

POLITECNICO DI TORINO

Collegio di Ingegneria Energetica e Nucleare

**Corso di Laurea Magistrale in
Ingegneria Energetica e Nucleare**

Tesi di Laurea Magistrale

**Data mining analysis per la previsione
dei profili di richiesta delle utenze del
teleriscaldamento**



Relatori:

Prof. Vittorio Verda

Ing. Elisa Guelpa

Candidato:

Federico Aiello

A.A. 2017-2018

Indice

Capitolo 1	4
INTRODUZIONE.....	4
1.1 Struttura della Tesi.....	4
Capitolo 2	6
GENERALITA' SUL TELERISCALDAMENTO	6
2.1 Componenti del teleriscaldamento	9
2.1.1 Centrali termiche	9
2.1.2 Rete di trasporto.....	12
2.1.3 Sottostazione	15
2.1.4 Sistema di pompaggio	16
2.2 Diffusione del teleriscaldamento nel mondo e in Europa ..	17
2.3 Teleriscaldamento in Italia e Torino.....	21
2.3.1 Teleriscaldamento a Torino	22
2.4 Quadro Normativo Italiano sul teleriscaldamento	24
Capitolo 3	27
MODELLO	27
3.1 Caratteristiche principali della richiesta termica giornaliera	27
3.2 Caratteristiche del modello.....	29
3.3 Determinazione dei coefficienti Beta e previsione dei profili di richiesta.....	35

3.3.1 Regressione lineare	36
3.3.2 Costruzione modello.....	37
3.4 Miglioramento delle prestazioni del modello	44
3.4.1 Clustering.....	44
3.4.2 Applicazione del Clustering.....	50
Capitolo 4	55
RISULTATI.....	55
4.1 Confronto dei risultati ottenuti.....	55
4.2 Confronto dei risultati ottenuti con i dati sperimentali.....	65
4.3 Conclusioni	69
BIBLIOGRAFIA	70

Capitolo 1

INTRODUZIONE

1.1 Struttura della Tesi

L'obiettivo che si persegue nella presente tesi è la creazione di un modello, attraverso il quale prevedere l'andamento dei profili di richiesta delle utenze allacciate alla rete di Teleriscaldamento di Torino.

Questo modello, che è stato creato analizzando utenze di cui è noto il profilo di richiesta, potrà essere utilizzato per prevedere l'anzidetto andamento anche per utenze per le quali, per insufficienza di dati, questo non è noto.

In queste studio sono state prese in considerazione le utenze dei baricentri 408,410,414,418 e 420 della rete di Torino.

Il modello, che consente di determinare il profilo di richiesta termica, è stato ottenuto utilizzando un metodo di regressione lineare i cui dati di input sono le ore di accensione giornaliere, la volumetria e il fabbisogno di energia termica annua delle utenze.

Successivamente, per riuscire ad avere una migliore previsione dei profili di richiesta, è stata effettuata un'analisi di Clustering in cui le utenze non vengono considerate in un unico insieme ma divise in più Cluster (in questa analisi 3) in modo tale da ottenere più gruppi contenenti utenze simili fra loro.

In questo modo, anziché costruire un solo modello per tutte le utenze, vengono costruiti più modelli a seconda del numero di Cluster presenti.

Nel Capitolo 2 viene esposto il sistema di teleriscaldamento; in cosa consiste, i componenti principali e come funziona. Infine viene descritta la diffusione di questa tecnologia nel mondo e in Italia, focalizzando l'attenzione sulla rete della città di Torino, mostrando anche il quadro normativo italiano.

Nel Capitolo 3 viene presentato il procedimento utilizzato per la creazione del modello, mostrandone le caratteristiche.

Viene successivamente esposta l'analisi di Clustering, discutendo quindi i metodi utilizzati per la determinazione dei singoli Cluster.

Nel Capitolo 4 vengono presentati i risultati ottenuti. Viene quindi mostrato, per le utenze analizzate, di cui è noto lo storico dei dati, il confronto fra i profili di richiesta ricavati dal modello e quelli noti giungendo a provare l'attendibilità del modello.

Capitolo 2

GENERALITA' SUL TELERISCALDAMENTO

Il teleriscaldamento, che testualmente significa riscaldamento a distanza [1], è un «*mezzo semplice, pulito, economico e sicuro per riscaldare gli edifici*» [2] ed utilizza il calore prodotto da una centrale termica, da un impianto di cogenerazione o da impianti che sfruttano energie rinnovabili.

Attraverso il teleriscaldamento è possibile sfruttare fonti di energia diversificate e disponibili in loco, esso può rappresentare anche un'occasione di vantaggio economico per i paesi che lo adottano. Infatti, con tale tecnologia, si riducono gli sprechi energetici in quanto si possono sfruttare diverse fonti di energia quali le risorse rinnovabili ed i processi industriali [3].

Esso dunque consiste nel trasporto a distanza di un fluido termovettore, generalmente acqua calda o acqua surriscaldata, attraverso una rete di tubazioni isolate, che giunge fino agli edifici allacciati alla rete stessa. Qui, mediante sottocentrali, il calore viene ceduto all'impianto dell'abitazione e ciò consente di riscaldare gli ambienti e produrre acqua calda sanitaria. Alla fine del trasferimento del calore, il fluido termovettore ritorna in centrale per essere riportato alla massima temperatura e ricominciare il ciclo [2].

Quindi i componenti principali di un sistema di teleriscaldamento sono:

- una o più centrali termiche, dove viene generato calore;
- una rete di trasporto;
- utenze, ciascuna con sottostazione di scambio;

- sistema di pompaggio, eventualmente suddiviso in più stazioni per vincere le perdite di carico nelle tubazioni [5].

Attraverso questa tecnologia, l'impianto di distribuzione, già presente all'interno degli edifici, rimane inalterato poiché la caldaia convenzionale viene sostituita da uno scambiatore collocato nelle sottostazioni. Se più utenze fossero allacciate a tale scambiatore, è necessario installare in ciascuna di esse un'apparecchiatura che consenta di gestire autonomamente la temperatura interna ai locali e i relativi consumi[4].

Il sistema di teleriscaldamento è caratterizzato da un'elevata affidabilità che consente di poterlo applicare ad interi territori urbani; in tal modo si realizza un effettivo servizio pubblico[4].

Nondimeno, affinché tale tecnologia risulti efficace, appare necessario individuare utenze che siano collocate in una ristretta area ben definita. Infatti non è consigliabile una rete di teleriscaldamento che colleghi utenze distanti fra loro in quanto le perdite termiche e i costi di pompaggio del fluido termovettore sarebbero troppo elevati, rendendo il sistema inefficiente[4].

Il teleriscaldamento svolge la propria funzione sulla base del principio di separazione fra produzione e consumo di calore. La produzione è di tipo centralizzato e si ottiene con impianti che sfruttano le migliori tecnologie, massimizzando il rendimento[4].

Dal punto di vista energetico, per ottenere una maggior efficienza energetica globale, è necessario che il calore immesso nella rete sia generato da un sistema cogenerativo. Infatti, in questo assetto, è possibile produrre energia elettrica e recuperare simultaneamente energia termica che nelle centrali convenzionali viene dispersa come "scarto" [4].

Di conseguenza l'utilizzo di una centrale ad assetto cogenerativo rispetto ad una centrale convenzionale permette, a parità di energia utile prodotta, un minor consumo di combustibile, un maggior sfruttamento delle risorse utilizzate e minori perdite[4]. E' possibile notare ciò confrontando le figure 2.1 e 2.2.

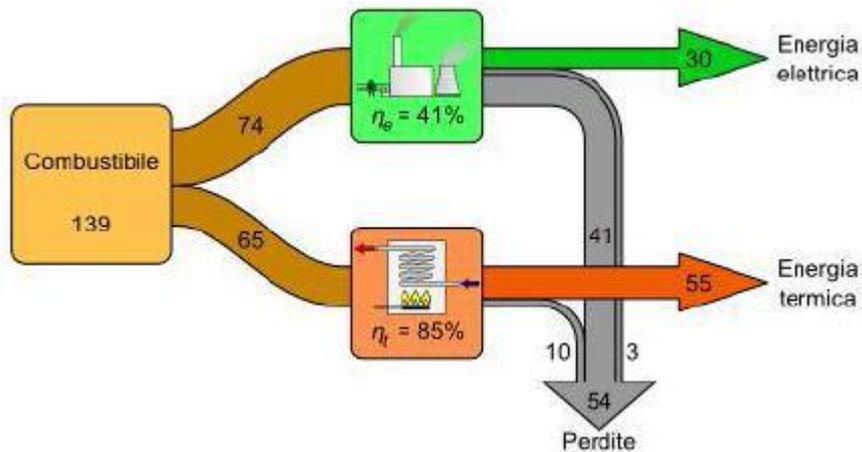


Figura 2.1 Efficienza energetica di un sistema convenzionale. Fonte [4]

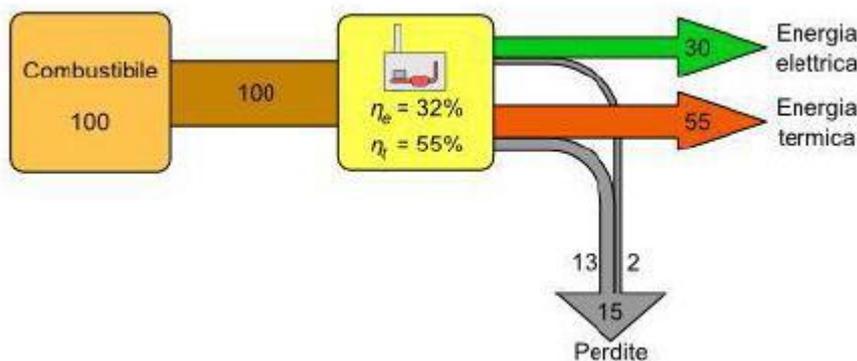


Figura 2.2 Efficienza energetica di un sistema cogenerativo Fonte [4]

Per ottenere la stessa energia utile finale, 30 unità di energia elettrica e 55 di calore, sono necessarie 139 unità di energia primaria nel caso di produzione separata (con rendimento elettrico medio del 41% e rendimento medio di

generazione di calore con una caldaia pari all'85%) a differenza delle 100 unità nel caso della cogenerazione con rendimento totale del 87% (32% elettrico e 55% termico) [4].

In aggiunta si può notare come le perdite energetiche diminuiscano notevolmente.

2.1 Componenti del teleriscaldamento

Come detto precedentemente i componenti principali di una rete di teleriscaldamento sono:

- una o più centrali termiche, dove viene generato calore;
- una rete di trasporto;
- utenze, ciascuna con sottostazione di scambio;
- sistema di pompaggio, eventualmente suddiviso in più stazioni per vincere le perdite di carico nelle tubazioni [5].

2.1.1 Centrali termiche

Generalmente la centrale utilizzata in una rete di teleriscaldamento, come si evince dalle ragioni spiegate nel paragrafo precedente, è ad assetto cogenerativo. Tuttavia anche l'utilizzo di impianti che sfruttano risorse rinnovabili (ad esempio geotermico e biomassa) risulta interessante.

La potenza e l'energia richieste dalle utenze determinano la configurazione della centrale. In genere per impianti di piccola potenza (alcune decine di MW), normalmente associati ad una temperatura di mandata bassa, si

possono utilizzare sistemi di cogenerazione di piccola taglia (es. turbogas), pompe di calore geotermiche, sistemi WTE, caldai a biomasse e collettori solari[5].

Per impianti di grande potenza (centinaia di MW), caratterizzati da elevate differenze tra temperatura di mandata e ritorno, si possono utilizzare sistemi di cogenerazione di grande taglia (cicli combinati), sistemi WTE, calore residuo da processi industriali, caldaie a biomasse[5].

La potenza prodotta dalla centrale termica è pari alla somma della richiesta delle utenze e delle perdite termiche della rete. Tali perdite risultano trascurabili in termini di potenza (1-2% della potenza massima), mentre raggiungono valori più elevati in termini di energia (8-10%). Questo perché esse sono costanti in tutta la rete e quindi il loro peso percentuale aumenta al diminuire della richiesta dell'utenza[5].

La potenza fornita dalla centrale è espressa nell'equazione 2.1:

$$\sum_{p=1}^p Q_{production} = Q_{loss} + \sum_{c=1}^c Q_{consumption} \quad (2.1)$$

Dove:

- $Q_{production}$ è la potenza termica totale prodotta dalla centrale
- Q_{loss} rappresenta le perdite termiche
- $Q_{consumption}$ è la richiesta termica dell'utenza

La dimensione della centrale dipende fortemente dall'energia richiesta dalle utenze allacciate alla rete. Per questo motivo è necessario calcolare il fabbisogno termico complessivo. Esso viene determinato a partire dai profili di richiesta di ciascuna utenza, che vengono cumulate fino ad individuare la curva di carico della centrale.

Essa non è semplicemente la somma dei fabbisogni delle utenze ed è influenzata da tre aspetti:

- la presenza di un picco di richiesta corrispondente a $2/5$ volte la richiesta stazionaria, ciò è cagionato dal fatto che, durante la notte, la temperatura del fluido termovettore, che si trova nelle tubazioni, diminuisce raggiungendo la temperatura ambiente e di conseguenza al mattino per portarlo alla temperatura di esercizio è necessaria una maggior potenza [5];
- una variazione della domanda associata ad una variazione della temperatura di ritorno del primario ha un effetto ritardato sul carico termico della centrale, dell'ordine di 15 minuti/km (considerando una velocità del fluido termovettore pari a 1m/s) [5];
- una variazione della domanda associata ad una variazione della portata invece ha un effetto praticamente istantaneo sul carico termico della centrale (ritardo di circa 0.5 s/km) [5].

Questi elementi generano la curva di carico della centrale. Un andamento tipico è quello mostrato in figura 2.3.

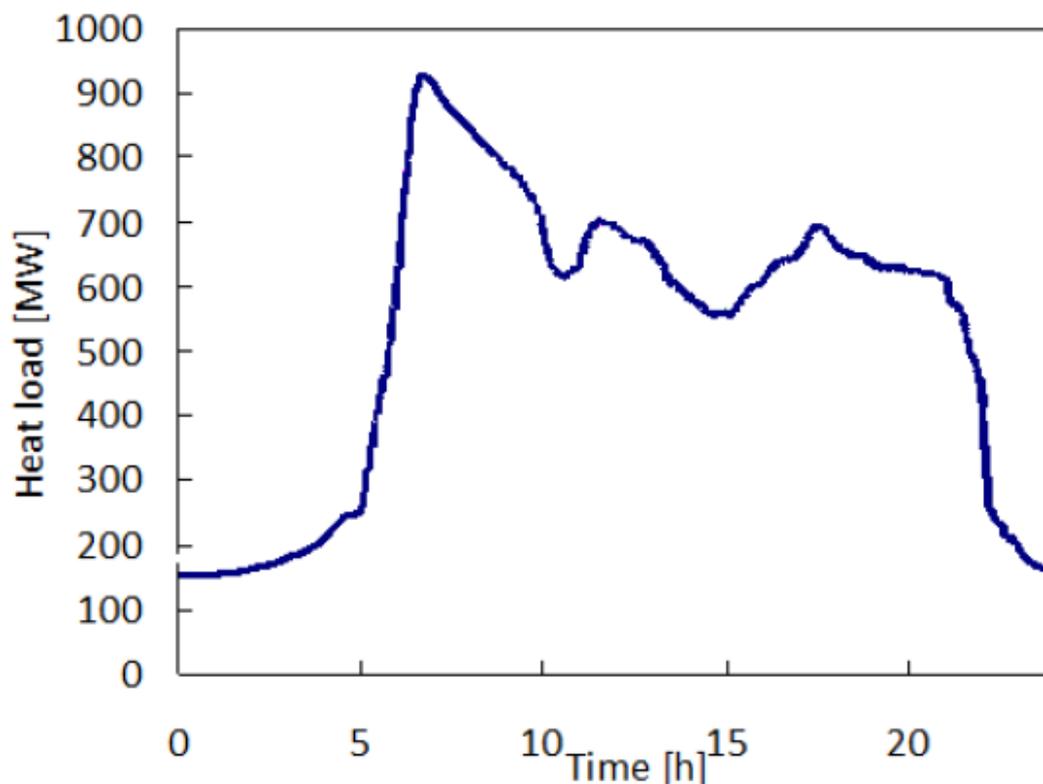


Figura 2.3 Curva di carico termico di una centrale termica in una tipica giornata invernale. Fonte [5]

In genere la centrale non viene dimensionata al valore del picco di richiesta, ma ad una potenza inferiore, potendo funzionare così a pieno carico per un numero elevato di ore all'anno, ammortizzando così l'elevato costo d'investimento [5].

Il picco viene poi soddisfatto attraverso caldaie ad integrazione o serbatoi di accumulo [5].

2.1.2 Rete di trasporto

La rete di trasporto o distribuzione rappresenta il mezzo con cui l'energia termica prodotta nella centrale viene distribuita alle utenze e costituisce la

parte più costosa del sistema di teleriscaldamento con un costo compreso fra il 50% e l'80% dell'investimento globale [4].

Essa è caratterizzata da una doppia tubazione: il tubo di mandata e quello di ritorno.

Il primo parte dalla centrale e si dirige verso le utenze mentre il secondo percorre il cammino inverso (torna alla centrale partendo dalle utenze). Il fluido termovettore che scorre nel tubo di mandata ha una temperatura, nonché un'entalpia, maggiore di quello presente nel tubo di ritorno [4].

Studiando più da vicino una rete si può notare come essa sia costituita da una "dorsale", tubazione principale, alla quale sono collegate tubazioni di minore diametro che vengono utilizzate per trasportare il calore alle utenze [4].

Le reti possono essere di due tipi:

- albero: esiste un unico percorso che collega la centrale e l'utenza e di conseguenza se a causa di un guasto un tratto di rete venisse chiuso si arresterebbe la fornitura di calore alle utenze collocate dopo il tratto interrotto [5];
- magliate: esistono più percorsi che collegano la centrale e l'utenza, quindi se a causa di un guasto un tratto di rete venisse chiuso la fornitura di calore verrebbe impedita solo al tratto chiuso. In aggiunta esse presentano minori cadute di pressione, riducendo così i costi di pompaggio [5].

Le due tipologie di rete sono rappresentate nelle figura 2.4 e 2.5.

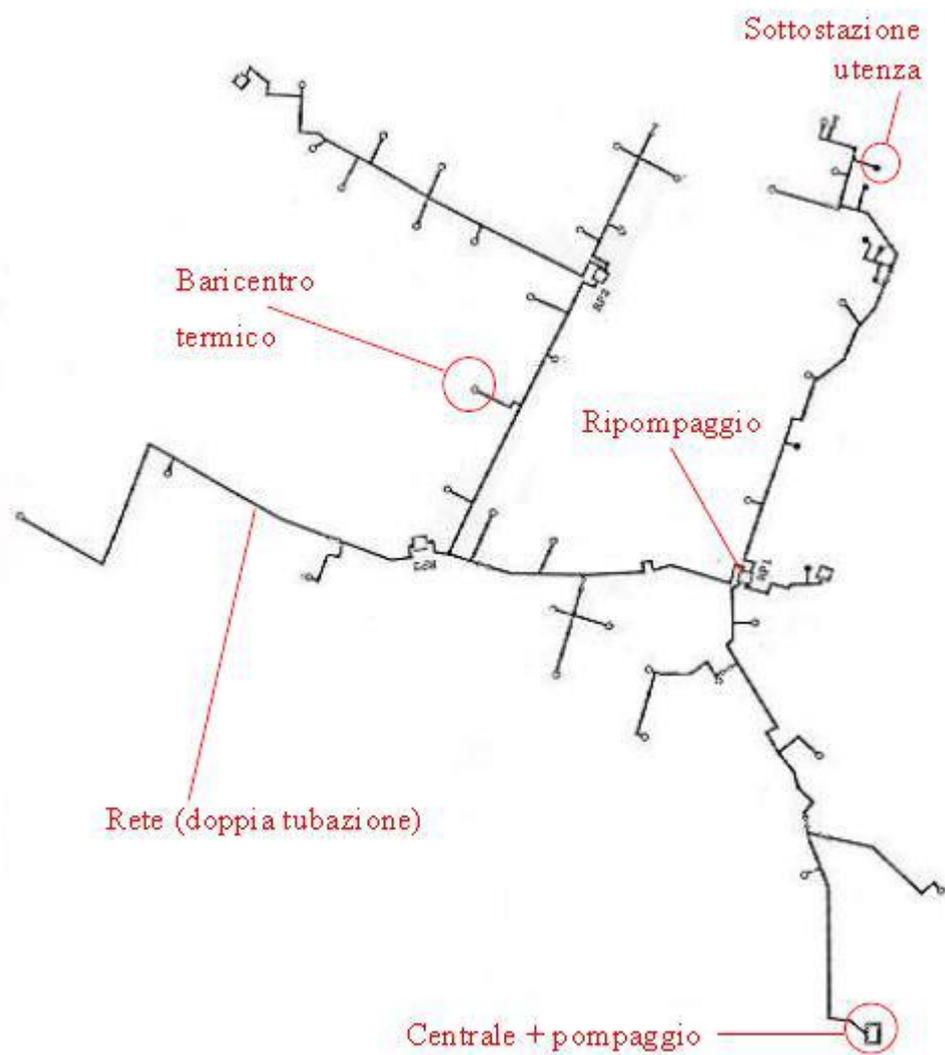


Figura 2.4 Rete di trasporto ad albero. Fonte [5]

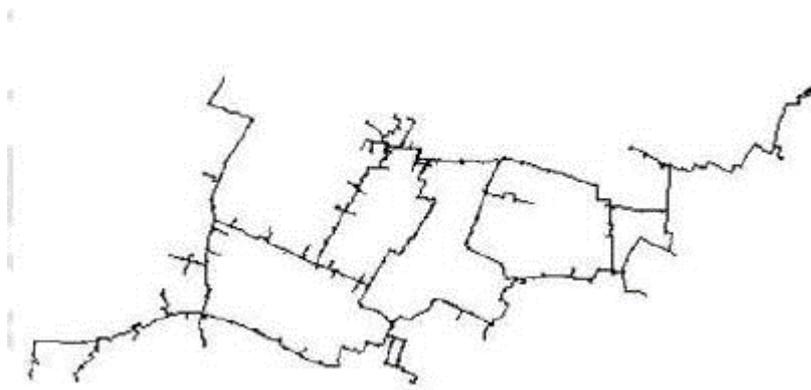


Figura 2.5 Rete di trasporto ad albero. Fonte [5]

2.1.3 Sottostazione

La sottostazione viene utilizzata per rimpiazzare le vecchie caldaie tradizionali e ogni sottostazione identifica un'utenza [4].

Gli elementi principali di una sottostazione sono:

- uno scambiatore di calore che separa il circuito primario (circuito della rete di teleriscaldamento) dal secondario (circuito di riscaldamento dell'utenza);
- una centralina termica che attraverso la temperatura esterna genera il set-point della temperatura di mandata del circuito secondario. Questo dispositivo viene utilizzato per mantenere la temperatura interna dell'edificio al valore desiderato (di solito 20°C) [5];
- una valvola termoregolatrice utilizzata per regolare la portata del circuito primario in modo tale che venga raggiunta la temperatura di set-point di mandata al secondario [5];

Per quanto riguarda le temperature di mandata e di ritorno al secondario, esse dipendono dall'impianto di riscaldamento presente nella struttura. Tali valori sono riassunti nella tabella 2.1 [5].

Tipologia di impianto	T_{mandata}	T_{ritorno}
Radiatori	70-80°C	60-70°C
Pannelli radianti	45-50°C	38-42°C
Fan coils	50°C	40°C

Tabella 2.1 Temperature di progetto lato secondario in base alla tipologia dell'impianto di riscaldamento installato. Fonte [5]

2.1.4 Sistema di pompaggio

Le centrali di pompaggio vengono utilizzate per supplire le perdite di pressione presenti nella rete garantendo così il raggiungimento di tutte le utenze con la portata di progetto [4] [5].

Tali centrali sono costituite da:

- elettropompe di circolazione;
- filtri;
- vaso di espansione;
- impianto di trattamento acqua;

- sistema di controllo automatico;
- serbatoio di riserva [4].

La velocità nelle tubazioni non deve essere troppo elevata ma neanche troppo bassa. Infatti nel primo caso si verificano perdite di carico elevate, mentre nel secondo si necessita di tubazioni con diametri di grande dimensione che comporta maggiori dispersioni termiche e costi [4].

2.2 Diffusione del teleriscaldamento nel mondo e in Europa

Il teleriscaldamento fu usato per la prima volta a New York nel 1876. Successivamente si diffuse anche in Europa in città come Amburgo e Parigi [6].

Oggi la nazione che usufruisce maggiormente di questa tecnologia è la Russia che presenta inoltre la rete più estesa al mondo a Mosca seguita da imponenti reti anche a San Pietroburgo [6].

Per determinare la rilevanza del teleriscaldamento all'interno dell'economia di un paese è utile analizzare alcuni parametri. Essi sono: le vendite di calore per il teleriscaldamento, la percentuale di abitanti che ne usufruisce, la lunghezza della rete di trasporto e distribuzione.

- Vendite di calore per il teleriscaldamento:

Come detto precedentemente la nazione più teleriscaldata è la Russia che infatti ha venduto, agli utenti del settore residenziale e dell'industria, quasi 7 milioni di TJ nel 2007. Dopo la Russia troviamo la Cina che, nel 2013, ha venduto più di 3 milioni di TJ, incrementando le vendite del 24.9% rispetto al 2009 come si può notare nella figura 2.6 [3].

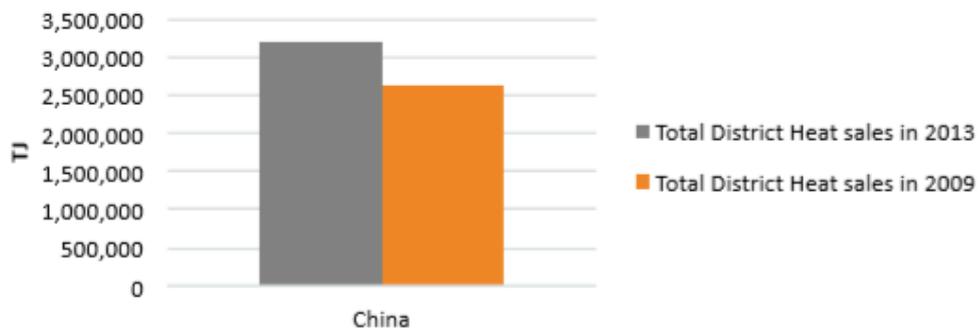


Figura 2.6 Vendite di calore per Teleriscaldamento per la Cina negli anni 2009 e 2013. Fonte [3]

Per quanto riguarda l'Unione Europea le vendite, nel 2013, ammontano a circa 1.5 milioni di TJ e sono diminuite del 4% rispetto al 2009. Analizzando la figura 2.7, in cui vengono mostrate le vendite per ogni nazione nel 2009 e nel 2013, si può notare come Germania, Polonia e Svezia siano i Paesi con maggiori vendite mentre l'Austria e l'Italia quelli in maggior crescita con un incremento del 28 % per la prima e del 41% per la seconda. Infine è possibile osservare che l'Ungheria, la Slovacchia e la Bulgaria sono le nazioni che hanno subito il maggior calo nelle vendite (-24%, -12%, -10% rispettivamente) [3].

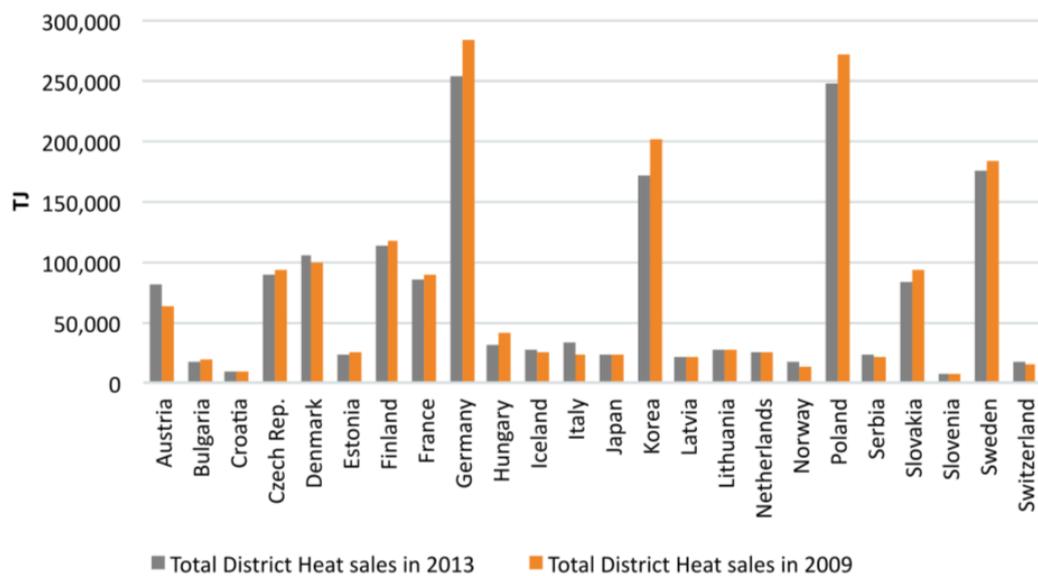


Figura 2.7 Vendite di calore per teleriscaldamento per le nazioni europee.
Fonte[3]

- Percentuale di abitanti che usufruisce del teleriscaldamento:

Come si può notare dalla figura 2.8 il numero di abitanti allacciati alla rete di teleriscaldamento varia da nazione a nazione. Ciò è dovuto al fatto che la richiesta termica dell'utenza può essere soddisfatta utilizzando diverse tecnologie. Infatti in alcuni paesi il teleriscaldamento è la tecnologia dominante sul mercato mentre in altri l'uso delle caldaie tradizionali [3].

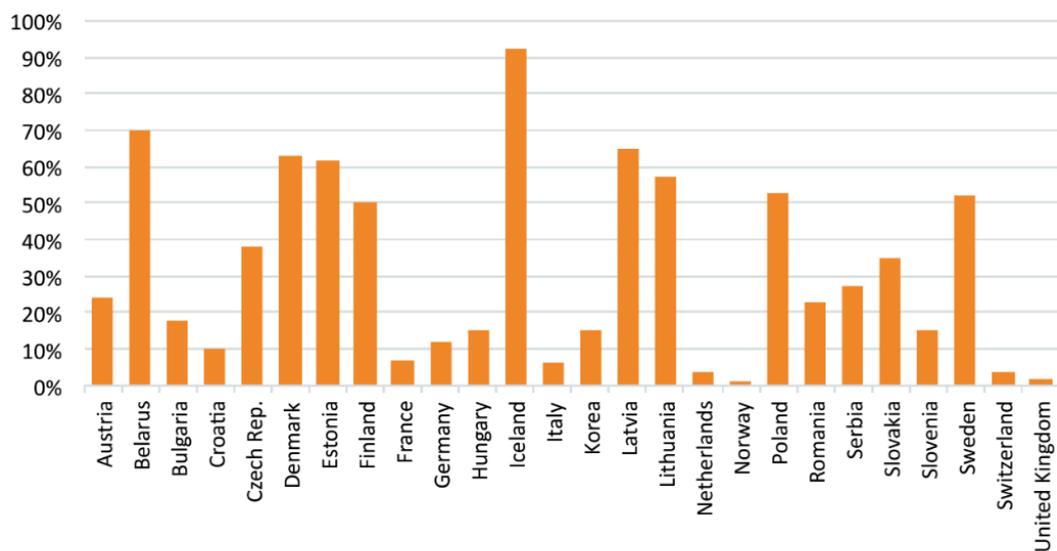


Figura 2.8 Percentuale abitanti che usufruisce del teleriscaldamento nelle nazioni europee. Fonte [3]

- Lunghezza della rete di trasporto e distribuzione:

Questo parametro è indipendente dall'energia annua consumata dagli utenti (a differenza delle vendite di calore per il teleriscaldamento) e quindi misura oggettivamente lo sviluppo di questa tecnologia in un dato Paese [3]. Analizzando la figura 2.9 si può notare la crescita della rete dal 2009 al 2013 nelle nazioni europee. L'Italia è il paese in maggior crescita con un incremento della lunghezza della rete del 58% [3].

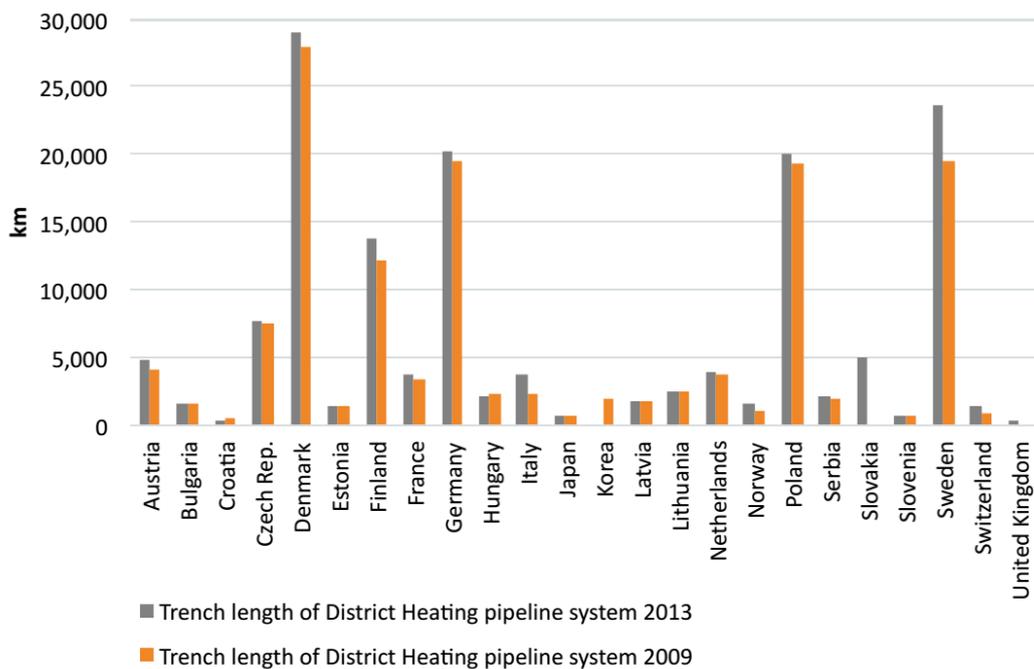


Figura 2.9 Lunghezza totale della rete di trasporto del teleriscaldamento per le nazioni europee negli anni 2009 e 2013. Fonte[3]

2.3 Teleriscaldamento in Italia e Torino

Il teleriscaldamento in Italia inizia a diffondersi dagli anni '70. La prima città ad essere teleriscaldata è Modena nel 1970, seguita da Brescia, dove nel 1971 viene attuato un progetto che oggi vede il 70% degli edifici residenziali e lavorativi allacciati alla rete di teleriscaldamento [7].

Altra prova della diffusione del teleriscaldamento in Italia è Reggio Emilia dove questa tecnologia nasce nel 1981 ed oggi vede metà delle abitazioni degli utenti allacciati alla rete [7].

Infine vi è la città di Torino che inizia ad essere teleriscaldata negli anni '80 ed oggi il 55% dei suoi abitanti beneficia del teleriscaldamento, mettendola tra i primi posti fra le città più teleriscaldate d'Europa [7].

Nel 2012, come riportato dal rapporto Legambiente del 2014 il teleriscaldamento è adottato in dieci regioni italiane che sono: Lombardia, Piemonte, Emilia Romagna, Trentino, Veneto, Liguria, Lazio, Toscana, Valle d'Aosta e Marche e le reti installate 192 in 150 città [8].

Di queste 150 città, 70, principalmente in Toscana e Trentino Alto Adige, utilizzano impianti a fonti rinnovabili. Nella prima viene sfruttata molto la geotermia mentre nella seconda le biomasse [8].

2.3.1 Teleriscaldamento a Torino

Come detto precedentemente Torino è la città più teleriscaldata d'Italia e IREN Energia è l'azienda che si occupa di tutto il teleriscaldamento.

La rete è in continua espansione come emerge dal fatto che è attualmente in corso l'allacciamento di circa 350 condomini che, una volta collegati alla rete, porteranno la volumetria sino a 85 milioni di metri cubi nella stagione corrente [10].

Il teleriscaldamento a Torino viene alimentato da tre impianti di cogenerazione situati due a Moncalieri e uno a Torino Nord con potenza elettrica installata complessiva di circa 1200 MWe e 740 MWth di potenza termica [9].

Nella centrale di Moncalieri sono presenti due impianti di cogenerazione a ciclo combinato con potenza elettrica complessiva di 800 MWe e potenza termica di 520 MWth. Essa presenta anche un impianto di integrazione e riserva di 141 MWth di potenza termica complessiva [9].

La centrale di Torino Nord è caratterizzata da 400 MWe e 220 MWth con un impianto di integrazione che ha una potenza termica installata di 340 MWth [9].

Lungo la rete sono presenti inoltre gli impianti di integrazione e riserva del Martinetto, del Politecnico di Torino e di Mirafiori Nord [9].

Nella figura 2.10 si può vedere l'area teleriscaldata di Torino.

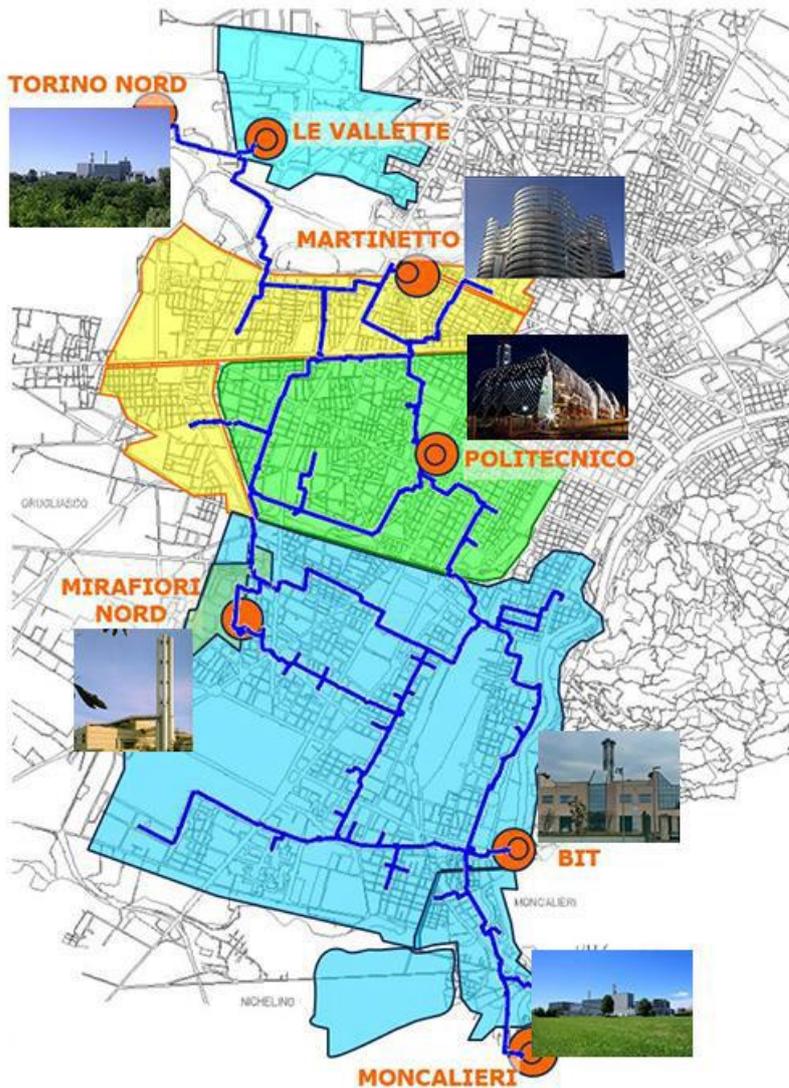


Figura 2.10 Schema dell'area teleriscaldata di Torino con i punti di produzione e integrazione. Fonte [9]

2.4 Quadro Normativo Italiano sul teleriscaldamento

Per molti anni il teleriscaldamento è rimasto privo di una disciplina normativa e di un'appropriate sistemazione giuridica [4].

Il teleriscaldamento viene definito per la prima volta con il decreto ministeriale del 24 ottobre 2005 "Direttive per la regolamentazione dell'emissione dei certificati verdi alle produzioni di energia di cui all'articolo 1, comma 71, della L. 23 agosto 2004 n. 239", nel quale si prevede che *l'«impianto di cogenerazione abbinato al teleriscaldamento è un sistema integrato, costituito dalle sezioni di un impianto di produzione combinata di energia elettrica e calore che rispettano i criteri definiti dall'Autorità per l'Energia Elettrica e il Gas ai sensi dell'articolo 2, comma 8, del decreto legislativo n. 79/1999, e da una rete di teleriscaldamento per la distribuzione del calore a una pluralità di edifici o ambienti per impieghi connessi prevalentemente con gli usi igienicosanitari e la climatizzazione, il riscaldamento, il raffrescamento, il condizionamento di ambienti»* [4].

In questo provvedimento vengono individuate alcune indispensabili caratteristiche proprie della rete:

- deve soddisfare la richiesta di una molteplicità di strutture;
- deve consentire l'allacciamento di ogni potenziale utenza (sempre nei limiti del sistema);
- l'erogazione dell'energia termica deve essere disciplinata mediante contratti di somministrazione e vendita [4].

Inoltre il decreto incentivava il risparmio energetico attraverso il rilascio di attestazioni denominati Certificati Verdi. La parte di energia termica

utilizzata per il teleriscaldamento che aveva diritto alla certificazione veniva calcolata con la formula sottostante [4]:

$$E_{cv} = H * C * T$$

Dove:

- E_{cv} rappresenta l'energia elettrica, espressa in MWh, che ha diritto ai certificati verdi;
- H è la parte di energia termica, espressa in MWh, utilizzata per il teleriscaldamento;
- C è un indice che dipende dal tipo di impianto cogenerativo;
- T è un indice che dipende dal tipo di intervento [4].

I valori assunti dagli indici C e T sono riassunti nelle tabelle 2.2 e 2.3.

Tipo di sezione	Valore dell'indice C
Ciclo combinato con recupero di calore	0,95
Termico a vapore con turbina a contropressione	0,45
Termico a vapore con turbina a condensazione	0,45
Turbina a gas con recupero di calore	0,55
Motore a combustione interna	0,75

Tabella 2.2 Valori assunti dall'indice C in base al tipo d'impianto. Fonte [4]

Tipo di intervento	Valore dell'indice T
Impianto di nuova costruzione, di cui all'articolo 2, comma 3, lettera <i>m</i>)	1,00
Rifacimento totale, di cui all'articolo 2, comma 3, lettera <i>j</i>)	1,00
Rifacimento parziale di cui all'articolo 2, comma 3, lettera <i>k</i>)	0,30
Realizzazione di una nuova rete con centrale esistente di cui all'articolo 2, comma 3, lettera <i>l</i>)	0,70

Tabella 2.3 Valori assunti dall'indice T in base alla tipologia di intervento. Fonte [4]

In questo modo, al teleriscaldamento, ogni 50 MWhel prodotti, veniva assegnato un Certificato Verde [4].

Con il successivo Decreto Ministeriale del 5 Settembre 2011 vengono introdotte ulteriori caratteristiche che la rete di teleriscaldamento deve possedere:

- la rete deve essere ampliata su terreni pubblici o privati ma non di proprietà esclusiva dell'operatore;
- l'allacciamento deve avvenire mediante contabilizzatori per misurare i flussi di energia termica;
- l'energia termica non deve essere distribuita direttamente all'operatore e si deve fondare su contratti di somministrazione e vendita [4].

Emerge come questo decreto, attraverso l'introduzione dei contabilizzatori, abbia rimarcato l'importanza della realizzazione di un controllo energetico [4].

Capitolo 3

MODELLO

3.1 Caratteristiche principali della richiesta termica giornaliera

In questo paragrafo, prima di descrivere la costruzione del modello si procederà all'identificazione delle caratteristiche principali della richiesta termica giornaliera, che rappresentano le fondamenta della realizzazione della previsione del profilo di richiesta.

Ciò viene fatto per poter prevedere attraverso il modello soltanto alcune grandezze invece di una funzione nel tempo, ottenendo così un modello compatto.

I dati di Input, come si può notare in figura 3.1, sono i dati Meteo e sono riferiti all'anno 2016-2017, più precisamente fra il 04/01/2016 e il 03/09/2017.

Dunque questo sarà il periodo temporale della nostra analisi.

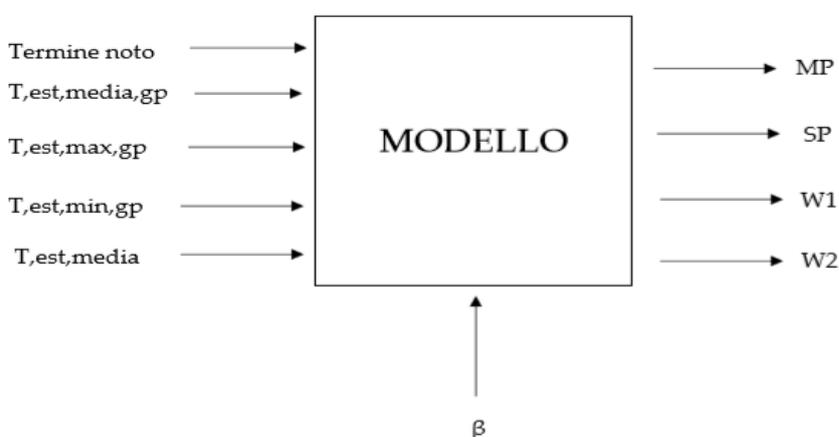


Figura 3.1 Modello

In particolare, studiando il giorno i -esimo, vengono inseriti nel modello i valori di temperatura media, massima e minima del giorno precedente a quello analizzato, il valore della temperatura media del giorno stesso e il termine noto.

Siccome le utenze, durante la giornata, possono accendersi e spegnersi più volte fino ad un massimo di tre, i punti cardine, necessari per determinare il profilo di richiesta, variano da un minimo di 5 fino ad un massimo di 15 e sono i seguenti:

1. Accensione;
2. Picco mattutino;
3. Inizio dello stazionario;
4. Fine dello stazionario;
5. Spegnimento.

Questi si ripetono nel caso di più accensioni e spegnimenti.

Gli output del modello, attraverso i quali è possibile ricavare l'ascissa e l'ordinata di questi punti, sono: l'ampiezza del picco (w_2), la distanza fra l'ascissa del picco e l'accensione (w_1), l'altezza del picco (MP) e dello stazionario (SP). Gli orari di accensione e spegnimento sono fissi per ogni giorno e quindi non costituiscono un output del modello.

A questo punto, unendo i punti appena descritti attraverso delle rette spezzate, è possibile determinare un'approssimazione del profilo di richiesta delle utenze.

Nella figura 3.2 viene mostrato l'andamento del profilo di richiesta reale e quello approssimato per un'utenza caratterizzata da tre accensioni in un tipico giorno invernale.

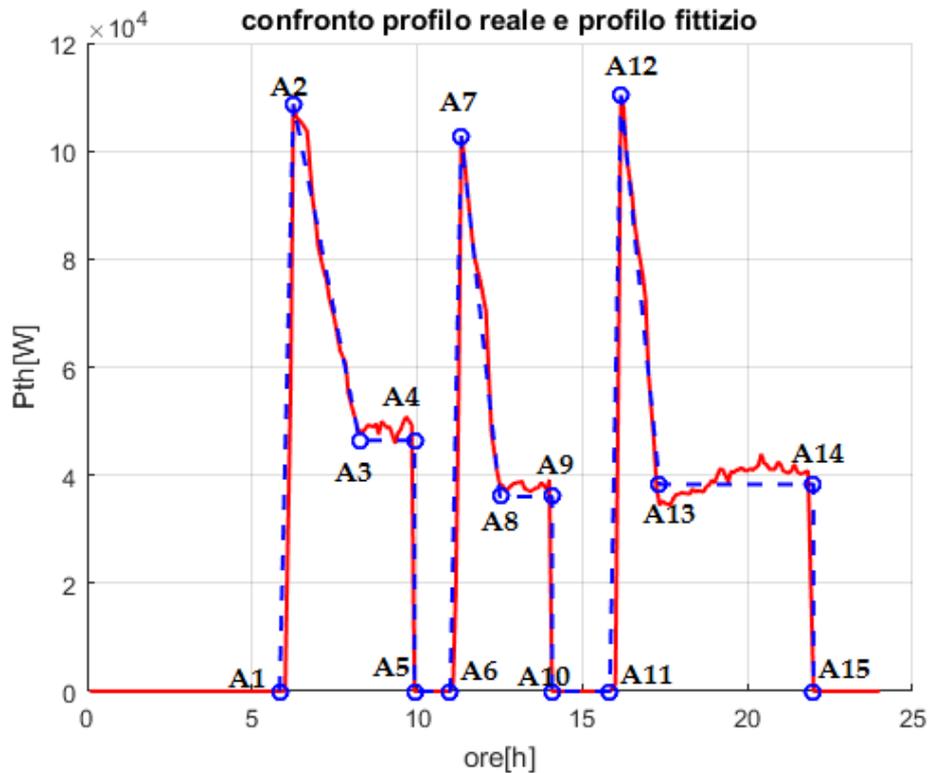


Figura 3.2 Confronto profilo richiesta reale e profilo richiesta fittizio

3.2 Caratteristiche del modello

Come detto precedentemente, per prevedere il profilo di richiesta si devono determinare i punti cardine sopra elencati.

Per fare ciò si parte dall'analisi dell'andamento del profilo di richiesta reale e successivamente attraverso il metodo di regressione lineare multipla multivariata è possibile ottenere il profilo di richiesta fittizio.

Il modello utilizzato viene costruito in più fasi e il software utilizzato è Matlab.

Inizialmente vengono raccolti, per ogni utenza analizzata, i dati relativi alla portata circolante nel circuito primario (G) e le temperature di mandata (T_1) e ritorno (T_2) sempre del primario. Questi valori sono ottenuti con un passo temporale di 5 minuti. La portata viene calcolata in m³/h mentre le temperature in °C.

Da questi dati è possibile determinare il flusso termico richiesto da ciascuna utenza usando la seguente formula:

$$\Phi = G * c_p * (T_1 - T_2) * \frac{\rho}{3600} \quad (3.1)$$

Dove ρ [kg/m³] è la densità dell'acqua (fluido termovettore utilizzato) e c_p [J/kg/K] il suo calore specifico. L'ultimo termine serve per convertire la portata da m³/h a kg/s ottenendo così il flusso termico espresso in W.

Dall'analisi del flusso termico, o della portata, è possibile risalire agli orari di accensione e spegnimento di ciascuna utenza. Essi quindi, come già anticipato nel paragrafo precedente, sono fissi e di conseguenza non rappresentano un output del modello.

Successivamente vi è la fase di pretrattamento dei dati, cioè vengono eliminati dall'analisi quei giorni in cui vi sono delle anomalie dovute ad errori nella fase di monitoraggio. Ad esempio può capitare che, dalla fase di raccolta dei dati, alcune utenze presentino più di dieci accensioni e spegnimenti e di conseguenza è chiaro che questi risultati derivino da errori commessi in questa fase.

Dopo la fase di pretrattamento vi è l'identificazione di alcune grandezze, nel profilo di richiesta reale, fondamentali per la determinazione dei punti cardine.

Per visualizzare meglio questa fase si prenda in considerazione l'andamento della domanda di una generica utenza, caratterizzata da tre accensioni e tre spegnimenti, rappresentato in figura 3.3

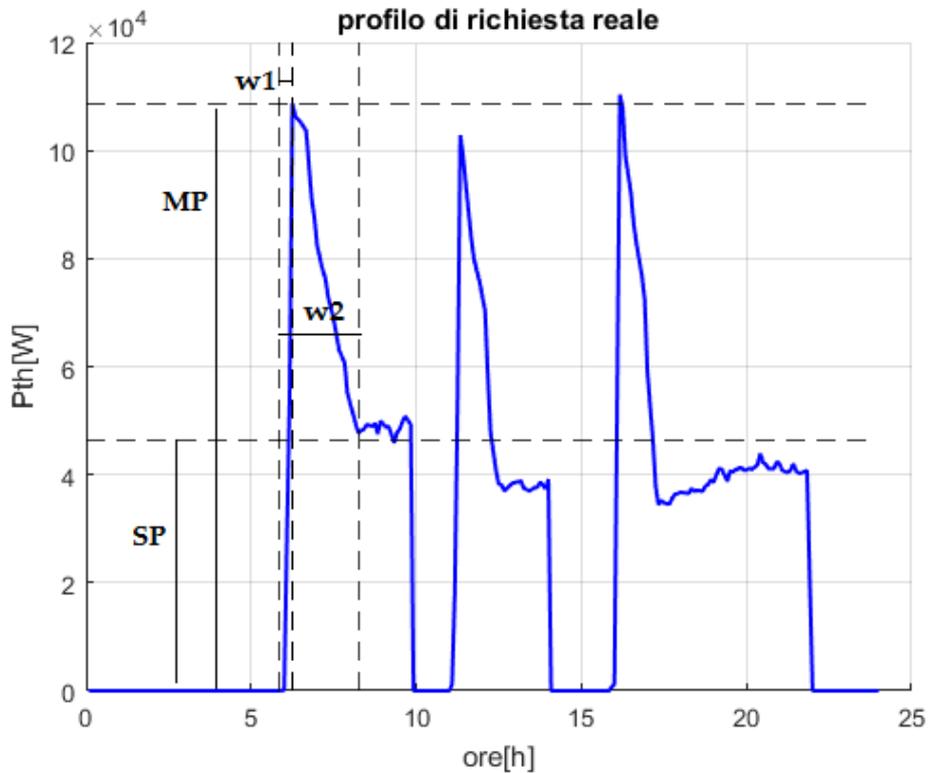


Figura 3.3 Profilo di richiesta reale di una generica utenza

Si può notare come vi siano tre parti fondamentali del profilo caratterizzate ciascuna da un'accensione, uno spegnimento, un picco di richiesta e una richiesta stazionaria.

Le grandezze da determinare per la prima accensione sono il MP, picco mattutino, SP, stazionario pomeridiano nella prima accensione, w_1 che rappresenta la distanza fra l'ascissa del picco mattutino (il tempo in cui avviene il picco) e la prima accensione e w_2 che è la larghezza totale del picco mattutino.

Per la seconda accensione occorre determinare il PP1, primo picco pomeridiano, SP1, stazionario pomeridiano nella seconda accensione, w_3

che è la distanza fra l'ascissa del primo picco pomeridiano e la seconda accensione e w_4 che è la larghezza totale del primo picco pomeridiano.

Infine per la terza accensione abbiamo il secondo picco pomeridiano (PP2), lo stazionario pomeridiano relativo alla terza accensione (SP2), la distanza fra l'ascissa del secondo picco pomeridiano e la terza accensione (w_5) e la larghezza totale del picco (w_6).

Giova chiarire il procedimento adottato per determinare queste grandezze.

Il picco mattutino è ottenuto calcolando il valore massimo assunto dal flusso termico fra la prima e la seconda accensione.

Prima di determinare lo stazionario pomeridiano relativo alla prima accensione è necessario calcolare l'istante di tempo in cui termina il picco ed inizia lo stazionario.

Per fare ciò è stata calcolata la derivata prima del flusso termico, a partire dal picco mattutino, rispetto al tempo.

Inizialmente essa assumerà valori negativi, dovuto al fatto che la curva è decrescente, per poi avere valore sempre maggiori fino ad arrivare a circa 0 in corrispondenza dello stazionario perché qui l'andamento è quasi orizzontale. Dunque analizzando i valori assunti dalla derivata è possibile determinare l'istante in cui inizia lo stazionario.

A questo punto il valore dello stazionario pomeridiano relativo alla prima accensione è ottenuto calcolando il valore medio del flusso termico fra l'inizio dello stazionario e il primo spegnimento.

W_1 è calcolato come differenza fra l'orario in cui avviene il picco mattutino e la prima accensione, mentre w_2 viene calcolata come differenza fra l'inizio dello stazionario e la prima accensione.

Per le grandezze relative alla seconda e terza accensione il procedimento è analogo.

Il modello è stato ottenuto applicando il metodo di regressione lineare multipla multivariata, quindi è stata imposta una correlazione lineare fra le variabili dipendenti ed indipendenti.

Le variabili dipendenti sono gli output del modello e quindi le grandezze sopra citate mentre le variabili indipendenti sono i dati di input.

A questo punto per determinare la funzione matematica che lega le variabili di input e quelle di output è necessario calcolare i coefficienti della regressione lineare, che chiameremo beta.

Essi sono ottenuti utilizzando il comando su matlab "mvregress" che li calcola utilizzando il metodo dei minimi quadrati.

Per ogni utenza viene fornita una matrice i cui elementi sono i coefficienti beta di regressione. Tale matrice avrà un numero di righe pari al numero dei dati di input (5 nel nostro caso) e un numero di colonne pari al numero di output (4, 8 o 12 a seconda che l'utenza abbia una, due o tre accensioni). Attraverso questo comando vengono fornite anche altre informazioni come la covarianza e i residui.

La prima grandezza indica una misura non standardizzata della correlazione fra le variabili dipendenti e indipendenti. In particolare se assume un valore diverso da zero allora le grandezze sono correlate. Più precisamente se la covarianza è positiva si dice che le variabili sono correlate positivamente (all'aumentare di una grandezza aumenta anche l'altra), se è negativa correlate negativamente (all'aumentare di una grandezza l'altra diminuisce) [13].

I residui invece sono definiti come la differenza fra il valore osservato e quello ottenuto con il modello. Come sopra accennato i coefficienti beta vengono determinati in modo tale che la retta di regressione minimizzi la somma dei residui al quadrato.

Una volta noti i coefficienti di regressione lineare è possibile ottenere le grandezze di output del modello attraverso la seguente equazione matriciale:

$$\beta' * T = Y \quad (3.2)$$

Dove T è un vettore colonna (5x1) contenente le variabili di input, beta è la matrice (5x4, 5x8 o 5x12) dei coefficienti di regressione e Y è il vettore (1x4, 1x8 o 1x12) contenente i risultati del modello, ossia la grandezze MP, SP, w1, w2, PP1, SP1, w3, w4, PP2, SP2, w5, w6.

A questo punto possono essere determinati i punti cardine per ottenere il profilo di richiesta fittizio. Essi sono calcolati come:

- A1(ACC1, 0);
- A2(ACC1+w1, MP);
- A3(ACC1+w2, SP);
- A4(SPEGN1, SP);
- A5(SPEGN1, 0).

Dove ACC1 e SPEGN1 sono rispettivamente la prima accensione e il primo spegnimento.

Per la seconda e terza accensione il procedimento è analogo.

Una volta determinati i punti cardine è possibile rappresentare il profilo di richiesta approssimato.

In alcuni casi però, a causa di mancanza di dati, non è possibile ricavare i coefficienti beta e quindi il profilo di richiesta.

Infatti lo scopo che ci si prefigge consiste proprio nella costruzione di un modello che sia in grado, attraverso i pochi dati disponibili, di determinare tali coefficienti per poter quindi prevedere il profilo di richiesta quando ignoto.

3.3 Determinazione dei coefficienti Beta e previsione dei profili di richiesta

Come anticipato nel precedente paragrafo, è possibile che per alcune utenze, a causa di mancanza di dati, non sia possibile ottenere i coefficienti beta, che sono fondamentali per la determinazione del profilo di richiesta.

In questo paragrafo viene mostrato il procedimento per la creazione di un modello atto ad ottenere tali coefficienti.

Le utenze analizzate sono quelle dei baricentri 408,410,414,418 e 420 (per un totale di 212 utenze) per le quali è noto lo storico dei dati (quindi anche i coefficienti beta) e di conseguenza è stato realizzato un confronto fra i profili fittizi e quelli determinati dal modello studiato, al fine di verificarne l'attendibilità.

Il metodo utilizzato è quello della regressione lineare.

3.3.1 Regressione lineare

La regressione è un procedimento statistico che lega una variabile dipendente y con una o più variabili indipendenti $x_1, x_2, x_3, \dots, x_n$ di cui si conosce il valore.

Per costruire il modello statistico in grado di prevedere y si deve analizzare il tipo di relazione funzionale fra le variabili dipendenti e le variabili indipendenti e quindi determinare la funzione matematica che le lega [11].

In questo studio è stata presa in considerazione una dipendenza, fra le variabili, di tipo lineare.

In base al numero di variabili dipendenti e indipendenti è possibile distinguere tre diverse tipologie di regressione lineare:

- Regressione lineare semplice, caratterizzata da una sola variabile in ingresso e una sola variabile in uscita;
- Regressione lineare multipla, caratterizzata da una sola variabile dipendente e più variabili indipendenti;
- Regressione lineare multipla multivariata, dove il numero di variabili dipendenti e indipendenti è maggiore di uno [12].

In questa tesi è stata utilizzata una regressione lineare multipla multivariata.

Nel caso generale in cui le variabili indipendenti fossero x_1, x_2, \dots, x_n , la variabile dipendente y viene calcolata attraverso la seguente equazione:

$$y_i = a + \gamma_1 * x_1 + \gamma_2 * x_2 + \dots + \gamma_n * x_n + \varepsilon \quad (3.3)$$

Dove il generico γ rappresenta il coefficiente di regressione lineare che viene calcolato attraverso il metodo dei minimi quadrati, cioè minimizzando la seguente quantità:

$$\sum_i (y_i - y_i^*)^2 \quad (3.4)$$

Dove:

- y_i è il valore osservato della variabile dipendente y ;
- y_i^* è il valore stimato attraverso il metodo di regressione multipla [12].

Attraverso questi coefficienti di regressione, che dipendono dall'unità di misura delle variabili indipendenti, è possibile stimare di quanto la variabile dipendente si modifichi per ogni incremento unitario della variabile indipendente [12].

3.3.2 Costruzione modello

La costruzione del modello prevede diverse fasi e il software utilizzato è Matlab.

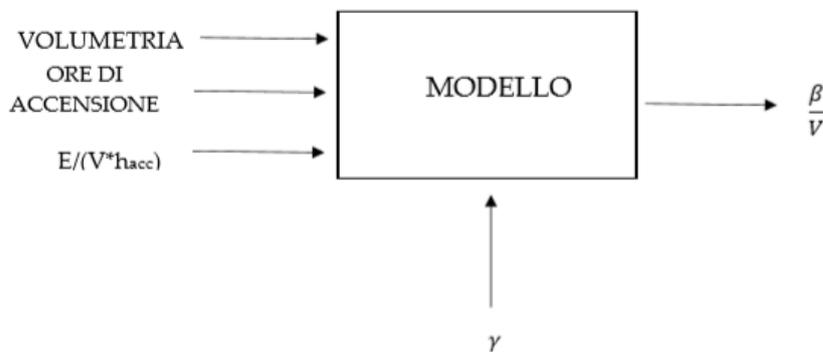


Figura 3.4 Modello

Inizialmente, visto che per ottenere il modello sono state adoperate utenze per cui è noto lo storico dei dati, è stata eseguita la raccolta degli stessi, in particolare la volumetria di ciascuna utenza, la portata circolante nel circuito primario con le relative temperature di mandata e di ritorno. I dati sono stati raccolti per due anni consecutivi, dal 2015 al 2017, escludendo le estati, con un time-step di 5 minuti. Dunque, più precisamente, l'arco temporale analizzato comprende il periodo dal primo Gennaio 2015 fino al 30 Aprile dello stesso anno per poi riprendere dal primo ottobre 2015 fino alla fine dell'anno (31/12/2015). Questo ciclo è ripetuto anche per il 2016 e per il 2017.

Attraverso questi valori come spiegato nel Capitolo 3.2 è possibile ottenere la potenza termica richiesta da ciascuna utenza, gli orari di accensione e spegnimento.

Inoltre per ogni utenza vengono forniti i coefficienti beta che saranno utilizzati per costruire il modello e per fare successivamente un confronto con quelli ottenuti da questa analisi.

Come chiarito nel Capitolo 3.2 la matrice beta può essere 5×4 , 5×8 o 5×12 in base al numero di accensioni dell'utenza. In questo studio però è necessario scomporre tale matrice in un vettore riga che potrà avere di conseguenza un numero di colonne pari a 20, 40 o 60 (5×4 , 5×8 , 5×12).

Successivamente è stato effettuato il pretrattamento dei dati poiché come evidenziato nel Capitolo 3.2 essi possono essere affetti da errori.

Per prima cosa quindi si è scelto di analizzare il giorno che presenta il numero maggiore di utenze con dati accettabili ovvero il 07/01/2016 in cui sono presenti 156 utenze "accettabili" su 212.

Dopo aver determinato il giorno di analisi, dalla conoscenza del flusso termico, è stato possibile determinare l'energia richiesta annualmente per ciascuna utenza.

L'energia termica infatti è stata ottenuta calcolando l'integrale della potenza termica nel tempo.

In seguito, noti gli orari di accensione e spegnimento, è stato possibile determinare il numero medio di ore di accensione di ciascuna utenza.

Queste due grandezze, energia termica richiesta annualmente e ore di accensione, insieme alla volumetria, rappresentano i dati di input del modello mentre i coefficienti beta sono gli output come si può osservare nella figura 3.4.

Più precisamente, visto che il fabbisogno di energia termica dipende dalla volumetria e dal numero di ore di accensione dell'utenza, si è pensato di non utilizzare come dato di input l'energia termica richiesta annualmente, ma il rapporto fra questa quantità e il prodotto fra il volume e le ore di accensione. Invece per quanto riguarda i valori di output si è deciso di determinare il rapporto fra i coefficienti beta e la volumetria. In particolare solo i parametri riguardanti le ordinate del picco e della richiesta termica stazionaria sono state divise per il volume al fine di adimensionalizzare questi coefficienti e di conseguenza ridurre la loro dipendenza dalla volumetria.

Fra le utenze prese in esame alcune sono state utilizzate per costruire il modello (81), altre invece per testarlo.

Il modello si basa sulla regressione lineare, quindi vi è un relazione lineare fra le variabili di input e output. L'obiettivo quindi è quello di determinare i coefficienti di regressione lineare che chiameremo gamma.

Inizialmente sono state identificate, fra le utenze utilizzate per costruire il modello, quelle con una sola accensione, quelle con due e infine quelle con tre accensioni.

Successivamente si è pensato di calcolare i coefficienti gamma tre volte; dapprima per determinare i coefficienti beta per la prima accensione, successivamente per ottenere quelli della seconda accensione ed infine per calcolare quelli della terza accensione.

Nel primo caso tutte le 81 utenze sono state analizzate, in quanto ognuna di esse presentava la prima accensione. Nel secondo solo le utenze con due o tre accensioni (45 su 81) sono state utilizzate, in quanto solo queste presentavano la seconda accensione. Infine nell'ultimo caso sono state studiate solo le utenze che presentavano tre accensioni nella giornata analizzata (23 su 81).

Questi coefficienti gamma sono stati calcolati utilizzando il comando di matlab "mvregress", che ci fornisce anche informazioni sulla covarianza e i residui.

La matrice contenente i coefficienti di regressione lineare gamma ha un numero di righe pari al numero dei dati di input, quindi 3, e un numero di colonne pari al numero di output.

Nel caso in analisi i coefficienti beta vengono calcolati tre volte, una per ogni accensione, quindi l'output del modello sarà un vettore riga 1×20 . Di conseguenza la matrice gamma avrà un numero di colonne pari a 20 e sarà quindi una matrice 3×20

Data la dimensione della matrice occorre che il numero minimo di utenze per costruire il modello sia maggiore di 20. Quindi è necessario avere

almeno 20 utenze che abbiano la terza accensione, altrimenti la matrice della covarianza risulta non definita positiva.

Una volta ottenute le tre matrici gamma è possibile calcolare i coefficienti beta per ciascuna accensione e successivamente unirli in un'unica matrice.

L'equazione per determinare il generico coefficiente β_i è la seguente:

$$\beta_i = a + \gamma_{1,i} * V + \gamma_{2,i} * \frac{E}{V * h_{acc}} + \gamma_{3,i} * h_{acc} \quad (3.5)$$

Dove V, E e h_{acc} sono la volumetria, l'energia richiesta annua e le ore di accensione giornaliere dell'utenza analizzata.

Questo procedimento viene ripetuto tre volte, utilizzando ogni volta una matrice gamma differente.

Nel primo caso il risultato è una matrice 212x20, cioè una matrice avente un numero di righe pari al numero totale delle utenze e un numero di colonne pari al numero di output.

Nel secondo invece il risultato è costituito da una matrice 122x20, in quanto il numero di output è 20 e il numero di utenze che si accendono almeno due volte nella giornata è pari a 122.

Infine dall'ultimo calcolo si ottiene una matrice 58x20 in quanto le utenze che si accendono tre volte in una giornata sono 58.

A questo punto queste tre matrici vengono unite ottenendone una di dimensioni 212x60. Per le utenze che non hanno la seconda e/o la terza accensione gli elementi dalla colonna 20 in poi saranno nulli.

Nelle figure 3.5 e 3.6 è rappresentato l'errore relativo rispetto ad alcune grandezze, in particolare il picco mattutino e la richiesta stazionaria

pomeridiana, che si commette utilizzando i coefficienti beta ottenuti dal modello rispetto a quelli forniti, applicando la seguente formula:

$$relerr = \frac{|Grandezza_{beta\ fornito} - Grandezza_{beta\ da\ modello}|}{|Grandezza_{beta\ fornito}|} \quad (3.6)$$

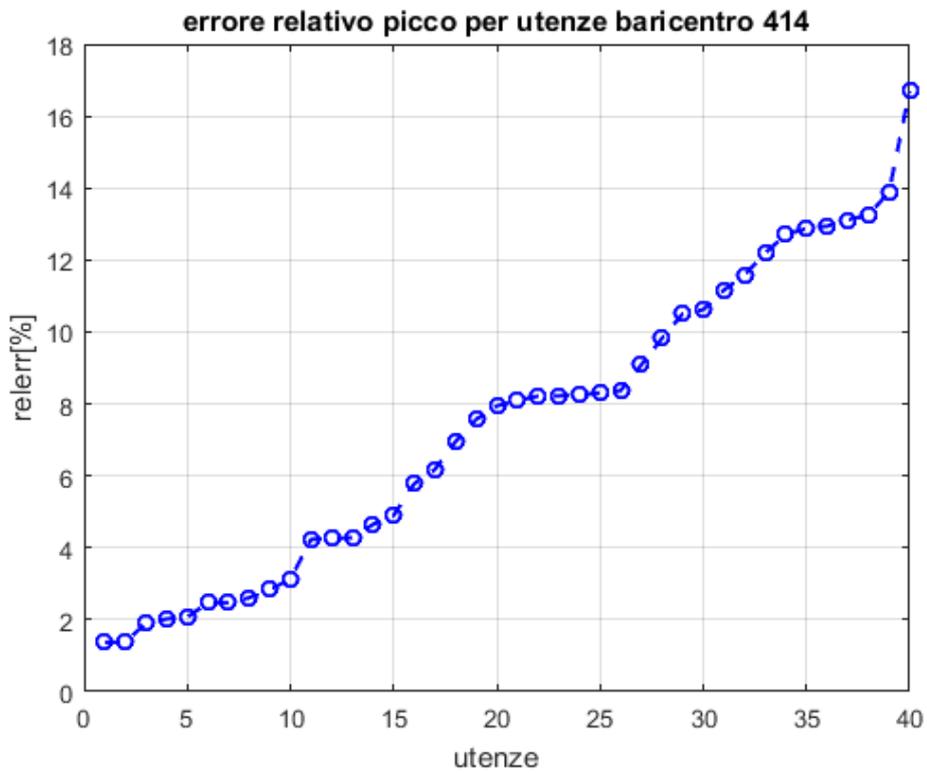


Figura 3.5 Errore relativo calcolato sul picco

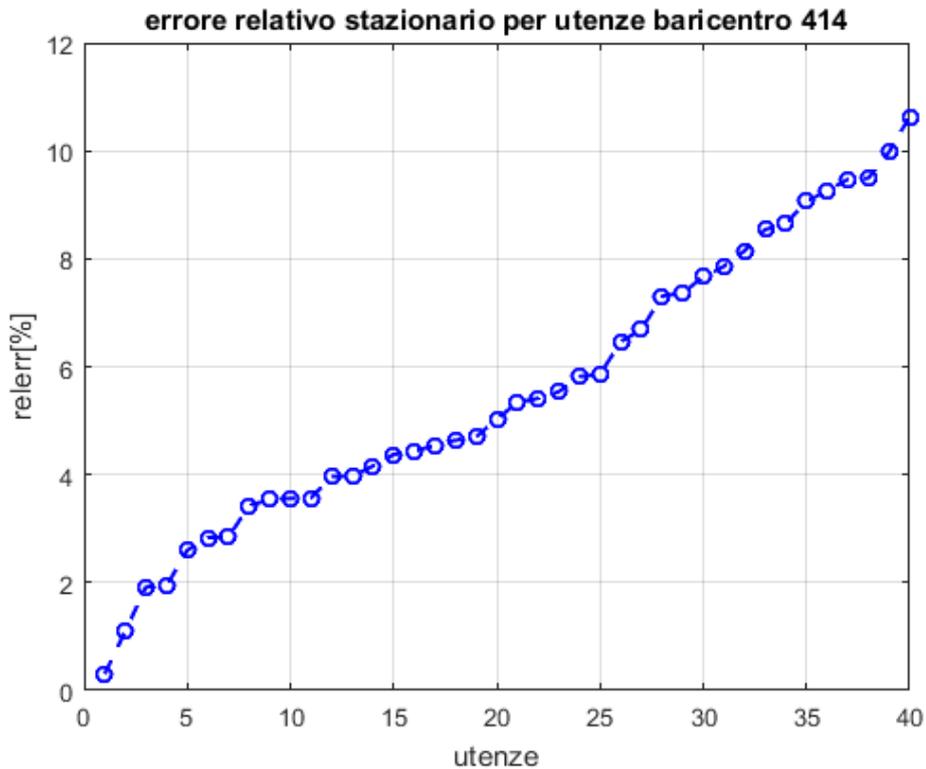


Figura 3.6 Errore relativo calcolato sulla richiesta stazionaria

Più precisamente nella prima figura viene rappresentato l'errore rispetto al picco mattutino mentre nella seconda figura quello relativo alla richiesta stazionaria e in entrambi i casi gli errori sono stati disposti in ordine crescente. In questa analisi sono state prese in considerazione alcune utenze del baricentro 414.

Una volta determinati i coefficienti beta è possibile ottenere il profilo di richiesta approssimato utilizzando il modello descritto nel Capitolo 3.2.

Più precisamente prima di poter applicare il modello è necessario modellare la matrice dei coefficienti. Infatti, per ciascuna utenza, i coefficienti sono raccolti in vettori riga (1x20, 1x40 o 1x60) mentre per il modello devono essere utilizzate matrici 5x4, 5x8 o 5x12.

Una volta scalate le matrici, inserendo i dati meteo è possibile ottenere il profilo di richiesta.

3.4 Miglioramento delle prestazioni del modello

In questo paragrafo si eseguirà un'ottimizzazione del modello, cioè esso verrà perfezionato al fine di migliorare la previsione dei profili di richiesta delle utenze.

Il procedimento effettuato è un'analisi di Clustering che, come si vedrà nei paragrafi successivi, porterà alla suddivisione delle utenze in più gruppi e alla creazione di più modelli, ciascuno per ogni gruppo.

3.4.1 Clustering

Il Clustering, che deriva dal verbo inglese "to cluster" che letteralmente significa "raggrupparsi" [14], è un insieme di metodi statistici che hanno come fine la creazione di uno o più gruppi di dati [15].

Pertanto, l'obiettivo dell'analisi è quello di raggruppare i dati studiati in gruppi di talché gli elementi di ciascun gruppo siano "simili" fra loro, secondo alcune leggi, mentre i gruppi risultino "diversi" [15].

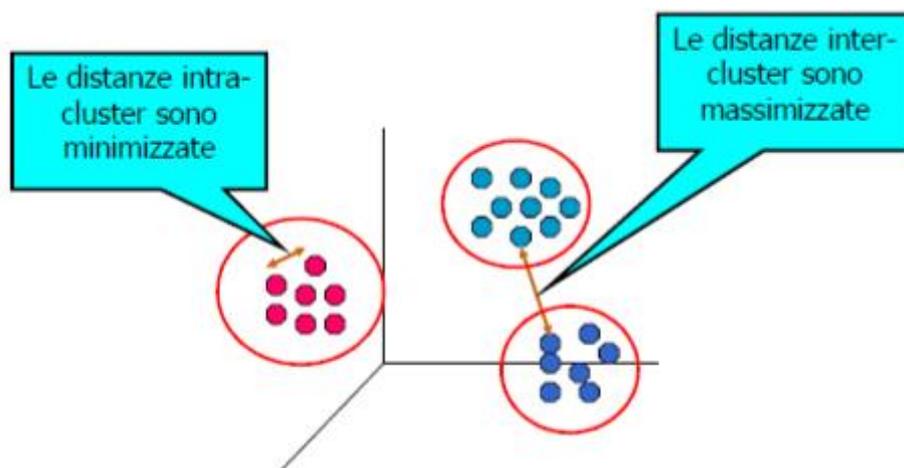


Figura 3.7 Rappresentazione grafica del Clustering. Fonte [16]

Nella figura 3.7 si può vedere come i dati vengono suddivisi nei vari Cluster (3 in questo caso).

L'analisi di Clustering può essere effettuata in diversi ambiti come la finanza, la biologia, la pianificazione territoriale, l'analisi del web ecc. [15].

I metodi di Clustering, per poter funzionare correttamente, devono possedere determinate caratteristiche fondamentali come:

- la capacità di trattare un gran numero di variabili;
- la capacità di trattare variabili di diverse tipologie (es. continue, categoriche);
- la capacità di trattare con *outlier*, dati che assumono valori al di fuori della norma, che devono essere eliminati;
- interpretabilità dei gruppi [15].

Tuttavia il Clustering presenta alcune problematiche fra cui, le più rilevanti sono:

- in alcune occasioni, i metodi di Clustering non soddisfano tutti i requisiti sopra menzionati;
- esistono diversi modi per calcolare la distanza tra elementi e gruppi, in base alla provenienza dei dati;

- l'utilizzo di metodi diversi può determinare la formazione di Cluster differenti e talvolta non è semplice individuare la suddivisione migliore;
- i risultati possono essere interpretati diversamente;
- il costo computazionale è elevato quando si lavora con una grande quantità di dati [15].

La prima fase dell'analisi di Clustering consiste nel pretrattamento dei dati. Infatti è possibile la presenza di dati rumorosi, *outlier*, dati mancanti o duplicati di osservazione [15].

In base al tipo di problema è possibile intervenire in diverse maniere; ad esempio gli *outlier* devono essere eliminati mentre i dati mancanti possono essere sostituiti o non essere presi in considerazione durante l'analisi [15].

In questa fase, inoltre, si deve prestare attenzione anche al tipo di variabili analizzate, in particolar modo alle loro unità di misura. È importante, infatti, effettuare una conversione tra scale di valori [15].

Come detto precedentemente, affinché un algoritmo di Clustering funzioni correttamente, la distanza fra i vari Cluster deve essere massimizzata mentre quella fra i dati all'interno di ciascun gruppo minimizzata [15].

Una grandezza utilizzata per capire la bontà di un algoritmo è la seguente:

$$R^2 = 1 - \frac{\sum_{j=1}^K \sum_{i \in C_j} \|x_i - \bar{x}_j\|^2}{\sum_{i=1}^n \|x_i - \bar{x}\|^2} \quad (3.7)$$

Dove K è il numero di gruppi C_j , n è il numero totale dei dati, \bar{x}_j è la media del Cluster C_j e \bar{x} è la media dei dati [15].

Affinché un algoritmo sia efficiente R^2 deve tendere a 1 per avere la varianza nei Cluster prossima a 0 [15].

Esistono diversi algoritmi di Clustering e la loro classificazione dipende dal tipo di Cluster che essi generano.

Questi sono:

- algoritmo esclusivo in cui un dato può appartenere ad un solo Cluster;
- algoritmo sovrapposto in cui un dato può appartenere a più Cluster;
- algoritmo gerarchico che si basa sull'unione di Cluster vicini fra loro;
- algoritmo probabilistico in cui la suddivisione dei Cluster avviene utilizzando un metodo probabilistico [15].

In questo studio sono stati applicati due algoritmi di Clustering: *K-means*, che è un algoritmo esclusivo, e un metodo gerarchico.

3.4.1.1 K-means

In questo algoritmo il numero di Cluster viene scelto *a priori*. Ogni gruppo è caratterizzato da un centroide e i dati vengono suddivisi nei vari Cluster in base alla loro distanza dal centroide stesso. Dopo un primo frazionamento nei vari gruppi si ricalcola il centroide, che tipicamente è il momento primo, e si procede nuovamente alla suddivisione dei dati. Questo procedimento viene ripetuto finché non si arriva ad avere centroidi fissi, minimizzando la seguente funzione obiettivo [15]:

$$J = \sum_{j=1}^K \sum_{i=1}^{n_j} \|x_i - c_j\|^2 \quad (3.8)$$

Il problema principale di questo algoritmo è la scelta dei centroidi iniziali, da cui dipende il risultato finale. Generalmente essi vengono scelti in maniera casuale [15].

Per risolvere questo problema si può pensare di ripetere più volte l'algoritmo cambiando i centroidi iniziali oppure calcolare quest'ultimi utilizzando un algoritmo gerarchico [16].

Il costo computazionale di questo metodo è pari a $O(tkmn)$, dove m è il numero di attributi, n il numero di dati analizzati, k il numero di gruppi, e t il numero di iterazioni. In genere i valori di k , m e t sono molto minori di n [17] [18].

3.4.1.2 Algoritmo gerarchico

Esistono due diversi tipi di algoritmi gerarchici:

- algoritmo agglomerativo in cui ogni dato rappresenta un Cluster e ad ogni iterazione i due Cluster che sono più vicini fra loro vengono uniti in un unico gruppo. Si procede fino ad ottenere un unico Cluster [15] [16];
- algoritmo divisivo in cui viene eseguito il procedimento inverso. Partendo da un unico Cluster, i dati vengono separati fino ad ottenere un numero di Cluster pari al numero totale di elementi [16].

In questa tesi è stato utilizzato un algoritmo gerarchico agglomerativo.

La distanza fra i vari Cluster può essere definita in diverse maniere:

- single-linkage: la distanza è la minima distanza tra gli elementi dei due cluster;
- complete-linkage: la distanza è la massima distanza tra gli elementi dei due Cluster;
- average-linkage: la distanza è la media delle distanze tra gli elementi dei due Cluster;

- aediana: la distanza è la distanza pesata dei centroidi dei due Cluster;
- centroidi: la distanza è la distanza dei centroidi dei due Cluster [15];
- ward: utilizza lo scarto quadratico medio [16].

In figura 3.8 è rappresentato un dendrogramma che è la rappresentazione grafica del risultato finale di questo algoritmo.

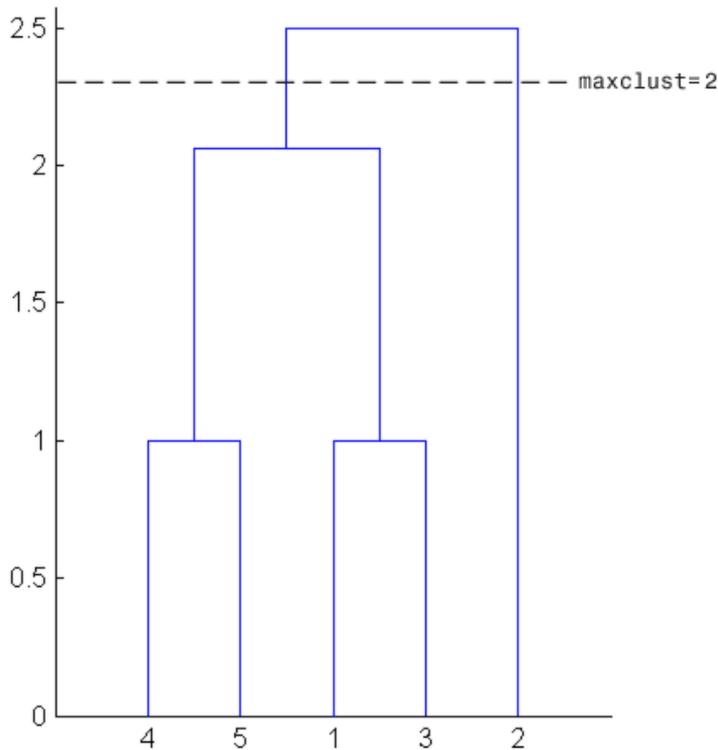


Figura 3.8 Dendrogramma

Si può notare che il numero di Cluster non deve essere definito a priori e può essere determinato troncando il dendrogramma al livello desiderato [16]. Nel caso illustrato in figura 3.9 il numero di Cluster voluto è pari a due; il primo costituito dai dati 4, 5, 1 e 3 mentre il secondo dal dato 2.

Un problema che si può riscontrare, applicando questo algoritmo, è la presenza di "tie", ovvero l'esistenza di due (o più) dati che si trovano alla stessa distanza da un Cluster. Ciò impone una scelta che darà luogo a due

(o più) possibili suddivisioni con la precisazione che non si potrà conoscere a priori quale sia l'opzione migliore [15].

Il costo computazionale di questo algoritmo è quadratico rispetto al numero di dati da analizzare $O(n^2)$ [17].

3.4.2 Applicazione del Clustering

In questo paragrafo viene mostrata l'applicazione dei metodi di Clustering, spiegati nel paragrafo precedente, per la costruzione del modello al fine di determinare i coefficienti di regressione lineare beta.

Il modello viene costruito in diverse fasi e il software utilizzato è Matlab.

Per le utenze analizzate è noto lo storico dei dati ed esse sono le medesime utilizzate per la costruzione del modello descritto nel Capitolo 3.3.2.

Inizialmente vi è una fase di raccolta dati seguita poi da una successiva fase di pretrattamento degli stessi.

Queste fasi, così come la determinazione degli orari di accensione e spegnimento, l'energia richiesta annualmente e il numero medio di ore di accensione di ciascuna utenza, sono analoghe a quelle rappresentate nel Capitolo 3.3.2.

Di conseguenza i dati di input e output del modello rimangono invariati.

L'unica differenza consiste nel fatto che le utenze in questa trattazione non verranno analizzate tutte insieme ma suddivise in Cluster attraverso algoritmi di Clustering.

Si è scelto di raggruppare le utenze in soli 3 Cluster in quanto non sono disponibili in elevata quantità.

Utilizzando l'algoritmo K-means un gruppo è costituito da 92 utenze, il secondo da 84 e il terzo da 36.

Invece, quando viene applicato l'algoritmo gerarchico, i tre Cluster sono formati da 92, 82 e 38 utenze. In figura 3.9 è rappresentato il dendrogramma.

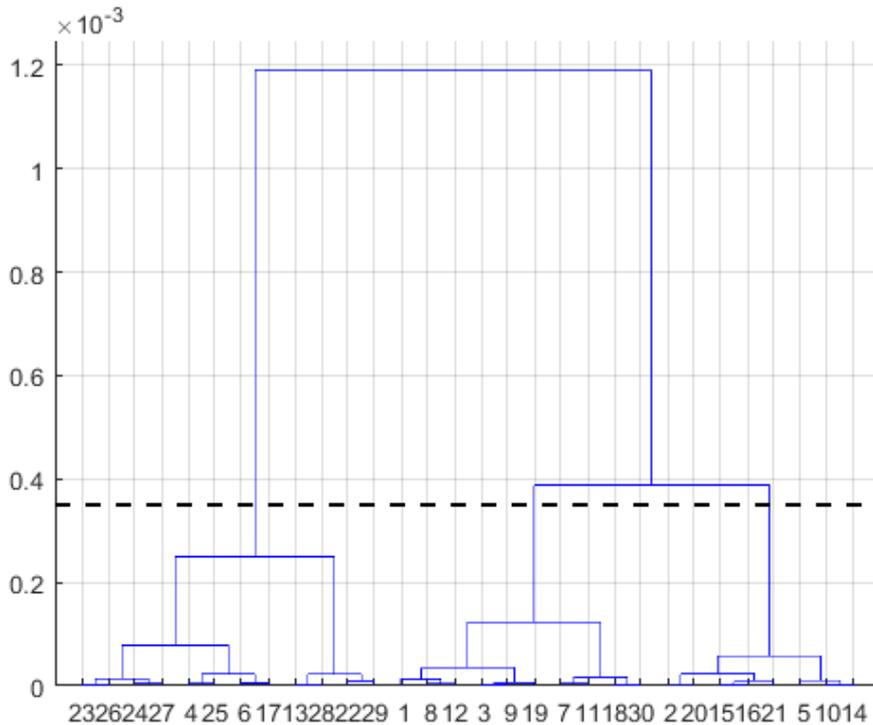


Figura 3.9 Dendrogramma

Sull'asse delle ascisse sono riportate le utenze mentre sull'asse delle ordinate la distanza fra i vari Cluster.

In quest'ultimo caso la distanza fra i cluster è stata imposta in misura pari alla average-linkage.

In entrambi i casi si è proceduto alla creazione di un modello per ogni cluster, analizzando così utenze simili fra loro.

Parte delle utenze sono state adoperate per costruire il modello altre per testarlo.

Il metodo utilizzato è sempre la regressione lineare multipla multivariata e quindi è necessario determinare i coefficienti gamma di regressione.

Si è pensato, a differenza del caso precedente, di calcolare i coefficienti gamma solo due volte.

Dapprima, per ottenere i coefficienti beta relativi alla prima accensione e, successivamente, per ottenere quelli relativi alla seconda. Infine, per calcolare i coefficienti beta relativi alla terza accensione sono stati utilizzati i coefficienti gamma relativi alla seconda.

Questo perché, come spiegato nel Capitolo 3.3.2, per costruire il modello è necessario avere un numero di utenze con la terza accensione maggiore di 20. Questo vale per ogni cluster per cui, fra le 212 utenze indagate, devono essere presenti almeno 60 utenze con la terza accensione.

Nel caso analizzato però questo non accade in quanto tali utenze sono 58 e dunque non è possibile calcolare i coefficienti gamma per determinare i rispettivi beta per la terza accensione.

Tuttavia, osservando i profili di richiesta delle utenze, si è notato che l'andamento del flusso termico per la seconda accensione è simile a quello della terza e quindi si è deciso di utilizzare la seconda matrice dei gamma per ottenere i coefficienti beta relativi alla terza accensione.

Uno dei motivi principali per cui si è deciso di suddividere i dati in soli tre Cluster è il fatto di avere un numero limitato di utenze con la terza accensione.

Le utenze utilizzate per costruire il modello sono 23 per ogni Cluster e ciascuna di esse presenta almeno due accensioni.

I coefficienti gamma sono stati calcolati utilizzando il comando di Matlab "mvregress" e il procedimento per la determinazione delle matrici beta è lo

stesso descritto nel Capitolo 3.3.2, solo che qui viene ripetuto per ogni cluster.

In figura 3.10 e 3.11 è rappresentato l'errore relativo rispetto ad alcune grandezze, in particolare il picco mattutino e la richiesta stazionaria pomeridiana, che si commette utilizzando i coefficienti beta ottenuti dal modello rispetto a quelli forniti, applicando la seguente formula:

$$relerr = \frac{|Grandezza_{beta\ fornito} - Grandezza_{beta\ da\ modello}|}{|Grandezza_{beta\ fornito}|} \quad (3.9)$$

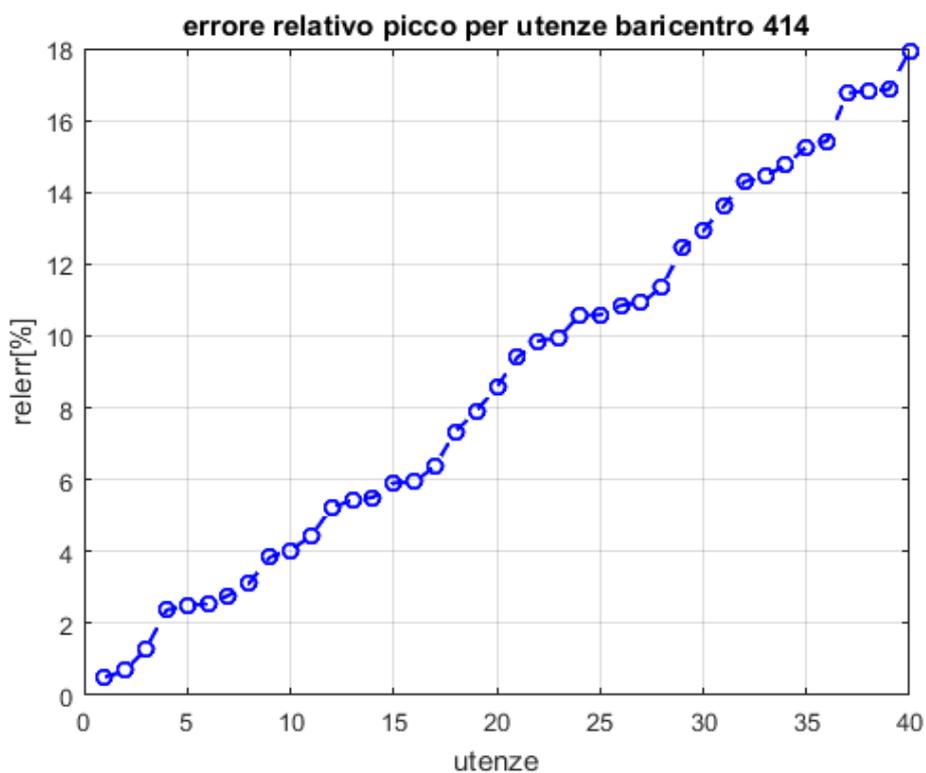


Figura 3.10 Errore relativo al picco mattutino applicando l'algoritmo K-means

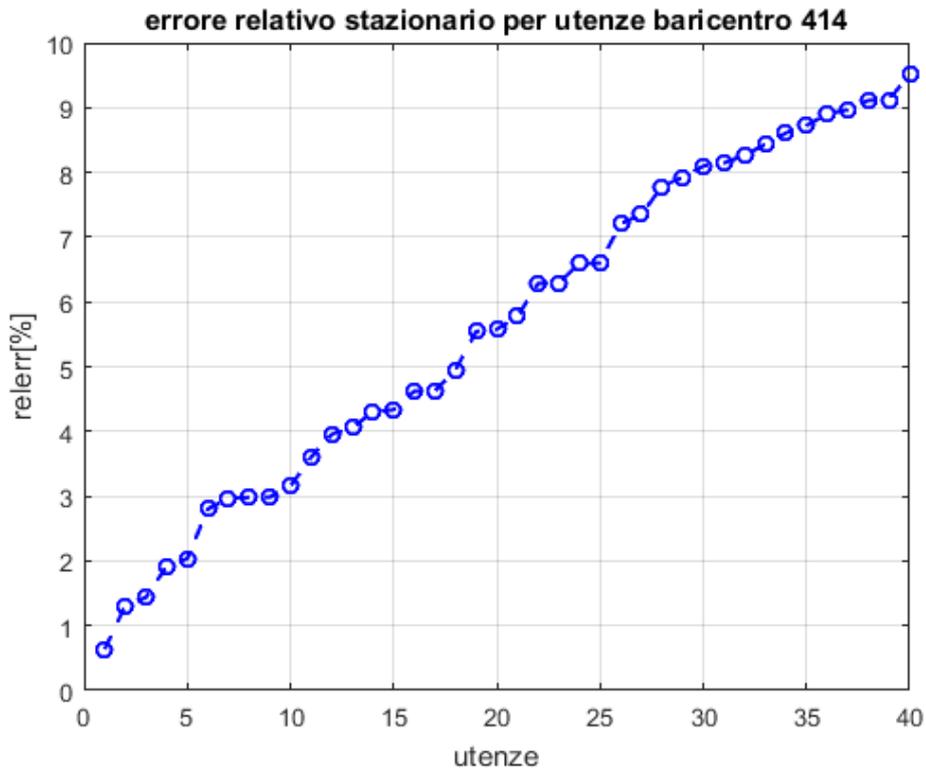


Figura 3.11 Errore relativo alla richiesta stazionaria applicando l'algoritmo gerarchico

In particolare nella prima figura è rappresentato l'errore commesso sul picco mattutino utilizzando l'algoritmo K-means, mentre nella seconda figura l'errore relativo alla richiesta stazionaria applicando l'algoritmo gerarchico e in entrambi i casi gli errori sono stati disposti in ordine crescente. Anche in questa analisi, come quella eseguita nel Capitolo 3.3.2, sono state analizzate utenze appartenenti al baricentro 414.

Capitolo 4

RISULTATI

4.1 Confronto dei risultati ottenuti

In questo paragrafo verranno messi a confronto i profili di richiesta ottenuti con le tre tecniche spiegate nel capitolo precedente.

Nella figura 4.1 viene illustrato il profilo di richiesta determinato senza l'utilizzo degli algoritmi di Clustering, mentre nelle figure 4.2 e 4.3 sono raffigurati i profili ottenuti utilizzando l'algoritmo K-means e l'algoritmo gerarchico.

In tutte e tre le figure si può notare il confronto con il profilo di richiesta fittizio.

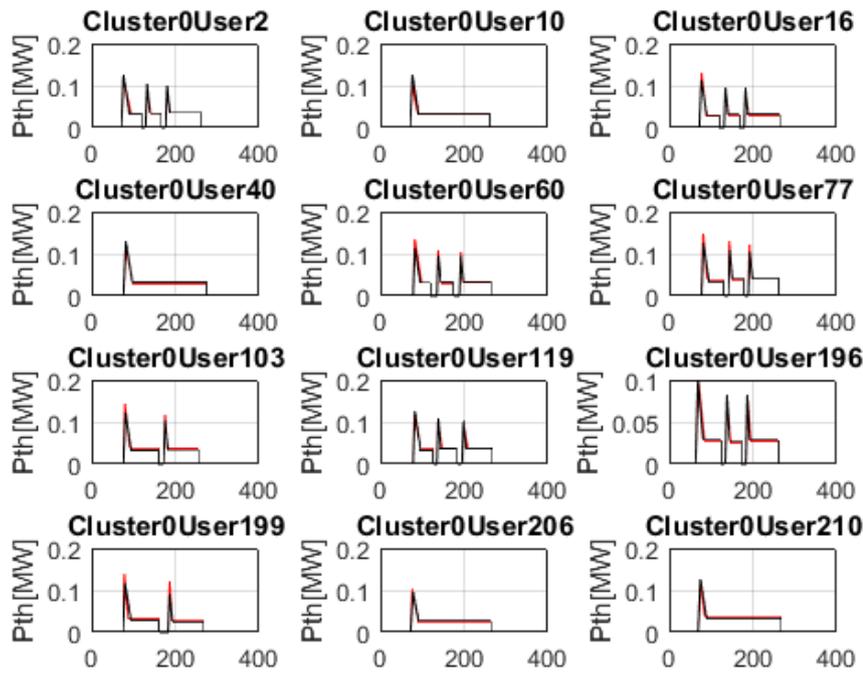


Figura 4.1 Confronto profili di richiesta: Caso senza Clustering

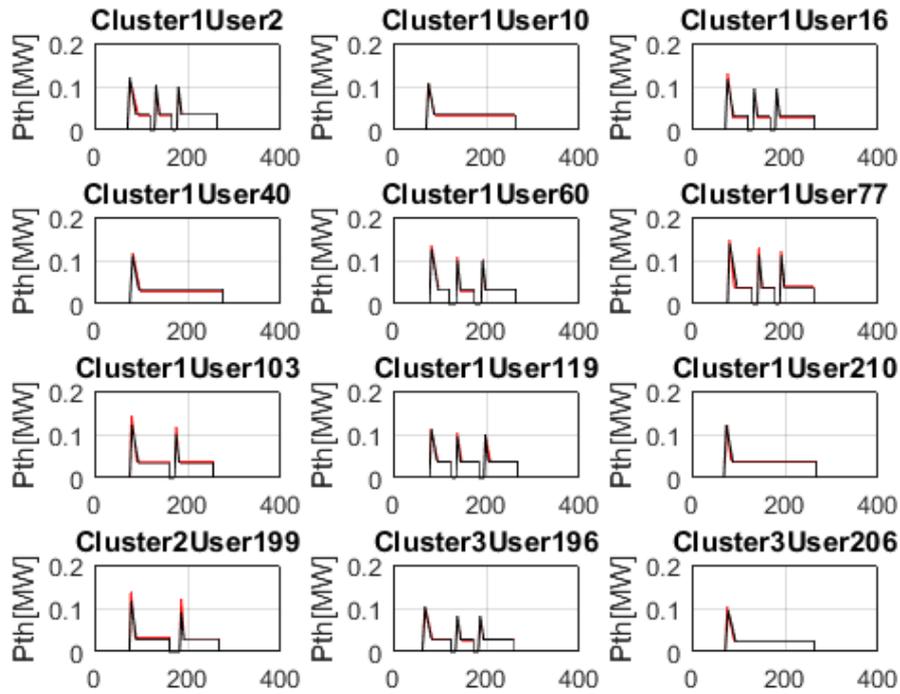


Figura 4.2 Confronto profili di richiesta: Caso K-means

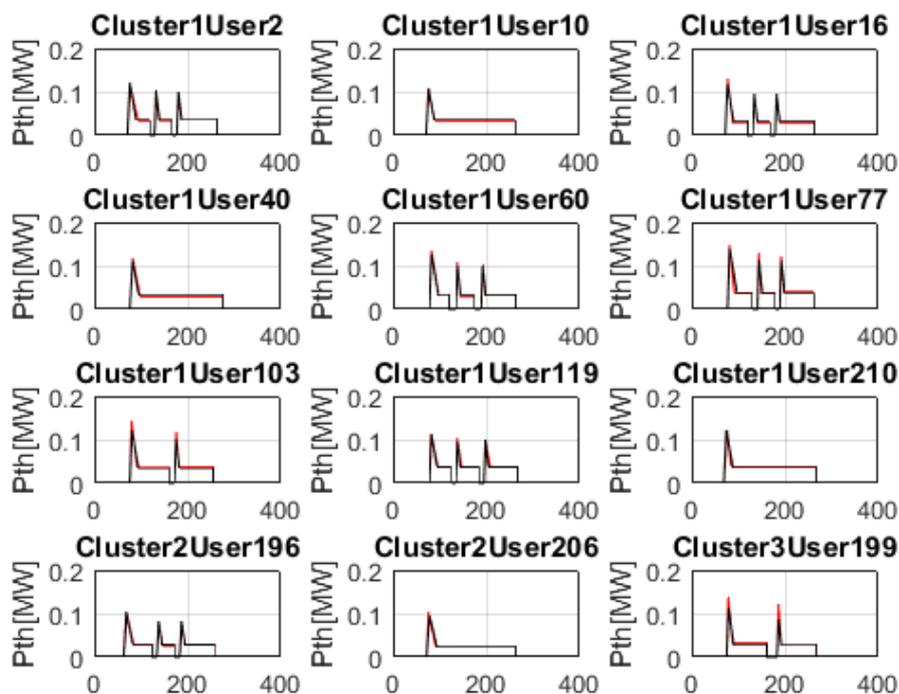


Figura 4.3 Confronto profili di richiesta: Caso algoritmo gerarchico

La curva in nero rappresenta il profilo di richiesta termica approssimato mentre la curva in rosso il profilo fittizio.

Le utenze numero 2, 10 e 16 rappresentate appartengono al baricentro 408, le utenze 40, 60 e 77 al 410, la 103 e la 119 al 414, la 196 e la 199 al 418 ed infine la 206 e la 210 al 420.

Già analizzando queste figure si nota come i profili ottenuti utilizzando gli algoritmi di Clustering tendano ad avvicinarsi in modo migliore alla richiesta delle utenze.

Infatti, ad esempio, analizzando l'utenza numero due si può notare come il valore del picco mattutino e della richiesta stazionaria siano rispettivamente di circa 490 kW e 168 kW mentre i valori determinati con i tre metodi siano rispettivamente 536 kW e 160kW, senza Clustering, e circa 500 kW e 169 kW (sia con il K-means che con l'algoritmo gerarchico). Come

si vede, con l'analisi di Clustering si riesce a prevedere meglio il profilo di richiesta.

Per verificare ciò si è deciso di calcolare gli errori relativi, rispetto al picco mattutino e alla richiesta stazionaria per ogni baricentro, ed, infine, un errore relativo globale, sempre rispetto al picco mattutino e alla richiesta stazionaria, pesato sulla potenza termica richiesta da ciascun baricentro.

Il procedimento utilizzato è il seguente.

- Errore relativo al picco mattutino per ciascun baricentro: per ogni baricentro analizzato si è determinato il valore globale del picco sia per il profilo fittizio (piccoF), sia per il profilo approssimato (piccoA) per cui l'errore è stato calcolato con la seguente formula:

$$relerr_{picco} = \frac{|piccoF - piccoA|}{|piccoF|} \quad (4.1)$$

- errore relativo alla richiesta stazionaria per ciascun baricentro: per ogni baricentro analizzato si è determinato il valore globale della richiesta stazionaria, sia per il profilo fittizio (stazF) che per il profilo approssimato (stazA) per cui l'errore è stato calcolato con la seguente formula:

$$relerr_{stazionario} = \frac{|stazF - stazA|}{|stazF|} \quad (4.2)$$

- errore relativo globale rispetto al picco mattutino pesato per la potenza termica: una volta determinato l'errore relativo al picco mattutino per ciascun baricentro e conoscendo la potenza termica richiesta da ciascuno di essi, si è calcolato l'errore utilizzando la seguente equazione:

$$relerr_{picco\ globale} = \sum_{i=1}^{nbar} \frac{relerr_{picco,i} * P_{th,i}}{P_{th}} \quad (4.3)$$

Dove $nbar$ è il numero totale dei baricentri (5), $P_{th,i}$ la potenza termica richiesta dall' i -esimo baricentro e P_{th} la potenza termica globale richiesta da tutti i baricentri.

- errore relativo globale rispetto alla richiesta stazionaria pesato per la potenza termica: una volta determinato l'errore relativo alla richiesta stazionaria per ciascun baricentro e conoscendo la potenza termica richiesta da ciascuno di essi, si è calcolato l'errore utilizzando la seguente equazione:

$$relerr_{stazionario_globale} = \sum_{i=1}^{nbar} \frac{relerr_{stazionario,i} * P_{th,i}}{P_{th}} \quad (4.4)$$

Nelle figure seguenti vengono mostrati gli errori relativi al picco mattutino e alla richiesta stazionaria per ciascun baricentro.

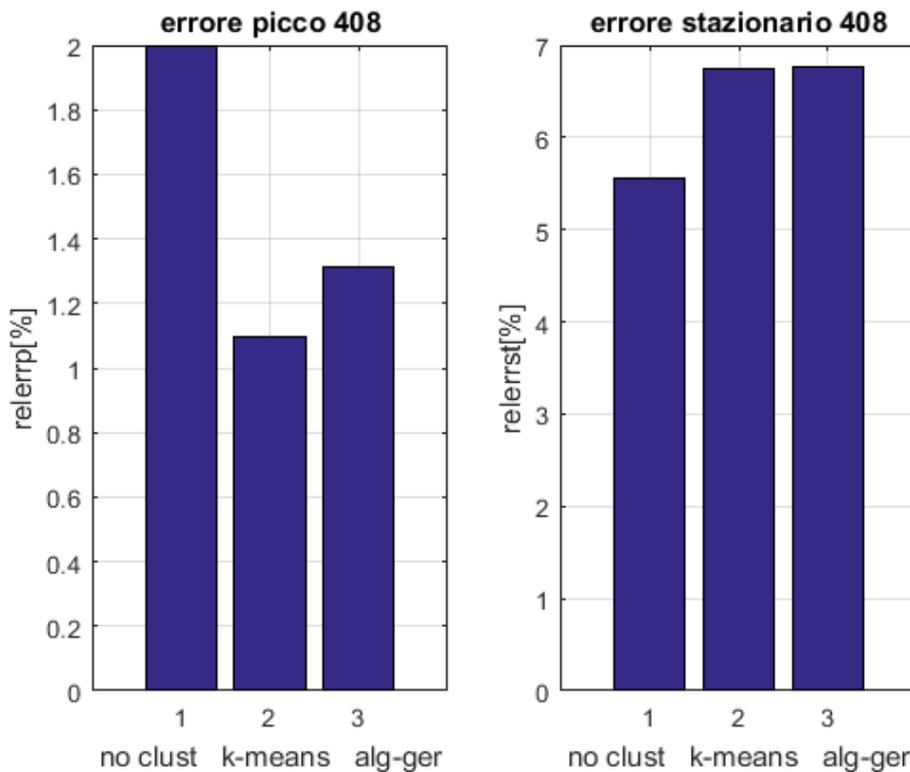


Figura 4.4 Errore calcolato al baricentro 408

Per il baricentro 408 si può notare come l'errore calcolato senza l'analisi di Clustering risulti maggiore rispetto a quello determinato con l'algoritmo K-means e l'algoritmo gerarchico, passando infatti, rispettivamente, da un

valore di circa il 2% nel primo caso ad un valore pari all' 1.1% nel secondo caso ed all' 1.3 % nel terzo.

L'opposto accade per la richiesta stazionaria dove l'errore calcolato senza Clustering è del 5.5% mentre con gli altri due algoritmi esso assume un valore maggiore di circa l'1%.

Per quanto riguarda i baricentri 410, 414 e 418, sia per il picco mattutino che per la richiesta stazionaria, l'analisi di Clustering conduce a risultati migliori.

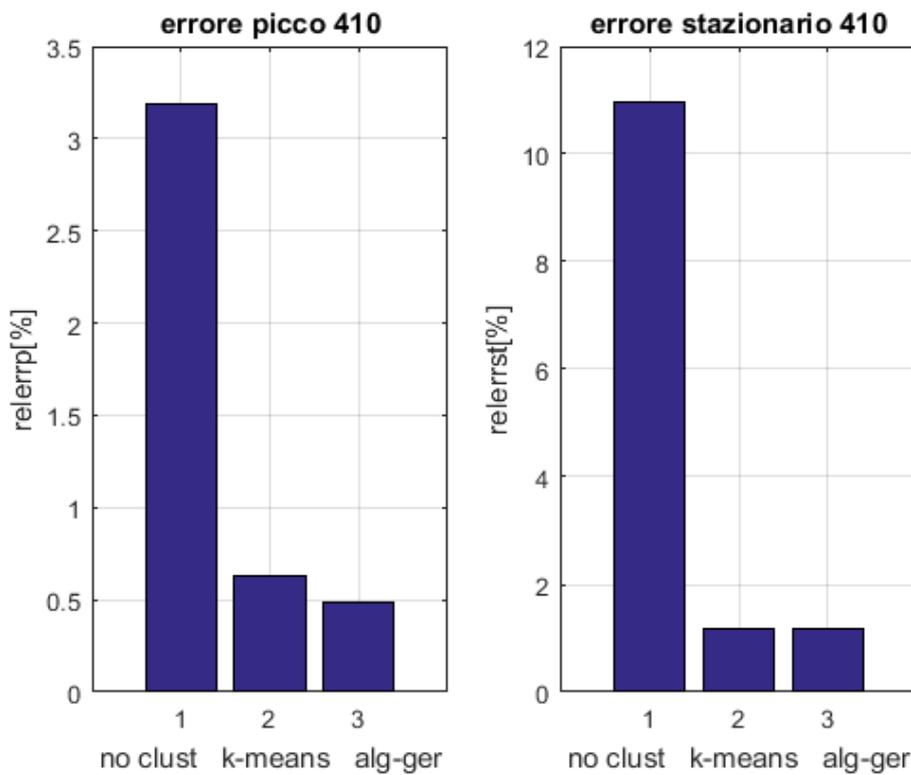


Figura 4.5 Errore calcolato al baricentro 410

Nel baricentro 410 l'errore rispetto alla richiesta stazionaria diminuisce di circa il 10% utilizzando il K-means e l'algoritmo gerarchico, giacché esso si riduce da circa l'11% a circa l'1%. Il miglioramento si ha anche sul picco mattutino dove applicando i due algoritmi di Clustering l'errore è pari a

0.6% e 0.5% a fronte di un valore di circa il 3.2% nell'analisi senza Clustering.

Nel baricentro 414 il miglioramento ottenuto sul picco mattutino è molto contenuto, poiché risulta inferiore allo 0.4%, mentre è più evidente nella richiesta stazionaria, dove appare in misura leggermente superiore al 3%.

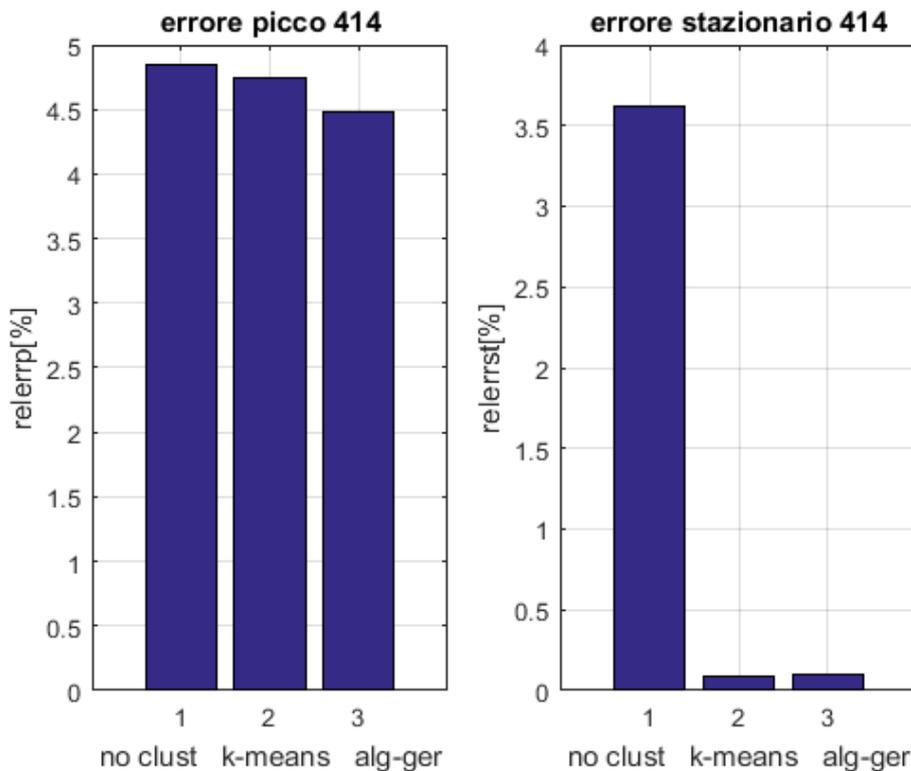


Figura 4.6 Errore calcolato al baricentro 414

Per il baricentro 418 l'errore calcolato sul picco mattutino e sulla richiesta stazionaria senza l'analisi di Clustering sono rispettivamente pari al 15.5% ed al 7.5%. Invece applicando i due algoritmi, K-means e gerarchico, si ottengono per il picco mattutino valori di circa 9% e 9.5% rispettivamente mentre, per la richiesta stazionaria, un errore pari a circa 2.6% in entrambi i casi. Dunque, anche in questa ipotesi, si può notare un buon miglioramento, corrispondente a circa il 6% per l'errore sul picco mattutino e pari a circa il 5% per la richiesta stazionaria.

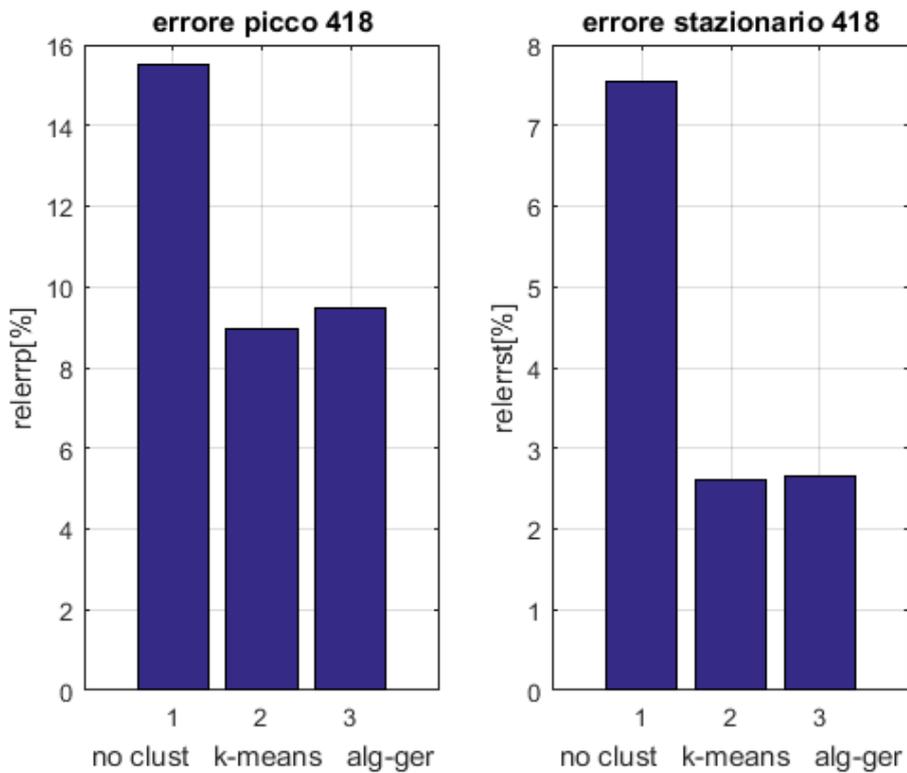


Figura 4.7 Errore calcolato al baricentro 418

Nel baricentro 420 si ha invece un peggioramento, per quanto riguarda l'errore calcolato sul picco mattutino, applicando l'analisi di Clustering. Questo peggioramento però è molto esiguo, ovvero pari a circa lo 0.25%. Si passa infatti da un valore dell'errore di circa 14.12% ad un valore di circa 14.37% per entrambi gli algoritmi di Clustering.

Invece si ha un piccolo miglioramento per quanto riguarda l'errore calcolato sulla richiesta stazionaria, corrispondente a circa lo 0.5%, in quanto si passa da un valore di circa il 3%, senza l'analisi di Clustering, ad un valore di circa il 2.5% sia utilizzando il K-means che l'algoritmo gerarchico.

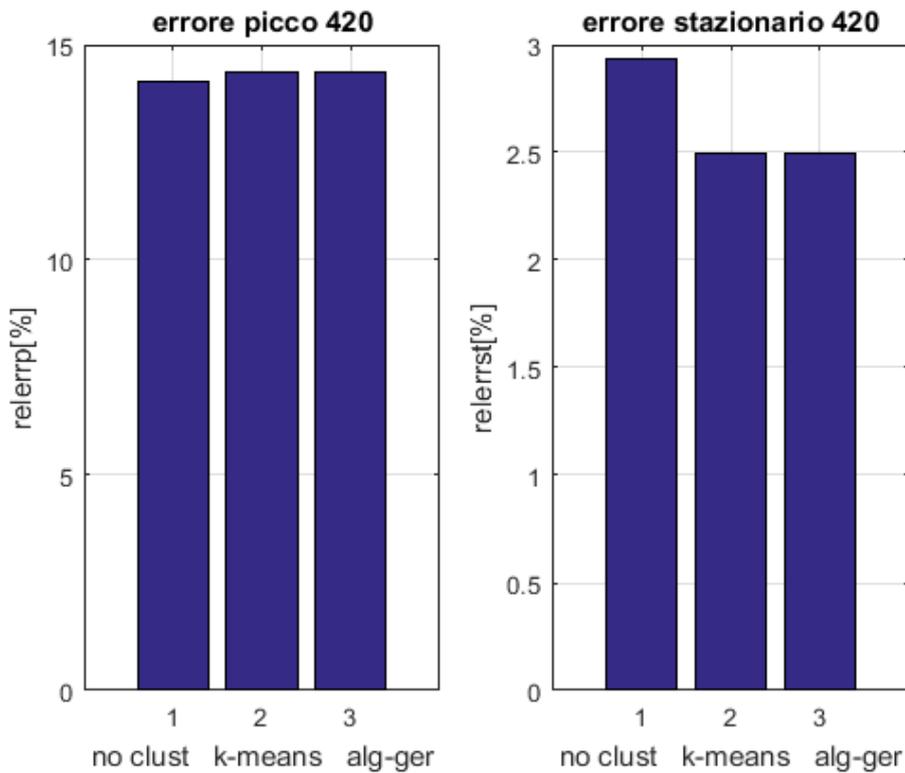


Figura 4.8 Errore calcolato al baricentro 420

Si può notare che per gli ultimi due baricentri analizzati, il 418 e 420, si ottengono errori, soprattutto sul picco mattutino, maggiori rispetto agli altri. Questo è dovuto principalmente al fatto che essi sono costituiti da poche utenze, 11 per il 418 e 10 per il 420 su 212 utenze analizzate, e quindi si ha un campione limitato su cui testare il modello.

Analizzando i valori assunti dagli errori appena studiati, si può dire che il modello costruito con l'analisi di Clustering conduce a prevedere meglio il profilo di richiesta. Ciò si può notare dal fatto che la maggior parte delle volte l'errore calcolato con gli algoritmi di Clustering risulta più basso. In aggiunta, quando si ha un peggioramento dell'errore, come ad esempio nel caso di richiesta stazionaria per il baricentro 408 e del picco mattutino per il baricentro 420, esso risulta abbastanza contenuto, in misura non superiore

all'1%, mentre quando si hanno miglioramenti essi raggiungono anche valori del 10%.

Per avvalorare ciò si è calcolato, sia per il picco mattutino che per la richiesta stazionaria, l'errore relativo globale pesato sulla potenza termica richiesta da ciascun baricentro. I risultati sono mostrati nelle figure 4.9 e 4.10.

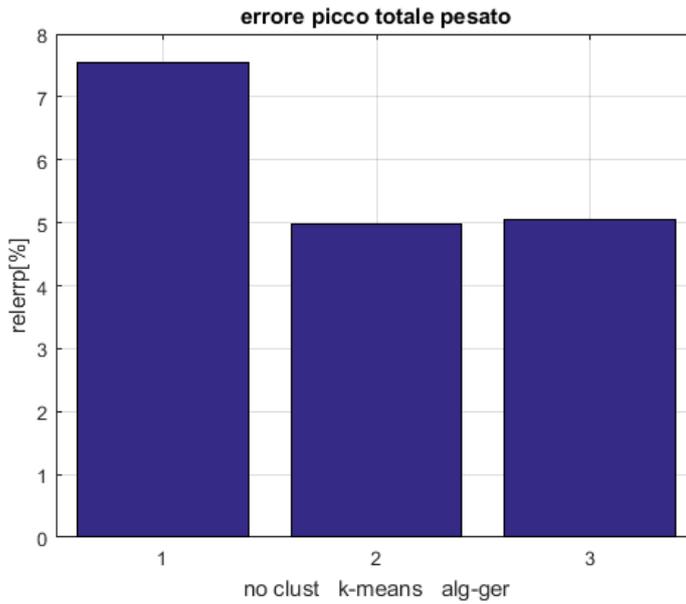


Figura 4.9 Errore relativo globale pesato calcolato sul picco

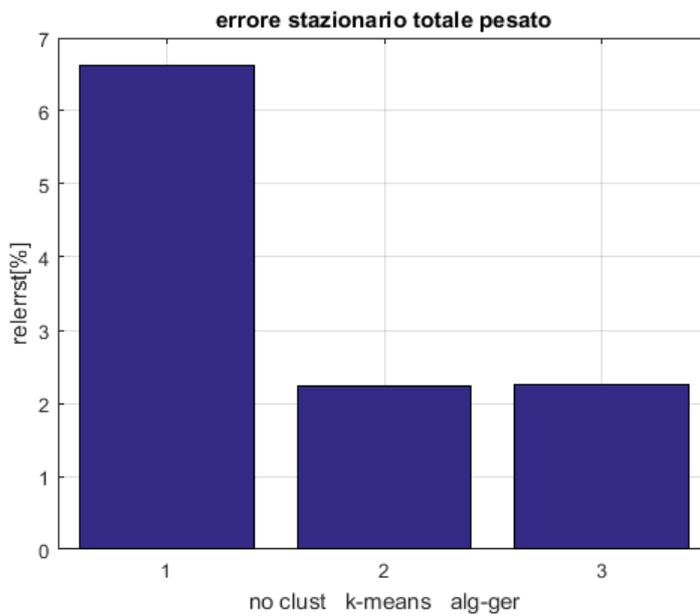


Figura 4.10 Errore globale pesato calcolato sulla richiesta stazionaria

Si può notare, in entrambi i casi, che l'errore calcolato con l'analisi di Clustering è più esiguo e che entrambi gli algoritmi applicati portano a risultati simili. Si passa infatti da un errore sul picco mattutino e sulla richiesta stazionaria del 7.5% e del 6.6%, senza l'analisi di Clustering, a valori, rispettivamente, pari circa al 5% e al 2.2%, applicando gli algoritmi di Clustering.

4.2 Confronto dei risultati ottenuti con i dati sperimentali

In questo paragrafo, come ultima analisi, viene eseguito un confronto tra i profili di richiesta ottenuti senza l'analisi di Clustering e l'algoritmo K-means con i profili di richiesta reali.

Nella figura 4.11 viene illustrato il confronto quando la domanda dell'utenza è determinata senza l'utilizzo degli algoritmi di Clustering, mentre nella figure 4.12 è raffigurato il raffronto quando il profilo di richiesta è ottenuto utilizzando l'analisi di Clustering.

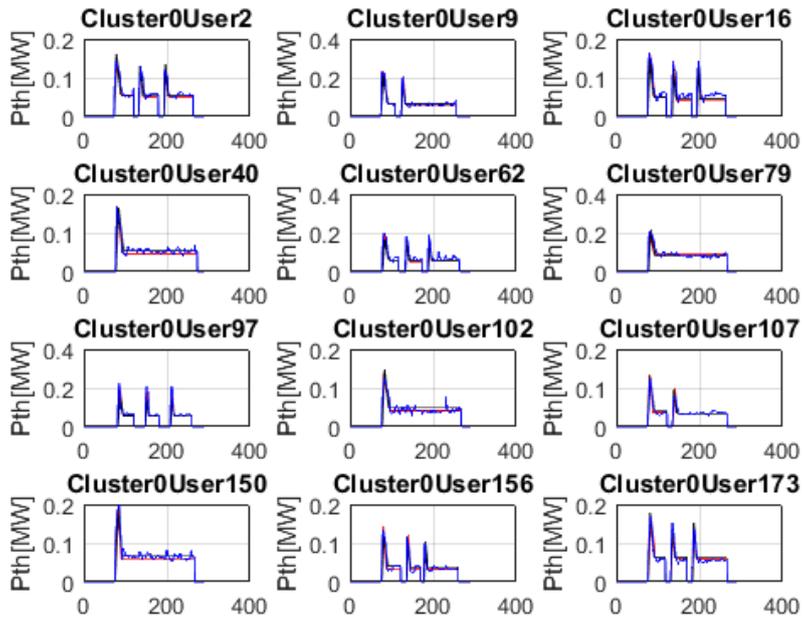


Figura 4.11 Confronto fra profilo di richiesta approssimato e reale: Caso senza Clustering

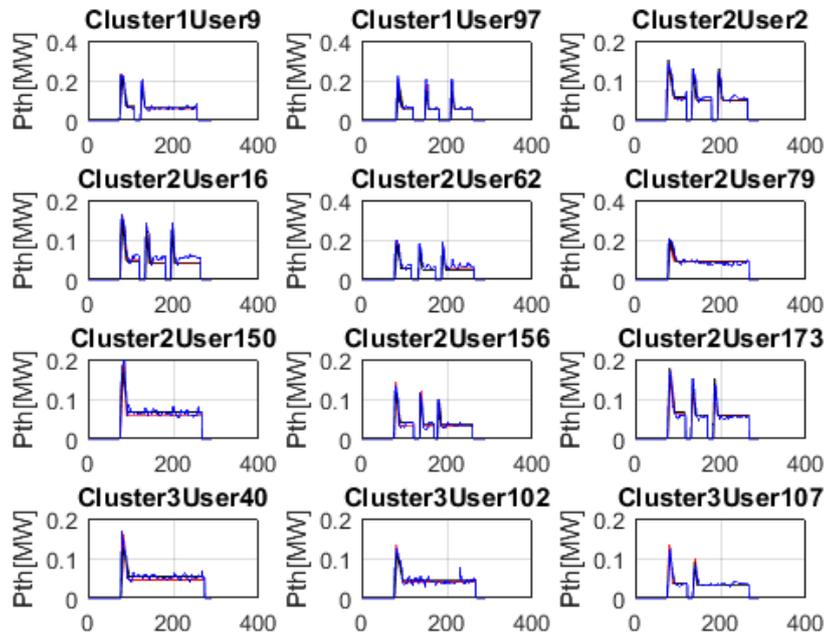


Figura 4.12 Confronto fra profilo di richiesta approssimato e reale: Caso con Clustering

La curva in nero rappresenta il profilo di richiesta approssimato, la curva in rosso quello fittizio mentre la curva in blu la domanda reale dell'utenza.

Anche in questo caso si può notare come l'analisi di Clustering porti a risultati più accurati, nel senso che i profili ottenuti utilizzando l'algoritmo di Clustering tendano ad avvicinarsi in modo migliore alla richiesta delle utenze.

Si è verificato ciò calcolando gli errori relativi globali rispetto al picco mattutino e alla richiesta stazionaria pesati per la potenza termica richiesta da ciascuno baricentro, utilizzando le formule spiegate nel Capitolo 4.1.

In figura 4.13 e 4.14 sono rappresentati i risultati ottenuti.

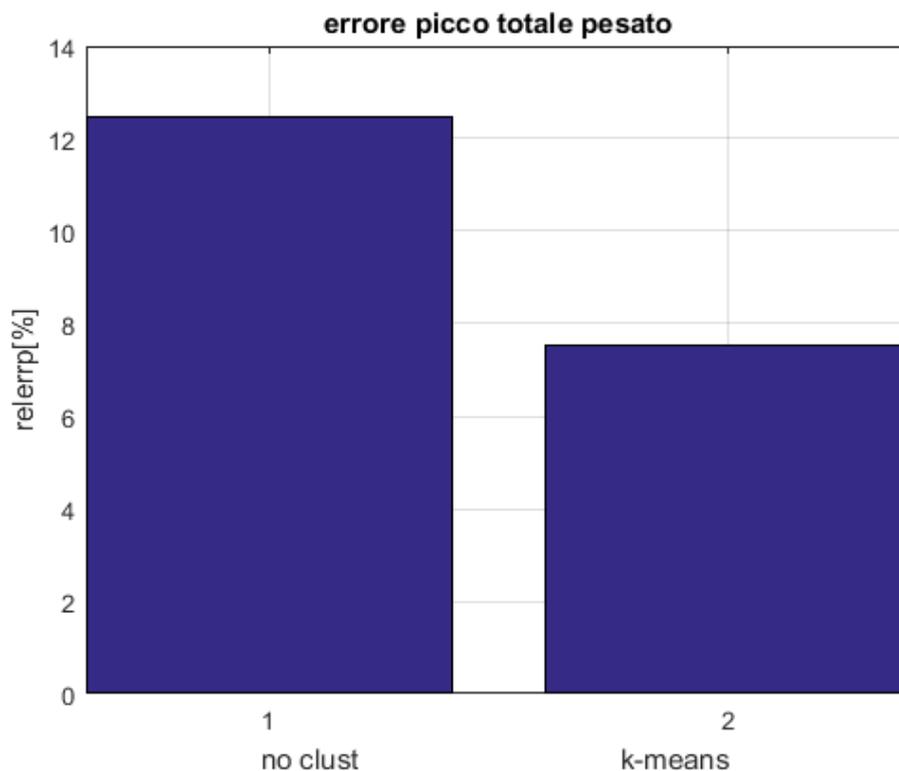


Figura 4.13 Errore relativo globale pesato calcolato sul picco mattutino

Si può vedere come l'errore calcolato sul picco senza eseguire l'analisi di Clustering sia circa 12.45% mentre quello determinato utilizzando l'algoritmo K-means sia circa 7.5%.

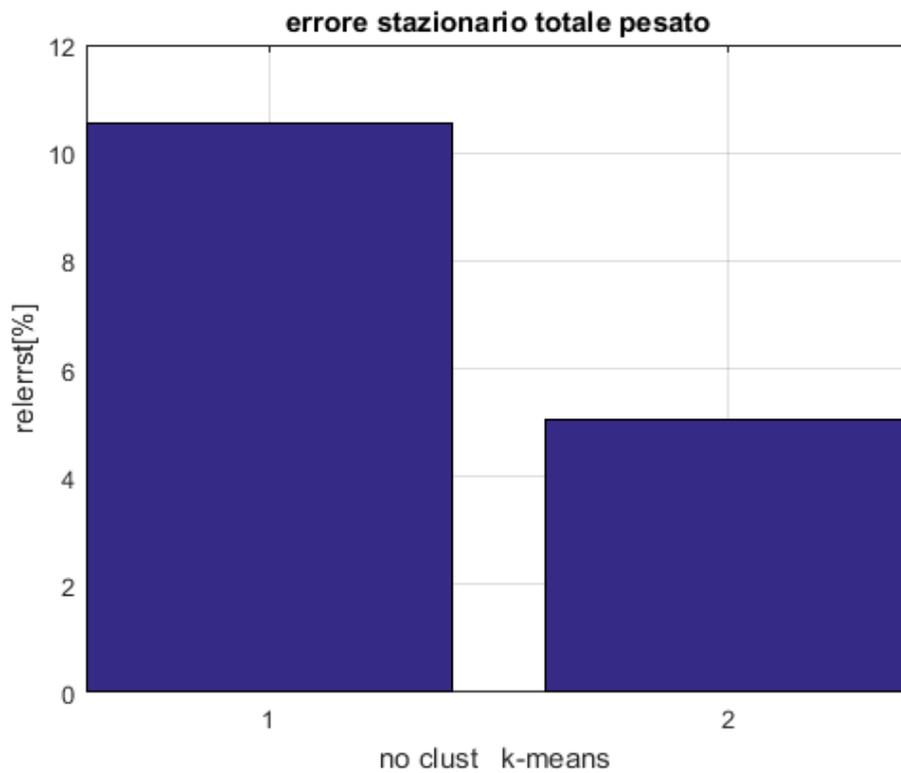


Figura 4.14 Errore globale pesato calcolato sulla richiesta stazionaria

Per quanto riguarda la richiesta stazionaria, eseguendo l'analisi di Clustering, l'errore viene quasi dimezzato. Si passa infatti da un valore leggermente superiore al 10%, senza l'analisi di Clustering, a un valore di circa il 5%, utilizzando l'algoritmo K-means.

4.3 Conclusioni

In questa Tesi si è proceduto alla creazione di un modello che consenta di prevedere il profilo di richiesta delle utenze allacciate alla rete di teleriscaldamento di Torino.

Dopo una prima analisi, in cui le utenze sono state studiate tutte insieme, si è eseguita un'analisi di Clustering per migliorare la previsione del profilo. Si è poi posto a confronto il profilo ottenuto con quello fittizio e la richiesta reale dell'utenza.

I risultati finali evidenziano come il modello creato, per quasi la totalità delle utenze, simuli in modo fedele gli andamenti reali del flusso termico.

Da un confronto delle tecniche utilizzate si può notare come l'analisi di Clustering porti a risultati più attendibili con un errore relativo globale sul picco e sulla richiesta stazionaria, pesato sulla potenza termica richiesta da ciascun baricentro, rispettivamente, del 5% e del 2.2 %, nel confronto con il profilo fittizio, e del 7.5% e 5% nel confronto con il profilo reale.

Da questi valori può comprendersi la validità dei risultati ottenuti con questa indagine.

Infine, per concludere, si può dire che fra gli algoritmi di Clustering utilizzati, il K-means è risultato il migliore in quanto a fronte del fatto che sono stati ottenuti risultati molto simili fra loro, di contro ha, come visto nei Capitoli 3.4.1.1 e 3.4.1.2, un costo computazionale inferiore, proporzionale al numero di dati analizzati.

BIBLIOGRAFIA

- [1] Vocabolario Treccani, teleriscaldamento s. m. [comp. di tele- e riscaldamento (propr. «Riscaldamento a distanza»)], (v. fig. a p. 466).
- [2] IREN energia, Teleriscaldamento, luglio 2017.
- [3] E. & Power, «District heating and cooling, Country by Country, 2015 Survey» 2015.
- [4] Marco Badami, Armando Portoraro “studio e caratterizzazione di reti termiche distribuite”, Report Ricerca di Sistema Elettrico, accordo di Programma Ministero dello Sviluppo Economico – ENEA. Settembre 2013.
- [5] V. Verda, “Dispense del corso Complementi di Energetica”, 2016.
- [6] Gruppo Hera, “Un mondo che si teleriscalda”, rapporto tecnico, 2013.
- [7] Gruppo Hera, “Anni '70: il teleriscaldamento arriva in Italia”, rapporto tecnico, 2013.
- [8] Debora Badii, “Il teleriscaldamento in Italia”, Articolo ARPAT.toscana.it, 21 Gennaio 2015.
- [9] IREN energia, Copertura di Torino, impianti di Torino, luglio 2017.
- [10] Redazione Ansa, Piemonte, “Iren, cresce teleriscaldamento a Torino”, 03 agosto 2017.
- [11] Prof. Claudio Capiluppi, “La regressione lineare”, 2007/2008.
- [12] Dott. Luca Filipponi, “La regressione lineare”.
- [13] Apogeo online, “Alcuni concetti di statistica: medie, varianze, covarianze e regressioni”, 2002.
- [14] <https://www.wordreference.com/enit/clustering>

- [15] D. Bonino, "Laboratorio del corso di statistica", 2013/2013
- [16] Prof. Matteo Golfarelli, "Clustering", Università di Bologna.
- [17] M. Steinbach, G. Karypis, V. Kumar, "A comparison of document Clustering Techniques", Report tecnico, University of Minnesota, 2000.
- [18] Prof. U. Villano, F. Pace, R. Di Candia, "K-means Data Clustering", Progetto, 2009/2010.