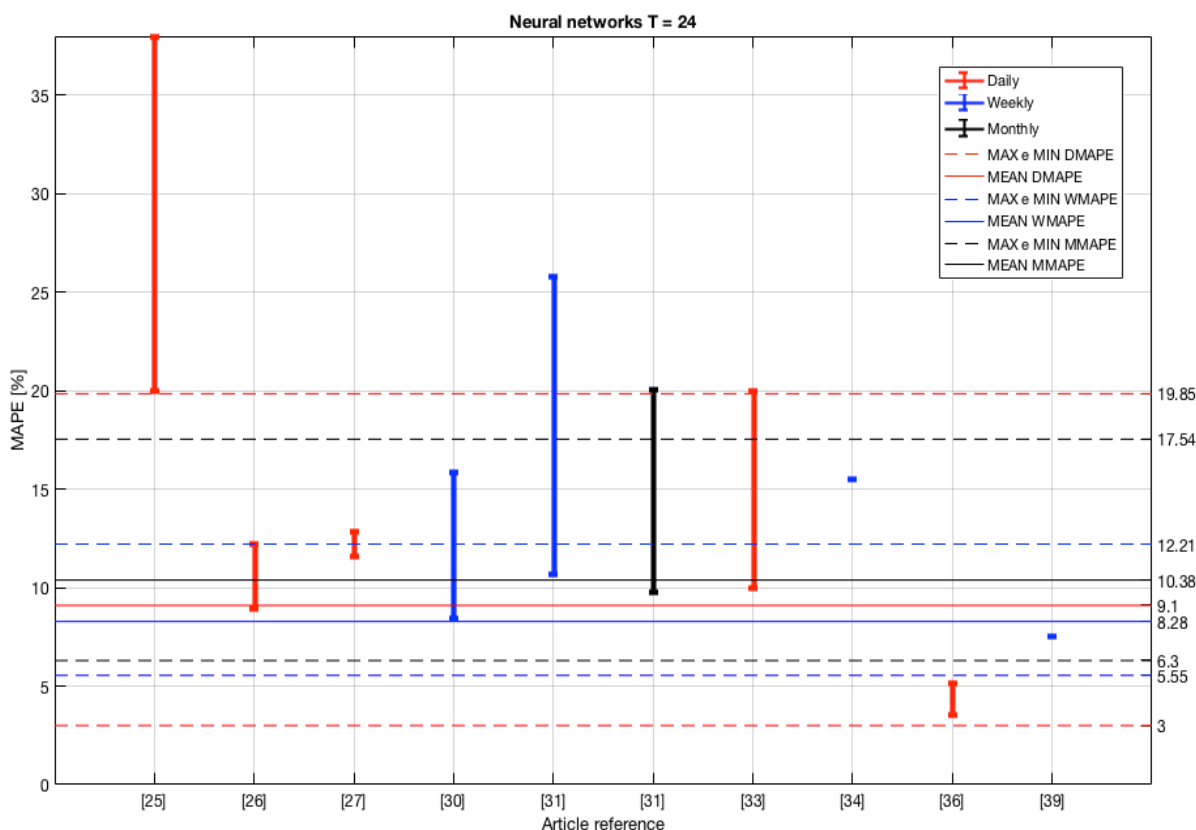


Figura 83 - Confronto con lo stato dell'arte [3] : Reti neurali

La Figura 83 compara i nostri risultati a pubblicazioni che riguardano le reti neurali, aiuta a inserire il modello in un contesto diverso per studiare una nuova classe di modelli e capire se le reti neurali sono categoricamente migliori di quello applicato qui. Al 2009 nessuna rete neurale analizzata da [3] riesce a fare di meglio della previsione non lineare utilizzata da questa analisi, solo [36] si avvicina e [39] la segue, tolti [25] e [31] nessun'altro sfiora oltre il massimo di 19,85%.

Le Tabella 1 e Tabella 2 riassumono il confronto con [21], riportano i risultati di un modello che considera l'effetto della temperatura sul mercato elettrico Italiano, purtroppo non si è sicuri di come l'autrice abbia calcolato il RMSPE perché i valori di tabella 2 sono veramente troppo bassi, si è comunque calcolato il RMSPE come indicato da Weron (sezione 3 di [1]) nella speranza di ottenere chiarimenti dall'autrice che però non hanno avuto risposta, i valori di RMPSE ottenuti da questa tesi sono compresi tra il 2,2% e il 40% ma il più dei test è stabile attorno al 5 - 10%. Si aggiunge un appunto per confrontare i risultati con [20], articolo del 2017 che implementa un innovativo metodo ibrido, i risultati sono fra i migliori in assoluto (articolo [20], tabelle 7 e 10), il DMAPE scende anche sotto al 2% con massimo a 12% ma la gran parte dei casi rimane entro il 6%, ottimo anche il RMSE.

Capitolo 6 : CONCLUSIONI

Questa tesi si pone l'obiettivo di investigare l'applicabilità del metodo di previsione non lineare alle complesse caratteristiche del mercato Italiano, si è studiato il meccanismo di formazione del prezzo analizzando i profili orari di più mercati, poi ne è seguita la previsione con la calibrazione dei parametri che sono stati utilizzati in seguito per quantificare le prestazioni del codice. Bisogna dire che la mancanza di riferimenti applicati al mercato italiano non permette di confrontare effettivamente le performance del metodo utilizzato con altri applicati nelle stesse condizioni, ogni mercato ha una sua struttura intrinseca che lo differenzia da tutti gli altri e per questo non è un confronto perfetto. Comunque la problematica rimane molto attuale e la ricerca propone miglioramenti, accortezze e punti deboli che lo rendono un campo promettente. E' molto importante proseguire in questa direzione perché è un problema che vede applicazioni immediate nel commerciale ed in secondo luogo promette miglioramenti sostanziali muovendosi correttamente. Il problema include talmente tante variabili che oltre un certo livello di dettaglio non è possibile ragionare i termini solo deterministici ma bisogna tenere conto di fattori che coinvolgono la natura caotica e stocastica delle relazioni fra soggetti umani come questioni geopolitiche o logistiche.

La calibrazione è un punto cardine del problema, essa dipende dal periodo perché a seconda della stagione i profili di prezzo hanno caratteristiche molto diverse, si è analizzata l'influenza di questo fenomeno calibrando su più periodi e la calibrazione più affidabile è quella globale perché include le caratteristiche proprie di estate e inverno pesandone i contributi, questa non sarà sempre la migliore ma fornisce il risultato più robusto. Molte circostanze riportano al stesso discorso, basta veramente poco per ricadere in minimi molto buoni ma non eccellenti, perciò un'analisi approfondita è necessaria per cogliere questi dettagli altrimenti non resta che affidarsi all'esperienza. Nei primi tentativi raramente la soluzione migliore coincide il minimo più robusto perché spesso sono minimi locali che lavorano bene in poche circostanze portando al problema di over-fitting, cioè una terna che eccelle solo durante la calibrazione. Ne è un esempio lampante la previsione a ventiquattr'ore, calibrando in estate otteniamo vettori corti con forte ritardo e molti vicini mentre, invece in inverno servono vettori lunghi il doppio a ritardo unitario con la metà dei vicini. Il punto in comune a tutte le calibrazioni è il che esiste un livello di informazione ottimo. La serie del PUN italiano raggiunge massima autocorrelazione su scala settimanale (vedi pagina 84) che corrisponde all'ottimo di informazione di cui il modello ha bisogno per produrre la migliore previsione. Per l'orizzonte unitario i vettori utilizzano dati successivi dall'ultimo valore fino poco più di sette giorni a ritroso, lo stesso per la previsione a 24h che ne richiede leggermente meno. L'autocorrelazione è un'informazione molto preziosa, poco prima o poco dopo il massimo le prestazioni migliorano e non accade solo a scala settimanale ma anche giornaliera sebbene il minimo assoluto abbia sempre a che fare con la settimana. L'unico caso che sembra discorde è la calibrazione estiva con orizzonte a 24 h, in quella situazione i vettori includono più di un mese per contribuire ad un minimo solido ma locale, bene, non un mese esatto ma una settimana in più, ecco che continua a pesare l'autocorrelazione settimanale. Nel caso in cui l'autocorrelazione dovesse assumere valori negativi si perderebbe informazione, questo non accade per nessuna delle calibrazioni raggiunge valori negativi.

Riepilogando i risultati, indipendentemente dalla terna ottima e su dalla scala temporale è sempre più facile compiere buone previsioni in estate piuttosto che in primavera, le altre stagioni sono danno risultati intermedi: ne troppo difficili ma neppure le migliori, anche se le situazioni sono molto variegata. Bisogna ricordare che il test estivo cade a luglio e certe giornate saranno comunque molto difficili da prevedere, comunque un profilo di prezzo estivo è meno variabile rispetto ad uno di un'altra stagione e questo pesa sulla previsione. Il centro del problema è la variabilità del profilo da predire, uno dei talloni d'Achille del settore è la previsione dei picchi che abbate le prestazioni nella maggioranza dei casi, con una discreta mole di dati è semplice prevedere la posizione dei picchi in una giorno feriali ma è complicato risalire al valore esatto. Invece, le giornate festive sono anomale, l'unico festivo con cadenza ciclica è la domenica che rimane difficile da prevedere anche con uno storico notevole, la situazione si complica con i festivi infrasettimanali.

E' interessante capire come si colloca questa tesi nella la ricerca scientifica, si sono concluse considerazioni di carattere quantitativo riguardo le prestazioni dei metodi che rientrano nella stessa categoria (metodi statistici) e un confronto con le reti neurali. Nel complesso i risultati ottenuti in questa tesi sono in linea con quelli resi noti in letteratura, sviluppi futuri includono ottime possibilità di miglioramento per affrontare il problema che non sta nella precisione ma nell'affidabilità: su scala giornaliera si può fare molto bene ed il passo successivo è cercare di abbassare il tetto massimo di errore, per la settimanale invece c'è chi è riuscito a fare di meglio. La sezione 4 del secondo capitolo è dedicata alle variabili endogene che contribuiscono alla formazione dei prezzi, in [21] [3] [17] [79] e quasi tutte le pubblicazioni fanno chiari riferimenti a come inserirle nell'analisi e la maggior parte delle pubblicazioni citate ne utilizzano almeno una. Per cui, ci si aspetta che attraverso uno studio meticoloso si arrivi a comprendere nel dettaglio la dinamica del mercato identificando le variabili a maggior impatto. La letteratura suggerisce moltissime alternative (si veda tabella 1 in [3]) che vanno dalla struttura di mercato al prezzo delle materie prime oppure variabili meteorologiche legate alla produzione da FER e non ultima la temperatura, che determina il fabbisogno da raffrescamento estivo. Il primo problema è capire da quali si può trarre maggior vantaggio e qui entra in gioco l'aspetto umano, il secondo è reperire i dati.

Come ogni metodo oltre i punti di forza la previsione non lineare presenta delle lacune, vi sono aspetti poco affrontati come le previsioni spot-price di valori di picco oppure indici di prestazione innovativi che tengano in conto molteplici variabili non strettamente quantificabili. Una grossa debolezza sta nella calibrazione, non esistono metodi immediati per indirizzare la ricerca verso la zona di minimo o perlomeno non applicabili se non in casi specifici in cui la teoria lo permette, di norma si spende moltissimo tempo per calibrare che si traduce in previsioni ad orizzonte maggiore quindi meno performance. A parte qualche test di diversa natura le possibilità più concrete in questo verso stanno nell'abilità dell'analista e il suo livello di esperienza nel settore.

Il lavoro svolto si è confermato molto interessante e soprattutto utile ai fini di ricerca e studio personale, i risultati sono decisamente interessanti e il confronto svolto nelle pagine precedenti lascia sperare per buoni miglioramenti, si conclude lasciando spazio a chiunque voglia cimentarsi in questo campo poco esplorato che promette stimolanti approfondimenti e un buon livello di complicazione includendo aspetti che considerano più campi in parallelo, non solamente l'aspetto matematico - scientifico fine a se stesso.

- [1] R. Weron, «Electricity price forecasting: A review of state-of-the-art with a look into the future,» *International journal of forecasting*, pp. 1030-1081, 2014.
- [2] M. Ventosa, Á. Baíllo, A. Ramos e M. Rivier, «Electricity market modeling trends,» *Energy policy*, pp. 897 - 913, 2005.
- [3] S. K. Aggarwal, L. M. Saini e A. Kumar, «Electricity price forecasting in deregulated markets: A review and evaluation,» *Electrical Power and Energy Systems*, pp. 13-22, 2009.
- [4] Batle, «A model for electricity generation risk analysis,» *Phd. Thesis*, 2002.
- [5] C. Batle e J. Barquin, «A strategic production costing model for electricity market price analysis,» 2005.
- [6] C. Day, B. Hobbs e J.-S. Pang, «Oligopolistic competition in power networks: a conjectured supply function approach,» *IEEE Transactions on Power Systems*, vol. 17, n. 3, p. 597-607, 2002.
- [7] M. Burger, B. Graeber e G. Schindlmayr, *Managing Energy Risk*, Wiley, 2007.
- [8] M. Barlow, «A diffusion model for electricity prices,» *Mathematical finance*, vol. 12, n. 4, pp. 287-298, 2007.
- [9] R. Weron, *Modeling and Forecasting Electricity Loads and Prices: A Statistical Approach*, Wiley, 2006.
- [10] G. E. P. Box e G. M. Jenkins, *Time series analysis*, Holden-Day, 1976.
- [11] G. E. P. Box, G. M. Jenkins e G. C. Reinsel, *Time series analysis : Forecasting and control*, 3rd ed., Englewood Cliffs: Prentice Hall, 1994.
- [12] H. Tong e K. S. Lim, «Threshold autoregression, limit cycles and cyclical data,» *Journal of the Royal Statistical Society*, vol. Series B, n. 42, p. 245-292, 1980.
- [13] R. Engle, «Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation,» *Econometrica*, vol. 50, n. 4, pp. 987-1007, 1982.
- [14] T. Bollerslev, «Generalized autoregressive conditional heteroskedasticity,» *Journal of Econometrics*, vol. 31, n. 3, pp. 307-327, 1986.
- [15] Vapnik, *The nature of statistical learning theory*, Springer, 1995.
- [16] D. C. Sansom, T. Downs e T. K. Saha, «Evaluation of support vector machine based forecasting tool in electricity price forecasting for Australian national electricity market participants,» *Journal of Electrical and Electronics Engineering*, vol. 22, n. 3, pp. 227-233, 2002.
- [17] J. Nowotarski e R. Weron, «Recent advances in electricity price forecasting: A review of probabilistic forecasting,» *Renewable and Sustainable Energy Reviews*, n. 81, pp. 1548 - 1568, 2018.
- [18] T. Hong, F. S. Z. Pinson, A. Troccoli e J. R. Hyndman, «Probabilistic energy forecasting: Global Energy Forecasting Competition 2014 and beyond,» *International Journal of Forecasting*, n. 32, pp. 896-913, 2016.
- [19] J. M. Bates e C. W. Granger, «The combination of forecasts,» *Operation Research Quarterly*, vol. 20, n. 4, pp. 451-468, 1969.
- [20] Z. Yang, L. Ce e L. Lian, «Electricity price forecasting by a hybrid model, combining wavelet transform, ARMA and kernel-based extreme learning machine methods,» *Applied Energy*, vol. 190, pp. 291-305, 2017.
- [21] S. Bigerna, «Estimating temperature effects on the Italian electricity market,» *Energy policy*, vol. 118, pp. 257-269, 2018.
- [22] J. Breuer, R. McNown e M. Wallace, «Misleading inferences from panel unit-root test with an illustration from purchasing power parity,» *Review of International Economics*, pp. 482-493, 2001.
- [23] G. Kapetanios, Y. Shin e A. Snell, «Testing for a unit root in the nonlinear STAR framework,» *Journal of Econometrics*, vol. 112, n. 2, pp. 359-379, 2003.
- [24] B. Szkuta, L. Sanabria e T. Dillon, «Electricity price short-term forecasting using artificial neural networks,» *IEEE Transactions on Power Systems*, vol. 14, n. 3, pp. 851 - 857, August 1999.
- [25] C. Rodriguez e G. Anders, «Energy price forecasting in the Ontario competitive power system market,» *IEEE Transactions on Power Systems*, vol. 19, n. 1, pp. 366 - 374, 19 February 2004.
- [26] A. Wang e B. Ramsay, «A neural network based estimator for electricity spot pricing with particular reference to weekend and public holidays,» *Neurocomputing*, vol. 23, pp. 47-57, 1998.
- [27] A. Wang e B. Ramsay, «Prediction of system marginal price in the UK power pool using neural networks,» *Proc IEEE Int Conf Neural Networks*, vol. 4, p. 2116-2120, 1997.
- [28] F. Gao, X. Cao e A. Papalexopoulos, «Forecasting power market clearing price and quantity using a neural network method,» *IEEE Power Eng Soc Summer Meet*, vol. 4, pp. 2183-2188, 2000.
- [29] R. Gareta, L. Romeo e G. A., «Forecasting of electricity prices with neural networks,» *Energy Convers Manage*, vol. 47, pp. 1770-1778, 2006.

- [30] P. Georgilakis, «Market clearing price forecasting in deregulated electricity markets using adaptively trained neural networks,» *SETN 2006. Lecture notes in artificial intelligence LNAI 3955*, p. 56–66, 2006.
- [31] P. Mandal, T. Senjyu e T. Funabashi, «Neural networks approach to forecast several hour ahead electricity prices and loads in deregulated market,» *Energy Convers Manage*, vol. 47, p. 2128–2142, 2006.
- [32] H. Yamin, S. Shahidehpour e Z. Li, «Adaptive short-term electricity price forecasting using artificial neural networks in the restructured power markets,» *Electrical Power Energy System*, vol. 26, p. 571–581, 2004.
- [33] Z. Hu, Y. Yu, Z. Wang, W. Sun, D. Gan e Z. Han, «Price forecasting using integrated approach,» *Proceedings of IEEE international conference on electric utility deregulation, restructuring and power technologies*, p. 28–31, April 2004.
- [34] Z. Hu, L. Yang, Z. Wang, D. Gan, W. Sun e K. Wang, «game theoretic model for electricity markets with tight capacity constraints,» *Electric Power Energy System*, vol. 30, n. 3, pp. 207–215, 2008.
- [35] A. T. Lora, J. R. Santos, J. L. M. Ramos e A. G. Exposito, «Electricity market price forecasting: neural networks versus weighted-distance k nearest neighbors,» *Lecture notes in computer science LNCS 2453*, p. 321–330, 2002.
- [36] C. Li e S. Wang, Next-day power market clearing price forecasting using artificial fish-swarm based neural networks, Berlin: Heidelberg: Springer-Verlag, 2006, p. 1290–1295.
- [37] A. Gonzalez, A. San Roque e J. Garcia-Gonzalez, «Modeling and forecasting electricity prices with input/output hidden Markov models,» *IEEE Trans Power System*, vol. 20, n. 1, p. 13–24, 2005.
- [38] J.-K. Lee, J.-B. Park, J.-R. Shin e K. Lee, «A system marginal price forecasting based on an artificial neural network adapted with rough set theory,» vol. 1, pp. 528–533, 2005.
- [39] N. Amjady, «Day ahead price forecasting of electricity markets by a new fuzzy neural network,» *IEEE Trans Power System*, vol. 21, n. 2, pp. 887–896, 2006.
- [40] W. Wu, J.-Z. Zhou, J. Yu, C.-J. Zhu e J.-J. Yang, «Prediction of spot market prices of electricity using chaotic time series,» *Proceedings of the 3rd international conference on machine learning and cybernetics*, pp. 888–893, 2004.
- [41] H. Yang e X. Duan, «Chaotic characteristics of electricity price and its forecasting model,» *Proceedings of the IEEE Canadian conference on electrical and computer engineering*, pp. 659–662, 2003.
- [42] S. Yao, Y. Song, L. Zhang e X. Cheng, «Prediction of system marginal price by wavelet transform and neural network,» *Electric Mach Power System*, vol. 28, pp. 537–549, 2000.
- [43] T. Niimura, H.-S. Ko e K. Ozawa, «A day-ahead electricity price prediction based on a fuzzy-neuaro autoregressive model in a deregulated electricity market,» *Proceedings of the 2002 international joint conference on neural networks*, vol. 2, pp. 1362–1366, 2002.
- [44] K. Chang-il, Y. In-Keun e S. Y.H., «Prediction of system marginal price of electricity using wavelet transform analysis,» *Energy Conversion and Management*, vol. 43, pp. 1839–1851, 2002.
- [45] J. Cuaresma, J. Hlouskova, S. Kossmeier e M. Obersteiner, «Forecasting electricity spot-prices using linear univariate time-series models,» *Applied Energy*, vol. 77, p. 87–106, 2004.
- [46] R. Weron e A. Misiorek, «Forecasting spot electricity prices with time series models,» in *Proceedings of international conference on the European electricity market EEM*, 2005.
- [47] R. Garcia, J. Contreras, M. Akkeren e J. van Garcia, «A GARCH forecasting model to predict day-ahead electricity prices,» *EEE Transaction on Power System*, vol. 20, n. 2, pp. 867–874, 2005.
- [48] F. Nogales, J. Contreras, A. Conejo e R. Espinola, «Forecasting next-day electricity prices by time series models,» *IEEE Transaction on Power Syst*, vol. 17, n. 2, pp. 342–348, 2002.
- [49] J. Contreras, R. Espinola, F. Nogales e A. Conejo, «ARIMA models to predict next- day electricity prices,» *EEE Transaction on Power System*, vol. 18, n. 3, pp. 1014–1020, 2003.
- [50] M. Zhou, Z. Yan, Y. Ni e G. Li, «An ARIMA approach to forecasting electricity price with accuracy improvement by predicted errors,» *Proceedings of IEEE power engineering society general meeting*, pp. 233–238, 6–10 June 2004.
- [51] M. Zhou, Z. Yan, Y. Ni, G. Li e Y. Nie, «Electricity price forecasting with confidence- interval estimation through an extended ARIMA approach,» *IEE Proceedings - Generation, Transmission and Distribution*, vol. 153, n. 2, pp. 187–195, 2006.
- [52] A. Conejo, M. Plazas, R. Espinola e A. Molina, «Day-ahead electricity price forecasting using the wavelet transform and ARIMA models,» *IEEE Transaction on Power System*, vol. 20, n. 2, pp. 87–106, 2005.
- [53] X. Haiteng e T. Niimua, «Short-term electricity price modeling and forecasting using wavelets and multivariate time series,» *IEEE power systems conference and exposition PES*, pp. 208–212.
- [54] A. Kian e A. Keyhani, «Stochastic price modeling of electricity in deregulated energy markets,» *Proceedings of the 34th Hawaii international conference on system sciences*, pp. 1–7, 2001.
- [55] S. Vucetic, K. Tomsovic e Z. Obradovic, «Discovering price – load relationships in California’s electricity market,» *IEEE Transaction Power System*, vol. 16, n. 2, pp. 280–286, 2001.

- [56] P. Mandal, T. Senjyu, N. Urasaki, T. Funabashi e A. K. Srivastava, «A Novel Approach to Forecast Electricity Price for PJM Using Neural Network and Similar Days Method,» *IEEE Transactions on Power Systems*, vol. 22, n. 4, pp. 2058 - 2065, 29 October 2007.
- [57] N. Pindoriya, S. Singh e S. Singh, «An Adaptive Wavelet Neural Network-Based Energy Price Forecasting in Electricity Markets,» *IEEE Transactions on Power Systems*, vol. 23, n. 3, pp. 1423 - 1432, 25 April 2008.
- [58] N. A. Shrivastava e P. B. Ketan, «A hybrid wavelet-ELM based short term price forecasting for electricity markets,» *International Journal of Electrical Power & Energy Systems*, vol. 55, pp. 41-50, February 2014.
- [59] Gestore del mercato elettrico S.p.A, «Vademecum della borsa elettrica,» Settembre 2012. [Online]. Available: <http://www.mercatoelettrico.org/It/MenuBiblioteca/documenti/20091028VademecumBorsaElettrica.pdf>. [Consultato il giorno 2018 Maggio 31].
- [60] Gestore dei Mercati Energetici S.p.A., «Mercato elettrico,» Novembre 2017. [Online]. Available: http://www.mercatoelettrico.org/It/MenuBiblioteca/Documenti/20160705Vademecum_PCE.pdf. [Consultato il giorno 11 Ottobre 2018].
- [62] ISTAT, «1. Ambiente e territorio,» in *Annuario statistico italiano 2010*, 2010.
- [63] ISTAT, «1. Ambiente e territorio,» in *Annuario statistico italiano 2011*, 2011.
- [64] ISTAT, «1. Ambiente e territorio,» in *Annuario statistico italiano 2012*, 2012.
- [65] ISTAT, «1. Ambiente e territorio,» in *Annuario statistico italiano 2013*, 2013.
- [66] ISTAT, «2. Ambiente ed energia,» in *Annuario statistico italiano 2014*, 2014.
- [67] ISTAT, «2. Ambiente ed energia,» in *Annuario statistico italiano 2015*, 2015.
- [68] ISTAT, «2. Ambiente ed energia,» in *Annuario statistico italiano 2016*, 2016.
- [69] ISTAT, «2. Ambiente ed energia,» in *Annuario statistico Italiano 2017*, 2017.
- [70] ISTAT, «Biblioteca digitale,» [Online]. Available: <https://ebiblio.istat.it/SebinaOpac/.do?idDoc=0010812&tabDoc=taboggd>. [Consultato il giorno Settembre 2018].
- [71] Gestore Servizi Energetici, «Dati e scenari,» [Online]. Available: https://www.gse.it/Dati-e-Scenari_site/statistiche_site. [Consultato il giorno 20 Agosto 2018].
- [72] EIA : Energy Information Administration, [Online]. Available: https://www.eia.gov/dnav/pet/pet_pri_spt_s1_a.htm. [Consultato il giorno 22 Agosto 2018].
- [73] Investing.com, [Online]. Available: <https://it.investing.com>. [Consultato il giorno 22 Agosto 2018].
- [74] D. Farmer e J. J. Sidorowich, «Predicting Chaotic Time Series,» *Physical review letters*, vol. 59, n. 8, pp. 845-848, 1987.
- [75] S. Tamea, F. Laio, A. Porporato e L. Ridolfi, *Approccio probabilistico alla previsione non lineare dei deflussi fluviali*.
- [76] F. Takens, «Detecting strange attractors in turbulence,» *Lectures Notes in Mathematics*, n. 898, pp. 366-381, 1981.
- [77] J. Bentley e J. Friedman, «data structure for range searching,» *ACM Computing Surveys*, vol. IV, n. 11, pp. 397-409, 1979.
- [78] A. Porporato e L. Ridolfi, «Nonlinear analysis of river flow time sequence,» *Water Resour. Res.*, vol. 33, n. 6, pp. 1353-1367, 1997.
- [79] J. Nowotarski e R. Weron, «On the importance of the long-term seasonal component in day-ahead electricity price forecasting,» *Energy economics*, vol. 57, pp. 228-235, 2016.
- [80] D. Le Comte e H. Warren, *Journal of applied meteorology*, vol. 20, pp. 1415-1419, 1981.
- [81] H. Cui e X. Song, «Research on Electricity Price Forecasting Based on Chaos Theory,» in *International Seminar on Future Information Technology and Management Engineering*, Baoding, 2008.
- [82] C. Unsuhay-Vila, A. Zambroni de Souza, J. Marangon-Lima e P. Balestrassi, «Electricity demand and spot price forecasting using evolutionary computation combined with chaotic nonlinear dynamic model,» *Electrical Power and Energy Systems*, vol. 32, pp. 108-116, 2010.
- [83] Y. Wang, C. Li and L. Zhang, *Short-term data forecasting based on wavelet transformation and chaos theory*, Beijing: IOP Publishing, 2017.
- [84] Istituto Superiore per la Protezione e la Ricerca Ambientale, «Produzione termoelettrica ed emissioni di CO2. Fonti rinnovabili e impianti soggetti a ETS,» 2011. [Online]. Available: <http://www.isprambiente.gov.it/it/pubblicazioni/rapporti/produzione-termoelettrica-ed-emissioni-di-co2>. [Consultato il giorno 4 Novembre 2018].

APPENDICE A – Tabelle capitolo 2

YEAR	MEAN [€/MWH]	SD [€/MWH]	VARIANCE [(€/MWH)^2]	MAX [€/MWH]	MIN [€/MWH]
TOT	63,09	26,54	704,26	378,47	0,00
2004	51,60	27,37	749,12	189,19	1,10
2005	58,59	27,67	765,56	170,61	10,42
2006	74,75	33,68	1134,49	378,47	15,06
2007	70,99	37,02	1370,17	242,42	21,44
2008	86,99	32,91	1082,87	211,99	21,54
2009	63,72	26,42	698,20	172,25	9,07
2010	64,12	18,36	337,27	174,62	10,00
2011	72,23	16,77	281,14	164,80	10,00
2012	75,48	22,18	491,82	324,20	12,14
2013	62,99	18,20	331,25	151,88	0,00
2014	52,08	15,56	241,97	149,43	2,23
2015	52,31	13,32	177,39	144,57	5,62
2016	42,78	13,02	169,45	150,00	10,94
2017	53,95	16,46	270,84	170,00	10,00
2018	54,21	15,66	245,25	159,40	10,00

YEAR	MEDIAN [€/MWH]	Q1 [€/MWH]	Q3[€/MWH]	VAS [€/MWH]	VAI [€/MWH]
TOT	59,40	45,00	76,04	122,61	0,00
2004	45,34	30,11	66,63	121,41	1,10
2005	51,81	35,67	78,48	142,70	10,42
2006	66,18	50,19	90,56	151,11	15,06
2007	65,46	46,36	89,02	153,03	21,44
2008	85,08	63,79	107,39	172,79	21,54
2009	61,28	45,93	80,07	131,27	9,07
2010	65,00	52,07	73,74	106,26	19,55
2011	71,39	62,59	81,33	109,45	34,47
2012	74,52	63,29	86,45	121,19	98,03
2013	64,07	53,48	72,71	101,55	24,63
2014	51,26	42,16	61,00	89,26	13,89
2015	51,06	43,47	58,46	80,93	21,00
2016	40,42	33,82	49,07	71,94	10,95
2017	50,60	44,00	59,51	82,77	20,73
2018	51,25	45,01	59,04	80,08	23,96

Tabella 22 – MGP: Grandezze statistiche di interesse

MGP	GEOGRAFICHE							VIRTUALI NAZIONALI				
	PUN	CNOR	CSUD	NORD	SARD	SICI	SUD	BRNN	FOGN	MFTV	PRGP	ROSN
2004	1	1	1	1	1	1	1	1	0	1	1	1
2005	1	1	1	1	1	1	1	1	0	1	1	1
2006	1	1	1	1	1	1	1	1	1	1	1	1
2007	1	1	1	1	1	1	1	1	1	1	1	1
2008	1	1	1	1	1	1	1	1	1	1	1	1
2009	1	1	1	1	1	1	1	1	1	1	1	1
2010	1	1	1	1	1	1	1	1	1	1	1	1
2011	1	1	1	1	1	1	1	1	1	1	1	1
2012	1	1	1	1	1	1	1	1	1	0	1	1
2013	1	1	1	1	1	1	1	1	1	0	1	1
2014	1	1	1	1	1	1	1	1	1	0	1	1
2015	1	1	1	1	1	1	1	1	1	0	1	1
2016	1	1	1	1	1	1	1	1	1	0	1	1
2017	1	1	1	1	1	1	1	1	1	0	1	1
2018	1	1	1	1	1	1	1	1	1	0	1	1

Tabella 23 - MGP : zone geografiche e virtuali nazionali commerciate di anno in anno

MGP	VIRTUALI ESTERE										
	AUST	BSP	CORS	COAC	FRAN	GREC	SLOV	SVIZ	XFRA	XAUS	MALT
2004	1	0	1	0	1	1	1	1	0	0	0
2005	1	0	1	0	1	1	1	1	0	0	0
2006	1	0	1	1	1	1	1	1	0	0	0
2007	1	0	1	1	1	1	1	1	0	0	0
2008	1	0	1	1	1	1	1	1	0	0	0
2009	1	0	1	1	1	1	1	1	0	0	0
2010	1	0	1	1	1	1	1	1	0	0	0
2011	1	1	1	1	1	1	1	1	0	0	0
2012	1	1	1	1	1	1	1	1	0	0	0
2013	1	1	1	1	1	1	1	1	0	0	0
2014	1	1	1	1	1	1	1	1	0	0	0
2015	1	1	1	1	1	1	1	1	1	1	1
2016	1	1	1	1	1	1	1	1	1	1	1
2017	1	1	1	1	1	1	1	1	1	1	1
2018	1	1	1	1	1	1	1	1	1	1	1

Tabella 24 - MGP : zone e virtuali estere commerciate di anno in anno

	NON ATTIVE						
MGP	CALB	PBNF	TBRV	E_CO	E_NE	E_NW	E_SD
2004	1	1	1	0	1	1	0
2005	1	1	1	0	1	1	1
2006	1	1	1	1	1	1	1
2007	1	0	1	1	1	1	1
2008	1	0	1	1	1	1	1
2009	0	0	0	0	0	0	0
2010	0	0	0	0	0	0	0
2011	0	0	0	0	0	0	0
2012	0	0	0	0	0	0	0
2013	0	0	0	0	0	0	0
2014	0	0	0	0	0	0	0
2015	0	0	0	0	0	0	0
2016	0	0	0	0	0	0	0
2017	0	0	0	0	0	0	0
2018	0	0	0	0	0	0	0

Tabella 25 - MGP : zone non attive commerciate di anno in anno

ZONE	MEAN [€/MWH]	SD [€/MWH]	VARIANCE [(€/MWH)^2]	MAX [€/MWH]	MIN [€/MWH]
AUST	42,34	15,65	244,79	160,00	10,00
BRNN	39,60	12,46	155,17	155,00	0,00
CNOR	42,68	14,73	216,92	155,00	8,50
COAC	41,05	13,19	174,08	155,00	5,03
CORS	39,12	18,48	341,47	155,00	0,00
CSUD	41,05	13,15	173,00	155,00	8,50
FOGN	39,60	12,46	155,17	155,00	0,00
FRANC	42,34	15,65	244,79	160,00	10,00
GREC	39,60	12,46	155,17	155,00	0,00
MALT	47,59	65,25	4257,63	3000,00	1,00
MFTV	42,34	15,65	244,79	160,00	10,00
NAT	40,85	13,10	171,67	155,00	10,00
NORD	42,34	15,65	244,79	160,00	10,00
PRGP	47,59	65,25	4257,63	3000,00	1,00
ROSN	37,89	13,92	193,77	155,00	0,00
SARD	41,05	13,19	174,08	155,00	5,03
SICI	47,59	65,25	4257,63	3000,00	1,00
SLOV	42,34	15,65	244,79	160,00	10,00
SUD	39,60	12,46	155,17	155,00	0,00
SVIZ	42,34	15,65	244,79	160,00	10,00

A

ZONE	MEDIAN [€/MWH]	Q1 [€/MWH]	Q3[€/MWH]
AUST	38,22	31,92	49,00
BRNN	37,12	31,54	45,00
CNOR	39,55	32,50	49,04
COAC	38,42	32,09	47,00
CORS	38,00	30,59	47,99
CSUD	38,41	32,10	47,00
FOGN	37,12	31,54	45,00
FRANC	38,22	31,92	49,00
GREC	37,12	31,54	45,00
MALT	44,05	35,89	53,58
MFTV	38,22	31,92	49,00
NAT	38,00	32,00	46,50
NORD	38,22	31,92	49,00
PRGP	44,05	35,89	53,58
ROSN	36,45	30,50	44,50
SARD	38,42	32,09	47,00
SICI	44,05	35,89	53,58
SLOV	38,22	31,92	49,00
SUD	37,12	31,54	45,00
SVIZ	38,22	31,92	49,00

Tabella 26 – MII : grandezze di sommario della distribuzione

ENERGIA PRODOTTA [<i>TWh</i>]										
	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016
IDRAULICA	32,82	41,62	49,14	51,12	45,82	41,88	52,77	58,55	45,54	42,43
EOLICA	4,03	4,86	6,54	9,13	9,86	13,41	14,90	15,18	14,84	17,69
FOTOVOLTAICO	0,04	0,19	0,68	1,91	10,80	18,86	21,59	22,31	22,94	22,10
GEOTERMICA	5,57	5,52	5,34	5,38	5,65	5,59	5,66	5,92	6,19	6,29
BIOENERGIE	5,44	5,97	7,63	9,44	10,83	12,49	17,09	18,73	19,40	19,51
BIOMASSE	3,99	4,30	4,44	4,31	4,73	4,75	5,89	6,19	6,29	6,54
BIOGAS	1,45	1,60	1,74	2,05	3,41	4,62	7,45	8,20	8,21	8,26
BIOLQUIDI	0,00	0,07	1,45	3,08	2,70	3,12	3,76	4,34	4,89	4,71
SOMMA	47,90	58,16	69,33	76,96	82,96	92,22	112,01	120,68	108,90	108,02

Tabella 27 – Energia prodotta da FER : dettaglio annuale per fonte.

APPENDICE B – Diagrammi di calibrazione

Si raccolgono in questa appendice i grafici necessari secondo il procedimento illustrato paragrafo B della sezione 4.2 a pagina 88. Rappresentano il passaggio 2 schematizzato da Figura 54 a pagina 91 che, a partire dalla calibrazione generale, permette di mantenere solamente le combinazioni di k_s , τ e m che si avvicinano all'ottimo e quindi di identificare almeno la zona di minimo globale. Le quattro figure a seguito sono divise per orizzonte e per set di calibrazione, precisamente:

- $T = 1 \rightarrow$ *set estivo* a pagina 92
- $T = 1 \rightarrow$ *set invernale* a pagina 102
- $T = 24 \rightarrow$ *set estivo* a pagina 97
- $T = 24 \rightarrow$ *set invernale* a pagina 106
- $T = 24 \rightarrow$ *set globale* a pagina 114

Le figure che seguono raggruppano, partendo dal basso : il valore di k_s , τ e m più frequenti e nel plot più in alto il rispettivo valore di RMSE della terna così costruita, in breve è da interpretare secondo sezioni verticali, ognuna di esse rappresenta la terna di parametri più frequenti (e rispettivo RMSE) risultati dall'imposizione del limite riportato in ascissa. Le unità di misura sono €/MWh ovunque si tratta di RMSE, in ascissa è riportato il limite di RMSE a prescindere dal grafico, l'ordinata del plot in alto indica sempre €/MWh di RMSE derivate dalla terna più frequente.

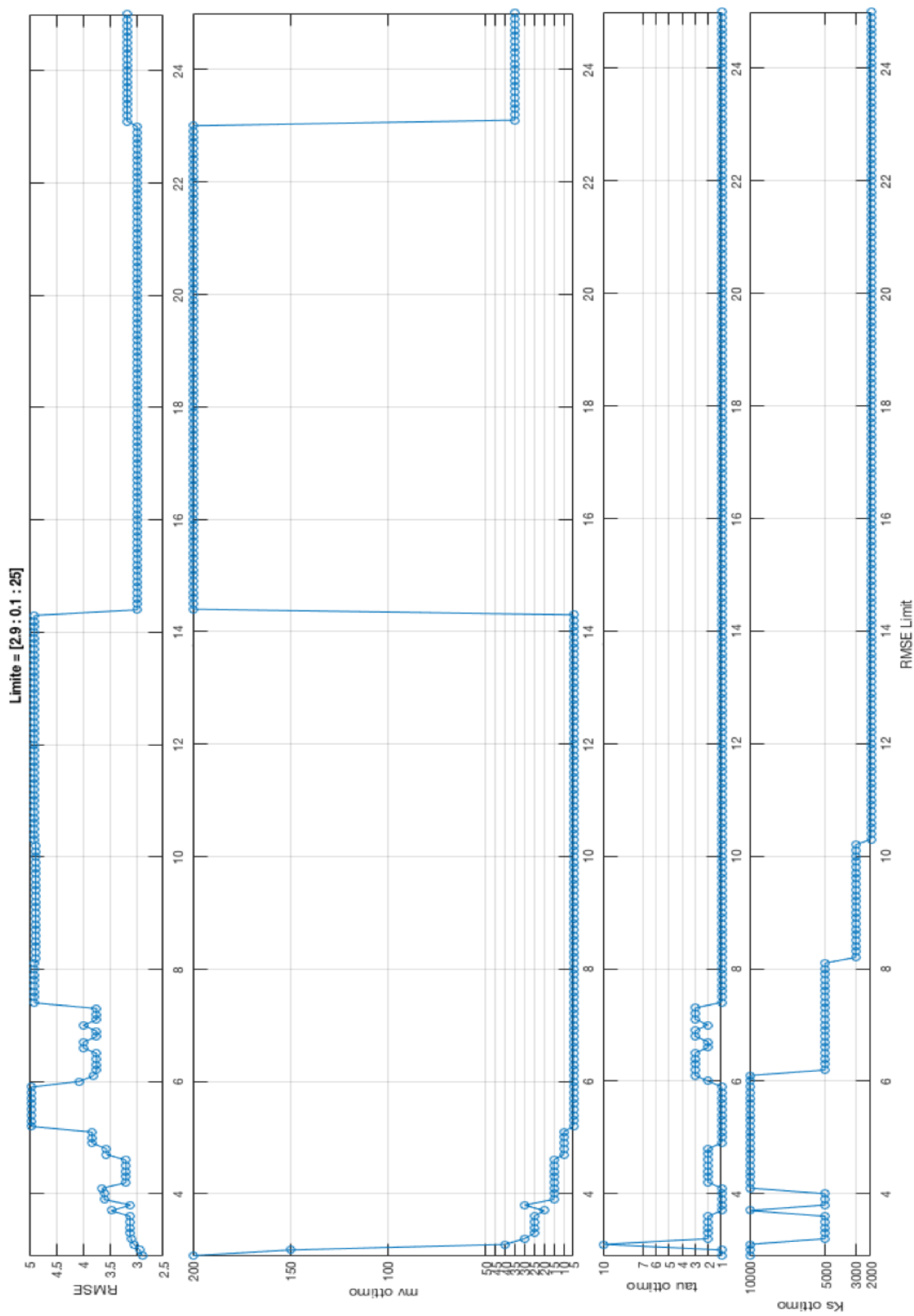


Figura 84 – Parametri più frequenti ed RMSE al variare del limite sul' errore , set estivo , $T=1$

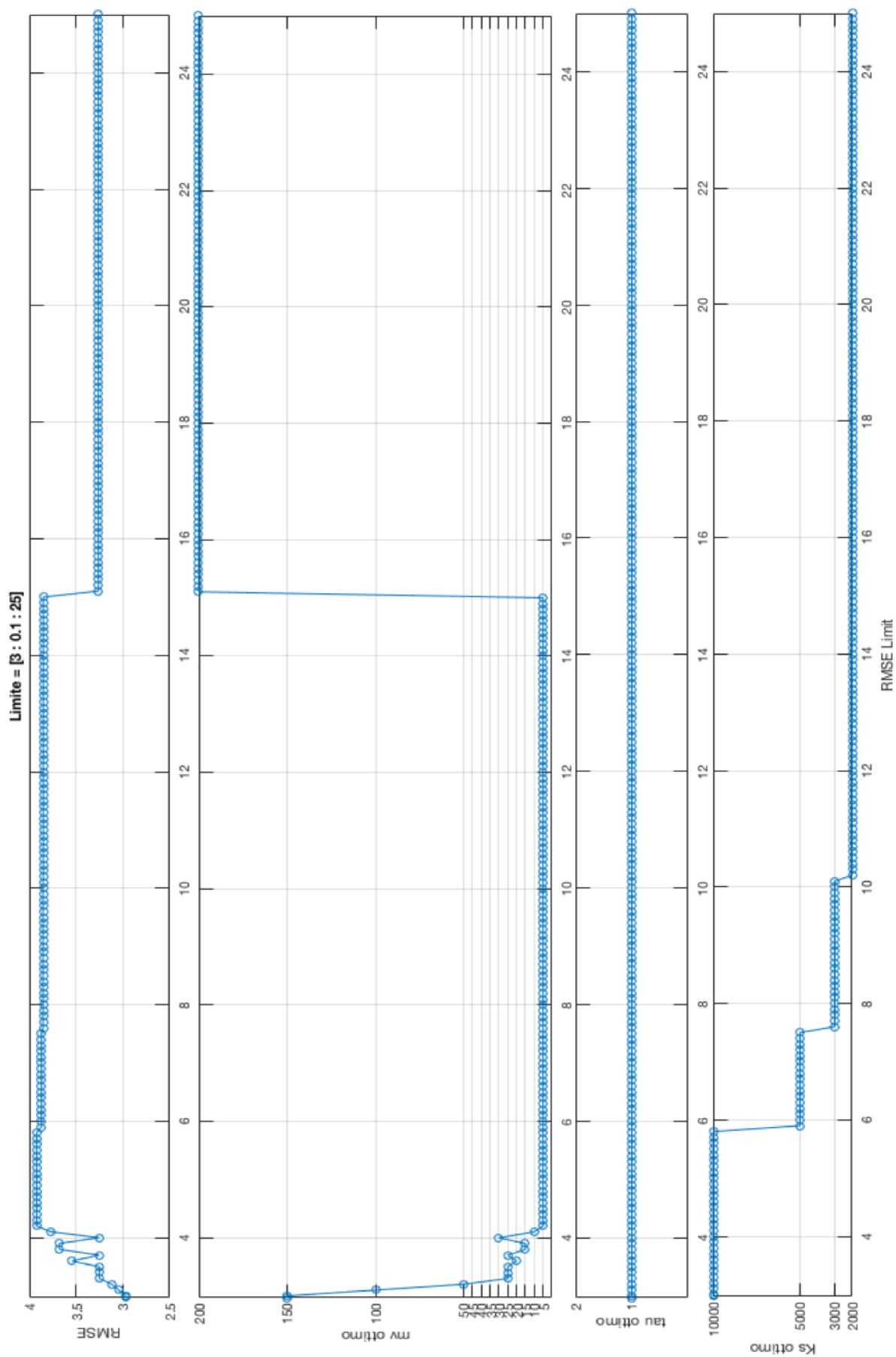


Figura 85 - Parametri più frequenti ed RMSE al variare del limite sul' errore , set invernale , $T=1$

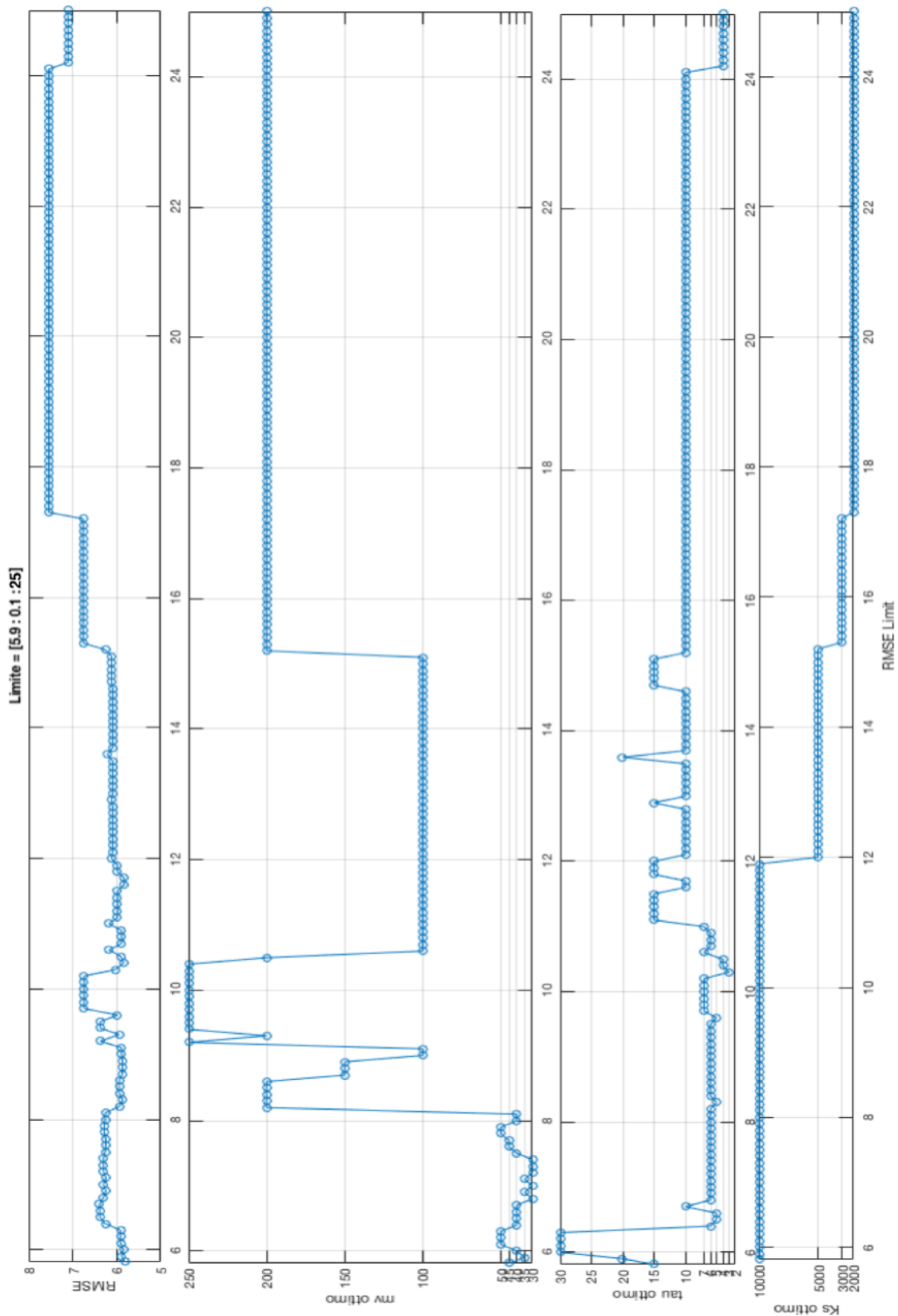


Figura 86 – Parametri più frequenti ed RMSE al variare del limite sul' errore , set estivo , $T=24$

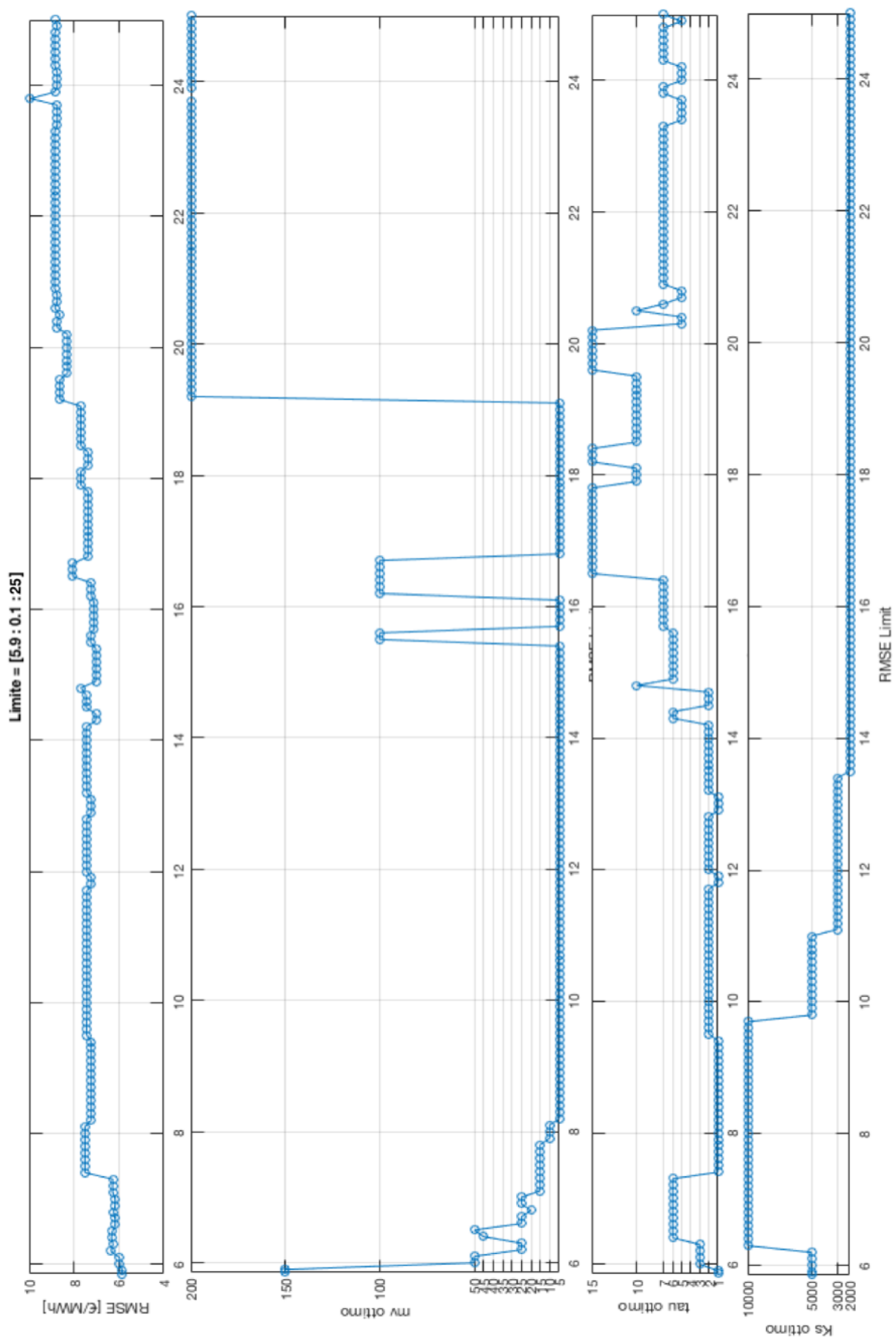


Figura 87 - Parametri più frequenti ed RMSE al variare del limite sul' errore , *set invernale* , $T=24$

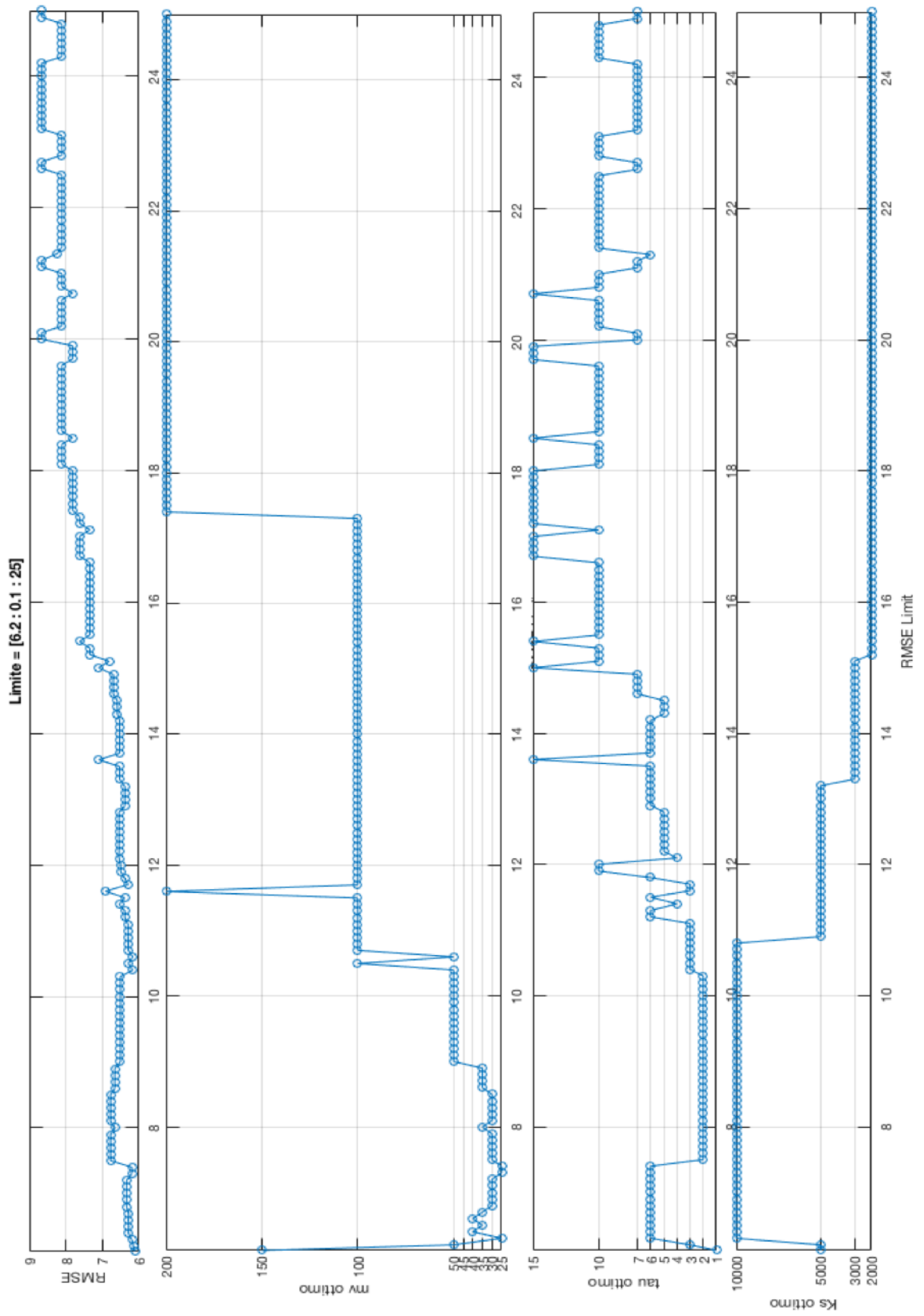


Figura 88 - Parametri più frequenti ed RMSE al variare del limite sul' errore , set globale , $T=24$