## POLITECNICO DI TORINO

### Corso di Laurea Magistrale in Ingegneria Edile

Tesi di Laurea Magistrale

## Analisi energetica di una centrale frigorifera del Politecnico di Torino attraverso analisi inversa



**Relatore** Candidato

Prof. Alfonso Capozzoli Mariachiara Pinto

#### Correlatori

Ing. Marco Savino Piscitelli

Ing. Silvio Brandi

Ing. Vincenzo Maria Gentile

Settembre 2018

## Indice

Indi	ce		1
Intro	duz	zione	3
1.	La g	gestione energetica degli edifici	5
1.1	1.	Contesto di riferimento	5
1.2	2.	Fattori influenzanti	13
	1.2.1	1. Fattori fisici	14
	1.2.2	2. Fattori umani	16
2.	Cara	atterizzazione e gestione di sistemi impiantistici di climatizzazione amb	oientale .19
2.1	1.	Individuazione del profilo di carico	20
2.2	2.	Riconoscimento e diagnosi delle anomalie	23
2.3	3.	Previsione dei consumi	26
3.	Poss	sibili approcci di diagnosi energetica del sistema edificio-impianti	31
3.1	1.	Approccio diretto	32
	3.1.1	1. Metodologia	32
3.2	2.	Approccio inverso	37
	3.2.1	1. Classificazione	38
	3.2.2	2. Analisi Clustering	46
4.	La c	centrale frigorifera del Politecnico di Torino.	56
4.1	۱.	Descrizione dell'impianto	57
4.2	2.	Descrizione del data-set	65
	4.2.1	1. Raccolta dei dati	67
	4.2.2	2. Pre – processamento dei dati	69

5. Risultati	86
5.1. Individuazione del profilo di carico tipico giornaliero	86
5.2. Classificazione	94
5.3. Costruzione del modello	108
5.3.1. Descrizione del modello	109
5.3.2. Calibrazione e validazione	114
Conclusioni	145
Allegato I	148
Allegato II	199
Bibliografia	205
Ringraziamenti	210

#### Introduzione

L'efficienza energetica degli edifici è un tema sempre più presente nelle politiche energetiche mondiali; circa il 40% del fabbisogno totale di energia deriva dal settore edile e al contempo, se si considera la sola Europa, circa il 75% degli edifici esistenti è energeticamente inefficiente.

Se, da una parte, tali politiche si focalizzano sulla fase di progettazione, promuovendo l'utilizzo di tecnologie costruttive ed impiantistiche eco-efficienti per gli edifici di nuova costruzione o soggetti a ristrutturazioni importanti, dall'altra esse si focalizzano anche sulla gestione degli stessi. A partire dal 2010 e nella fattispecie con la Direttiva 2010/31/UE viene incentivato l'utilizzo di sistemi di misurazione intelligenti per gli edifici di nuova costruzione o soggetti a ristrutturazioni importanti in modo da poter monitorare i consumi energetici e, al contempo, poter utilizzare i dati operativi per tracciare un quadro fedele della prestazione energetica globale ed individuare delle strategie di retrofit.

Grazie al crescente sviluppo di sistemi ICT ed alla riduzione dei costi della sensoristica, l'edificio non viene più considerato unicamente come un sistema ad alta intensità energetica ma anche come un sistema da cui attingere informazioni chiave utili. I dati operativi diventano il punto di partenza per la caratterizzazione prestazionale del sistema edificio-impianti, andando quindi a determinare le relazioni esistenti fra i consumi energetici ed i fattori influenzanti, l'identificazione di strategie di retrofit e, allo stesso tempo, essi divengono un termine di paragone per valutare l'efficacia di tali strategie.

È proprio in tale contesto che si pone il presente lavoro di tesi. Nella fattispecie, essendo gli impianti HVAC responsabili di circa il 50% dell'energia consumata da un edificio non residenziale, lo scopo coincide con l'effettuare un'analisi energetica della centrale frigorifera 1 del Politecnico di Torino attraverso l'approccio data-driven e al contempo creare un modello calibrato sui dati storici che rispecchi quanto più fedelmente possibile il caso studio. La fase di caratterizzazione dell'impianto è stata effettuata utilizzando le

tecniche di data mining di clustering e classificazione ed il software open-source R, mentre la fase di costruzione del modello simulativo ha visto l'utilizzo del software Trnsys.

Al contrario dell'approccio diretto o law-driven che si avvale di software simulativi che tengono conto di condizioni operative standard, l'approccio data-driven permette di tenere in considerazione non solo l'influenza dei fattori fisici sul consumo finale di energia, ma anche quelli umani, come il grado di comfort interno richiesto, il comportamento dell'utenza e lo stato manutentivo dei sistemi impiantistici.

In definitiva, il presente lavoro di tesi si compone delle seguenti sezioni:

- 1. Inquadramento dal punto di vista normativo e descrizione dei fattori che influiscono sul consumo energetico degli edifici;
- 2. Literature review sugli studi presenti in letteratura che si focalizzano sugli impianti HVAC utilizzando l'approccio data-driven;
- 3. Descrizione dei possibili approcci di diagnosi energetica, law-driven o data-driven, con riferimento alle metodologie e alle tecniche utilizzabili;
- 4. Descrizione impiantistica dal caso studio e del data-set utilizzato per la sua caratterizzazione, ponendo particolare attenzione al pre-processamento dei dati;
- 5. Risultati ottenuti dalle fasi di caratterizzazione dell'impianto e di costruzione del modello.

Nei due allegati vengono inoltre riportati rispettivamente lo script utilizzato per la fase di caratterizzazione dell'impianto e il file costruito per la calibrazione dell'impianto.

#### 1. La gestione energetica degli edifici

#### 1.1. Contesto di riferimento

L'edilizia rappresenta uno dei settori maggiormente preoccupanti in tema di sostenibilità ambientale. Secondo IEA, *International Energy Agency*, esso infatti è il principale settore energivoro con un consumo superiore ad un terzo di quello globale, e con una conseguente produzione ugualmente importante di CO2, che si prevede aumenterà del 50% nel 2050 se non si considerano delle misure correttive al riguardo [01]. Se restringiamo il campo alla sola Europa, esso rappresenta il primo consumatore singolo, con un assorbimento di circa il 40% del fabbisogno totale di energia: circa il 75% degli edifici esistenti è infatti energeticamente inefficiente [02].

A fronte di ciò e in conformità agli obiettivi europei (espressi all'interno del piano "20-20-20" e coincidenti con la moderazione delle emissioni di gas serra, la limitazione dei consumi energetici e l'aumento di produzione di energia rinnovabile) sono state redatte una serie di normative volte appunto a ridurre il consumo energetico degli edifici, promuovere l'utilizzo di fonti rinnovabili, e favorire una corretta gestione degli stessi. In particolar modo:

• Il D.lgs. 28/2011 [03], che recepisce la Direttiva Europea 28/2009/CE, "definisce gli strumenti, i meccanismi, gli incentivi, il quadro istituzionale e finanziario e giuridico, necessari per il raggiungimento degli obiettivi fino al 2020 in materia di quota complessiva di energia da fonti rinnovabili sul consumo finale lordo di energia e di quota di energia da fonti rinnovabili nei trasporti".<sup>2</sup> Nell'allegato 3 sono inoltre riportati gli obblighi per gli edifici di nuova costruzione o soggetti a ristrutturazioni importanti. In particolar modo esso impone che gli impianti di produzione di energia termica, alimentati da fonti rinnovabili, devono garantire la

<sup>&</sup>lt;sup>1</sup> Il piano "20-20-20" corrisponde ad un insieme di norme vincolanti il cui fine è quello di garantire, per l'Unione Europea, il raggiungimento degli obiettivi in materia di clima ed energia entro il 2020. I settori a cui si riferisce il piano sono: edilizia, industria, agricoltura e trasporti.

<sup>&</sup>lt;sup>2</sup> Art.1- Finalità, Decreto legislativo 3 marzo 2011, n. 28.

copertura del 50% dei consumi previsti per la sola acqua calda sanitaria e di una quota che varia dal 20% al 50%, in funzione della data entro cui si presenta la richiesta del titolo edilizio, per la somma dei consumi dovuti ad acqua calda sanitaria, riscaldamento e raffrescamento.

- La legge n. 90 del 3 Agosto 2013, che recepisce la direttiva 31/2010/UE, "promuove il miglioramento della prestazione energetica degli edifici tenendo conto delle condizioni locali e climatiche esterne, nonché delle prescrizioni relative al clima degli ambienti interni e all'efficacia sotto il profilo dei costi"<sup>3</sup>. Essa stabilisce dei requisiti minimi in termini di prestazione energetica e indica che le metodologie, da usarsi per il calcolo della stessa, coincidono con quelle riportate nelle UNI/TS 11300, parti da 1 a 4, nella raccomandazione CTI 14/2013 e nella UNI EN 15193. Inoltre, introduce l'attestato di prestazione energetica, all'interno del quale devono essere riportati tutti i dati relativi all'efficienza energetica dell'edificio, e definisce un piano di azione per la promozione degli edifici ad energia quasi zero [04].
- Il Decreto legislativo n. 102 del 4 Luglio 2014 [05], che recepisce la direttiva 2012/27/EU [06], "stabilisce un quadro di misure per la promozione e il miglioramento dell'efficienza energetica che concorrono al conseguimento dell'obiettivo nazionale di risparmio energetico".4
- Il Decreto ministeriale del 26 Giugno 2015, Applicazione delle metodologie di calcolo delle prestazioni energetiche e definizione delle prescrizioni e dei requisiti minimi degli edifici [07].
- UNI CEI EN ISO 50001, Sistemi di gestione dell'energia.

Sebbene le prime due normative sopracitate hanno come tema principale il miglioramento delle prestazioni energetiche, in termini di requisiti minimi e utilizzo di fonti rinnovabili, all' interno delle stesse vengono dati degli input per una gestione razionale dell'utilizzo di energia. In entrambe viene infatti previsto, per gli edifici di nuova costruzione e per quelli soggetti a ristrutturazioni importanti, l'utilizzo di sistemi di controllo attivo, come i sistemi di automazione, controllo e monitoraggio finalizzati al risparmio energetico. In questo

<sup>&</sup>lt;sup>3</sup> Art.1- Finalità, Legge 3 agosto 2013, n. 90.

<sup>&</sup>lt;sup>4</sup> Art.1 – Finalità, Decreto Legislativo 4 luglio 2014, n. 102.

modo i dati ottenuti da tali sistemi di controllo vengono registrati in modo tale da poter essere utilizzati come dati storici su cui valutare il grado di efficienza dell'edificio e, al contempo, i miglioramenti nel tempo.

Il D.lgs. 102/2014 si rivolge a tutte le Aziende energivore, sia Pubbliche che Private, inclusi i condominii, a cui spetta l'obbligo di effettuare un audit energetico, definito come "una procedura sistematica finalizzata a ottenere un'adeguata conoscenza del profilo di consumo energetico di un edificio o gruppo di edifici, di una attività o impianto industriale o commerciale o di servizi pubblici o privati, a individuare e quantificare le opportunità di risparmio energetico sotto il profilo costi-benefici e a riferire in merito ai risultati". <sup>5</sup> Gli audit energetici devono tenere conto delle norme tecniche internazionali, come le ISO 50001 e le ISO 140001, e rispettare i criteri minimi, definiti nell'Allegato II del presente Decreto, e riportati di seguito:

- a) sono basati su dati operativi relativi al consumo di energia aggiornati, misurati e tracciabili e (per l'energia elettrica) sui profili di carico;
- b) comprendono un esame dettagliato del profilo di consumo energetico di edifici o di gruppi di edifici, di attività o impianti industriali, ivi compreso il trasporto;
- c) ove possibile, si basano sull'analisi del costo del ciclo di vita, invece che su semplici
  periodi di ammortamento, in modo da tener conto dei risparmi a lungo termine,
  dei valori residuali degli investimenti a lungo termine e dei tassi di sconto;
- d) sono proporzionati e sufficientemente rappresentativi per consentire di tracciare un quadro fedele della prestazione energetica globale e di individuare in modo affidabile le opportunità di miglioramento più significative.<sup>6</sup>

Il Titolo II del presente decreto, "Efficienza nell'uso dell'energia", affronta il tema del miglioramento prestazionale. In particolar modo esso prescrive che:

• le Pubbliche Amministrazioni si impegnano alla raccolta dei dati, al monitoraggio degli interventi di riqualificazione e alla stesura di piani di efficientamento, e al

.

<sup>&</sup>lt;sup>5</sup> Art.2-Definizioni, Capo I, Direttiva 2012/27/UE del 25 Ottobre 2012.

<sup>&</sup>lt;sup>6</sup> Allegato 2-Criteri minimi per gli audit energetici, compresi quelli realizzati nel quadro dei sistemi di gestione dell'energia, Decreto Legislativo 4 luglio 2014, n. 102.

contempo, a partire dal 2014 e sino al 2020, all'attuazione di interventi volti ad ottenere una riqualificazione energetica, pari almeno al 3 per cento annuo della superficie utile coperta climatizzata, o un risparmio energetico, cumulato nel periodo 2014-2020, di almeno 0.04 Mtep. Le Regioni e gli enti locali devono inoltre promuovere l'introduzione di sistemi di gestione dell'energia.

- le Imprese hanno l'obbligo di effettuare, con cadenza quadriennale, un audit condotto da società di servizi energetici, esperti in gestione dell'energia, o da ISPRA, Istituto Superiore per la Protezione e la Ricerca Ambientale, relativamente allo schema volontario EMAS<sup>7</sup>, nei siti produttivi localizzati sul territorio nazionale. Fanno eccezione le imprese che hanno adottato sistemi di gestione conformi EMAS e alle norme ISO 50001.
- l'articolo 9 sancisce che: negli edifici dotati di sistemi centralizzati di riscaldamento, raffrescamento o fornitura di acqua calda sanitaria, sia resa obbligatoria la computazione dei consumi individuali basati su dati reali. Le imprese distributrici hanno infatti l'obbligo, entro il 31 dicembre 2016, di fornire sistemi di contabilizzazione che rispecchino il consumo reale e diano informazioni sul tempo effettivo di utilizzo di energia e sulle relative fasce temporali.

Da quanto detto, emerge che il punto chiave del D.lgs. 102/2014 è il monitoraggio dei consumi. Esso infatti è alla base sia dell'articolo 9, precedentemente descritto, e sia della definizione stessa di audit che, basandosi sui dati reali, deve essere un punto di partenza per l'identificazione di piani strategici, volti al miglioramento prestazionale, e un termine di paragone per valutare i progressi ottenuti nel tempo. Inoltre, il presente Decreto individua la certificazione ISO 50001 come una delle possibilità per ottemperare gli obblighi di legge.

Il Decreto ministeriale del 26 Giugno 2015 [07] riprende quanto affermato nel D.lgs. 102/2014 e, nella sezione 3 dell'Allegato 1, prescrive "Al fine di ottimizzare l'uso

<sup>&</sup>lt;sup>7</sup> EMAS, Eco-Management and Audit Scheme, è un sistema comunitario di eco-gestione e audit a cui possono aderire volontariamente le imprese e le organizzazioni, sia pubbliche che private, aventi sede nel territorio della Comunità Europea o al di fuori di esso, che desiderano impegnarsi nel valutare e migliorare la propria efficienza ambientale.

dell'energia negli edifici, per gli edifici ad uso non residenziale è reso obbligatorio un livello minimo di automazione per il controllo, la regolazione e la gestione delle tecnologie dell'edificio e degli impianti termici (BACS), corrispondente alla Classe B, come definita nella Tabella 1 della norma UNI EN 15232 8 e successive modifiche o norma equivalente."9 La Tabella 1 viene mostrata di seguito in Table 1.1-1.

		Definition of classes							
		Residential			Non residential				
		D	С	В	А	D	С	В	
TOM	ATIC CONTROL								
ATIN	IG AND COOLING CONTROL								
Emi	ssion control								
	The control system is installed at the emittor or room level, for car	se f or	ne syste	m can c	control s	everal	rooms		
0	No automatic control								
1	Central automatic control								
2	Individual room automatic control by thermostatic valves or electronic controller								
3	Individual room control with communication between controllers								
Con	ntrol of distribution network water temperature (supply or ret	um)							
	Similar function can be applied to the control of direct electric hea	ting ne	tworks						
0	No automatic control								Г
1	Outside temperature compensated control								
2	Indoor temperature control								
Con	itrol of distribution pumps								
	The controlled pumps can be installed at different levels in the ne	foork							
0	No control								Г
1	On off control								
2	Variable speed pump control with constant Δp								
3	Variable speed pump control with proportional Δp								
Inte	rmittent control of emission and/or distribution								
	One controller can control different rooms/zone having same occur	upancy	pattern	is .					
0	No automatic control;								
1	Automatic control with fixed time program;								
2	Automatic control with optimum start/stop.								
Inte	rlock between heating and cooling control of emission and/	or dist	ribution	1					
0	No interlock;								Г
1	Partial interlock (dependant of the HVAC system)								
2	Total interlock.								
Ger	nerator control								
0	Constant temperature								
1	Variable temperature depending on outdoor temperature								
2	Variable temperature depending on the load								
Seq	uencing of different generators								
0	Priorities only based on loads								П
	Priorities based on loads and generator capacities			_					

Table 1.1-1: Function list and assignment to energy performance classes, UNI EN 15232.

Si vuole inoltre aggiungere che la sigla BACS, *Building Automation and Control System*, i.e. Sistemi di controllo e automazione dell'edificio, si riferisce a tutti i sistemi centralizzati che monitorano, controllano e registrano il funzionamento degli impianti tecnici. Affinché un

<sup>&</sup>lt;sup>8</sup> La normativa UNI EN 15232 del Febbraio 2012, abroga la UNI EN 15232:2007 e fornisce informazioni circa le funzioni di regolazione, automazione e gestione tecnica degli edifici, una metodologia per definire i requisiti minimi da applicare e per calcolare l'incidenza di queste funzioni su un determinato edificio.

<sup>&</sup>lt;sup>9</sup> Paragrafo 3, Allegato 1, Decreto ministeriale del 26 Giugno 2015.

sistema sia in classe B, i dispositivi di regolazione degli ambienti devono poter comunicare con il sistema di automazione dell'edificio.

La norma internazionale ISO 50001 del 2011 [08] fornisce a tutte le organizzazioni, di qualsiasi tipologia e dimensione, i requisiti per creare, avviare, mantenere e migliorare un sistema di gestione dell'energia, SGE, in modo da incrementare costantemente, attraverso un approccio sistematico, la propria prestazione energetica e dunque l'uso e il consumo dell'energia.

Essa si basa sul ciclo PDCA, *Plan-Do-Check-Act*, visibile in figure 1.1-1 e consistente in un approccio sequenziale e iterativo basato su quattro fasi:

- Plan: realizzare l'analisi energetica e stabilire il valore di riferimento, gli indicatori
  di prestazione energetica (EnPls), gli obiettivi, i traguardi e i piani di azione
  necessari a fornire i risultati che portano a migliorare la prestazione energetica in
  conformità alla politica energetica dell'organizzazione.
- Do: attuare i piani d'azione della gestione dell'energia.
- Check: sorvegliare e misurare i processi e le caratteristiche chiave delle operazioni,
   che determinano le prestazioni energetiche rispetto alla politica energetica e agli
   obiettivi, e riportarne i risultati.
- Act: intraprendere azioni per migliorare in continuo la prestazione energetica e il sistema di gestione dell'energia.<sup>10</sup>

All'interno del processo, la pianificazione energetica assume un ruolo fondamentale in quanto deve coinvolgere tutte le entità che influiscono sulla prestazione energetica. Essa inizia dunque con una analisi energetica che, attraverso lo studio degli usi e dei consumi, basati su misurazioni o altri dati, mira ad individuare le fonti principali di energia, le aree di uso più significative, le priorità e le opportunità di miglioramento.

\_

<sup>&</sup>lt;sup>10</sup> Introduzione, UNI CEI EN ISO 50001:2011.

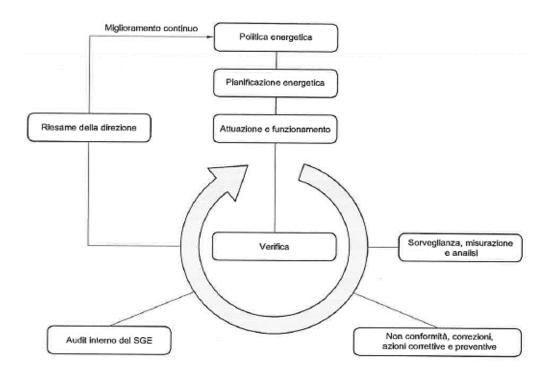


Figure 1.1-1Energy management system model for this International Standard, UNI CEI EN ISO 50001:2011

Sulla base dei risultati ottenuti dall'analisi energetica devono inoltre essere stabiliti il consumo di riferimento, considerando un periodo adatto a rappresentare l'uso di energia nell'organizzazione, e gli indici prestazionali appropriati per il monitoraggio continuo delle prestazioni. Nella figura 1.1.2 mostrata di seguito è possibile vedere un diagramma esplicativo del processo di pianificazione.

Altra fase molto importante nel processo è la fase di verifica: le organizzazioni devono infatti prevedere, ad intervalli predefiniti, che le performance energetiche siano monitorate, misurate ed analizzate. In questo modo infatti è possibile valutare l'efficacia dei piani di azione, e quindi il raggiungimento degli obiettivi fissati.

Nel caso in cui ci siano delle non conformità allora è previsto un riesame delle stesse andando a individuarne le cause e applicando delle azioni correttive o preventive, che saranno comunque soggette ad ulteriore analisi in modo da verificarne l'efficacia.

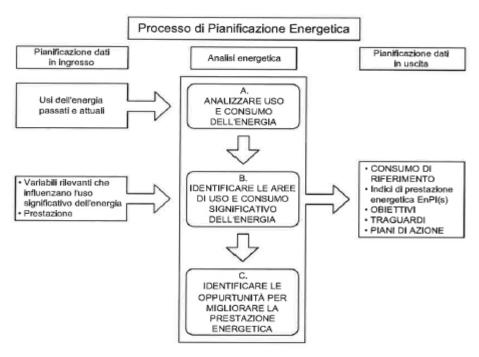


Figure 1.1-2 Energy planning process concept diagram, UNI CEI EN ISO 50001:2011

In virtù di quanto detto, emerge che la legislazione sta spingendo sempre di più verso gli edifici smart, dotati di sistemi di automazione e controllo, che permettono di collezionare i dati reali relativi all'intero edificio e di migliorarne le prestazioni in termini energetici, nonché di sistemi di management dell'energia, BEMS. Ciò è possibile grazie al forte sviluppo del settore ICT<sup>11</sup>, la riduzione dei costi della sensoristica, la diffusione delle tecnologie wireless, e dell'IoT<sup>12</sup>.

L'acronimo ICT viene utilizzato per indicare le tecnologie dell'informazione e della comunicazione, in inglese Information and Communications Technology, ovvero l'insieme dei metodi e delle tecniche che permettono agli utenti di creare, collezionare e scambiare informazioni.
L'acronimo IoT, Internet of Things, la cui traduzione letteraria è Internet degli oggetti, indica una rete di apparecchiature e dispositivi connessi ad internet. Vengono utilizzati in molti campi inclusa la domotica.

#### 1.2. Fattori influenzanti

L'edificio può essere considerato come un sistema termodinamico aperto, dove per questo si intende una porzione materiale, confinata da una superficie di controllo immaginaria o reale, rigida o deformabile, caratterizzata da scambi di energia e massa con l'ambiente esterno e con gli occupanti presenti all'interno. Per una corretta gestione energetica, occorre quindi identificare quali sono i fattori che influiscono sul consumo reale dell'edificio, al fine di migliorarne la performance attraverso una serie di strategie.

Il problema principale del raggiungimento dell'efficienza energetico nel settore edile consiste nella carenza di conoscenze dei fattori che stabiliscono l'uso dell'energia, con un conseguente sfasamento tra il progetto e l'utilizzo reale. In generale possiamo individuare due macro-categorie, strettamente inter-correlate fra di loro, ovvero:

- Fattori fisici: il clima, le caratteristiche dell'edificio e dei sistemi impiantistici;
- Fattori umani: il grado di qualità dell'ambiente interno richiesto, il comportamento dell'utenza e le modalità di esercizio e lo stato manutentivo dei sistemi impiantistici.

Se infatti i fattori fisici ci permettono di stimare la performance energetica dell'edificio, solo con l'aggiunta dei fattori umani otteniamo il consumo reale di energia [09], Figure 1.2-1.

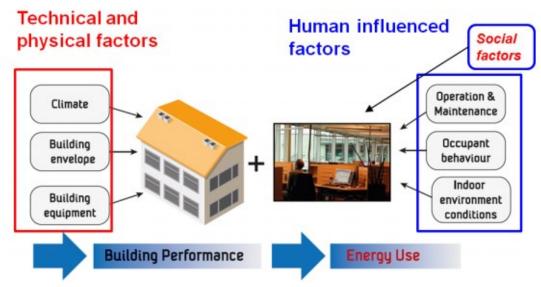


Figure 1.2-1- Six factors influencing total energy use in buildings. Annex 53-IEA project.

#### 1.2.1. Fattori fisici

I fattori fisici sono quelli più noti, sui cui si basano la maggior parte delle normative esistenti volte al miglioramento delle prestazioni energetiche, e che vengono considerati soprattutto nella fase di progettazione. La valutazione della loro influenza è facilmente ottenibile grazie ad un approccio deterministico, che tiene conto di tutti gli scambi termici, che avvengono nel sistema ambiente-edificio-impianto in condizioni standard, e che possono avvenire per:

- conduzione, trasferimento di calore o energia attraverso un solido, senza che ci sia un trasferimento di materia. Gli scambi per conduzione sono principalmente quelli che si hanno attraverso l'involucro edilizio, opaco e trasparente, causati dalla differenza di temperatura delle due facce della parete: più questa sarà maggiore e più lo sarà l'entità degli scambi termici.
- convezione, trasferimento di calore o energia attraverso un fluido associato al trasferimento di materia. Gli scambi per convenzione sono principalmente dovuti alla presenza di corpi scaldanti, all'interno dell'ambiente, come ad esempio gli impianti e l'utenza.
- irraggiamento, trasferimento di calore o energia che avviene anche in assenza di materia, tra corpi distanti, per mezzo di onde elettromagnetiche. Gli scambi per irraggiamento sono principalmente quelli che riguardano la radiazione solare e di conseguenza l'involucro, incidendo maggiormente su quello trasparente.

Appare chiaro dunque che il clima e le caratteristiche dell'edificio sono fortemente correlati fra di loro. Tutti gli scambi termici sono sì dovuti alle condizioni climatiche, come temperatura e radiazione solare, e di conseguenza l'area geografica e l'orientamento, ma anche alle caratteristiche dell'involucro edilizio, in quanto mezzo attraverso cui si propaga il calore. Nello specifico, essi possono essere dovuti a:

• la trasmittanza termica, definita come "il coefficiente di scambio globale di calore, in regime stazionario, tra due fluidi separati da una parete, espresso (in unità SI) in watt a metro quadrato (di area della parete) e a kelvin (di differenza di temperatura tra le due facce della parete), W/(m2 K)." Siano U la trasmittanza termica ed Ri le

resistenze termiche dei materiali formanti l'involucro edilizio, incluse le resistenze termiche superficiali, essa si ottiene con la seguente formula:  $U = 1/\sum Ri [W/mq^{\circ}K]$ . Minore è la trasmittanza termica, maggiore è il grado di isolamento che il componente offre.

- l'inerzia termica, o massa termica, ovvero la capacità di un materiale ad opporsi al
  passaggio di calore, attraverso un accumulo dello stesso e un successivo rilascio
  con uno sfasamento temporale. Essa è direttamente proporzionale alla capacità
  termica, definita come massa per calore specifico, e inversamente proporzionale
  alla conducibilità termica.
- la presenza di ponti termici, definiti come "parte dell'involucro edilizio dove la resistenza termica, altrove uniforme, cambia in modo significativo per effetto di compenetrazione totale o parziale di materiali con conduttività termica diversa nell'involucro edilizio e/o variazione dello spessore della costruzione e/o differenze tra l'area della superficie disperdente sul lato interno e quella del lato esterno". <sup>13</sup>I ponti termici possono essere causati da discontinuità geometrica, differenza tra le aree delle superfici disperdenti, ad esempio un angolo che si viene a formare dall'incontro di due pareti esterne, o materica, dovuta alla presenza di un elemento costruttivo avente caratteristiche termiche diverse, ad esempio la presenza di un pilastro, o di un serramento, all'interno della parete.
- i ricambi di aria e le modalità di ventilazione, controllata o non.

Allo stesso modo, i sistemi impiantistici, dovendo garantire le condizioni interne desiderate, dipendono strettamente dalle condizioni ambientali e costruttive precedentemente elencate. Per un'analisi delle loro prestazioni si considerano i valori di efficienza energetica sia dei generatori di calore, dei gruppi frigoriferi e dei produttori di acqua calda sanitaria, nonché degli ausiliari e dei sottosistemi di distribuzione, sia dei corpi illuminanti e sia degli elettrodomestici o delle macchine per ufficio.

\_

<sup>&</sup>lt;sup>13</sup> UNI EN ISO 10211:2008. Ponti temici in edilizia- Calcolo dei flussi termici e delle temperature superficiali.

#### 1.2.2. Fattori umani

I fattori umani rappresentano la causa principale del divario che si ha fra i consumi reali e quelli stimati. Lo studio della loro incidenza è un tema molto complesso in quanto sono fortemente variabili e inoltre non è semplice riuscire ad identificare il contributo che ciascuno di essi fornisce singolarmente sul consumo energetico. Data la loro aleatorietà, al fine di valutarne l'incidenza sul consumo finale, è necessario ricorrere non solo ad un approccio deterministico ma anche probabilistico. Attraverso la comprensione dei meccanismi che regolano questi fattori e la loro descrizione quantitativa, utilizzando modelli matematici, si possono identificare delle nuove metodologie di stima dei consumi energetici, che tengano conto dei fattori precedentemente descritti, al fine di creare dei benchmarks<sup>14</sup> dei consumi ed elaborare dei modelli predittivi che consentano la valutazione dei benefici ottenibili.

Il comportamento dell'utenza, e di conseguenza il comfort interno, è di natura fortemente stocastica. Esso influisce sul consumo energetico sia in funzione della destinazione d'uso, e quindi il livello di occupazione dell'edificio, sia per altri fattori fisici e socioeconomici, si pensi banalmente all'apertura e alla chiusura di finestre, che modificano le condizioni ambientali interne, o all'utilizzo di elettrodomestici o apparecchiature da ufficio. Gli studi in letteratura si concentrano su due temi fondamentali: l'identificazione e la valutazione quantitativa delle relazioni che sussistono tra la domanda e l'utilizzo energetico.

Ad esempio, Ouyang e Hokao [10] hanno investigato sui possibili risparmi energetici ottenibili da un utilizzo consapevole di energia, considerando 124 abitazioni in Cina e suddividendole in due gruppi: uno di essi è stato educato in modo da evitare sprechi di energia mentre all'altro è stato chiesto di mantenere il loro stesso comportamento. Confrontando i consumi elettrici, si è evinto che un uso consapevole di energia ha provocato una riduzione degli stessi in media superiore al 10%.

Yu et al. [11] ha fornito una metodologia basata su tecniche di clustering, di cui si parlerà nel capitolo 3, per identificare gli effetti del comportamento degli occupanti sul consumo

<sup>&</sup>lt;sup>14</sup> Il termine benchmark, la cui traduzione letteraria è banco di prova, viene utilizzato per indicare dei valori quantitativi di riferimento da utilizzare come temine di paragone.

di energia. Per validare la metodologia, uno studio è stato condotto su un gruppo di edifici residenziali, in sei diversi distretti del Giappone, e le variabili considerate consistono in dodici fattori di influenza non correlati con il comportamento dell'utente. I risultati ottenuti dimostrano che il raggiungimento del comfort in generale si traduce in un elevato consumo energetico, e pertanto bisognerebbe trovare un compromesso.

D'Oca e Hong [12] hanno proposto una metodologia basata su tre tecniche di data mining, come clustering, modelli decisionali e regole di associazione, in modo da classificare gli occupanti in funzione del loro profilo di utenza, e definire quali sono i fattori che lo spingono ad interagire con i serramenti, apertura e chiusura. I risultati ottenuti dall'applicazione di tale metodologia su 16 uffici privati, situati in Francoforte, mostrano che i motivi per cui l'occupante interagisce con l'apertura e la chiusura dei serramenti possono essere: una sensazione di discomfort o un comportamento legato alla routine giornaliera e alle abitudini. In particolar modo, dalle regole di associazione, è emerso che l'impatto sulla ventilazione naturale è minore nel momento in cui l'utente è spinto da fattori fisici, come temperatura interna ed esterna, a ristabilire la qualità dell'aria all'interno dell'ambiente, rispetto a quando esso è invece guidato da un comportamento legato alle abitudini.

Capozzoli et al. [13] ha invece sviluppato una metodologia volta a ridurre il consumo energetico del sistema HVAC utilizzando le tecniche di clustering e di riconoscimento dei pattern<sup>15</sup>, in modo tale da identificare e raggruppare gli impiegati, aventi modelli di occupazione simili, nella stessa zona termica. La fase successiva è consistita nella valutazione dell'impatto del programma di accensione e spegnimento degli impianti HVAC, attraverso una simulazione calibrata sulla base dei dati storici relativi ai consumi energetici. L'output di questo modello è stato poi utilizzato come riferimento per valutare un nuovo programma operativo che tiene conto, per quanto riguarda gli orari di accensione e spegnimento, dei livelli di occupazione. Dal confronto dei due modelli è

<sup>15</sup> Il riconoscimento dei pattern, in inglese pattern recognition, è una tecnica di machine learning,

apprendimento automatico, che si basa sull'individuazione e la classificazione di modelli all'interno di un dataset.

emerso che l'utilizzo del nuovo programma operativo comporta una riduzione dei consumi di circa il 14%.

Ricapitolando quanto detto nel presente capitolo possiamo affermare che il tema dell'efficienza energetica sta assumendo sempre più importanza a livello normativo non solo durante la fase di progettazione ma anche durante la fase di esercizio, ponendo particolare attenzione alla gestione del sistema edificio-impianti. Viene infatti promosso l'utilizzo di sistemi di automazione e controllo che permettano, attraverso l'acquisizione dei dati reali operativi, la caratterizzazione prestazionale del sistema edificio-impianti, l'identificazione di strategie di retrofit e, allo stesso tempo, un termine di paragone per valutare l'efficacia di tali strategie. Tali dati sono inoltre fondamentali per determinare le relazioni esistenti fra i consumi energetici ed i fattori influenzanti, ponendo particolare attenzione a quelli umani: in virtù della loro natura aleatoria la loro determinazione deve basarsi su un approccio sì deterministico ma anche probabilistico. Nella fattispecie, così come nel paragrafo 1.2.2 si è mostrato come i dati reali possono essere adoperati per valutare l'incidenza dell'utenza, allo stesso modo nel prossimo capitolo verrà mostrato come essi possono essere utilizzati per stimare l'influenza delle modalità di esecuzione e manutenzione dei sistemi impiantistici, con riferimento agli impianti HVAC.

# 2. Caratterizzazione e gestione di sistemi impiantistici di climatizzazione ambientale

Le modalità di esercizio determinano la maggior parte del consumo energetico durante l'intero ciclo di vita degli edifici, e il potenziale risparmio è considerevole se si tiene conto dell'ampia gamma di carenze operative esistenti. Il crescente utilizzo di sistemi di monitoraggio e controllo ha reso gli edifici moderni ad alta intensità circa le informazioni che essi riescono a fornire, e non solo dal punto di vista energetico. I dati acquisiti rispecchiano le performance energetiche reali e permettono di estrapolare informazioni utili, su cui basare delle strategie attuabili di efficientamento energetico[14].

Molti degli studi in letteratura sull'efficientamento energetico mirano a definire dei metodi che, attraverso l'utilizzo di tecniche di data mining, permettano una conoscenza più dettagliata delle relazioni, che sussistono fra i fattori influenzanti, al fine di:

- Caratterizzare il sistema edificio-impianti in funzione del suo profilo di carico, energetico o termico, i.e. load profiling;
- Individuare e diagnosticare le anomalie durante l'esercizio dei sistemi impiantistici, i.e. Fault Detection and Diagnosis;
- Prevedere i consumi energetici a breve termine, e.g. scala giornaliera o subgiornaliera.

Occorre inoltre aggiungere che ciascuno dei tre ambiti sopra riportati rappresenta una possibile fase del processo di efficientamento energetico. Il load profiling infatti corrisponde alla fase di audit e quindi di conoscenza preliminare del sistema analizzato attraverso cui andare ad individuare e diagnosticare le eventuali anomalie operative che si sono verificate nel periodo considerato. Allo stesso modo l'analisi degli output di entrambe le fasi permette l'individuazione di strategie di retrofit che è possibile valutare attraverso le procedure di stima dei consumi energetici.

Delle tecniche di data mining si discuterà maggiormente nel capitolo successivo, e ai fini del presente lavoro di tesi solo le tecniche di classificazione e clustering verranno ulteriormente approfondite. Per il momento ci si limita a dire che esse possono essere:

- Supervisionate: mirano a stabilire delle relazioni fra gli input e gli output basandosi sui dati reali. Esempi di tecniche supervisionate sono: regressione, classificazione Naive Bayes, machine a supporto vettoriale (Support Vector Machine, SVM), reti neurali artificiali (Artificial Neural Network, ANN).
- Non supervisionate: non sono guidate da un obiettivo esplicito ma mirano, attraverso un approccio esplorativo, a riconoscere delle relazioni nascoste o non note fra le variabili. Esempi di tecniche non supervisionate sono: clustering, analisi dei componenti principali (Principal Component Analysis, PCA), Symbolic Aggregate approXimation (SAX), regole di associazione.

#### 2.1. Individuazione del profilo di carico

Il load profiling consiste nel raggruppare in sequenze temporali i consumi energetici al fine di comprendere il comportamento tipico di un sistema impiantistico, un edificio o degli occupanti [15]. Il profilo di carico tipico varia in funzione delle caratteristiche dell'edificio, della destinazione d'uso, e quindi orari di accensione e spegnimento, dell'utenza e delle condizioni climatiche. Inoltre, la sua identificazione consente non solo di acquisire una maggiore conoscenza del sistema edificio-impianti considerato, ma permette anche di:

- Caratterizzare il funzionamento energetico impiantistico, riconoscendone le dinamiche e le eventuali anomalie, [17,18,19,20,21,23]
- Comparare edifici simili in termini di consumo energetico, [19,20,22];
- Stabilire delle procedure di benchmark [16,20,21];
- Proporre delle strategie di retrofit [16,22,23].

Le tecniche utilizzate sono molteplici: in generale si può affermare che l'analisi clustering è quella maggiormente utilizzata, in tutte le sue eccezioni, anche se studi in letteratura dimostrano il possibile utilizzo di altre tecniche quali: regole di associazione temporali, SAX.

Fan et al. [17] nel 2015 ha proposto una metodologia al fine di identificare delle relazioni temporali, basandosi su un dataset ottenuto dall'utilizzo di un sistema di controllo automatizzato degli impianti, inclusi i sistemi HVAC. In particolar modo, il procedimento, basato su quattro fasi, prevede l'alternanza di tecniche quali SAX e clustering, per la fase di pre-processamento, al fine di migliorare la qualità dei dati e raggrupparli in modo da avere la massima similarità all'interno di ciascun cluster. Successivamente, attraverso il riconoscimento di motif<sup>16</sup> e l'utilizzo di regole di associazione temporali, sono state individuate le relazioni che intercorrono all'interno del sistema edificio-impianti. L'utilizzo di tale approccio ha permesso di effettuare una valutazione energetica delle prestazioni dell'edificio, identificare le dinamiche, i modelli operativi e le anomalie.

Miller et al [18] introduce un processo di digitazione giornaliera, chiamato *DayFilter*, che utilizza le tecniche di SAX, estrazione di motif e discord, e clustering per rilevare la struttura intrinseca dei dati di performance degli edifici. In particolar modo gli schemi frequenti vengono rilevati e raggruppati attraverso il k-means, mentre quelli infrequenti vengono filtrati e contrassegnati per un'analisi più approfondita e dettagliata in modo da individuare le potenziali opportunità di risparmio energetico. Il processo, basato sulla somiglianza delle prestazioni quotidiane, trasforma i dati grezzi quantitativi in sottogruppi qualitativi e li visualizza utilizzando tecniche espressive.

Nello stesso anno McLoughlin et al. [19], considerando circa 4000 abitazioni in Irlanda, ha individuato una metodologia atta a classificare ciascun nucleo famigliare, e di conseguenza il loro utilizzo energetico, in funzione delle loro caratteristiche (ad esempio composizione del nucleo famigliare ed età del capofamiglia). In una prima fase di clustering, tre differenti tecniche sono state considerate, k-means, k-medoid e Self Organising Maps, SOM. Attraverso l'utilizzo dell'indice di Davies-Bouldin quest'ultima risulta essere quella che meglio classifica i consumi energetici. La procedura di clusterizzazione è stata effettuata

<sup>&</sup>lt;sup>16</sup> I motif sono dei modelli sequenziali ricorrenti che si ottengono dall'applicazione della tecnica SAX. In contrasto ad esso vi sono i discord che invece sono modelli sequenziali non frequenti.

separatamente per ogni giorno, considerando un periodo di sei mesi, e nove cluster sono stati ottenuti. Basandosi su queste analisi sono state individuate 10 classi di profili di utenza e attraverso un modello regressivo multi-nominale ciascuna classe è stata associata ai nuclei famigliari.

Do Carmo e Christensen [20] attraverso il clustering e la regressione lineare hanno valutato il profilo energetico tipico giornaliero, dovuto al riscaldamento, di 139 abitazioni in Danimarca dotate di una pompa di calore, e la relazione che sussiste fra la domanda energetica e i fattori fisici e socioeconomici. In particolar modo la fase di clustering ha portato alla segmentazione dei dati in base all'intensità della domanda, alta, media e bassa, e alla tipologia di giorno, lavorativo o feriale. Per ciascuna segmentazione sono stati identificati due cluster: il primo, quello più comune, mostra un andamento piuttosto costante, il secondo invece è caratterizzato da una maggiore variazione durante il giorno. Il modello regressivo ha dimostrato una precisione nella previsione dei consumi di circa il 77% e indica che il modello tipico giornaliero dipende: dalle dimensioni e dall'anno di costruzione dell'edificio, dalle caratteristiche del sistema di distribuzione e da fattori socioeconomici come il numero di bambini o adolescenti. Allo stesso modo, Rhodes et al. [21] considerando 103 case situate ad Austin, Texas, ha determinato una curva di domanda energetica media stagionale per ciascuna residenza, i cluster distintivi per ogni stagione, e infine, attraverso un modello di regressione binaria ha identificato le variabili che contribuiscono maggiormente al consumo energetico.

Ma et al. [22] attraverso l'utilizzo di tecniche di cluster, ha identificato il profilo di carico tipico giornaliero, relativo al consumo di energia termica per il riscaldamento, di 19 edifici scolastici. In particolar modo il coefficiente di correlazione di Pearson, PCC, è stato utilizzato per raggruppare i profili in funzione della loro variazione, durante l'arco di tempo considerato, e non della loro intensità. Attraverso l'utilizzo di un cluster gerarchico, considerando la distanza euclidea come misura di dissimilarità, sono poi stati categorizzati gli edifici aventi un comportamento simile circa l'utilizzo energetico. L'uso di tale strategia ha permesso una conoscenza dettagliata del profilo energetico tipico: essa fornisce informazioni circa i picchi, la domanda giornaliera e quindi gli orari di accensione e spegnimento e permette inoltre lo sviluppo di strategie di retrofit. L'utilizzo di tale

coefficiente di dissimilarità fa sì che i profili di carico all'interno di ciascun cluster siano simili in termini di variazione e non di intensità, come invece accade utilizzando la distanza euclidea, e inoltre permette un migliore confronto fra edifici diversi.

Capozzoli et al. [23], invece, attraverso l'utilizzo di clustering gerarchico, con metodo di Ward, ha caratterizzato il funzionamento tipico giornaliero di una centrale termica del Politecnico di Torino: tramite l'utilizzo di misure qualitative come l'indice di Davies Bouldin e di Silhouette quattro cluster sono stati individuati. In una prima fase, ciascun profilo è stato normalizzato sul suo massimo: in questo modo i cluster ottenuti tengono conto della forma dei profili e non della loro intensità, in termini di potenza. Attraverso il metodo CART e considerando come variabili esplicative il mese e il giorno, i quattro cluster sono stati classificati, all'interno dell'anno considerato, con una precisione di circa il 96%. Una seconda fase è invece consistita nel riapplicare la procedura ad uno dei profili precedentemente ottenuti: utilizzando i dati non normalizzati, sono stati ottenuti tre differenti cluster, aventi intensità diversa.

#### 2.2. Riconoscimento e diagnosi delle anomalie

L'individuazione e la diagnosi delle anomalie si basa sul riconoscimento di funzionamenti atipici all'interno dell'intero edificio o di uno specifico sistema impiantistico, ad esempio l'impianto HVAC o quello di illuminazione. In generale se consideriamo un'associazione del tipo A→B, l'anomalia si può individuare nel momento in cui l'osservazione considerata partendo dall'antecedente A non porta al conseguente B [24]. Un esempio di anomalia può essere un picco non frequente all'interno del profilo tipico giornaliero o la presenza di rumori. Individuare le anomalie e soprattutto le condizioni al contorno che le caratterizzano permette di attuare delle strategie preventive e quindi migliorare la gestione del sistema edificio-impianti. Le tecniche utilizzate in questo campo sono molteplici: clustering, LDA¹¹ e molti dei processi di riconoscimento degli errori prevedono come fase

\_

<sup>&</sup>lt;sup>17</sup> La sigla LDA viene utilizzata per indicare la Linear Discriminant Analysis, ovvero analisi discriminante lineare.

iniziale l'individuazione di un profilo di carico tipico che viene utilizzato come termine di paragone per classificare un evento come anomalia o non.

Nel 2011 Han et al. [25] ha utilizzato un approccio ibrido che vede l'utilizzo di tecniche quali PCA, per estrarre le caratteristiche significative dai dati a disposizione, e SVM per il rilevamento e la diagnosi dei guasti dei sistemi di refrigerazione. Considerando che SVM non può essere utilizzato per risolvere direttamente problemi di classificazione multiclasse, sono stati analizzati e confrontati diversi algoritmi classici multi-SVM e infine è stato adottato l'algoritmo "Uno contro gli altri". Il modello ibrido PCA-SVM è stato presentato e convalidato dai dati storici. I risultati mostrano che tale approccio riesce ad isolare il normale funzionamento e ha un tasso diagnostico non inferiore al 98,57%, che è migliore del modello SVM senza PCA. Un approccio ibrido è stato utilizzato anche da Chou e Telega [26]. Attraverso l'algoritmo k-means i profili di consumo giornaliero sono stati classificati all'interno della settimana e successivamente le anomalie sono state identificate in due fasi, tramite un modello ibrido di rete ARIMA<sup>18</sup>. La prima fase è consistita nella creazione di un modello predittivo dei consumi reali. La seconda fase invece, si basa sull'identificazione delle anomalie come differenza fra i consumi stimati e quelli reali. I risultati sperimentali, considerando un ufficio e un lasso temporale pari a 17 settimane, confermano che il metodo è in grado di rilevare valori anomali in tempo reale.

Mulumba et al. [27] propone un metodo di diagnosi dei guasti basato sull'applicazione delle tecniche di SVM per modellare i parametri, utili per il riconoscimento di anomalie, ricorsivamente calcolati da uno stimatore online. Quest'ultimo presuppone un modello auto regressivo di serie temporali con variabili esogene, ARX<sup>19</sup>, che quindi osserva il comportamento dei soli input e output e non la relazione fra gli stessi. Tale metodologia è stata applicata ad un dataset relativo ad un'unità di trattamento aria al fine di individuare, sotto forma di vettore, l'effetto marginale inosservabile che ciascuna variabile esogena ha sulle quelle dipendenti, proprie del sistema. Ciascun vettore viene ottenuto combinando,

<sup>18</sup> La sigla ARIMA corrisponde a AutoRegressive Integrated Moving Average, ovvero un modello autoregressivo integrato a media mobile.

<sup>&</sup>lt;sup>19</sup> La sigla ARX corrisponde a AutoRegresive model with eXternal input, ovvero un modello autoregressivo con input esogeni

alle variabili provenienti dai dati originari e misurate con un timestep regolare, intervalli opportunamente selezionati tramite una preelaborazione con modello ARX. L'insieme di vettori di parametri è suddiviso in classi da SVM, consentendo la classificazione dell'errore.

Li et al. [28] nel 2016 ho proposto una metodologia, atta a diagnosticare le anomalie che possono occorrere durante il funzionamento di un chiller, che vede l'utilizzo di tecniche quali clustering e analisi discriminante lineare (LDA) e si suddivide in due fasi. Durante la prima fase, attraverso l'algoritmo k-means, otto cluster sono stati individuati: uno di questi corrisponde al normale funzionamento mentre i restanti sette sono collegati ai tipi di malfunzionamento del chiller. L'algoritmo LDA è poi applicato in modo da ridurre dimensionalmente il dataset senza però perdere le informazioni importanti in esso contenute. La tipologia di anomalia e il livello di severità vengono diagnosticati se esse appartengono ad uno dei cluster precedentemente classificati come malfunzionamenti. L'utilizzo di tale approccio permette non solo la conoscenza dell'errore di funzionamento ma anche la sua incidenza. Nello stesso anno Beghi et al. [29] ha proposto una strategia diversa che vede l'utilizzo della PCA per individuare il funzionamento tipico di un chiller e le condizioni al contorno che lo caratterizzano. In particolar modo attraverso un indice ottenuto dalla combinazione della variabile casuale T<sup>2</sup> di Hotteling e l'errore quadratico di previsione, SPE, vengono fissati dei limiti di controllo per il normale funzionamento. Se l'indice dell'osservazione x non rientra nei limiti allora essa sarà classificata come anomalia.

Guo et al. [30] ha invece applicato una strategia di diagnosi a un sistema di condizionamento a portata variabile, VRF, considerando quattro anomalie tipiche. Innanzitutto, il dataset è stato dimensionalmente ridotto: attraverso un preprocessamento dei dati ed una analisi della correlazione, le variabili irrilevanti o ridondanti sono state rimosse. Successivamente, l'applicazione del metodo di regole di associazione ha portato alla selezione delle caratteristiche ottimali. Dalla combinazione di queste fasi, cinque set di caratteristiche sono stati ottenuti e utilizzati come input per il modello di diagnosi delle anomalie basato su una rete neurale. I risultati mostrano che l'efficacia del metodo è strettamente dipendente dal set di caratteristiche utilizzato.

Capozzoli et al. [31] ha proposto una metodologia innovativa che, attraverso le tecniche di SAX e CART, mira ad una classificazione del consumo energetico e ad una individuazione automatica delle anomalie, al fine di definire delle strategie di gestione energetica. Una novità introdotta riguarda l'intervallo delle finestre temporali, che non viene considerato costante ma variabile, la cui determinazione si ha attraverso l'utilizzo di alberi di regressione. Il grande vantaggio di tale approccio è la sua caratteristica di autoregolazione del modello che permette di identificare il numero e l'ampiezza delle suddette finestre al fine di ridurre le serie temporali. Inoltre, il numero di simboli, propri dell'algoritmo SAX, che servono per codificare i profili temporali, è stato eguagliato al numero di finestre temporali giornaliere. Ciascun profilo temporale ottenuto viene utilizzato come input per la fase d'individuazione delle anomalie. In particolar modo, un funzionamento non normale verrà classificato come tale se la probabilità di occorrenza, di ciascun simbolo in ogni finestra temporale e sotto specifiche condizioni al contorno, è bassa.

#### 2.3. Previsione dei consumi

La previsione dei consumi energetici è di fondamentale importanza per la riduzione degli stessi in quanto permette di valutare diverse alternative di progettazione e di funzionamento, in termini di efficienza energetica, degli edifici. Tuttavia, la sua determinazione resta un compito impegnativo a causa della varietà dei fattori influenzanti. I metodi principali per la sua determinazione sono due: un approccio fisico, basato sulla modellazione del sistema considerato e strettamente correlato alle leggi termodinamiche che lo caratterizzano, e un approccio basato sui dati, data-driven, che al contrario non richiede una conoscenza dettagliata del sistema ma apprende dai dati storici [32]. In letteratura si trovano molti studi effettuati con entrambe le metodologie. Ad esempio, il software EnergyPlus è stato utilizzato per valutare: i vantaggi, in termini di risparmio energetico, ottenibili dall'installazione di un impianto HVAC condiviso, per un edificio residenziale ed uno adibito ad uffici commerciali [33] o l'incidenza sul consumo energetico di varie tipologie di impianti di riscaldamento e raffrescamento, ventilconvettori, sistemi a

volume d'aria variabile [34] e a portata variabile [35]. Tuttavia, tali studi considerano delle condizioni operative standard e non tengono conto dell'aleatorietà dei fattori umani.

Per quanto riguarda invece l'utilizzo dell'approccio data-driven, gli studi in letteratura si concentrano soprattutto sull'individuazione della tecnica migliore. Zhao et. Al [36] ha testato tre tecniche di data mining per stimare i consumi energetici di sistemi a volume di refrigerazione variabile, VRV, considerando come indicatore l'intensità di consumo energetico, in modo da poter confrontare edifici differenti fra di loro per area e locazione. Tre modelli predittivi sono stati sviluppati considerando gli algoritmi di ANN, SVM, e ARIMA, e l'accuratezza di ciascuno di essi è stata valutata tramite lo scarto quadratico medio RMSE, l'errore quadratico medio MSE, e l'errore medio assoluto percentuale MAPE. I risultati mostrano un'accuratezza maggiore per il modello costruito tramite ANN, seguito da quello ottenuto tramite SVM. Il modello ottenuto tramite ARIMA è in grado di stimare con buona precisione il funzionamento del sistema quando esso lavora a regime, ma presenta delle deviazioni significative quando l'impianto è spento o lavora a carico parziale.

Yan et al. [37] ha invece effettuato una previsione delle prestazioni di una pompa di calore, GSHP, installata all'interno di un edificio terziario. I dati di monitoraggio sono stati preprocessati in modo da considerare solo le variabili importanti da usare come input per i tre modelli, che variano in base al modo di funzionamento e alla stagione di riferimento. Per ognuno di essi è stato costruito un modello predittivo a breve termine, tramite l'algoritmo back-propagation delle reti neurali, BPNN, che viene successivamente utilizzato come input per la previsione a lungo termine. I risultati ottenuti dallo studio dei dati di monitoraggio mostrano che la performance del sistema GSHP è insoddisfacente a causa dell'elevata potenza di pompaggio che riduce l'efficienza del sistema mentre quelli relativi ai modelli neurali, costruiti sui dati monitorati, mostrano una buona accuratezza.

Li et al. [38] ha invece confrontato le tecniche di SVM e di BPNN per predire il carico di raffreddamento di un edificio terziario. Dal confronto dei risultati emerge che lo scarto quadratico medio e l'errore relativo medio del modello ottenuto tramite SVM sono inferiori di circa il 50% di quelli ottenuti tramite BPNN. In [39] per la previsione del consumo energetico, dovuto al riscaldamento di un campus universitario, varie reti neurali artificiali

sono state utilizzate: rete neurale feed forward backpropagation (FFNN), rete neurale radial base function (RBFN) e rete neuro-fuzzy (ANFIS). I dati misurati effettivi vengono utilizzati per la formazione e il test dei modelli e per ogni tipo di rete neurale sono stati analizzati tre modelli, ognuno avente un diverso numero di input. Al fine di migliorare la precisione della previsione, viene creata ed esaminata una rete neurale *ensemble*<sup>20</sup> caratterizzata da tre diverse combinazioni di output. I risultati mostrano che tutte le tecniche utilizzate sono in grado di stimare con elevata precisione il consumo energetico e la rete ensemble si contraddistingue per una migliore accuratezza:

Inoltre, anche nel campo della previsione non mancano gli approcci ibridi. In [41] viene proposta una metodologia che prevede l'integrazione di PCA e WSVM, dove la sigla W indica Weighted ovvero pesato, per la predizione del carico di raffrescamento. La PCA viene utilizzata per ridurre il dataset originario. I dati ottenuti diventano i nuovi input per il modello pesato creato attraverso SVM e vengono inoltre usati per trovare una correlazione fra gli input e gli output. Questo metodo ha il vantaggio di auto adattarsi per fissare i pesi degli indici. Inoltre, ha una ottima capacità di apprendimento e di generalizzazione e una bassa dipendenza dai dati di input se confrontato con altri modelli basati sull'uso di PCA-SVM o sola SVM.

Lee et al. [42] propone un approccio probabilistico per stimare i possibili risparmi energetici ottenibili dalla sostituzione dei chiller in edifici esistenti, tenendo conto dei fattori che influiscono sul risparmio energetico e che includono i carichi di raffreddamento dell'edificio, i sistemi di controllo e funzionamento del sistema, nonché le caratteristiche dell'impianto. L'approccio proposto comporta: l'uso dell'analisi di correlazione per l'identificazione di tali parametri, il software EnergyPlus per simulare l'uso di energia dell'impianto di refrigerazione, e un approccio Monte Carlo per simulare la probabilità di risparmi energetici post-retrofit. Tale approccio è stato testato su un edificio commerciale dove i refrigeratori raffreddati ad aria sono stati sostituiti con quelli raffreddati ad acqua.

<sup>&</sup>lt;sup>20</sup> La rete neurale d'ensemble, ovvero di insieme, è un paradigma di apprendimento in cui un numero finito di reti neurali viene addestrato per lo stesso compito. In generale, la sua costruzione prevede due fasi, cioè, l'addestramento delle reti che lo compongono e la combinazione degli output. [40]

I risultati mostrano che le variazioni nei risparmi energetici annuali, ottenibili dalla sostituzione dei chiller, possono essere stimati con una buona accuratezza. Inoltre, considerando il caso studio, i risparmi ottenibili variano da 37.6% al 49.2%. Tale approccio può essere utilizzato dalle ESCO e dai proprietari degli edifici per valutare le possibili soluzioni di efficientamento.

Santamouris et al. [43] presenta una metodologia atta a classificare degli edifici scolastici: la valutazione energetica fornisce informazioni specifiche sulla loro efficienza, il confronto con altri edifici di natura simile e consente inoltre una migliore pianificazione di interventi volti a migliorarne le prestazioni energetiche. Il lavoro è organizzato in tre fasi: durante la prima fase i dati sui consumi sono stati raccolti attraverso indagini energetiche condotte in 320 scuole in Grecia, nella seconda fase gli edifici sono stati valutati attraverso l'utilizzo di clustering fuzzy (ad ogni cluster corrisponde una classe energetica) e nella terza fase 10 scuole sono state selezionate per esaminare maggiormente nel dettaglio il loro grado di efficienza energetica e la qualità ambientale. Considerando le scuole selezionate nell'ultima fase, nove possibili strategie di efficientamento sono state individuate e testate, in modo da valutare i risparmi ottenibili, attraverso l'utilizzo del software TRNSYS, calibrato tramite i dati storici.

Nel presente capitolo sono stati forniti numerosi esempi presenti in letteratura riguardanti la caratterizzazione e la gestione del sistema edificio-impianti, ponendo particolare attenzione agli impianti HVAC. In particolare, tutti i casi descritti si basano sull'utilizzo di dati reali e forniscono indicazioni circa le tecniche e le procedure più consone per l'analisi dei dati stessi ottenuti dai vari sistemi di monitoraggio. In particolar modo possiamo affermare che la tecnica maggiormente utilizzata per l'individuazione del profilo di carico e delle anomalie coincide con il clustering in tutte le sue eccezioni, e.g. misura di similarità e metodo utilizzato, mentre per la branchia relativa alla previsione e quindi alla stima dei consumi energetici di un edificio si ha una maggiore variabilità delle tecniche utilizzate: si passa dall'utilizzo dell'approccio fisico, che quindi si avvale di software simulativi come ad esempio EnergyPlus, all'utilizzo di un approccio data-driven che si avvale di tecniche quali reti neurali (ANN o BPNN o FFNN), SVM, ARIMA. Non mancano inoltre approcci

ibridi basati sulla simulazione calibrata e che vedono quindi l'utilizzo di entrambi gli approcci, law-driven e data-driven. Nella fattispecie se consideriamo il presente lavoro di tesi possiamo affermare che la fase di caratterizzazione dell'impianto e individuazione delle anomalie operative ha visto l'utilizzo delle tecniche di clustering mentre la fase di previsione dei consumi energetici ha visto l'utilizzo di un approccio ibrido basato sulla simulazione calibrata.

Nel capitolo successivo verranno descritti in maniera più approfondita i possibili approcci utilizzabili, law-driven e data-driven nonché le tecniche di classificazione e clustering.

## 3. Possibili approcci di diagnosi energetica del sistema edificioimpianti.

Per diagnosi energetica si intende una analisi prestazionale sul sistema edificio-impianto al fine di determinarne i flussi energetici, termici ed elettrici, attraverso un modello matematico, che rappresenti al meglio il sistema stesso e che si può affermare essere caratterizzato da tre componenti:

- Input: tutto ciò che agisce sul sistema, come ad esempio i fattori fisici;
- Struttura del sistema e sue proprietà: una descrizione dal punto di vista fisico;
- Output: variabili dipendenti che descrivono la reazione del sistema in funzione degli input, ad esempio il consumo di energia.

I possibili approcci di audit energetico sono due e si differenziano in funzione dello scopo [44]:

- Approccio diretto, o forward approach, prevede una modellazione dettagliata dell'edificio e degli impianti HVAC al fine di ottimizzarne il design;
- Approccio indiretto, o data-driven approach, prevede una modellazione dell'uso di energia di edifici esistenti per stabilire dei valori di riferimento e calcolare dei retrofit di risparmio.

L'immagine sottostante mostra la differenza dal punto di vista applicativo dei sopracitati approcci, Figure 3.1, di cui si parlerà nei seguenti paragrafi.

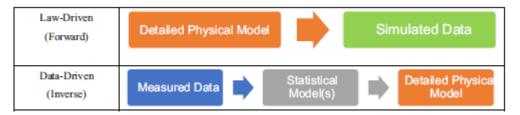


Figure 3-1: Different methods of Forward and Data-Driven approach [45].

#### 3.1. Approccio diretto

L'approccio diretto, o forward approach, è una metodologia che si avvale di software di simulazione energetica per stimare gli output, o variabili dipendenti, del sistema edificio-impianto, noti input e struttura del sistema stesso. I modelli ottenuti da tale metodo vengono anche definiti law driven, in quanto essi applicano un determinato insieme di leggi, fenomeni fisici, che governano il sistema edificio-impianti, ad esempio la conduzione del calore, per prevederne il comportamento energetico e termico [45]. Esso presuppone una conoscenza dettagliata dei fenomeni fisici che regolano gli scambi che avvengono all'interno del sistema edificio-impianti, della loro intensità, dell'edificio che si vuole studiare e del software utilizzato, metodologia e limiti, in quanto l'incomprensione di uno di questi aspetti può portare a risultati errati o irreali.

Nato in ambito accademico esso si è poi diffuso anche nel settore professionale per i notevoli benefici che esso presenta fra cui emerge soprattutto la possibilità di stimare il comportamento dell'edificio anche se esso non è ancora stato costruito: viene infatti ampiamente utilizzato in fase di progettazione al fine di scegliere le migliori soluzioni tecnologiche, sia dal punto di vista impiantistico che architettonico.

#### 3.1.1. Metodologia

Al fine di una più corretta comprensione, prima di parlare della metodologia di cui si avvale l'approccio diretto occorre fornire dei cenni sulle tipologie di simulazione possibili e i relativi software.

Per quanto concerne i programmi simulativi, si può affermare che essi nascono intorno al 1970 dall'implementazione di procedure di tipo manualistico. Clarke nel 2001 [46] individua quattro generazioni, le cui peculiarità vengono schematizzate nella tabella sottostante, Table 3.1-1

Prima	- Approccio da manuale	- Risultati indicativi			
generazione	- Semplificato	- Applicazioni limitate			
8	- Utilizzo limitato a specifici campi	- Difficoltà di utilizzo			
Seconda	- Dinamica più importante				
generazione	- Meno semplificato				
8	-Ancora di utilizzo limitato				
Terza	- Soluzione diversa in base al tipo di				
generazione	problema				
8	- Utilizzo di metodi numerici				
	- Modellazione integrata				
	- Interfaccia utente grafica				
	- Interoperabilità parziale delle funzioni	~			
Quarta	- Interoperabilità completa delle funzioni	- Predittivo			
generazione	- Metodi numerici avanzati	- Generalizzato			
8	- Basato su una conoscenza intelligente	- Facile da usare			
	- Software ingegneristici avanzati				

Table 3.1-1Evolution of design tools, [46].

I metodi utilizzati dai programmi delle prime due generazioni sono semplificati in quanto essi operano in regime stazionario: le funzioni matematiche implementate, basandosi su molte ipotesi semplificative, sono basilari. Inoltre, il loro scopo non è quello di rappresentare fedelmente i flussi termici ed energetici, che si verificano all'interno di un edificio, ma solo di fornire delle indicazioni generali sulle performance o risolvere problematiche specifiche come ad esempio il dimensionamento dei canali d'aria o la determinazione dei carichi termici. L'assenza di una interfaccia grafica inoltre li rende anche di difficile utilizzazione.

I software di terza e quarta simulazione, utilizzati attualmente, corrispondono invece ai metodi dinamici: l'intero sistema edificio-impianti viene simulato attraverso procedure analitiche e numeriche in modo da modellare simultaneamente non solo i flussi termici ed elettrici ma anche quelli luminosi ed acustici, nonché il comportamento dell'utenza. Nonostante la presenza di un'interfaccia grafica che rende più intuitiva la modellazione, tali software richiedono un'esperienza da parte dell'utilizzatore, soprattutto nella fase di inserimento dei dati di input.

Per quanto riguarda invece le simulazioni energetiche, esse solitamente vengono classificate in funzione delle semplificazioni che vengono adottate per la rappresentazione

del sistema edificio-impianti, dal punto di vista fisico, e dell'unità temporale che viene impiegata per analizzare i fenomeni coinvolti. È proprio quest'ultimo fattore che influenza maggiormente il livello di accuratezza dei risultati, e che porta all'individuazione di tre tipologie di simulazioni energetiche, ovvero:

- Simulazioni energetiche stazionarie, l'unità temporale in questo caso coincide con l'intera stagione che si sta considerando, riscaldamento o raffrescamento. Sono le più semplici dal punto di vista applicativo in quanto richiedono dei valori medi per i dati di input, come ad esempio le temperature esterne ed interne. Tuttavia, essi sono anche i meno accurati proprio in virtù delle ipotesi semplificative che essi assumono.
- Simulazioni energetiche semi-stazionarie, si discostano leggermente dalle precedenti in quanto le analisi vengono effettuate su base mensile ma assumono le stesse ipotesi semplificative. In entrambe queste tipologie il bilancio energetico viene effettuato considerando delle condizioni stazionarie nell'arco della giornata: fenomeni come l'inerzia termica o lo sfasamento, di cruciale importanza per i consumi finali ed il comfort totale, non vengono quindi considerati.
- Simulazioni energetiche dinamiche, essendo l'unità di tempo oraria o sub-oraria, tali simulazioni forniscono risultati più accurati. Esse infatti considerano la variabilità dei fattori che incidono sul comportamento di un edificio come l'inerzia termica, gli apporti interni, le condizioni di utilizzo degli ambienti, gli apporti solari e le condizioni climatiche. Tuttavia, a maggiore accuratezza dei risultati corrisponde anche una maggiore complessità nella costruzione del modello: i dati richiesti sono molteplici e talvolta non reperibili.

Nonostante i diversi regimi in cui è possibile effettuare una simulazione energetica, e quindi il grado di complessità dei modelli, dal punto di vista processuale l'iter resta invariato e si compone principalmente di 3 fasi:

 Calcolo del carico termico richiesto dall'ambiente, dove per carico termico si intende la quantità di calore che deve essere fornita o sottratta dall'ambiente in modo da mantenere una condizione di comfort. Le simulazioni stazionarie o semistazionarie calcolano tale parametro considerandolo come funzione della sola differenza fra le temperature esterne ed interne. Quelle dinamiche invece considerano anche gli apporti solari e quelli interni, l'inerzia termica e l'effetto del vento.

- Calcolo del carico richiesto dai sistemi secondari. Durante questa fase vengono
  calcolate tutte le perdite, o i guadagni, degli impianti di distribuzione e
  trasmissione. Vengono inoltre determinate tutte le forme di energie richieste dal
  sistema secondario, come ad esempio ventilatori o pompe.
- 3. Calcolo del carico richiesto dal sistema principale in modo da poter soddisfare i carichi richiesti e la domanda di picco. Tale fase fa riferimento ai sistemi di generazione e considera il grado di efficienza degli stessi come carico parziale richiesto. Inoltre, è necessario considerare le differenti energie, elettrica, a gas naturale o ad olio: in alcuni casi infatti, come ad esempio in Italia, occorre convertire tali energie in energia primaria, considerando dunque la fonte energetica utilizzata, e.g. energia elettrica, gas naturale, biomasse, etc.

Spesso inoltre il processo simulativo si conclude con un'analisi economica in modo da valutare l'efficacia, dal punto di vista economico, delle misure conservative [44].

Le leggi fisiche che governano tale iter sono i fenomeni di conduzione, convenzione e irraggiamento di cui si è parlato nel paragrafo 1.2. Per quanto riguarda gli input, l'ASHRAE [44] fornisce un diagramma di flusso, riportato nella figura sottostante, Figure 3.1-1, rappresentante l'iter processuale e i dati che ciascuna fase necessita, sostanzialmente essi corrispondono ai fattori fisici e umani di cui si è discusso nel paragrafo 1.2.

Gli input necessari per il calcolo del carico termico richiesto dall'ambiente, fase 1, corrispondono ad una descrizione più o meno dettagliata dell'edificio, in base al tipo di simulazione che si sta effettuando. Occorre pertanto fornire indicazioni circa la località dell'edificio, le caratteristiche architettoniche e fisiche, i materiali e le loro proprietà che compongono l'involucro opaco, trasparente e le tramezzature interne, i carichi interni, la ovvero la presenza di corpi scaldanti o illuminanti, il profilo di utenza e quindi la destinazione d'uso e le infiltrazioni. Durante questa fase inoltre l'edificio viene suddiviso per zone termiche, dove con questo termine si intende parte dell'edificio servito dallo stesso impianto, mantenuta a temperatura uniforme e avente stessa destinazione d'uso

nonché profilo di occupazione. Per la fase 1 inoltre è necessario fornire i dati metereologici della località dell'edificio: tali dati solitamente sono già presenti nel software anche se nella maggior parte dei casi si limitano ai capoluoghi di provincia e spesso derivano da interpolazioni.

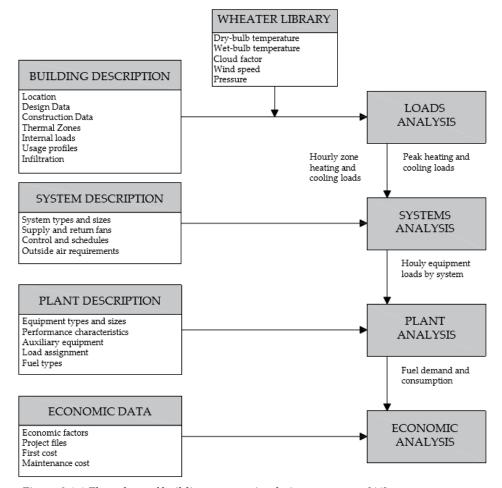


Figure 3.1-1 Flow chart of building energy simulation programs, [44]

Per le fasi 2 e 3 gli input richiesti coincidono con una descrizione degli impianti, quindi dei sistemi di distribuzione, trasmissione, emissione, nonché degli ausiliari e dei sistemi di generazione: vengono indicati non solo le caratteristiche fisiche, ad esempio le dimensioni delle condutture e i materiali che le compongono, ma anche le logiche di controllo, e.g. temperature di set-point o orari di accensione e spegnimento, nonché il tipo di combustibile.

Come già accennato nei capitoli 1 e 2 il problema di tale approccio consiste nella non coincidenza fra i consumi reali e questi stimati ottenuti dalla simulazione. I modelli

considerano delle condizioni standard sia per l'edificio, ad esempio il profilo di utilizzo viene assunto in funzione della sola destinazione d'uso, che per gli impianti, non considerando lo stato manutentivo degli stessi e quindi il reale grado di efficienza.

# 3.2. Approccio inverso

L'analisi energetica attraverso l'approccio inverso prevede la conoscenza preliminare delle variabili in ingresso e in uscita, input e output, attraverso cui andare a determinare il modello matematico del sistema considerato. Questo tipo di approccio viene anche denominato data-driven intendendo con questo termine quella metodologia in cui l'iter processuale è guidato dai dati aventi a disposizione, e non a conoscenze empiriche o a intuizioni [44]. I modelli data-driven per le analisi energetiche possono essere classificati in tre tipologie:

- Black box o empirico: in questo caso il sistema viene considerato come una scatola nera e le relazioni fra gli input e gli output vengono ricavate attraverso delle regressioni semplici o multilineari. Tali regressioni possono essere puramente statistiche o possono basarsi su formulazioni ingegneristiche sull'utilizzo di energia all'interno dell'edificio;
- Simulazione calibrata: questo approccio prevede l'utilizzo di un software di simulazione energetica calibrato appunto attraverso i dati reali in modo tale che gli output stimati corrispondano il più possibile ai dati reali;
- Grey-box: tale approccio prevede come primo step la formulazione di un modello fisico che rappresenti la struttura del sistema analizzato e come secondo step l'identificazione e l'analisi statistica dei parametri maggiormente rappresentativi.

Nonostante le differenze dal punto di vista applicativo, tutti i modelli presuppongono un analisi volta all'estrazione delle relazioni e dunque della conoscenza insita all'interno dei dati. Ciò è possibile grazie alle tecniche di data mining, la cui traduzione è appunto estrazione di dati. Inoltre, come già accennato in precedenza, il presente lavoro di tesi prevede la costruzione di un modello calibrato sui dati storici al fine di poter stimare con

maggiore precisione i consumi energetici della centrale frigorifera del Politecnico di Torino (CF1). Tale operazione presuppone tuttavia una conoscenza approfondita delle modalità operative della CF1. Pertanto, al fine di caratterizzare l'impianto oggetto del caso studio sono state utilizzate le tecniche di Data Mining di classificazione e clustering, spiegate nei successivi paragrafi.

### 3.2.1. Classificazione

Per classificazione si intende il processo che ha il compito di apprendere una funzione target, f, in grado di assegnare a ciascuna istanza, x, un'etichetta di classe y [47]. Se ad esempio volessimo classificare i vertebrati come mammiferi o non, allora questi ultimi corrisponderebbero alle etichette di classe y mentre gli attributi imposti, x, potrebbero essere la temperatura corporea o il metodo di riproduzione. Le tecniche di classificazione sono molteplici, si passa dalle reti neurali, ai classificatori bayesiani ed agli alberi decisionali. In generale esse prevedono un approccio sistematico che, attraverso un algoritmo di apprendimento, sia in grado di individuare un modello classificatore che si adatti al meglio ai dati di input e che può essere:

- Descrittivo: esso è quindi uno strumento esplicativo avente il compito di distinguere istanze appartenenti a classi differenti, come nell'esempio precedente;
- Predittivo: utilizzato per predire la classe di appartenenza di record sconosciuti.

Al contempo l'algoritmo di apprendimento deve avere buone capacità di generalizzazione e deve essere in grado di carpire le relazioni, o regole, che intercorrono fra gli input e gli output: solitamente tali regole sono del tipo "se è vero una certa condizione dell'attributo allora la classe di appartenenza sarà questa". In figura 3.2-1 viene mostrato l'approccio generico utilizzato per la costruzione di un modello classificatore generico.

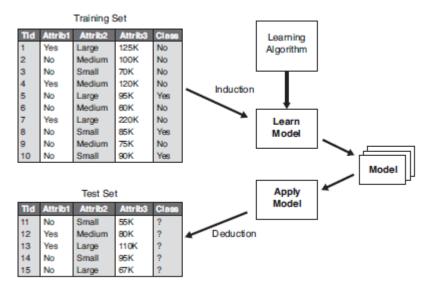


Figure 3.2-1General approach to build a classification model, [47].

#### 3.2.1.1. Modelli decisionali ad albero

Gli alberi di decisione rappresentano il modo più semplice per classificare degli oggetti o pattern in un numero finito di classi [48]. Essi sono caratterizzati da una struttura gerarchica del tipo top-down che vede l'alternanza di nodi, etichettati con il nome degli attributi, e archi, ovvero i rami dell'albero definiti dai possibili valori degli attributi stessi. In particolare, un albero di decisione sarà formato da un solo nodo radice, dei nodi interni, rappresentanti gli attributi attraverso cui suddividere le istanze in categorie, e dei nodi foglia, coincidenti con le categorie stesse. Se ad esempio volessimo classificare, tramite un modello ad albero, l'esempio precedente riguardante la suddivisione di vertebrati come mammiferi o non allora il modello corrispondente sarebbe quello riportato nell'immagine sottostante, figura 3.2-2.

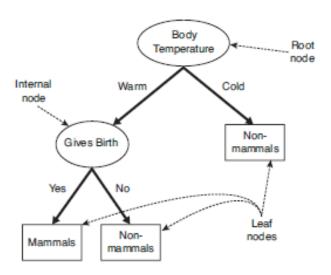


Figure 3.2-2: A decision tree for the mammals classification problem, [47].

Uno dei problemi principali della costruzione dei modelli classificatori è che il numero di alberi decisionali che si possono costruire, a partire da un set di attributi, varia in maniera esponenziale, pertanto la costruzione del modello ottimale è computazionalmente infattibile. Tuttavia, sono stati elaborati degli algoritmi in grado di costruire, in tempi ridotti, dei modelli classificatori aventi buona accuratezza: essi, di solito, utilizzano una strategia locale per decidere quale attributo suddivide in maniera ottimale i dati. Uno di questi è l'algoritmo di Hunt che è alla base di tecniche quali ID3, C4.5 e CART.

Nell'algoritmo di Hunt l'albero decisionale viene costruito ripartendo i record di allenamento in sottoset sempre più omogenei, attraverso l'applicazione recursiva di 2 step. In particolar modo, sia  $D_t$  il training set di record associati al nodo t e  $y = \{y_1, y_2, ..., y_c\}$  le etichette di classe, allora:

STEP 1: Se tutti gli oggetti di  $D_t$  appartengono alla classe  $y_t$  allora t è un nodo foglia.

STEP 2: Se  $D_t$  contiene oggetti che appartengono a classi differenti allora un altro attributo verrà scelto per suddividere i record in sottoset più piccoli. Un nodo figlio viene creato per ciascun outcome e i record in  $D_t$  vengono distribuiti all'interno dei nodi figli.

In generale quindi un algoritmo di apprendimento deve affrontare due problemi:

- 1. Come splittare, e quindi separare, gli oggetti di training? Ciascuno step recursivo nella costruzione dell'albero deve selezionare un attributo per dividere i records in dataset più piccoli.
- 2. Quando si deve interrompere la procedura di split? Una condizione di stop deve essere fornita all'algoritmo.

Risolvere il primo problema equivale a costruire un albero decisionale, andando a fornire un metodo che specifichi una condizione di test per i differenti tipi di attributi ed una misura obiettiva in grado di valutare la bontà di ciascuna condizione di split. Nello specifico, gli attributi possono essere:

- binari: ciascuna condizione di test produce due risultati.
- nominali: poiché tali attributi, per loro natura, possono avere molti valori, la condizione di test può essere espressa in due modi ovvero considerando o uno split multiplo, dove il numero di outcome dipende dal numero di valori distinti dell'attributo, oppure una suddivisione binaria, come avviene per il CART in quanto esso considera 2<sup>(k-1)</sup> -1 modi per splittare, dove k è il numero di attributi.
- ordinati: caratterizzati da serie dotate di un ordine, ad esempio l'altezza come: basso, medio, alto. Possono produrre split binari o multipli e la suddivisione continua finché il raggruppamento raggiunto non vìola le proprietà dell'ordine dei valori degli attributi.
- continui: per questi attributi la condizione di test può essere espressa come un test
  di comparazione (A < v) o (A ≥ v) con split binari o come dei range v<sub>i</sub> ≤ A < v<sub>i+1</sub>,
  dove i= 1,...,k. In realtà l'utilizzo di tali attributi presuppone un modello di
  regressione in quanto la classificazione si ha per attributi discreti. Tuttavia, ci sono
  algoritmi, come il CART, che sono in grado di maneggiare contemporaneamente
  sia attributi discreti che continui.

Per quanto riguarda le misure che possono essere scelte per selezionare il miglior partizionamento dei dati, esse sono di solito definite considerando il grado di impurità dei nodi figli. Siano p(i|t) la frazione di record che appartengono alla classe i in un dato nodo t e c il numero delle classi, esempi di misure di impurità possono essere:

Entropia(t) = 
$$-\sum_{i=0}^{c-1} p(i|t) \log_2 p(i|t)$$
  
Gini(t) =  $1 \cdot 1 - \sum_{i=0}^{c-1} [p(i|t)]^2$ 

Errore di classificazione(t) = 1-  $\max_{i}[p(i|t)]$ 

Un ulteriore misura, utilizzata per determinare la bontà della condizione di split, è il gain  $\Delta$ , i.e. guadagno. Siano I () la misura di impurità in un dato nodo, N il numero totale di oggetti nel nodo genitore, parent node, k il numero di attributi ed N ( $v_j$ ) il numero di record del nodo figlio ( $v_i$ ), il gain si ottiene come:

$$\Delta = I(parent) - \sum_{j=1}^{k} \frac{N(v_j)}{N} I(v_j)$$

Il gain, dunque, confronta il grado di impurità del nodo genitore e quella del nodo figlio rispettivamente prima e dopo la procedura di split: maggiore sarà la differenza fra i due, e quindi Δ, migliore sarà la performance e la media ponderata delle impurità dei nodi figli sarà minima. Occorre inoltre aggiungere che la scelta dell'attributo ottimale varia in funzione della misura di impurità considerata: l'algoritmo CART ad esempio utilizza l'indice di Gini per valutare la migliore condizione di split, mentre il C4.5 e l'ID3 utilizzano l'entropia.

Dopo aver selezionato gli attributi che, per ciascuno split, suddividono al meglio il dataset di partenza in sotto set più piccoli occorre fornire una condizione di stop per terminare la procedura iterativa [49]. Essenzialmente ci sono due possibili strategie: la prima consiste nel continuare a creare dei nodi finché tutti i record appartengono alla stessa classe, la seconda invece prevede di interrompere la procedura recursiva quando il numero di record è inferiore ad una determinata soglia. Il processo di classificazione termina con una fase di pruning, ovvero potatura: un albero decisionale di dimensioni eccessive oltre ad essere di difficile interpretazione dal punto di vista predittivo, può anche essere soggetto al fenomeno di Overfitting, sovradattamento.

# 3.2.1.2. Overfitting

L'obiettivo di un modello classificatore è quello di riuscire ad applicare ad un set di oggetti non noti, le conoscenze acquisite durante la fase di allenamento. Affinché ciò avvenga esso dovrà essere caratterizzato da un basso errore sia di training, dove per esso si intende un errore dovuto ad una errata classificazione durante la fase di allenamento, che di generalizzazione durante la fase predittiva, basata dunque su oggetti non noti. L'entità di tali errori, e di conseguenza i fenomeni di overfitting e underfitting, dipendono essenzialmente dalla complessità del modello. Un modello complesso solitamente è caratterizzato da un errore di training molto basso in quanto esso si è perfettamente, o quasi, adattato ai dati utilizzati durante la fase di apprendimento. Tale adesione fra il modello e il set di training comporta una diminuzione della capacità di generalizzazione, e quindi un alto errore di test. Allo stesso modo un modello poco complesso porta invece ad avere dei valori alti per entrambi gli errori in quanto il set di record, fornito per la fase di apprendimento, non è sufficiente a far emergere la vera struttura dei dati. Il modello così creato pertanto non godrà né di adattabilità che di generalizzazione.

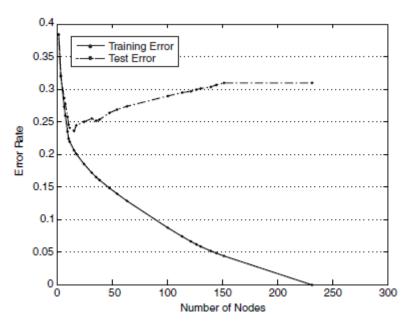


Figure 3.2-3 Training and test error rates. When the size of the tree is very small both the errors are large, other way when the tree is too large the test error begin to increase even if its training error continues to decrease [47].

Se consideriamo l'immagine precedente, figure 3.2 -3, riportante gli andamenti dei due errori in funzione del numero di nodi, notiamo che quando l'albero è molto piccolo entrambi gli errori sono elevati.

In questo caso si parla di underfitting, ovvero sotto-adattamento del modello. Tale situazione è dovuta al fatto che il modello non ha ancora colto la vera struttura dei dati, pertanto non godrà delle proprietà di adattabilità, durante la fase di training, e di generalizzazione, durante la fase predittiva. All'aumentare dei nodi entrambi gli errori diminuiscono sino a che, arrivati ad una certa dimensione del modello, l'errore di allenamento continuerà a diminuire, mentre quello di test ricomincerà ad aumentare. Tale condizione prende il nome di overfitting.

Il fenomeno di overfitting è tipico dei modelli che inconsapevolmente considerano parte del rumore, presente all'interno dei dati, come se esso stesso rappresenti la struttura del modello. Esso può essere infatti causato dalla presenza di rumori, e quindi oggetti etichettati in modo errato, oppure a causa di una carenza di campioni significativi, che quindi non permette di cogliere la vera struttura dei dati e delle regole che li caratterizzano. Il fenomeno di overfitting risulta quindi un problema sotto più punti di vista in quanto, oltre a ridurre l'accuratezza del modello, ne aumenta lo spazio di archiviazione e di computazione e di conseguenza la complessità delle previsioni rendendo il modello stesso più difficile da interpretare. Inoltre, i modelli sovra-adattati possono avere una minore precisione sui nuovi dati, rispetto ai modelli non influenzati dal fenomeno di overfitting [50].

Per evitare il fenomeno di overfitting vi sono due possibili strategie attuabili:

• Pre-pruning, o regola dell'arresto precoce. Questo approccio prevede di interrompere l'algoritmo prima che l'albero decisionale generato si adatti perfettamente ai dati di training, ad esempio stoppare l'espansione dei nodi foglia quando il gain ricade al di sotto di una certa soglia. Il vantaggio di tale approccio consiste nell'ottenere alberi meno complessi anche se la scelta del limite ottimale, per la procedura di arresto precoce, non è di semplice computazione. Un valore troppo alto di tale soglia può portare al fenomeno di underfitting, così come un

- valore troppo basso può non essere sufficiente per evitare il sovradattamento del modello.
- Post-pruning. Questa strategia prevede di tagliare l'albero decisionale, inizialmente costruito in modo tale che esso abbia la massima dimensione, con un approccio botton-up, dal basso verso l'alto. La potatura può essere effettuata attraverso la sostituzione di un sottoalbero tramite o un nuovo nodo foglia, la cui etichetta di classe corrisponde alla categoria a cui appartengono la maggior parte dei record associati al sottoalbero considerato, o tramite il ramo utilizzato più frequentemente sempre nel sottoalbero considerato. La procedura termina quando non ci sono ulteriori miglioramenti e in generale offre migliori risultati rispetto al pre-pruning.

### 3.2.1.3. Valutazione del modello

La valutazione del modello di classificazione si basa sul conteggio dei valori correttamente, o non, predetti. Tali risultati possono essere riassunti in una tabella che prende il nome di matrice di confusione, all'interno della quale ciascuna variabile fij denota il numero di oggetti di classe i predetti come classe j. Se consideriamo la matrice di confusione mostrata nella tabella sottostante, table 3.2 -1, e relativa ad un set di dati avente da tre classi, C1, C2 e C3, allora le tre seguenti regole possono essere utilizzate per analizzarla:

	Predicted Values			
Actual Values		<b>C</b> <sub>1</sub>	C <sub>2</sub>	C <sub>3</sub>
	C <sub>1</sub>	C <sub>11</sub>	C <sub>12</sub>	C <sub>13</sub>
	C <sub>2</sub>	C <sub>21</sub>	C22	C23
	C <sub>3</sub>	C <sub>31</sub>	C <sub>32</sub>	C33

Table 3.2-1 Example of a three-class confusion matrix.

- 1. I valori riportati nella diagonale principale corrispondono alle classificazioni corrette;
- 2. I valori riportati nella i-esima riga corrispondono ai casi appartenenti alla classe C<sub>i</sub>, se ad esempio i = 3 allora gli oggetti della classe C<sub>3</sub> saranno quelli associati alle celle

- C31, C32 e C33. Il numero di casi erroneamente classificati come classe 3, sarà dato dalla somma dei valori di C31 e C32.
- 3. I valori riportati nella i-esima colonna corrispondono ai casi appartenenti alla classe Ci, se ad esempio i = 3 allora gli oggetti della classe C3 saranno quelli associati alle celle C13, C23 e C33. Il numero di casi erroneamente classificati come classe 3, sarà dato dalla somma dei valori di C13 e C23. [51].

Nonostante la matrice di confusione fornisce le informazioni necessarie per determinare la bontà di un modello di classificazione, per semplificare una comparazione fra più modelli possono essere utilizzate delle metriche di performance in modo da riassumere le informazioni contenute nella matrice di confusione con un unico valore. Tali metriche possono essere:

$$Accuratezza = \frac{numero\ di\ previsioni\ corrette}{numero\ totale\ di\ previsioni}$$
 
$$Tasso\ di\ errore = \frac{numero\ di\ previsioni\ non\ corrette}{numero\ totale\ di\ previsioni}$$

Se quindi consideriamo l'esempio precedente avremo:

$$Accuratezza = \frac{C_{11} + C_{22} + C_{33}}{C_{11} + C_{12} + C_{13} + C_{21} + C_{22} + C_{23} + C_{31} + C_{32} + C_{33}}$$

$$Tasso\ di\ errore = \frac{C_{12} + C_{13} + C_{21} + C_{23} + C_{31} + C_{32}}{C_{11} + C_{12} + C_{13} + C_{21} + C_{22} + C_{23} + C_{31} + C_{32} + C_{33}}$$

# 3.2.2. Analisi Clustering

L'analisi clustering è una delle tecniche non supervisionate di data mining più note ed utilizzate che consiste nel suddividere i dati in gruppi significativi, cluster, in base al loro grado di similarità o vicinanza. L'obiettivo è quello di fare in modo che gli oggetti abbiano massima similarità intra-cluster e minima similarità inter-cluster. Per alcuni versi le operazioni di clustering possono essere paragonate a quelle di classificazione seppur con una differenza: se infatti per queste ultime il processo di categorizzazione si avvale di

etichette note precedentemente assegnate agli item presenti all'interno del dataset, nelle operazioni di clustering la classificazione si ottiene solo in funzione del cluster cui appartengono e le etichette derivano dunque solo dai dati. Le tecniche di clustering quindi possono essere utilizzate come mezzo attraverso cui comprendere i dati, e quindi analizzare e descrivere il dataset, in virtù delle loro proprietà di sintesi e compressione dei dati nonché della loro capacità di individuare in modo efficiente gli oggetti più simili fra loro [55].

Le tecniche di clustering sono molteplici e la loro classificazione può essere effettuata considerando la logica di controllo che li governa o i tipi di cluster ottenibili [52]. Se consideriamo il primo di tali attributi allora le classificazioni ottenibili possono essere:

- Hierarchical vs Partitional: un clustering si dice partizionale, partitional, se gli oggetti verranno suddivisi in cluster, non sovrapposti: in questo modo ciascun oggetto apparterrà ad un unico cluster. Se invece consideriamo un cluster come un gruppo di sotto-cluster allora otteniamo un clustering gerarchico, hierarchical: i cluster verranno organizzati in modo da formare una struttura ad albero dove ciascun cluster, corrispondente ad un nodo, sarà dato dall'unione dei suoi sotto-cluster. I cluster gerarchici possono essere visti come un sequenziamento di cluster partitivi ed un particolare cluster può essere ottenuto tagliando l'albero gerarchico a qualsiasi livello e prendendo quindi tutti i membri di quella sequenza.
- Exclusive vs Overlapping vs Fuzzy: in generale si possono avere dei clustering esclusivi, non esclusivi, overlapping, o sfumati, fuzzy. I clustering esclusivi prevedono che ogni oggetto appartenga ad un unico cluster, ci sono tuttavia delle situazioni in cui un oggetto può appartenere contemporaneamente a più gruppi: se pensiamo ad esempio ad una persona all'interno di un'università, allora essa potrà essere tanto uno studente quanto un impiegato, pertanto invece di assegnarlo arbitrariamente a solo una delle due categorie, esso si fa appartenere ad entrambe. E poi vi sono i clustering fuzzy dove tutti gli oggetti appartengono a tutti i cluster e ciascuno di questi è dotato di un indice, peso, che varia da 0, assolutamente non appartiene, a 1, assolutamente appartiene. Nei clustering fuzzy di solito viene imposta la condizione aggiuntiva che la somma dei pesi di ciascun oggetto sia

uguale ad 1: tale strategia è la più appropriata per evitare l'arbitrarietà di assegnare un oggetto ad un solo cluster e può quindi diventare un clustering esclusivo andando poi a valutare i pesi dell'oggetto e assegnandolo al cluster in cui il peso, o la probabilità, è più alto.

Complete vs Partial: il clustering completo prevede l'assegnazione di tutti gli
oggetti ai cluster mentre quello parziale no. La motivazione di tale strategia risiede
nel fatto che alcuni oggetti potrebbero non appartenere ad un gruppo ben definito
in quanto rumori, valori anomali o oggetti non interessanti.

Se invece volessimo classificarli in funzione dei cluster ottenibili allora avremmo:

- Well Separeted: i cluster "ben separati" coincidono con dei set di oggetti aventi similarità intra-cluster massima e inter-cluster minima. Solitamente questa condizione si ha per dati aventi dei cluster naturali.
- Prototype Based: ogni cluster basato su prototipo sarà caratterizzato da oggetti più simili al prototipo definente il proprio cluster, che ai prototipi degli altri cluster. Per i dati con attributi continui il prototipo del cluster è un centroide, che coincide con la media di tutti i punti nel cluster, mentre per dati aventi attributi categorici allora il prototipo sarà il medoide, che corrisponde al punto più rappresentativo del cluster. Un'ulteriore differenza fra il centroide e il medoide consiste nel fatto che mentre il secondo di essi coincide, per sua definizione, con un oggetto appartenente al dataset iniziale, il centroide è un oggetto artificiale. In generale quindi un prototipo può essere visto come l'oggetto che si posiziona più centralmente all'interno di un cluster e, in virtù di ciò, spesso ci si riferisce a questa tipologia come center-based cluster, cluster basati sul centro.
- Graph- Based: se i dati sono rappresentati come un grafico dove i nodi sono gli oggetti e i link le connessioni fra gli oggetti allora un cluster può essere definito come un componente connesso. Un esempio importante sono i cluster basati sulla contiguità dove 2 oggetti sono connessi solo se sono posizionati entro una certa distanza l'uno dall'altro. Questo tipo di clustering è utile quando si ha a che fare con dati irregolari o intrecciati ma può provocare difficoltà se sono presenti rumori.

- Density Based: il cluster si può vedere come una regione di oggetti ad alta densità circondata da una regione a bassa densità. Questo tipo di definizione è spesso utilizzata quando i cluster sono irregolari o intrecciati e sono presenti anche rumori e valori anomali.
- Shared Property or Conceptual cluster: per cluster concettuali si intendono tutti i set di oggetti che condividono le stesse proprietà, da qui shared property. Tale definizione è la più generale in quanto racchiude al suo interno tutte le altre descritte in precedenza: e.g. gli oggetti di un cluster center-based condividono lo stesso centroide o medoide a cui sono più vicini.

Nonostante le varie tipologie di clustering possibili, tutte le tecniche presuppongono l'utilizzo di un indice, misura, di similarità in modo da raggruppare gli oggetti più vicini fra loro all'interno dello stesso cluster. Alcune delle misure più utilizzate sono:

 Distanza euclidea: corrisponde al concetto geometrico di distanza nello spazio multidimensionale e la sua formulazione è la seguente:

$$d(x,y) = \sqrt{\sum_{k=1}^{n} (x_k - y_k)^2} = ||x - y||$$

 Il quadrato della distanza euclidea: viene utilizzato solitamente quando si vuole assegnare un peso progressivamente maggiore agli oggetti che stanno oltre una certa distanza, esso corrisponde dunque a:

$$d(x,y) = \sum_{k=1}^{n} (x_k - y_k)^2$$

• Distanza di Manhattan, o assoluta: corrisponde alla differenza media fra le dimensioni e pertanto verrà calcolata come:

$$d(x,y) = \sum_{k=1}^{n} |x_k - y_k|$$

• Similarità del coseno: viene utilizzata soprattutto per il confronto di testi e la sua formulazione è la seguente:

$$similarity = cos(\theta) = \frac{x \cdot y}{||x|| ||y||}$$

Dove x e y sono due unità statistiche di dimensione n.

Due delle principali tecniche maggiormente utilizzate sono il K-means e il Hierarchial clustering. In particolar modo la tecnica utilizzata nel presente lavoro di tesi coincide con il clustering gerarchico e pertanto solo quest'ultima verrà ulteriormente approfondita.

## 3.2.2.1. Hierarchical Clustering

Come già accennato in precedenza i metodi gerarchici considerano i cluster come unione di più sotto-cluster e il cui risultato finale non è dunque una singola partizione degli n dati inziali ma una serie di partizioni visibili graficamente attraverso un dendogramma o diagramma ad albero avente sull'asse delle ascisse le singole unità e sull'asse delle ordinate il livello di distanza. Tale diagramma quindi mostra le relazioni fra i cluster e i sotto-cluster e l'ordine con cui essi vengono accoppiati o disaccoppiati. I metodi gerarchici infatti possono avere due tipi di approccio:

- Agglomerativo, o bottom-up: la condizione iniziale di tali tecniche considera ogni
  cluster composto da una singola unità e attraverso la fusione iterativa dei cluster
  più vicini fra loro, si arriva ad ottenere un unico cluster onnicomprensivo.
- Divisivo, o top-down: la condizione iniziale prevede un unico cluster onnicomprensivo che viene suddiviso ad ogni step sino a che ogni cluster sia composto da una singola unità.

Nella figura sottostante, figura 3.2-3 viene riportato un esempio di dendogramma e i diversi approcci possibili.

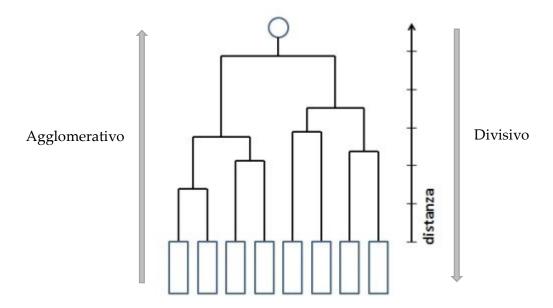


Figure 3.2-4 Example of generic dendogram, with indication of the possible approaches: agglomerative and divisive.

Gli algoritmi gerarchici presenti in letteratura differiscono fra loro per il criterio di prossimità che essi utilizzano per la valutazione delle distanze fra i gruppi ai fini dell'aggregazione. Tuttavia, i loro algoritmi si compongono delle seguenti fasi:

- 1. Computa la matrice di prossimità
- 2. Ripeti lo step 1
- 3. Unisci i due cluster più vicini
- 4. Aggiorna la matrice di prossimità in modo che rifletta la vicinanza fra i cluster nuovi e quelli originali
- 5. Ripeti finché non resta un unico cluster.

# Single Linkage

Il metodo del link singolo utilizza come misura di prossimità la distanza minima fra due qualsiasi punti appartenenti a differenti cluster. Esso predilige la differenza tra i cluster piuttosto che la somiglianza intra-cluster: il dendogramma costruito in questo modo è caratterizzato da rami più corti e da una forma più compatta. Tale tecnica è buona per manipolare forme non ellittiche ma è sensibile a rumori e valori anomali. Nell'immagine

sottostante, figure 3.2-4, è possibile vedere in formato grafico la logica che regola tale algoritmo.

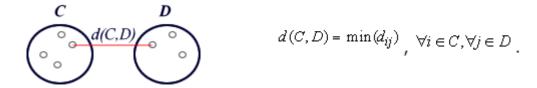


Figure 3.2-5 An illustration of how single linkage method works.

## **Complete Linkage**

In questo algoritmo la prossimità viene definita come la massima distanza tra due punti appartenenti a cluster diversi. Si inizia con un punto per ogni cluster e si aggiungono, uno per volta, i collegamenti fra punti iniziando da quelli più vicini. Questo metodo privilegia l'omogeneità intra-cluster: i dendogrammi presentano rami più lunghi e i gruppi si formano a distanze maggiori. È meno suscettibile ai rumori e ai valori anomali ma favorisce forme globulari e quindi rischia di spezzare dei cluster grandi. Nell'immagine sottostante, figure 3.2-5, è possibile vedere in formato grafico la logica che regola tale algoritmo.

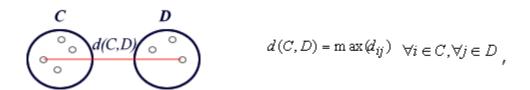


Figure 3.2-6 An illustration of how complete linkage method works.

## **Average Linkage**

È un approccio intermedio fra i due precedenti e la prossimità è intesa come il valore medio delle distanze tra tutti gli elementi appartenenti ai diversi cluster. L'unione fra due cluster avviene sempre considerando la minima distanza così definita. In figure 3.2-6 è possibile vedere in formato grafico la logica che regola tale algoritmo.

$$d(C,D) = \frac{1}{n_1 n_2} \sum_{i=1}^{n_i} \sum_{j=2}^{n_i} d_{ij}, \forall i \in C, \forall j \in D,$$

Figure 3.2-7 An illustration of how average linkage method works.

#### Centroid Method

Tale algoritmo si basa sulla preliminare determinazione dei centroidi di tutti i cluster e la distanza fra essi verrà assunta uguale alla distanza fra i rispettivi centroidi. Nell'immagine sottostante, figure 3.2-7, è possibile vedere in formato grafico la logica che regola tale algoritmo: in particolare  $\overline{X_C}$  e  $\overline{X_D}$  sono i centroidi dei cluster C e D.



Figure 3.2-8 An illustration of how centroid method works.

## Ward's Method

Per il metodo di Ward la prossimità fra due cluster è definita come l'incremento nello scarto quadratico, SSE, che risulta quando due cluster vengono uniti. Come è noto, la varianza può essere scomposta e quindi considerata come la somma tra la varianza intra-cluster e la varianza inter-cluster. Ad ogni step essa verrà dunque calcolata e verranno uniti i cluster dalla cui fusione deriva il minimo incremento della varianza interna, in modo da avere cluster omogenei. Nonostante possa sembrare che questa caratteristica faccia discostare questo metodo dagli altri clustering gerarchici si può dimostrare matematicamente che il metodo Ward è molto simile al link medio quando la prossimità fra due punti è ritenuta come il quadrato della distanza euclidea fra essi [52].

Tutto il processo di aggregazione è visibile tramite il dendogramma: in particolar modo il numero dei k cluster si ottiene attraverso un "taglio" dello stesso ad un dato livello della distanza di fusione. Poiché di solito l'obiettivo è quello di ottenere il minor numero di

cluster con la massima omogeneità intra-cluster l'approccio più usuale consiste nel tagliare il dendogramma lì dove i rami sono più lunghi. Tuttavia, esistono dei metodi come ad esempio l'elbow method, approfondito nel paragrafo successivo, che forniscono una indicazione circa il miglior numero di cluster k a cui tagliare il dendogramma.

I metodi gerarchici permettono dunque di studiare il processo che porta alla creazione dei cluster. Occorre però aggiungere che tali metodi sono onerosi dal punto di vista computazionale e richiedono molto spazio di memoria in quanto tutte le fasi vengono memorizzate.

#### 3.2.2.2. Elbow Method

L'elbow method, i.e. metodo del gomito, è un metodo grafico che considera la variazione totale dello scarto quadratico d'errore interno, WSSE, come funzione del numero di cluster ottenibili. Esso si basa sull'idea che il numero k di cluster scelti dovrebbe essere tale che l'aggiunta di un ulteriore cluster non fornisca una maggiore modellazione dei dati e dunque non migliori il WSSE [53].

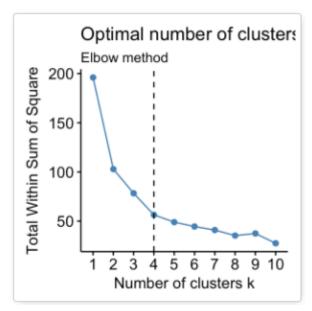


Figure 3.2-9: Example of elbow method applied to a generic dataset.

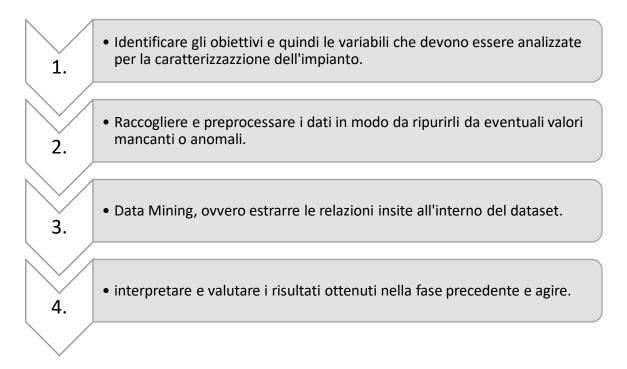
Se consideriamo l'immagine sovrastante riportante un esempio di elbow method applicato ad un generico dataset possiamo notare che lo scarto quadratico totale è inversamente proporzionale al numero di cluster ottenibili: il WSSE diminuisce all'aumentare di k. In particolare, per k=1 il WSSE ha un valore di circa 200 che si riduce drasticamente sino a k=4 dove esso assume un valore pari a 50. Allo stesso tempo per k > 4 possiamo notare un andamento differente in quanto l'aggiunta di ulteriori cluster non modifica in maniera rilevante lo scarto quadrato totale che si mantiene pressoché costante. Il numero di cluster ottimali coinciderà dunque con k=4. Per tale valore inoltre possiamo notare la presenza di un gomito all'interno della curva, da qui criterio del gomito. Tuttavia, occorre anche aggiungere che tale metodo non fornisce sempre una soluzione univoca in quanto il punto di gomito potrebbe non essere evidente oppure vi potrebbero essere più punti di gomito [54].

Il presente capitolo ha fornito delle informazioni circa i possibili approcci di diagnosi energetica, law-driven e data-driven, delle loro peculiarità e dell'iter processuale che ciascuno di essi prevede. In particolar modo, il primo di essi prevede una simulazione del comportamento energetico dell'edificio considerato note le sue caratteristiche. Il secondo, invece, prevede la formulazione di un modello statistico basato sui dati reali che permetta a sua volta la costruzione di un modello fisico dettagliato. Uno strumento utile per il trattamento dei dati e per l'estrazione delle informazioni che essi contengono coincide con le tecniche di Data mining. In definitiva possiamo affermare che sino ad ora è stata fornita un'introduzione alle tematiche e alle tecniche che verranno utilizzate per questo lavoro di tesi. Nei prossimi capitoli verrà invece esaminato il caso studio e l'iter processuale utilizzato.

# 4. La centrale frigorifera del Politecnico di Torino.

Come già accennato nel capitolo precedente l'approccio inverso richiede una caratterizzazione preliminare dell'oggetto che si vuole studiare attraverso l'analisi dei suoi dati caratteristici e quindi l'estrazione delle relazioni insite al loro interno. Tale fase tuttavia si colloca all'interno di un processo che prende il nome di Knowledge Discovery in Dataset, KDD, ovvero estrazione di conoscenza attraverso i dati.

Il processo di KDD si articola in 4 fasi.



Come in ogni ricerca la prima fase consiste nel fissare gli obiettivi che si intendono perseguire. L'obiettivo del presente lavoro di tesi consiste in una valutazione energetica della centrale frigorifera 1 del Politecnico di Torino, andando quindi ad identificare i profili di carico tipici, le eventuali anomalie che si sono verificate e le possibili soluzioni di risparmio energetico. In particolar modo l'iter processuale vedrà una prima fase di caratterizzazione dell'impianto attraverso le tecniche di Data mining precedentemente descritte e l'utilizzo del software open-source R in modo da individuare e classificare i profili di carico caratteristici su base giornaliera e cogliere le relazioni ad essi intrinseche. Una seconda fase prevede invece la costruzione e la validazione di un modello calibrato

sui dati storici, attraverso l'utilizzo del software TRNSYS, che rispecchi quanto più fedelmente possibile l'impianto in modo da poter essere utilizzato per testare eventuali scenari di ottimizzazione e gestione.

Prima di passare alla fase di caratterizzazione vera e propria occorre innanzitutto fornire una descrizione del caso studio dal punto di vista impiantistico e del dataset utilizzato, e quindi le variabili analizzate per caratterizzarlo dal punto vista operativo.

# 4.1. Descrizione dell'impianto

Il caso studio del presente lavoro di tesi consiste in una centrale frigorifera del Politecnico di Torino, Centrale Frigorifera 1 o CF1, caratterizzata da un impianto geotermico operante in sola climatizzazione estiva e servente un edificio adibito ad uffici.

In generale per impianto geotermico si intende un impianto composto da una pompa di calore accoppiata ad una sorgente termica, la cui funzione è quella di produrre acqua, o aria, calda o fredda, in funzione del regime operante e che utilizza come sorgente termica il suolo, l'acqua di falda o i grandi specchi di acqua di superficie (mare, laghi o fiumi). L'utilizzo di tali sorgenti termiche porta innumerevoli vantaggi rispetto ai classici impianti ad aria: nonostante quest'ultima tipologia sia quella più utilizzata, in virtù della disponibilità illimitata di tale fonte, è anche quella meno vantaggiosa dal punto di vista prestazionale. La temperatura dell'aria esterna varia durante l'anno ed è molto bassa nei periodi di maggior fabbisogno termico, causando quindi un decremento del rendimento e delle capacità [55]. Al contrario invece le sorgenti termiche utilizzate dagli impianti geotermici sono caratterizzate da temperature pressoché costanti durante l'anno. Nella figura sottostante, figure 4.1-1, sono riportati gli andamenti delle temperature durante l'anno dell'aria esterna, del suolo, e delle acque di falda.

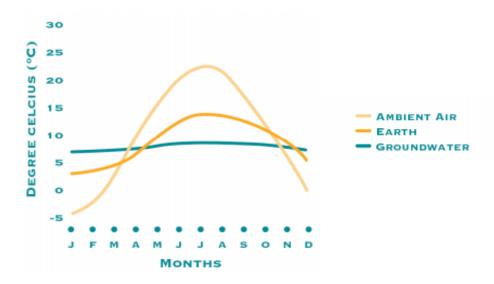


Figure 4.1-1. Development of ambient air, earth and groundwater temperatures troughout year.

L'ASHRAE identifica tre tipologie di impianti geotermici a pompa di calore, in funzione della sorgente termica con cui sono accoppiate [56]:

1. Pompe di calore accoppiate al terreno, Ground-Coupled Heat Pumps o GCHPs, sono caratterizzate da un circuito idraulico chiuso, scambiatore di calore, affogato nel terreno e all'interno del quale circola il fluido termovettore. Tali scambiatori possono essere disposti in verticale o in orizzontale. Nel primo caso, i circuiti coincidono con tubazioni ripiegate ad U in modo da far scendere il fluido in profondità all'interno del pozzo e di risalire dopo aver scambiato potenza termica con il terreno. La loro progettazione deve tener conto dell'influenza termica reciproca dei tratti, essi solitamente sono posti a distanza di circa 7-8 metri, dei materiali utilizzati per il rivestimento delle tubazioni, sia in termini di resistenza termica che di manutenzione richiesta, e delle caratteristiche del terreno. Tale soluzione è quella economicamente meno vantaggiosa in quanto i pozzi sono caratterizzati da profondità elevate, solitamente superiori ai 25 metri. Gli scambiatori orizzontali consistono invece in circuiti che corrono orizzontalmente all'interno di trincee, più o meno ampie e profonde in funzione della tipologia, ad una profondità minima di 1.5 metri e che differiscono fra di loro per la disposizione. In particolare, essi possono essere a tubi singoli, a tubi multipli, coppie di tubazioni di mandata e di ritorno che possono essere sovrapposte o affiancate, o a spirale, tubazioni a spirali contenute sempre all'interno di trincee verticali. A prescindere dalla disposizione, tale soluzione richiede una ampia estensione del terreno.

Nella figura sottostante, figure 4.1-2, vengono mostrate le soluzioni relative all'utilizzo di scambiatori termici disposti in modo verticale, a sinistra, e in modo orizzontale, a destra.



Figure 4.1-2 An example of vertical GCHP, on the left, orizontal GCHP, in the center, and spiral GCHP on the right.

2. Pompe di calore ad acque di superficie, Surface Water Heat Pumps o SWHPs: utilizzano come fonte termica le acque superficiali e possono essere corredate da sistemi ad anello aperti o chiusi. Nel primo caso lo scambio termico avviene attraverso tubazioni immerse nel corso d'acqua che provvedono sia alla fase di prelievo che a quella di scarico. Nel secondo caso invece, vengono impiegate delle tubazioni a bobina immerse nei corpi d'acqua e all'interno delle quali scorre un fluido termovettore. In generale l'accoppiamento con le acque superficiali presenta diverse problematiche. In primo luogo bisogna prevedere delle soluzioni resistenti, per via della natura corrosiva tipica di tali acque, e la disposizione di filtri. Inoltre, in questo caso le perdite di carico sono più elevate, i consumi energetici dovuti alle pompe di circolazione saranno maggiori, ed inoltre le condizioni ambientali sono maggiormente variabili rispetto alle altre soluzioni, limitando dunque la performance dell'impianto. In figure 4.1-3 viene mostrato un esempio di pompa di calore ad acqua di superficie con circuito aperto, a sinistra, e chiuso, a destra.

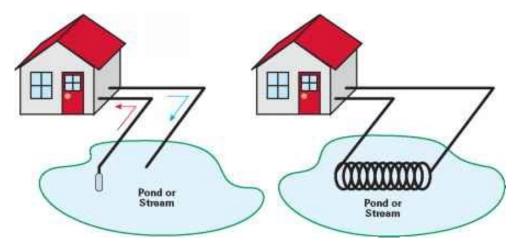


Figure 4.1-3: An example of surface water heat pump with an open loop, on the left, and a closed loop, on the right.

3. Pompe di calore ad acqua di falda, Ground Water Heat Pumps o GWHPs: corrispondono a sistemi a circuito aperto che utilizzano l'acqua di falda come mezzo di scambio termico diretto. La principale attrattiva di tale soluzione consiste nel fatto che sono necessari al massimo due pozzi, uno di presa ed uno di restituzione, anche se vi sono soluzioni che prevedono la restituzione in superficie, laghetti o corsi d'acqua, e altre chiamate standing column, la cui traduzione letteraria sarebbe colonna in piedi, che prevedono la re-immissione dell'acqua di falda all'interno del pozzo utilizzato per l'estrazione. Lo spazio occupato quindi è solitamente limitato rispetto alle precedenti soluzioni e una sola coppia di pozzi può ottemperare alle esigenze anche di grandi edifici; inoltre tali soluzioni hanno un costo inferiore e quando è possibile si preferisce utilizzare tale tecnologia rispetto alle precedenti [55]. I problemi principali di tale tecnologia riguardano le quantità di portate estratte, che devono essere compatibili con i tempi di ricarica della falda, la qualità dell'acqua e soprattutto la percentuale di ferro che potrebbe portare ad una corrosione o ad uno sporcamento delle tubazioni nel momento in cui non si utilizza acqua depurata. Tale inconveniente può essere tuttavia risolto dotando gli impianti di uno scambiatore disposto in posizione centrale tra la sorgente e il circuito lato pompa di calore[57]. Un ulteriore problema è invece di natura normativa in quanto non sempre è possibile utilizzare le acque di falda ed inoltre bisogna porre particolare attenzione alla re-immissione delle stesse, soprattutto per quanto concerne le temperature, in modo da non modificare la natura chimica, fisica e biologica delle falde. Occorre inoltre aggiungere che vi è una forte carenza legislativa su quest'ultimo aspetto infatti, solitamente, ogni caso viene trattato in modo autonomo previo studio e progetto preliminare. La figura sottostante, figure 4.1-4, mostra un esempio di pompe di calore ad acqua di falda a due pozzi, a destra, e con re-immissione in acque superficiali.

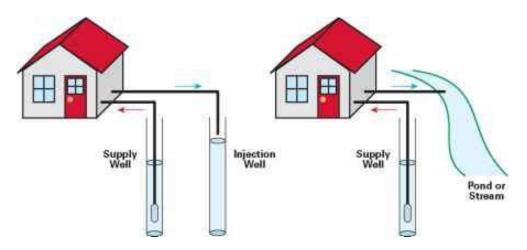


Figure 4.1-4: An example of ground water open loop heat pump with two wells, on the left, and with single well, on the right.

Fra le sorgenti termiche sopra brevemente descritte, la configurazione dell'impianto termico analizzato coincide con la terza, ovvero una pompa di calore ad acqua di falda e dotata di due pozzi, uno di estrazione ed uno di re-immissione. A questo punto occorre fornire delle informazioni circa il funzionamento di una pompa di calore e dei componenti di cui si caratterizza.

Le pompe di calore sono delle macchine termiche che devono il loro nome al fatto di essere in grado di trasferire calore da un livello termico più basso ad un livello termico più alto, così come una pompa solleva un fluido da una quota inferiore ad una superiore. Il ciclo termodinamico che esse utilizzano, intendendo con tale termine una serie di trasformazioni termodinamiche in cui si ha il passaggio da uno stato iniziale ad uno finale coincidente con il primo, è quello frigorifero, ovvero inverso a quello che avverrebbe in natura in maniera spontanea, in quanto altrimenti si violerebbe il secondo principio della termodinamica, ed è possibile solo fornendo energia al sistema e quindi compiendo lavoro.

Una rappresentazione schematica riferita ad entrambi i regimi, estivo ed invernale, viene fornita nell'immagine sottostante, figure 4.1-5.

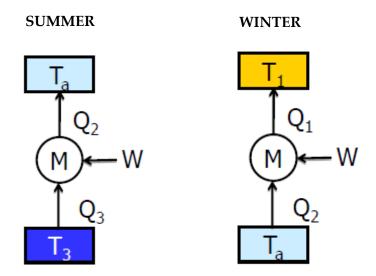


Figure 4.1-5: Heating and cooling production with an heat pump. In particular  $T_a$  is the external temperature,  $T_1$  is the hot source temperature,  $T_3$  is the cold source temperature, M is the heat pump and M is the work that the heat pump needs.

Il loro principio di funzionamento coincide dunque con quello delle macchine frigorifere seppur con scopo diverso: se le macchine frigorifere, e quindi i chiller nel caso di climatizzazione ambientale, hanno il solo ruolo di raffrescare l'ambiente interno e quindi sottrarre calore, le pompe di calore hanno il compito di riscaldare l'ambiente interno e quindi fornire calore. Occorre tuttavia aggiungere che le pompe di calore sono nate come macchine frigorifere per il raffrescamento dotate di una valvola che permetteva, attraverso l'inversione del ciclo, di riscaldare anziché raffrescare. Tuttavia, con il passare degli anni ci si rese conto che i livelli termici a cui i componenti lavorano nei diversi regimi, invernale ed estivo, sono molto differenti fra loro e pertanto sono stati introdotti degli appositi componenti ed accessori per assicurare un elevata efficienza operativa [55].

Operando in solo regime estivo, la CF1 non corrisponde ad una pompa di calore ma bensì ad un impianto frigorifero: essa infatti si compone di 3 chiller acqua-acqua ovvero:

- 1. Climaveneta TECS2-W/LC 0912;
- 2. M.T.A. AQTC 220;
- 3. RC FRIGO TURBO FL 560 T2 P4;

Ai fini di una maggiore comprensione verranno prima fornite indicazioni generiche circa il funzionamento di un chiller e dei componenti che lo riguardano, e solo successivamente verranno descritte le caratteristiche dei chiller oggetto di studio del presente lavoro di tesi.

Se consideriamo lo schema impiantistico relativo ad un chiller mostrato nell'immagine sottostante, figure 4.1-6 possiamo individuare le seguenti fasi caratteristiche del regime estivo.



Figure 4.1-6: Chiller operating cycle [58]

All'interno del compressore, A, il fluido allo stato gassoso e a bassa pressione, proveniente dall'evaporatore, D, viene compresso in modo da aumentarne pressione e temperatura. A queste condizioni il gas arriva al condensatore, B, dove viene raffreddato, ritornando allo stato liquido e cedendo parte del calore all'ambiente esterno, nel nostro caso tale cessione avviene attraverso i pozzi di emungimento e smaltimento dell'acqua di falda. Segue il passaggio del fluido attraverso una valvola di laminazione, C, che riduce repentinamente la pressione, provocando una evaporazione di parte del gas e al contempo un abbassamento di temperatura. La miscela così ottenuta passando attraverso l'evaporatore, D, assorbe calore dall'ambiente da climatizzare: il fluido a questo punto è ritornato alle stesse condizioni di partenza ed ha quindi completato il ciclo termodinamico.

I componenti caratteristici di un chiller sono dunque 4, ovvero:

• 2 scambiatori di calore, evaporatore e condensatore, ovvero elementi all'interno del quale si realizza uno scambio termico fra il fluido refrigerante interno, in

cambiamento di fase, e il fluido esterno a cui è accoppiato, acqua o aria [59]. Esistono diverse tipologie costruttive che differiscono fra loro in funzione del tipo di fluido intermedio utilizzato. Nel caso in cui il fluido sia aria la soluzione più utilizzata coincide con batterie alettate, mentre nel caso di acqua le soluzioni più comuni possono essere scambiatori a tubi concentrici, a piastra o a fascio tubiero. Quest'ultima soluzione coincide con quella utilizzata da tutti e tre i chiller contenuti all'interno della CF.

- Compressore, corrisponde all'elemento fondamentale del ciclo frigorifero in quanto attraverso esso viene fornito lavoro al sistema. Il suo scopo è quello di innalzare la pressione di un fluido: se consideriamo il ciclo termodinamico descritto precedentemente, figure 4.1-6, allora il suo scopo è quello di portare il fluido frigorigeno a bassa pressione e in uscita dall'evaporatore, alla pressione di condensazione. L'incremento di pressione può aversi tramite una compressione del fluido in un volume via via più piccolo, in questo caso il compressore sarà di tipo volumetrico, oppure tramite un processo che viene definito dinamico ovvero al fluido viene inizialmente fornita un'energia cinetica, attraverso un aumento della velocità, che si trasforma poi in incremento di pressione statica: in questo caso il compressore sarà di tipo centrifugo.
- Valvola di laminazione, ovvero un organo statico il cui scopo è quello di abbassare la temperatura del fluido attraverso una riduzione della pressione. In particolar modo, essendo un organo di strozzamento, degrada l'energia di pressione in attrito e permette di mantenere costante l'entalpia del fluido. Essa corrisponde dunque ad un regolatore di portata che si adatta al valore risultante dall'equilibrio degli altri organi caratteristici del circuito: compressore, evaporatore e condensatore [60].

Le caratteristiche dei tre chiller contenuti all'interno della CF1 vengono riassunte nella tabella sottostante.

	Climaveneta TECS2-	M.T.A AQTC	RC FRIGO TURBO
	W/LC 0912	220	FL 560 T2
Potenza termica [kW]	879	600	560
Potenza elettrica [kW]	172	136.6	110.9
СОР	5.12	4.39	5,05
Compressore	Centrifugo	Centrifugo	Centrifugo
Scambiatori	Fascio tubiero	Fascio tubiero	Fascio tubiero

Table 4.1-1: Summary table of the characteristics of the chillers in the CF1.

# 4.2. Descrizione del data-set

Prima di procedere con la descrizione del data-set utilizzato per delineare il comportamento operativo del caso studio occorre introdurre quali sono le variabili che entrano in gioco durante il suo funzionamento.

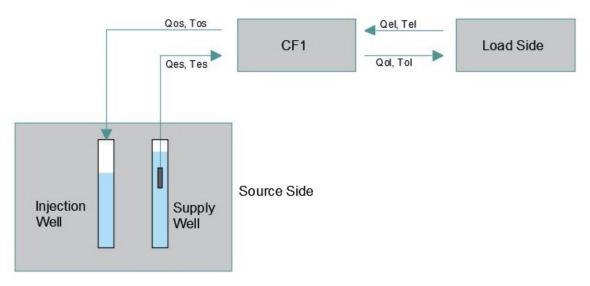


Figure 4.2-1: Simplified represensentation of cooling plant.

Se consideriamo lo schema semplificativo dell'impianto di condizionamento mostrato nell'immagine sovrastante, figure 4.2-1, possiamo individuare le seguenti variabili caratteristiche:

- Qes: Entering Source Flowrate i.e. portata in ingresso dal lato geotermico. Essa coincide con la portata di acqua emunta dalla falda ed in ingresso al condensatore;
- Tes: Entering Source Temperature i.e. temperatura in ingresso dal lato geotermico. Essa corrisponde alla temperatura dell'acqua di falda emunta e in ingresso al condensatore;
- Qos: Outlet Source Flowrate i.e. portata in uscita dal lato geotermico. Essa coincide dunque con la portata di acqua in uscita dal condensatore e re-immessa in falda.
- Tos: Outlet Source Temperature i.e. temperatura in uscita dal lato geotermico.
   Essa corrisponde alla temperatura dell'acqua in uscita dal condensatore che viene re-immessa in falda.
- Qel: Entering Load Flowrate i.e. portata di ritorno dall'ambiente ed in ingresso all'evaporatore;
- Tel: Entering Load Temperature, i.e. temperatura di ritorno dall'ambiente ed in ingresso all'evaporatore;
- Qol: Outlet Load Flowrate, portata di mandata all'ambiente da raffrescare ed in uscita dall'evaporatore;
- Tol: Outlet Load Temperature, temperatura di mandata all'ambiente da raffrescare ed in uscita dall'evaporatore.

A tali variabili si aggiungono inoltre la potenza elettrica assorbita dall'intero impianto, considerando anche quindi i valori necessari per l'emungimento dell'acqua di falda, e la potenza termica prodotta, calcolata come rapporto fra il calore ceduto all'ambiente nell'unità di tempo.

## 4.2.1. Raccolta dei dati

I dati necessari per la caratterizzazione dell'impianto sono stati acquisiti dal Living Lab del Politecnico di Torino e si riferiscono all'intera stagione di raffrescamento che va dal 15/06/2017 al 20/09/2017.

In particolar modo la CF1 è dotata di un sistema di monitoraggio che si compone di:

- Due contaimpulsi di cui uno rileva la quantità di volume di acqua emunta e reimmessa in falda, cumulata sull'intera stagione, e le rispettive temperature (rispettivamente Qes, Qos, Tes e Tos), e uno invece che rileva la sola portata di ritorno dall'ambiente in m³/h, Qel, e le temperature di mandata e ritorno dall'ambiente, rispettivamente Tol e Tel;
- un meter che rileva la potenza elettrica assorbita dall'impianto, in kW, con un timestep di 1 minuto.

Dopo aver elencato i dati forniti, occorre a questo punto effettuare una precisazione circa il funzionamento dei contaimpulsi. Il principio di funzionamento di tali sensori si basa sull'effetto Doppler che produce una variazione di frequenza di un segnale ad ultrasuoni nel momento in cui esso viene riflesso da una discontinuità all'interno di un fluido in movimento, come ad esempio la presenza di particelle solide o bolle di aria all'interno del liquido o la turbolenza del fluido stesso. Ogni qualvolta si ha una condizione del genere viene quindi registrato un impulso. In particolar modo la variazione di frequenza dipende dalla velocità del flusso ed il conteggio di tali impulsi si ottiene attraverso una conversione dell'energia cinetica, del fluido in movimento, in energia elettrica sotto forma di impulsi digitali binari che possono quindi assumere valore 0 o 1. I periodi corrispondenti ai valori 0 e 1 possono quindi essere misurati per determinare la velocità del flusso e la portata o il volume di acqua, nota la sezione. L'impostazione di tale sensore quindi non si basa su un timestep temporale ma sulla quota di energia immagazzinata. Nel momento in cui tale energia supera una determinata soglia, variabile in funzione del sistema considerato, allora l'impulso verrà registrato e di conseguenza anche la portata, o il volume, ad esso corrispondenti. Nel nostra caso le impostazioni del contaimpulsi fanno sì che il timestep coincida con 35 minuti.

Tuttavia, il contaimpulsi che rileva i volumi di acqua emunti e reimmessa in falda è sovradimensionato. Se infatti si osserva la figura sottostante relativa al volume di acqua
prelevata, cumulata sull'intera stagione considerata, figure 4.2-2, si nota un andamento a
gradini, e non lineare, non solo durante i giorni di inattività ma per tutto il periodo
considerato. Il timestep considerato per il calcolo delle portate di emungimento e reimmissione pertanto non sarà di 35 minuti ma bensì di 70 minuti.

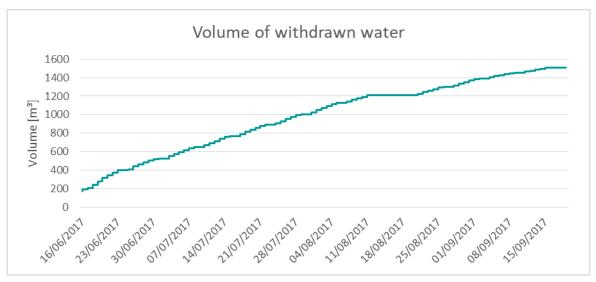


Figure 4.2-2: Development of withdrawn water's volume during cooling season.

La tabella riportata di seguito, table 4.2-1, mostra in maniera riassuntiva i dati ricevuti. Occorre tuttavia precisare che tutte le variabili verranno pre-processate in modo da conferire uniformità al data-set, così come indicato nel paragrafo successivo, ma solo alcune di esse verranno utilizzate per la caratterizzazione dell'impianto, ovvero: le portate e le rispettive temperature in ingresso alla CF1, sia dal lato condensatore che dal lato evaporatore e quindi Qes, Qel, Tes e Tel e la potenza elettrica e termica. Le restanti variabili, quindi Qos, Tos e Tol, verranno invece utilizzate successivamente per la fase di validazione del modello costruito tramite TRNSYS di cui si parlerà nel prossimo capitolo.

	Metodo di rilevamento	Unità di misura	Timestep
Qes	Ricavata	m³/s	70 minuti
Tes	Data logger	°C	35 minuti
Qos	Ricavata	m³/s	70 minuti
Tos	Data logger	°C	35 minuti

Qel	Contaimpulsi	m³/h	35 minuti
Tel	Data logger	°C	35 minuti
Tol	Data logger	°C	35 minuti
Potenza Elettrica	Meter	kW	1 minuto
Potenza Termica	Ricavata	kW	35 minuti

Table 4.2-1: Summary data table with indication of detection method, unit of measurement and timestep.

# 4.2.2. Pre – processamento dei dati

La fase successiva alla raccolta dei dati coincide con il pre-processamento degli stessi. Tale fase è di fondamentale importanza in un processo di KDD in quanto solitamente i dati reali, non trattati, possono contenere al loro interno delle irregolarità, e possono quindi essere [51]:

- Incompleti: presenza di valori nulli o mancanti;
- Fastidiosi: presenza di rumore o outlier;
- Incoerenti: presenza di discrepanze.

La presenza di tali caratteristiche all'interno del dataset di partenza può infatti portare ad una riduzione di qualità dell'intero processo di analisi. Pertanto, durante la fase di preprocessamento i dati dovranno essere studiati in modo da cogliere le differenze dimensionali e granulometriche, indentificare gli eventuali valori mancanti e i rumori presenti.

La prima fase di consolidamento dei dati è consistita nel far coincidere gli stessi dal punto di vista dimensionale e temporale. Dal punto di vista dimensionale tutti i dati riferiti alla stessa tipologia di variabili quindi saranno caratterizzati dalle stesse unità di misura, ovvero le portate saranno in m³/s, le temperature in °C e le potenze in kW.

Allo stesso modo la corrispondenza temporale si avrà nel momento in cui tutti i dati saranno caratterizzati dallo stesso timestep, condizione che al momento non è verificata in quanto i dati ricevuti sono caratterizzati da 3 passi temporali differenti ovvero 1, 35 e 70 minuti. Poiché la variabile principale del nostro dataset corrisponde alla potenza elettrica mediata nei 15 minuti, è stata necessaria una ricostruzione delle serie temporali, relative

alle restanti grandezze, tramite interpolazione lineare ai 15 minuti in modo che tutte le variabili considerate siano caratterizzate dalla stessa granulometria. Si ricorda inoltre che la scala temporale utilizzata per descrivere il comportamento dell'impianto è quella giornaliera e se si considera un timestep di 15 minuti allora ciascun giorno potrà essere considerato come un vettore avente 96 componenti.

La scelta della potenza elettrica come variabile principale è dovuta al fatto che essa è quella che meglio riesce a descrivere il comportamento dell'impianto: ci permette infatti non solo di cogliere gli orari di accensione e spegnimento ma anche le fasi di picco ed eventuali funzionamenti anomali o infrequenti.

Dopo aver reso il dataset uniforme si è poi proceduto a verificare l'eventuale presenza di dati mancati e valori anomali. Per quanto concerne la prima di queste due verifiche il primo e l'ultimo giorno della stagione, rispettivamente il 15/06/2017 ed il 20/09/2017, sono stati rimossi. In tali giorni i dati registrati, ad esclusione di quelli relativi alla potenza elettrica, non ricoprivano l'intero arco temporale della giornata e nella fattispecie il 15/06 i primi dati rilevati si hanno a partire dalle ore 4:00 mentre il 20/09 il rilevamento dei dati si è interrotto alle 11:00. Il periodo considerato pertanto andrà dal 16/06/2017 al 19/09/2017.

## 4.2.2.1. Individuazione dei profili anomali

Già nel capitolo 2 si è parlato dell'individuazione dei profili di carico e delle anomalie che possono occorrere durante il funzionamento di un impianto di climatizzazione. In linea generale si può affermare che nella maggior parte dei casi le procedure utilizzate sono le medesime. Nonostante il differente fine delle ricerche relative a tali ambiti esse si basano entrambe sullo studio dei profili aventi a disposizione e nel riconoscimento di pattern caratteristici o meno: solo attraverso un confronto fra i vari profili aventi a disposizione si può infatti determinare se il profilo relativo ad un determinato giorno è anomalo o comune.

Una delle tecniche più utilizzate in letteratura per l'individuazione di un profilo, a prescindere dal fatto che esso sia atipico o meno, corrisponde al clustering. Nel presente lavoro di tesi la tipologia di clustering che verrà utilizzata coincide con quello gerarchico:

il motivo di tale scelta risiede essenzialmente nel fatto che tale tipologia non richiede a priori la determinazione dei k cluster.

Pertanto, il processo di identificazione dei profili anomali si compone delle seguenti fasi:

- Clustering gerarchico con metodo di Ward sull'intero dataset della potenza elettrica in modo da avere una panoramica dei profili che caratterizzano il nostro dataset;
- 2. Clustering con metodo single linkage su ciascun cluster ottenuto dalla fase 1.

Inoltre, per tale fase, così come anche per la fase di individuazione dei profili caratteristici e di classificazione, che verranno descritte nel capitolo successivo, è stato utilizzato il software open-source R di statistica e grafica.

La prima fase ha visto dunque l'applicazione dell'algoritmo di clustering gerarchico all'intero dataset della potenza elettrica. Nella fattispecie i dati sono stati normalizzati sul valore massimo giornaliero in modo da cogliere differenze di forma e non di intensità. In particolare, ricordando che ciascun i-esimo giorno considerato può essere considerato come un vettore avente 96 componenti, siano  $Pel_i$  il singolo componente della potenza elettrica dell'i-esimo giorno e  $Pel\_max_i$  il componente della potenza elettrica avente maggiore intensità per il giorno i-esimo, allora il valore normalizzato di potenza elettrica per il singolo componente sarà:  $Pel\_norm_i = \frac{Pel_i}{Pel\_max_i}$ . Inoltre, la misura di similarità considerata coincide con la distanza euclidea ed il metodo utilizzato è quello di Ward.

Nella figura sottostante, figure 4.2 -3, viene mostrato a sinistra il dendogramma risultante e a destra l'elbow method indicante il numero ottimale di cluster per tale variabile. In particolar esso suggerisce di tagliare il dendogramma in modo da ottenere 6 cluster. Inoltre, sempre all'interno di tale dendogramma si può notare che il gruppo a sinistra è caratterizzato, ad eccezione del dato pari al giorno 10, da una distanza pari a 0: ciò significa che non solo i record appartengono allo stesso cluster ma sono anche identici fra loro. Come si può notare dalla figure 4.2 -4 tale gruppo corrisponde dunque al cluster 5.

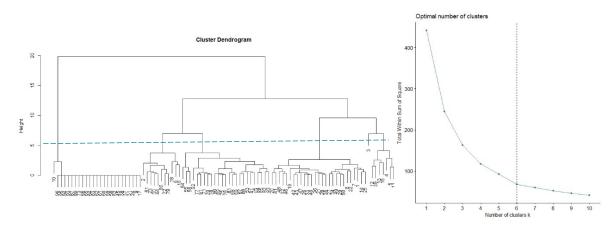


Figure 4.2-4: Cluster dendogram obtained with Ward method for normalized electrical power, on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster to which such dendogram has been cut.

Tagliando per k = 6 i cluster ottenuti vengono mostrati nell'immagine sottostante: a sinistra quelli relativi ai dati normalizzati sul valore massimo giornaliero, mentre a destra quelli ottenuti per i cluster non normalizzati, figure 4.2-4.

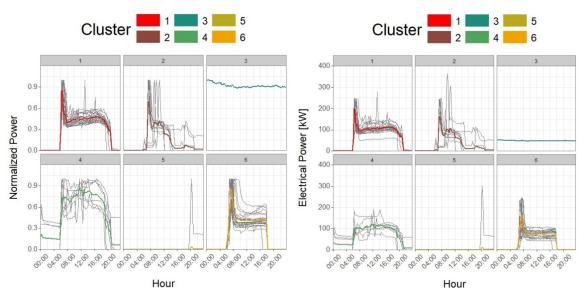


Figure 4.2-3: Clusters achieved by cutting precedent dendogram for k=6, with normalized values, on the left, and not normalized values, on the right. Clusters 3 and 4 are classified as outlier.

# In essa è possibile notare che:

• i cluster 1 e 6 presentano delle caratteristiche comuni, ovvero una fase accensione caratterizzata da un picco massimo, una fase di funzionamento a regime, caratterizzata da un andamento pressoché costante, ed una fase di spegnimento. Tuttavia, all'interno del cluster 1 si nota del rumore;

- i cluster 3 e 4, al contrario, sono caratterizzati da profili anomali in quanto presentano non solo un funzionamento diurno ma anche notturno. I profili appartenenti al cluster 4 inoltre non presentano una fase di picco ma sono caratterizzati da una intensità constante durante la fase di attività;
- i cluster 2 e 5 presentano del rumore al loro interno, specialmente all'interno del cluster 2, in quanto non solo si nota un funzionamento notturno ma anche un'accensione posticipata. Al fine di non incorrere nel fenomeno di overfitting tali cluster verranno ulteriormente analizzati in modo da rimuovere solo i profili effettivamente infrequenti.

A valle di ciò i cluster 3 e 4 vengono classificati come outlier e separati dal dataset iniziale per una successiva analisi. I cluster 1, 2 e 5 vengono ulteriormente analizzati in modo da eliminare gli eventuali profili anomali che potrebbero compromettere la qualità delle analisi successive.

#### Cluster 1

Per il cluster 1 di figure 4.2-4 il metodo considerato consiste nel single linkage. La scelta di tale metodo consiste nel fatto che, come già detto nel capitolo precedente, esso predilige le differenze inter-cluster e pertanto riesce a cogliere maggiormente gli eventuali profili anomali. L'immagine sottostante, figure 4.2-5 mostra a sinistra il dendogramma ottenuto dall'applicazione di tale metodo considerando sempre la distanza euclidea ed i valori normalizzati su valore massimo giornaliero, e a destra l'elbow method. In questo caso l'elbow method è di più difficile interpretazione in quanto non è visibile un vero e proprio gomito all'interno della curva, tuttavia, si può notare che per k>4 lo scarto quadratico decresce più lentamente all'aumentare di k pertanto il numero di cluster scelti coinciderà con k=4.

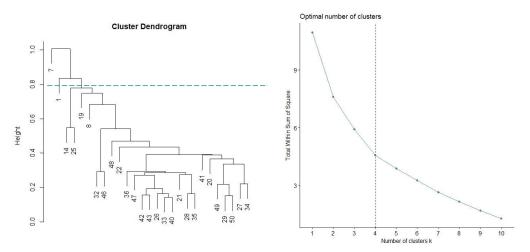


Figure 4.2-5: Cluster dendogram obtained with single linkage for normalized electrical power on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster.4 in this case, to which such dendogram has been cut.

Tagliando per k = 4, i cluster ottenuti vengono mostrati nell'immagine sottostante, figure 4.2 -6: anche in questo caso a sinistra vengono riportati quelli relativi ai dati normalizzati su massimo giornaliero, mentre a destra quelli ottenuti considerando i dati non normalizzati. Fra questi solo il cluster 2 presenta un andamento infrequente in quanto si nota la presenza di un secondo picco durante il funzionamento a regime e pertanto verrà classificato come outlier. L'effettivo andamento infrequente verrà confermato anche dalle altre variabili e specialmente dalla portata di ritorno dall'ambiente.

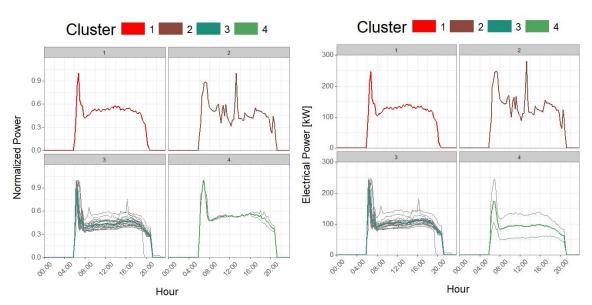


Figure 4.2-6: Clusters achieved by cutting precedent dendogram for k=4 with normalized values, on the left, and not normalized values, on the right. Cluster 2 is classified as outlier.

# Cluster 2

Così come per il cluster 1, anche per il cluster 2 di figure 4.2-4 il metodo utilizzato per l'ulteriore fase di analisi coincide con il single linkage. In particolar modo in figure 4.2-7 è possibile vedere il dendogramma ottenuto, a sinistra, e l'elbow method, a destra. Il dendogramma è stato tagliato per k= 6, così come suggerito dall'elbow method.

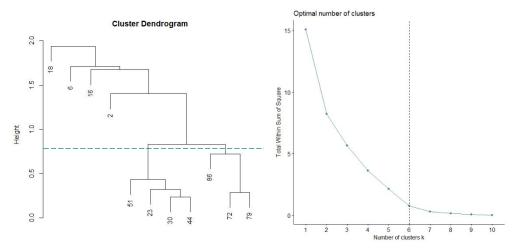


Figure 4.2-7: Cluster dendogram obtained with single linkage for normalized electrical power on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster,6 in this case, to which such dendogram has been cut.

I cluster ottenuti vengono mostrati nell'immagine sottostante, figure 4.2-8. Fra questi i primi 4 vengono classificati come outlier in quanto caratterizzati da un funzionamento notturno, cluster 1, o da un picco, solitamente coincidente con la fase di accensione, posticipato.

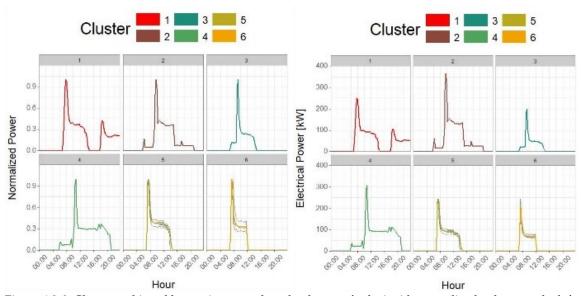


Figure 4.2-8: Clusters achieved by cutting precedent dendogram for k=6 with normalized values, on the left, and not normalized values, on the right. Cluster 1, 2, 3, 4 are classified as outliers.

# Cluster 5

L'immagine sottostante, figure 4.2-9, mostra a sinistra il dendogramma ottenuto applicando il single linkage al cluster 5 di figure 4.2- 4 e a destra l'elbow method.

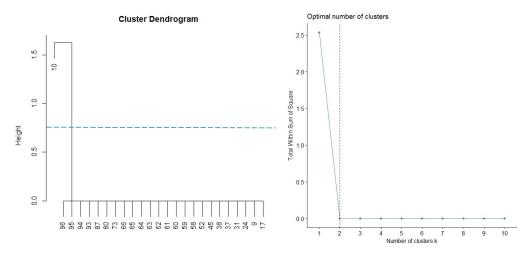


Figure 4.2-9: Cluster dendogram obtained with single linkage for normalized electrical power, on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster, to which such dendogram has been cut.

Tagliando il dendogramma per k=2, così come indicato dall'elbow method, i cluster risultanti sono quelli mostrati nell'immagine seguente, figure 4.2 -10. Fra questi il secondo viene classificato come outlier, in quanto mostra un funzionamento notturno.

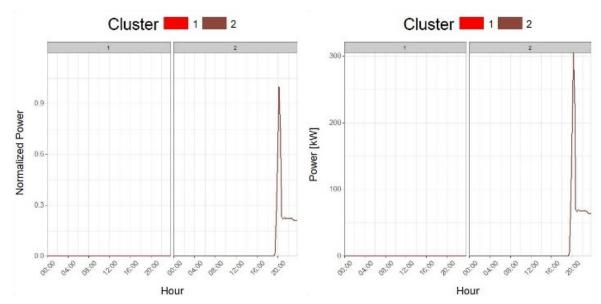


Figure 4.2-10: Clusters achieved by cutting precedent dendogram for k=2, with normalized values, on the left, and not normalized values, on the right. Cluster 2 is classified as outlier.

I profili anomali sino ad ora individuati sono stati analizzati in modo da individuare dei cluster caratteristici e permettere un più semplice confronto con le altre variabili caratteristiche del comportamento dell'impianto in modo da verificare l'effettiva anomalia dei profili individuati, ovvero: potenza termica, portata di emungimento e di ritorno dall'ambiente e le relative temperature.

Nell'immagine sottostante, figure 4.2-11, viene mostrato a sinistra il dendogramma ottenuto considerando il metodo di Ward, la distanza euclidea ed i valori normalizzati per la potenza elettrica, e a destra l'elbow method.

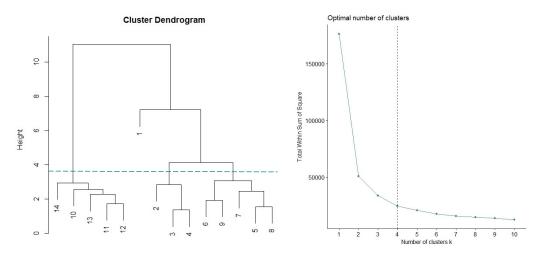


Figure 4.2-11: Cluster dendogram obtained with Ward method for normalized electrical power, on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster, to which such dendogram has been cut.

Tagliando tale dendogramma per k=4, così come indicato dall'elbow method, i cluster dei profili anomali ottenuti vengono mostrati nell'immagine sottostante, figure 4.2-12.

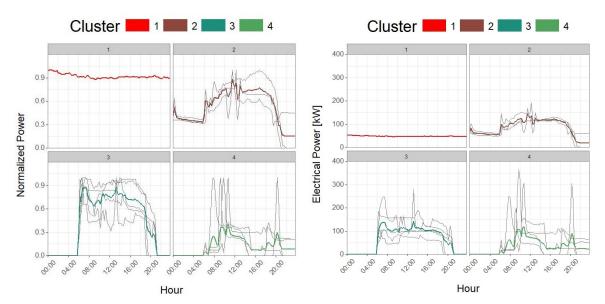


Figure 4.2-12: Clusters achieved for outlier by cutting precedent dendogram for k=4, with normalized values, on the left, and not normalized values, on the right.

### Potenza termica

Considerando il dataset iniziale, comprensivo degli outlier individuati in funzione della potenza elettrica, il dendogramma ottenuto considerando il metodo di Ward e la distanza euclidea viene riportato sulla sinistra nell'immagine sottostante, figure 4.2-13.

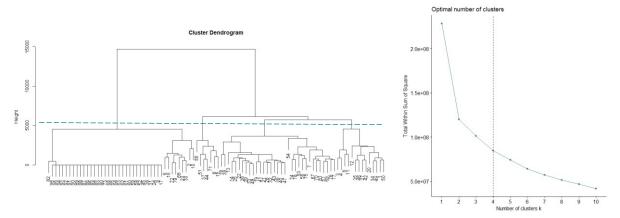


Figure 4.2-13: Cluster dendogram obtained with Ward method for thermal power, on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster, to which such dendogram has been cut.

Tagliando tale dendogramma per k= 4, così come indicato dall'elbow method mostrato a destra in figure 4.2-13 i cluster ottenuti sono riportati in figure 4.2-14. Anche all'interno di tali cluster si nota la presenza di rumore: all'interno dei cluster 1 e 2 vi sono dei profili aventi un funzionamento notturno.

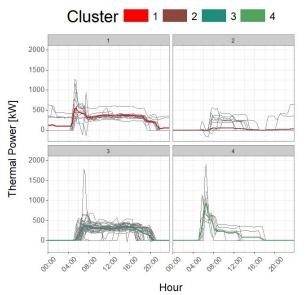


Figure 4.2-14: Clusters achieved for thermal power by cutting precedent dendogram for k=4.

Allo stesso modo se consideriamo i cluster definitivi degli outlier, mostrati in figure 4.2-12, e li confrontiamo con l'immagine sottostante, figure 4.2-15, riportante i profili caratteristici per gli stessi giorni, suddivisi nei medesimi cluster, possiamo notare delle congruenze. In particolar modo i cluster 1, 2 e 4 contengono al loro interno dei profili aventi un funzionamento notturno. Allo stesso modo il cluster 3 si caratterizza dall'assenza di un vero e proprio picco e da una maggiore variabilità dei profili. All'interno di quest'ultimo cluster si nota inoltre che il centroide si riduce e si mantiene costante per circa un'ora intono alle 12:00: ciò è dovuto alla presenza di un profilo in cui la potenza termica si è interrotta. Tale giorno coincide con l'outlier rimosso dal cluster 1.

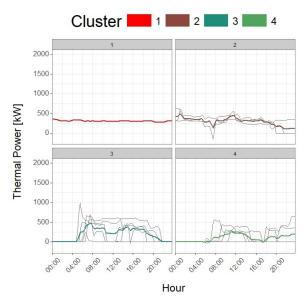


Figure 4.2-15: Final outlier obtained for thermal power considering the same cluster achieved for electrical power.

# Portata di emungimento, Qes

Considerando il dataset iniziale, comprensivo degli outlier individuati in funzione della potenza elettrica, il dendogramma ottenuto considerando il metodo di Ward e la distanza euclidea viene riportato sulla sinistra nell'immagine sottostante, figure 4.2-16. Sulla destra di figure 4.2-16 viene invece riportato l'elbow method: in questo caso il punto di gomito non è molto accentuato, tuttavia si può notare che per k>4 lo scarto quadratico decresce più lentamente all'aumentare di k. Allo stesso modo se consideriamo il dendogramma notiamo

la presenza di 4 gruppi ben separati fra di loro, pertanto il numero di cluster scelti coinciderà con k=4.

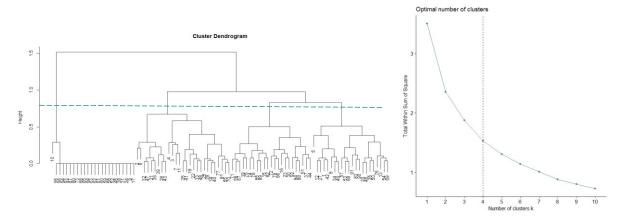


Figure 4.2-16: Cluster dendogram obtained with Ward method for entering source flowrate, on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster, to which such dendogram has been cut.

I cluster ottenuti vengono mostrati nell'immagine sottostante, figure 4.2-17. Anche in questo caso notiamo la presenza di profili anomali caratterizzati da un funzionamento notturno nei cluster 1, 2 e 3.

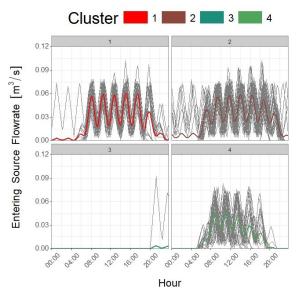


Figure 4.2-17: Clusters achieved for entering source flowrate by cutting precedent dendogram for k=4.

Allo stesso modo se consideriamo i cluster definitivi degli outlier individuati per la potenza elettrica, mostrati in figure 4.2-12, e li confrontiamo con l'immagine sottostante, figure 4.2-18, riportante i profili caratteristici per gli stessi giorni, suddivisi nei medesimi cluster, possiamo notare delle caratteristiche comuni. I cluster 1, 2 e 4 contengono al loro interno

dei profili aventi un funzionamento giornaliero continuo. Questi profili sono tuttavia presenti anche nel cluster 3, nonostante tale condizione operativa non fosse visibile all'interno del medesimo cluster relativo alla potenza elettrica.

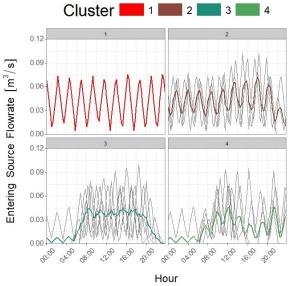


Figure 4.2-17: Final outlier obtained for entering source flowrate considering the same cluster achieved for electrical power.

# Portata di ritorno dall'ambiente, Qel

Considerando il dataset iniziale, comprensivo degli outlier individuati in funzione della potenza elettrica, il dendogramma ottenuto considerando il metodo di Ward e la distanza euclidea viene riportato sulla sinistra nell'immagine sottostante, figure 4.2-18.

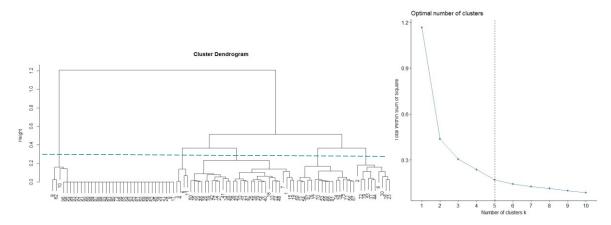


Figure 4.2-18: Cluster dendogram obtained with Ward method for entering load flowrate, on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster, to which such dendogram has been cut.

Tagliando per k=5, così come indicato dall'elbow method mostrato a destra in figure 4.2-18, i cluster ottenuti sono visibili nell'immagine riportata in basso a sinistra, figure 4.2-19. Se osserviamo tale immagine possiamo affermare che il cluster 5 corrisponde al normale funzionamento dell'impianto in quanto al suo interno si possono notare le fasi di accensione, di spegnimento e di funzionamento a regime costante. Al contrario, i restanti cluster sono caratterizzati dalla presenza di profili anomali caratterizzati da un funzionamento notturno, cluster 2, 3 e 4, o da una interruzione durante la fase di funzionamento a regime, cluster 1. Tali profili anomali, riportati a destra in figure 4.2-19, corrispondono con quelli individuati per la potenza elettrica, figure 4.2-12, e sono inoltre suddivisi nei medesimi cluster.

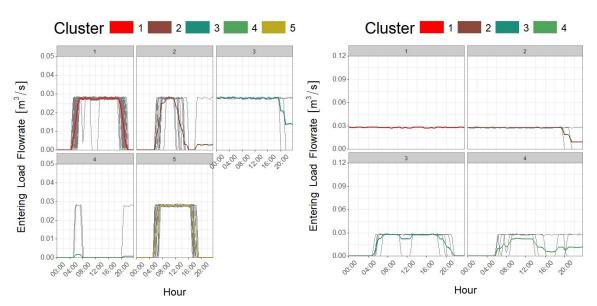


Figure 4.2-19: Clusters achieved for entering load flowrate by cutting precedent dendogram for k=5, on the left, and final outlier obtained for entering source flowrate considering the same cluster achieved for electrical power, on the right.

### Temperatura di emungimento, Tes

Considerando il dataset iniziale, comprensivo degli outlier individuati in funzione della potenza elettrica, il dendogramma ottenuto considerando il metodo di Ward e la distanza euclidea viene riportato sulla sinistra nell'immagine sottostante, figure 4.2-20. Sulla destra di figure 4.2-20 viene invece riportato l'elbow method che, in questo caso, suggerisce di tagliare il dendogramma per k=4.

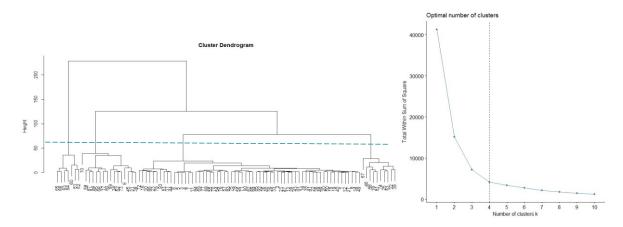


Figure 4.2-20: Cluster dendogram obtained with Ward method for entering source temperature, on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster, to which such dendogram has been cut.

L'immagine sottostante, figure 4.2-21 mostra i cluster ottenuti tagliando il precedente dendogramma per k=4. In particolar modo, come si vedrà in maniera più approfondita nel capitolo successivo, solitamente le temperature di ingresso alla CF1 da ambo i lati, geotermico e ambiente, sono caratterizzate da temperature più alte ed in aumento durante la fase di non funzionamento a cui segue un rapido abbassamento ed un funzionamento a temperatura pressoché costante durante la fase di attività. In virtù di ciò possiamo dunque affermare che i cluster 1 e 3 corrispondono ai giorni lavorativi mentre i cluster 2 e 4 corrispondo ai giorni festivi.

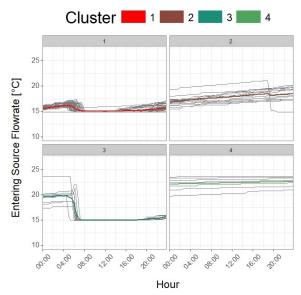


Figure 4.2-21: Clusters achieved for entering source temperature by cutting precedent dendogram for k=4.

Tuttavia, anche in questo caso possiamo notare la presenza di profili anomali specialmente nei cluster 1, 2 e 3. Tali profili, visibili in figure 4.2-22, coincidono con quelli individuati per

la potenza elettrica, figure 4.2-12, e sono inoltre suddivisi nei medesimi cluster. Se si effettua un confronto fra tali cluster e quelli ottenuti per la potenza elettrica si nota una congruenza circa le modalità operative dell'impianto.

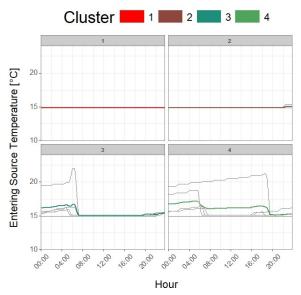


Figure 4.2-22: Final outlier obtained for entering source temperature considering the same clusters achieved for electrical power.

# Temperatura di ritorno dall'ambiente

Considerando il dataset iniziale, comprensivo degli outlier individuati in funzione della potenza elettrica, il dendogramma ottenuto considerando il metodo di Ward e la distanza euclidea viene riportato sulla sinistra nell'immagine sottostante, figure 4.2-23. Sulla destra, della medesima immagine, viene invece riportato l'elbow method che, in questo caso, suggerisce di tagliare il dendogramma per k=4.

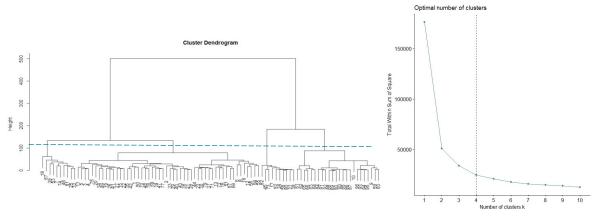


Figure 4.2-23: Cluster dendogram obtained with Ward method for entering load temperature, on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster, to which such dendogram has been cut

Tagliando il dendogramma per k=4, i cluster ottenuti coincidono con quelli mostrati a sinistra in figure 4.2-24. Così come per le altre variabili, anche in questo caso si può notare la presenza di rumore in ciascun cluster. Tali profili anomali, riportati nell'immagine a destra di figure 4.2-24, corrispondono a quelli individuati per la potenza elettrica e sono suddividisi nei medesimi cluster. Fra questi il cluster 1 e 2 si contraddistinguono per un andamento pressochè costante durante l'intera giornata, mentre i cluster 3 e 4 sono caratterizzati da profili maggiormente variabili soprattutto in corrispondenza degli orari di accensione.

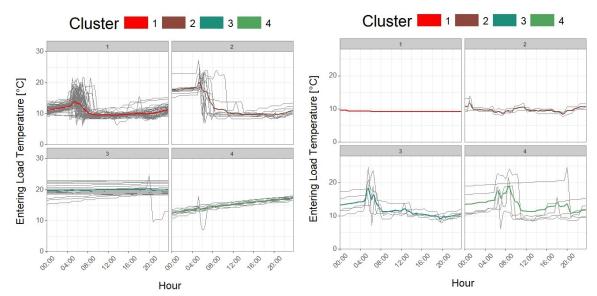


Figure 4.2-24: Clusters achieved for entering load temperature by cutting precedent dendogram for k=4, on the left, and final outlier obtained for entering load temperature considering the same clusters achieved for electrical power, on the right.

In definitiva il numero totale di giorni classificati come outlier e rimossi dal dataframe iniziale corrisponde a 14 a cui si aggiungono i due giorni rimossi poiché contenenti valori mancanti: il dataset pre-processato quindi si comporrà di 82 giorni. Nel prossimo capitolo verranno discusse le fasi di caratterizzazione dell'impianto, quindi individuazione e classificazione dei profili caratteristici di ciascuna variabile, e di costruzione del modello rappresentativo della CF1.

### 5. Risultati

Il presente capitolo descrive in maniera approfondita l'iter processuale svolto e i risultati ottenuti. Esso si articola delle seguenti sezioni: individuazione del profilo di carico tipico, paragrafo 5.1, estrazione delle regole attraverso un processo di classificazione, paragrafo 5.2, costruzione di un modello che rispecchi il comportamento dell'impianto oggetto del caso studio, paragrafo 5.3.

# 5.1. Individuazione del profilo di carico tipico giornaliero

Nel capitolo 2 e nella fattispecie nel paragrafo 2.1 si è già introdotto il tema del load profiling, dei benefici che esso comporta e delle tecniche maggiormente utilizzate. Così come per l'individuazione degli outlier, anche in questo caso la strategia adottata consiste nell'applicazione delle tecniche di clustering gerarchico alle variabili considerate per la caratterizzazione della CF1. Il processo di individuazione del profilo di carico tipico giornaliero e di classificazione è stato effettuato attravero l'utilizzo del software R e lo script utilizzato viene riportato in Allegato I.

Ricapitolando l'algoritmo del clustering gerarchico prevede i seguenti step:

- 1. Creazione della matrice dei dati: nel nostro caso poiché la scala temporale considerata è quella giornaliera con timestep di 15 minuti le sue dimensioni saranno 82x96, dove 82 è il numero di giorni restanti dopo la fase di preprocessamento;
- 2. Calcolo della matrice di prossimità considerando la distanza euclidea;
- 3. Inizialmente ogni punto, nel nostro caso ogni giorno, viene considerato come un singolo cluster;
- 4. Repeat;

- 5. Unione dei cluster più vicini in funzione dello scarto quadratico medio essendo il metodo considerato quello di Ward;
- 6. Ricalcolo della matrice di prossimità;
- 7. *Until* fino a che non rimane un singolo cluster.

Di seguito vengono forniti i risultati ottenuti dalle analisi effettuate. Si ricorda inoltre che i giorni caratterizzati da funzionamenti anomali sono stati rimossi dal dataset.

### Potenza elettrica

Come già detto in precedenza la variabile principale del dataset coincide con la potenza elettrica e pertanto le prime analisi sono state condotte su di essa. Anche in questo caso il clustering gerarchico è stato effettuato su valori normalizzati su massimo giornaliero. Nell'immagine sottostante, figure 5.1-1, viene riportato a sinistra il dendogramma ottenuto considerando il metodo di Ward e a destra l'elbow method indicante il numero ottimale di cluster per tale variabile. In particolar esso suggerisce di tagliare il dendogramma in modo da ottenere 3 cluster: tale indicazione viene anche confermata dal dendogramma all'interno del quale si notano 3 gruppi ben distinti fra loro.

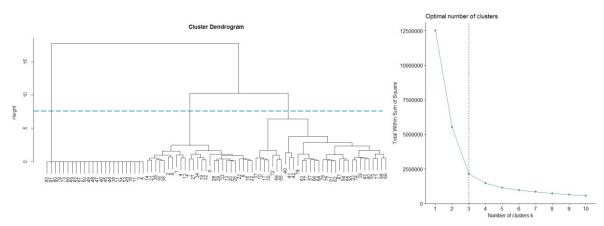


Figure 5.1-1: Cluster dendogram obtained with Ward method for normalized electrical power, on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster to which such dendogram has been cut.

Tagliando per k = 3 i cluster ottenuti vengono mostrati nell'immagine sottostante: a sinistra quelli relativi ai dati normalizzati su massimo giornaliero, mentre a destra quelli ottenuti per i valori non normalizzati, figure 5.1-2.

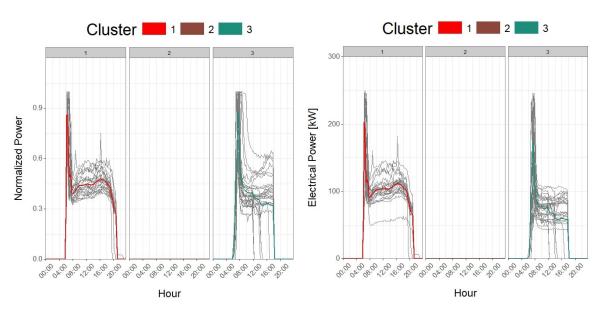


Figure 5.1-2: Clusters of electrical power achieved by cutting precedent dendogram for k=3, with normalized values, on the left, and not normalized values, on the right.

Se osserviamo i cluster sopra riportati possiamo notare che al cluster 2, essendo la potenza elettrica pari a 0 kW durante tutto l'arco della giornata, appartengono i giorni in cui l'impianto non è stato attivo. Al contrario invece, i cluster 1 e 3 corrispondono ai giorni in cui l'impianto è stato attivo ed in particolar modo, prima del processo di classificazione, possiamo affermare che ciò che li differenzia consiste negli orari di funzionamento ed in particolar modo di spegnimento: se i giorni del cluster 1 sono caratterizzati da un funzionamento continuo sino alle 20:00 circa, i giorni del cluster hanno degli orari variabili fra le 12:00 e le 18:00 circa.

# Potenza termica

Il dendogramma ottenuto dall'applicazione del clustering gerarchico con metodo di Ward viene mostrato nell'immagine sottostante affiancato dall'elbow method, figure 5.1-3. Quest'ultimo suggerisce un numero di k cluster ottimali compresi fra 2 e 4. Tuttavia, se

osserviamo il dendogramma notiamo la presenza di 3 macro-gruppi pertanto il numero k di cluster scelti sarà uguale a 3.

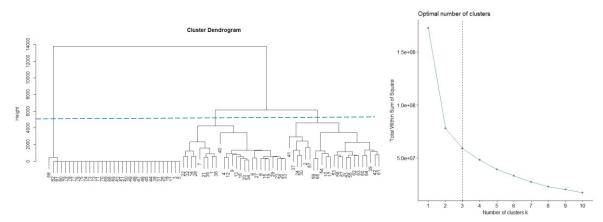


Figure 5.1-3: Cluster dendogram obtained with Ward method for thermal power, on the left, and elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster to which such dendogram has been cut.

I cluster risultanti vengono mostrati nell'immagine sottostante, figure 5.1-4. Così come per la potenza elettrica, due dei tre cluster ottenuti, 1 e 2, corrispondono ai giorni in cui l'impianto è stato attivo e differiscono fra di loro per gli orari di spegnimento. Al restante cluster invece appartengono i giorni di non funzionamento ad eccezione di uno in cui l'impianto è stato attivo per circa 4 ore ovvero dalle 4:00 alle 8:00: la presenza di un giorno leggermente dissimile dai restanti era inoltre già visibile dal dendogramma mostrato precedentemente.

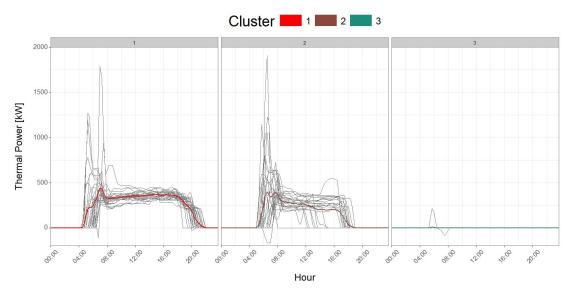


Figure 5.1-4: Clusters of thermal power achieved by cutting precedent dendogram for k=3.

# Portata di emungimento, Qes

L'immagine sottostante, figure 5.1-5, mostra il dendogramma ottenuto dall'applicazione del metodo di Ward ai valori della portata di emungimento ed in ingresso alla CF1 dal condensatore, sulla sinistra, e l'elbow method, sulla destra. Quest'ultimo suggerisce k compreso fra 2 e 4. Tuttavia, il dendogramma mostra la presenza di 3 cluster principali pertanto il numero di cluster scelti coincide con k=3.

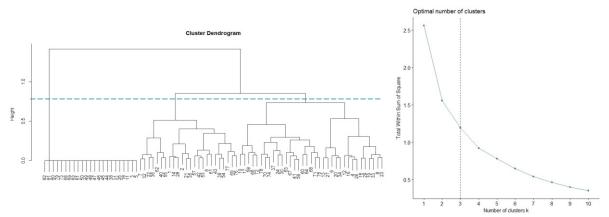


Figure 5.1-5: Cluster dendogram obtained with Ward method for entering source flowrate, on the left, and elbow method with indication of optimal number of clusters on the right. The dash line represents the number of cluster to which such dendogram has been cut.

I cluster ottenuti tagliando per k=3 vengono mostrati nell'immagine sottostante, figure 5.1-6. Anche in questo caso due fra i tre cluster corrispondono ai giorni di attività dell'impianto e ciò che li distingue è uno leggero sfasamento temporale.

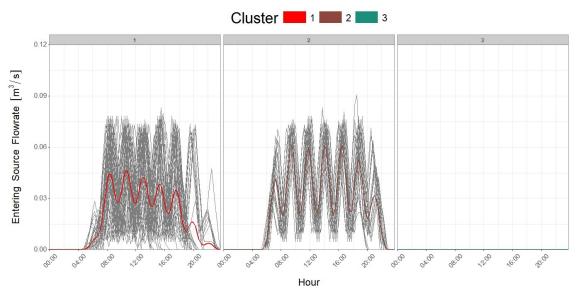


Figure 5.1-6: Clusters of entering load flowrate achieved by cutting precedent dendogram for k=3.

### Portata di ritorno dall'ambiente, Qel

L'immagine sottostante, figure 5.1-7, mostra il dendogramma ottenuto dall'applicazione del metodo di Ward ai valori della portata di ritorno dall'ambiente, affiancato dall'elbow method, all'interno del quale il gomito si ha per k compreso fra 2 e 5. Tuttavia, all'interno del dendogramma si nota la presenza di 3 cluster principali pertanto il numero di k scelti coinciderà con 3.

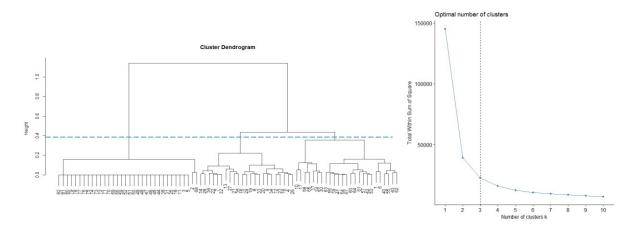


Figure 5.1-6: Cluster dendogram obtained with Ward method for entering load flowrate, on the left, and the elbow method with indication of optimal number of clusters, on the right. The dash line represents the number of cluster to which such dendogram has been cut.

L'immagine sottostante, figure 5.1-8, mostra i cluster ottenuti tagliando il dendogramma per k=3.

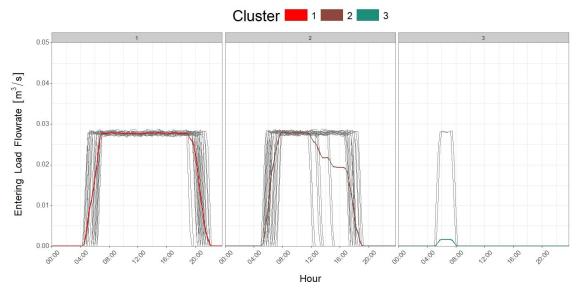


Figure 5.1-7: Clusters of entering load flowrate achieved by cutting precedent dendogram for k=3.

In particolar modo possiamo notare che, anche in questo caso, due cluster,1 e 3, si riferiscono ai giorni di attività dell'impianto seppur con orari di accensione e spegnimento differenti, così come i cluster ottenuti in precedenza per le altre variabili. Se invece consideriamo il cluster 2 unitamente al dendogramma da cui esso è stato generato possiamo affermare che esso in generale si riferisce ai giorni di non attività dell'impianto, la distanza a cui la maggior parte dei giorni che lo compongono sono uniti è pari a 0, ad eccezione di 2 giorni, dei quali uno coincidente con quello avente valori diversa da 0 nel cluster 3 della potenza termica.

#### Temperatura di emungimento, Tes

L'immagine sottostante, figure 5.1-9, mostra il dendogramma ottenuto dall'applicazione del metodo di Ward ai valori della temperatura di emungimento ed in ingresso alla CF1 dal condensatore, a sinistra, e l'elbow method a destra. Quest'ultimo indica di tagliare il dendogramma per k=3.

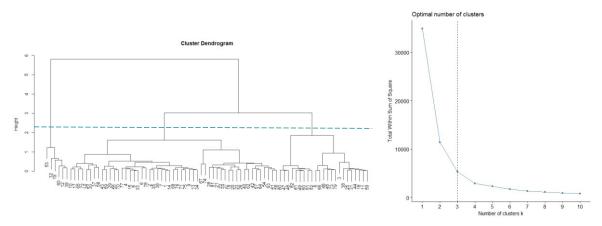


Figure 5.1-8: Cluster dendogram obtained with Ward method for entering source temperature, on the left, and elbow method with indication of the optimal number of clusters. The dash line represents the number of cluster to which such dendogram has been cut.

Fra i 3 cluster ottenuti, visibili in figure 5.1-10, due, 1 e 3, sono caratterizzati da un abbassamento ed un rialzamento in corrispondenza degli orari di accensione e spegnimento mostrati dai cluster della potenza elettrica ed in generale anche dalle altre variabili, mentre il restante presenta un andamento crescente durante l'arco della giornata.

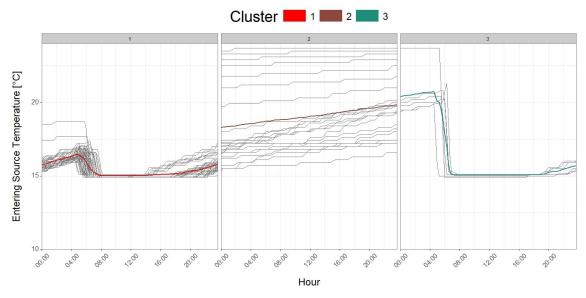


Figure 5.1-9: Clusters of entering source temperature achieved by cutting precedent dendogram for k=3.

# Temperatura di ritorno dall'ambiente, Tel

L'immagine sottostante, figure 5.1-11, mostra il dendogramma ottenuto dall'applicazione del metodo di Ward ai valori della temperatura di ritorno dall'ambiente, affiancato dall'elbow method, a destra, il quale indica di tagliare il dendogramma per k=3.

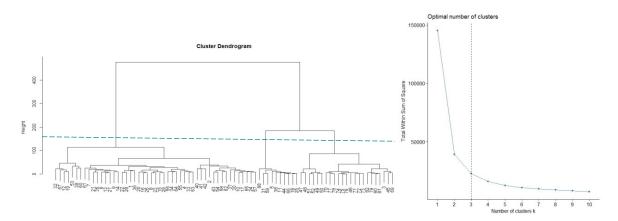


Figure 5.1-11: Cluster dendogram obtained with Ward method for entering load temperature, on the left, and elbow method with indication of the optimal number of clusters. The dash line represents the number of cluster to which such dendogram has been cut.

Fra i 3 cluster, visibili in figure 5.1-12, due, cluster 2 e 3, presentano un andamento costante o al più crescente durante l'intero arco della giornata. Il restante invece è caratterizzato da un abbassamento ed un aumento delle temperature in corrispondenza degli orari di accensione e spegnimento.

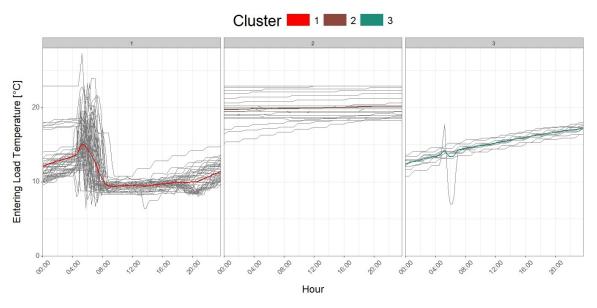


Figure 5.1-12: Clusters of entering load temperature achieved by cutting precedent dendogram for k=3.

### 5.2. Classificazione

Dopo aver individuato i cluster occorre ora effettuare una classificazione degli stessi in modo da cogliere le relazioni intrinseche esistenti dal punto di vista operativo. La strategia deve quindi prevedere la scelta di un algoritmo, che permetta la costruzione di modelli decisionali ad albero, e degli attributi da utilizzare come variabili predittive.

L'algoritmo utilizzato coincide con il CART in quanto, essendo un albero di classificazione e regressione, permette di trattare attributi categorici e numerici, tipologie presenti in egual misura fra gli attributi considerati. Ricordando che tale algoritmo produce solo split binari ed utilizza come misura di partizionamento l'indice di Gini, possiamo affermare che il suo iter si basa su tre punti fondamentali, ovvero:

- Trovare il migliore split per ciascun attributo: a prescindere dalla natura di quest'ultimo, continuo, discreto o nominale, egli passa in rassegna tutti i valori candidati come punti di divisione e sceglie quello che massimizza maggiormente il criterio di divisione quando il nodo viene suddiviso in base ad esso;
- 2. Trovare il migliore split per ciascun nodo: fra le migliori soluzioni individuate nello step 1 sceglie quello che massimizza il criterio di suddivisione;

 Dividere il nodo utilizzando la migliore suddivisione identificata nello step 2 se le condizioni di arresto non sono soddisfatte [61]. Nel nostro caso la condizione di arresto coincide con l'imporre un numero minimo di record per ciascun nodo foglia.

Per quanto riguarda invece le variabili predittive, essendo il nostro scopo quello di classificare ciascun cluster precedentemente ottenuto in funzione di condizioni esterne, o ambientali, ed interne, o operative, gli attributi considerati coincidono con quelli riportati nella tabella sottostante, table 5.2-1.

Variabili esterne	Tipo di giorno: tale attributo può avere valore compreso fra 1 e 8,
	dove quest'ultimo corrisponde ai giorni festivi
	Mese di appartenenza, indicato in formato numerico e che quindi
	potrà variare fra 6, giugno, e 9, settembre
	Temperature esterne giornaliere: massima, T_max, minima,
	T_min, e media, T_mean
Variabili operative	Valori medi giornalieri di temperatura di emungimento,
	Tes_mean, temperatura di ritorno dall'ambiente, Tel_mean,
	portata di emungimento, Qes_mean, portata di ritorno
	dall'ambiente, Qel_mean, Potenza termica, Pt_mean, e Potenza
	elettrica, Pel_mean.
	Cluster precedentemente ottenuti per temperatura di
	emungimento, clust_Tes, temperatura di ritorno dall'ambiente,
	clust_Tel, portata di emungimento, clust_Qes, portata di ritorno
	dall'ambiente, clust_Qel, Potenza termica, clust_Pt, e Potenza
	elettrica, clust_Pel.

Table 5.2-1: Summary table of attributes used for classification.

Contrariamente a quanto effettuato in precedenza, per la fase di classificazione la potenza elettrica sarà l'ultima variabile ad essere utilizzata in quanto essa, così come anche la potenza termica, è strettamente dipendente dalle temperature e le portate di esercizio che pertanto saranno le prime variabili ad essere esaminate. Inoltre, tutte le variabili saranno

inizialmente classificate in funzione delle condizioni esterne e successivamente in funzione di quelle operative.

# Portata di emungimento, Qes

Ricordando i cluster ottenuti per tale variabile e riportati in figure 5.1-6 il modello classificatore ottenuto considerando le sole variabili esterne viene riportato nell'immagine sottostante corredato della sua matrice di confusione, figure 5.2-1. Tale modello conferma la non attività dell'impianto per i giorni appartenenti al cluster 3, in quanto corrispondenti al giorno 8 della settimana, e quindi festivo, mentre identifica lavorativi i cluster 1 e 3 e li differenzia in funzione della temperatura esterna massima e minima. La matrice di confusione tuttavia mostra una maggior errore di classificazione per il cluster 2 che infatti ha un richiamo del 56%.

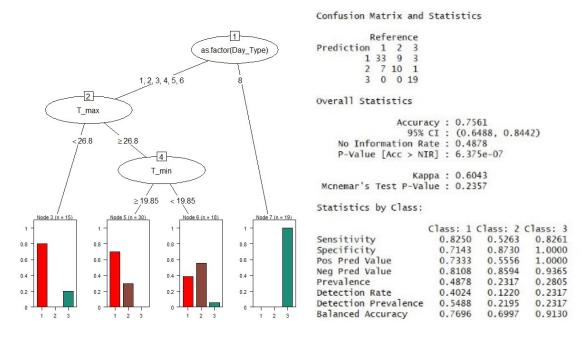


Figure 5.2-1: Classification of Qes according to external attributes.

Se invece consideriamo l'immagine sottostante relativa alla classificazione basata sulle variabili operative, figure 5.1-2, possiamo già notare una relazione con i cluster ottenuti per la temperatura di emungimento, Tes e la portata media giornaliera di ritorno dall'ambiente. In particolar modo il cluster 3 che precedentemente era stato classificato come festivo adesso trova una correlazione con un richiamo del 100% con il cluster 2 della temperatura di emungimento. La differenza fra i cluster 1 e 2 invece corrisponde ad una

differente intensità della portata media giornaliera di ritorno dall'ambiente, Qel, minore per i giorni appartenenti al cluster 1.

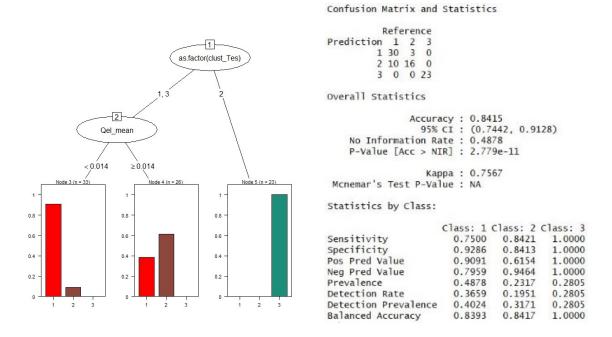


Figure 5.2-2: Classification of Qes according to operational attributes.

### Portata di ritorno dall'ambiente, Qel

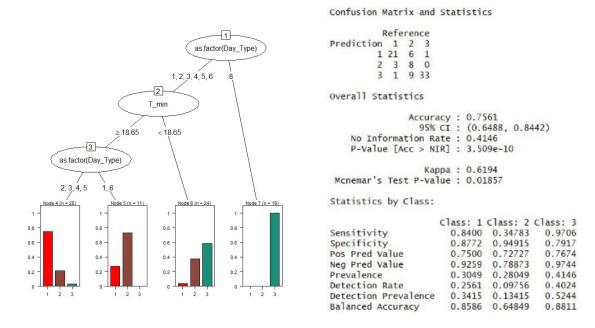


Figure 5.2-3: Classification of Qel according to external attributes.

Tenendo presente i cluster ottenuti per tale variabile, figure 5.1-8, se consideriamo l'immagine precedente relativa alla classificazione basata sulle variabili esterne, figure 5.2-3, possiamo affermare che anche in questo caso il cluster 3 corrisponde ai giorni festivi, mentre i cluster 1 e 2 differiscono fra loro in funzione dell'entità della temperatura minima giornaliera e per i tipi di giorni che essi contengono: i differenti orari di spegnimento visibili nel cluster 2 sono dovuti alla presenza di sabati al suo interno. Dalla matrice di confusione inoltre emerge un errore generale di classificazione per tutti i cluster e in special modo per il cluster 2, con un richiamo di circa il 73%.

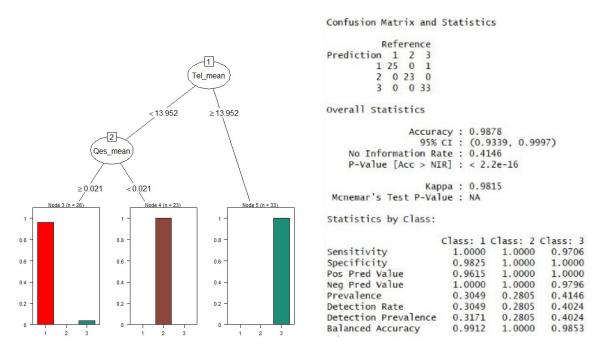


Figure 5.2-4: Classification of Qel according to operational attributes.

Il CART riportato nell'immagine sovrastante e relativo alle condizioni operative, figure 5.2-4, mostra invece con una precisione di circa il 99% una relazione fra tale variabile e i valori medi giornalieri della temperatura di ritorno dall'ambiente e della portata di emungimento. In particolar modo i cluster 1 e 2, precedentemente classificati come lavorativi, sono caratterizzati da una temperatura di ritorno dall'ambiente media giornaliera superiore ai 14 °C e si differenziano per l'intensità della portata di emungimento, maggiore per il cluster 1.

# Temperatura di emungimento, Tes

Ricordando i cluster ottenuti per tale variabile, figure 5.1-10, e considerando il CART mostrato di seguito, figure 5.2-6, possiamo affermare che il cluster 2 corrisponde ai giorni festivi, mentre i cluster 1 e 3 corrispondono ai giorni lavorativi e si differenziano per il tipo di giorno. Il cluster 3 corrisponde ai lunedì e questo spiega il valore maggiore di temperatura prima del calo in corrispondenza dell'orario di accensione: i giorni festivi sono infatti caratterizzati da un andamento crescente di tale variabile.

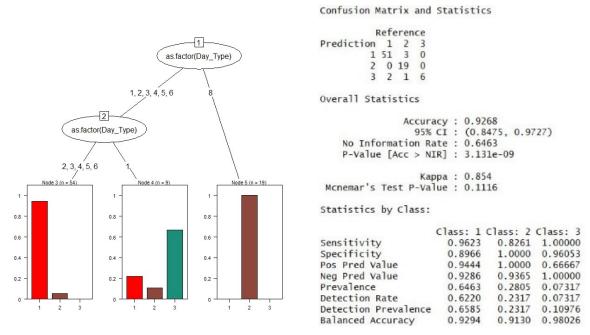


Figure 5.2-5: Classification of Tes according to external attributes.

Se invece si considera il CART ottenuto considerando le variabili operative, figure 5.2-6, possiamo affermare che i cluster 1 e 3 si differenziano in funzione della temperatura di ritorno dall'ambiente media giornaliera e per la portata di emungimento media giornaliera. In particolar modo se la portata di emungimento è minore di 0.019 m³/s allora il profilo della temperatura di emungimento corrisponderà a quelli appartenenti al cluster 1 di tale temperatura, mentre se è maggiore di 0.019 m³/s allora corrisponderà al cluster 3. I giorni lavorativi appartenenti al cluster 2 invece sono caratterizzati da una portata di emungimento minore di 0.003 m³/s.

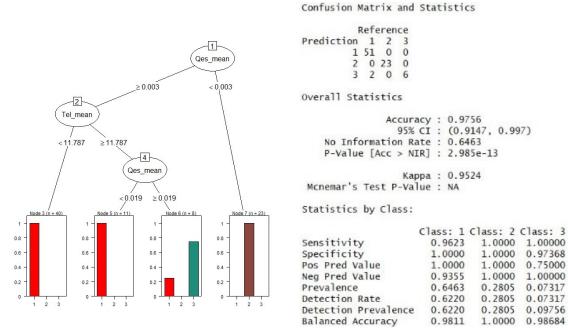


Figure 5.2-6: Classification of Tes according to operational attributes.

# Temperatura di ritorno dall'ambiente, Tel

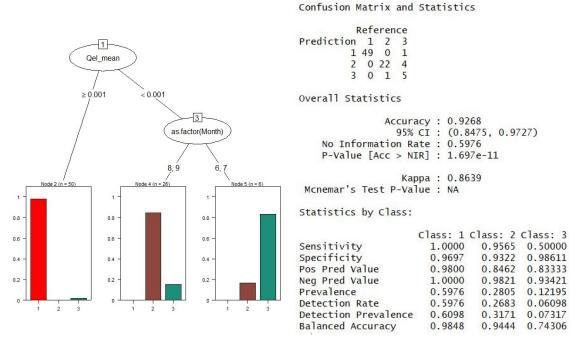


Figure 5.2-7: Classification of Tel occording to exterrnal and operation attributes.

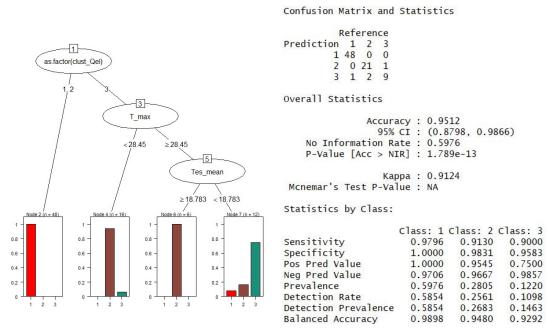


Figure 5.2-8: Classification of Tel according to external and operational attributes.

Ricordando i cluster ottenuti per tale variabile, figure 5.1-12, e osservando i modelli classificatori riportati nelle immagini sovrastanti, figure 5.2-7 e 5.2-8, possiamo affermare che il cluster 1 corrisponde ai giorni lavorativi mentre i clsuter 2 e 3 corrispondono ai giorni festivi. Tali affermazioni trovano riscontro in entrambi i CART sopra riportati. Infatti, il cluster 1 trova una forte correlazione con i cluster 1 e 2 della rispettiva portata, e quindi quando la media giornaliera di tale variabile è superiore a 0.001 m³/s. I cluster 2 e 3 corrispondono al cluster 2 sempre della stessa portata e differiscono fra loro per la temperatura di emungimento media giornaliera, minore per il cluster 3 ed i mesi di appartenenza, agosto e settembre per il cluster 2 e giugno e luglio per il cluster 3. La matrice di confusione mostra tuttavia un errore di classificazione maggiore il cluster 3.

# Potenza termica

Ricordando i cluster ottenuti per tale variabile, visibili in figure 5.1-4, e osservando il CART riportato di seguito, figure 5.2-8, e relativo alle condizioni esterne possiamo affermare che il cluster 3 corrisponde ai giorni festivi o lavorativi aventi temperatura minima giornaliera inferiore di 18,6°C. I restanti cluster invece corrispondono ai giorni lavorativi in cui la temperatura esterna minima giornaliera è superiore a 18,6°C e ciò che li differenzia

consiste, anche in questo caso, nella presenza di sabati all'interno del cluster 2. Occorre tuttavia aggiungere che quest'ultimo cluster è quello maggiormente affetto da errore di

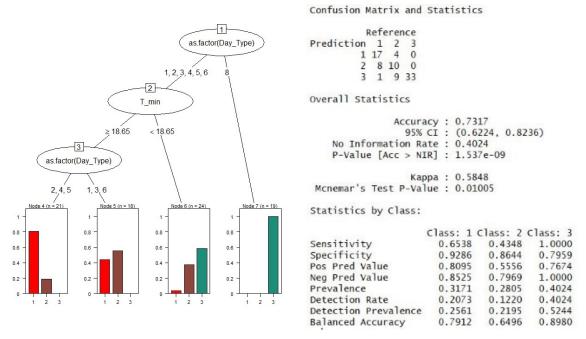


Figure 5.2-8: Classification of Thermal Power according to external conditions.

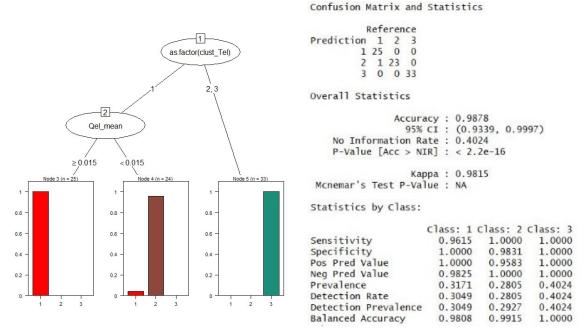


Figure 5.2-9: Classification of Thermal Power according to operational attributes. classificazione.

Se invece consideriamo figure 5.2-9 e figure 5.2-10 notiamo la forte correlazione esistente fra la potenza termica, la temperatura e la portata di ritorno dall'ambiente e la potenza

elettrica. Nella fattispecie figure 5.2-9 mostra come il cluster 3 corrisponde ai cluster 2 e 3 della Tel, mentre i cluster 1 e 2 differiscono fra di loro per l'intensità della portata media giornaliera, maggiore per il cluster 1. Il CART di figure 5.2-10 conferma la relazione fra i cluster della potenza termica e della temperatura di ritorno dall'ambiente ed in più mostra una correlazione con un richiamo minimo del 96% fra i cluster 1 delle due potenze ed il cluster 2 della potenza termica con i cluster 2 e 3 di quella elettrica.

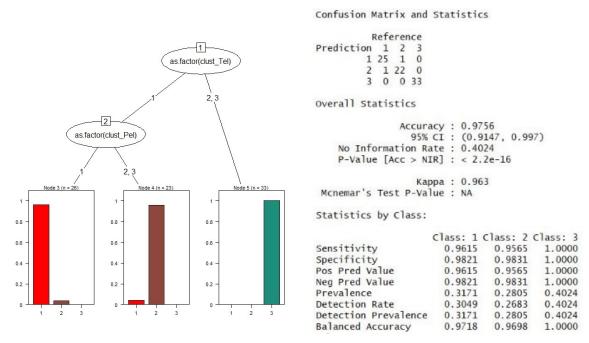


Figure 5.2-10: Classification of Thermal Power according to operational attributes.

#### Potenza elettrica

Sino ad ora la potenza elettrica è stata considerata come predittore solo della potenza termica in quanto entrambe sono strettamente dipendenti dalle modalità operative, coincidenti con le portate e le temperature. Anche in questo caso i CART sono prima stati elaborati considerando dapprima le variabili esterne: tenendo presente i cluster precedentemente ottenuti e visibili in figure 5.1-2, possiamo notare dall'immagine sottostante, figure 5.2-11, che al cluster 2 appartengono i giorni festivi, mentre i cluster 1 e 3 differiscono fra loro per temperatura massima giornaliera, maggiore per il cluster 1, e per la presenza di sabati all'interno del cluster 3.

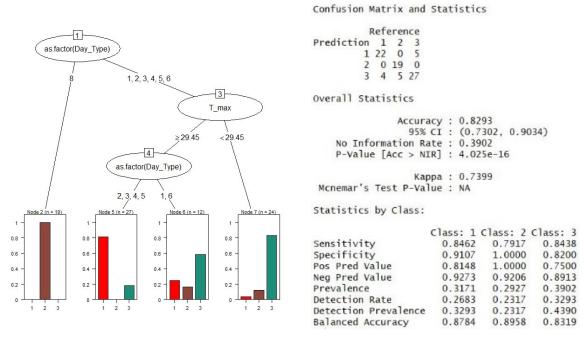


Figure 5.2-11: Classification of Electrical Power according to external attributes.

Se invece consideriamo l'immagine sottostante, figure 5.2-12, possiamo notare come il cluster 1 e 3 differiscano fra di loro per la potenza termica media prodotta, maggiore per il cluster 1. Tuttavia, la matrice di confusione mostra un maggiore errore per la classificazione del cluster 3.

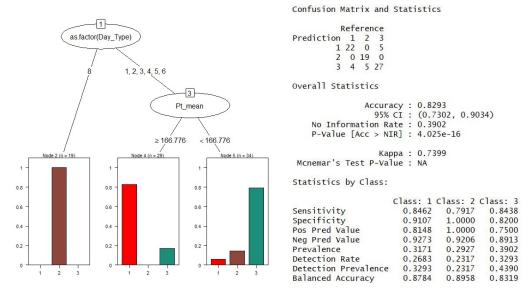


Figure 5.2-12: Classification of Electrical Power according to Day\_Type and Thermal\_Power

Allo stesso modo le immagini sotto riportate mostrano i CART ottenuti considerando le variabili operative. In particolar modo figure 5.2-13 e 5.2-14 mostrano come il cluster 1 della

potenza elettrica sia caratterizzato da una portata di emungimento e di ritorno dall'ambiente maggiore e figure 5.1-15 mostra con un richiamo del 96% come esso corrisponda al cluster 1 della potenza termica. Per quanto riguarda il cluster 2 esso corrisponde ai cluster 2 e 3 della portata di ritorno dall'ambiente e al medesimo cluster della temperatura di emungimento: viene quindi confermata la non attività dell'impianto per tali giorni. Il cluster 3 invece è quello maggiormente affetto da errore di classificazione in tutti i CART elaborati e corrisponde ai cluster 2 e 3 della portata di emungimento e ai cluster 1 e 3 della temperatura di emungimento.

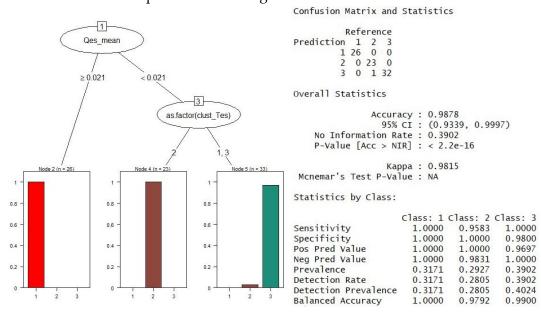


Figure 5.2-13: Classification of Electrical Power ccording to operational conditions.

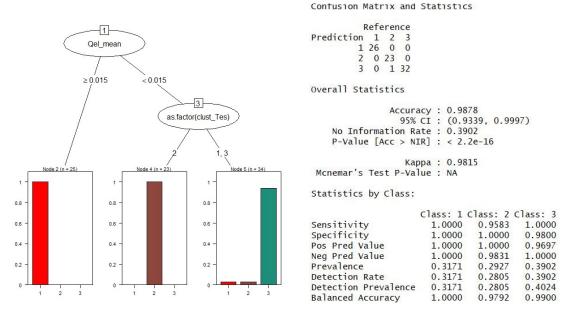


Figure 5.2-14: Classification of Electrical Power ccording to operational conditions.

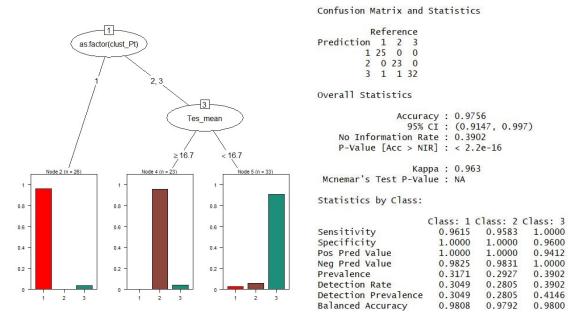


Figure 5.2-15: Classification of Electrical Power ccording to operational conditions.

Ricapitolando, le relazioni intrinseche alle variabili considerate, individuate attraverso il processo di classificazione, vengono riportate nella tabella sottostante, nella forma tipica di tale regole ovvero if—then.

If	Then
Day_Type = {1,2,3,4,5,6}  T_min \ge 20 °C  T_max < 27 °C  Qel_mean < 0.014 m <sup>3</sup> /s  Clust_Ts = {1,3}	Clust_Qes = 1
Day_Type = {1,2,3,4,5,6} T_min < 20 °C T_max ≥ 27 °C Qel_mean ≥ 0.014 m³/s Clust_Ts = {1,3}	Clust_Qes = 2
Day_Type = 8 Clust_Ts = 2	Clust_Qes = 3
Day_Type = {2,3,4,5} T_min ≥ 19 °C Tel_mean < 14 °C Qes_mean ≥ 0.021 m³/s	Clust_Qel = 1

Day_Type = $\{1,6\}$ T_min $\geq 19$ °C Tel_mean $< 14$ °C Qes_mean $\geq 0.021$ m <sup>3</sup> /s	Clust_Qel = 2
Day_Type = 8 T_min < 19 °C Tel_mean ≥ 14 °C	Clust_Qel = 3
Day_Type = {2,3,4,5,6} Qes_mean ≥0,003 m³/s Tel_mean < 12 °C Tel_mean ≥12 °C e Qes_mean < 0.019 m³/s	Clust_Tes = 1
Day_Type = 8 $Qes_mean < 0.003 \text{ m}^3/\text{s}$	Clust_Tes = 2
Day_Type = 1 Tel_mean ≥12 °C Qes_mean ≥0.019 m³/s	Clust_Tes = 3
Day_Type = {1,2,3,4,5,6} Clust_Qel = {1,2}	Clust_Tel = 1
Day_Type = 8 T_max < 29 °C T_max ≥29 °C e Tes_mean≥ 19 °C Clust_Qel = 3	Clust_Tel = 2
Day_Type = 8 T_max ≥29 °C Tes_mean ≥19 °C Clust_Qel = 3	Clust_Tel = 3
Day_Type = {1,2,3,4,5}  T_min ≥19 °C  Qel_mean ≥0.015 m³/s  Clust_Tel = 1  Clust_Pel = 1	Clust_Pt = 1
Day_Type = {1,2,3,4,5,6}  T_min ≥19 °C  Qel_mean ≥0.015 m³/s  Clust_Tel = 1  Clust_Pel = {2,3}	Clust_Pt = 2

Day_Type = 8 T_min < 19 °C Clust_Tel = {2,3}	Clust_Pt = 3
Day_Type = {1,2,3,4,5}  T_max ≥29 °C  Pt_mean ≥166.776 kW  Qes_mean ≥0.021 m³/s  Qel_mean ≥0.015 m³/s  Clust_Pt = 1	Clust_Pel= 1
Day_Type = 8 Qes_mean < $0.021 \text{ m}^3/\text{s}$ e Clust_Tes = 2 Qel_mean < $0.015 \text{ m}^3/\text{s}$ e Clust_Tes = 2 Clust_Pt = {2,3} e Tes_mean $\geq 17 \text{ °C}$	Clust_Pel = 2
Day_Type = {1,2,3,4,5,6} T_max < 29 °C Pt_mean < 166.776 kW Qes_mean < 0.021 m³/s e Clust_Tes = {1,3} Qel_mean < 0.015 m³/s e Clust_Tes = {1,3} Clust_Pt = {2,3} e Tes_mean < 17 °C	Clust_Pel = 3

Table 5.2-2: Table containing all the relationship individualized through classification phase.

## 5.3. Costruzione del modello

Come già accennato in precedenza, la seconda fase del presente lavoro di tesi consiste nella creazione di un modello che rispecchi il comportamento della CF1, attraverso una calibrazione dello stesso basata sui dati storici. In particolar modo tale fase è stata eseguita attraverso l'utilizzo del software TRNSYS, Transient System Simulation Tool, i.e. strumento di simulazione a regime transitorio, che permette di modellare non solo sistemi energetici, termici ed elettrici, ma anche altri sistemi dinamici come il flusso del traffico o i processi biologici. Esso si compone di due parti:

 un elaboratore, chiamato kernel, in grado di leggere ed elaborare i file di input, risolvere iterativamente il sistema in modo da determinarne la convergenza, eseguire regressioni lineari, interpolare file esterni e tracciare le variabili del sistema. • una libreria di componenti, type, che permette di modellare le prestazioni di una parte di sistema, come ad esempio pompe, edifici multizona, sistemi HVAC.

La peculiarità di tale software consiste nel fatto che ciascun utente può aggiungere nuovi componenti, utilizzando il linguaggio di programmazione Fortran, o modificare quelli già esistenti permettendone dunque la calibrazione.

Ai fini di una maggiore comprensione, il presente paragrafo verrà organizzato come segue: descrizione del modello costruito e dei componenti che lo compongono, paragrafo 5.3.1, descrizione della fase di calibrazione e validazione, paragrafo 5.3.2.

#### 5.3.1. Descrizione del modello

Il modello costruito per la simulazione della CF1 viene riportato nella figura sottostante, figure 5.3.1-1 e al suo interno possiamo notare 5 componenti diversi, ovvero:

- 1 Type 927, rinominato CF1;
- 6 Data Reader, ovvero: T\_es, Q\_es, T\_el, Q\_el, Thermal Power, Electrical Power;
- 2 stampanti: online plotter e printer;
- 1 file meteo, contenente i dati metereologici della città di Torino.

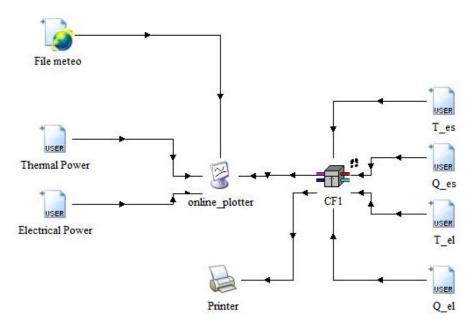


Figure 5.3-1: Trnsys model built to simulate CF1.

L'elemento fondamentale del presente modello coincide con il Type 927, ovvero una pompa di calore a singolo stadio, operante in regime invernale e/o estivo e la cui logica di controllo è in funzione della temperatura: quando il segnale di controllo, definito dall'utente, indica che l'unità deve essere accesa in modalità riscaldamento o raffreddamento, allora esso funzionerà al suo livello di potenza termica finché i valori del segnale di controllo non cambiano.

Tale type non è un modello primario ma si basa invece sui dati di catalogo resi disponibili dai produttori di pompe di calore. Il cuore di tale componente consiste in due file di testo esterni, uno per ogni stagione considerata, forniti dall'utente e corrispondenti ad una mappatura dell'impianto. Essi devono essere costruiti seguendo un ordine specifico descritto nel capitolo 7 del manuale di programmazione di TRNSYS [62] e devono contenere i dati di catalogo relativi alla potenza elettrica e termica, variabili dipendenti, in funzione delle temperature e delle portate in ingresso sia dal lato ambiente che dal lato geotermico, variabili indipendenti. Nello specifico, tale documento dovrà contenere:

- m valori normalizzati di portata in ingresso dall'ambiente, Qelnorm;
- n valori normalizzati di portata di emungimento, Qesnorm;
- j valori della temperatura di ritorno dall'ambiente espressi in °C, Tel;
- k valori della temperatura di emungimento espressi in °C, Tes
- x valori normalizzati di potenza termica;
- x valori normalizzati di potenza elettrica

#### Dove:

- $\{m, n, j, k\} \ge 3$ ;
- x = m\*n\*j\*k, ovvero un valore per ogni combinazione ottenibile dalle variabili indipendenti;
- i valori delle portate e delle potenze dovranno essere normalizzati sui rispettivi valori nominali, se quindi consideriamo ad esempio la portata di emungimento avremo che  $Qes_{norm} = \frac{Qes_i}{Qes_{nominale}}$ ;
- i valori delle temperature dovranno rispettare l'ipotesi che il calore specifico sia costante nel range considerato.

Allo stesso modo, all'interno dell'ambiente di lavoro dovranno essere forniti una serie di parametri ed input, riassunti nella tabella sottostante.

Parametri	Input
Caratteristiche del fluido, densità e calore specifico	Valore della temperatura di emungimento, Tes
Il numero degli j, k valori inseriti per entrambe le temperature sia entrambe le modalità di funzionamento (riscaldamento e raffreddamento)	Valore della portata di emungimento, Qes
Il numero dei m, n valori inseriti per entrambe le portate	Valore della temperatura di ritorno dall'ambiente, Tel
Valori nominali di potenza termica per il riscaldamento ed il raffrescamento	Valore della portata di ritorno dall'ambiente, Qel
Valori nominali di potenza elettrica per il riscaldamento ed il raffrescamento	Segnale di comando per il raffrescamento $(1 = on, 0 = off)$
Valori nominali di portata volumetrica in ingresso dal condensatore e dall'evaporatore	Fattore di scala (può essere utilizzato per simulare più pompe di calore connesse in parallelo e che possono essere controllate individualmente)
Numero di pompe identiche che deve essere modellato con tale type	Numero delle identiche pompe di calore in funzione

Table 5.3-1: Parameters and inputs that must be specify in the job environment.

Attraverso la funzione *InterpolateData*, TRNSYS legge i dati di input (Tes, Tel, Qes, Qel) e basandosi sui range indicati all'interno dei file esterni, andrà ad interpolarli in modo multidimensionale restituendo i rispettivi valori delle variabili dipendenti (P\_elettrica e P\_termica). Occorre tuttavia aggiungere che tale funzione è solo in grado di interpolare all'interno dei range specificati nei file esterni e non di estrapolare: pertanto se un dato input ricade al di sotto o al di sopra dell'intervallo specificato allora ad esso verrà assegnato rispettivamente il valore minimo o massimo presente all'interno del file esterno.

Gli output di tale componente saranno:

temperatura di mandata, Tol [°C];

- portata di mandata, Qol [kg/s];
- temperatura di re-immissione in falda, Tos [°C];
- portata di re-immissione in falda, Qos [kg/s];
- potenza elettrica, P\_elettrica [kW];
- potenza termica prodotta dai chiller, P\_termica [kW];
- potenza termica prodotta dall'impianto geotermico, P\_geo [kW];
- COP [-].

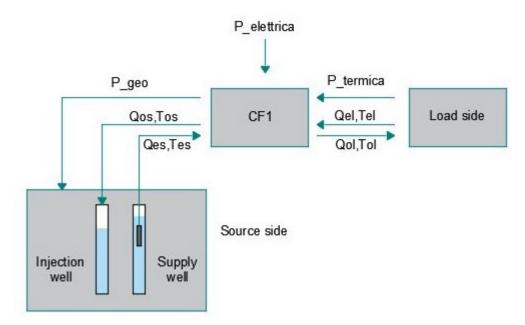


Figure 5.3-2: Simplified representation of cooling plant with indication of entering and outgoing flow.

Al fine di una più corretta comprensione delle equazioni riportate di seguito ed utilizzate da Trnsys per determinare gli output, l'immagine sovrastante (figure 5.3-2) mostra una rappresentazione schematica della CF1 con indicazioni circa i flussi termici, entranti ed uscenti, ed il flusso elettrico.

Come già detto in precedenza, i valori di potenza termica prodotta dalla pompa di calore ed elettrica vengono calcolati tramite la funzione *InterpolateData* mentre il valore di potenza termica prodotta dall'impianto geotermico viene calcolata considerando il primo principio della termodinamica ed effettuando dunque un bilancio energetico. Esso sarà:

$$P\_geo = P\_termica + P\_elettrica$$

Noti tutti i valori di potenza elettrica e termica, le temperature in uscita dall'impianto da ambo i lati, evaporatore e condensatore, verranno calcolate tramite le seguenti equazioni:

$$Tos = Tes - \frac{P\_geo}{Mes * cp}$$

$$Tol = Tel + \frac{P\_termica}{Mel * cp}$$

I valori delle portate verranno calcolati in funzione del primo principio della termodinamica, ovvero:

$$Qol = \frac{P\_termica}{cp * (Tel - Tol)}$$

$$Qos = \frac{P\_geo}{cp * (Tes - Tos)}$$

Il coefficiente di performance verrà invece calcolato come rapporto fra la potenza termica e quella elettrica, ovvero:

$$COP = \frac{P\_termica}{P\ elettrica}$$

Dove:

- Mes = portata massica di emungimento, [kg/hr];
- Mel = portata massica di ritorno dall'ambiente, [kg/hr];
- Cp = 4.186 [J/kgK], calore specifico del fluido considerato, ovvero acqua.

Un altro componente presente all'interno del modello coincide con il type 9e, la cui funzione è quella di leggere file di dati ad intervalli di tempo regolari e allo stesso tempo convertirli in un formato leggibile per tutti gli altri componenti. Il timestep a cui i dati devono essere letti viene indicato all'interno dell'area di lavoro e deve essere costante per ciascun file di dati. Tale type viene utilizzato per fornire i dati di input al type 927, ovvero file T\_es, Q\_es, T\_el e Q\_el; nello specifico le temperature dovranno essere espresse in °C mentre le portata in kg/s. Allo stesso modo, il type 9e viene utilizzato per confrontare la

performance del type 927 attraverso i file contenenti i dati reali relativi a potenza elettrica e termica, Thermal Power e Electrical Power. Se si osserva infatti figure 5.3.1-1 si nota infatti che tali file non sono collegati al type 927 ma al plotter online, il cui scopo è quello di rendere visibile l'andamento delle variabili ad esso collegato mentre si sta effettuando la simulazione.

Allo stesso modo, per permettere la stampa e quindi il salvataggio dei dati di simulazione in un file esportabile in altri software è stata inserita all'interno del modello una stampante in cui devono essere fornite indicazioni circa il timestep, le unità di misura e il periodo di simulazione considerato.

Inoltre, nell'ambiente di lavoro devono essere fornite le indicazioni circa il periodo di simulazione considerato ed il timestep con cui la simulazione deve essere effettuata, nel nostro caso l'intervallo di tempo considerato sarà di 15 minuti e coinciderà con quello dei Data Reader.

#### 5.3.2. Calibrazione e validazione

Dopo aver descritto i componenti del modello costruito ed il loro funzionamento, occorre ora descrivere la fase di calibrazione che coincide principalmente con la creazione del file esterno che regolerà il type 927 durante il periodo di raffrescamento (si ricorda che la CF1 opera in sola climatizzazione estiva) e con il settaggio dello stesso.

Prima di procedere con la descrizione della fase di costruzione del file esterno occorre effettuare una serie di premesse.

Nei precedenti capitoli abbiamo già accennato al fatto che i valori di potenza elettrica forniti dal Living Lab sono riferiti all'intera CF1 pertanto il type 927 modellerà l'intera centrale frigorifera a prescindere dal numero di chiller di cui essa si compone. Per tale motivo quindi non ci si potrà avvalere dei dati di catalogo dei rispettivi chiller e i valori utilizzati per calibrare tale type coincideranno con quelli ottenuti dalle fasi di clustering e di classificazione. Queste ultime due fasi infatti sono state di fondamentale importanza in

quanto conoscere il profilo tipico giornaliero equivale a conoscere i range che le variabili assumono durante il periodo di attività, e non, dell'impianto. Se infatti consideriamo i cluster relativi ai giorni lavorativi di tali variabili possiamo avere una panoramica circa i range all'interno dei quali andare a ricercare i valori caratteristici di ciascuna di esse. Se consideriamo le sole ore lavorative si può osservare che:

- la temperatura di emungimento, Tes, assume valori nell'intorno di 15 °C;
- la temperatura di ritorno dall'ambiente, Tel, è caratterizzata da un range più ampio infatti durante la fase di attività essa varia fra gli 8 °C e i 10°C circa;
- la portata di emungimento ha un range molto elevato, essa infatti come si nota dai rispettivi cluster ha un andamento oscillante che varia dai 0 m³/s ai 0.08 m³/s.
- la portata di ritorno dall'ambiente al contrario si caratterizza di valori più stazionari durante il funzionamento che corrispondono all'incirca a 0.027 m<sup>3</sup>/s.

Noti i range di partenza occorre ora andare ad individuare i valori caratteristici: le difficoltà relative a tale fase, tuttavia, sono state molteplici poiché, come già detto in precedenza, per ciascuna variabile indipendente devono essere individuati almeno 3 valori caratteristici. Ciascuno di essi deve essere collegato ai restanti valori delle altre variabili: ovvero, affinché un valore sia considerato caratteristico e quindi utilizzato come input all'interno del file esterno, esso deve essere presente in ciascuna delle combinazioni ottenibili per i valori caratteristici, di tutte le variabili, individuati. Tali combinazioni inoltre devono corrispondere a condizioni operative che sono effettivamente avvenute nel normale funzionamento dell'impianto perché se così non fosse non si potrebbero indicare i valori di potenza elettrica e termica corrispondenti e quindi non si potrebbe calibrare l'impianto.

Abbiamo già accennato in precedenza che il file esterno deve essere costruito seguendo un ordine ben specifico. Esso corrisponde ad un file .dat costruibile attraverso Excel all'interno del quale le prime 4 righe corrispondono ai valori delle variabili indipendenti, separati da uno spazio, ed ordinati in ordine crescente. Un file di esempio fornito da Trnsys, parzialmente rappresentano in figure 5.3-2, indica inoltre l'ordine in funzione delle righe con cui tali variabili devono essere inserite, ovvero:

1ª riga: valori normalizzati della portata di ritorno dall'ambiente, Qel\_norm;

- 2 ª riga: valori normalizzati della portata di emungimento, Qes\_norm;
- 3 ª riga: valori relativi alla temperatura di ritorno dall'ambiente, Tel;
- 4 ª riga: valori relativi alla temperatura di emungimento, Tes.

Dalla 5 ª riga in poi devono invece inseriti rispettivamente i valori normalizzati di potenza termica ed elettrica, anche questi con un ordine ben preciso.

U2	2	~	$\times$	~	fx							
4	Α	В	С	D	Е	F	G	Н	1	J	K	L
1	0,5429	1	1,2714		!Normaliz	ed Load Fl	ow Rate					
2	0,5429	1	1,2714		!Normaliz	ed Source	Flow Rate					
3	7,2222	12	18,3333	23,8889	!Entering	Load Temp	eratures					
4	10	15,5556	21,1111	26,6667	32,2222	37,7778	43,3333	48,8889	Entering S	ource Tem	peratures	
5	0,9043	0,7176			!Normalized Capacity and Normalized Power at 10,0 C (T_Source_In),					rce_In), 7,22		
6	0,8612	0,8092				!Normalized Capacity and Normalized Power at 15,56 C (T_Source_In), 7,:					urce_In), 7,2	
7	0,8325	0,916				!Normaliz	ed Capacit	ty and Nor	malized Po	ower at 21	,1 C (T_Sou	rce_In), 7,22
8	0,8086	1,0382				!Normaliz	ed Capacit	ty and Nor	malized Po	ower at 26	,67 C (T_So	urce_In), 7,2
9	0,7943	1,1832				!Normaliz	ed Capacit	ty and Nor	malized Po	ower at 32	,2 C (T_Sou	rce_In), 7,22
10	0,7799	1,3511				!Normaliz	ed Capacit	ty and Nor	malized Po	ower at 37	78 C (T_So	urce_In), 7,2
11	0,7703	1,5344										
12	0,7608	1,7328										
13	0,9952	0,7328				!Normaliz	ed Capacit	ty and Nor	malized Po	ower at 10	,0 C (T_Sou	rce_ln), 12,0
14	0,9474	0,8168										
15	0,9187	0,9237										
16	0,8947	1,0534										

Figure 5.3-3: Partial view of Trnsys example file.

Considerando l'immagine sovrastante, nella tabella sottostante, table 5.3-1, viene riportato un esempio di come i valori normalizzati di potenza elettrica e termica devono essere inseriti all'interno del file.

Riga	Valori normalizzati di potenza elettrica e termica	Condizioni operative
5	0.9043 – 0.7176	Tes= 10 Tel=7.2222 Qes_norm=0.5429 Qel_norm=0.5429
6	0.8612 – 0.8092	Tes= 15.5556 Tel=7.2222 Qes_norm=0.5429 Qel_norm=0.5429
7	0.8325 - 0.916	Tes= 21.111 Tel=7.2222 Qes_norm=0.5429 Qel_norm=0.5429

8	0.8086 - 1.0382	Tes= 26.667 Tel=7.2222 Qes_norm=0.5429 Qel_norm=0.5429
13	0.9952 – 0.7328	Tes= 10 Tel=12 Qes_norm=0.5429 Qel_norm=0.5429
14	0.9474 - 0.8168	Tes= 15.5556 Tel=12 Qes_norm=0.5429 Qel_norm=0.5429

Table 5.3-2: Example of how normalized values of thermal and electrical power have to be written in Trnsys external file considering figure 5.3-2.

A fronte di ciò appare dunque chiaro che la prima operazione da effettuare consiste nell'individuazione dei valori nominali in modo da poter normalizzare entrambe le portate e le potenze.

Per quanto riguarda i valori nominali relativi alle due portate, Q\_es e Q\_el, e alla potenza elettrica, essi sono stati ricavati considerando la definizione stessa di valore nominale, ovvero il valore massimo teorico a cui l'impianto dovrebbe lavorare e per cui è stato progettato. Per tali variabili esso coinciderà dunque con il valore massimo presente all'interno del dataset.

Differente è invece stato l'approccio considerato per la potenza termica in quanto il suo valore massimo, pari a circa 1900 kW, non è frequente durante l'intero periodo considerato. Se infatti si confrontano i cluster lavorativi di queste 4 variabili si nota che mentre il valore di picco delle portate e della potenza elettrica è pressoché costante durante tutta la stagione, il valore di picco della potenza termica è molto più variabile. Al fine di migliorare la fase di calibrazione tale valore di picco è stato quindi individuato attraverso box plot o diagramma a scatola e baffi, ovvero un metodo di rappresentazione che descrive la distribuzione dei valori di ciascuna variabile a partire dai 5 numeri di sintesi: minimo, 1° quartile ( $Q_1$ ), mediana, 3° quartile ( $Q_3$ ), massimo. Esso permette inoltre di individuare gli eventuali valori anomali assunti da una variabile, ovvero sia IQR = ( $Q_3$ - $Q_1$ ) la distanza interquartilica indicante la dispersione della distribuzione, se un determinato valore x è tale che x  $\geq$  ( $Q_3$ +1.5 IQR) o x  $\leq$  ( $Q_1$ -1.5 IQR) allora tale valore sarà considerato anomalo.

Nell'immagine sottostante, figure 5.3-2, viene riportato il box plot ottenuto considerando i valori della potenza termica durante il funzionamento dell'impianto. Al suo interno possiamo notare che il valore massimo del dataset, 1909 kW, corrisponde ad un valore anomalo mentre il valore massimo corrisponde a 519.5 kW.

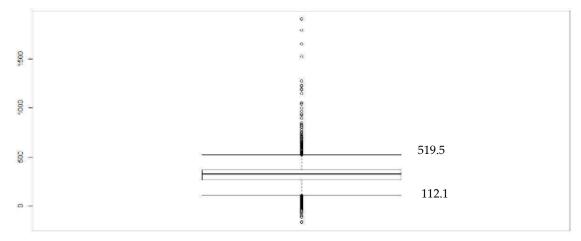


Figure 5.3-4: Thermal Power Boxplot with indication of minimun and maximum values.

Ricapitolando i valori nominali considerati vengono riportati in table 5.3-2.

Potenza elettrica	246.8 kW
Potenza termica	519.5 kW
Portata di emungimento, Qes	313200 l/hr
Portata di ritorno dall'ambiente, Qel	103680 l/hr

Table 5.3-3: Nominal values for electrical/thermal power and source/load flowrate.

Una volta individuati i valori nominali si è poi proceduto alla normalizzazione di tali variabili e quindi alla costruzione del file. Per tale fase inoltre, sono stati considerati solo i cluster relativi ai giorni lavorativi ed in particolar modo solo gli intervalli temporali in cui l'impianto è in funzione e tali per cui P\_elettrica≠0. In questo modo si è ridotto notevolmente lo spazio all'interno del quale ricercare i valori caratteristici delle modalità operative.

Se riconsideriamo i cluster ottenuti per le variabili indipendenti, mostrati nel paragrafo 5.1, si nota come i valori meno variabili durante il funzionamento della CF1 corrispondono a quelli relativi alla portata di ritorno dall'ambiente. Essa pertanto coincide con la prima variabile analizzata al fine di individuare i valori più frequenti. Se consideriamo

l'immagine sottostante, figure 5.3-4, notiamo come i valori normalizzati di tale variabile siano molto prossimi ad 1 ed in particolar modo si collochino nell'intorno di 0.95.

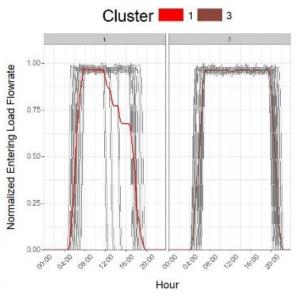


Figure 5.3-5: Working day clusters of normalized entering load flowrate values.

L'ipotesi su cui si basa l'individuazione dei valori caratteristici consiste nel supporre che tali valori corrispondono a quelli più frequenti. Pertanto, essi sono stati indentificati considerando la moda di ciascuna variabile sotto determinate condizioni e l'ordine con cui tali variabili sono state analizzate è in funzione della variabilità dei valori che esse assumono durane il funzionamento e pertanto sarà:

- 1. Valori normalizzati della portata di ritorno dall'ambiente (Qel\_norm),
- 2. Valori relativi alla temperatura di emungimento (Tes),
- 3. Valori relativi alla temperatura di ritorno dall'ambiente (Tel),
- 4. Valori normalizzati della portata di emungimento (Qes\_norm).

Il primo valore individuato corrisponde dunque alla moda dei valori normalizzati di portata di ritorno dall'ambiente ovvero 0.97. A questo punto, attraverso la visualizzazione filtrata, si è considerata la sola porzione del dataset tale per cui la portata di ritorno dall'ambiente corrisponde a 0.97 e si è proceduto all'individuazione della moda dei valori relativi alla temperatura di emungimento che corrisponde a 15.1.

Considerando nuovamente la porzione del dataset tale per cui Qel\_norm=0.97 e Tes=15.1 si è poi individuata la moda per le restanti variabili. In particolar modo i valori più frequenti di tali variabili corrispondono a Tel=9.6 e Qes\_norm=0.5. Il valore di Tel=9.6 verrà tuttavia successivamente scartato in quanto corrisponde a condizioni operative molto limitate per le altre variabili a cui è accoppiato ed in particolar modo per la portata di emungimento che assume valori compresi nell'intervallo [0.4-0.6]. Il valore di Qes\_norm = 0.5 viene invece inserito all'interno del file e quindi utilizzato per la fase di calibrazione, riga 3 di figure 5.3-5 e dell'Allegato II.

Il secondo step è consistito nel considerare il dataset tale per cui Qel\_norm=0.97 e Tes≠15.1. Il nuovo valore più frequente per tale variabile coincide con Tel=15. Considerando ora la porzione di dataset per cui Qel\_norm=0.97 e Tes=15 le mode per le restanti variabili sono Tel= 8.6 e Qes\_norm=0.8. Anche quest'ultimo valore è stato scartato in quanto esso si ha solo per Tes= [14.9, 15.0].

Allo stesso modo il terzo step è consistito nel considerare Qel\_norm=0.97, Tes≠15 e Qes\_norm≠0.8. I valori più frequenti per queste condizioni sono: Tes=14.9, Tel=8.6 e Qes\_norm=0.2. Il valore di 8.6 per la temperatura di ritorno dall'ambiente viene nuovamente confermato e pertanto assume una maggiore rilevanza. Esso verrà quindi successivamente utilizzato per verificare la candidabilità di nuovi valori caratteristici. Tali valori individuati per la portata di emungimento e la temperatura di ritorno dall'ambiente vengono inoltre inseriti all'interno del file esterno come è visibile in figure 5.3-5 e in Allegato II, rispettivamente riga 2 e 3. Allo stesso modo anche i valori individuati per la temperatura di emungimento vengono utilizzati per lo stesso scopo in quanto essa assume valori pari a [14.9, 15, 15.1] durante il funzionamento dell'impianto. Come si può notare da figure 5.3-5 e dall'Allegato II tali valori della temperatura di emungimento sono stati utilizzati per la calibrazione del modello.

Se invece si elimina il filtro relativo a Qel\_norm e si considera dunque il dataset tale per cui Qel\_norm≠0.97 il nuovo valore più frequente coincide con 0.98. Per tale valore la moda di Tes coincide con 15.1 e tenendo presente entrambe le condizioni i valori più frequenti per le restanti variabili sono: Tel= 9.9 e Qes\_norm=0.5. Questi ultimi valori individuati

vengono confermati ed inseriti all'interno del file, come è visibile in figure 5.3-5 e in Allegato II, riga 3 e 2 rispettivamente.

Come si può intuire, l'iter processuale considerato è stato temporalmente dispendioso in quanto affinché ciascun probabile valore x di qualunque variabile esaminata sia considerato caratteristico esso deve essere in relazione con i probabili valori delle restanti variabili che sono stati individuati prima e dopo di esso. L'obiettivo è sì quello di individuare almeno 3 valori per ogni variabile ma anche fare i modo che tali valori ricoprano un range più ampio possibile in modo tale da permettere un miglior funzionamento del type.

In generale possiamo dunque affermare che il procedimento effettuato ha visto i seguenti step:

- 1. Trova la moda per Qel\_norm
- Considerando il dataset tale per cui Qel\_norm coincide con il valore individuato nello step 1, trova la moda per Tes
- 3. Considerando i valori ottenuti dallo step 1 e 2, trova la moda per Tel e Qes\_norm.
- 4. Considerando il dataset tale per cui Qel\_norm coincide con il valore individuato nello step 1, trova il secondo valore più frequente per Tes
- 5. Considerando i valori ottenuti dallo step 1 e 4, trova la moda per Tel e Qes\_norm.
- 6. Confronta i valori ottenuti dallo step 5 con quelli ottenuti nello step 3.
- 7. Considerando il dataset tale per cui Qel\_norm coincide con il valore individuato nello step 1, trova il terzo valore più frequente per Tes
- 8. Considerando i valori ottenuti dallo step 1 e 7, trova la moda per Tel e Qes\_norm.
- 9. Confronta i valori ottenuti dallo step 8 con quelli ottenuti nello step 3 e 5.
- 10. Considera dataset tale per cui Qel\_norm ≠step 1
- 11. Ripeti finché non vengono individuati almeno 3 valori per ciascuna variabile.

Tali step vengono espressi in formato grafico nel flowchart mostrato di seguito, figure 5.3-6.

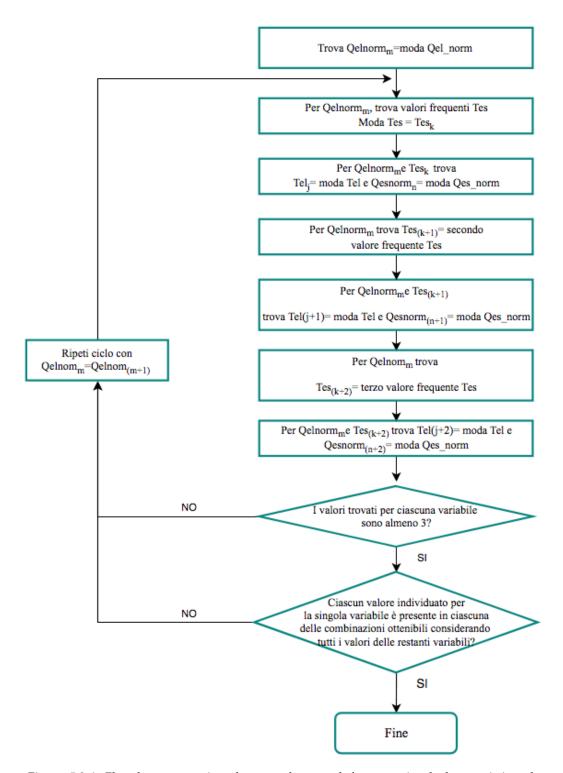


Figure 5.3-6: Flowchart concerning the procedure used for operational characteristic values identification.

In definitiva i valori tipici del funzionamento della CF1 individuati sono:

• Tes = 14.9 - 15.0 - 15.1 [°C]

- Tel = 8.6 9 °C 9.2 9.5 9.9 [°C]
- Qes\_norm = 0.2 0.4 0.5 0.7
- Qel norm = 0.95 0.96 0.97 0.98

Ad essi inoltre si aggiungono 192 valori di potenza termica ed elettrica normalizzati sul loro valore nominale. Nell'immagine sottostante viene riportata una rappresentazione parziale del file costruito, figure 5.3-7, mentre in Allegato II è possibile vedere l'intero file.

T2	23	<b>▼</b> : []	×	fx					
4	Α	В	С	D	Е	F	G	Н	1
1	0.95	0.96	0.97	0.98		!Normaliz	ed Load Fl	ow Rate	
2	0.2	0.4	0.5	0.7		!Normaliz	ed Source	Flow Rate	
3	8.7	9	9.2	9.5	9.9	!Entering	Load Tem	peratures	
4	14.9	15	15.1			!Entering	Source Te	mperatures	
5	0.4674	0.3258				!14.9-8.7-	0.2-0.95		
6	0.4449	0.3291							
7	0.4706	0.3258							
8	0.5118	0.3451				!14.9-9-0.	2-0.95		
9	0.2759	0.3446							
10	0.5095	0.2665							
11	0.5600	0.4357				!14.9-9.2-	0.2-0.95		
12	0.5635	0.3574							
13	0.5459	0.3540							
14	0.6252165	0.39405				!14.9-9.5-	0.2-0.95		
15	0.6241	0.3836							
16	0.6169393	0.3649066	67						
17	0.7113	0.4462				!14.9-9.9-	0.2-0.95		
18	0.7091	0.4494							
19	0.7037536	0.45025							
20	0.4579	0.3223				!14.9-8.7-	0.2-0.96		
21	0.4768	0.3066							

Figure 5.3-7: Partial view of Trnsys external file built for CF1.

Occorre tuttavia aggiungere che tali valori sono tipici del funzionamento a regime dell'impianto e non del funzionamento a carico parziale o transitorio. Quest'ultimo infatti è caratterizzato da condizioni operative maggiormente variabili, rispetto alla fase a regime, e da un intervallo temporale molto ridotto, circa il 10% delle ore di funzionamento giornaliere. Pertanto, per riuscire a calibrare l'impianto durante tale fase occorrerebbe una maggiore densità dei dati: il timestep di 15 minuti è sufficiente per calibrare l'impianto

durante il funzionamento a regime ma non durante quello transitorio. Si ricorda inoltre che i valori delle variabili termiche, e di conseguenza della potenza termica, sono sì caratterizzate da un timestep di 15 minuti ma sono state ottenute tramite interpolazione lineare di dati aventi un timestep iniziale di 35 e 70 minuti a causa delle problematiche relative ai contaimpulsi di cui si è parlato nel capitolo precedente. Per cui, molte delle informazioni che si sarebbero potute ottenere con un timestep reale di 15 minuti sono state perse.

Dopo aver completato la costruzione del file esterno occorre ora inserire all'interno dell'area di lavoro i parametri e gli input precedentemente descritti. Il loro settaggio viene mostrato nella tabella sottostante.

Parametri	Input		
Densità = 1004 kg/m³	I valori di Tes sono forniti tramite il		
Calore specifico= 4.190 kJ/ kg K	type 9e in °C		
Numero di valori per Tes= 3	I valori di Qes sono forniti tramite il		
Numero di valori per Tel= 5	type 9e in kg/s		
Numero di valori iniseriti per Qes= 3	I valori di Tel sono forniti tramite il		
Numero di valori inseriti per Qel=4	type 9e in °C		
Valore reminele di notonza termica- E10 E LIVI	I valori di Qel sono forniti tramite il		
Valore nominale di potenza termica= 519.5 kW	type 9e in kg/s		
Valore peminale di petenza elettrica 246 8 kW	Segnale di controllo per il		
Valore nominale di potenza elettrica= 246.8 kW	raffrescamento = 1, ovvero On		
Valore nominale di Qes= 0.087 m³/s	Fattore di scala =1		
Numero di nomno di coloro identiche 1	Numero di pompe di calore in		
Numero di pompe di calore identiche= 1	funzione= 1		

Table 5.3-4: Entered values for parameters and input in Trnsys workspace.

Il passo successivo alla creazione del modello consiste nella validazione dello stesso utilizzando come misure di accuratezza l'errore medio assoluto percentuale, MAPE (Mean Absolute Percentage Error), e l'errore medio assoluto, MAE (Mean Absolute Error).

Il primo di tali indici è stato scelto in quanto esso rappresenta uno degli indici maggiormente utilizzati per il confronto tra i dati reali e quelli predetti. In particolar modo esso corrisponde alla media aritmetica degli errori relativi, presi in valore assoluto e moltiplicati per 100, ovvero:

$$MAPE = \left(\frac{1}{N} \sum_{k=1}^{N} \frac{|Actual_k - Forecast_k|}{|Actual_k|}\right) * 100$$

La peculiarità di tale indice consiste nel fatto che esso media l'errore sulla sua rilevanza effettiva in quanto tiene conto del numero totale di campioni utilizzati, N. Tuttavia, esso è molto sensibile ai valori nulli o molto prossimi allo zero producendo dunque una distorsione dell'errore complessivo.

Il MAE invece è stato scelto in modo da valutare l'effettivo scostamento fra valori reali e quelli predetti. Esso corrisponde ad una media degli errori assoluti, considerati in valore assoluto e sarà dato dunque dalla seguente equazione:

$$MAE = \left(\frac{1}{N} \sum_{k=1}^{N} |Actual_k - Forecast_k|\right)$$

Inoltre, prima della fase di validazione l'intero dataset è stato sottoposto ad un ulteriore pre-processamento: circa il 50% dei campioni è stato rimosso in quanto caratterizzato da una incongruenza temporale delle variabili. Il nuovo dataset sarà quindi composto da 42 giorni.

Nell'immagine sottostante, figure 5.3-8, è possibile vedere i differenti andamenti fra i valori reali e simulati per la potenza elettrica, in alto, e l'andamento dell'errore relativo, in basso, relativi ad una tipica giornata lavorativa estiva, in particolare il 18 Luglio. Nella fattispecie, possiamo notare come durante la fase di funzionamento a regime il modello si adatti bene al reale funzionamento della CF1: la distribuzione degli errori relativi, in funzione delle ore giornaliere, mostra infatti come tali errori siano inferiori al 20%. La stessa cosa non può affermarsi per quanto riguarda il funzionamento a carico parziale: durante tale fase infatti il modello deriva e l'errore relativo raggiunge anche il 100%. Tuttavia, tale valore per l'errore relativo non ha un vero e proprio significato in quanto durante tale fase il modello non è calibrato. Infatti, tutti i valori di input ricadono al di fuori dei range specificati all'interno del file esterno ed il valore di picco tipico della fase di accensione come verrà confermato anche in seguito, non verrà mai raggiunto.

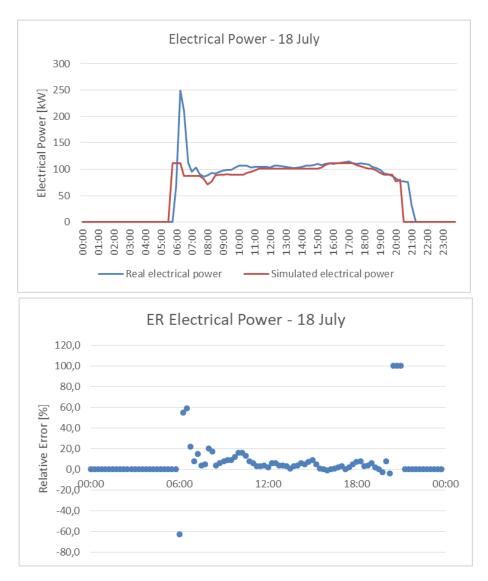


Figure 5.3-8: Evolution of simulated electrical power, in red, and real electrical power, in blue, for the 18th of July, on top, and scatter plot of relative error for the same day, on bottom.

Se invece consideriamo figure 5.3-9, riportante gli andamenti dei valori simulati e reali della potenza termica, in alto, ed dell'errore relativo, in basso, riferiti al medesimo giorno, possiamo notare che anche in questo caso il modello si adatta bene durante il funzionamento a regime, l'errore relativo è inferiore al 10%, ma deriva durante la fase transitoria dove tale errore raggiunge un valore del -600%. Anche in questo caso il valore dell'errore relativo non ha un vero e proprio significato in quanto il modello non è calibrato per tale fase. Inoltre, entrambe le fasi di accensione e spegnimento avvengono per Trnsys in maniera repentina ed immediata e non in modo graduale così come avviene nella realtà: nel momento in cui la portata assume valori diversi da zero, il valore della potenza termica, in questo caso, passa da 0 a 350 kW. Si ricorda in aggiunta che se i valori delle variabili

indipendenti cadono al di fuori del range inserito all'interno del file esterno, allora a tali variabili viene assegnato il valore minimo o massimo presente all'interno del file. Tale situazione coincide con quella avvenuta durante la fase transitoria per il giorno considerato, sia per quanto riguarda le portate che le temperature.

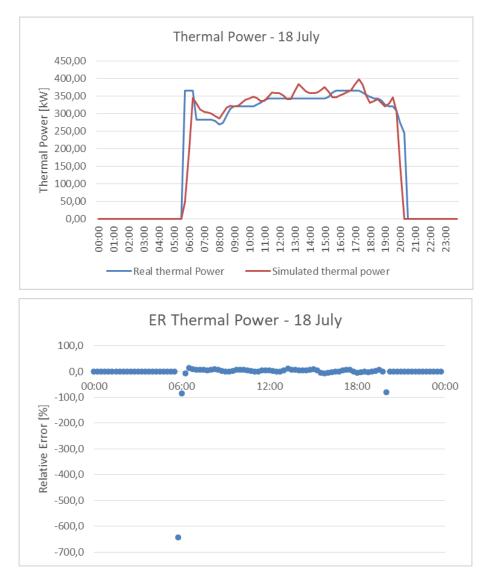


Figure 5.3-9: Evolution of simulated thermal power, in red, and real electrical power, in blue, for the 18th of July, on top, and scatter plot of relative error for the same day, on bottom.

Le stesse considerazioni possono effettuarsi considerando gli andamenti del coefficiente di performance reale e simulato e dell'errore relativo, mostrato in figure 5.3-10. In particolar modo il picco durante la fase transitoria è dovuto alla stretta dipendenza di tale coefficiente dai valori di potenza termica ed elettrica che porta quindi ad una deriva da parte del

modello. A ciò si aggiunge inoltre un non perfetto match temporale fra i valori reali e simulati, già visibile nelle precedenti immagini. Inoltre anche i valori reali del coefficiente di performance sono anomali in quanto solitamente per una pompa di calore essi assumono un valori compresi fra 3 e 4, mentre qui assumono valori nell'intorno di 3 e durante la fase transitoria scendono nell'intono di 1.5. Tali valori, che come si vedrà anche in seguito non rappresentano una eccezione, confermano la scarsa qualità dei dati termici durante il regime transitorio.

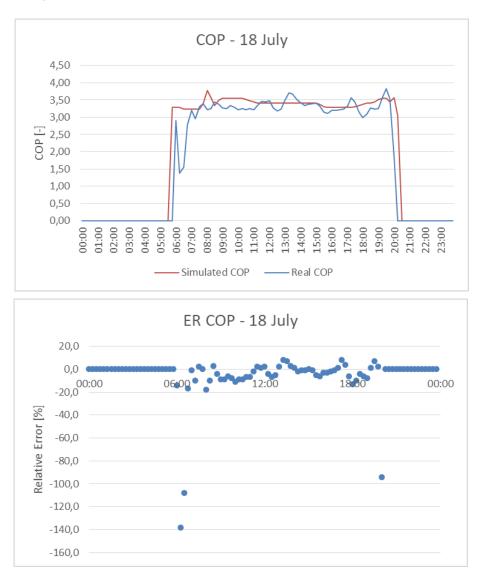


Figure 5.3-10: Evolution of simulated coefficient of performance, in red, and real coefficient of performance, in blue, for the 18th of July, on top, and scatter plot of relative error for the same day, on bottom.

Dopo aver verificato l'effettiva taratura del modello durante il funzionamento a regime per il tipico giorno lavorativo, il modello è stato applicato all'intero dataset. In particolar modo

le variabili considerate per la fase di validazione corrispondono agli output restituiti da trnsys, ad eccezione della portata di mandata all'ambiente in quanto non presente nel dataset iniziale.

Le immagini mostrate nella presente pagina ed in quella successiva mostrano rispettivamente gli andamenti della potenza elettrica, figure 5.3-11, e termica, figure 5.3-12, relative alla settimana compresa fra il 10 e 15 Luglio: tale settimana si può considerare rappresentativa dell'intero dataset. In generale possiamo notare un maggior scostamento del modello dalla realtà durante la fase transitoria a causa delle problematiche precedentemente descritte. Il valore evidenziato in arancio, circa -900%, è dovuto allo spegnimento repentino ed immediato e non graduale.

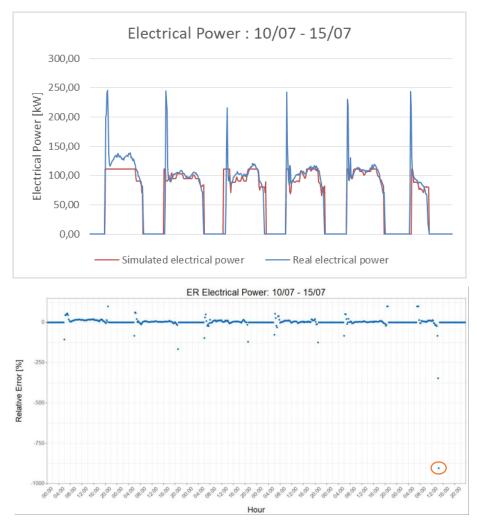


Figure 5.3-11: Evolution of simulated electrical power, in red, and real electrical power, in blue, for the week from the 10th till the 15th of July, on top, and scatter plot of relative error for the same week, on bottom.

Inoltre, se si considera il primo giorno di figure 5.3-11 e 5.3-12 si nota un discostamento fra i valori reali e simulati anche durante la fase a regime. Ciò è dovuto al fatto che durante tale giorno, il 10 luglio, i valori della Tel ricadono al di fuori del range di valori che è stato specificato nel file esterno con cui il modello è calibrato. All'interno del file infatti l'intervallo specificato per tale temperatura è  $8.7 \le Tel \le 9.9$  [°C], mentre il 10 Luglio tale temperatura è circa è maggiore o uguale a 10.1 °C. Allo stesso modo l'errore relativo superiore al -4000% (evidenziato in arancio) è dovuto ad un non perfetto match temporale delle variabili che può essere causato dall'interpolazione lineare effettuata durante la fase di pre-processamento.

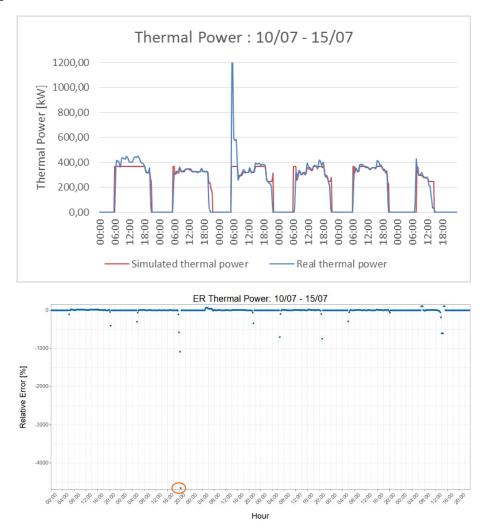


Figure 5.3-12: Evolution of simulated thermal power, in red, and real thermal power, in blue, for the week from the 10th till the 15th of July, on top, and scatter plot of relative error for the same week, on bottom.

Se consideriamo gli andamenti del coefficiente di performance relativi alla medesima settimana, riportati in figure 5.3-13, possiamo notare anche in questo caso un anomalia per

i valori reali che assumono valori superiori a 10 durante la fase transitoria. In corrispondenza di tale valore di picco si può inoltre notare che l'errore relativo assume un valore superiore a -4000 [%], evidenziato in arancio. Inoltre per tutti giorni appartenenti alla settimana considerata si può notare uno scostamento fra il reale ed il simulato durante la fase transitoria a causa della non calibrazione del modello durante tale regime di funzionamento.

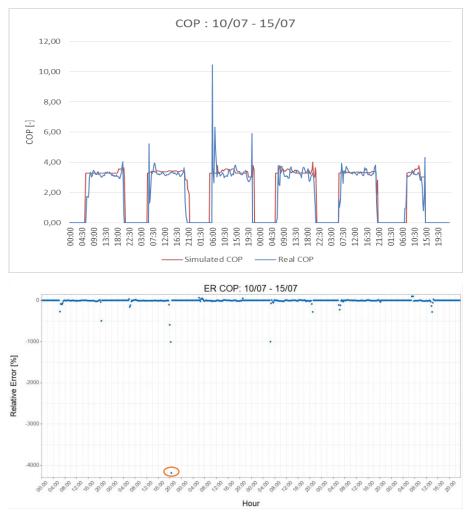


Figure 5.3-13: Evolution of simulated COP in red, and real COP, in blue, for the week from the 10th till the 15th of July, on top, and scatter plot of relative error for the same week, on bottom

Se sino ad ora abbiamo considerato solo le differenze di andamento fra i valori reali e simulati di entrambe le potenze e del coefficiente di performance per ciascun giorno, considerando quindi l'errore relativo espresso in percentuale, adesso esaminiamo i valori del MAPE e del MAE considerando l'intero dataset e di conseguenza tutte le variabili, ovvero:

- Potenza elettrica;
- Potenza termica;
- Coefficiente di prestazione, COP;
- Portata di re-immissione in falda, Qos;
- Temperatura di re-immissione in falda, Tos;
- Temperatura di mandata all'ambiente.

### Potenza elettrica

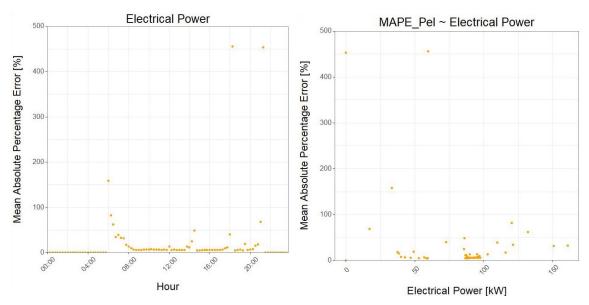


Figure 5.3-14: Scatter plot of MAPE of electrical power depending on daily hours, on the left, and depending on electrical power, on the right.

Le immagini sovrastanti, figure 5.3-14, mostrano rispettivamente, andando da sinistra verso destra, come il MAPE si distribuisce in funzione delle ore giornaliere e dei valori reali della potenza elettrica, mediati sul tempo. In particolar modo, se consideriamo l'immagine a sinistra, possiamo notare come lo scostamento fra il reale ed il simulato sia massimo durante la fase transitoria, in virtù delle problematiche precedentemente elencate (non calibrazione del modello durante la fase transitoria ed accensione/spegnimento immediato e repentino) e con picchi del MAPE superiori al 400%, e minimo durante il funzionamento a regime, con valori del MAPE nell'intorno del 10%. Tenendo presente i cluster ottenuti

per tale variabile, se osserviamo ora l'immagine a destra notiamo come i valori massimi si hanno per una potenza elettrica compresa fra 0 e 60 kW, condizione che corrisponde per l'appunto alle fasi di accensione e spegnimento, mentre sono minimi per una potenza elettrica pari a circa 90 kW, valore che solitamente tale variabile assume durante la fase di funzionamento a regime. Quanto appena detto trova riscontro anche se consideriamo figure 5.3-15, riportante la distribuzione del MAE in funzione delle ore giornaliere, a sinistra, e dei valori reali di potenza elettrica, mediati sul tempo. In particolar modo possiamo notare che i valori massimi si hanno anche in questo caso durante la fase transitoria dove la potenza elettrica assume valori nell'intorno di 0 o maggiori di 110 kW. Quest'ultimo valore corrisponde infatti al valore massimo di potenza elettrica che Trnsys riesce a simulare in funzione del file esterno costruito.

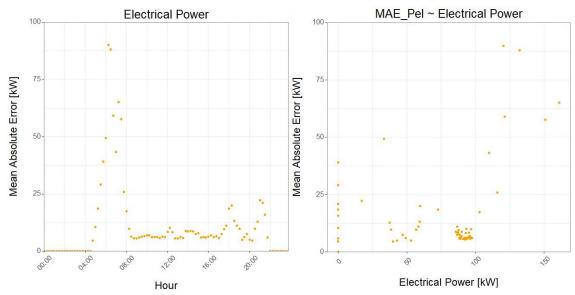


Figure 5.3-15: Scatter plot of MAE of electrical power depending on daily hours, on the left, and depending on electrical power, on the right.

Inoltre, se osserviamo figure 5.3-16 riportante il box plot ottenuto per entrambi gli indici di validazione utilizzati considerando il solo periodo di attività dell'impianto, tale per cui P\_elettrica ≠0, notiamo come i valori tipici della fase transitoria siano considerati outlier in virtù della loro bassa frequenza. I valori caratteristici di tali boxplot vengono invece riportati in table 5.3-5.

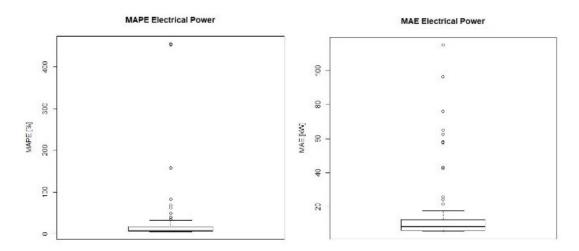


Figure 5.3-16: Box plot of MAPE, on the left, and MAE, on the right, obtained for electrical power.

Potenza elettrica						
Valore Primo Mediana Terzo Massim quartile						
MAPE [%]	5.0	6.0	7.0	17.1	32.8	
MAE [kW]	4.7	6.2	7.5	11.2	18.6	

Table 5.3-5: Summary table with indication of characteristic values for box plot shown in figure 5.3-16.

## Potenza termica

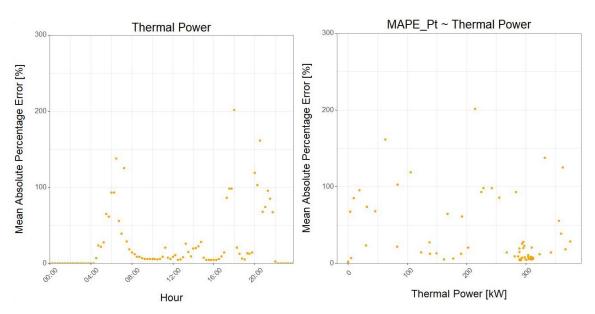


Figure 5.3-17: Scatter plot of MAPE of thermal power depending on daily hours, on the left, and depending on thermal power, on the right.

Le stesse considerazioni effettuate precedentemente per la potenza elettrica valgono anche per la potenza termica. Anche in questo caso infatti, come è visibile in figure 5.3-17 e 5.3-18, è possibile notare come entrambi gli indici utilizzati siano massimi durante la fase transitoria. In particolar modo il MAPE è massimo durante la fase di spegnimento, con un valore superiore al 150%, e minimo durante la fase di funzionamento a regime, con valori inferiori al 10%. Il valore massimo per il MAE si ha invece durante la fase di accensione, con un valore superiore a 200. Tuttavia, se consideriamo le distribuzioni del MAPE e del MAE in funzione dei valori reali di potenza termica, mediati sul tempo, notiamo come esse siano maggiormente variabili e non dipendano strettamente dai valori assunti dalla potenza termica.

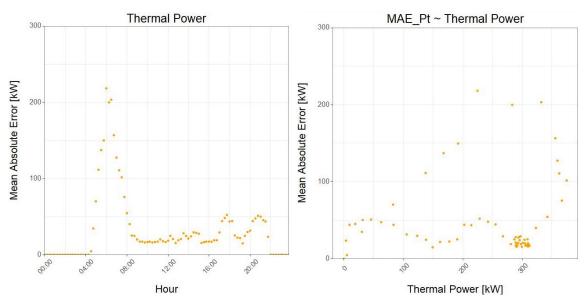


Figure 5.3-18: Development of MAE of thermal power during the day, on the left, and scatter plot of MAE depending on thermal power, on the right.

Anche in questo, se consideriamo i box plot ottenuti per entrambi tali indici considerando le sole ore lavorative dell'impianto, figure 5.3-19, notiamo che i valori caretteristici delle fasi transitorie vengono riconosciuti come outlier in virtù della loro bassa frequenza. Tuttavia gli errori sono in generale maggiori rispetto a quelli ottenuti per la potenza elettrica. Ciò è dovuto alle problematiche relative al contaimpulsi e di conseguenza alla cronometria lasca con il quale le variabili termiche, e quindi la potenza termica, sono state rilevate.

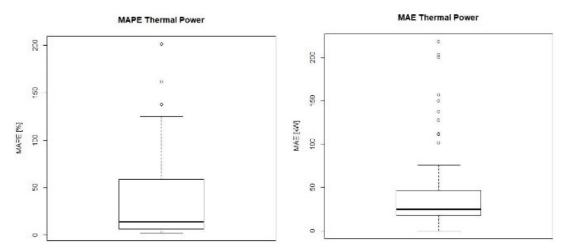


Figure 5.3-19: Box plot of MAPE, on the left, and MAE, on the right, obtained for thermal power.

## I valori rassuntivi di tali box plot vengono riportati in table 5.3-6

Potenza termica						
Valore Primo Mediana Terzo Masseminimo quartile Massemana quartile						
MAPE [%]	2.4	6.3	14.2	58.6	125.3	
MAE [kW]	0.1	17.7	25.1	46.4	75.8	

Table 5.3-6: Summary table with indication of characteristic values of box plot shown in figure 5.3-19.

# Temperatura di re-immissione in falda, Tos

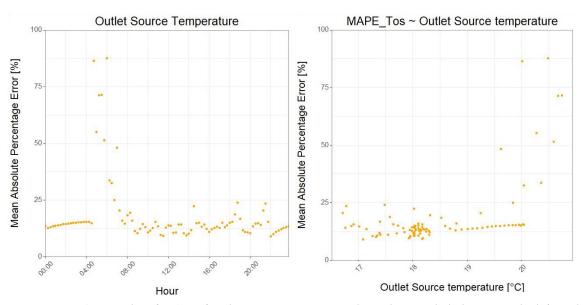


Figure 5.3-20: Scatter plot of MAPEof outlet source temperature depending on daily hours, on the left, and depending on outlet source temperature, on the right.

Ricordando le equazioni che Trnsys utilizza per il calcolo di tale variabile, riportate nel paragrafo 5.3.1, si può affermare che l'output di tale variabile, così come anche delle successive, è strettamente dipendente dagli output della potenza elettrica e termica. Pertanto, anche in questo caso, come viene mostrato dall'immagine sottostante figure 5.3-20, il modello deriva durante la fase transitoria, dove il MAPE raggiunge anche il 75% durante la fase di accensione, mentre si adatta durante il funzionamento a regime, dove il medesimo indice assume valori compresi tra il 10% ed il 20%.

Allo stesso modo, se osserviamo l'andamento dell'errore medio assoluto durante il giorno, figure 5.3-21, notiamo come il maggior scostamento fra i valori reali e simulati si ha durante la fase di accensione dell'impianto.

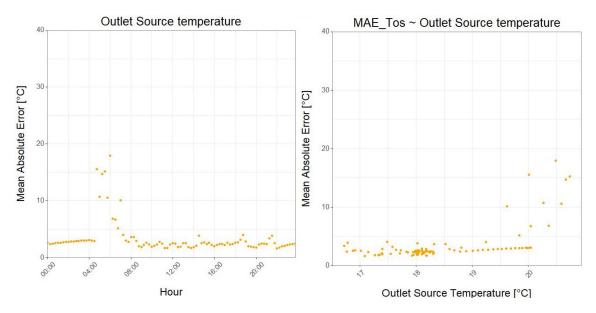


Figure 5.3-21: Scatter plot of MAE of outlet source temperature depending on daily hours, on the left, and depending on outlet source temperature, on the right.

Così come per le precedenti variabili analizzate, anche in questo caso se osserviamo i boxplot riportati nell'immagine sottostante, figure 5.3-22, notiamo che i valori caratteristici della fase transitoria vengono considerati anomali in virtù della loro bassa frequenza. I valori caratteristici di tali boxplot vengono riportati in table 5.3-7.

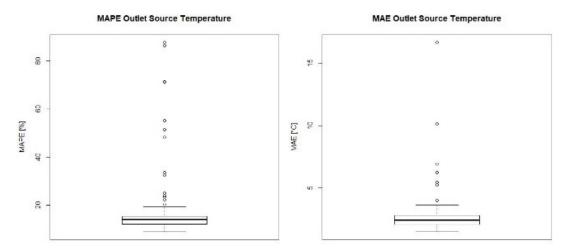


Figure 5.3-22: Box plot of MAPE, on the left, and MAE, on the right, obtained for outlet source temperature.

Temperatura di re-immissione in falda						
Valore Primo Mediana Terzo Massimo quartile						
MAPE [%]	9.0	12.3	14.1	15.5	19.4	
MAE [°C]	1.5	2.1	2.4	2.8	3.6	

Table 5.3-7: Summary table with indication of characteristic values of box plot shown in figure 5.3-22.

## Temperatura di mandata, Tol

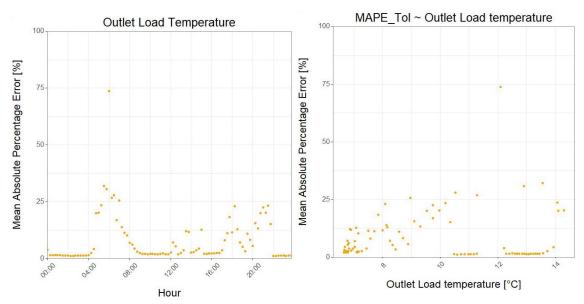


Figure 5.3-23: Scatter plot of MAPE of outlet load temperature depending on daily hours, on the left, and depending on outlet load temperature, on the right.

Così come per la temperatura di re-immissione in falda, anche per la temperatura di mandata gli indici di validazione utilizzati mostrano come il modello si discosti maggiormente durante la fase di accensione dell'impianto dove il MAPE assume un valore massimo pari al 75%, figure 5.3-23, ed il MAE assume un valore pari a 10, figure 5.3-24. Inoltre se consideriamo i grafici a dispersione di tali indici, in funzione della temperatura di mandata, notiamo che essi non dipendono strettamente dai valori assunti da tale variabile.

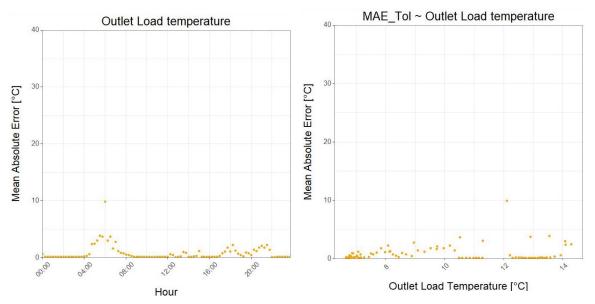


Figure 5.3-24: Scatter plot of MAE of outlet load temperature depending on daily hours, on the left, and depending on outlet load temperature, on the right.

Anche in questo caso i valori del MAPE e del MAE relativi alla fase transitoria vengono classificati come outlier nei box plot riportati in figure 5.3-25.

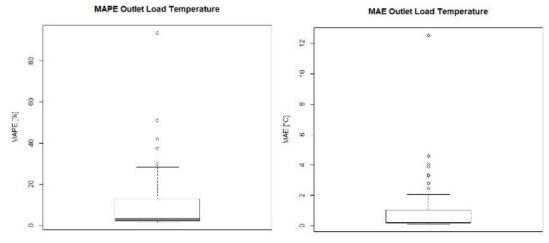


Figure 5.3-25: Box plot of MAPE, on the left, and MAE, on the right, obtained for outlet load temperature.

Temperatura di mandata								
	Valore	Primo	Mediana	Terzo	Massimo			
	minimo	quartile		quartile				
MAPE [%]	1.6	2.2	3.0	12.8	28.3			
MAE [°C]	0.11	0.15	0.2	1.0	2.0			

I valori caratteristici di tali boxplot sono visibili in table 5.3-8.

Table 5.3-8: Summary table with indication of characteristic values of box plot shown in figure 5.3-25.

### Portata di re-immissione in falda, Qos

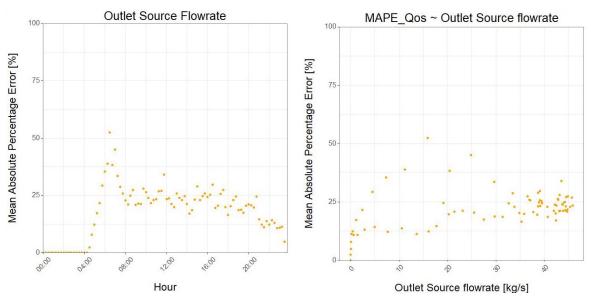


Figure 5.3-26: Scatter plot of MAPE of outlet source flowrate depending on daily hours, on the left, and depending on outlet source flowrate, on the right.

Figure 5.3-26 mostra l'andamento del MAPE per la portata di re-immissione in falda. In questo caso notiamo un valore di picco pari a circa il 50%, durante la fase di accensione, ed un andamento pressoché costante, nell'intorno del 25%, durante le ore lavorative giornaliere. Se invece consideriamo figure 5.3-27 e quindi il MAE possiamo osservare uno scostamento generale fra i valori simulati e reali nell'ordine di 10 kg/s. Dal grafico a dispersione relativo al MAE si nota inoltre come tale errore aumenti all'aumentare della portata.

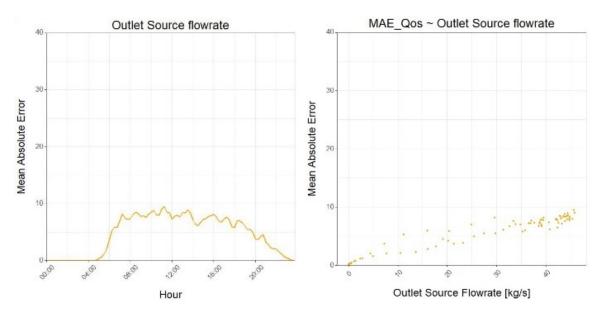


Figure 5.3-27: Scatter plot of MAE of outlet source flowrate depending on daily hours, on the left, and depending on outlet source flowrate, on the right.

In figure 5.3-28 vengono riportati i box plot ottenuti per il MAPE ed il MAE considerando le sole ore lavorative. In generale possiamo notare che i valori che tali variabili assumono durante la fase transitoria siano considerati anomali a causa della loro bassa frequenza. Tuttavia, tali valori, calcolati a partire dai valori simulati di potenza elettrica e termica, sono maggiori di quelli ottenuti per le temperature.

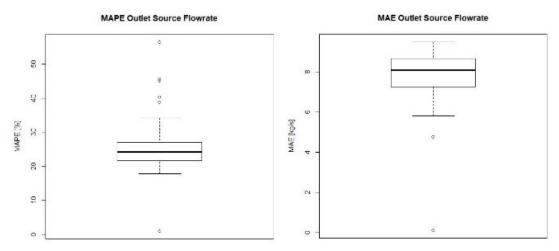


Figure 5.3-28 Box plot of MAPE, on the left, and MAE, on the right, obtained for outlet source flowrate.

La tabella sottostante, table 5.3-9, riporta i valori caratteristici dei box plot mostrati in figure 5.3-28.

Portata di re-immissione in falda								
	Valore minimo	Primo quartile	Mediana	Terzo quartile	Massimo			
MAPE [%]	17.9	21.8	24.3	27.0	34.2			
MAE [kg/s]	5.8	7.3	8.1	8.7	9.5			

Table 5.3-9: Summary table with indication of characteristic values of box plot shown in figure 5.3-28.

# Coefficiente di prestazione, COP

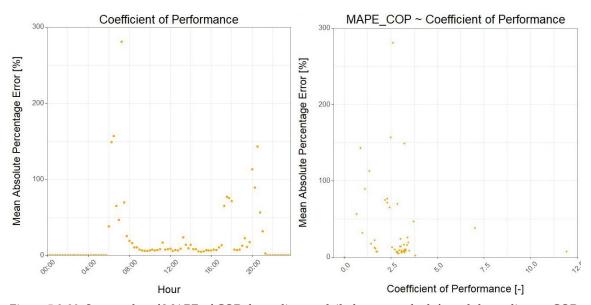


Figure 5.3-29: Scatter plot of MAPE of COP depending on daily hours, on the left, and depending on COP, on the right.

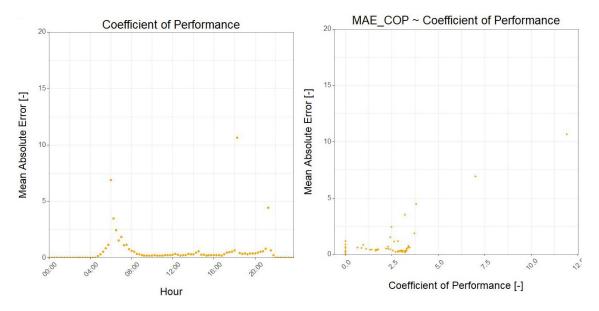


Figure 5.3-29: Scatter plot of MAE of COP depending on daily hours, on the left, and depending on COP, on the right.

Se consideriamo le immagini sovrastanti, figure 5.3-29 e figure 5.3-30, riportanti le distribuzioni del MAPE e del MAE relativo al coefficiente di prestazione in funzione delle ore giornaliere e dei valori reali di tale variabile, possiamo notare come, anche in questo caso, tali indici siano massimi durante la fase di funzionamento a carico parziale. Se si considerano i grafici a dispersione, relativi ad entrambi gli indici, si nota inoltre un'anomalia per quanto riguarda i valori reali che tale variabile assume. Infatti, il COP, che solitamente per impianti geotermici molto efficienti arriva ad assumere un valore massimo pari a 6, in questo caso ha un valore massimo pari a 12. Ricordando che tale coefficiente deriva dal rapporto fra la potenza termica e quella elettrica, possiamo dunque affermare che le tale anomalia deriva dalle problematiche relative al contaimpulsi e di conseguenza dalla cronometria lasca a cui le variabili termiche sono state rilevate. Inoltre se si osserva l'immagine sottostante, figure 5.3-31, si nota come anche in questo caso i valori assunti da entrambi gli indici di validazione durante la fase transitoria siano considerati anomali, a causa della loro bassa frequenza.

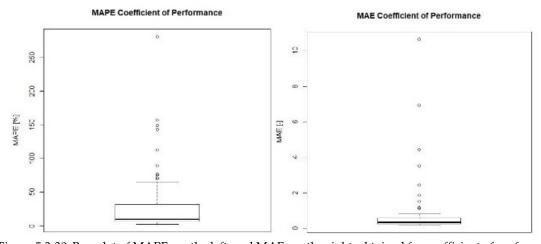


Figure 5.3-30: Box plot of MAPE, on the left, and MAE, on the right, obtained for coefficient of performance. La tabella sottostante, table 5.3-10, riporta i valori caratteristici dei box plot mostrati in figure 5.3-30.

Coefficiente di Prestazione					
	Valore minimo	Primo quartile	Mediana	Terzo quartile	Massimo
MAPE [%]	2.3	7	9.7	31.9	65.2
MAE [-]	0.2	0.24	0.3	0.6	0.8

Table 5.3-10: Summary table with indication of characteristic values for box plot shown in figure 5.3-31.

Infine, il modello è stato testato anche considerando il consumo energetico giornaliero. Osservando il grafico a dispersione sottostante, figure 5.3-32, relativo al confronto fra i dati reali e quelli simulati, si nota una buona correlazione fra le due variabili, dimostrata da un coefficiente di correlazione  $R^2$  = 0.9.

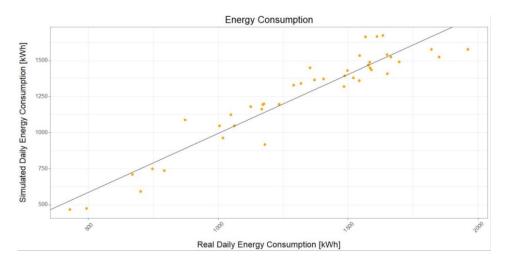


Figure 5.3-31: Regression plot of simulated and real daily energy consumption.

La buona aderenza del modello ai dati reali si nota inoltre in figure 5.3-33, all'interno del quale possiamo osservare come il MAPE assuma valori compresi fra 0 e 20 e la cui mediana corrisponde a 6. Ciò è dovuto al fatto che, come già detto in precedenza, la fase transitoria è molto breve rispetto a quella a regime.

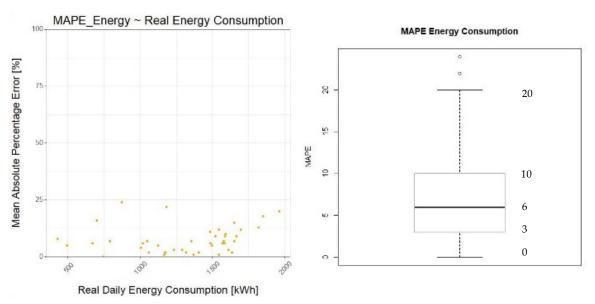


Figure 5.3-32: Scatter plot of MAPE depending on real daily energy consumption, on the left, and box plot of MAPE with indication of minimum, first quartile, median, third quartile and maximum, on the right.

## Conclusioni

Il presente lavoro di tesi fornisce una metodologia in grado di caratterizzare dal punto vista di vista operativo un impianto di raffrescamento e al contempo creare un modello simulativo, calibrato sui dati storici, che ne rispecchi il comportamento.

In particolar modo, la fase di caratterizzazione si compone a sua volta di tre sotto-fasi, ovvero: pre-processamento, individuazione del profilo di carico e classificazione di quest'ultimo.

La fase di pre-processamento è di fondamentale importanza in quanto i dati non trattati possono portare ad una riduzione di qualità dell'intero processo di analisi poiché talvolta caratterizzati dalla presenza di valori mancanti, incoerenti o anomali. Durante tale fase i dati sono quindi stati consolidati in modo da ottenere uniformità dal punto di vista dimensionale e temporale. Poiché la variabile principale considerata consiste con la potenza elettrica (rilevata al minuto e mediata sui 15 minuti), in quanto essa è quella che meglio riesce a descrivere le modalità operative (orari di accensione e spegnimento nonché le fasi di picco ed eventuali comportamenti infrequenti), le serie temporali delle restanti variabili sono state ricostruite attraverso interpolazione lineare. A causa dell'errato settaggio del contaimpulsi, le variabili termiche relative alle portate ed alle temperature del fluido in ingresso alla CF1 da ambo i lati, geotermico ed ambiente, e di conseguenza la potenza termica sono state rilevate con una cronometria lasca. I profili infrequenti sono invece stati individuati attraverso clustering gerarchico considerando il metodo del single linkage ed i valori di potenza elettrica normalizzati su valore massimo giornaliero, prediligendo così variazioni di forma e non di intensità. Tali profili, 14 in tutto, sono stati rimossi dal dataset in modo da non compromettere la qualità delle successive analisi di caratterizzazione dell'impianto nonché la fase di costruzione del modello simulativo calibrato sui dati storici.

Anche la fase di individuazione del profilo di carico ha visto l'utilizzo delle tecniche di clustering gerarchico seppur con metodo diverso. In questo caso infatti è stato considerato

il metodo di Ward in quanto esso massimizza l'omogeneità intra-cluster. Il numero di cluster per ciascuna variabile è inoltre stato determinato in funzione dell'elbow method. La fase di classificazione, effettuata attraverso l'applicazione dell'algoritmo CART ha fornito delucidazioni circa i cluster ottenuti dalla fase precedente. In particolar modo i cluster di ciascuna variabile sono stati classificati in funzione di variabili esterne (tipo di giorno, mese di appartenenza e temperature esterne giornaliere) ed operative (valori medi giornalieri e cluster di tutte le variabili considerate). Tale fase ha quindi permesso di ricavare i costrutti tipologici relativi agli orari di accensione e spegnimento ed allo stesso modo individuare i range che le variabili considerate assumono durante il periodo di attività, e non, dell'impianto.

Noti tali range ed in generale le dinamiche impiantistiche, si è poi proceduto alla fase di costruzione di un modello simulativo calibrato sui dati storici attraverso l'utilizzo del software Trnsys. Il cuore di tale modello coincide con una mappa delle performance che prevede l'inserimento dei valori normalizzati di potenza elettrica e termica in funzione delle portate e delle temperature caratteristiche del funzionamento dell'impianto. L'ipotesi su cui si basa l'individuazione dei valori caratteristici è consistita nel supporre che tali valori corrispondano a quelli più frequenti e pertanto alla moda di ciascuna variabile. Inoltre, l'ordine con cui tali variabili sono state analizzate per l'identificazione dei valori caratteristici è in funzione della variabilità dei valori che esse assumono durante il funzionamento, informazione desunta sempre considerando i cluster. La validazione del modello è stata effettuata considerando come indici statistici il MAPE ed il MAE. Da tale fase è emersa una buona aderenza del modello durante il funzionamento a regime ma non durante quello a carico parziale. Ciò è dovuto alla cronometrica lasca a cui le variabili termiche e soprattutto la potenza termica sono state rilevate. I dati di input inseriti per la calibrazione del modello sono tipici del funzionamento a regime e non di quello a carico parziale, caratterizzato da condizioni operative maggiormente variabili rispetto al funzionamento a regime e da un intervallo temporale molto ridotto, nel nostro caso circa il 10% delle ore di funzionamento giornaliere. Pertanto, per riuscire a calibrare il modello durante tale fase occorrerebbe una maggiore densità dei dati. Se infatti anche le variabili termiche e la potenza termica fossero state rilevate con un timestep di 1 minuto il modello non solo si sarebbe adattato meglio durante la fase di funzionamento a regime ma soprattutto sarebbe stato in grado di simulare efficacemente anche la fase transitoria. La scarsa qualità dei dati termici, temporalmente parlando, si evince inoltre dai valori anomali del coefficiente di performance che arriva ad assumere valori pari ad 1 o addirittura a 12. Come ultima fase il modello è stato inoltre testato considerando il consumo energetico su base giornaliera. L'analisi della regressione effettuata considerando il confronto fra i valori reali e quelli simulati mostra un coefficiente di determinazione pari a 0.9 ed un MAPE inferiore al 10% per il 75% dei campioni.

In definitiva, possiamo affermare che tale metodologia non solo è in grado di cogliere i costrutti tipologici relativi agli orari di accensione e spegnimento, le modalità operative, le relazioni esistenti fra le variabili ed i range che tale variabili assumono durante il normale funzionamento, ma permette inoltre di costruire un modello simulativo che si adatti al suo reale comportamento tenendo quindi conto del reale stato di efficienza impiantistico. Numerosi studi in letteratura mostrano inoltre come tali modelli calibrati sui dati storici possano essere utilizzati al fine di individuare eventuali problematiche operative, proporre strategie di retrofit di risparmio energetico e valutare nuove strategie di gestione.

## Allegato I

Il presente allegato coincide con lo script elaborato su R al fine di caratterizzare l'impianto attraverso l'utilizzo delle tecniche di clustering e classificazione. Esso si può considerare suddiviso in 4 sezioni ovvero: creazione dei dataset, individuazione dei profili anomali o infrequenti, individuazione dei profili di carico e loro caratterizzazione.

```
#installo librerie
library(rpart)
library(tidyr)
library(dplyr)
library(plyr)
library(partykit)
library(ggplot2)
library(scales)
library(caret)
library(e1071)
library(zoo)
library(lubridate)
library(stats)
library(RColorBrewer)
library(wesanderson)
library(car)
library(NbClust)
library(factoextra)
#imposto palette colori
pals<-wes_palette("Darjeeling1",10,type ="continuous")</pre>
```

```
#leggo database potenza elettrica
df<-read.csv("elettrico_ok.csv",sep=";",dec=",")
df <- df %>%
mutate(
  date_time = as.POSIXct(strptime(paste(paste(Year, Month, Day, sep = "/"), paste(Hour,
Minute*100, "00", sep = ":")), format = "%Y/%m/%d %H:%M:%S", tz="Etc/GMT+12")),
  date = as.Date((date_time), format = "%Y/%m/%d", tz="Etc/GMT+12"),
  time = strftime(date_time, format = "%H:%M:%S", tz="Etc/GMT+12")
 )%>%
 select(date,time,Day_Type,Month,POWER)
#leggo database temperature esterne
temp<-read.csv("T_est.csv",sep=";",dec=",")
temp$date<-as.Date(with(temp,paste(Year,Month,Day,sep = "/")),"%Y/%m/%d")
temp$Year<-NULL
temp$Day<-NULL
term<- read.csv("pred_term.csv", sep = ";", dec = ",") #leggo database variabili termiche
term <- term %>%
 mutate(
  date_time = as.POSIXct(strptime(paste(paste(Year, Month, Day, sep = "/"), paste(Hour,
Minute, "00", sep = ":")), format = "%Y/%m/%d %H:%M:%S", tz="Etc/GMT+12")),
  date = as.Date((date_time), format = "%Y/%m/%d", tz="Etc/GMT+12"),
  time = strftime(date_time, format = "%H:%M:%S", tz="Etc/GMT+12")
 )%>%
 select(date,time,T_es,T_el,Q_el,P_term)
#leggo database portata di emungimento
```

```
Q_es<-read.csv("Q_es.csv",sep=";",dec=",")
Q es <- Q es %>%
 mutate(
  date_time = as.POSIXct(strptime(paste(paste(Year, Month, Day, sep = "/"), paste(Hour,
Minute, "00", sep = ":")), format = "%Y/%m/%d %H:%M:%S", tz="Etc/GMT+12")),
  date = as.Date((date_time), format = "%Y/%m/%d", tz="Etc/GMT+12"),
  time = strftime(date_time, format = "%H:%M:%S", tz="Etc/GMT+12")
 )%>%
 select(date,time,Q_es)
# creo dataframe date e orari per interpolazione
date<-seq(as.POSIXct("2017-06-16 00:00:00",tz="Etc/GMT+12"),as.POSIXct("2017-09-19
23:45:00",tz="Etc/GMT+12"),by=5*60)
dat<-
data.frame(date=strftime(date,format="%Y/%m/%d",tz="Etc/GMT+12"),time=strftime(date
,format = "%H:%M:%S",tz="Etc/GMT+12"))
# interpolo tutte le variabili termiche
df_qes<-merge(Q_es,dat,all=TRUE)
df_qes<-df_qes %>% mutate(Q_es=na.approx(Q_es,na.rm=FALSE))
df_qes$time<-strptime(df_qes$time, "%H:%M:%S")
df_qes<-subset(df_qes, format(df_qes$time, "%M") %in% c('00', '15', '30', '45'))
df_qes$Time <- strftime(as.POSIXct(df_qes$time,format = "%H:%M:%S"),format =
"%H:%M:%S")
df_qes$time<-NULL
colnames(df_qes)[colnames(df_qes)=='Time']<-'time'
df_term<-merge(term,dat,all=TRUE)</pre>
df_term<-df_term%>% mutate(T_es=na.approx(T_es,na.rm=FALSE),
              T_el=na.approx(T_el,na.rm=FALSE),
```

```
Q_{el}=na.approx(Q_{el},na.rm = FALSE),
              P_term=na.approx(P_term,na.rm = FALSE))
df_term$time<-strptime(df_term$time, "%H:%M:%S")
df_term<-subset(df_term, format(df_term$time, "%M") %in% c('00', '15', '30', '45'))
df_term$Time<-format(as.POSIXct(df_term$time ,format = "%H:%M:%S"))</pre>
df_term$Time<-strftime(df_term$Time,format = "%H:%M:%S")</pre>
df_term$time<-NULL
colnames(df_term)[colnames(df_term)=='Time']<-'time'
df_term<-merge.data.frame(df_term,df_qes)</pre>
df<-merge.data.frame(df,df_term)</pre>
#calcolo valori medi giornalieri delle variabili
P_el<-select(df,date,POWER)
P_el<-ddply(P_el,c("date"),summarise,Pel_mean=mean(POWER))
P_el$date<-as.Date(P_el$date)
df<-merge.data.frame(df,P_el)
term<- ddply(term, c("date"), summarise, Tes_mean = mean(T_es),
       Tel_mean=mean(T_el),Qel_mean=mean(Q_el),Pt_mean=mean(P_term))
df<-merge.data.frame(df,term)</pre>
Q es<-ddply(Q es,c("date"),summarise,Qes mean=mean(Q es))
Q_es$date<-as.Date(Q_es$date)
df<-merge.data.frame(df,Q_es)
df<-merge.data.frame(df,temp) #il dataset completo è pronto
rm(df_qes,df_term,Q_es,P_el,temp,term) #rimuovo df inutili
```

```
#creo df per ciascuna variabile
df_pel<-select(df,-c(T_es,T_el,Q_es,Q_el,P_term,Pel_mean))
df_pt<-select(df,-c(T_es,T_el,Q_es,Q_el,POWER,Pt_mean))
df_ts<-select(df,-c(T_el,Q_es,Q_el,POWER,P_term,Tes_mean))
df_tl<-select(df,-c(T_es,Q_es,Q_el,POWER,P_term,Tel_mean))
df qs<-select(df,-c(T_el,T_es,Q_el,POWER,P_term,Qes_mean))</pre>
df_ql<-select(df,-c(T_el,T_es,Q_es,POWER,P_term,Qel_mean))
#esamino portenza elettrica normalizzata su max giornaliero e non normalizzata
df1_pel<-spread(df_pel,time,POWER)
normal <- function(x) { x / max(x) }
df1_pel[12:107]<-adply(df1_pel[12:107],1,normal)
df1_pel[is.na(df1_pel)]<-0
dfpel_norm<-gather(df1_pel,time,POWER,"00:00:00":"23:45:00")
dist_pel_all<-dist(df1_pel[12:107])
hcl_pel_all<-hclust(dist_pel_all,method = "ward.D2")</pre>
dev.new()
plot(hcl_pel_all)
dev.new()
fviz_nbclust(df1_pel[12:107], hcut, method = "wss") +
geom_vline(xintercept = 6, linetype = 2)
df1 pel$clust Pel<-cutree(hcl pel all,6)
dfpel_norm<-merge.data.frame(dfpel_norm,df1_pel[c("date","clust_Pel")])
df_pel<-merge.data.frame(df_pel,df1_pel[c("date","clust_Pel")])
centr_pel_norm_all<-
ddply(dfpel_norm,c("clust_Pel","time"),summarise,POWER=mean(POWER))
centr_pel_all<-ddply(df_pel,c("clust_Pel","time"),summarise,POWER=mean(POWER))
clplot_Pel_norm_all <- ggplot()+</pre>
```

```
geom_line(data= dfpel_norm, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
 geom_line(data= centr_pel_norm_all, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color = as.factor(clust_Pel)), size=1)+
 scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y_continuous(limits = c(0,ceiling(max(dfpel_norm$POWER)/1.2)*1.2),
expand=c(0,0)+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Normalized Power", color="Cluster")
dev.new()
plot(clplot_Pel_norm_all)
```

```
clplot_Pel_all <- ggplot()+
 geom line(data= df pel, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
 geom_line(data= centr_pel_all, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color =as.factor(clust Pel)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
scale_y = c(0,ceiling(max(df_pel$POWER)/100)*100), expand=c(0,0))+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Electrical Power [kW]", color="Cluster")
dev.new()
```

```
plot(clplot_Pel_all)
df_pel2<-df_pel #creo copie da usare dopo aver rimosso outlier
dfpel_norm2<-dfpel_norm
#cluster panoramico sulle variabili termiche
df1_pt<-spread(df_pt,time,P_term)
dist_pt_all<-dist(df1_pt[12:107])
hcl_pt_all<-hclust(dist_pt_all,method = "ward.D2")</pre>
dev.new()
plot(hcl_pt_all)
dev.new()
fviz_nbclust(df1_pt[12:107], hcut, method = "wss") +
geom_vline(xintercept = 4, linetype = 2)
df1_pt$clust_Pt<-cutree(hcl_pt_all,4)
df_pt<-merge.data.frame(df_pt,df1_pt[c("date","clust_Pt")])</pre>
centr_pt_all<-ddply(df_pt,c("clust_Pt","time"),summarise,P_term=mean(P_term))</pre>
clplot_pt_all <- ggplot()+
 geom_line(data= df_pt, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = P_term, group = date), size=0.65, color="grey50")+
 geom_line(data= centr_pt_all, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = P_term, color =as.factor(clust_Pt)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date breaks("4 hour"))+
 scale_y\_continuous(limits = c(-180,ceiling(max(df_pt$P_term)/100)*100))+
 theme_bw()+
 facet_wrap(~clust_Pt)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
```

```
legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
          axis.title.x = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.title.y = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.text.x = element_text(size=12, angle=45, vjust = .5),
          axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour",y="Thermal Power [kW]", color="Cluster")
dev.new()
plot(clplot_pt_all)
df1_qs<-spread(df_qs,time,Q_es)
dist_qs_all < -dist(df1_qs[12:107])
hcl_qs_all<-hclust(dist_qs_all,method = "ward.D2")</pre>
dev.new()
plot(hcl_qs_all)
dev.new()
fviz_nbclust(df1_qs[12:107], hcut, method = "wss") +
geom_vline(xintercept = 4, linetype = 2)
df1_qs$clust_Qes<-cutree(hcl_qs_all,4)
df_qs<-merge.data.frame(df_qs,df1_qs[c("date","clust_Qes")])
```

```
centr_qs_all<-ddply(df_qs,c("clust_Qes","time"),summarise,Q_es=mean(Q_es))
clplot_qs_all <- ggplot()+
 geom_line(data= df_qs, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_es, group = date), size=0.65, color="grey50")+
 geom line(data=centr qs all, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_es, color =as.factor(clust_Qes)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y\_continuous(limits = c(0,ceiling(max(df_qs$Q_es)/0.12)*0.12), expand=c(0,0),name
= expression(Entering ~~ Source ~~ Flowrate ~~ group("[",m^3/s, "]")))+
 theme bw()+
 facet_wrap(~clust_Qes)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", color="Cluster")
```

```
dev.new()
plot(clplot_qs_all)
df1_ql<-spread(df_ql,time,Q_el)
dist_ql_all < -dist(df1_ql[12:107])
hcl ql_all<-hclust(dist_ql_all,method = "ward.D2")</pre>
dev.new()
plot(hcl_ql_all)
dev.new()
fviz_nbclust(df1_ql[12:107], hcut, method = "wss") +
geom_vline(xintercept = 5, linetype = 2)
df1_ql$clust_Qel<-cutree(hcl_ql_all,5)
df_ql<-merge.data.frame(df_ql,df1_ql[c("date","clust_Qel")])
centr\_ql\_all <-ddply (df\_ql, c("clust\_Qel", "time"), summarise, Q\_el=mean(Q\_el))
clplot_ql_all <- ggplot()+
 geom_line(data= df_ql, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_el, group = date), size=0.65, color="grey50")+
 geom_line(data= centr_ql_all, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_el, color =as.factor(clust_Qel)), size=1)+
 scale x datetime(expand=c(0,0), labels = date format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y = c(0,ceiling(max(df_qlQ_el)/0.05)*0.05), expand=c(0,0),name
= expression(Entering ~~ Load ~~ Flowrate ~~ group("[",m^3/s, "]")))+
 theme_bw()+
 facet_wrap(~clust_Qel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
```

```
legend.box = "vertical",
         legend.title = element text(size = 28),
          axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.title.y = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.text.x = element_text(size=12, angle=45, vjust = .5),
          axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", color="Cluster")
dev.new()
plot(clplot_ql_all)
df1_ts<-spread(df_ts,time,T_es)
dist_ts_all<-dist(df1_ts[12:107])
hcl_ts_all<-hclust(dist_ts_all,method = "ward.D2")
dev.new()
plot(hcl_ts_all)
dev.new()
fviz_nbclust(df1_ts[12:107], hcut, method = "wss") +
geom_vline(xintercept = 4, linetype = 2)
df1_ts$clust_Tes<-cutree(hcl_ts_all,4)
df_ts<-merge.data.frame(df_ts,df1_ts[c("date","clust_Tes")])
centr_ts_all<-ddply(df_ts,c("clust_Tes","time"),summarise,T_es=mean(T_es))
clplot_ts_all <- ggplot()+
```

```
geom_line(data= df_ts, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_es, group = date), size=0.65, color="grey50")+
geom_line(data= centr_ts_all, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_es, color =as.factor(clust_Tes)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
scale_y\_continuous(limits = c(10,ceiling(max(df_ts$T_es)/27)*27))+
 theme_bw()+
 facet_wrap(~clust_Tes)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour",y="Entering Source Flowrate [°C]", color="Cluster")
dev.new()
plot(clplot_ts_all)
```

```
df1 tl<-spread(df tl,time,T el)
dist_tl_all<-dist(df1_tl[12:107])
hcl_tl_all<-hclust(dist_tl_all,method = "ward.D2")
dev.new()
plot(hcl_tl_all)
dev.new()
fviz_nbclust(df1_tl[12:107], hcut, method = "wss") +
geom_vline(xintercept = 4, linetype = 2)
df1_tl$clust_Tel<-cutree(hcl_tl_all,4)
df_tl<-merge.data.frame(df_tl,df1_tl[c("date","clust_Tel")])
centr_tl_all<-ddply(df_tl,c("clust_Tel","time"),summarise,T_el=mean(T_el))
clplot_tl_all <- ggplot()+
 geom_line(data= df_tl, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_el, group = date), size=0.65, color="grey50")+
 geom_line(data= centr_tl_all, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_el, color =as.factor(clust_Tel)), size=1)+
 scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date breaks("4 hour"))+
 scale_y\_continuous(limits = c(0,ceiling(max(df_tl$T_el)/30)*30), expand=c(0,0))+
 theme bw()+
 facet_wrap(~clust_Tel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
```

```
axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.title.y = element text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.text.x = element_text(size=12, angle=45, vjust = .5),
          axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Entering Load Temperature [°C]", color="Cluster")
dev.new()
plot(clplot_tl_all)
#separo outlier ottenuti per la p_elettrica
out_pel<-subset(df1_pel,clust_Pel==3|clust_Pel==4)
out_pel_norm<-subset(df1_pel,clust_Pel==3|clust_Pel==4)
#esamino cluster1 della potenza elettrica
clust1<-subset(df1_pel,clust_Pel==1)</pre>
clust1$clust_Pel<-NULL
dist clust1<-dist(clust1[12:107])
hcl<-hclust(dist_clust1,method="single")</pre>
dev.new()
plot(hcl)
dev.new()
fviz nbclust(clust1[12:107], hcut, method = "wss") +
 geom_vline(xintercept = 4, linetype = 2)
```

```
clust1$clust_Pel<-cutree(hcl,4)</pre>
dfpel norm$clust Pel<-NULL
df_pel$clust_Pel<-NULL
dfpel_norm<-merge.data.frame(dfpel_norm,clust1[c("date","clust_Pel")])
df_pel<-merge.data.frame(df_pel,clust1[c("date","clust_Pel")])
centr1 Pelnorm<-
ddply(dfpel_norm,c("clust_Pel","time"),summarise,POWER=mean(POWER))
centr1_Pel<-ddply(df_pel,c("clust_Pel","time"),summarise,POWER=mean(POWER))
clplot_Pel1_norm <- ggplot()+
geom_line(data= dfpel_norm, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
 geom_line(data= centr1_Pelnorm, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color =as.factor(clust_Pel)), size=1)+
 scale x datetime(expand=c(0,0), labels = date format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y_continuous(limits = c(0,ceiling(max(dfpel_norm$POWER)/1.2)*1.2),
expand=c(0,0)+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
```

```
panel.background = element_rect(fill="white"),
         panel.border = element rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Normalized Power", color="Cluster")
dev.new()
plot(clplot_Pel1_norm)
clplot_Pel1 <- ggplot()+
geom_line(data= df_pel, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
 geom_line(data= centr1_Pel, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color =as.factor(clust_Pel)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
scale_y\_continuous(limits = c(0,ceiling(max(df\_pel$POWER)/100)*100), expand=c(0,0))+
 theme_bw()+
 facet wrap(~clust Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
```

```
axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Electrical Power [kW]", color="Cluster")
dev.new()
plot(clplot_Pel1)
out1<-subset(clust1,clust_Pel==2)
#esamino cluster2 potenza elettrica
clust2<-subset(df1_pel,clust_Pel==2)</pre>
dist_clust2<-dist(clust2[12:107])
hcl_clust2<-hclust(dist_clust2,method="single")</pre>
dev.new()
plot(hcl_clust2)
dev.new()
fviz_nbclust(clust2[12:107], hcut, method = "wss") +
geom_vline(xintercept = 6, linetype = 2)
clust2$clust Pel<-NULL
clust2$clust_Pel<-cutree(hcl_clust2,6)</pre>
dfpel_norm<-dfpel_norm2
df_pel<-df_pel2
dfpel_norm$clust_Pel<-NULL
df_pel$clust_Pel<-NULL
dfpel_norm<-merge.data.frame(dfpel_norm,clust2[c("date","clust_Pel")])
```

```
df_pel<-merge.data.frame(df_pel,clust2[c("date","clust_Pel")])
centr2 Pelnorm<-
ddply(dfpel_norm,c("clust_Pel","time"),summarise,POWER=mean(POWER))
centr2_Pel<-ddply(df_pel,c("clust_Pel","time"),summarise,POWER=mean(POWER))
clplot_Pel2_norm <- ggplot()+
 geom_line(data= dfpel_norm, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
geom_line(data= centr2_Pelnorm, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color =as.factor(clust_Pel)), size=1)+
 scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date breaks("4 hour"))+
 scale_y_continuous(limits = c(0,ceiling(max(dfpel_norm$POWER)/1.2)*1.2),
expand=c(0,0)+
 theme_bw()+
 facet wrap(~clust Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
```

```
)+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Normalized Power", color="Cluster")
dev.new()
plot(clplot_Pel2_norm)
clplot_Pel2 <- ggplot()+
 geom line(data= df pel, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
geom_line(data= centr2_Pel, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color =as.factor(clust_Pel)), size=1)+
 scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_v_continuous(limits = c(0,ceiling(max(df_pel$POWER)/100)*100), expand=c(0,0))+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
```

```
plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Electrical Power [kW]", color="Cluster")
dev.new()
plot(clplot_Pel2)
out2<-subset(clust2,clust_Pel<=4)
#esamino cluster 5 potenza elettrica
clust5<-subset(df1_pel,clust_Pel==5)</pre>
clust5$clust_Pel<-NULL
dist_clust5<-dist(clust5[12:107])
hcl_clust5<-hclust(dist_clust5,method="single")
dev.new()
plot(hcl_clust5)
fviz_nbclust(clust5[12:107], hcut, method = "wss") +
geom_vline(xintercept = 2, linetype = 2)
clust5$clust_Pel<-cutree(hcl_clust5,2)</pre>
dfpel_norm<-dfpel_norm2
df_pel<-df_pel2
dfpel_norm$clust_Pel<-NULL
df pel$clust Pel<-NULL
dfpel_norm<-merge.data.frame(dfpel_norm,clust5[c("date","clust_Pel")])
df_pel<-merge.data.frame(df_pel,clust5[c("date","clust_Pel")])
centr5_Pelnorm<-
ddply(dfpel_norm,c("clust_Pel","time"),summarise,POWER=mean(POWER))
centr5_Pel<-ddply(df_pel,c("clust_Pel","time"),summarise,POWER=mean(POWER))
clplot_Pel5_norm <- ggplot()+
```

```
geom_line(data= dfpel_norm, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
geom_line(data= centr5_Pelnorm, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color =as.factor(clust_Pel)), size=1)+
 scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y_continuous(limits = c(0,ceiling(max(dfpel_norm$POWER)/1.2)*1.2),
expand=c(0,0)+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Normalized Power", color="Cluster")
dev.new()
plot(clplot_Pel5_norm)
```

```
clplot_Pel5 <- ggplot()+
 geom line(data= df pel, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
 geom_line(data= centr5_Pel, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color =as.factor(clust Pel)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
scale_y = c(0,ceiling(max(df_pel$POWER)/100)*100), expand=c(0,0))+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Electrical Power [kW]", color="Cluster")
dev.new()
```

```
plot(clplot_Pel5)
out5<-subset(clust5,clust Pel==2)
#unisco outlier in unico df
outlier<-rbind(out_pel,out1,out2,out5)</pre>
dfout_pel<-gather(outlier,time,POWER,"00:00:00":"23:45:00")
dfout_pel$clust_Pel<-NULL
df_pel$clust_Pel<-NULL
df1out_pel<-spread(dfout_pel,time,POWER)</pre>
dist_out_pel<-dist(df1out_pel[12:107])
hcl_out_pel<-hclust(dist_out_pel,method = "ward.D2") #ottengo cluster definitivi outlier</pre>
dev.new()
plot(hcl_out_pel)
dev.new()
fviz_nbclust(df1_tl[12:107], hcut, method = "wss") +
geom_vline(xintercept = 4, linetype = 2)
df1out_pel$clust_Pel<-cutree(hcl_out_pel,4)</pre>
dfout_pel_norm<-merge.data.frame(dfout_pel,df1out_pel[c("date","clust_Pel")])
dfout_pel_norm2<-dfout_pel_norm
dfout_pel_norm2$POWER<-NULL
dfout_pel_nn<-merge.data.frame(df_pel2,dfout_pel_norm2[c("date","time")])
rm(dfout_pel_norm2)
centr out Pelnorm<-
ddply(dfout_pel_norm,c("clust_Pel","time"),summarise,POWER=mean(POWER))
centr_out_Pel<-
ddply(dfout_pel_nn,c("clust_Pel","time"),summarise,POWER=mean(POWER))
clplot_Pelout_norm <- ggplot()+</pre>
geom_line(data= dfout_pel_norm,aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
```

```
geom_line(data= centr_out_Pelnorm, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color =as.factor(clust_Pel)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date breaks("4 hour"))+
 scale_y_continuous(limits = c(0,ceiling(max(dfout_pel_norm$POWER)/1.2)*1.2),
expand=c(0,0)+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Normalized Power", color="Cluster")
dev.new()
plot(clplot_Pelout_norm)
clplot_Pelout <- ggplot()+
```

```
geom_line(data= dfout_pel_nn, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
geom_line(data= centr_out_Pel, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color = as.factor(clust_Pel)), size=1)+
 scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y_continuous(limits = c(0,ceiling(max(dfout_pel_nn$POWER)/100)*100),
expand=c(0,0)+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Electrical Power [kW]", color="Cluster")
dev.new()
plot(clplot_Pelout)
```

```
#esamino outlier p termica
dfout_pt<-df_pt
dfout_pt$clust_Pt<-NULL
dfout_pt<-merge(dfout_pt,dfout_pel_norm)
centrout_pt<-ddply(dfout_pt,c("time","clust_Pel"),summarise,P_term=mean(P_term))</pre>
clplot_pt_out <- ggplot()+
 geom_line(data= dfout_pt, aes(x = as.POSIXct(time, format="%H:%M:%S" ,
tz="Etc/GMT+12"), y = P_term, group = date), size=0.65, color="grey50")+
geom_line(data= centrout_pt, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = P_term, color =as.factor(clust_Pel)), size=1)+
 scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y\_continuous(limits = c(-180,ceiling(max(df_pt$P_term)/100)*100))+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element text(size = 28),
         axis.title.x = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
```

```
strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour",y="Thermal Power [kW]", color="Cluster")
dev.new()
plot(clplot_pt_out)
#esamino outlier qes
dfout_qs<-df_qs
dfout_qs$clust_Qes<-NULL
dfout_qs<-merge(dfout_qs,dfout_pel_norm)</pre>
centrout_qs<-ddply(dfout_qs,c("time","clust_Pel"),summarise,Q_es=mean(Q_es))</pre>
clplot_qs_out <- ggplot()+
 geom_line(data= dfout_qs, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_es, group = date), size=0.65, color="grey50")+
 geom_line(data= centrout_qs, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_es, color =as.factor(clust_Pel)), size=1.05)+
 scale x datetime(expand=c(0,0), labels = date format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
scale_y_continuous(limits = c(0,0.12), expand=c(0,0),name=expression(Entering ~~
Source ~~ Flowrate ~~ group("[",m^3/s,"]")))+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
```

```
legend.title = element_text(size = 28),
         axis.title.x = element text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour",color="Cluster")
dev.new()
plot(clplot_qs_out)
#esamino outlier qel
dfout ql<-df ql
dfout_ql$clust_Qel<-NULL
dfout_ql<-merge(dfout_ql,dfout_pel_norm)</pre>
centrout_ql<-ddply(dfout_ql,c("time","clust_Pel"),summarise,Q_el=mean(Q_el))
clplot_ql_out <- ggplot()+
 geom line(data= dfout gl, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_el, group = date), size=0.65, color="grey50")+
 geom_line(data= centrout_ql, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_el, color =as.factor(clust_Pel)), size=1.05)+
 scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale y continuous(limits = c(0,0.12), expand=c(0,0),name=expression(Entering ~~ Load
~~ Flowrate ~~ group("[",m^3/s,"]")))+
 theme_bw()+
```

```
facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
          axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.text.x = element_text(size=12, angle=45, vjust = .5),
          axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour",color="Cluster")
dev.new()
plot(clplot_ql_out)
#esamino outlier tes
dfout_ts<-df_ts
dfout_ts$clust_Tes<-NULL
dfout_ts<-merge(dfout_ts,dfout_pel_norm)</pre>
centrout_Ts<-ddply(dfout_ts,c("time","clust_Pel"),summarise,T_es=mean(T_es))</pre>
clplot_Ts_out <- ggplot()+
```

```
geom_line(data= dfout_ts, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_es, group = date), size=0.65, color="grey50")+
 geom_line(data= centrout_Ts, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_es, color =as.factor(clust_Pel)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
scale_y\_continuous(limits = c(10,ceiling(max(df_ts$T_es)/1)*1), expand=c(0,0))+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Entering Source Temperature [°C]", color="Cluster")
dev.new()
plot(clplot_Ts_out)
```

```
#esamino outlier tel
dfout_tl<-df_tl
dfout_tl$clust_Tel<-NULL
dfout_tl<-merge(dfout_tl,dfout_pel_norm)</pre>
centrout_Tl<-ddply(dfout_tl,c("time","clust_Pel"),summarise,T_el=mean(T_el))
clplot_Tl_out <- ggplot()+
 geom_line(data= dfout_tl, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_el, group = date), size=0.65, color="grey50")+
geom_line(data= centrout_Tl, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_el, color =as.factor(clust_Pel)), size=1)+
 scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y = c(0,ceiling(max(df_tl\$T_el)/1)*1), expand=c(0,0))+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
```

```
strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Entering Load Temperature [°C]", color="Cluster")
dev.new()
plot(clplot_Tl_out)
dfpel_norm<-dfpel_norm2
df_pel<-df_pel2
dfpel_norm$clust_Pel<-NULL
df_pel$clust_Pel<-NULL
dfpel_norm<-dfpel_norm[!dfpel_norm$date %in% dfout_pel$date,]
df_pel<-df_pel[!df_pel$date %in% dfout_pel$date,]</pre>
df_pt<-df_pt[!df_pt$date %in% dfout_pel$date,]</pre>
df_ql<-df_ql[!df_ql$date %in% dfout_pel$date,]
df_qs<-df_qs[!df_qs$date %in% dfout_pel$date,]
df_tl<-df_tl[!df_tl$date %in% dfout_pel$date,]
df_ts<-df_ts[!df_ts$date %in% dfout_pel$date,] #gli outlier sono stati rimossi da tutti i df
dfpel_norm2<-dfpel_norm
df_pel2<-df_pel
#analizzo potenza elettrica
df1pel_norm<-spread(dfpel_norm,time,POWER)
df1_pel<-spread(df_pel,time,POWER)
dist_pel<-dist(df1pel_norm[12:107])
hcl_pel<-hclust(dist_pel,method = "ward.D2")
```

```
dev.new()
plot(hcl_pel)
dev.new()
fviz_nbclust(df1_pel[12:107], hcut, method = "wss") +
geom_vline(xintercept = 3, linetype = 2)
df1pel_norm$clust_Pel<-cutree(hcl_pel,3)
dfpel_norm<-gather(df1pel_norm,time,POWER,"00:00:00":"23:45:00")
dfpel_norm2<-dfpel_norm
dfpel_norm2$POWER<-NULL
df_pel<-merge.data.frame(df_pel,dfpel_norm2[c("date","time","clust_Pel")])
centr_pel_norm<-
ddply(dfpel_norm,c("clust_Pel","time"),summarise,POWER=mean(POWER))
centr_pel<-ddply(df_pel,c("clust_Pel","time"),summarise,POWER=mean(POWER))
clplot_Pel_norm <- ggplot()+
 geom_line(data= dfpel_norm, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
 geom_line(data= centr_pel_norm, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color =as.factor(clust Pel)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y_continuous(limits = c(0,ceiling(max(dfpel_norm$POWER)/1.2)*1.2),
expand=c(0,0)+
 theme_bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
```

```
legend.title = element_text(size = 28),
         axis.title.x = element text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Normalized Power", color="Cluster")
dev.new()
plot(clplot_Pel_norm)
clplot_Pel <- ggplot()+
 geom_line(data= df_pel, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, group = date), size=0.65, color="grey50")+
 geom_line(data= centr_pel, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = POWER, color =as.factor(clust_Pel)), size=1)+
 scale x datetime(expand=c(0,0), labels = date format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y_continuous(limits = c(0,ceiling(max(df_pel$POWER)/100)*100), expand=c(0,0))+
 theme bw()+
 facet_wrap(~clust_Pel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
```

```
legend.box = "vertical",
         legend.title = element text(size = 28),
          axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.title.y = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.text.x = element_text(size=12, angle=45, vjust = .5),
          axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Electrical Power [kW]", color="Cluster")
dev.new()
plot(clplot_Pel)
#esamino potenza termica
df1_pt<-spread(df_pt,time,P_term)
dist_pt < -dist(df1_pt[12:107])
hcl_pt<-hclust(dist_pt,method="ward.D2")
dev.new()
plot(hcl_pt)
dev.new()
fviz_nbclust(df1_pt[12:107], hcut, method = "wss") +
geom_vline(xintercept = 3, linetype = 2)
df1_pt$clust_Pt<-cutree(hcl_pt,3)
df_pt<-merge.data.frame(df_pt,df1_pt[c("date","clust_Pt")])</pre>
centr_pt<-ddply(df_pt,c("clust_Pt","time"),summarise,P_term=mean(P_term))</pre>
```

```
clplot_Pt <- ggplot()+
 geom line(data= df pt, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = P_term, group = date), size=0.65, color="grey50")+
 geom_line(data= centr_pt, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = P_term, color =as.factor(clust_Pt)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
scale_y = c(-200,ceiling(max(df_ptP_term)/100)*100), expand=c(0,0))+
 theme_bw()+
 facet_wrap(~clust_Pt)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Thermal Power [kW]", color="Cluster")
dev.new()
```

```
plot(clplot_Pt)
#esamino Tes
df1_ts<-spread(df_ts,time,T_es)
df1_ts[12:107]<-adply(df1_ts[12:107],1,normal)
dist_ts<-dist(df1_ts[12:107])
hcl_ts<-hclust(dist_ts,method="ward.D2")</pre>
dev.new()
plot(hcl_ts)
dev.new()
fviz_nbclust(df1_ts[12:107], hcut, method = "wss") +
geom_vline(xintercept = 3, linetype = 2)
df1_ts$clust_Tes<-cutree(hcl_ts,3)
df_ts<-merge.data.frame(df_ts,df1_ts[c("date","clust_Tes")])
centr_ts<-ddply(df_ts,c("clust_Tes","time"),summarise,T_es=mean(T_es))
clplot_Ts <- ggplot()+
geom_line(data= df_ts, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_es, group = date), size=0.65, color="grey50")+
tz="Etc/GMT+12"), y = T_es, color =as.factor(clust_Tes)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
scale_y = c(10,ceiling(max(df_tsT_es)/1)*1), expand=c(0,0))+
 theme_bw()+
 facet_wrap(~clust_Tes)+
 ggplot2::theme(legend.text=element_text(size=18),
        legend.position = "top",
        legend.key.width = unit(1.8, "cm"),
        legend.key.height = unit(0.6, "cm"),
```

```
legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
          axis.title.x = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.text.x = element_text(size=12, angle=45, vjust = .5),
          axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour", y="Entering Source Temperature [°C]", color="Cluster")
dev.new()
plot(clplot_Ts)
#esamino Tel
df1_tl<-spread(df_tl,time,T_el)
dist_tl<-dist(df1_tl[12:107])
hcl_tl<-hclust(dist_tl,method="ward.D2")</pre>
dev.new()
plot(hcl_tl)
dev.new()
fviz_nbclust(df1_tl[12:107], hcut, method = "wss") +
geom_vline(xintercept = 3, linetype = 2)
df1_tl$clust_Tel<-cutree(hcl_tl,3)
df_tl<-merge.data.frame(df_tl,df1_tl[c("date","clust_Tel")])
```

```
centr_tl<-ddply(df_tl,c("clust_Tel","time"),summarise,T_el=mean(T_el))
clplot_Tl <- ggplot()+
 geom line(data= df tl, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_el, group = date), size=0.65, color="grey50")+
 geom line(data= centr tl, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = T_el, color =as.factor(clust_Tel)), size=1)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
scale_y = c(0,ceiling(max(df_tl\$T_el)/1)*1), expand=c(0,0)+
 theme_bw()+
 facet_wrap(~clust_Tel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element text(size = 28),
         axis.title.x = element text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide legend(override.aes = list(size=10)))+
labs(x="Hour", y="Entering Load Temperature [°C]", color="Cluster")
```

```
dev.new()
plot(clplot Tl)
#esamino Qes
df1_qs<-spread(df_qs,time,Q_es)
dist_qs<-dist(df1_qs[12:107])
hcl_qs<-hclust(dist_qs,method="ward.D2")
dev.new()
plot(hcl_qs)
dev.new()
fviz_nbclust(df1_qs[12:107], hcut, method = "wss") +
geom_vline(xintercept = 3, linetype = 2)
df1_qs$clust_Qes<-cutree(hcl_qs,3)
df_qs<-merge.data.frame(df_qs,df1_qs[c("date","clust_Qes")])
centr_qs<-ddply(df_qs,c("clust_Qes","time"),summarise,Q_es=mean(Q_es))
clplot_qs <- ggplot()+
 geom_line(data= df_qs, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_es, group = date), size=0.65, color="grey50")+
 geom_line(data= centr_qs, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_es, color =as.factor(clust_Qes)), size=1.05)+
 scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_y_continuous(limits = c(0,0.12), expand=c(0,0),name=expression(Entering ~~
Source ~~ Flowrate ~~ group("[",m^3/s,"]")))+
 theme_bw()+
 facet_wrap(~clust_Qes)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
```

```
legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
          axis.title.x = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
          axis.text.x = element_text(size=12, angle=45, vjust = .5),
          axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour",color="Cluster")
dev.new()
plot(clplot_qs)
#esamino Qel
df1_ql<-spread(df_ql,time,Q_el)
dist_ql < -dist(df1_ql[12:107])
hcl_ql<-hclust(dist_ql,method="ward.D2")</pre>
dev.new()
plot(hcl_ql)
dev.new()
fviz_nbclust(df1_ql[12:107], hcut, method = "wss") +
geom_vline(xintercept = 3, linetype = 2)
df1_ql$clust_Qel<-cutree(hcl_ql,3)
df_ql<-merge.data.frame(df_ql,df1_ql[c("date","clust_Qel")])
```

```
centr_ql<-ddply(df_ql,c("clust_Qel","time"),summarise,Q_el=mean(Q_el))
clplot_ql <- ggplot()+</pre>
 geom_line(data= df_ql, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_el, group = date), size=0.65, color="grey50")+
 geom line(data= centr ql, aes(x = as.POSIXct(time, format="%H:%M:%S",
tz="Etc/GMT+12"), y = Q_el, color =as.factor(clust_Qel)), size=1.05)+
scale_x_datetime(expand=c(0,0), labels = date_format("%H:%M", tz="Etc/GMT+12"),
breaks = date_breaks("4 hour"))+
 scale_v_continuous(limits = c(0,0.05), expand=c(0,0),name=expression(Entering \sim Load)
~~ Flowrate ~~ group("[",m^3/s,"]")))+
 theme bw()+
 facet_wrap(~clust_Qel)+
 ggplot2::theme(legend.text=element_text(size=18),
         legend.position = "top",
         legend.key.width = unit(1.8, "cm"),
         legend.key.height = unit(0.6, "cm"),
         legend.direction = "horizontal",
         legend.box = "vertical",
         legend.title = element_text(size = 28),
         axis.title.x = element_text(size=18, margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.title.y = element_text(size=18,margin = margin(t = 18, r = 20, b = 0, l = 0)),
         axis.text.x = element_text(size=12, angle=45, vjust = .5),
         axis.text.y = element_text(size=12, vjust=.3),
         panel.background = element_rect(fill="white"),
         panel.border = element_rect(fill = NA,colour = "grey50"),
         strip.background = element_rect(fill = "grey80",colour = "grey50"),
         plot.background = element_rect(colour = "grey50")
 )+ scale_colour_manual(values = pals)+
 guides(colour = guide_legend(override.aes = list(size=10)))+
labs(x="Hour",color="Cluster")
```

```
dev.new()
plot(clplot_ql)
#creo dataset completo di tutto
dfall<-select(df_pel,date,time,Day_Type,Month,T_mean,T_min,T_max,
       Tes_mean, Tel_mean, Qel_mean, Qes_mean, Pt_mean, POWER, clust_Pel)
dfall<-merge.data.frame(dfall,df_pt)
dfall<-merge.data.frame(dfall,df_ts)
dfall<-merge.data.frame(dfall,df_tl)
dfall<-merge.data.frame(dfall,df_qs)
dfall<-merge.data.frame(dfall,df_ql)
#creo nuovi df per CART
df_pel<-select(dfall,-c(T_es,T_el,Q_es,Q_el,P_term))
df_pt<-select(dfall,-c(T_es,T_el,Q_es,Q_el,POWER))
df_tl<-select(dfall,-c(T_es,Q_es,Q_el,P_term,POWER))
df_ts<-select(dfall,-c(T_el,Q_es,Q_el,P_term,POWER))
df_qs<-select(dfall,-c(T_el,T_es,Q_el,P_term,POWER))
df_ql<-select(dfall,-c(T_el,Q_es,T_es,P_term,POWER))
df1_pel<-spread(df_pel,time,POWER)
df1_pt<-spread(df_pt,time,P_term)
df1_ts<-spread(df_ts,time,T_es)
df1_tl<-spread(df_tl,time,T_el)
df1_qs<-spread(df_qs,time,Q_es)
df1_ql<-spread(df_ql,time,Q_el)
#CART Qes
```

```
ct_qes1<-
rpart(as.factor(clust_Qes)~as.factor(Day_Type)+T_mean+T_min+T_max,data=df1_qs,cont
rol = rpart.control(minbucket = 10))
ct_party_qes1 <- as.party(ct_qes1)
dev.new()
plot(ct_party_qes1,terminal_panel = node_barplot(ct_party_qes1, fill=pals, widths=3
_{gp=gpar(fontsize=8, lwd = 0.7))}
  gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_qes1,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_qs$clust_Qes)</pre>
confusionMatrix(pr,df.p)
ct qes2<-
rpart(as.factor(clust_Qes)~Tes_mean+Tel_mean+Qel_mean+as.factor(clust_Tes)+as.factor(
clust_Tel)+as.factor(clust_Qel),data=df1_qs,control = rpart.control(minbucket = 14))
ct_party_qes2 <- as.party(ct_qes2)
dev.new()
plot(ct_party_qes2,terminal_panel = node_barplot(ct_party_qes2, fill=pals, widths=3)
,gp=gpar(fontsize=8, lwd = 0.7)),
   gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_qes2,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_qs$clust_Qes)
confusionMatrix(pr,df.p)
#CART Qel
ct_qel1<-
rpart(as.factor(clust_Qel)~as.factor(Day_Type)+T_mean+T_min+T_max,data=df1_ql,contr
ol = rpart.control(minbucket = 10))
ct_party_qel1 <- as.party(ct_qel1)
dev.new()
```

```
plot(ct_party_qel1,terminal_panel = node_barplot(ct_party_qel1, fill=pals, widths=3
_{gp=gpar}(fontsize=8, lwd = 0.7)),
  gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_qel1,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_ql$clust_Qel)
confusionMatrix(pr,df.p)
ct_qel2<-
rpart(as.factor(clust_Qel)~Tes_mean+Tel_mean+Qes_mean+T_min+T_mean+T_max+as.fa
ctor(clust_Qes)+as.factor(clust_Tel)+as.factor(clust_Tes),data=df1_ql,control =
rpart.control(minbucket = 8))
ct_party_qel2 <- as.party(ct_qel2)
dev.new()
plot(ct_party_qel2,terminal_panel = node_barplot(ct_party_qel2, fill=pals, widths=3
,gp=gpar(fontsize=8, lwd = 0.7)),
  gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_qel2,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_ql$clust_Qel)
confusionMatrix(pr,df.p)
#CART Tes
ct_tes1<-
rpart(as.factor(clust_Tes)~as.factor(Day_Type)+as.factor(Month)+T_mean+T_min+T_max,
data=df1_ts,control = rpart.control(minbucket = 6))
ct_party_tes1 <- as.party(ct_tes1)
dev.new()
plot(ct_party_tes1,terminal_panel = node_barplot(ct_party_tes1, fill=pals,widths=3
,gp=gpar(fontsize=8, lwd = 0.7)),
   gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_tes1,type = "vector")</pre>
```

```
pr<-as.factor(pr)
df.p<-as.factor(df1_ts$clust_Tes)</pre>
confusionMatrix(pr,df.p)
ct tes2<-
rpart(as.factor(clust_Tes)~Tel_mean+Qes_mean+as.factor(clust_Qel)+as.factor(clust_Tel)+
as.factor(clust_Qes),data=df1_ts,control = rpart.control(minbucket = 6))
ct_party_tes2 <- as.party(ct_tes2)
dev.new()
plot(ct_party_tes2,terminal_panel = node_barplot(ct_party_tes2, fill=pals, widths=3
,gp=gpar(fontsize=8, lwd = 0.7)),
  gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_tes2,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_ts$clust_Tes)
confusionMatrix(pr,df.p)
#CART Tel
ct tel1<-
rpart(as.factor(clust_Tel)~as.factor(Day_Type)+T_mean+T_min+T_max,data=df1_tl,contro
l = rpart.control(minbucket = 10))
ct_party_tel1 <- as.party(ct_tel1)
dev.new()
plot(ct_party_tel1,terminal_panel =
node_barplot(ct_party_tel1,fill=pals,widths=3,gp=gpar(fontsize=8, lwd = 0.7)),
   gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_tel1,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_tl$clust_Tel)
confusionMatrix(pr,df.p)
ct_tel2<-rpart(as.factor(clust_Tel)~T_max+Qes_mean+Tes_mean +as.factor(clust_Tes)+
```

```
as.factor(clust_Qel)+as.factor(clust_Qes),data=df1_tl,
        control = rpart.control(minbucket = 5))
ct_party_tel2 <- as.party(ct_tel2)
dev.new()
plot(ct_party_tel2,terminal_panel = node_barplot(ct_party_tel2, fill=pals, widths=3
,gp=gpar(fontsize=8, lwd = 0.7)),
   gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_tel2,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_tl$clust_Tel)
confusionMatrix(pr,df.p)
ct_tel3<-rpart(as.factor(clust_Tel)~as.factor(Day_Type)+
as.factor(Month)+T_mean+T_min+T_max,data=df1_tl,control = rpart.control(minbucket =
6))
#CART P_termica
ct_pt1<-
rpart(as.factor(clust_Pt)~as.factor(Day_Type)+T_mean+T_min+T_max,data=df1_pt,contro
l = rpart.control(minbucket = 12))
ct_party_pt1 <- as.party(ct_pt1)
dev.new()
plot(ct_party_pt1,terminal_panel = node_barplot(ct_party_pt1, fill=pals, widths=3)
,gp=gpar(fontsize=8, lwd = 0.7)),
   gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_pt1,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_pt$clust_Pt)</pre>
confusionMatrix(pr,df.p)
ct pt2<-
rpart(as.factor(clust_Pt)~as.factor(clust_Tes)+as.factor(clust_Tel)+as.factor(clust_Qes)+as.f
```

```
actor(clust_Qel)+as.factor(clust_Pel)+Pel_mean+Tes_mean+Tel_mean+Qes_mean+Qel_me
an,data=df1_pt,control = rpart.control(minbucket = 12))
ct_party_pt2 <- as.party(ct_pt2)
dev.new()
plot(ct_party_pt2,terminal_panel = node_barplot(ct_party_pt2 , fill=pals ,widths=3
,gp=gpar(fontsize=8, lwd = 0.7)),
  gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_pt2,type = "vector")</pre>
pr<-as.factor(pr)
df.p<-as.factor(df1_pt$clust_Pt)
confusionMatrix(pr,df.p)
ct_pt3<-
rpart(as.factor(clust_Pt)~as.factor(clust_Tes)+as.factor(clust_Tel)+as.factor(clust_Pel)+as.fa
ctor(clust_Qes)+as.factor(clust_Qel)+Tes_mean+Tel_mean,data=df1_pt,control
= rpart.control(minbucket = 12))
ct_party_pt3 <- as.party(ct_pt3)
dev.new()
plot(ct_party_pt3,terminal_panel = node_barplot(ct_party_pt3 , fill=pals ,widths=3
,gp=gpar(fontsize=8, lwd = 0.7)),
  gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_pt3,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_pt$clust_Pt)
confusionMatrix(pr,df.p)
#CART P_elettrica
ct_pel1<-
rpart(as.factor(clust_Pel)~as.factor(Day_Type)+T_mean+T_min+T_max,data=df1_pel,cont
rol = rpart.control(minbucket = 8))
ct_party_pel1 <- as.party(ct_pel1)
dev.new()
```

```
plot(ct_party_pel1,terminal_panel = node_barplot(ct_party_pel1, fill=pals, widths=3
_{gp=gpar}(fontsize=8, lwd = 0.7)),
  gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_pel1,type = "vector")</pre>
pr<-as.factor(pr)
df.p<-as.factor(df1_pel$clust_Pel)
confusionMatrix(pr,df.p)
ct_pel2<-
rpart(as.factor(clust_Pel)~as.factor(Day_Type)+T_mean+T_min+T_max+Pt_mean,data=df
1_pel,control = rpart.control(minbucket = 8))
ct_party_pel2 <- as.party(ct_pel2)
dev.new()
plot(ct_party_pel2,terminal_panel = node_barplot(ct_party_pel2, fill=pals, widths=3
,gp=gpar(fontsize=8, lwd = 0.7)),
   gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_pel1,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_pel$clust_Pel)</pre>
confusionMatrix(pr,df.p)
ct_pel3<-
rpart(as.factor(clust_Pel)~as.factor(clust_Tes)+as.factor(clust_Tel)+as.factor(clust_Qes)+as.
factor(clust_Qel)+as.factor(clust_Pt)+Tes_mean+Tel_mean+Qes_mean+Qel_mean+Pt_mea
n,data=df1_pel,control = rpart.control(minbucket = 6))
ct_party_pel3 <- as.party(ct_pel3)
dev.new()
plot(ct_party_pel3,terminal_panel = node_barplot(ct_party_pel2, fill=pals,widths=3
,gp=gpar(fontsize=8, lwd = 0.7)),
   gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_pel2,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_pel$clust_Pel)
```

```
confusionMatrix(pr,df.p)
ct_pel4<-
rpart(as.factor(clust_Pel)~as.factor(clust_Tes)+as.factor(clust_Tel)+as.factor(clust_Pt)+Pt_
mean+Tes_mean+Tel_mean+Qel_mean,data=df1_pel,control = rpart.control(minbucket =
6))
ct_party_pel4 <- as.party(ct_pel4)
dev.new()
plot(ct_party_pel4,terminal_panel = node_barplot(ct_party_pel4, fill=pals, widths=3
,gp=gpar(fontsize=8, lwd = 0.7)),
  gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_pel3,type = "vector")</pre>
pr<-as.factor(pr)
df.p<-as.factor(df1_pel$clust_Pel)
confusionMatrix(pr,df.p)
ct_pel5<-
rpart(as.factor(clust_Pel)~as.factor(clust_Tel)+as.factor(clust_Pt)+Pt_mean+Tes_mean+Tel
_mean,data=df1_pel,control = rpart.control(minbucket = 6))
ct_party_pel5 <- as.party(ct_pel5)
dev.new()
plot(ct_party_pel5,terminal_panel = node_barplot(ct_party_pel5, fill=pals, widths=3
,gp=gpar(fontsize=8, lwd = 0.7)),
   gp=gpar(fontsize=12, lwd = 0.7))
pr<-predict(ct_pel4,type = "vector")</pre>
pr<-as.factor(pr)</pre>
df.p<-as.factor(df1_pel$clust_Pel)
confusionMatrix(pr,df.p)
```

## Allegato II

Il presente allegato coincide con il file esterno creato per la calibrazione del type 927 e contenente i valori caratteristici del solo funzionamento a regime. Come già ampiamente spiegato nel paragrafo 5.3, le prime 4 righe corrispondono alle variabili indipendenti, mentre le restanti righe contengono rispettivamente i valori normalizzati di potenza termica ed elettrica corrispondenti a ciascuna combinazione. Tali combinazioni sono inoltre riportate sulla destra di ogni riga precedute dal punto esclamativo. Se ad esempio consideriamo la quinta riga avremo che i valori normalizzati di potenza termica= 0.4674 ed elettrica= 0.3258 si ottengono quando Tes= 14.9, Tel=8.7, Qes\_norm=0.2 e Qel\_norm=0.95.

0.95	0.96	0.97	0.98		!Normalized Load Flow Rate
0.2	0.4	0.5	0.7		!Normalized Source Flow Rate
8.7		9.2	9.5	9.9	!Entering Load Temperatures
14.9		15.1			!Entering Source Temperatures
0.4674	0.3258				!14.9-8.7-0.2-0.95
0.4449	0.3291				
0.4706	0.3258				
0.5118	0.3451				!14.9-9-0.2-0.95
0.2759	0.3446				
0.5095	0.2665				
0.5600	0.4357				!14.9-9.2-0.2-0.95
0.5635	0.3574				
0.5459	0.3540				
0.625216554	0.39405				!14.9-9.5-0.2-0.95
0.6241	0.3836				
0.616939365	55 0.364906667				
0.7113	0.4462				!14.9-9.9-0.2-0.95
0.7091	0.4494				
0.703753609	0.45025				
0.4579	0.3223				!14.9-8.7-0.2-0.96
0.4768	0.3066				
0.4681	0.3371				
0.5209	0.5039				!14.9-9-0.2-0.96
0.527237729	0.35475				
0.5182	0.3451				
0.5557	0.3915				!14.9-9.2-0.2-0.96
0.565447546	0.33825				

0.5745	0.007000007	
0.5745	0.367066667	14400=000
0.6327	0.3933	!14.9-9.5-0.2-0.96
0.6792	0.382763636	
0.6479	0.4098	
0.7386	0.2231	!14.9-9.9-0.2-0.96
0.733301251		
0.7198	0.4575	
0.4346	0.3098	!14.9-8.7-0.2-0.97
0.4527	0.3291	
0.4674	0.3130	
0.5217	0.3507	!14.9-9-0.2-0.97
0.1839	0.2491	
0.5273	0.2721	
0.5498	0.4678	!14.9-9.2-0.2-0.97
0.5449	0.3291	
0.617901829	0.36915	
0.632	0.3993	!14.9-9.5-0.2-0.97
0.6439	0.400733333	
0.631087584	0.39285	
0.7236	0.4543	!14.9-9.9-0.2-0.97
0.7507	0.2263	
0.715784408	0.492	
0.4397	0.3274	!14.9-8.7-0.2-0.98
0.4680	0.3398	
0.464870067	0.31785	
0.526895519	0.350966667	!14.9-9-0.2-0.98
0.55587103	0.346303	
0.533461662	0.346135	
0.5729	0.3547	!14.9-9.2-0.2-0.98
0.601362083	0.364738462	
0.604170677	0.577866667	
0.662389049	0.387282667	!14.9-9.5-0.2-0.98
0.660827719	0.3515	
0.657401347	0.405300928	
0.7321	0.4462	!14.9-9.9-0.2-0.98
0.741706769	0.224013	
0.7657	0.4478	
0.4619	0.3234	!14.9-8.7-0.4-0.95
0.4348	0.3339	
0.4564	0.32665	
	0.348833333	!14.9-9-0.4-0.95
0.5009	0.3419	
	0.289466667	
0.6062	0.5955	!14.9-9.2-0.4-0.95
0.5690	0.3895	5.2 5.1 5.55
3.3030	0.0000	

0.5750	0.4350	
0.6204	0.3772	!14.9-9.5-0.4-0.95
0.622329163	0.3981	
0.6456	0.3612	
0.7126	0.4543	!14.9-9.9-0.4-0.95
0.7138	0.4077	
0.6678	0.4414	
0.4477	0.305	!14.9-8.7-0.4-0.96
0.4485	0.3307	
0.406939167	0.327983333	
0.5167	0.3494	!14.9-9-0.4-0.96
0.53923151	0.359815385	
0.515688162	0.25845	
0.5388	0.5297	!14.9-9.2-0.4-0.96
0.5869	0.3323	
0.5623	0.3499	
0.6539	0.4109	!14.9-9.5-0.4-0.96
0.649919795	0.420311049	
0.6346	0.4037	
0.7147	0.4494	!14.9-9.9-0.4-0.96
0.702454283	0.39522447	
0.720666506	0.449205	
0.4666	0.3098	!14.9-8.7-0.4-0.97
0.464677575	0.34915	
0.4780	0.3323	
0.5234	0.3499	!14.9-9-0.4-0.97
0.5838	0.4462	
0.526306545	0.2976325	
0.5688	0.3579	!14.9-9.2-0.4-0.97
0.573648808	0.366031667	
0.596102021	0.41562254	
0.6603	0.3884	!14.9-9.5-0.4-0.97
0.633686237	0.39805	
0.623387873	0.402366667	
0.7328	0.28505	!14.9-9.9-0.4-0.97
0.7448	0.4254	
0.727	0.469505714	
0.4794	0.3186	!14.9-8.7-0.4-0.98
0.4103	0.3276	
0.4735	0.3188	
0.5245	0.3515	!14.9-9-0.4-0.98
0.5731	0.3563	
0.397160731	0.291594444	
0.5474	0.5714	!14.9-9.2-0.4-0.98
0.5702	0.3499	

0.5709	0.3563	
0.6679	0.3911	!14.9-9.5-0.4-0.98
0.6387	0.3788	
0.5796	0.3879	
0.731761309	0.4374	!14.9-9.9-0.4-0.98
0.737969201	0.276075	
0.725761951	0.45665	
0.465351299	0.3082	!14.9-8.7-0.5-0.95
0.4279	0.3307	
0.4654	0.3291	
0.51039461	0.3507	!14.9-9-0.5-0.95
0.519441771	0.33862	
0.5091	0.3435	
0.5839	0.4842	!14.9-9.2-0.5-0.95
0.5542	0.3547	
0.5559	0.3579	
0.5894	0.2039	14.9-9.5-0.5-0.95
0.621559192	0.409309791	
0.6316	0.3884	
0.7070	0.4318	14.9-9.9-0.5-0.95
0.68825794	0.371105	
0.5567	0.1750	
0.4689	0.3291	!14.9-8.7-0.5-0.96
0.4491	0.3339	
0.4724	0.3242	
0.516843118	0.348833333	!14.9-9-0.5-0.96
0.5427	0.38205	
0.5143	0.3419	
0.5388	0.5402	!14.9-9.2-0.5-0.96
0.5405	0.354033333	
0.5598	0.3547	
0.6510	0.3997	!14.9-9.5-0.5-0.96
0.638917487	0.481581935	
0.6052	0.3965	
0.7274	0.2301	!14.9-9.9-0.5-0.96
0.7428	0.4142	
0.6939	0.4446	
0.4747	0.3258	!14.9-8.7-0.5-0.97
0.4629	0.3339	
0.4751	0.3307	
0.5221	0.3507	!14.9-9-0.5-0.97
0.5519	0.3949	
0.5058	0.2371	
0.5650	0.3596	!14.9-9.2-0.5-0.97
0.5638	0.3483	

0.5007	0.0005	
0.5807	0.3886	14400505050
0.6591	0.3884	!14.9-9.5-0.5-0.97
0.6343	0.4157	
0.6477	0.378940741	
0.7270	0.4393	!14.9-9.9-0.5-0.97
0.7510	0.4098	
0.7590	0.4254	
0.4820	0.3274	!14.9-8.7-0.5-0.98
0.4312	0.3211	
0.4981	0.3647	
0.4945	0.3419	!14.9-9-0.5-0.98
0.3374	0.3557	
0.4953	0.3403	
0.5942	0.3660	!14.9-9.2-0.5-0.98
0.6256	0.3494	
0.560250241	0.3665	
0.6526	0.40045	!14.9-9.5-0.5-0.98
0.6380	0.3799	
0.630606352	0.4005	
0.7311	0.4387	!14.9-9.9-0.5-0.98
0.7538	0.4141	
0.7523	0.5136	
0.4645	0.3226	!14.9-8.7-0.7-0.95
0.4447	0.3371	
0.4683	0.3291	
0.5128	0.3499	!14.9-9-0.7-0.95
0.5582	0.3644	
0.504331088	0.339766667	
0.5634	0.3852	!14.9-9.2-0.7-0.95
0.567853705	0.3661	
0.5523	0.3740	
0.6468	0.4045	!14.9-9.5-0.7-0.95
0.628039782	0.4301	
0.6273	0.3724	
0.7135	0.4345	!14.9-9.9-0.7-0.95
0.694642284	0.4251	
0.6874	0.4478	
0.465768367	0.31814	!14.9-8.7-0.7-0.96
0.445428296	0.3371	
0.4720	0.3307	
0.515591915	0.3491	!14.9-9-0.7-0.96
0.528055823		
0.5169	0.1742	
0.5609	0.3740	!14.9-9.2-0.7-0.96
0.5636	0.3579	2 312 217 210
3.555		

0.5844	0.3676	
0.6223	0.3804	!14.9-9.5-0.7-0.96
0.636766121	0.3975	
0.616297722	0.411433333	
0.735514918	0.3363	!14.9-9.9-0.7-0.96
0.7402	0.4015	
0.6987	0.4462	
0.4764	0.329404	!14.9-8.7-0.7-0.97
0.436669875	0.3307	
0.4668	0.3579	
0.5188	0.3531	!14.9-9-0.7-0.97
0.558421559	0.3732	
0.520667308	0.2441	
0.5444	0.5656	!14.9-9.2-0.7-0.97
0.570259865	0.35635	
0.542637151	0.3421	
0.6576	0.3981	!14.9-9.5-0.7-0.97
0.6321	0.3917	
0.638460058	0.402070471	
0.7253	0.4318	!14.9-9.9-0.7-0.97
0.726082772	0.2825	
0.7471	0.49035	
0.4789	0.3274	!14.9-8.7-0.7-0.98
0.4278	0.3312	
0.4887	0.3596	
0.519634264	0.3483	!14.9-9-0.7-0.98
0.4651	0.3419	
0.5272	0.3435	
0.5932	0.3632	!14.9-9.2-0.7-0.98
0.5478	0.3483	
0.6228	0.4216	
0.683926853	0.5281	!14.9-9.5-0.7-0.98
0.6932	0.3900	
0.6175	0.3933	
0.7332	0.4398	!14.9-9.9-0.7-0.98
0.7473	0.2305	
0.7347	0.4703	

## Bibliografia

- [01] Organisation for Economic Co-operation and Development, Transition to Sustainable Buildings: Strategies and Opportunities to 2050. (2013). OECD.
- [02] Proposta di DIRETTIVA DEL PARLAMENTO EUROPEO E DEL CONSIGLIO che modifica la direttiva 2010/31/UE sulla prestazione energetica nell'edilizia, C. -2. (s.d.).
- [03] Decreto legislativo 3 marzo 2011, n. 2. (s.d.).
- [04] Italiana, R. (. (s.d.). LEGGE 3 agosto 2013, n. 90. Conversione in legge, con modificazioni, del decreto-legge, (63).
- [05] Decreto Legislativo 4 luglio 2014, n. 1. (s.d.).
- [06] Europeo, P. (. (s.d.). Gazzetta ufficiale dell'Unione Europea n. L, 315(1).
- [07] Governo Italiano, D. M.-n.-n. (s.d.).
- [08] ISO 5001 Energy Managment Systems, (. (s.d.).
- [09] Filippi, M. C. (2010). Annex 53-IEA "Total energy use in buildings-Analysis and evaluation methods".
- [10] Ouyang, J. &. (2009). Energy-saving potential by improving occupants' behavior in urban residential sector in Hangzhou City, China. *Energy and Buildings*, 41(7), 711-720.
- [11] Yu, Z. F. (2011). A systematic procedure to study the influence of occupant behavior on building energy consumption. *Energy and Buildings*, 43(6), 1409-1417.
- [12] D'Oca, S. &. (2014). A data-mining approach to discover patterns of window opening and closing behavior in offices. *Building and Environment*, 82., 726-739.
- [13] Capozzoli, A. P. (2017). Data analytics for occupancy pattern learning to reduce the energy consumption of HVAC systems in office buildings. *Sustainable Cities and Society*, 35, 191-208.

- [14] Fan, C. X. (2017). Unsupervised Data Analytics in Mining Big Building Operational Data for Energy Efficiency Enhancement: A Review. *Energy and Buildings*.
- [15] Wei, Y. Z. (2018). A review of data-driven approaches for prediction and classification of building energy consumption. *Renewable and Sustainable Energy Reviews*, 82, 1027-1047.
- [16] Panapakidis, I. P. (2014). Pattern recognition algorithms for electricity load curve analysis of buildings. *Energy and Buildings*, 73, 137-145.
- [17] Fan, C. X. (2015). Temporal knowledge discovery in big BAS data for building energy management. *Energy and Buildings*, 109, 75-89.
- [18] Miller, C. N. (2015). Automated daily pattern filtering of measured building performance data. *Automation in Construction*, 49, 1-17.
- [19] McLoughlin, F. D. (2015). A clustering approach to domestic electricity load profile characterisation using smart metering data. *Applied energy*, 141, 190-199.
- [20] do Carmo, C. M. (2016). Cluster analysis of residential heat load profiles and the role of technical and household characteristics. *Energy and Buildings*, 125, 171-180.
- [21] Rhodes, J. D. (2014). Clustering analysis of residential electricity demand profiles. *Applied Energy*, 135, 461-471.
- [22] Ma, Z. Y. (2017). A variation focused cluster analysis strategy to identify typical daily heating load profiles of higher education buildings. *Energy*, *134*, 90-102.
- [23] Capozzoli, A. P. (2017). Mining typical load profiles in buildings to support energy management in the smart city context. *Energy Procedia*, 134, 865-874.
- [24] Fan, C. X. (2017). Unsupervised Data Analytics in Mining Big Building Operational Data for Energy Efficiency Enhancement: A Review. *Energy and Buildings*.
- [25] Han, H. G. (2011). Fault diagnosis for refrigeration systems based on principal component analysis and support vector machine. *Journal of Shanghai Jiaotong University*, 45(9),, 1355-1361.

- [26] Chou, J. S. (2014). Real-time detection of anomalous power consumption. *Renewable* and Sustainable Energy Reviews, 33, 400-411.
- [27] Mulumba, T. A. (2015). Robust model-based fault diagnosis for air handling units. *Energy and Buildings*, 86, 698-707.
- [28] Li, D. H. (2016). A data-driven strategy for detection and diagnosis of building chiller faults using linear discriminant analysis. *Energy and Buildings*, 128, 519-529.
- [29] Beghi, A. B. (2016). Data-driven fault detection and diagnosis for HVAC water chillers. *Control Engineering Practice*, *53*, 79-91.
- [30] Guo, Y. L. (2017). Optimized neural network-based fault diagnosis strategy for VRF system in heating mode using data mining. *Applied Thermal Engineering*, 125, 1402-1413.
- [31] Capozzoli, A. P. (2018). Automated load pattern learning and anomaly detection for enhancing energy management in smart buildings. *Energy*.
- [32] Amasyali, K. &.-G. (2018). A review of data-driven building energy consumption prediction studies. *Renewable and Sustainable Energy Reviews*, 81, 1192-1205.
- [33] Cho, S. L. (2013). Energy simulation modeling and savings analysis of load sharing between house and office. *Renewable energy*, 54, 70-77.
- [34] Korolija, I. M.-H. (2011). Influence of building parameters and HVAC systems coupling on building energy performance. *Energy and Buildings*, 43(6), 1247-1253.
- [35] Zhou, Y. P. (2007). Energy simulation in the variable refrigerant flow air-conditioning system under cooling conditions. *Energy and Buildings*, 39(2), 212-220.
- [36] Zhao, D. Z. (2016). Energy consumption predicting model of VRV (Variable refrigerant volume) system in office buildings based on data mining. *Energy*, 102, 660-668.
- [37] Yan, L. H. (2016). The performance prediction of ground source heat pump system based on monitoring data and data mining technology. *Energy and Buildings*, 127, 1085-1095.

- [38] Li, Q. M. (2009). Applying support vector machine to predict hourly cooling load in the building. *Applied Energy*, 86(10), 2249-2256.
- [39] Jovanović, R. Ž. (2015). Ensemble of various neural networks for prediction of heating energy consumption. *Energy and Buildings*, *94*, 189-199.
- [40] Zhou, Z. H. (2002). Ensembling neural networks: many could be better than all. *Artificial intelligence*, 137(1-2), 239-263.
- [41] Jinhu, L. X. (2010). Applying principal component analysis and weighted support vector machine in building cooling load forecasting. 2010 International Conference On, Vol.1, 434-437.
- [42] Lee, P. L. (2016). Analysis of an air-cooled chiller replacement project using a probabilistic approach for energy performance contracts. *Applied energy*, 171, 415-428.
- [43] Santamouris, M. M. (2007). Using intelligent clustering techniques to classify the energy performance of school buildings. *Energy and buildings*, 39.
- [44] ASHRAE, H. F. (2009). "Energy estimating and modelling methods".
- [45] Coakley, D. R. (2014). A review of methods to match building energy simulation models to measured data. *Renewable and sustainable energy reviews*, 37, 123-141.
- [46] Clarke, J. (2007). Energy simulation in building design. Routledge.
- [47] Tan, P. N. (2006). Introduction to data mining, Chapter 4. Pearson Education India.
- [48] Dulli, S. F. (2009). Data mining: metodi e strategie. Springer Science & Business Media.
- [49] Yu, Z. F. (2013, June). Extracting knowledge from building-related data—A data mining framework. *In Building Simulation*, Vol. 6, No. 2, pp. 207-222.
- [50] Jensen, D. D. (2000). Multiple comparisons in induction algorithms. *Machine Learning*, 38(3), 309-338.
- [51] Roiger, R. J. (2004). Introduzione al data mining. McGraw-Hill Italia.
- [52] Tan, P. N. (2006). Introduction to data mining, Chapter 8. Pearson Education India.

- [53] Bholowalia, P. &. (2014). EBK-means: A clustering technique based on elbow method and k-means in WSN. *International Journal of Computer Applications*, 105(9).
- [54] Ketchen, D. J. (1996). The application of cluster analysis in strategic management research: an analysis and critique. *Strategic management journal*, *17*(6), 441-458.
- [55] Basta, S. &. (2008). Geotermia e pompe di calore: guida pratica agli impianti geotermici di climatizzazione. Associazione Geotermia. Org.
- [56]. (1995). Geothermal Energy, ASHRAE Handbook-Applications. ASHRAE.
- [57] Starace, G. (2003). Le Pompe di Calore Geotermiche. AICARR" Le moderne tecnologie negli impianti e nei componenti per il riscaldamento Padova, Bari, Catania. AICARR, Italy.
- [58]. (s.d.). efficienzaenergetica.enea.it. Tratto da http://www.efficienzaenergetica.enea.it/Cittadino/climatizzatori/solo-freddo-e-pompe-di-calore
- [59] Grazzini, G. &. (2017). Tecnica del Freddo. Società Editrice Esculapio.
- [60] ALDEGHERI, A. (2012). Studio di pompe di calore a CO2 con o senza eiettore.
- [61] Breiman, L. F. (1984).
- [62] Klein, S. A.-a.-M. (s.d.).

## Ringraziamenti

Si ringrazia il Living Lab del Politecnico di Torino per aver fornito i dati utilizzati per il presente lavoro di tesi.