



Politecnico di Torino

Corso di laurea Magistrale in Ingegneria Meccanica

A.A. 2025/2026

Sessione di Laurea: marzo-aprile 2026

Automazione CAD/CFD e Ottimizzazione Multi-Obiettivo Basata su Reti Neurali di Dissipatori Pin-Fin a Liquido per Elettronica di Potenza

Relatore:

Prof. Salvatore Musumeci

Co-relatori:

Dott. Fausto Stella

Dott. Luigi Solimene

Candidato:

Filippo Lepre

(matricola 320021)

Sommario

ABSTRACT	1
1. Introduzione	2
1.1 Obiettivi e contributi della tesi.....	2
1.2 Struttura della tesi	3
2. Stato dell'arte e fondamenti	4
2.1 Principi di trasferimento di calore rilevanti per dissipatori.....	4
2.1.1 Conduzione.....	4
2.1.2 Convezione	7
2.1.3 Irraggiamento	10
2.2 Tipologia di dissipatori per elettronica	11
2.2.1 Evoluzione delle geometrie e dei metodi di progettazione.....	13
3. Validazione SolidWorks Flow Simulation.....	18
3.1 Setup della simulazione CFD	19
3.2 Calibrazione dell'area riscaldata tramite R_{th}	21
3.3 Confronto R_{th} al variare della portata e sintesi validazione	23
4. Sviluppo dell'automazione MATLAB–SolidWorks per l'ottimizzazione di un dissipatore Pin-Fin a liquido.....	26
4.1 Ricostruzione del modello CAD di base e creazione del progetto termico	27
4.2 Integrazione MATLAB–SolidWorks	30
4.3 Tentativo preliminare di ottimizzazione diretta con Pattern Search.....	32
4.3.1 Discussione dei risultati del primo tentativo.....	37
4.3.2 Tentativo di mitigazione: area ristretta e mesh locale.....	42
4.3.3 Discussione dei risultati del secondo tentativo e conclusioni	45
5. Ottimizzazione multi-obiettivo assistita da surrogato	47
5.1 Setup	48
5.1.1 Parametrizzazione compatta.....	48
5.1.2 Preparazione del modello CAD per la generazione del dataset.....	50
5.1.3 Studio preliminare della Mesh locale: accuratezza vs costo.....	52
5.2 Generazione del dataset mediante sequenze di Sobol	56
5.3 Modello surrogato basato su rete neurale artificiale feed-forward.....	59
5.3.1 Analisi preliminari di correlazione.....	61
5.3.2 Suddivisione del dataset e pre-processing dei dati	62
5.3.3 Architettura della rete neurale	63
5.3.4 Valutazione del modello surrogato	67
5.4 Ottimizzazione multi-obiettivo assistita da surrogato basata sull'algoritmo genetico NSGA-II	69

5.4.1	Implementazione della procedura di ottimizzazione.....	71
5.4.2	Risultati dell'ottimizzazione e validazione CFD ad alta fedeltà	73
5.5	Conclusioni.....	80
6.	Applicazione del metodo a un caso più realistico: ottimizzazione della forma dei pin	82
6.1	Setup e dataset	82
6.1.1	Parametrizzazione compatta.....	85
6.1.2	Preparazione del modello CAD per la generazione del dataset.....	87
6.1.3	Generazione del dataset	89
6.2	Modello surrogato	90
6.2.1	Analisi preliminari di correlazione.....	90
6.2.2	Valutazione del modello surrogato	93
6.3	Ottimizzazione multi-obiettivo assistita da surrogato	95
6.3.1	Risultati dell'ottimizzazione e validazione CFD ad alta fedeltà	96
6.3.2	Esempio di due soluzioni significative e confronto con cilindri tradizionali	104
6.4	Conclusioni.....	114
7.	Conclusioni generali e sviluppi futuri	115
	Riferimenti	118

ABSTRACT

La gestione termica è un aspetto cruciale nella progettazione di sistemi elettronici ad alta densità di potenza, poiché la temperatura di esercizio influenza prestazioni, affidabilità e vita utile dei componenti.

Questa tesi propone un metodo per l'esplorazione e l'ottimizzazione di dissipatori, basato sull'integrazione tra simulazioni CFD e modelli surrogati, con l'obiettivo di rendere praticabile l'analisi di un elevato numero di configurazioni progettuali.

Il lavoro inizia con lo sviluppo di una pipeline automatizzata MATLAB–SolidWorks che consente di generare in modo parametrico le geometrie (pin-fin), impostare automaticamente il modello di simulazione in SolidWorks Flow Simulation, ed estrarre gli output termici di interesse.

Tale automazione è finalizzata alla costruzione di un dataset di simulazioni, utilizzato per addestrare e validare una rete neurale feed-forward in grado di predire la risposta termica del dissipatore a fronte delle variabili di progetto.

Una volta ottenuto il modello surrogato, questo viene impiegato all'interno di un processo di ottimizzazione multi-obiettivo per individuare configurazioni che bilancino prestazioni termiche e complessità geometrica, producendo un insieme di soluzioni di compromesso (fronte di Pareto).

L'intero metodo viene dapprima applicato a un caso semplificato, scelto per dimostrarne la fattibilità, e successivamente trasferito a un problema più realistico, in cui l'ottimizzazione riguarda la forma del singolo pin-fin.

1. Introduzione

La crescente diffusione di convertitori e azionamenti elettrici ad alta densità di potenza (ad esempio in trazione elettrica, rinnovabili e automazione industriale) impone vincoli sempre più stringenti sulla gestione termica: la miniaturizzazione e l'integrazione spinta rendono necessario dissipare quantità di calore elevate in volumi ridotti. La temperatura di esercizio è un fattore determinante per prestazioni e affidabilità dei dispositivi di potenza, e la letteratura evidenzia come il controllo della temperatura di giunzione e delle sollecitazioni termiche sia centrale per prevenire degradazione e guasti prematuri ^[1].

In questo scenario, i dissipatori a scambio convettivo rappresentano una soluzione consolidata, ma la scelta della configurazione (convezione naturale, forzata o raffreddamento a liquido) e della geometria non è banale: occorre bilanciare prestazioni termiche, ingombri, costi e, nei sistemi fluidodinamici, anche aspetti termo-idraulici (ad esempio perdite di carico e potenza di pompaggio). All'interno di questa classe di soluzioni, i dissipatori pin-fin hanno ricevuto notevole attenzione grazie alla loro capacità di fornire elevate superfici di scambio e una buona uniformità di raffreddamento; studi recenti ne discutono sistematicamente le prestazioni e il compromesso tra scambio termico ed energia richiesta dal fluido, mostrando come forma, passo e densità dei pin influenzino in modo significativo la risposta del sistema ^[2]. Inoltre, in ambito di raffreddamento a liquido, configurazioni con pin cilindrici e passo variabile sono state investigate anche con approcci avanzati di progettazione/ottimizzazione, evidenziandone potenziali vantaggi congiunti su scambio termico e prestazioni complessive ^[3].

Dal punto di vista metodologico, l'uso della CFD (Computational Fluid Dynamics) consente di analizzare con dettaglio il comportamento termo-fluidodinamico e di valutare molteplici varianti geometriche. Tuttavia, l'ottimizzazione "diretta" con CFD nel loop risulta spesso proibitiva per tempi computazionali. Per questo motivo, negli ultimi anni si sono diffusi approcci che combinano: generazione strutturata di campioni DoE (Design of Experiments), modelli surrogati addestrati su risultati CFD e ottimizzazione multi-obiettivo sul surrogato, ottenendo set di soluzioni con costi drasticamente inferiori rispetto ad analisi CFD iterative ^[4].

1.1 Obiettivi e contributi della tesi

Questo lavoro si colloca nell'ambito dei metodi di progettazione assistita da simulazione, dove l'esigenza principale è trasformare l'analisi termica da attività "manuale" e poco ripetibile a procedura sistematica, adatta a confrontare molte varianti progettuali in tempi compatibili con un iter di sviluppo. L'approccio integra:

1. Automazione della generazione geometrica e del setup di simulazione, finalizzata a produrre dataset numerici su larga scala;
2. Modello surrogato (rete neurale) addestrato su dataset CFD per predire rapidamente l'output termico;
3. Ottimizzazione multi-obiettivo sul surrogato per individuare il fronte di Pareto, riducendo in modo considerevole il costo computazionale rispetto all'ottimizzazione diretta su CFD.

In particolare, i principali contributi del lavoro sono:

- Una procedura di validazione dell'affidabilità dell'analisi CFD attraverso un caso studio confrontato con resistenza termica di datasheet;

- Lo sviluppo di un'interfaccia MATLAB–SolidWorks per: generazione parametrica dei pin, definizione automatica delle impostazioni, esecuzione batch delle simulazioni ed estrazione degli output;
- L'applicazione del metodo a un caso a complessità controllata (parametrizzazione a zone) per dimostrare la concretezza dell'approccio end-to-end (dataset → NN → ottimizzazione multi-obiettivo) e porre le basi per estensioni a problemi più realistici e complessi.
- La verifica della trasferibilità del metodo a un problema più applicativo, in cui l'ottimizzazione riguarda la forma del singolo pin-fin, con confronto finale rispetto a pin cilindrici tradizionali e discussione dei possibili meccanismi fisici alla base dei vantaggi osservati.

1.2 Struttura della tesi

La tesi è organizzata come segue.

Il capitolo 2 presenta i fondamenti teorici necessari: principi di trasferimento di calore, principali tipologie di dissipatori (convezione naturale, forzata e raffreddamento a liquido) e relativi sviluppi.

Il capitolo 3 descrive la procedura di validazione delle simulazioni CFD, basata sul confronto della resistenza termica ottenuta numericamente con i valori riportati nei datasheet dei costruttori, tramite un caso studio rappresentativo (dissipatore a liquido).

Il capitolo 4 illustra l'automazione sviluppata per integrare MATLAB e SolidWorks: generazione parametrica delle geometrie pin-fin, impostazione automatica delle simulazioni in SolidWorks Flow Simulation, gestione di esecuzioni batch, estrazione degli output. Riporta inoltre un tentativo di ottimizzazione “diretta” con simulazioni CFD in the loop, basato sul posizionamento di un numero limitato di cilindri, evidenziandone i limiti computazionali che hanno portato alla riformulazione del problema.

A partire da tali considerazioni, il Capitolo 5 sposta il *focus* dalla ricerca dell'ottimo in sé allo sviluppo e alla validazione di un metodo generale di ottimizzazione multi-obiettivo assistita da surrogato, fondato sulla generazione automatizzata di un dataset CFD, sull'addestramento di una rete neurale e sul successivo impiego di un algoritmo genetico per l'esplorazione del fronte di Pareto. Il metodo viene applicato a un caso di studio semplificato, scelto per consentire una verifica chiara e controllata della procedura prima di affrontare applicazioni di maggiore interesse pratico.

Il capitolo 6 estende il metodo a un problema più realistico, in cui l'ottimizzazione riguarda la forma del singolo pin-fin. In particolare, il *workflow* viene applicato a un nuovo spazio di progetto per verificarne la trasferibilità. I risultati ottenuti vengono analizzati anche mediante confronto con pin cilindrici tradizionali, così da valutarne la rilevanza progettuale e fornire una lettura fisica delle configurazioni ottimizzate.

Infine, il capitolo 7 conclude il lavoro con una riflessione complessiva sul metodo sviluppato, sui suoi limiti e sulle possibili direzioni di estensione futura.

2. Stato dell'arte e fondamenti

Sul piano del metodo di progettazione, lo stato dell'arte distingue tradizionalmente tre strumenti principali: correlazioni e modelli semplificati, modellazione numerica (CFD/CHT) e validazione sperimentale. Le correlazioni e i modelli a parametri concentrati (ad esempio reti di resistenze termiche) sono utili nella fase preliminare e per la lettura delle schede tecniche, perché consentono di stimare rapidamente l'ordine di grandezza delle temperature e di confrontare alternative. L'uso della CFD permette invece di analizzare campi tridimensionali di temperatura e flusso, valutando effetti geometrici non catturabili con semplicità dalle correlazioni (interazioni tra getti, scie, bypass, distribuzione non uniforme del flusso, effetti locali).

Tuttavia, la simulazione non può essere considerata una “black box”: i risultati devono essere interpretati criticamente, verificando coerenza con ordini di grandezza attesi e con i meccanismi fisici dominanti, ed eventualmente confrontati con correlazioni o dati sperimentali quando disponibili. In questo senso, rimane fondamentale una solida comprensione dei principi di trasferimento di calore alla base del funzionamento dei dissipatori (conduzione nel solido, convezione naturale/forzata e, quando rilevante, irraggiamento). Tale conoscenza consente non solo di impostare correttamente il modello, ma anche di riconoscere risultati non fisici o configurazioni apparentemente “migliori” che in realtà derivano da artefatti numerici, semplificazioni inappropriate o condizioni operative non realistiche. Inoltre, la familiarità con le principali tipologie di dissipatori e con i relativi trend di sviluppo in letteratura aiuta a comprendere quali scelte geometriche siano ragionevoli e quali trade-off (ad esempio prestazioni termiche vs perdite di carico/potenza ausiliaria) debbano essere considerati nella valutazione delle soluzioni.

Per queste ragioni, nelle sezioni seguenti vengono richiamati i principi di trasferimento di calore maggiormente rilevanti per i dissipatori (Sez. 2.1) e viene presentata una panoramica delle principali configurazioni impiegate in elettronica e dei relativi sviluppi (Sez. 2.2).

2.1 Principi di trasferimento di calore rilevanti per dissipatori

Un dissipatore realizza una catena di trasferimento termico dalla base di accoppiamento all'ambiente o al fluido di raffreddamento attraverso una combinazione di conduzione nel solido, convezione all'interfaccia solido-fluido e, in misura variabile, irraggiamento. In applicazioni elettroniche la convezione è spesso il termine dominante nello smaltimento verso l'esterno, ma la conduzione nella base e nelle alette determina quanto efficacemente la superficie estesa “viene alimentata” dal calore. È stato evidenziato come l'incremento delle prestazioni si ottenga tipicamente: aumentando l'area di scambio, aumentando il coefficiente convettivo (ad es. con flusso forzato o turbolenza/mescolamento), e riducendo le resistenze di conduzione e di contatto [5].

2.1.1 Conduzione

La conduzione governa la distribuzione del flusso termico nella base e lungo le alette/pin, di seguito ne vengono riportate le equazioni fondamentali.

La legge costitutiva è la legge di Fourier:

$$q = -k\nabla T \quad (1)$$

dove q è il vettore flusso termico $[W/m^2]$, k la conducibilità $[W/(m K)]$, T la temperatura.

Combinando Fourier con la conservazione dell'energia su un volume di controllo si ottiene la equazione del calore (forma generale):

$$\rho c \frac{\partial T}{\partial t} = \nabla \cdot (k \nabla T) + \dot{q}''' \quad (2)$$

dove ρ è la densità, c il calore specifico $[J/(kg K)]$ e $\dot{q}'''$ la generazione volumetrica di calore $[W/m^3]$.

L'equazione del calore da sola non basta a determinare il campo di temperatura, poiché descrive come la temperatura varia dentro il dominio, ma per avere una soluzione unica bisogna specificare come il dominio interagisce con l'esterno. Questo è il ruolo delle condizioni al contorno, che quindi "chiudono" il problema e rappresentano la fisica ai bordi.

Le condizioni al contorno tipiche sono:

1. Dirichlet (temperatura imposta):

Si assegna direttamente la temperatura sul bordo:

$$T = T_s \quad (3)$$

È il caso di una superficie a contatto con un termostato o con un corpo molto grande che mantiene T praticamente costante (serbatoio termico).

2. Neumann (flusso imposto):

Si assegna il flusso termico normale alla superficie:

$$-k \frac{\partial T}{\partial n} = q'' \quad (4)$$

Dove n è la normale uscente e q'' è il flusso termico per unità di area $[W/m^2]$. Include anche il caso adiabatico come caso particolare con $q'' = 0$, quindi:

$$\frac{\partial T}{\partial n} = 0 \quad (5)$$

(nessuno scambio di calore attraverso quel bordo).

3. Robin o "mista" (scambio convettivo):

Il flusso conduttivo al bordo è uguale allo scambio con un fluido per convezione:

$$-k \frac{\partial T}{\partial n} = h (T_s - T_\infty) \quad (6)$$

Rappresenta un bordo che scambia calore con un fluido a temperatura T_∞ (temperatura del fluido “lontano” dallo strato limite) con coefficiente di scambio termico convettivo $h [W/(m^2K)]$. È la condizione tipica per superfici esposte ad aria/acqua, dove la temperatura superficiale T_s non è nota a priori.

Molti problemi reali possono essere ricondotti a un modello quasi-1D se:

- il gradiente principale è lungo una sola coordinata (geometria “snella” o pareti sottili),
- la sezione di passaggio A è circa costante (o varia lentamente),
- k è costante o “equivalente” (media/effettiva).

In queste ipotesi, in regime stazionario e senza generazione interna di calore, il flusso è costante e si ottiene la classica resistenza termica:

(analogia con Ohm: $\Delta V = R I \Rightarrow \Delta T \leftrightarrow \Delta V, \dot{Q} \leftrightarrow I$)

$$R_{\text{cond}} = \frac{\Delta T}{\dot{Q}} \quad (7)$$

Dove $\dot{Q}$ è la potenza termica $[W]$, la resistenza si misura in $[K/W]$ e ΔT è la differenza di temperatura tra le due superfici del solido attraverso cui avviene la conduzione.

Casi notevoli sono:

- Parete piana (spessore L , area A):

$$R = \frac{L}{kA} \quad (8)$$

- Cilindro cavo (radiale, lunghezza L , raggi $r_1 \rightarrow r_2$):

$$R = \frac{\ln(r_2/r_1)}{2\pi kL} \quad (9)$$

La forza del modello a resistenze è che consente reti in serie/parallelo. Il limite è che non cattura deviazioni 2D/3D importanti (cambi bruschi di sezione “spreading/constriction resistance”, sorgenti localizzate, bordi “non uniformi”).

Quando il flusso non è 1D, un modo molto usato per stimare rapidamente la conduzione stazionaria e senza generazione, è l'introduzione del “conduction shape factor” S , che pur essendo utile per ottenere un modello compatto, è spesso difficile da determinare^[6] per geometrie reali e condizioni al contorno complesse; di conseguenza, l'uso di valori tabulati o stimati può introdurre incertezze significative e portare a risultati solo qualitativi o con accuratezza limitata.

2.1.2 Convezione

La convezione è lo scambio di calore tra una superficie solida e un fluido in moto (forzato o naturale). A scala macroscopica viene descritta tramite la legge di raffreddamento di Newton:

$$q'' = h (T_s - T_\infty) \Rightarrow \dot{Q} = h A (T_s - T_\infty) \quad (10)$$

I termini mantengono il significato già introdotto nel paragrafo sulla conduzione. Questa relazione non è ricavata direttamente dai principi di conservazione, ma costituisce una legge fenomenologica: il coefficiente h ingloba gli effetti del moto del fluido, della geometria e delle proprietà termo-fisiche, e viene in genere determinato tramite analisi dello strato limite, correlazioni adimensionali o dati sperimentali.

In linea di principio, una descrizione “completa” della convezione richiede: un modello per le proprietà termo-fisiche del fluido, la soluzione accoppiata delle equazioni di conservazione della massa, della quantità di moto (Navier–Stokes) e dell’energia (advezione–diffusione), eventualmente coniugate con la conduzione nel solido; le correlazioni per ricavare h e l’Eq. (10) possono essere viste come una formulazione ingegneristica di tale problema.

Prima di introdurre le correlazioni per il coefficiente di scambio termico convettivo, è opportuno distinguere tra convezione naturale e convezione forzata, poiché il meccanismo fisico che genera il moto del fluido è diverso. Nella convezione forzata il flusso è imposto da dispositivi esterni (ventole, pompe), e l’intensità dello scambio dipende principalmente dalla velocità e dal regime di moto; nella convezione naturale il moto è invece indotto dalle forze di galleggiamento dovute a variazioni di densità causate dal gradiente termico. Di conseguenza cambiano i parametri adimensionali che governano il fenomeno.

Per rendere generalizzabile il problema si usano i numeri adimensionali.

- Numero di Nusselt

$$\text{Nu} = \frac{hL}{k} \Rightarrow h = \frac{k}{L} \text{Nu} \quad (11)$$

dove: L è una lunghezza caratteristica (es. lunghezza piastra, diametro cilindro, diametro idraulico nei canali) [m]; k è la conducibilità termica del fluido [$W/(m K)$].

Significato fisico di Nu: esprime il rapporto tra il flusso di scambio termico convettivo e quello per conduzione nel fluido. In pratica è un indice dell’efficacia dello scambio convettivo.

- Numero di Prandtl

$$\text{Pr} = \frac{\mu c_p}{k} = \frac{\nu}{\alpha} \quad (12)$$

dove: c_p è il calore specifico a pressione costante [$J/(kg K)$], $\alpha = k/(\rho c_p)$: diffusività termica [m^2/s] del fluido, μ è la viscosità dinamica [$Pa s$], $\nu = \mu/\rho$ è la viscosità cinematica [m^2/s].

Significato fisico di Pr: confronto tra diffusione di quantità di moto e diffusione termica; determina il rapporto tra spessori dello strato limite di velocità e di temperatura (es. aria $\text{Pr} \sim 0.7$, acqua ~ 7).

Questo concetto è illustrato in Figura 1: δ indica lo spessore dello strato limite di velocità, cioè la regione adiacente alla parete in cui il profilo di velocità è significativamente influenzato dalla presenza della parete; δ_T rappresenta invece lo spessore dello strato limite termico, definito in modo analogo per il campo di temperatura.

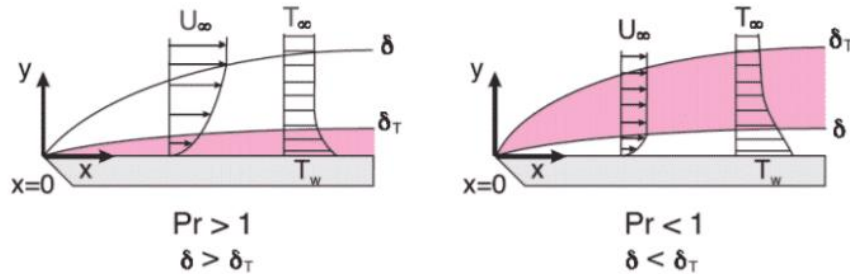


Figura 1 ^[7]: Rappresentazione del significato fisico di Pr

- Numero di Reynolds (per convezione forzata)

$$Re = \frac{\rho UL}{\mu} = \frac{UL}{\nu} \quad (13)$$

dove: U è la velocità caratteristica $[m/s]$.

Significato fisico di Re : rapporto tra forze inerziali e viscosi; governa il regime del moto (laminare/transitorio/turbolento) e quindi l'intensità dello scambio, molte correlazioni sono del tipo $Nu = f(Re, Pr, \dots)$.

- Numero di Grashof (per convezione naturale)

$$Gr = \frac{g \beta (T_s - T_\infty) L^3}{\nu^2} \quad (14)$$

Dove: g è l'accelerazione di gravità, β è il coefficiente volumetrico di dilatazione termica $[1/K]$ (per gas ideali viene spesso stimato come $1/T_{media,gas}$).

Significato fisico di Gr : analogo di Reynolds per la convezione naturale; misura il rapporto tra forze di galleggiamento e forze viscosi.

- Numero di Rayleigh

$$Ra = Gr Pr \quad (15)$$

Significato fisico di Ra : parametro chiave che governa l'intensità della convezione naturale e l'eventuale transizione di regime; molte correlazioni sono del tipo $Nu = f(Ra, Pr, \dots)$.

Il coefficiente h si ottiene quasi sempre da correlazioni per Nu , perché la soluzione analitica completa delle equazioni di moto + energia è disponibile solo in pochi casi ideali.

Alcuni esempi di correlazioni sono:

- Dittus–Boelter (turbolento, condotti lisci; forma classica):

$$Nu = 0.023 Re^{0.8} Pr^n \quad (16)$$

con $n \approx 0.4$ per fluido riscaldato e $n \approx 0.3$ per fluido raffreddato (scelta dipende dalla convenzione e dal verso del trasferimento). La correlazione è storicamente attribuita al lavoro originale su radiatori tubolari [8].

- Correlazioni per piastra piana:

$$\bar{Nu}_L = 0.664 Re_L^{1/2} Pr^{1/3} \text{ valida per } Re_L < 10^5 \quad (17)$$

$$\bar{Nu}_L = (0.037 Re_L^{4/5} - 871) Pr^{1/3} \text{ valida per } 10^5 \leq Re_L \leq 10^8 \quad (18)$$

- Mc Adams (convezione naturale, parete verticale e cilindri verticali)

$$\bar{Nu}_L = 0.59 Ra_L^{1/4} \text{ valida per } Ra_L < 10^9 \quad (19)$$

$$\bar{Nu}_L = 0.13 Ra_L^{1/3} \text{ valida per } Ra_L > 10^9 \quad (20)$$

Come per la conduzione anche in questo caso si può ricavare una resistenza convettiva equivalente (regime stazionario). In molti modelli si usa la resistenza:

$$R_{\text{conv}} = \frac{1}{hA} \Rightarrow \dot{Q} = \frac{T_s - T_\infty}{R_{\text{conv}}} \quad (21)$$

che si combina in serie/parallelo con conduzione e (eventuali) resistenze di contatto. La resistenza convettiva equivalente è utilissima nei modelli compatti, ma ha limiti chiari ed è “diversa” dalla resistenza conduttiva perché non è una proprietà puramente del materiale/geometria, bensì dipende dal fluido e dal moto. Dipendendo da accoppiamenti termo-fluidodinamici h può variare nello spazio e con il punto di funzionamento (portata, regime, proprietà del fluido), limitando l'accuratezza dei modelli compatti quando il campo termo-fluidodinamico risulta fortemente non uniforme o accoppiato. Inoltre, in applicazioni reali non è sempre immediato individuare una correlazione pienamente applicabile: le correlazioni sono infatti derivate per geometrie ideali e intervalli di validità specifici (regime laminare/turbolento, condizioni di ingresso, parete a temperatura o flusso imposto, ecc.). Ne consegue che esiste il rischio di selezionare una correlazione non appropriata o di applicarla fuori dal proprio campo di validità, introducendo errori anche significativi. Anche quando la correlazione è corretta, il valore di h ottenuto rappresenta in genere una stima media (di ordine ingegneristico) e non un valore “esatto”, poiché la convezione reale è influenzata da non uniformità locali, tolleranze geometriche e condizioni operative difficili da modellare in modo perfettamente aderente.

Va infine osservato, che, non è sempre immediato definire univocamente la differenza di temperatura ΔT da associare allo scambio convettivo. Nel caso di superfici alettate (tipico dei

dissipatori), l'incremento di area non si traduce automaticamente in un aumento proporzionale dello scambio termico (come suggerirebbe l'Eq. 21); lungo l'aletta si instaura infatti una caduta di temperatura dovuta alla conduzione interna, per cui non tutta la superficie alettata lavora alla stessa differenza ($T_s - T_\infty$). A tal proposito, per valutare in modo quantitativo l'utilità delle superfici alettate si introducono due grandezze complementari: l'efficacia dell'aletta ε_f , che indica se conviene inserire l'aletta confrontando il flusso termico scambiato dalla superficie alettata con quello che si otterrebbe dalla sola superficie piana di base, e l'efficienza dell'aletta η_f , che misura quanto l'aletta riesca ad avvicinarsi al comportamento ideale. In particolare, l'ideale corrisponde al caso in cui l'intera aletta si trovi a temperatura uniforme, pari a quella alla radice (base), così che tutta la sua superficie lavori alla massima differenza di temperatura rispetto al fluido. Le espressioni analitiche di η_f ed ε_f sono ricavabili sotto ipotesi semplificate e rappresentano quindi un criterio di stima utile ma idealizzato.

Riassumendo, quindi, la valutazione dello scambio convettivo può essere affrontata secondo tre livelli di modellazione:

1. Correlazioni e modelli 1D: si adotta una correlazione $Nu = f(Re, Pr, \dots)$, si ricava $h = \frac{k}{L} Nu$ e si usa R_{conv} (approccio rapido, ma approssimato: fornisce stime globali/medie, non cattura le non uniformità locali, ed è vincolato ai range di validità).
2. Modelli di strato limite: si risolvono forme semplificate delle equazioni di quantità di moto ed energia nello strato limite vicino alla parete (più fisico, ma applicabile soprattutto a geometrie/regimi ideali).
3. CFD/CHT: si risolvono numericamente le equazioni di Navier–Stokes e dell'energia, eventualmente accoppiate alla conduzione nel solido (Conjugate Heat Transfer), ottenendo campi termo-fluidodinamici 3D (più generale, ma computazionalmente più oneroso).

2.1.3 Irraggiamento

L'irraggiamento è lo scambio di energia tramite onde elettromagnetiche emesse dalle superfici a temperatura T . A differenza della conduzione e della convezione, non richiede un mezzo materiale (avviene anche nel vuoto). Ogni corpo con temperatura superiore allo zero assoluto ($T > 0$ K) emette radiazione termica; tuttavia, lo scambio termico rilevante è il flusso netto, che dipende dalla differenza tra la radiazione emessa dalla superficie e quella assorbita dall'ambiente circostante, e quindi dalle rispettive temperature.

Nella modellazione termica dei dissipatori per elettronica, il contributo dell'irraggiamento è stato spesso trascurato, sia perché talvolta secondario rispetto alla convezione forzata, sia soprattutto per la difficoltà di stimarne correttamente l'entità in geometrie reali. In particolare, l'inclusione della radiazione introduce non linearità (dipendenza in T^4), aumentando sensibilmente la complessità del modello ^[9].

Nel presente lavoro l'irraggiamento non costituisce un meccanismo dominante (raffreddamento a liquido) e, non essendo un punto focale dei capitoli successivi, non viene approfondito oltre. È tuttavia opportuno considerarlo (o quantificarne l'ordine di grandezza) nei casi in cui può diventare comparabile con la convezione, ad esempio: temperature superficiali elevate, convezione debole (convezione naturale, aria ferma, spazi confinati, vuoto), superfici ampie e ad alta emissività (vernici opache/anodizzazioni), oppure quando la superficie ha un elevato fattore di vista verso un ambiente significativamente più freddo. In tali condizioni l'irraggiamento può contribuire in modo non trascurabile allo scambio complessivo e può valere la pena approfondirne la modellazione ^[10].

2.2 Tipologia di dissipatori per elettronica

Nella gestione termica dell'elettronica, il termine “dissipatore” include una famiglia di soluzioni che hanno lo stesso obiettivo: ridurre la resistenza termica complessiva tra sorgente e ambiente (o tra sorgente e un fluido di raffreddamento), rispettando vincoli di ingombro, massa, costo (anche dei materiali), affidabilità e rumorosità. In prima approssimazione, le architetture più comuni possono essere ricondotte a tre categorie: dissipatori ad aria in convezione naturale, dissipatori ad aria in convezione forzata, e dissipatori a liquido (cold plate/microcanali, ecc.).

In Figura 2 è mostrato uno schema tipico di montaggio di un dissipatore su un modulo IGBT (Insulated Gate Bipolar Transistor), evidenziando la stratigrafia del pacchetto (chip, saldature, substrato, DCB: Direct Copper Bonded, baseplate) e l'interfaccia verso il dissipatore tramite pasta/grasso termico. Le grandezze Z_{jc} rappresentano le impedenze termiche lungo il percorso di conduzione dalla giunzione al case (spesso modellate come resistenze equivalenti, si vedano l'Eq. 7,8 del paragrafo sulla conduzione, talvolta anche dinamiche); mentre R_{ch} descrive la resistenza di contatto case-dissipatore attraverso il materiale d'interfaccia. Infine, Z_h indica l'impedenza termica del dissipatore, cioè il contributo complessivo del dissipatore (heat sink) nello smaltimento del calore verso l'ambiente o il fluido di raffreddamento (per Z_h fare riferimento anche all'Eq. 21 del paragrafo sulla convezione).

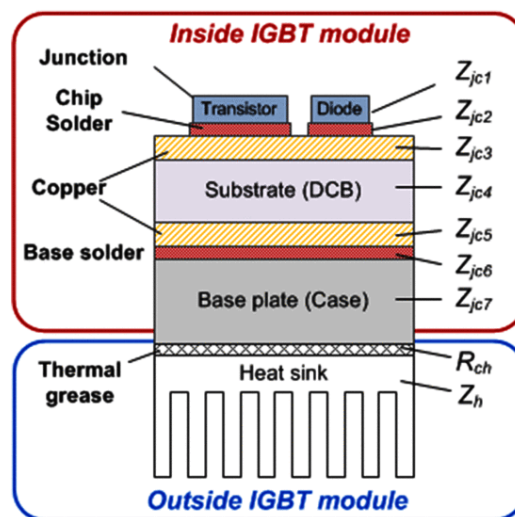


Figura 2 ^[11]: Schema della catena di smaltimento termico

- Convezione naturale (passiva)

È la soluzione più semplice e affidabile: non richiede componenti attivi (ventole/pompe), quindi offre vantaggi su rumore, manutenzione e failure rate ^[12], ma soffre di coefficienti di scambio relativamente bassi e di una forte sensibilità a orientamento e ostacoli al moto convettivo (sensibilità che può essere ridotta adottando dissipatori a pin-fin, generalmente meno direzionali e quindi più “omnidirezionali” rispetto ai plate-fin tradizionali ^[13]). Per questo è tipica per dissipare potenze moderate e in ambienti dove silenziosità/affidabilità sono prioritari.

- Convezione forzata (aria con ventola/soffiante)

A parità di ingombro, l'aria forzata consente in genere h più elevati rispetto alla naturale, ma introduce costi/complessità (ventole, filtri, vibrazioni, rumore) e soprattutto un vincolo progettuale essenziale: la perdita di carico del dissipatore deve essere compatibile con la curva della ventola, come rappresentato in Figura 3.

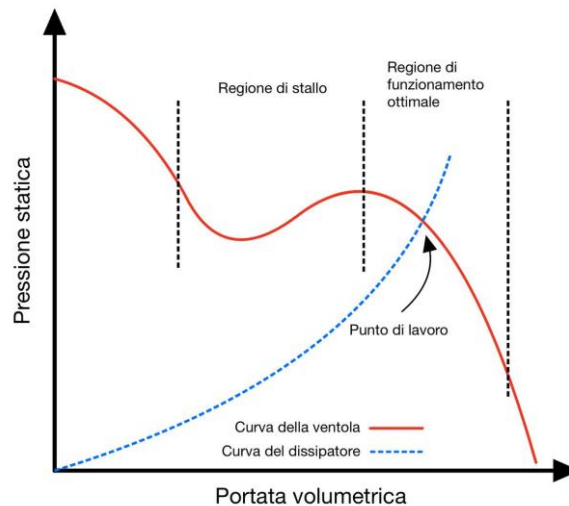


Figura 3: Un esempio di curva caratteristica di una ventola

Dunque, oltre alla stima del coefficiente di scambio h tramite correlazioni, è comune stimare anche la perdita di carico Δp mediante correlazioni analoghe basate sul fattore d'attrito $f = f(Re, \dots)$, così da collegare rapidamente geometria $\leftrightarrow h \leftrightarrow \Delta p$ nelle fasi preliminari di progetto^[14].

- Raffreddamento a liquido

Quando il carico termico o la densità di potenza crescono oltre ciò che l'aria può gestire con limiti accettabili di rumorosità e ingombro, la soluzione a liquido diventa spesso preferibile: maggiore capacità termica del fluido e possibilità di configurazioni molto compatte (cold plates, microcanali). Per contro aumentano complessità, rischi e costi (pompa, tenute, potenziali perdite, manutenzione). In questo ambito è utile distinguere due architetture principali che si vedranno nei capitoli successivi: cold plate "separato" e raffreddamento diretto/integrato nel modulo. Nel caso del cold plate, il dissipatore a liquido è un componente esterno (una piastra con canali interni) accoppiato meccanicamente al modulo tramite un'interfaccia termica (TIM) e, spesso, una *baseplate*: il calore attraversa quindi più strati solidi e l'interfaccia prima di raggiungere il fluido. Nella Figura 4 è riportata l'immagine di uno scambiatore di calore (cold plate). Nel caso di modulo a raffreddamento diretto/integrato (direct cooling), invece, la geometria di scambio (ad es. pin-fin, ribs o canali) è realizzata direttamente sul lato posteriore del modulo e viene bagnata dal refrigerante, riducendo o eliminando alcuni strati intermedi (ad esempio baseplate/TIM) e quindi potenzialmente la resistenza termica^[15]; di contro, questa soluzione richiede maggiore integrazione tra design termico e packaging, e introduce vincoli più stringenti su tenute, compatibilità dei materiali e producibilità/affidabilità del modulo. Un esempio è presentato in Figura 5.

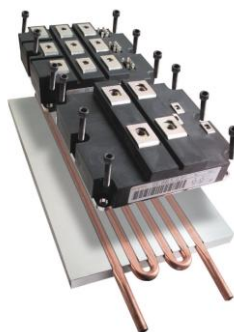


Figura 4: Scambiatore di calore (cold plate), (Riferimento: Aavid, divisione termica di Boyd Corporation)



Figura 5: Raffreddamento diretto (direct cooling). HybridPACK™ Drive Power Module (a sinistra, fronte e retro) e alloggiamento (cooler) con modulo IGBT montato in trasparenza (a destra), (Riferimento: adattato da Infineon Technologies AG, “Assembly Instructions for the HybridPACK™ Drive Performance”)

Nel paragrafo successivo verrà presentata una panoramica delle geometrie di dissipatori impiegate nelle tre classi considerate, insieme a un inquadramento dell’evoluzione delle metodologie di progetto adottate per l’ottenimento e l’ottimizzazione di tali geometrie. Va specificato che non esiste una scelta “migliore” in senso assoluto, né tra le tre categorie sopra citate né tra le diverse soluzioni all’interno di ciascuna di esse: la soluzione ottimale dipende dai requisiti di progetto e dalla funzione obiettivo che si intende privilegiare (ad esempio minimizzare la temperatura massima, limitare massa e ingombri, contenere la perdita di carico e la potenza di pompaggio/ventilazione, ridurre costi, rumore e manutenzione). Di conseguenza, il confronto deve essere impostato come un problema di ottimizzazione multicriterio, in cui prestazioni termiche, vincoli fluidodinamici e aspetti produttivi vengono bilanciati in modo coerente con l’applicazione. In tutti e tre gli approcci descritti la leva progettuale principale rimane, nella sostanza, la stessa: definire la geometria delle superfici estese (alette/pin-fin/canali) e la loro disposizione spaziale rispetto alla sorgente di calore e al percorso del fluido per ottimizzare il requisito desiderato.

2.2.1 Evoluzione delle geometrie e dei metodi di progettazione

In modo trasversale alle tre tipologie, la progettazione dei dissipatori è passata da:

1. modelli compatti e correlazioni per il predimensionamento e l’analisi parametrica (approccio storico, tuttora impiegato nelle fasi preliminari di progetto);
2. simulazioni CFD/CHT per catturare effetti 3D (bypass, ricircoli, non uniformità di portata/temperatura) e validare le scelte (fase moderna);
3. ottimizzazione automatica (algoritmi di ottimizzazione per stabile dimensioni/forma e layout delle superfici estese), più recentemente, topology optimization (TO) che non impone a priori una geometria predefinita, e approcci data-driven/surrogati quando lo spazio di progetto cresce e/o le simulazioni diventano costose (sviluppi promettenti) ^[16].

La diffusione della TO è stata favorita dai recenti progressi nella fabbricazione additiva, che rende realizzabili geometrie interne complesse e topologie “free-form” difficilmente ottenibili con processi tradizionali; inoltre, l’AM sta abilitando l’impiego di materiali alternativi ai metalli (ad es. polimeri/coating e ceramiche), utili in condizioni corrosive o ad alta temperatura, ampliando ulteriormente lo spazio di progetto e di ricerca ^[17].

Di seguito vengono riportate alcune immagini (vedi esempi in Figura 6) che illustrano l’evoluzione dei dissipatori, dalle configurazioni più “classiche” a soluzioni moderne, evidenziando le principali motivazioni che hanno guidato tale sviluppo: aumento delle densità di potenza, vincoli di ingombro e integrazione, controllo delle perdite di carico/potenza di pompaggio e, non da ultimo, riduzione della massa e del materiale impiegato (aspetto rilevante sia per costi sia per applicazioni weight-sensitive).

- Convezione naturale (passiva).

In convezione naturale il limite è spesso l'intensità del moto convettivo stesso: la geometria deve favorire la formazione e la risalita dei pennacchi convettivi e ridurre l'interferenza tra strati limite. Di conseguenza, orientazione, spaziatura e forma delle alette sono variabili critiche (progettazione "classica", parametrica); inoltre, il progetto deve bilanciare l'aumento di superficie con la necessità di non "strozzare" il moto dell'aria e con l'incremento di massa dal materiale alettato.

L'articolo di Alexandersen et al. ^[18] mostra che, usando l'ottimizzazione topologica con un modello termo-fluidodinamico completo, è possibile generare automaticamente dissipatori 3D in convezione naturale con geometrie ramificate non convenzionali (Figura 6, immagine a destra). Al crescere del numero di Grashof le soluzioni ottimali diventano più compatte e multi-ramificate, migliorando lo scambio termico grazie ad una maggiore superficie efficace e a geometrie che accelerano localmente il moto del fluido. Inoltre, le topologie ottenute confermano conclusioni già note in letteratura riguardo ai rapporti lunghezza/spessore delle alette (tassi più bassi di convezione naturale richiedono poche alette più lunghe, mentre tassi più elevati un numero maggiore di alette più corte), ma indicano anche che geometrie complesse e opportunamente "ritagliate" possono migliorare in modo significativo il raffreddamento rispetto a semplici geometrie alettate.



Figura 6: Dissipatore tradizionale ad alette parallele (plate-fin, a sinistra) e dissipatore topologicamente ottimizzato per convezione naturale (esempio accademico, a destra) ^[18]

- Convezione forzata (aria con ventola/soffiante).

In convezione forzata la prestazione non è legata soltanto al coefficiente di scambio h : la geometria deve risultare compatibile con un budget di pressione e con la potenza di ventilazione disponibile. L'evoluzione delle geometrie è quindi guidata dal compromesso tra incremento dello scambio termico e contenimento di Δp , insieme a vincoli pratici quali rumorosità, affidabilità e riduzione di massa a parità di prestazione. In Figura 7 si riporta l'immagine di un dissipatore topologicamente ottimizzato per convezione forzata ottenuto nello studio di Zeng et al. ^[19].

Il vantaggio del TO sopra citato deriva soprattutto dal raffreddamento nella zona a valle, ottenuto mediante interruzione ripetuta dello strato limite e mixing, mentre nel canale rettilineo classico (Figura 6 a sinistra: Dissipatore tradizionale ad alette parallele) lo scambio degrada progressivamente con lo sviluppo degli strati limite. Rispetto a un dissipatore convenzionale, mostra una migliore capacità di scambio termico con una penalizzazione della caduta di pressione ritenuta accettabile. In termini quantitativi, il dissipatore TO consente un risparmio del 54,9% della potenza di pompaggio mantenendo la stessa temperatura del dissipatore classico a parità di potenza termica applicata.



Figura 7^[19]: Dissipatore topologicamente ottimizzato per convezione forzata

- Raffreddamento a liquido.

Il raffreddamento a liquido consente di gestire densità di potenza più elevate e soluzioni più compatte, ma introduce vincoli stringenti su portata, pompaggio e distribuzione del flusso nei canali, oltre alla necessità di limitare la perdita di carico. In presenza di sorgenti multiple o hotspot, diventa inoltre importante garantire uniformità termica e ridurre fenomeni di maldistribuzione ^[16]. Per questo l'ottimizzazione non riguarda solo la micro-geometria (canali/pin-fin), ma anche il manifolding e la configurazione globale dei passaggi, cercando al contempo di minimizzare il materiale impiegato e quindi peso e costo del sistema.

Vengono proposti due esempi, uno di ottimizzazione di cold plates, e uno per il direct cooling a pin-fin.

1. Cold plates

Nella ricerca di Chen et al. ^[20], l'adozione di configurazioni ottenute tramite ottimizzazione topologica consente di superare i limiti delle geometrie tradizionali a canali rettilinei (RCP, Rectangular-Channel Cold Plate) o serpentine (SCP, Serpentine-Channel Cold Plate), perché la TO non impone a priori un pattern di canali ma distribuisce il dominio fluido/solido per migliorare contemporaneamente uniformità termica e prestazioni idrauliche. In Figura 8 è riportato un confronto tra le geometrie convenzionali RCP e SCP e le corrispondenti configurazioni derivate da ottimizzazione topologica. Le soluzioni TCP (Topology-optimized Cold Plate) proposte mostrano una distribuzione di temperatura più uniforme rispetto alle configurazioni convenzionali: a parità di pressione di ingresso, la riduzione della non-uniformità è marcata, mentre la temperatura massima rimane circa costante (diminuzione dell'ordine di ~1% o meno). Dal punto di vista fluidodinamico, le TCP risultano anche più favorevoli perché presentano minore resistenza al flusso (e quindi maggiore portata/velocità a parità di pressione) e, allo stesso tempo, un coefficiente di scambio termico più elevato. In sintesi, rispetto a RCP e SCP, le geometrie TCP migliorano l'efficacia del raffreddamento soprattutto grazie a una migliore gestione globale del percorso, che riduce la maldistribuzione e rende più efficace l'asportazione del calore sull'intero componente.

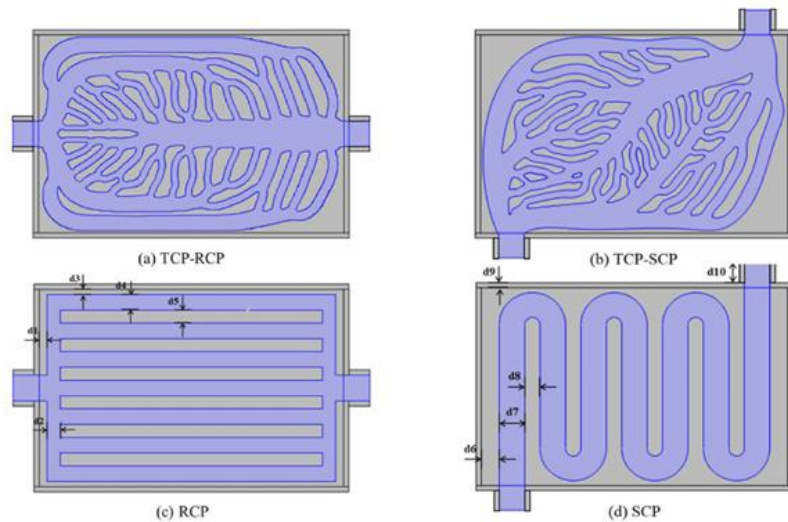


Figura 8 ^[20]: Ottimizzazione topologica dei canali nei cold plates

2. Pin-Fin

Per i dissipatori a pin-fin, l'articolo di Nguyen et al. ^[21] presenta un approccio che mostra alcune analogie con il presente lavoro. Nel loro studio, la configurazione dell'array è mantenuta fissa e l'ottimizzazione agisce sulla forma del singolo pin-fin, parametrizzata tramite sette variabili di input. In Figura 9 è possibile osservare la definizione delle sette variabili (X_1 – X_7) che determinano la geometria del pin-fin, insieme al confronto tra il pin cilindrico convenzionale e quello ottimizzato; la figura riporta inoltre le mappe di temperatura sulla superficie di dissipazione ottenute con le due configurazioni, evidenziandone l'impatto sulle prestazioni termiche. Il workflow dello studio segue uno schema "CFD → modello surrogato → ottimizzazione": i dati generati da simulazioni vengono usati per addestrare un regressore ML capace di predire rapidamente (nell'ordine di ~ 1 secondo) le metriche di prestazione termo-idraulica (temperatura e caduta di pressione). Il surrogato consente la rapida esplorazione del design space, permettendo di accoppiare l'ottimizzazione ad un algoritmo genetico (GA) con popolazioni più grandi e quindi una ricerca più efficace. Il risultato principale è l'identificazione di una geometria non convenzionale (pin-fin "a imbuto") che a parità di condizioni di lavoro migliora lo scambio ($\approx +20\%$ di coefficiente di scambio termico) senza un incremento evidente della penalità di Δp rispetto al cilindro standard. Viene dunque mostrato come l'accoppiamento surrogato ML + ottimizzatore sia particolarmente adatto quando l'obiettivo è cercare forme nuove in spazi di progetto ampi e costosi da esplorare via CFD diretta.

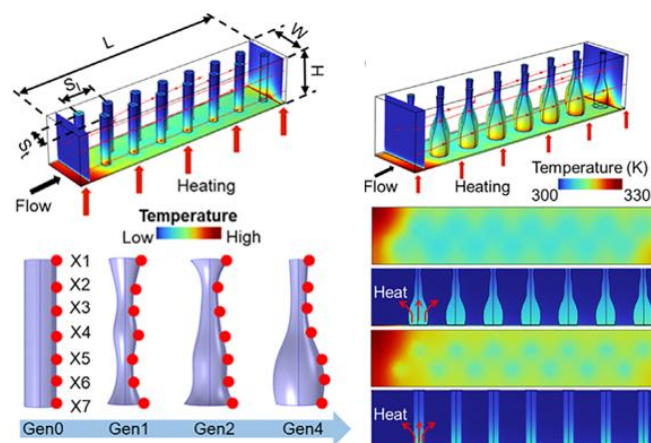


Figura 9 ^[21]: Ottimizzazione geometria dei Pin-Fin mediante modelli surrogati e algoritmi genetici

L'analogia con il presente lavoro va tuttavia intesa in senso generale, cioè con riferimento alla struttura del *workflow*, fondata sull'integrazione tra simulazioni CFD, modello surrogato e ottimizzazione. Le modalità con cui tali passaggi sono stati realizzati nella presente tesi sono infatti state sviluppate autonomamente. Va inoltre precisato che l'unico elemento effettivamente ripreso dallo studio di Nguyen et al. ^[21] è la parametrizzazione geometrica del singolo pin-fin. In una prima fase della tesi, infatti, l'ottimizzazione è stata applicata a un problema differente, riguardante la distribuzione dei pin sulla superficie del dissipatore; solo in un secondo momento la parametrizzazione a sette variabili qui proposta è stata assunta come spunto per estendere il *workflow* a un nuovo problema progettuale, di carattere più applicativo.

3. Validazione SolidWorks Flow Simulation

Prima di utilizzare SolidWorks Flow Simulation come strumento principale per le simulazioni CFD necessarie a questa tesi, è stata svolta una fase di validazione preliminare con lo scopo di verificare l'attendibilità dei risultati ottenibili tramite il software.

L'obiettivo della validazione è stato controllare che l'impostazione della simulazione produca risultati fisicamente coerenti, confrontando la resistenza termica di un dissipatore a liquido (cold plate) calcolata in simulazione con quella riportata nel datasheet del produttore, al variare della portata.

Questa fase ha permesso di aumentare la confidenza nell'utilizzo di SolidWorks Flow Simulation per la generazione del dataset e per le analisi comparative svolte nei capitoli successivi.

Come caso studio è stato selezionato un dissipatore commerciale per cui era già disponibile il modello CAD (Computer Aided Design) fornito dal produttore (Boyd Corporation): 416201U00000G – Unfinished, Standard 305 mm Plate Length, di cui si riporta l'immagine in Figura 10.



Figura 10: 416201U00000G – Unfinished, Standard 305 mm Plate Length, (Riferimento: Boyd-Hi-Contact-6-Pass-Datasheet)

Sebbene le simulazioni oggetto della tesi siano svolte su dissipatori a direct cooling, questa differenza non è critica nella fase di validazione: lo scopo qui non è confrontare specifiche soluzioni progettuali, ma valutare l'affidabilità del software e del workflow di simulazione.

Occorre fare una precisazione prima di passare alla sezione successiva, SolidWorks Flow Simulation mette a disposizione una serie di linee guida operative (SolidWorks Flow Simulation User Guide) e tutorial (SolidWorks Flow Simulation Tutorial) per impostare correttamente le simulazioni termofluidodinamiche, con l'obiettivo di ridurre errori comuni e garantire risultati numericamente stabili e fisicamente realistici. Le indicazioni comprendono esempi applicativi che coprono scenari industriali tipici, inclusi casi legati al raffreddamento di componenti elettronici. Nel presente lavoro, tali linee guida e tutorial sono stati seguiti come riferimento metodologico per l'impostazione delle simulazioni.

Di conseguenza, non viene riportata nel dettaglio la procedura “passo per passo” di creazione del progetto termico (impostazioni del wizard, selezione dei menu e operazioni di interfaccia), ci si concentra invece su ciò che è rilevante dal punto di vista ingegneristico e riproducibile, ovvero:

- Fenomeni di scambio termico considerati
- Direzione della gravità
- Materiali (solidi)

- Tipo di fluido
- Portata del fluido
- Moto del fluido in ingresso
- Temperatura di ingresso del fluido
- Condizioni al contorno dell'intero dissipatore
- Potenza termica applicata
- Definizione dell'area riscaldata
- Considerazioni sulla Mesh
- Criteri di arresto della simulazione

Questi fattori sono definiti in accordo con i dati disponibili nel datasheet, così da ottenere le grandezze di interesse e rendere possibile il confronto dei risultati. L'elenco sopra riportato sarà fondamentale anche nelle simulazioni presenti nei capitoli successivi.

3.1 Setup della simulazione CFD

Il dissipatore appena presentato (Figura 10) è stato importato in SolidWorks e utilizzato per impostare un'analisi termica.

Di seguito vengono definiti i fattori, categorizzandoli in base al metodo con cui sono stati impostati.

1. Fattori da datasheet del dissipatore (Boyd-Hi-Contact-6-Pass-Datasheet):
 - Materiali: base in alluminio, tubi in rame
 - Fluido: Acqua
 - Portata del fluido (volumetrica, Q) variabile nelle diverse simulazioni: 0,5-1-1,5-2-2,5-3,5 galloni al minuto (GPM), viene mantenuta l'unità di misura del datasheet per fare confronti rapidi, per convertire da GPM a m^3/s basta moltiplicare per $6,309 \times 10^{-5}$.
2. Fattori che dipendono da ipotesi fisiche del problema:
 - Fenomeni di scambio termico considerati: conduzione e convezione, non viene incluso l'irraggiamento in quanto trascurabile, ancor di più considerando lo scopo delle simulazioni in questa fase.
 - Direzione della gravità: direzione y, verso negativo, si veda Figura 11 per comprendere la direzione relativa al dissipatore.
 - Moto del fluido in ingresso: completamente sviluppato, si immagina ci sia una pompa all'ingresso del dissipatore.
 - Temperatura di ingresso del fluido (T_{in} , media di massa): 35 °C, valore scelto assumendo che il circuito esterno di raffreddamento non riesca a riportare l'acqua alla temperatura ambiente.
 - Condizioni al contorno dell'intero dissipatore: pareti adiabatiche (fatta eccezione per l'area riscaldata), viene trascurato lo scambio termico del dissipatore con l'aria, perché ciò che interessa è valutare l'efficacia con cui il liquido di raffreddamento asporta il calore.
 - Potenza termica applicata (ϕ): 6000 W, una stima ragionevole considerando le dimensioni del dissipatore e la tipologia (dissipatore a liquido).
3. Fattori derivati da considerazioni sulle simulazioni

- Considerazioni sulla Mesh: SolidWorks Flow Simulation identifica in automatico la dimensione più piccola che deve essere discretizzata dalla Mesh globale in base a geometria, condizioni al contorno e altri elementi di progetto. Per queste simulazioni la Mesh automatica è risultata soddisfacente, infatti il foro del tubo, che è la geometria più piccola, è risolto con molte celle di fluido come mostrato in Figura 11. Perciò non si è andati a modificare alcun parametro lasciando quelli di default e identificati dal software.

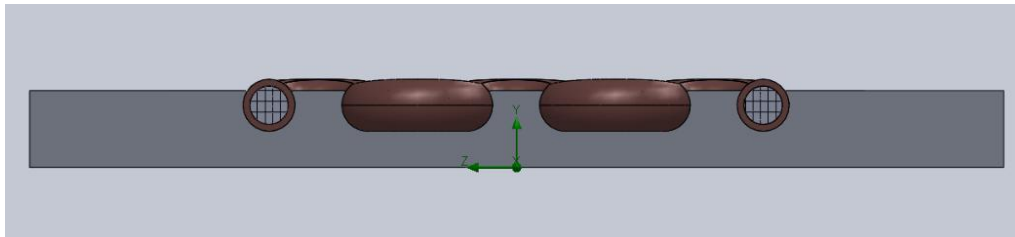


Figura 11: Modello CAD del dissipatore, visuale sulla Mesh all'ingresso e all'uscita dei tubi

- Criteri di arresto della simulazione: SolidWorks Flow Simulation considera inizialmente qualsiasi problema di flusso in regime stazionario come un problema dipendente dal tempo; per questo motivo è opportuno definire dei Goals (grandezze di interesse) la cui convergenza possa essere assunta come criterio di arresto (raggiungimento del regime stazionario). Nel presente lavoro sono stati impostati come Goals: la temperatura media sulla superficie di interesse (superficie riscaldata, Temperature Average T_{av}), e la temperatura del fluido in uscita (T_{out} , media di massa). In particolare, la temperatura di uscita del fluido consente di verificare indirettamente il raggiungimento del regime attraverso il bilancio energetico: a regime, la potenza termica immessa nel sistema deve risultare coerente con la potenza asportata dal fluido (tutte le altre pareti sono adiabatiche per imposizione), stimabile dalla variazione di temperatura tra ingresso e uscita (a portata nota).
- Definizione dell'area riscaldata: qui sta il fulcro della validazione; perciò, verrà discussa in dettaglio nella sezione successiva.

Quanto appena descritto viene riassunto in Tabella 1.

Setup CFD	
Fenomeni di scambio termico	Conduzione, Convezione
Gravità	Direzione: y Verso: negativo
Materiali	Base: Alluminio Tubi: Rame
Fluido	Acqua
Q	0,5-1-1,5-2-2,5-3,5 GPM
Moto del fluido in ingresso	Completamente sviluppato
T_{in}	35 °C
Condizioni al contorno	Pareti adiabatiche
ϕ	6000 W
Mesh	Default di SolidWorks
Goals	T_{av} , T_{out}

Tabella 1: Setup CFD per la validazione

3.2 Calibrazione dell'area riscaldata tramite R_{th}

Il datasheet fornisce la resistenza termica R_{th} in funzione della portata.

La R_{th} è definita in modo analogo alle resistenze termiche introdotte per conduzione (Eq. 7) e convezione (Eq. 21): il suo scopo è riassumere un problema tridimensionale di scambio termico in una grandezza scalare equivalente, così da avere una stima rapida delle prestazioni del dissipatore. In base all'analogia di Ohm applicata al trasferimento di calore, essa mette in relazione il flusso termico con il corrispondente salto di temperatura, secondo la relazione

$$R_{th} = \frac{\Delta T}{\phi} \quad (22)$$

dove ϕ è la potenza termica scambiata [W] (precedentemente indicata con $\dot{Q}$, ma in elettronica viene spesso utilizzato ϕ) e ΔT è la differenza di temperatura tra i due riferimenti scelti, nel caso in esame tra la temperatura media della superficie riscaldata e la temperatura del fluido in ingresso, come indicato nel grafico del datasheet riportato in Figura 12.

PERFORMANCE

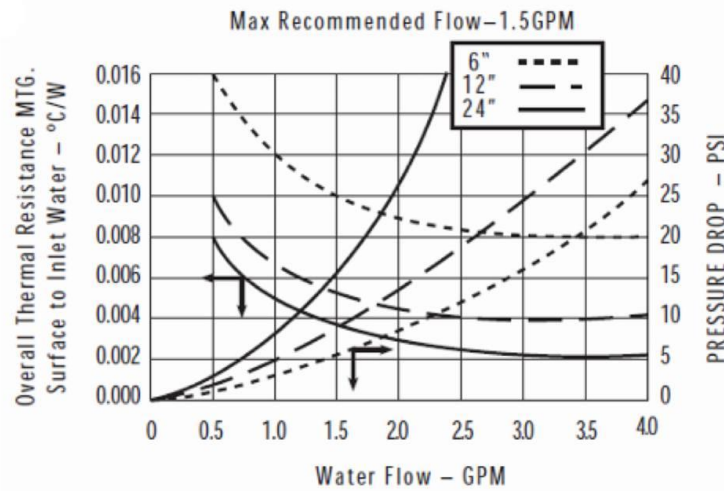


Figura 12: R_{th} in funzione della portata volumetrica, caso preso in esame 12" (305 mm), (Riferimento: Boyd-Hi-Contact-6-Pass-Datasheet)

L'incognita più rilevante per replicare il dato di datasheet riguarda la superficie effettiva di scambio termico tra sorgente di calore e dissipatore, non specificata dal produttore. Per questo motivo è stata svolta una procedura di calibrazione utilizzando come riferimento la condizione a 1 GPM:

- Dal datasheet (ds) è possibile ricavare a 1 GPM una resistenza pari a $R_{th,ds} = 0,0067 \text{ } ^\circ\text{C/W}$; (notare che essendo la resistenza termica definita tramite un ΔT è indifferente utilizzare come unità di misura $^\circ\text{C}$ o K).
- Con $\phi = 6000 \text{ W}$, il salto termico atteso (ds) è: $\Delta T_{ds} = R_{th,ds} \cdot \phi = 40,2 \text{ } ^\circ\text{C}$.

La superficie riscaldata è stata quindi modificata iterativamente (la dimensione) fino a ottenere in simulazione (sim) un valore di ΔT_{sim} coerente con quello atteso (ΔT_{ds}) alla portata di 1 GPM ($\Delta T_{sim} = 40,27 \text{ } ^\circ\text{C}$). La superficie stimata è risultata ragionevole dal punto di vista geometrico (posizionamento e dimensioni compatibili con una tipica area di contatto reale), come si può notare dalla Figura 13.

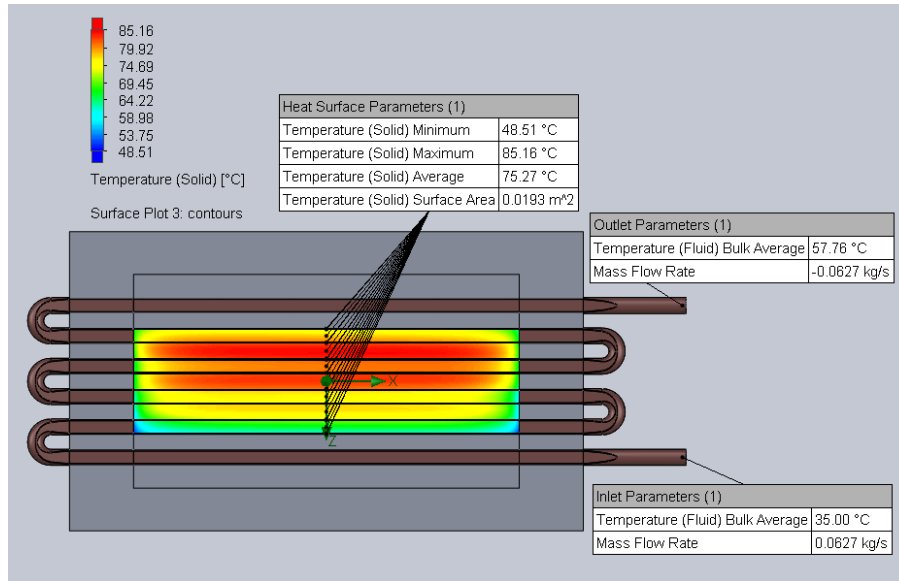


Figura 13: Risultato della calibrazione dell'area riscaldata a 1 GPM. L'area riscaldata coincide con la regione su cui è visualizzata la mappa di temperatura

È importante sottolineare che, in questa prima fase, la validazione non consiste in un confronto diretto della R_{th} (poiché il modello viene volutamente calibrato sul punto a 1 GPM), ma nella verifica che l'area riscaldata stimata risulti fisicamente plausibile nel momento in cui la simulazione riproduce la R_{th} di datasheet (ossia quando $\Delta T_{sim} = \Delta T_{ds} \Rightarrow R_{th,sim} = R_{th,ds}$ per 1 GPM). In altre parole, il criterio adottato è che la corrispondenza con il datasheet sia ottenuta mediante una superficie di scambio coerente con un contatto reale, evitando aggiustamenti non realistici.

Come precedentemente accennato, si è anche effettuato un controllo sul raggiungimento delle condizioni stazionarie utilizzando il bilancio energetico (in sostanza la potenza termica dissipata deve essere uguale a quella assorbita dal fluido):

$$\phi = \dot{Q}_{\text{fluido}} = \dot{m} c_p \Delta T_{\text{fluido}} \quad (23)$$

dove $\Delta T_{\text{fluido}} = T_{\text{out}} - T_{\text{in}}$ (22,76 °C nel caso proposto), c_p è il calore specifico dell'acqua a pressione costante (assunto pari a $c_p = 4186 \text{ J/(kg K)}$ ovvero il valore a 25°C), e $\dot{m}$ è la portata in massa [kg/s]. Quest'ultima è stata calcolata come:

$$\dot{m} = \rho Q \quad (24)$$

con ρ densità dell'acqua ($\rho = 1000 \text{ kg/m}^3$ ovvero il valore a 25°) e Q portata volumetrica [m^3/s]. La portata Q è stata ottenuta convertendo la portata espressa in GPM.

Così facendo si è ottenuto $\dot{Q}_{\text{fluido}} = 6010,8 \text{ W}$, la lieve differenza rispetto a $\phi = 6000 \text{ W}$ è dovuta alle assunzioni su ρ e c_p che in realtà variano con la temperatura, come mostrano le look-up-table di SolidWorks in Figura 14.

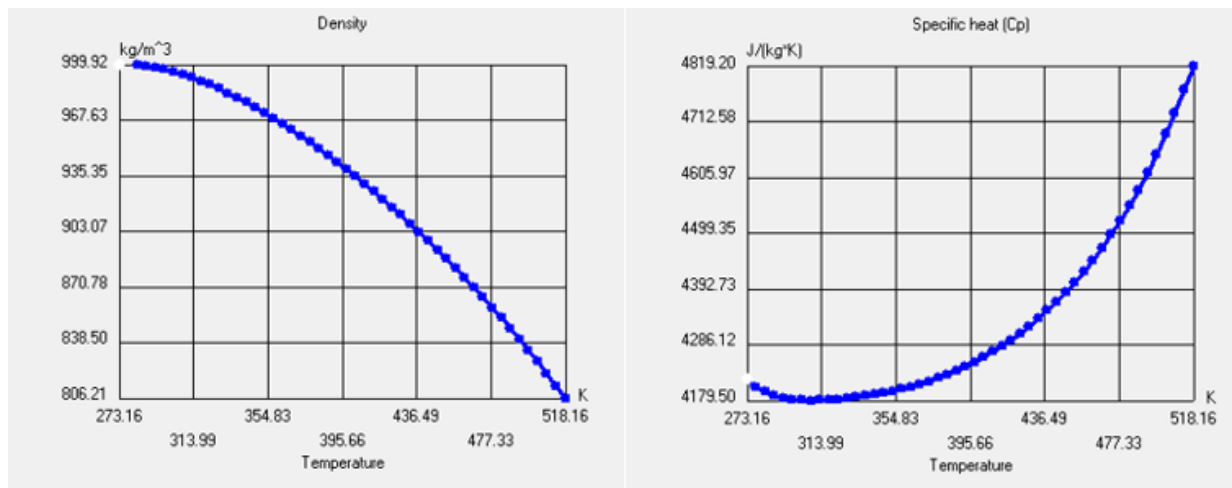


Figura 14: Look-up-table di densità e calore specifico dell'acqua in funzione della temperatura (Riferimento: SolidWorks)

3.3 Confronto R_{th} al variare della portata e sintesi validazione

Una volta calibrato il modello sul caso a 1 GPM, la portata è stata variata per verificare se il modello riuscisse a predire in modo coerente la resistenza termica anche lontano dal punto di calibrazione.

Le simulazioni sono state eseguite alle seguenti portate (e a parità di altre condizioni):

- 0,5 – 1,5 – 2 – 2,5 – 3,5 GPM

Per ciascun caso la R_{th} è stata ricavata utilizzando l'Eq. (22), con i parametri presenti nell'equazione definiti come precedentemente spiegato.

Di seguito, in Figura 15 vengono riportati i confronti delle simulazioni a 0,5 e 1,5 GPM, con alcune considerazioni.

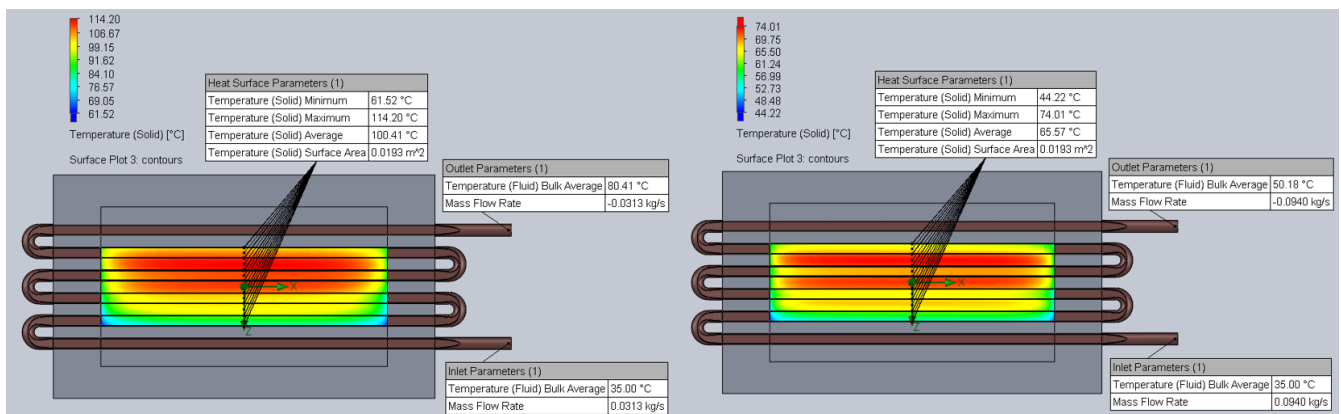


Figura 15: Risultati simulazione a 0,5 GPM (a sinistra) e 1,5 GPM (a destra)

Dal punto di vista qualitativo, le mappe di temperatura risultano simili a colpo d'occhio; tuttavia, questo dipende dal fatto che la scala colori viene adattata automaticamente caso per caso. In entrambi i casi le temperature massime si concentrano nella parte superiore della superficie riscaldata: ciò è coerente con l'andamento termico del fluido lungo il percorso, poiché l'acqua si riscalda progressivamente man mano che assorbe calore e, di conseguenza, la capacità di asportazione locale risulta maggiore nelle zone in cui il fluido è più freddo e minore nelle zone attraversate da fluido già riscaldato. Inoltre, aumentando la portata la distribuzione di temperatura tende a risultare meno uniforme: si accentuano differenze locali legate alla geometria interna, con zone più calde in corrispondenza delle regioni non direttamente "servite" dal canale (assenza di

raffreddamento convettivo diretto) rispetto alle aree adiacenti. Al contrario, a portate più basse la mappa appare complessivamente più uniforme, perché la conduzione nel solido ha un ruolo relativamente più importante nello “spalmare” i gradienti di temperatura, rendendo le differenze locali meno marcate rispetto al contributo convettivo.

Impostando una scala di temperatura comune per entrambe le configurazioni, con estremi pari alla temperatura minima e massima complessive delle due configurazioni, il confronto, proposto in Figura 16, diventa immediato. In queste condizioni si osserva chiaramente il comportamento atteso: all’aumentare della portata la temperatura superficiale diminuisce.

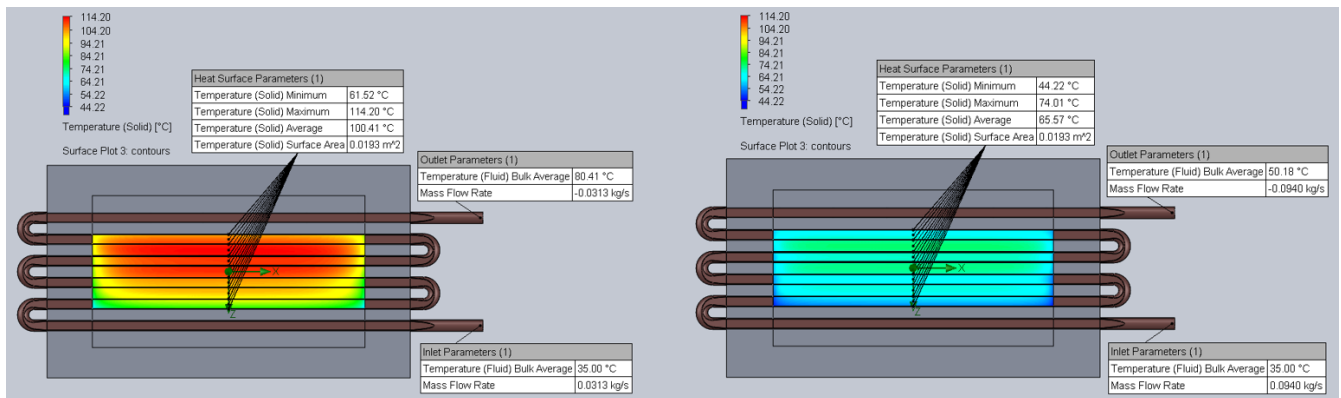


Figura 16: Risultati simulazione a 0,5 GPM (a sinistra) e 1,5 GPM (a destra), stessa scala di temperatura

I valori di resistenza termica ottenuti dalle simulazioni, e riportati in Figura 17, risultano complessivamente abbastanza fedeli a quelli di datasheet anche variando la portata, pur avendo calibrato il modello solo sul punto a 1 GPM.

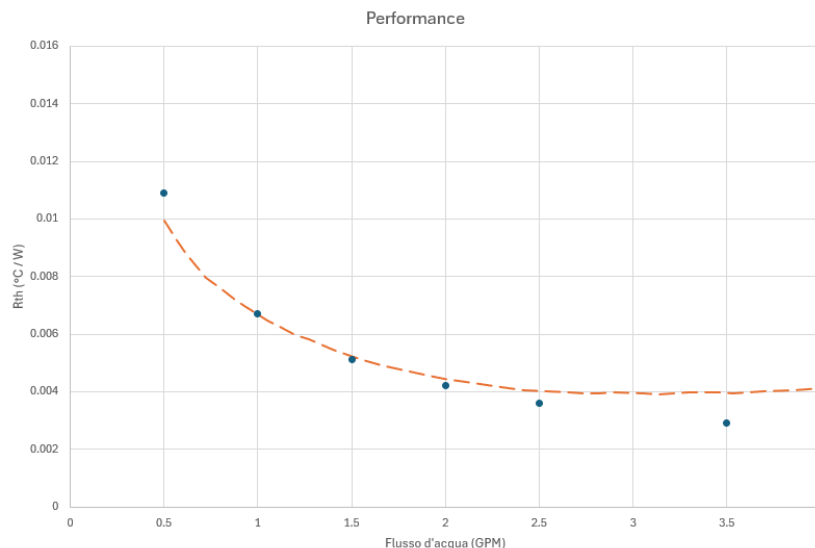


Figura 17: Resistenza termica in funzione del flusso d’acqua, la curva arancione è stata estratta dal datasheet di Figura 12, i punti blu sono le R_{th} ottenute dalle varie simulazioni

Dal grafico si osserva che, al variare della portata, il modello CFD in SolidWorks Flow Simulation riproduce correttamente l’andamento atteso della resistenza termica R_{th} : aumentando il flusso d’acqua, R_{th} diminuisce rapidamente alle portate più basse, mentre alle portate maggiori la riduzione diventa progressivamente meno significativa, tendendo a un valore limite. Nel complesso, i valori simulati risultano in buon accordo con l’andamento del datasheet, nonostante il modello sia inevitabilmente semplificato a causa delle assunzioni adottate (condizioni di prova non

completamente note) e del fatto che il modulo di potenza sia rappresentato in modo equivalente solo come sorgente di potenza termica.

Alla luce di queste evidenze e delle verifiche discusse nel capitolo, SolidWorks Flow Simulation può essere considerato uno strumento affidabile per svolgere le analisi termofluidodinamiche impiegate nei capitoli successivi.

4. Sviluppo dell'automazione MATLAB–SolidWorks per l'ottimizzazione di un dissipatore Pin-Fin a liquido

Questo capitolo descrive il flusso di lavoro di automazione sviluppato per supportare l'ottimizzazione di un dissipatore a raffreddamento a liquido in un contesto applicativo realistico di elettronica di potenza. In particolare, il caso studio considerato riguarda un modulo a raffreddamento diretto (direct cooling), che, come discusso nella sezione 2.2 (e mostrato in Figura 5), presenta la geometria di scambio termico direttamente sul lato posteriore (in questo caso pin-fin) ed è montato su un cooler (alloggiamento) all'interno del quale scorre il liquido di raffreddamento. Per rappresentare una configurazione industrialmente rilevante, il modello CAD del cooler è stato ricostruito sulla base di disegni e dimensioni di un componente reale, mentre le condizioni operative (potenza dissipata, portata, ...) sono state definite facendo riferimento a un modulo di potenza tipicamente accoppiato a tale architettura.

Di conseguenza, l'automazione non è finalizzata a “posizionare cilindri in modo astratto”, ma a generare e valutare configurazioni geometriche all'interno di un assieme completo (cooler, connessioni idrauliche, piastra rappresentativa del modulo, ...), in cui le prestazioni termo-fluidodinamiche dipendono dall'interazione tra geometria di scambio e flusso. In questa prima fase, l'obiettivo è ricercare una disposizione dei pin che migliori le prestazioni termiche, mantenendo coerente il contesto fisico del problema.

Nel dettaglio, la procedura integra tre elementi: la generazione parametrica della geometria in SolidWorks, l'esecuzione di simulazioni termo-fluidodinamiche in SolidWorks Flow Simulation e la gestione della generazione dati e dell'ottimizzazione in MATLAB. L'obiettivo del presente capitolo è duplice: documentare in modo riproducibile l'infrastruttura CAD/CFD automatizzata e riportare un tentativo preliminare di ottimizzazione diretta, evidenziando i limiti emersi nella formulazione iniziale del problema e le motivazioni che hanno portato ad adottare una strategia successiva più robusta e trattabile.

Nel seguito, la sequenza operativa della procedura è riassunta nel diagramma a blocchi in Figura 18, che chiarisce i passaggi principali di scambio dati e di esecuzione tra MATLAB e SolidWorks ed anticipa gli argomenti delle sezioni successive.

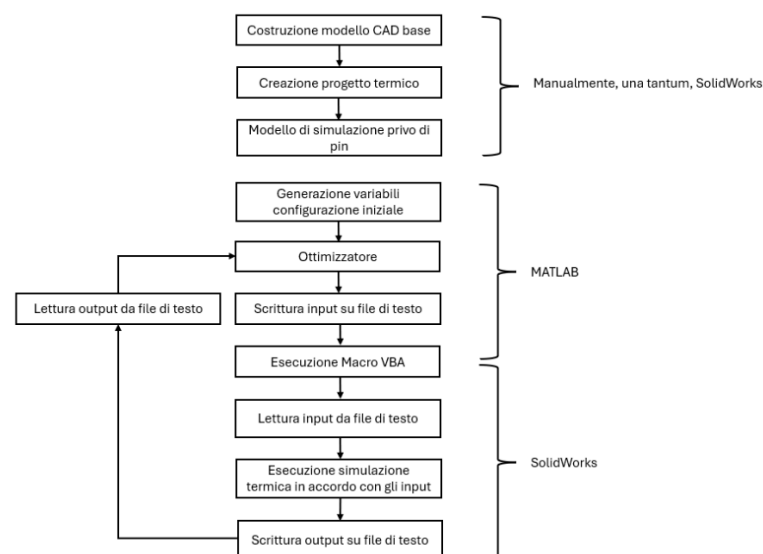


Figura 18: Diagramma a blocchi della procedura automatizzata MATLAB–SolidWorks per il tentativo preliminare di ottimizzazione diretta

4.1 Ricostruzione del modello CAD di base e creazione del progetto termico

The image displays four views of a mechanical component, likely a rectangular housing or base plate. The top view shows a rectangular plate with a recessed rectangular area in the center, defined by a raised border. The side view shows the profile of the plate, which is relatively flat with a slight thickness. The front view shows the plate with a recessed rectangular area in the center, defined by a raised border. The isometric view shows the plate from a three-dimensional perspective, highlighting its rectangular shape and the recessed area. Each view includes a green coordinate system (X, Y, Z axes) for reference.

27

È stato poi costruito l'assieme di base, inserendo l'anello di tenuta, e il modulo di potenza come si può vedere nell'esploso di Figura 21.

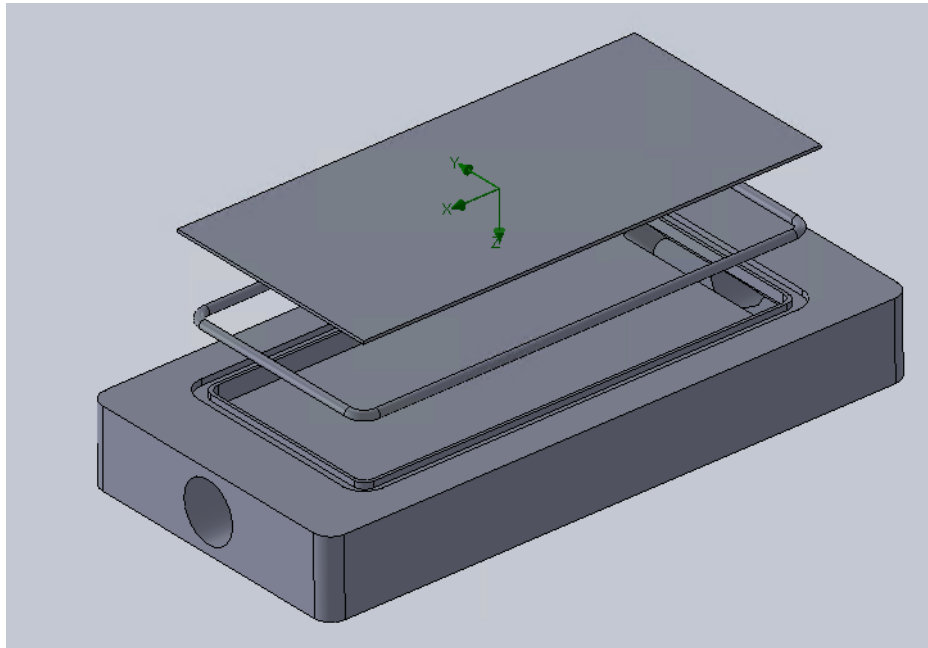


Figura 21: Esploso del modello CAD base (senza Pin)

Il modulo di potenza mostrato in Figura 22 è stato modellato in modo semplificato come una piastra di spessore pari a 1 mm (in sostanza la sola *baseplate*), inizialmente priva dei pin integrati, che verranno aggiunti successivamente tramite la procedura di automazione. Lo spessore ridotto è stato scelto deliberatamente e tale assunzione può influenzare la resistenza termica del componente (in particolare la componente di “spreading/constriction resistance”, i cambi di sezione rallentano il flusso termico). Tuttavia, nel contesto di questo lavoro, la piastra sottile consente di rendere più marcata (e quindi più osservabile) la differenza prestazionale indotta dalla disposizione dei pin. La struttura interna del modulo non è stata modellata, in quanto avrebbe introdotto complessità aggiuntiva non necessaria rispetto agli obiettivi delle analisi svolte. Si noti in figura 22 che la superficie su cui sono posizionati i pin ha dimensioni 137,4 mm x 60 mm (questa è presa come area utile, verranno posizionati sulla stessa superficie nel modello).

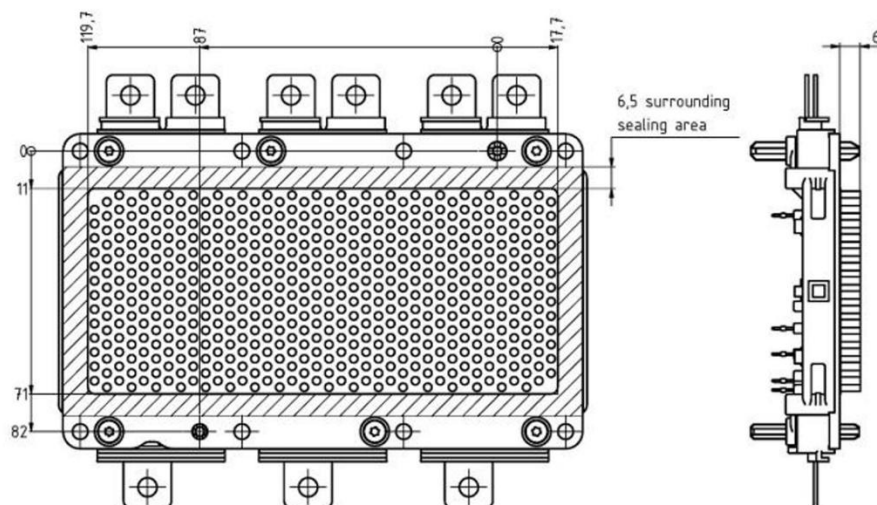


Figura 22: HybridPACK™ Drive Power Module, (Riferimento: Infineon Technologies AG, “Assembly Instructions for the HybridPACK™ Drive Performance”)

Il fine di questa ricostruzione non è la riproduzione estetica del componente, ma l'ottenimento di una geometria sufficientemente fedele dal punto di vista fluidodinamico e termico, tale da consentire simulazioni coerenti con il sistema reale.

Ottenuto il CAD di base è stato creato il progetto termico manualmente in SolidWorks Flow Simulation. Questa scelta è motivata da un vincolo pratico delle API (Application Programming Interface, abilitano i programmatori ad accedere direttamente alle funzionalità di SolidWorks). Esse consentono infatti di modificare parametri e lanciare simulazioni in un progetto esistente, ma non ne permettono la creazione da zero interamente via codice. Sono stati impostati i parametri di Setup della simulazione CFD, spiegati nella sezione 3.1, in accordo con i dati disponibili nel datasheet del modulo di potenza (Infineon Technologies AG, FS950R08A6P2B HybridPACK™ Drive Module), e nell'application note in cui si parla del cooler (Infineon Technologies AG, "Assembly Instructions for the HybridPACK™ Drive Performance"). Vengono riportati in Tabella 2.

Setup CFD	
Fenomeni di scambio termico	Conduzione, Convezione
Gravità	Direzione: z Verso: positivo (Si veda Figura 21 per comprendere la direzione relativa al componente)
Materiali	Cooler: Alluminio, (Application Note: AlMgSi0.5) Piastra e Pin-Fin: Rame, (Datasheet: Rame con placcatura in Nichel) Anello di tenuta: Isolante, (Application Note: EPDM 70) Tubature inlet/outlet: Alluminio (aggiunte in seguito)
Fluido	50% Acqua, 50% Glicole etilenico (da Datasheet) Direzione: x Verso: positivo (Si veda Figura 21 per comprendere la direzione relativa al componente)
$\dot{Q}$	$10 \text{ dm}^3/\text{min}$ (da Datasheet)
Moto del fluido in ingresso	Completamente sviluppato
T_{in}	$75 \text{ }^\circ\text{C}$ (da Datasheet)
Condizioni al contorno	Pareti adiabatiche
ϕ	870 W (da Datasheet)
Definizione dell'area riscaldata	Superficie ridotta sulla piastra (50 mm x 30 mm, si veda Figura 23)
Mesh	Default di SolidWorks (considerazioni successive)
Goals	$T_{av}, T_{out}, p_{in}, p_{out}$

Tabella 2: Setup CFD per il tentativo preliminare di ottimizzazione diretta

Facendo riferimento a Tabella 2, vengono discusse alcune impostazioni. Poiché alcuni materiali specifici non sono presenti nella libreria di SolidWorks Flow Simulation, essi sono stati sostituiti con il materiale di base più rappresentativo ai fini delle analisi. Per quanto riguarda il fluido di raffreddamento, non essendo disponibile nel Database di SolidWorks il glicole etilenico, è stato aggiunto manualmente definendone le proprietà termofisiche^{[22][23]} alla temperatura di riferimento di $75 \text{ }^\circ\text{C}$, coerentemente con le condizioni operative considerate. La sorgente di dissipazione termica è applicata su una regione rettangolare localizzata sulla piastra (sulla faccia superiore), che è una stima dell'area deputata allo scambio termico nel componente reale. Per essere sicuri di essere giunti alle condizioni di regime si è anche aggiunto (rispetto a quanto fatto nella sezione 3.1) che le pressioni di ingresso e uscita (p_{in}, p_{out} , medie di massa) del fluido raggiungessero la convergenza.

Durante le prime analisi di verifica del setup CFD, è emersa la necessità di introdurre tratti di condotto in ingresso e in uscita, in accordo con le linee guida di SolidWorks Flow Simulation, per rendere stabili le simulazioni. In particolare, lato uscita, l'applicazione della condizione di pressione troppo vicino al dissipatore ha evidenziato la presenza di vortici in prossimità dell'apertura,

situazione che può compromettere l'affidabilità dei risultati; per questo motivo è stato aumentato lo sviluppo del condotto di uscita, allontanando la sezione di applicazione della condizione al contorno dalla zona di ricircolo. Analogamente, lato ingresso, l'imposizione della temperatura del fluido su una sezione troppo prossima alla regione riscaldata ha portato a mappe di temperatura non realistiche; è stato quindi aggiunto un tratto di tubo in ingresso per spostare a monte la sezione di applicazione della condizione termica e garantire un campo di temperatura più regolare. Si è così ottenuto il modello di simulazione privo di pin (Figura 23), su cui verranno svolte le analisi termiche.

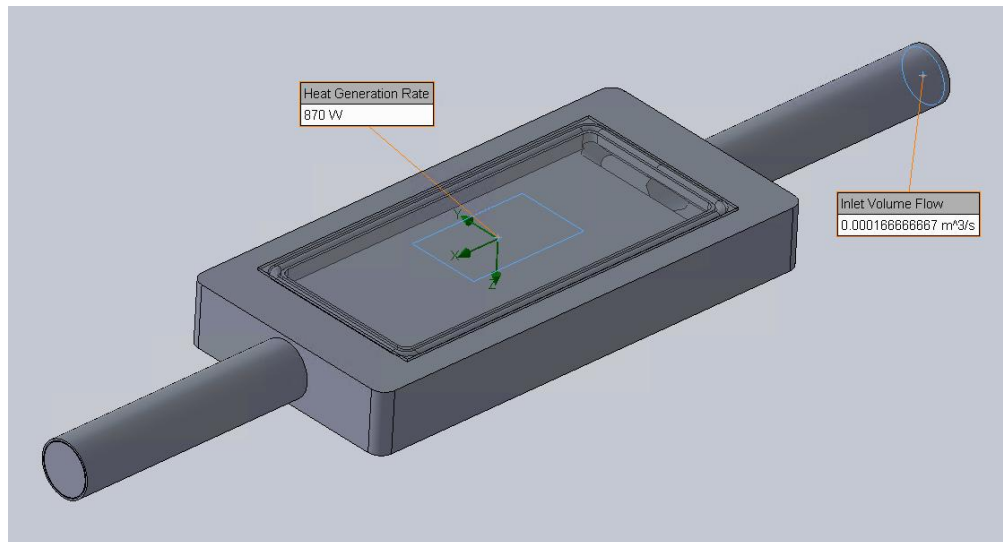


Figura 23: Modello di simulazione privo di pin (piastra in trasparenza)

4.2 Integrazione MATLAB–SolidWorks

MATLAB è stato adottato come ambiente di orchestrazione principalmente per la sua elevata flessibilità e per la possibilità di integrare, all'interno dello stesso framework, tutte le attività necessarie al workflow: generazione sistematica degli input, gestione e archiviazione di dataset, analisi e visualizzazione dei risultati, utilizzo di algoritmi di ottimizzazione, addestramento ed utilizzo di modelli basati su reti neurali come surrogati. Un approccio equivalente in VBA (Visual Basic for Applications, linguaggio di programmazione) risulterebbe meno immediato e meno scalabile.

SolidWorks, d'altra parte, mette a disposizione un ambiente di automazione impostato su macro (sequenza automatizzata di comandi spesso basate su VBA) strettamente integrato con il CAD. In particolare, il Macro Recorder consente di registrare le operazioni svolte nell'interfaccia grafica e tradurle automaticamente in codice VBA. Questo approccio facilita notevolmente lo sviluppo delle macro, perché consente di ottenere una base di codice a partire dalle azioni eseguite a schermo, riducendo la necessità di consultare in modo estensivo la documentazione delle API di SolidWorks. Va tuttavia evidenziato che il Macro Recorder non garantisce una copertura completa di tutte le operazioni: gli Add-In come SolidWorks Flow Simulation e alcune funzioni lato CAD richiedono il ricorso diretto alla documentazione API (specifica dell'Add-In nel caso di Flow Simulation).

Per quanto evidenziato, le macro VBA sono state impiegate esclusivamente per le operazioni strettamente legate a SolidWorks (aggiornamento CAD ed esecuzione delle simulazioni), mentre MATLAB è stato utilizzato per il controllo generale del processo.

Il collegamento tra i due ambienti è stato implementato tramite uno scambio dati basato su file di testo, soluzione semplice e robusta. La sequenza operativa è la seguente:

1. MATLAB si collega a SolidWorks tramite interfaccia COM (Component Object Model), Codice 1.

```
% Connessione a SolidWorks
swApp = actxGetRunningServer('SldWorks.Application');
```

Codice 1: MATLAB, collegamento a SolidWorks

2. MATLAB scrive un file di input contenente i valori necessari alla macro (nell'esempio le impostazioni della simulazione CFD e il percorso del modello), Codice 2.

```
% Scrittura file txt
fid = fopen(iniPath, 'w');
fprintf(fid, 'BCValue=%g\n', BCValue);
fprintf(fid, 'HSValue=%g\n', HSValue);
fprintf(fid, 'BCtemperatureValue=%g\n', BCtemperatureValue);
fprintf(fid, 'MeshValue=%g\n', MeshValue);
fprintf(fid, 'ModelPath=%s\n', ModelPath);
fclose(fid);
```

Codice 2: MATLAB, creazione file txt con i dati di input

3. MATLAB avvia la macro VBA in SolidWorks tramite interfaccia COM, attraverso: percorso, nome del modulo e nome della procedura da eseguire (tutte informazioni sulla macro), Codice 3.

```
% Esecuzione macro
Ok = invoke(swApp, 'RunMacro', macroPath, moduleName, procedureName);
```

Codice 3: MATLAB, esecuzione della Macro

4. La macro VBA legge il file di input generato da MATLAB, ed esegue ciò per cui è stata programmata, Codice 4.

```
' Caricamento parametri dal file scritto da MATLAB
LoadParamsFromFile iniPath, BCValue, HSValue, BCtemperatureValue, MeshValue, ModelPath
```

Codice 4: VBA, caricamento dei parametri letti da file txt

In caso di operazioni preliminari, finalizzate alla sola preparazione ed aggiornamento della geometria del modello (ad esempio l'inserimento dei pin), la procedura si arresta a questo punto. Qualora invece sia richiesta l'esecuzione della simulazione, il flusso prosegue con il passo successivo, Codice 5.

5. La macro estrae le grandezze di interesse e le scrive su un file di output.

```
' Apertura nuovo file di testo e scrittura dei risultati
fileNum = FreeFile()
Open risPath For Output As #fileNum

' Scorrimento di tutti i Goal calcolati
Set FDAGoal = EnumFDAGoals.Next
Do While Not FDAGoal Is Nothing
    For i = 1 To 4
        If FDAGoal.GetGoalName = GoalNames(i) Then
            GoalValue = FDAGoal.GetLastCalculatedValue
            Print #fileNum, "Goal Name: " & GoalNames(i) & " | Value: " & Format(GoalValue, "0.000")
        End If
    Next i
    Set FDAGoal = EnumFDAGoals.Next
Loop

Close #fileNum
```

Codice 5: VBA, creazione file txt con i risultati

6. MATLAB legge il file di output e salva i dati, Codice 6.

```

% Lettura file risultati
txt = fileread(risPath);
% Estrazione risultati
tokens = regexp(txt, '([-+]?[d*[\.,]?\d+)', 'match');
% Conversione in double
values = str2double(tokens);
% Assegnazione valori
Tav = values(2);
Pin = values(4);
Pout = values(6);
Tout = values(8);

```

Codice 6: MATLAB, lettura e caricamento dei risultati

In linea di principio la preparazione ed aggiornamento dell'assieme (operazioni lato CAD) potrebbero essere eseguiti in background (senza interfaccia grafica), con un beneficio in termini di tempi complessivi. Tuttavia, l'Add-In di SolidWorks Flow Simulation forza l'apertura dell'interfaccia grafica per l'esecuzione delle simulazioni. Durante i primi test condotti, nelle sequenze ripetute in cui si alternano generazione della geometria e avvio della simulazione, la chiusura e la successiva riattivazione dell'interfaccia hanno causato instabilità e crash sporadici. Al fine di massimizzare l'affidabilità della catena di esecuzione, si è scelto di mantenere l'interfaccia grafica sempre attiva, privilegiando la robustezza e la ripetibilità delle simulazioni rispetto alla sola velocità.

4.3 Tentativo preliminare di ottimizzazione diretta con Pattern Search

Come primo approccio, il problema di ottimizzazione è stato formulato come la ricerca della disposizione ottimale di un insieme di 8 cilindri (diametro pari a 5 mm, altezza pari a 6 mm) sulla piastra di scambio termico, con l'obiettivo di migliorare la prestazione termica (riduzione della temperatura media sulla superficie riscaldata, T_{av}).

Per rendere automatizzabile l'esplorazione delle configurazioni, sono state sviluppate due macro VBA distinte: una macro dedicata alla generazione della geometria (inserimento dei pin) e una macro dedicata al lancio delle simulazioni termiche e all'estrazione dei risultati.

La macro di generazione geometrica opera nei seguenti step: crea il pin delle dimensioni indicate (se non è presente nella cartella specificata), apre il modello di base privo di pin ed inserisce i cilindri nelle posizioni richieste, infine salva un nuovo assieme. In Figura 24 si mostra la coerenza tra le posizioni generate da MATLAB e quelle ottenute nel modello tramite automazione (alcune immagini di SolidWorks sono state specchiate per agevolare il confronto grafico con l'immagine di MATLAB).

Sul nuovo assieme la macro di simulazione termica imposta i principali parametri operativi e numerici, in particolare: portata, potenza termica applicata, temperatura del fluido in ingresso e dimensione più piccola che deve essere discretizzata dalla Mesh. Nelle campagne descritte in questa tesi tali parametri sono mantenuti costanti; tuttavia, reimpostarli automaticamente a ogni *run* riduce il rischio di errori dovuti a modifiche manuali del progetto dimenticate (ad esempio variazioni dei valori tra un set di prove e un altro) e garantisce quindi la coerenza delle condizioni di confronto tra configurazioni. Inoltre, questa scelta mantiene la procedura flessibile e riutilizzabile per possibili analisi future, ad esempio studi al variare della portata, della temperatura di ingresso, e della potenza termica dissipata, oppure verifiche di sensibilità rispetto alla Mesh.

In questo capitolo i parametri sono quelli precedentemente descritti in Tabella 2.

Al termine di ciascuna simulazione l'automazione estrae e salva le grandezze di interesse: la temperatura media sulla superficie riscaldata T_{av} , le pressioni medie di massa in ingresso e in uscita (p_{in} e p_{out} , da cui ricavare Δp) e la temperatura del fluido in uscita T_{out} . Nel presente lavoro

l'ottimizzazione è condotta considerando esclusivamente T_{av} ; la registrazione delle ulteriori grandezze consente tuttavia di mantenere la procedura generale e di supportare, in possibili analisi successive, valutazioni termo-idrauliche (ad esempio il trade-off tra prestazioni termiche e perdite di carico) e controlli di bilancio energetico (come precedentemente spiegato nella sezione 3.2).

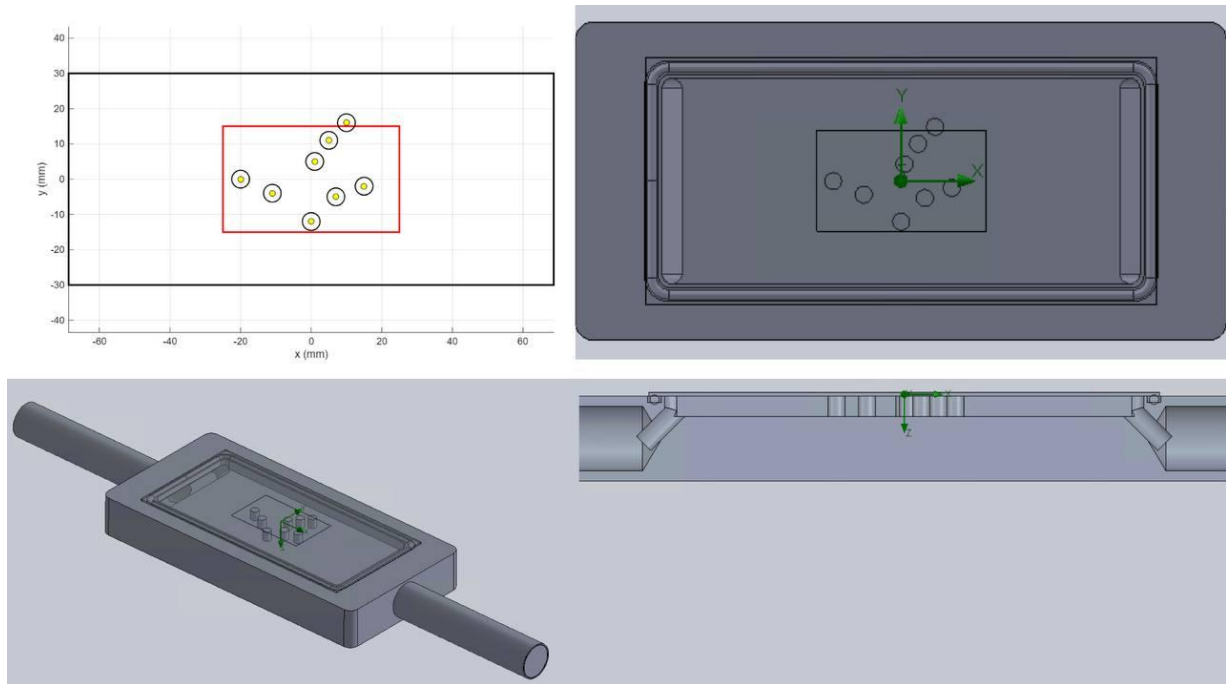


Figura 24: Verifica di coerenza tra MATLAB e SolidWorks: l'automazione genera correttamente la configurazione geometrica richiesta. Alcune immagini di SolidWorks sono state specchiate per agevolare il confronto grafico con l'immagine di MATLAB. In alcune immagini le tubazioni sono state volutamente rimosse (taglio grafico) per migliorare la leggibilità e ridurre l'ingombro visivo

L'ottimizzazione è stata gestita in MATLAB mediante l'algoritmo “*patternsearch*”^[24], disponibile all'interno dell'Add-On Global Optimization Toolbox. Pattern Search appartiene alla famiglia dei metodi Derivative-Free, in particolare alla sottoclasse dei Direct-Search, algoritmi che cercano miglioramenti provando punti attorno alla soluzione corrente secondo un insieme di direzioni (Polling)^[25]. Questi metodi sono particolarmente adatti quando la funzione obiettivo, ossia la grandezza da minimizzare in funzione delle variabili di progetto, non è disponibile in forma analitica. Nel presente caso, infatti, la funzione obiettivo non è una funzione matematica, ma un'intera procedura di valutazione che accoppia MATLAB e SolidWorks: dato un vettore di variabili di progetto $Pos_{mm} = (x_1, y_1, x_2, y_2, \dots, x_8, y_8)$ (posizioni dei cilindri, vettore $\in \mathbb{R}^{16}$), la procedura genera l'assieme corrispondente, avvia la simulazione termica e restituisce T_{av} . Di seguito viene descritto come è stato utilizzato Pattern Search, l'obiettivo non è descrivere in modo esaustivo l'algoritmo, il funzionamento viene richiamato solo nei suoi aspetti essenziali, necessari per interpretare i risultati discussi in seguito. I parametri significativi riportati nel Codice 7 vengono introdotti progressivamente nei punti in cui risultano rilevanti.

```
% Impostazione opzioni Pattern Search
options = optimoptions("patternsearch", Display='iter', MeshTolerance=0.03125);
% Utilizzo di Pattern Search
[Posmm_ott, Tav_ott, exitflag, output] = patternsearch(Tav, Posmm0, [], [], [], [], lb, ub, [], options);
```

Codice 7: MATLAB, impostazione e utilizzo di Pattern Search

- L'algoritmo richiede un vettore iniziale $Pos_{mm,0}$, fornito tramite l'argomento Posmm0. Nel presente lavoro tale configurazione iniziale è generata in modo pseudo-casuale fissando un

determinato Seed tramite il comando MATLAB rng (a Seed fissato la generazione random è sempre la stessa), così da garantire riproducibilità e consentire confronti tra inizializzazioni diverse. A partire da $Pos_{mm,0}$ Pattern Search valuta la funzione obiettivo (l'argomento Tav passato a patternsearch, è una funzione che racchiude la procedura automatizzata descritta in precedenza) ottenendo il valore $T_{av,0}$; il vettore di input e il risultato corrispondente costituiscono la miglior soluzione disponibile all'inizio della procedura e vengono utilizzate come soluzione corrente ($Pos_{mm,best}, T_{av,best}$).

- A ogni iterazione, Pattern Search considera configurazioni alternative generate perturbando le variabili di progetto attorno alla soluzione corrente. Nel caso in esame tali perturbazioni possono essere interpretate come spostamenti che agiscono su una coordinata alla volta, prima sommati e poi sottratti lungo le direzioni di Polling; ad esempio:

$$\begin{aligned}
 Pos_{mm,1} &= Pos_{mm,best} + (\Delta x, 0, 0, \dots, 0) \\
 Pos_{mm,2} &= Pos_{mm,best} + (0, \Delta y, 0, \dots, 0) \\
 Pos_{mm,3} &= Pos_{mm,best} + (0, 0, \Delta x, \dots, 0) \\
 &\vdots \\
 &\vdots \\
 &\vdots \\
 Pos_{mm,17} &= Pos_{mm,best} + (-\Delta x, 0, 0, \dots, 0)
 \end{aligned}$$

e così via per un totale di 32 perturbazioni, durante queste valutazioni possono verificarsi due situazioni:

1. Quando Pattern Search individua una configurazione migliorativa (polling “riuscito”, cioè $T_{av} < T_{av,best}$), si ferma, non effettuando tutte le perturbazioni. La soluzione corrente viene aggiornata, e riparte dall'inizio di questo punto ma con l'ampiezza delle perturbazioni raddoppiate (espansione della Mesh, $2\Delta x, 2\Delta y$). In questo modo l'algoritmo evita di concentrarsi prematuramente in un intorno troppo ristretto e favorisce l'esplorazione di regioni più lontane dello spazio di progetto, potenzialmente più promettenti.
2. Viceversa, quando nessuna delle configurazioni candidate fornisce un miglioramento (polling “non riuscito”), la soluzione corrente rimane invariata, e si riparte dall'inizio di questo punto ma con l'ampiezza della perturbazione dimezzata (contrazione della Mesh $\Delta x/2, \Delta y/2$). Ciò concentra la ricerca in un intorno più vicino alla soluzione corrente, aumentando la risoluzione e la precisione della stima dell'ottimo locale. Se l'ampiezza della Mesh scende al di sotto della soglia impostata (Mesh Tolerance), l'algoritmo arresta la ricerca e la soluzione corrente diventa l'ottimo estratto da Pattern Search.

È importante sottolineare che la “Mesh” citata in questo contesto è la griglia di esplorazione dell'algoritmo in MATLAB (legata all'ampiezza delle perturbazioni delle variabili) e non la Mesh di discretizzazione utilizzata in SolidWorks Flow Simulation.

- Nel presente lavoro sono stati imposti solo vincoli di tipo bound (limiti) tramite i vettori lb (Lower Bounds) e ub (Upper Bounds), in modo da garantire che ogni cilindro resti posizionato all'interno della piastra di riferimento. In particolare, facendo riferimento alla geometria del modulo di potenza (Figura 22), i pin sono interamente vincolati entro l'area utile della piastra, includendo un margine minimo dai bordi di 1 mm. I vettori hanno struttura:

$$\begin{aligned}
 lb &= [x_{min}, y_{min}, x_{min}, y_{min}, \dots] \\
 ub &= [x_{max}, y_{max}, x_{max}, y_{max}, \dots]
 \end{aligned}$$

con ripetizione per ciascun cilindro. Le configurazioni candidate che porterebbero anche una sola coordinata al di fuori dell'intervallo non vengono automaticamente considerate dall'algoritmo.

- I possibili valori dei passi Δx e Δy sono multipli e frazioni di potenze di 2 (compreso 2^0). Il valore massimo dei passi si basa sui limiti imposti alle variabili: nel caso in esame si ottengono valori massimi pari a $\Delta x_{max} = 64 = 2^6$ mm e $\Delta y_{max} = 32 = 2^5$ mm che sono anche quelli di partenza quando si valutano le perturbazioni intorno a $Pos_{mm,0}$. Poiché l'aggiornamento dei passi è gestito internamente dall'algoritmo e Δx e Δy evolvono in modo accoppiato, non è possibile imporre direttamente un valore minimo indipendente per ciascuna direzione. Lo scopo era quello di valutare configurazioni in cui le posizioni non variassero meno di 1 mm; perciò, la tolleranza di arresto della Mesh è stata impostata a $MeshTolerance = 1/2^5 = 0,03125$: con tale impostazione il passo minimo lungo y non scende sotto $\Delta y = 1$ mm. Questa scelta implica che, nelle fasi finali, il passo minimo lungo x risulti pari a $\Delta x = 2$ mm.
- Vanno infine fatte alcune precisazioni, poiché tramite Pattern Search non sono stati imposti vincoli interni, nella funzione obiettivo T_{av} è stato incluso un controllo di fattibilità geometrica basato sulla distanza reciproca tra i cilindri. In particolare, una configurazione è considerata ammissibile solo se i cilindri non si sovrappongono e se risulta rispettata una distanza minima tra le superfici pari a 0,5 mm, introdotta per evitare condizioni di contatto o quasi-contatto tra elementi adiacenti. In caso contrario la procedura non realizza la geometria e non viene avviata la simulazione CFD, ma viene restituito un valore penalizzato pari a 305 °C, in modo che la configurazione non possa risultare migliorativa e non venga mai accettata come successful poll. Questa scelta consente di ridurre il tempo computazionale complessivo ed evita possibili errori o instabilità legati alla generazione di geometrie non fisiche. Il controllo di intersezione è applicato anche alla configurazione iniziale $Pos_{mm,0}$: essa viene generata in modo pseudo-casuale e validata, ripetendo l'estrazione finché non risulta geometricamente fattibile. Inoltre, all'interno della funzione obiettivo è stato implementato il plotting e salvataggio dei plot delle configurazioni valutate (stile figura 24 in alto a sinistra, il titolo di salvataggio riporta coordinate e T_{av}) così da consentire una verifica immediata dell'evoluzione dell'ottimizzazione e del corretto funzionamento dell'intera catena automatizzata durante l'esecuzione.

Al fine di valutare la dipendenza dall'inizializzazione dell'ottimizzazione diretta, la procedura è stata eseguita ripetendo l'ottimizzazione per tre configurazioni iniziali diverse, generate fissando *Seed* differenti (Seed 1–3), mostrate in Figura 25.

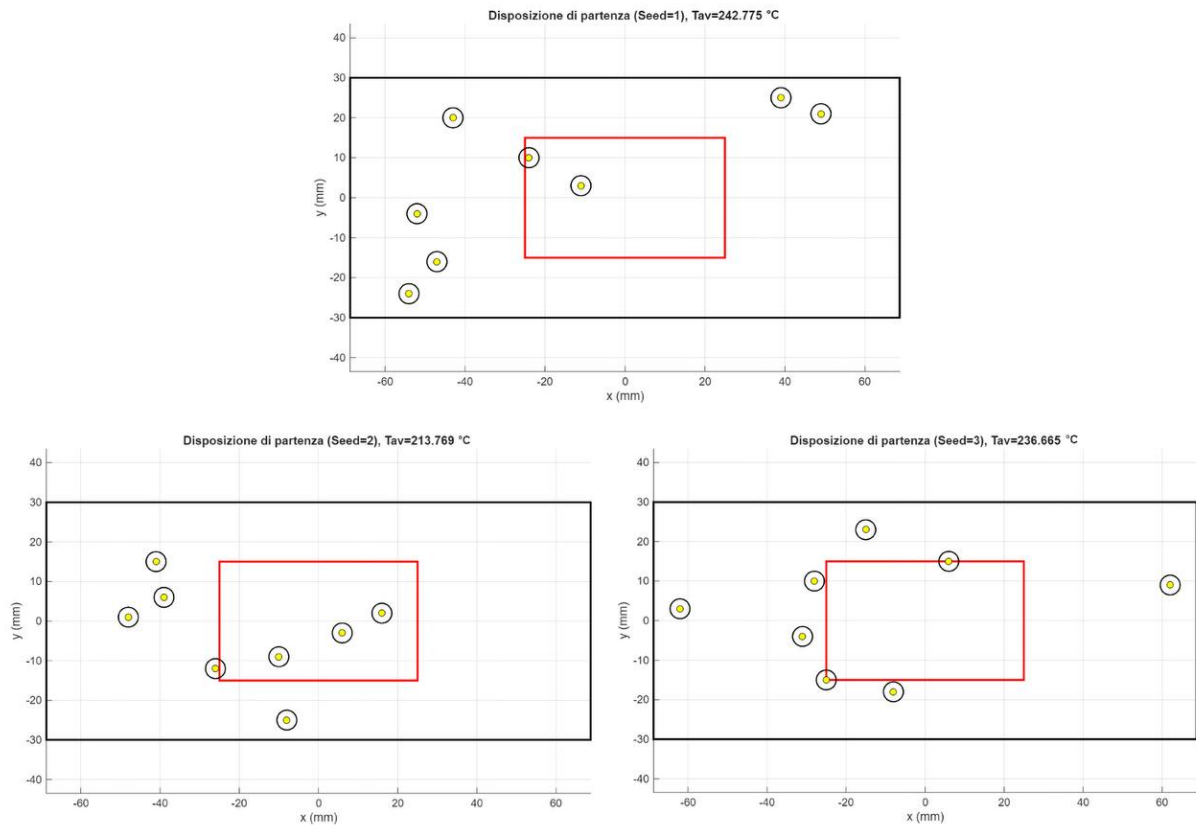


Figura 25: Prime tre configurazioni iniziali generate in modo pseudo-casuale (seed 1–3) per l'avvio dell'ottimizzazione; i cilindri sono posizionati entro i Bounds imposti

A parità di Bounds e parametri CFD, le soluzioni finali ottenute vengono presentate in Figura 26. Per coerenza con quanto discusso in merito alla catena MATLAB–SolidWorks, viene inoltre riportato in Figura 27 un esempio di risultato CFD in SolidWorks relativo a una delle configurazioni ottimizzate. La rappresentazione della configurazione in MATLAB è stata specchiata al solo scopo di facilitare il confronto visivo con l'orientamento della vista in SolidWorks.

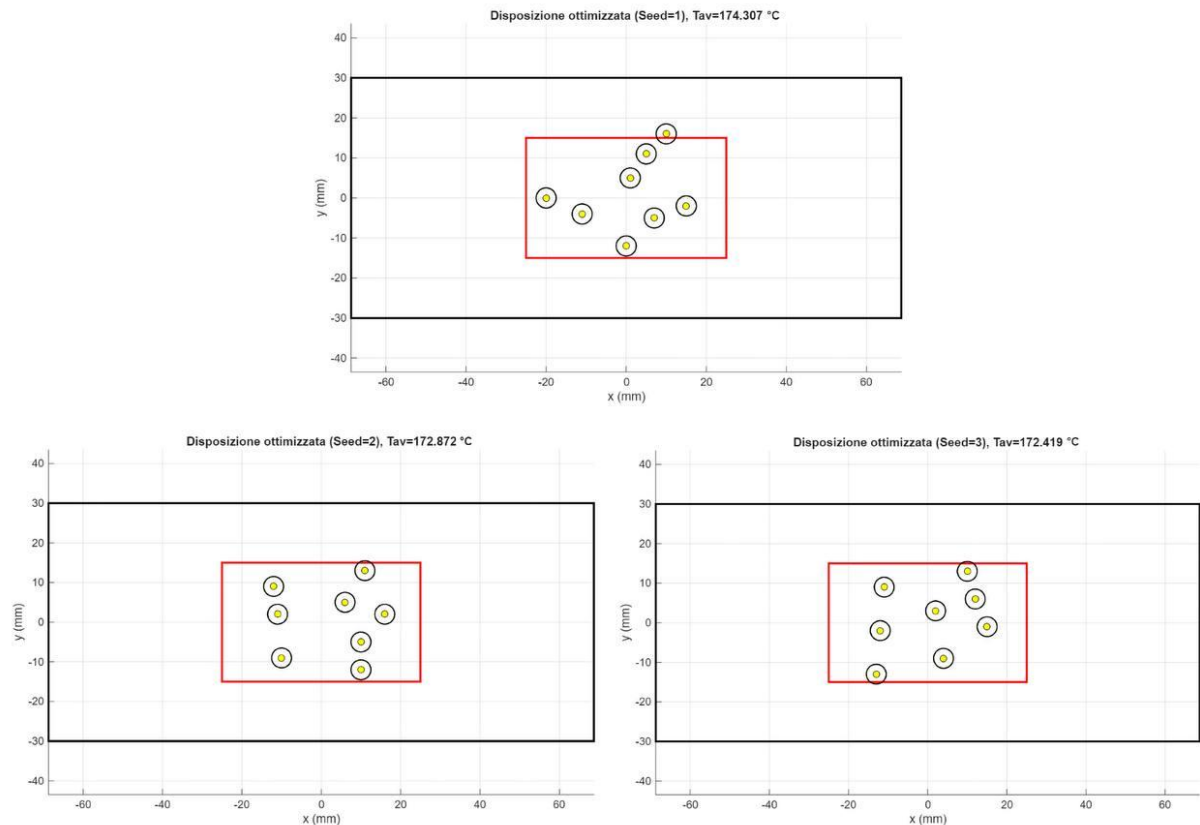


Figura 26: Configurazioni finali ottenute con Pattern Search a partire da Seed differenti (Seed 1–3), con i corrispondenti valori di T_{av} ; le soluzioni risultano differenti, evidenziando dipendenza dall'inizializzazione

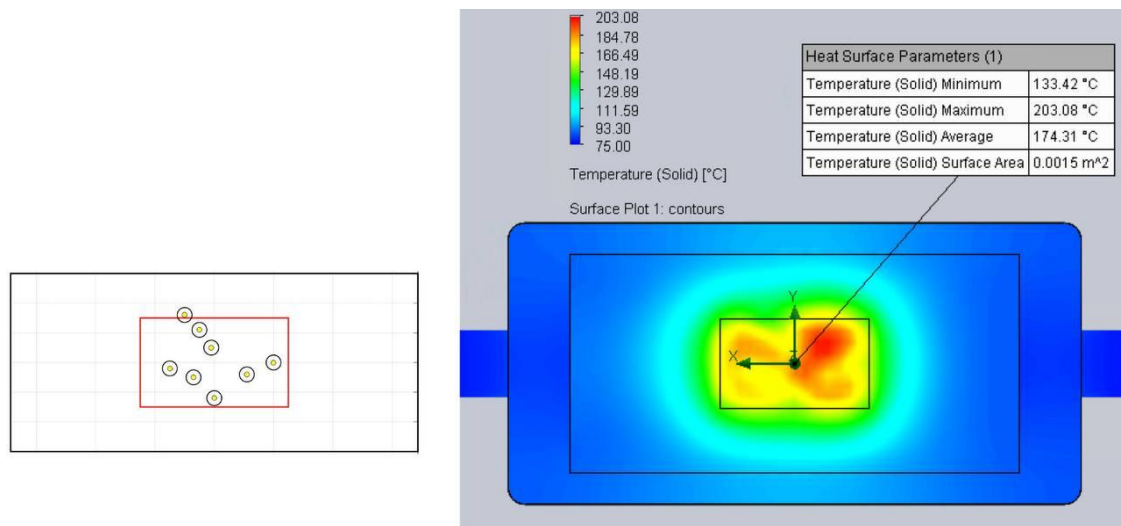


Figura 27: Esempio di risultato CFD in SolidWorks Flow Simulation relativo alla configurazione ottimizzata con Seed=1. La rappresentazione della stessa configurazione in MATLAB è stata specchiata al solo scopo di facilitare il confronto visivo con la mappa termica su SolidWorks

4.3.1 Discussione dei risultati del primo tentativo

Prima di commentare i risultati, in Tabella 3 vengono riepilogate le impostazioni delle simulazioni, poiché verranno successivamente discusse e modificate nel paragrafo successivo. In Tabella 4 sono invece riportati i risultati delle campagne svolte, al fine di fornire un quadro sintetico e confrontabile dei dati che verranno commentati nel seguito.

Impostazioni Generali Del Processo	
Mesh Tolerance (MATLAB)	$1/2^5 = 0,03125$ $\Delta x_{min} = 2 \text{ mm}, \Delta y_{min} = 1 \text{ mm}$
Bounds (MATLAB)	Cilindri interamente contenuti con margine di 1 mm nell'area utile (137,4 mm x 60 mm)
Mesh (SolidWorks)	Default di SolidWorks

Tabella 3: Impostazioni del primo tentativo di ottimizzazione diretta

Seed	Configurazioni di partenza	Configurazioni ottimizzate	Func-Count Pattern Search Tempo di ottimizzazione Criterio di Stop
1	$Pos_{mm,0}=(-11;3;-47;-16;39;25;-24;10;49;21;-54;-24;-43;20;-52;-4) \text{ mm}$ $T_{av,0}=242.775 \text{ }^{\circ}\text{C}$	$Pos_{mm,ott}=(5;11;15;-2;7;-5;0;-12;1;5;10;16;-11;-4;-20;0) \text{ mm}$ $T_{av,ott}=174.307 \text{ }^{\circ}\text{C}$	1096 20 h Mesh Tolerance
2	$Pos_{mm,0}=(-8;-25;6;-3;-10;-9;-39;6;-26;-12;16;2;-48;1;-41;15) \text{ mm}$ $T_{av,0}=213.769 \text{ }^{\circ}\text{C}$	$Pos_{mm,ott}=(6;5;10;-5;-10;-9;-11;2;10;-12;16;2;-12;9;11;13) \text{ mm}$ $T_{av,ott}=172.872 \text{ }^{\circ}\text{C}$	1534 42 h Mesh Tolerance
3	$Pos_{mm,0}=(-62;3;-31;-4;-28;10;-8;-18;6;15;-25;-15;-15;23;62;9) \text{ mm}$ $T_{av,0}=236.665 \text{ }^{\circ}\text{C}$	$Pos_{mm,ott}=(2;3;15;-1;12;6;-12;-2;4;-9;-13;-13;-11;9;10;13) \text{ mm}$ $T_{av,ott}=172.419 \text{ }^{\circ}\text{C}$	1934 41 h Mesh Tolerance

Tabella 4: Risultati del primo tentativo di ottimizzazione diretta

Per l'interpretazione di Tabella 4 si precisa che Func-Count corrisponde al numero di valutazioni della funzione obiettivo effettuate da Pattern Search, cioè al numero di chiamate alla funzione Tav. Nel presente lavoro, ciascuna valutazione di Tav attiva la procedura descritta nelle sezioni precedenti. Il tempo medio necessario per una singola valutazione completa, comprendente aggiornamento della geometria ed esecuzione della simulazione CFD, è dell'ordine di 2 minuti. Tuttavia, il tempo complessivo di ottimizzazione non è semplicemente dato dal prodotto "Func-Count × tempo per simulazione", poiché una parte delle valutazioni viene scartata tramite penalizzazione (configurazioni non fattibili), restituendo un valore fittizio di T_{av} senza avviare la simulazione e riducendo quindi il tempo medio di valutazione. D'altra parte, il tempo complessivo risente anche dell'*overhead* legato alla catena di automazione, in particolare apertura e chiusura ripetuta dei modelli, ricostruzione e sovrascrittura dei file e gestione dell'interfaccia, possono introdurre rallentamenti cumulativi durante campagne molto lunghe. Per tali ragioni, i tempi riportati devono essere intesi come valori indicativi del costo computazionale complessivo della procedura, più che come una misura direttamente riconducibile al solo tempo di calcolo CFD.

Le prime tre ottimizzazioni, avviate da configurazioni iniziali differenti (Seed 1–3), convergono a soluzioni finali diverse sia in termini di disposizione dei cilindri sia in termini di valore di $T_{av,ott}$. Questo comportamento suggerisce un paesaggio della funzione obiettivo caratterizzato da più bacini di attrazione e una marcata dipendenza dall'inizializzazione, coerente con la presenza di minimi locali. Le configurazioni ottimizzate, pur mostrando la tendenza generale a collocare i cilindri nella regione di maggiore carico termico, non evidenziano un pattern geometrico chiaramente interpretabile.

È importante osservare che l'assenza di un pattern univoco non implica necessariamente che le soluzioni siano prive di significato fisico: data l'elevata dimensionalità del problema, è plausibile che un'interpretazione robusta richieda l'esplorazione di un numero molto maggiore di bacini di attrazione, ripetendo l'ottimizzazione con molti Seed per verificare l'eventuale ricorrenza di caratteristiche comuni. Tuttavia, un approccio di questo tipo risulta immediatamente poco praticabile, poiché ogni Seed comporta numerose valutazioni della funzione obiettivo basate su simulazioni CFD. Con soli tre Seed, il tempo di calcolo complessivo supera già i quattro giorni.

Perciò, prima di attribuire questa mancanza di pattern esclusivamente alla necessità di una campagna multi-Seed estesa, si è ritenuto opportuno analizzare alcune cause alternative o che comunque possono contribuire a rendere poco leggibili i risultati. Nel seguito vengono quindi discusse due possibili cause e le relative azioni di mitigazione adottate.

1. Prima possibile causa.

A questo punto è stata considerata una prima possibile causa legata alla granularità della ricerca. In particolare, si è verificato se la discretizzazione imposta da Mesh Tolerance potesse limitare la capacità dell’algoritmo di affinare la soluzione; per questo motivo, per il caso Seed = 1 è stata ripetuta l’ottimizzazione mantenendo l’impostazione di default di MATLAB (Mesh Tolerance= 10^{-6} , ovvero la massima riduzione è $1/2^{19}$). In Tabella 5 vengono riportate le impostazioni modificate rispetto alle simulazioni precedenti, mentre in Tabella 6 sono riassunti i risultati.

Impostazioni Modificate Del Processo	
Mesh Tolerance (MATLAB)	$1/2^{19} = 1,907 \times 10^{-6}$ $\Delta x_{min} = 0,00012 \text{ mm}, \Delta y_{min} = 0,00006 \text{ mm}$

Tabella 5: Variazione di Mesh Tolerance per la simulazione con Seed=1

Seed	Configurazioni di partenza	Configurazioni ottimizzate	Func-Count Pattern Search Tempo di ottimizzazione Criterio di Stop
1	$Pos_{mm,0}=(-11;3;-47;-16;39;25;-24;10;49;21;-54;-24;-43;20;-52;-4) \text{ mm}$ $T_{av,0}=242.775 \text{ }^{\circ}\text{C}$	$Pos_{mm,opt}=(5;12;14.9995;-2;7;-5;0.0625;-12.25;-0.5;5;9;16;-11;-3.25;-20;0.25) \text{ mm}$ $T_{av,opt}=171.956 \text{ }^{\circ}\text{C}$	2037 51 h Mesh Tolerance

Tabella 6: Risultati simulazione con Seed=1 e Mesh Tolerance modificato

In tale condizione l’algoritmo esegue un numero maggiore di valutazioni, e valuta configurazioni più “vicine” tra loro. Tuttavia, la configurazione finale risulta sostanzialmente sovrapponibile a quella ottenuta con le impostazioni precedenti, come si può osservare in Figura 28, mentre il valore di T_{av} mostra una variazione significativa. Una differenza abbastanza marcata di T_{av} ($171,956 \text{ }^{\circ}\text{C}$ vs $174.307 \text{ }^{\circ}\text{C}$) a fronte di configurazioni geometricamente quasi identiche risulta inattesa dal punto di vista fisico, poiché piccoli spostamenti dovrebbero produrre variazioni più contenute della temperatura media. Per questo motivo si è ritenuto necessario approfondire il comportamento della procedura mediante ulteriori analisi.

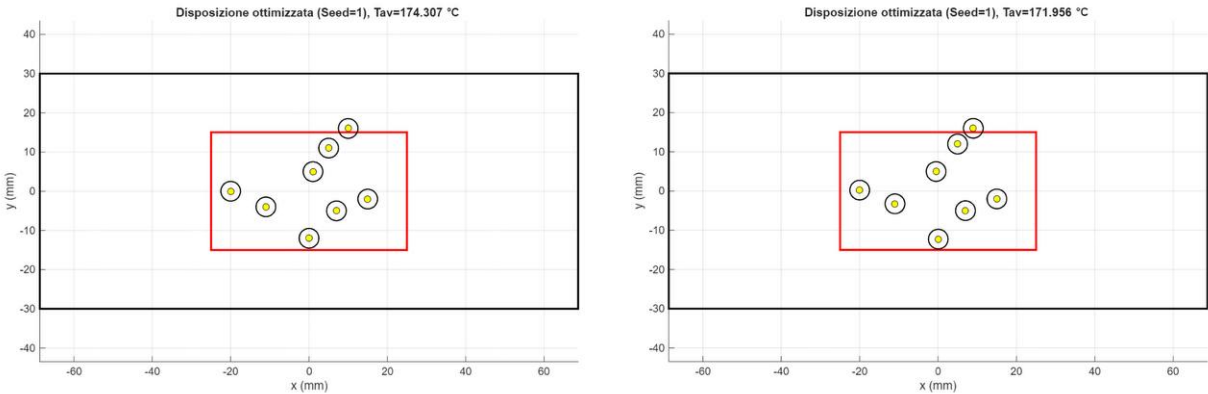


Figura 28: Disposizione ottimizzata a parità di Seed (Seed=1), variando Mesh Tolerance, a sinistra Mesh Tolerance= $1/2^5$, a destra Mesh Tolerance= $1/2^{19}$

Per interpretare meglio i risultati si è analizzata l’evoluzione dell’ottimizzazione utilizzando le informazioni riportate nei titoli dei plot salvati durante l’esecuzione, inizialmente impiegati come

strumento di monitoraggio del corretto funzionamento della procedura (sezione precedente). Si precisa che tali immagini non riportano tutte le valutazioni: per la logica di Pattern Search può accadere che una stessa configurazione venga valutata più volte e, in tal caso, il salvataggio in sovrascrittura fa sì che la configurazione compaia solo nell'ultima occorrenza. Un effetto analogo riguarda la configurazione ottimizzata, che viene visualizzata nuovamente a fine esecuzione e sovrascrive l'immagine relativa all'iterazione in cui è stata effettivamente individuata. Nonostante queste limitazioni, la sequenza riportata in Figura 29 fornisce una descrizione qualitativa dell'andamento della ricerca e consente di discutere il comportamento dell'algoritmo nelle fasi finali ovvero la sensibilità di T_{av} a piccole perturbazioni geometriche.

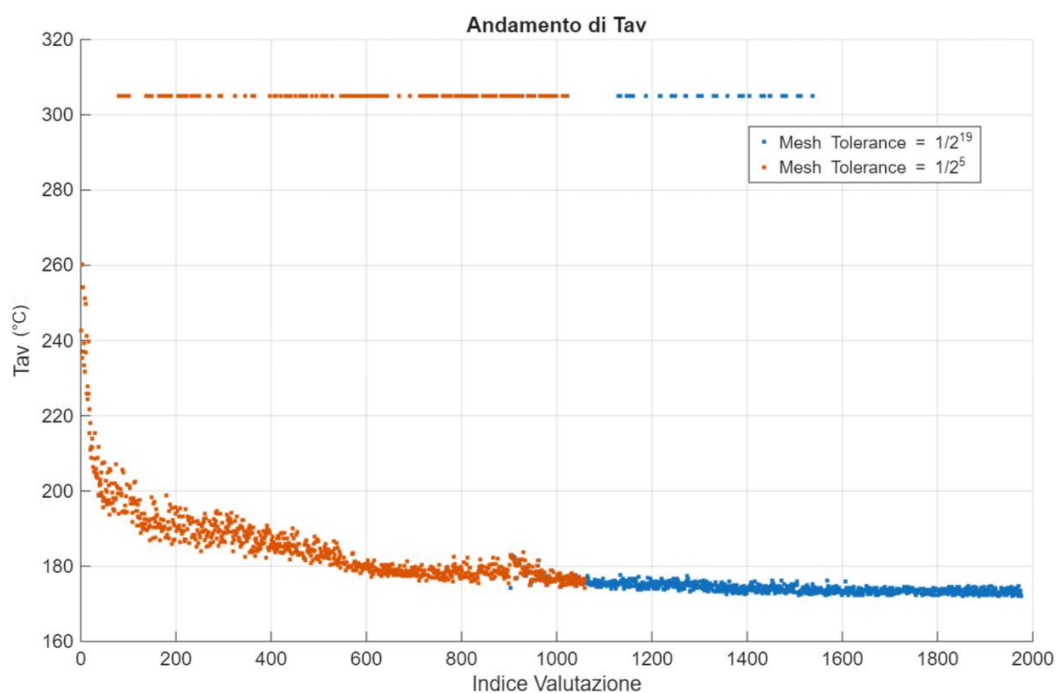


Figura 29: Evoluzione qualitativa dell'ottimizzazione (Seed=1)

Il comportamento osservato in Figura 29 conferma quanto già emerso dal confronto in Figura 28. In particolare, sia nelle fasi finali con MeshTolerance = $1/2^5$, sia soprattutto nel caso MeshTolerance = $1/2^{19}$ (in cui le perturbazioni introdotte dall'algoritmo diventano estremamente piccole, inferiori al micron), si osservano variazioni marcate di T_{av} a fronte di configurazioni praticamente indistinguibili. Lo stesso si osserva per Seed 2-3, qui non riportati per non appesantire il grafico. Tale risposta risulta difficilmente giustificabile dal punto di vista fisico e suggerisce che la differenza osservata sia dominata da effetti numerici, plausibilmente legati alla Mesh della simulazione in SolidWorks Flow Simulation. Perciò, prima di proseguire con ulteriori campagne di ottimizzazione, si è verificata in modo almeno preliminare l'influenza della discretizzazione CFD sul valore calcolato di T_{av} .

A tal fine è stata sviluppata un'automazione per svolgere un'analisi di sensibilità al variare del parametro che identifica la dimensione più piccola che deve essere discretizzata dalla Mesh globale di SolidWorks. Si riportano in Figura 30 i risultati per una configurazione ottimizzata rappresentativa (Seed=1, MeshTolerance = $1/2^5$), ma l'analisi è stata eseguita per tutti e 3 i Seed.

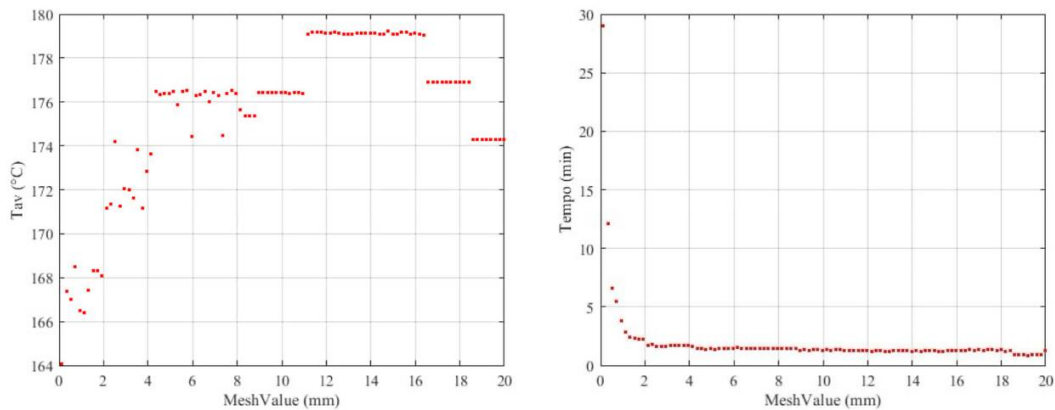


Figura 30: Analisi di sensibilità della temperatura media T_{av} e dei tempi di simulazione al variare della dimensione minima discretizzata dalla mesh globale di SolidWorks, per una configurazione ottimizzata rappresentativa (Seed=1, MeshTolerance = $1/2^5$)

I risultati mostrano che T_{av} tende a diminuire al diminuire della Mesh, senza tuttavia raggiungere una piena convergenza nel range esplorato, mentre il tempo di simulazione cresce sensibilmente per Mesh più fini. Ne consegue che un infittimento della discretizzazione risulta necessario per ottenere una valutazione più stabile di T_{av} , ma deve essere perseguito mantenendo sotto controllo il costo computazionale, poiché l'ottimizzazione richiede un'elevata numerosità di valutazioni. Conclusa questa prima analisi, si può passare alla seconda possibile causa che rende i risultati dell'ottimizzazione diretta poco significativi.

2. Seconda possibile causa

La formulazione iniziale, basata sulle coordinate di 8 cilindri liberi di muoversi sull'intera piastra, genera uno spazio di progetto estremamente ampio: anche imponendo spostamenti minimi di 1 mm, il numero di configurazioni possibili rimane elevatissimo. A ciò si aggiunge una ridondanza strutturale legata al fatto che i cilindri sono fisicamente indistinguibili: le permutazioni delle etichette (ad esempio scambiando le coordinate del cilindro 1 con quelle del cilindro 5) non modificano la configurazione fisica, ma vengono interpretate come soluzioni diverse dall'algoritmo. In linea teorica, una stessa configurazione può essere rappresentata in fino a $8! = 40320$ permutazioni equivalenti, moltiplicando ulteriormente la complessità effettiva dell'esplorazione.

Per mitigare entrambe le criticità evidenziate, da un lato l'eccessiva ampiezza dello spazio di progetto, dall'altro la necessità di aumentare la risoluzione CFD senza rendere proibitivo il costo computazionale, si è adottata una riformulazione operativa del problema. In particolare, l'area ammissibile di posizionamento dei cilindri è stata ridotta a una regione pari a $1,4^2$ volte l'area calda (entrambi i lati della regione sono 1,4 volte i lati dell'area calda), concentrando la ricerca nella zona fisicamente più rilevante e riducendo il numero di configurazioni esplorabili. Questa scelta ha inoltre permesso di introdurre un raffinamento numerico mirato tramite Mesh locale in SolidWorks. La Mesh locale è stata posizionata nella nuova regione di posizionamento dei cilindri, un esempio è presentato in Figura 31. L'inserimento della Mesh locale è stato eseguito manualmente nell'assieme di simulazione di base (quello in cui tramite automazione vengono inseriti i Pin), poiché SolidWorks non ne consente l'aggiunta automatica tramite macro.

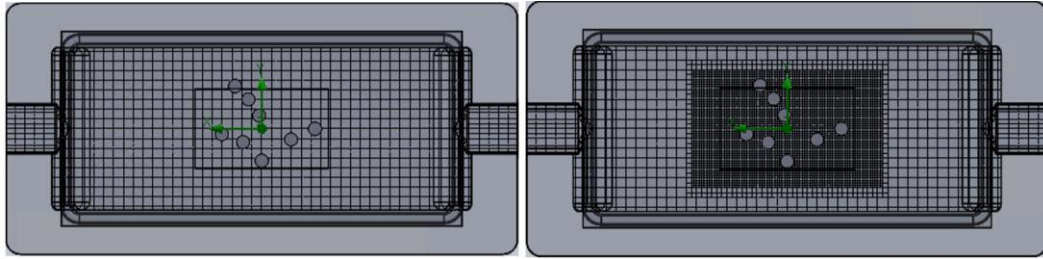


Figura 31: Confronto tra discretizzazione con sola mesh globale (sinistra) e con raffinamento mediante mesh locale nella regione di interesse (destra)

Sulla base dei test preliminari (Figura 30), la Mesh locale è stata impostata con un livello di raffinamento pari a 3 (su una scala da 1 a 9), definito in relazione alla dimensione delle celle della Mesh globale. La Mesh globale è stata mantenuta alle impostazioni di default di SolidWorks, concentrando l'incremento di risoluzione nella sola area di interesse. In questo modo si ottiene un compromesso tra accuratezza e tempi di calcolo, migliorando la stabilità dei risultati rispetto a piccoli spostamenti geometrici senza compromettere l'esecuzione iterativa richiesta dall'ottimizzazione. Per fornire un riscontro qualitativo dell'effetto del raffinamento numerico, in Figura 32 è riportato un confronto tra le mappe di temperatura del solido ottenute con sola Mesh globale e con l'introduzione della Mesh locale.

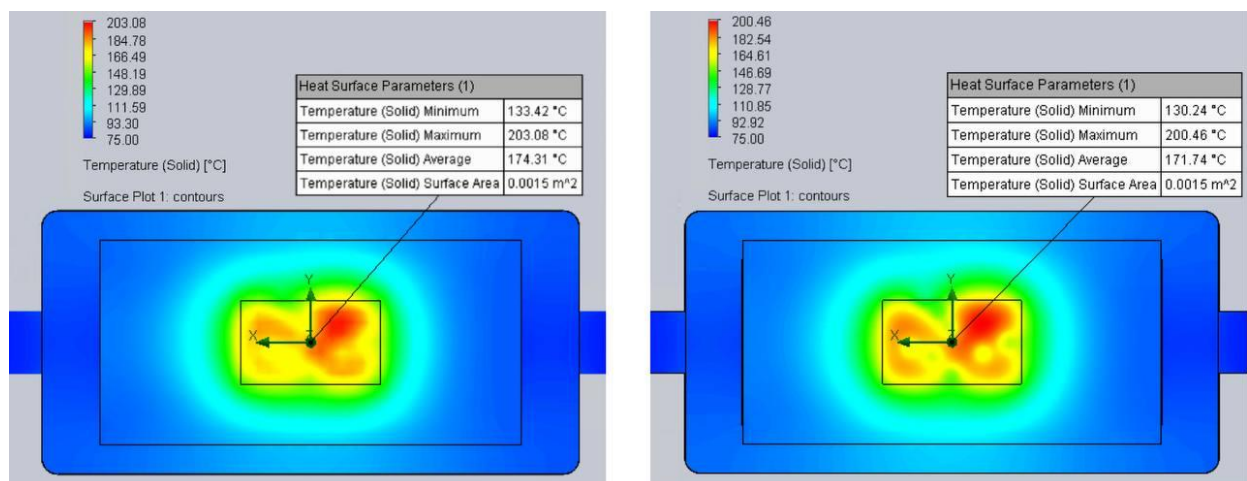


Figura 32: Confronto qualitativo delle mappe di temperatura del solido ottenute con sola Mesh globale (sinistra) e con raffinamento mediante Mesh locale nella regione di interesse (destra)

Nel caso con sola Mesh globale (Figura 32 a sinistra), il campo termico risulta meno regolare e l'influenza dei cilindri è meno evidente, con una distribuzione che può risentire maggiormente della discretizzazione. L'applicazione della Mesh locale, concentrando la risoluzione dove avviene lo scambio termico, restituisce invece una mappa più graduale e leggibile, in cui l'effetto dei cilindri sulla distribuzione di temperatura risulta più chiaramente distinguibile. Questo confronto supporta l'ipotesi che fosse necessario infittire la Mesh, giustificando l'adozione di un raffinamento locale nelle analisi successive.

4.3.2 Tentativo di mitigazione: area ristretta e mesh locale

Come anticipato nella conclusione del paragrafo precedente, per mitigare le criticità emerse nel tentativo iniziale (ampiezza dello spazio di progetto e sensibilità numerica legata alla Mesh di SolidWorks) il problema è stato riformulato introducendo: una riduzione dell'area ammissibile di

posizionamento dei cilindri (tramite i Bounds) e un raffinamento locale della Mesh nella regione di interesse. Le principali variazioni rispetto al caso precedente sono riassunte in Tabella 7.

Impostazioni Nuove Del Processo	
Mesh Tolerance (MATLAB)	$1/2^5 = 0,03125$ $\Delta x_{min} = 1 \text{ mm}, \Delta y_{min} = 0,5 \text{ mm}$ (considerazioni successive)
Bounds (MATLAB)	Cilindri interamente contenuti con margine di 1 mm nella nuova area utile (70 mm x 42 mm) ovvero 1,4 volte le dimensioni dell'area calda (50 x 30 mm)
Mesh (SolidWorks)	Globale: Default di SolidWorks Locale: Raffinamento livello 3

Tabella 7: Impostazioni del secondo tentativo di ottimizzazione diretta

A seguito della modifica dei Bounds di Pattern Search, la variazione minima teorica lungo y scende a 0,5 mm nelle fasi finali della ricerca. Tuttavia, poiché la risoluzione geometrica di interesse è stata fissata a 1 mm (sia lungo x sia lungo y), le coordinate candidate generate dall'algoritmo sono state arrotondate ai millimetri. Questa scelta consente di mantenere una discretizzazione uniforme delle variabili e di evitare valutazioni su spostamenti sub-millimetrici. L'arrotondamento aumenta però la possibilità che, nelle fasi finali, l'algoritmo proponga più volte configurazioni già valutate. Per mantenere efficiente la procedura, è stato quindi implementato un controllo sulle configurazioni già esplorate: se la configurazione proposta è già presente nello storico delle valutazioni, la simulazione non viene rilanciata e il valore di T_{av} viene recuperato dai risultati precedentemente calcolati. In questo modo si riduce il numero di simulazioni CFD ripetute, migliorando il tempo complessivo dell'ottimizzazione senza alterare il comportamento dell'algoritmo in termini di confronto tra soluzioni.

Definite queste modifiche, la procedura di ottimizzazione rimane invariata rispetto alla sezione precedente: l'ottimizzazione è stata eseguita ripetendo la ricerca a partire da tre Seed differenti (Seed 1–3), al fine di valutare nuovamente la dipendenza dall'inizializzazione. In Figura 33 e 34 vengono riportate rispettivamente le configurazioni di partenza e quelle ottimizzate.

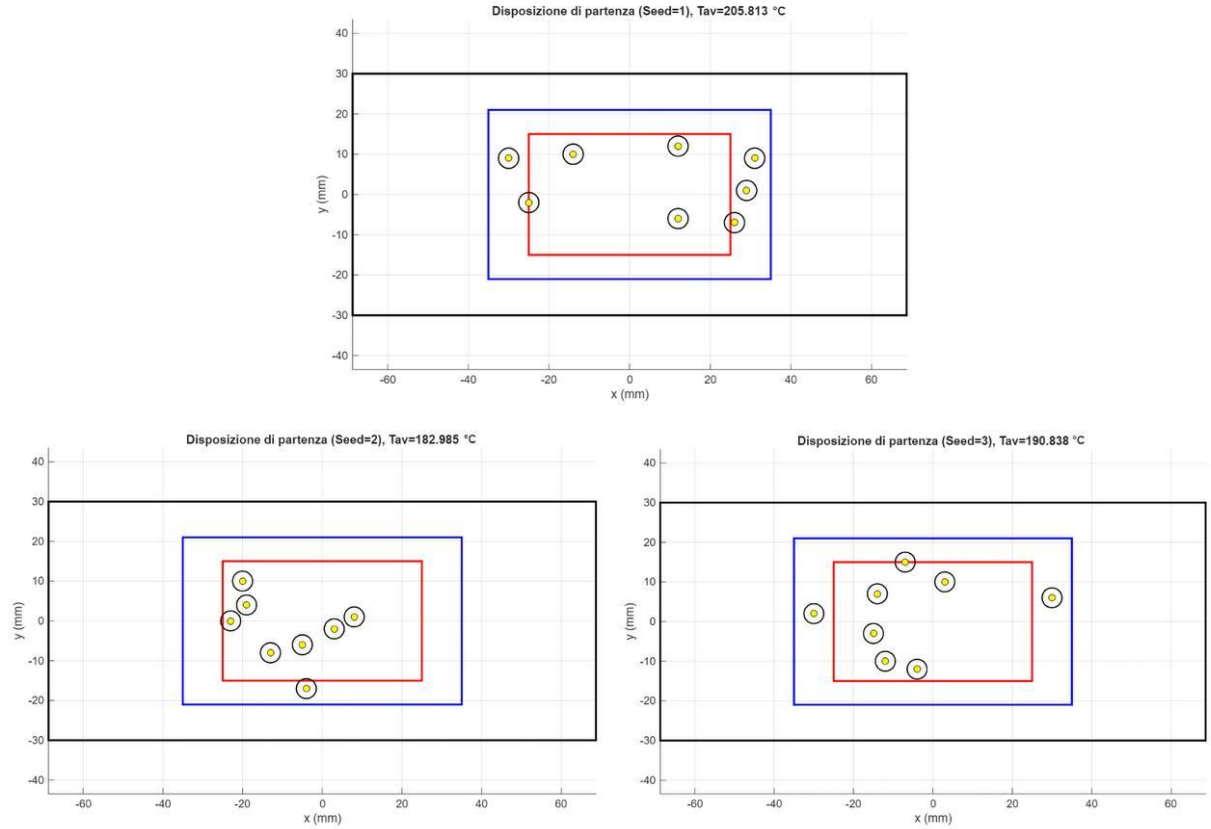


Figura 33: Configurazioni iniziali generate in modo pseudo-casuale (seed 1–3) per l'avvio dell'ottimizzazione; i cilindri sono posizionati entro i nuovi Bounds imposti, rappresentati dal rettangolo blu

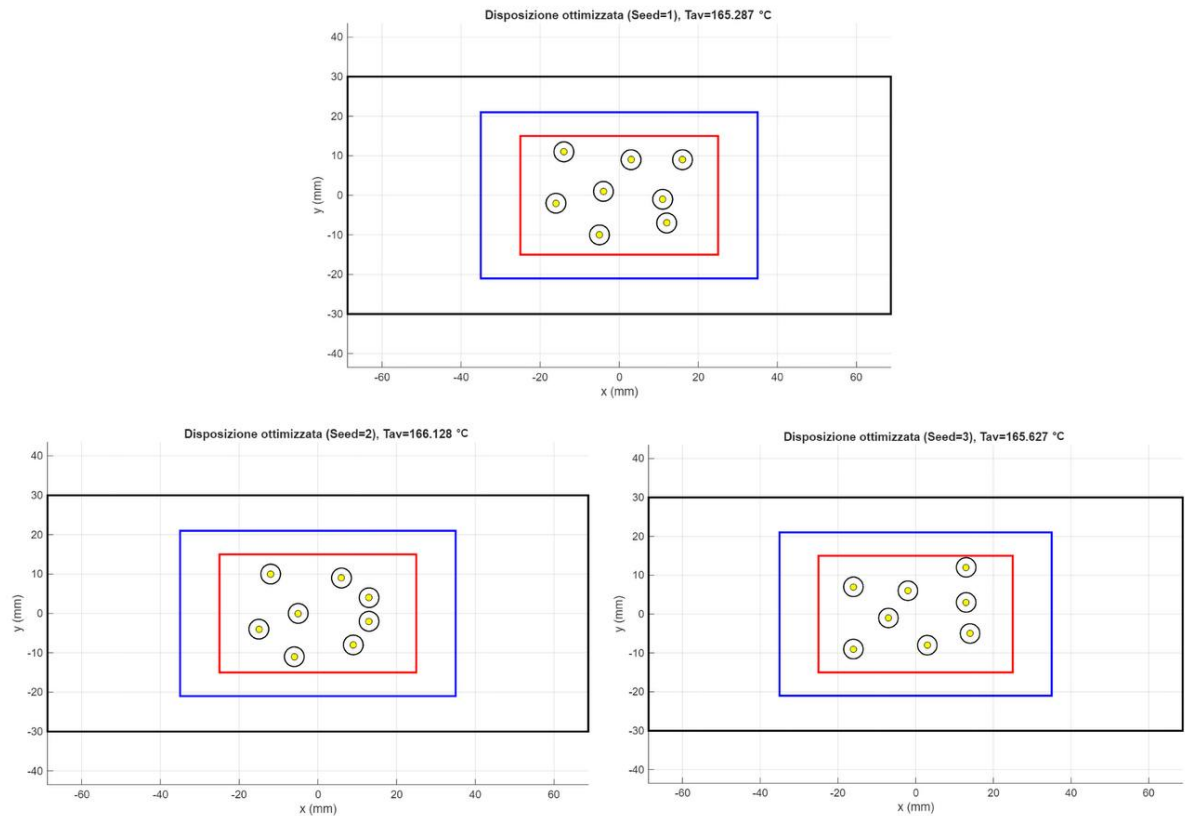


Figura 34: Configurazioni finali ottenute con Pattern Search a partire da Seed differenti (Seed 1–3), con i corrispondenti valori di T_{av} ; le soluzioni risultano differenti, evidenziando dipendenza dall'inizializzazione

4.3.3 Discussione dei risultati del secondo tentativo e conclusioni

In Tabella 8 sono riportati i risultati delle campagne svolte, al fine di fornire un quadro sintetico dei dati che verranno commentati nel seguito.

Seed	Configurazioni di partenza	Configurazioni ottimizzate	Func-Count Pattern Search Tempo di ottimizzazione Criterio di Stop
1	$Pos_{mm,0}=(29;1;12;-6;12;12;-30;9;31;9;-14;10;-25;-2;26;-7)$ mm $T_{av,0}=205.813$ °C	$Pos_{mm,ott}=(-4;1;-16;-2;16;9;12;-7;3;9;-14;11;11;-1;-5;-10)$ mm $T_{av,ott}=165.287$ °C	1403 37 h Mesh Tolerance
2	$Pos_{mm,0}=(-4;-17;3;-2;-5;-6;-19;4;-13;-8;8;1;-23;0;-20;10)$ mm $T_{av,0}=182.985$ °C	$Pos_{mm,ott}=(-6;-11;13;-2;-5;0;13;4;9;-8;6;9;-15;-4;-12;10)$ mm $T_{av,ott}=166.128$ °C	1511 49 h Mesh Tolerance
3	$Pos_{mm,0}=(-30;2;-15;-3;-14;7;-4;-12;3;10;-12;-10;-7;15;30;6)$ mm $T_{av,0}=190.838$ °C	$Pos_{mm,ott}=(14;-5;13;3;-16;7;13;12;3;-8;-16;-9;-7;-1;-2;6)$ mm $T_{av,ott}=165.627$ °C	1216 55 h Mesh Tolerance

Tabella 8: Risultati del secondo tentativo di ottimizzazione diretta

Le nuove campagne confermano la dipendenza dall'inizializzazione, infatti, le soluzioni finali risultano ancora diverse tra loro.

Tuttavia, rispetto al caso precedente, si osserva una riduzione della variabilità di T_{av} nelle fasi finali dell'ottimizzazione: a parità di piccoli spostamenti, la risposta termica risulta più coerente, come evidenziato in Figura 35. Questo elemento conferma che un'analisi della Mesh è essenziale per ottenere risultati numericamente affidabili. Tale miglioramento comporta però un incremento del costo computazionale: il tempo medio per una valutazione completa (aggiornamento geometria + simulazione CFD) passa da circa 2 a circa 3 minuti.

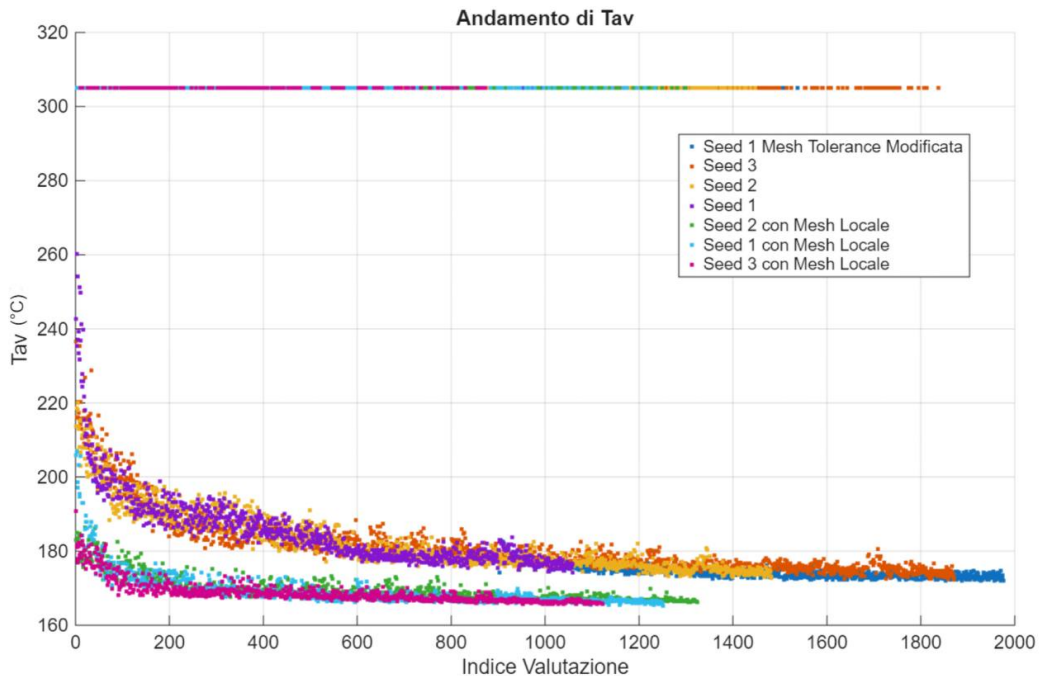


Figura 35: Evoluzione qualitativa delle ottimizzazioni

Nonostante quanto appena sottolineato, le soluzioni ottimizzate ottenute non presentano un pattern geometrico particolarmente significativo, oltre alla tendenza generale a collocare i cilindri nella regione di maggiore carico termico, come nel caso precedente. A livello qualitativo, le configurazioni risultano complessivamente più "ragionevoli", mostrando una distribuzione più

uniforme e alcuni elementi di simmetria globale (ad esempio un bilanciamento tra quadranti e tra semipiano superiore e inferiore), coerenti con una disposizione meno casuale dei pin-fin. Tuttavia, tre configurazioni non sono sufficienti per trarre conclusioni concrete.

In sintesi, la mitigazione “area ristretta + mesh locale” contribuisce a rendere più stabile l’andamento finale dell’algoritmo, ma non elimina la criticità strutturale della formulazione: lo spazio di progetto resta comunque troppo ampio. Alla luce di quanto discusso in precedenza, l’ipotesi che per far emergere un pattern robusto sia necessaria una campagna multi-Seed rimane quindi la più probabile. Tuttavia, le correzioni introdotte nel secondo tentativo rendono tale strategia ancora meno praticabile dal punto di vista computazionale: a parità di numero di Seed, l’aumento del costo per singola valutazione porta il tempo complessivo a crescere sensibilmente, passando da oltre quattro giorni (primo tentativo) a circa sei giorni.

A conclusione, si riassumono le principali criticità emerse che hanno motivato la riformulazione del problema:

- Sensibilità numerica alla Mesh: T_{av} risulta influenzato dal livello di discretizzazione; è quindi necessario un approccio che consenta un infittimento mirato senza rendere proibitivi i tempi di calcolo.
- Spazio di progetto troppo ampio: il posizionamento “libero” genera un numero elevatissimo di configurazioni, rendendo difficile far emergere un pattern univoco con poche ottimizzazioni.
- Costo computazionale proibitivo: una campagna multi-Seed estesa sarebbe necessaria per esplorare molti bacini di attrazione, ma risulta impraticabile poiché ogni Seed richiede numerose valutazioni CFD (pochi Seed comportano già diversi giorni di calcolo).

Per queste ragioni, nei capitoli successivi il focus si sposta dalla ricerca dell’ottimo “fine a sé stessa” con ottimizzazione diretta CFD, allo sviluppo di un metodo di ottimizzazione più robusto e scalabile: il problema viene riformulato riducendo i gradi di libertà (parametrizzazione più compatta), rendendo sostenibile la generazione di un dataset e l’adozione di modelli surrogati e ottimizzazione multi-obiettivo. Tale approccio viene prima verificato su un caso semplificato e successivamente esteso a scenari più applicativi.

5. Ottimizzazione multi-obiettivo assistita da surrogato

Nel Capitolo 4 è emerso come un approccio di ottimizzazione diretta diventi rapidamente proibitivo: il costo di valutazione di una singola configurazione rende poco scalabile l'esplorazione iterativa dello spazio di progetto. Inoltre, l'elevata dimensionalità del problema rende lo spazio di progetto difficilmente esplorabile e poco interpretabile.

Per superare tali limiti, questo capitolo introduce un *workflow* alternativo, basato su tre idee chiave:

1. riformulare il problema con una parametrizzazione discreta e compatta;
2. investire la CFD principalmente in una fase iniziale di generazione del dataset (Design of Experiments);
3. sostituire la CFD durante l'ottimizzazione con un modello surrogato basato su rete neurale, addestrato sul dataset generato.

Una volta addestrato, il modello è in grado di predire la temperatura media sulla superficie di interesse ($\hat{T}_{av}$) a partire dalle configurazioni di input con tempi di valutazione drasticamente inferiori rispetto a quelli richiesti dalle simulazioni CFD. In questo modo, la CFD viene impiegata principalmente nella fase di costruzione del dataset e successivamente riservata alla validazione finale dell'intero insieme di soluzioni.

L'obiettivo principale non è tanto “trovare l'ottimo” di questo specifico caso di studio, quanto sviluppare e validare un metodo generale di ottimizzazione. Il caso applicativo considerato è volutamente semplice dal punto di vista dell'intuizione umana (ad esempio, è naturale aspettarsi che convenga concentrare i pin nella zona più calda). Tuttavia, questa “ovvietà” non è disponibile al modello: la rete neurale non conosce la fisica del problema, ma apprende esclusivamente dai dati. In tale contesto, riuscire a discriminare configurazioni efficienti all'interno di uno spazio combinatorio enorme (dell'ordine di 15^9 possibili combinazioni) non è affatto scontato. Il valore del capitolo risiede quindi nel mostrare che il *workflow* è in grado di guidare in modo sistematico l'esplorazione del progetto e di individuare soluzioni competitive, dimostrando la potenzialità dell'approccio per problemi più complessi e meno “intuitivi”.

Il flusso complessivo del metodo, applicato al caso studio in esame, è sintetizzato nel diagramma a blocchi riportato in Figura 36, che anticipa la struttura del capitolo e i principali argomenti affrontati nelle sezioni successive. In particolare, il *workflow* comprende: la definizione di una parametrizzazione vincolata con variabili intere, la generazione del Design of Experiments, la valutazione automatica CFD delle configurazioni campionate, l'addestramento del modello surrogato, l'ottimizzazione multi-obiettivo sul surrogato per ottenere un insieme di soluzioni di Pareto e infine la verifica CFD delle configurazioni ottenute.

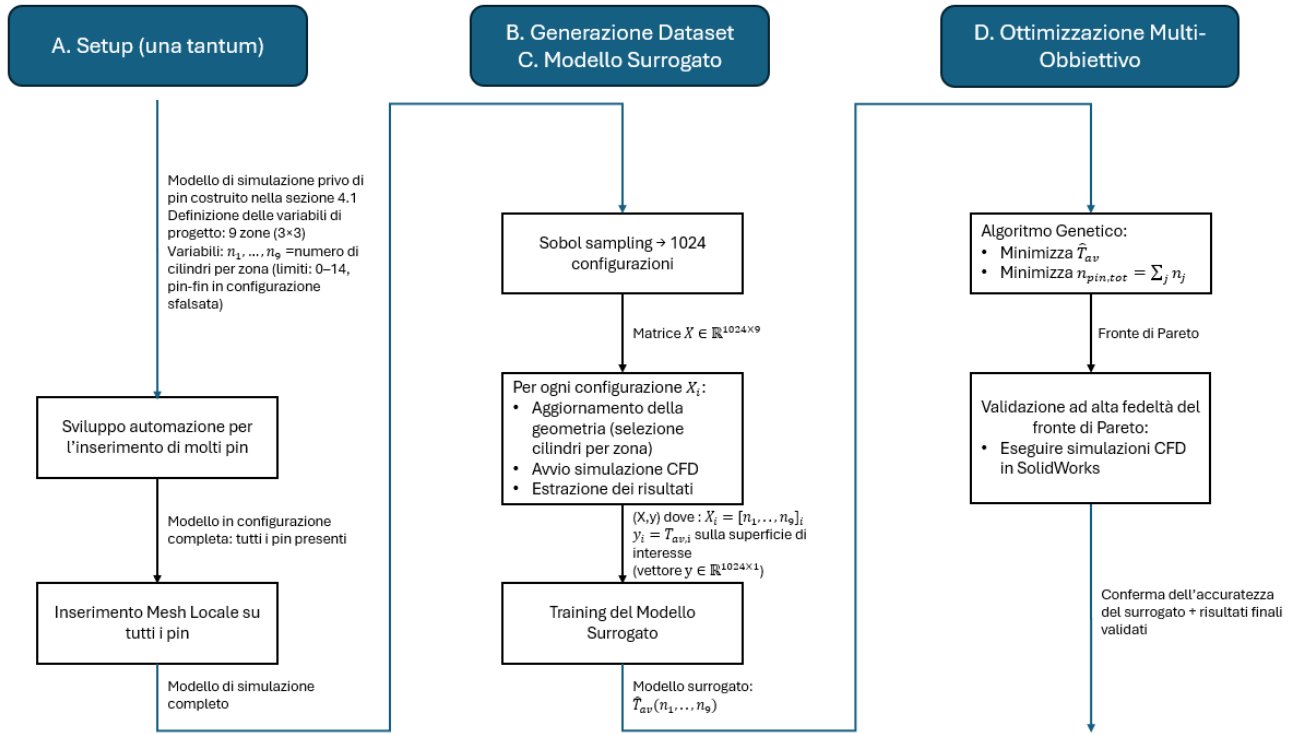


Figura 36: Diagramma a blocchi del Workflow complessivo proposto

5.1 Setup

Per rendere il problema trattabile senza ridurre eccessivamente la varietà delle configurazioni è stata adottata una parametrizzazione a zone. Invece di descrivere la geometria tramite la posizione di ogni singolo pin (approccio che porterebbe a un numero di combinazioni difficilmente gestibile), la superficie del dissipatore viene suddivisa in un numero limitato di regioni, e ogni configurazione è definita assegnando a ciascuna zona un numero intero di pin, inseriti in ordine e posizioni prestabilite. Questa scelta mantiene uno spazio di progetto ancora molto ampio, ma più strutturato e interpretabile. In questo contesto, lo scopo dell'ottimizzazione diventa quello di minimizzare la temperatura media sulla superficie di interesse T_{av} e, al tempo stesso, il numero complessivo di cilindri $n_{pin,tot}$ (multi-obiettivo). Perciò il problema consiste sostanzialmente nel determinare quanti pin distribuire per zona, in modo da collocarli nelle regioni più efficaci dal punto di vista termico.

Il modello di base è stato mantenuto invariato rispetto a quello introdotto nella Sezione 4.1 (Figura 23), così come le dimensioni dei pin (diametro pari a 5 mm, altezza pari a 6 mm, presentate nella Sezione 4.3). Le modifiche apportate in questo capitolo riguardano principalmente la gestione parametrica e automatizzata dell'inserimento/rimozione dei pin. Anche le impostazioni della simulazione CFD sono state mantenute coerenti con quelle adottate nella Sezione 4.1 (Tabella 2), l'unica eccezione riguarda la Mesh, che è stata rivista per gestire correttamente la presenza dei pin e per contenere il costo computazionale, come discusso nelle sezioni successive.

5.1.1 Parametrizzazione compatta

Il dissipatore è stato suddiviso in una griglia 3x3, generando 9 regioni rettangolari numerate (zone), come rappresentato in Figura 37.

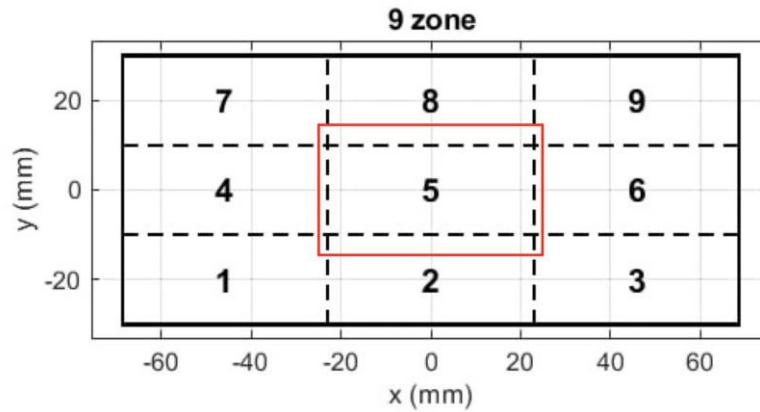


Figura 37: Suddivisione della superficie del dissipatore in 9 zone (griglia 3x3) utilizzata per la parametrizzazione a zone. In rosso è evidenziata la zona calda, così da chiarirne la posizione rispetto alle nove regioni considerate

Il vettore decisionale è definito come:

$$x = [n_1, n_2, \dots, n_9]$$

dove n_j è il numero intero di pin da inserire nella regione j . Per garantire la fattibilità geometrica, ogni variabile è vincolata dalla capacità c della regione:

$$0 \leq n_j \leq 14$$

Poiché le nove regioni hanno uguale estensione, il valore massimo ($c = 14$) risulta identico per ciascuna zona. I cilindri sono inoltre posizionati in modo da ottenere, nel caso di riempimento completo, una configurazione sfalsata sull'intera superficie (disposizione tipicamente impiegata nei dissipatori a pin-fin). Ciò implica che, oltre a un ordine di inserimento locale all'interno di ogni zona, esiste anche un criterio di ordinamento globale coerente tra le diverse regioni, tale da preservare lo schema sfalsato complessivo. All'interno di ciascuna zona, i pin vengono attivati secondo una sequenza deterministica basata su una scansione per colonne: si parte dalla colonna centrale della zona e si procede verso le colonne più esterne, all'interno di ogni colonna, l'ordine è definito dall'alto verso il basso; infine, per colonne equidistanti dalla colonna centrale, viene assegnata priorità alla colonna a sinistra e successivamente a quella a destra. Questo criterio è illustrato in Figura 38, dove i cilindri sono numerati in due zone rappresentative, evidenziando la sequenza di attivazione delle posizioni.

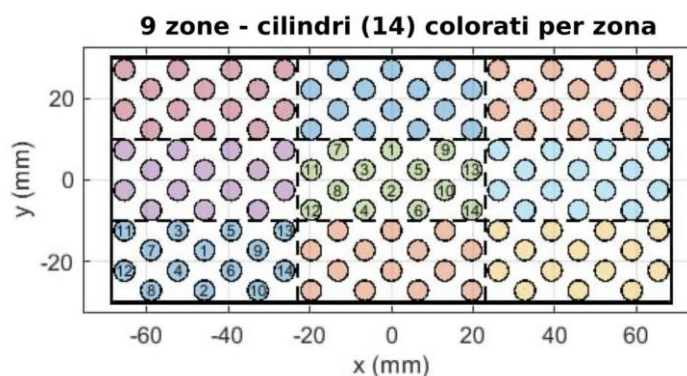


Figura 38: Configurazione a riempimento completo (tutti i pin presenti) e relativo ordine deterministico di inserimento all'interno delle zone, mostrato su due zone rappresentative

In Figura 39 sono riportate le distanze relative tra i pin. In particolare, tra colonne successive i pin non risultano perfettamente centrati rispetto ai due pin della colonna precedente: è stata infatti introdotta una lieve traslazione della griglia di posizionamento per garantire che ciascun cilindro sia interamente contenuto all'interno della propria zona, evitando casi in cui un pin possa ricadere a cavallo del confine tra due regioni.

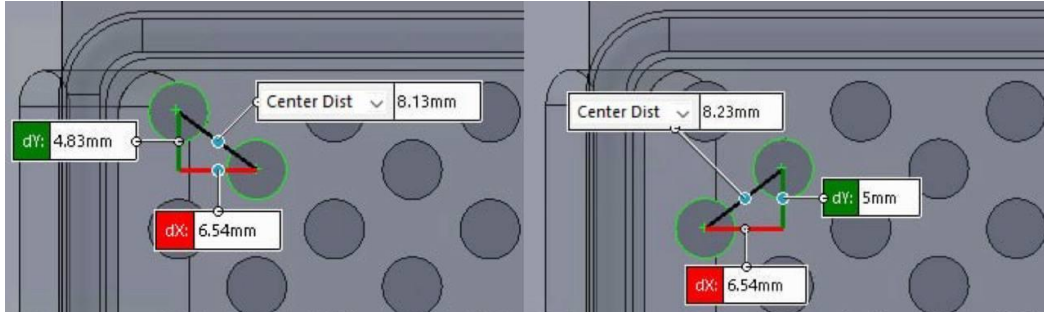


Figura 39: Dettaglio del posizionamento dei pin-fin

In Tabella 9 sono infine riassunte tutte le quote geometriche per il posizionamento dei pin-fin, così da rendere completamente esplicita la configurazione di riferimento e garantirne la riproducibilità all'interno della procedura automatizzata implementata in MATLAB.

Posizionamento Pin-Fin globale	
Zona globale utile posizionamento Pin	137,4 mm x 60 mm
Zona globale effettiva posizionamento Pin	135,8 mm x 59 mm
Distanza centro Pin esterni - bordi globali (sinistro e destro)	2,5 mm (raggio) + 0,8 mm (spazio) = 3,3 mm
Distanza centro Pin esterni - bordi globali (superiore e inferiore)	2,5 mm (raggio) + 0,5 mm (spazio) = 3 mm
Distanza tra centri Pin colonne adiacenti lungo x ('dx' Figura 39)	6,54 mm
Distanza tra centri Pin righe adiacenti (su stessa colonna, somma 'dy' Figura 39)	9,83 mm
Sfalsamento verticale tra colonne adiacenti ('dy' Figura 39, sinistra)	4,83 mm
Colonne totali di Pin	21
Righe di Pin per colonna	6

Tabella 9: Quote posizionamento pin

5.1.2 Preparazione del modello CAD per la generazione del dataset

Alla luce di quanto emerso nel Capitolo 4, la Mesh rappresenta un elemento determinante sia per l'affidabilità numerica di T_{av} , sia per il costo computazionale della singola valutazione. Perciò si sono fatte diverse considerazioni:

- Un infittimento globale della Mesh aumenterebbe l'accuratezza, ma a fronte di un incremento di tempo non sostenibile su grandi campagne di simulazioni.
- Un raffinamento locale "a zona" non è praticabile in modo efficace perché, per come è formulato il problema, non è noto a priori dove verranno collocati i cilindri nelle diverse configurazioni. Per garantire che il raffinamento includa sempre i pin-fin, la regione di Mesh locale dovrebbe quindi coprire una porzione molto estesa del dominio (di fatto quasi tutta l'area utile), annullando il vantaggio del raffinamento locale. Inoltre, in generale, introduce celle aggiuntive anche dove non strettamente necessario.
- Il compromesso ritenuto più efficace è l'adozione di una Mesh locale direttamente sui cilindri, in modo da catturare correttamente il loro contributo fluidodinamico e termico limitando al minimo l'impatto sul resto del dominio e quindi sui tempi di calcolo.

Tuttavia, nel contesto dell'automazione, SolidWorks non consente di applicare la Mesh locale ai singoli cilindri tramite API. Di conseguenza è stata adottata una strategia operativa in tre passi:

1. generare automaticamente una configurazione a riempimento completo, inserendo tutti i cilindri;
2. impostare manualmente la Mesh locale su un cilindro e replicarla su tutti gli elementi identici, operazione rapida tramite interfaccia;
3. ottenere le configurazioni richieste per sottrazione, partendo ogni volta dalla configurazione completa di riferimento, eliminando automaticamente i cilindri non necessari e salvando la configurazione risultante in un file separato, che viene sovrascritto ad ogni iterazione; la configurazione completa, invece, viene mantenuta inalterata come *template* permanente.

Questa strategia si discosta in modo significativo da quanto adottato nel Capitolo 4. In quel caso la geometria veniva aggiornata inserendo un numero limitato di cilindri, richiamando ripetutamente una macro VBA in un ciclo (un "for" lato MATLAB). Tale schema è adeguato quando il numero di elementi da aggiungere è ridotto, ma degrada rapidamente passando a decine o centinaia di pin, poiché l'*overhead* delle chiamate VBA diventa dominante e rende l'aggiornamento geometrico eccessivamente lento.

Nel presente capitolo, poiché le configurazioni vengono generate per sottrazione a partire dalla configurazione a riempimento completo, l'operazione critica dal punto di vista prestazionale (lato CAD) è la rimozione dei cilindri, che viene eseguita ad ogni nuova configurazione. Per questo motivo è stata sviluppata un'automazione dedicata in grado di eliminare tutti i cilindri non necessari con un'unica chiamata, evitando cicli di cancellazione che renderebbero i tempi complessivi proibitivi. L'inserimento dei cilindri, essendo invece un'operazione svolta una sola volta (per creare la configurazione completa su cui impostare la Mesh locale), non richiederebbe strettamente la stessa ottimizzazione; tuttavia, per completezza del processo, anche questa fase è stata implementata tramite una singola invocazione della macro.

Per rendere l'operazione di eliminazione non solo rapida ma anche corretta e ripetibile, è stata inoltre implementata una strategia di indicizzazione univoca dei cilindri. In pratica, a ciascun pin è stato assegnato un identificatore stabile che permette di selezionare ed eliminare esattamente gli elementi corrispondenti alle posizioni "non attive" per una data configurazione. In questo modo, la rimozione dei cilindri segue in modo coerente la stessa logica deterministica di posizionamento descritta nel paragrafo precedente, garantendo la corrispondenza uno a uno tra vettore decisionale in MATLAB e geometria effettiva in SolidWorks.

In Figura 40 è riportato il modello di simulazione nella configurazione completa, in cui viene anche evidenziato visivamente l'effetto del raffinamento di Mesh locale applicato ai cilindri.

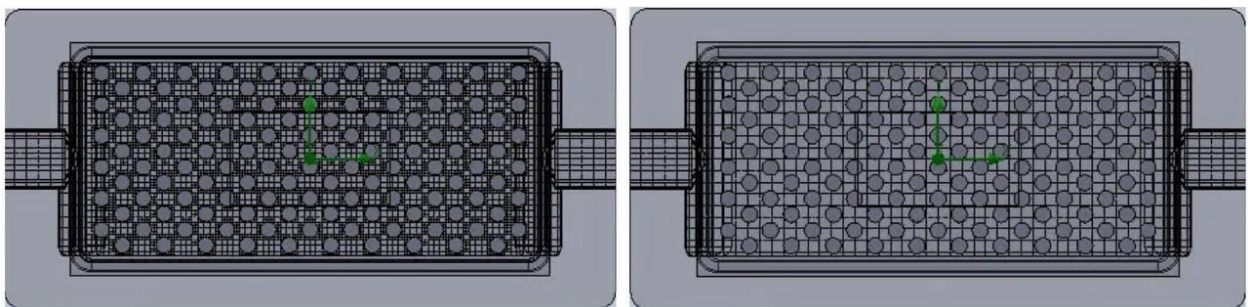


Figura 40: Configurazione a riempimento completo (tutti i cilindri presenti), con Mesh locale sui cilindri (a sinistra) e senza (a destra); immagini speculari per avere coerenza con MATLAB

5.1.3 Studio preliminare della Mesh locale: accuratezza vs costo

Poiché T_{av} è risultato sensibile alla Mesh nel capitolo precedente, è stato necessario stabilire un livello di infittimento che garantisse risultati affidabili senza rendere proibitivi i tempi di calcolo. Per questo motivo, prima di produrre il dataset completo, è stato condotto uno studio preliminare di sensibilità della Mesh locale.

Lo studio è stato condotto inizialmente su una parametrizzazione più densa, considerando fino a 21 pin per zona, poiché in una prima fase si era ipotizzato di utilizzare tale configurazione per la generazione del dataset. Successivamente, tuttavia, questa soluzione è risultata eccessivamente densa dal punto di vista progettuale e si è scelto di adottare una configurazione più “spaziata”, pari a 14 pin per zona. La scelta di mantenere il livello di Mesh locale selezionato a partire dalle prove iniziali è giustificata da due considerazioni.

1. Le configurazioni con 21 pin per zona rappresentano un caso più gravoso e conservativo dal punto di vista numerico: la presenza di un maggior numero di cilindri e di interstizi più stretti aumenta la complessità dei campi locali (gradienti più marcati e interazioni più ravvicinate), rendendo la discretizzazione più critica. Di conseguenza, un livello di Mesh che risulta adeguato in tali condizioni può essere considerato una scelta cautelativa anche per configurazioni meno dense.
2. Il raffinamento adottato è definito localmente sui cilindri e quindi è principalmente legato alla risoluzione della geometria elementare (diametro, regioni di separazione e strato limite) più che al numero massimo di pin presenti. Il passaggio a 14 pin per zona modifica la densità globale della matrice di pin, ma non altera la scala geometrica del singolo elemento su cui la Mesh locale è applicata.

Per queste ragioni, si è ritenuto non necessario ripetere lo studio di sensibilità, mantenendo il livello selezionato come compromesso affidabile tra accuratezza e costo computazionale.

Chiarita questa premessa, nel seguito vengono descritti la metodologia adottata per l'analisi di sensibilità della Mesh locale e i principali risultati ottenuti, con l'obiettivo di motivare la scelta finale del livello di discretizzazione impiegato per la generazione del dataset.

Poiché l'analisi ha coinvolto un numero elevato di combinazioni (più geometrie e più livelli di discretizzazione), è stata sviluppata anche in questo caso un'automazione dedicata, in grado di gestire la preparazione delle configurazioni, l'impostazione del livello di Mesh locale, l'avvio delle simulazioni e l'estrazione dei risultati (T_{av} sulla zona di interesse e i tempi di simulazione). In particolare, sono state predisposte 10 geometrie differenti: 8 configurazioni casuali, più due casi scelti intenzionalmente come estremi, ovvero una configurazione completamente popolata (21 cilindri in ciascuna zona) e una configurazione poco popolata, con 21 cilindri solo nella zona 5 (zona centrale) e le restanti zone vuote; questo per valutare tempi e sensibilità anche in condizioni limite. Per ciascuna geometria è stata poi eseguita una campagna di simulazioni variando il livello di Mesh locale sui cilindri da 1 a 6, dove il parametro di infittimento è definito in modo coerente con quanto introdotto nel capitolo precedente (Paragrafo 4.3.1). Non si è esteso lo studio oltre il livello 6 perché i tempi di calcolo raggiungevano già l'ordine delle ore per singola simulazione, rendendo poco utile analizzare ulteriori livelli di accuratezza non sostenibili nell'ambito della generazione del dataset.

In figura 41 si può osservare l'esempio di 3 casi (random, poco popolata, completamente popolata).

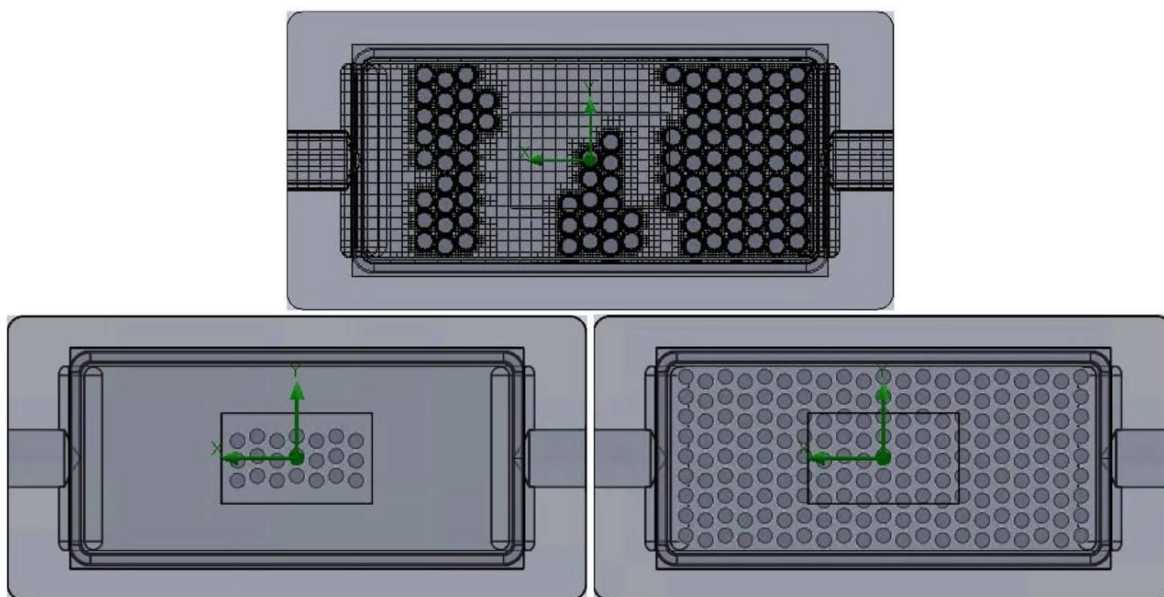


Figura 41: Esempio di tre configurazioni utilizzate nello studio di sensibilità della Mesh, in ordine: una configurazione random, una poco popolata (21 pin solo nella zona centrale) e una completamente popolata. Per la configurazione random è inoltre mostrato il dettaglio della Mesh locale applicata ai cilindri con livello di infittimento 6

Per evitare un'eccessiva frammentazione grafica, non vengono mostrati i risultati per ogni configurazione singolarmente. A titolo esemplificativo, viene invece riportato in Figura 42 l'andamento completo dell'analisi su una delle configurazioni casuali, mentre le figure successive sintetizzano l'intervallo di variazione (errori e tempi) osservato sull'intero insieme di geometrie considerate.

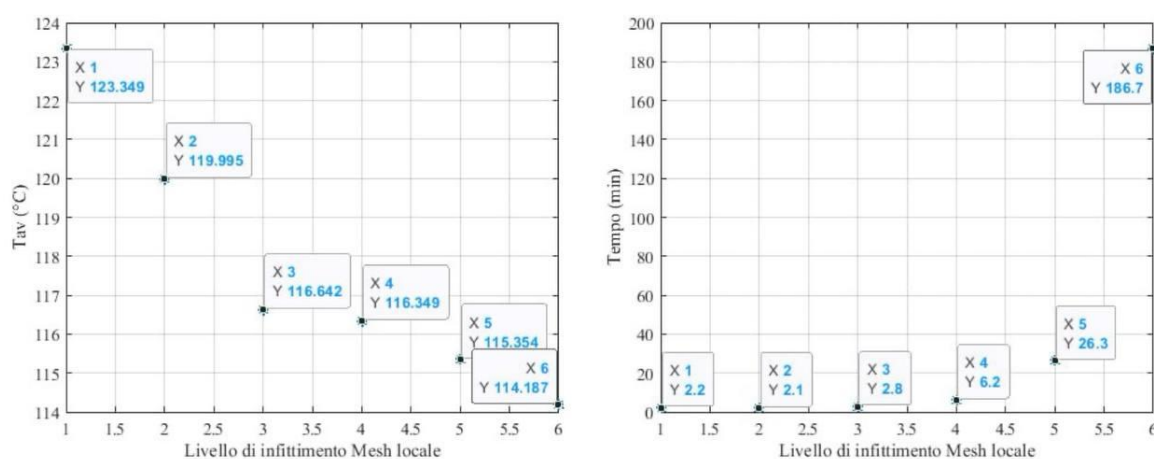


Figura 42: Esempio analisi della Mesh su una configurazione casuale: andamento di T_{av} (a sinistra) e del tempo di simulazione (a destra) al variare del livello di infittimento della mesh locale (1-6)

I risultati mostrano un comportamento tipico:

- T_{av} varia al crescere del livello della Mesh locale, ma oltre un certo infittimento le variazioni diventano progressivamente meno marcate.
- Il tempo di simulazione aumenta all'aumentare del livello della Mesh locale, con un andamento non lineare.

Si osserva tuttavia un'eccezione tra i livelli di infittimento 1 e 2, per cui il tempo risulta leggermente inferiore al livello 2 rispetto al livello 1. Questo andamento non è legato alla Mesh in sé, ma al fatto che l'infittimento 1 coincide con la prima esecuzione sul modello: in questa fase è necessario completare operazioni di preparazione (inizializzazione dello studio, creazione di cartelle e file temporanei, ecc.). Nelle esecuzioni successive, variando solo il livello della Mesh, tali strutture risultano già predisposte e il tempo complessivo si riduce.

Nelle Figure 43 e 44 i risultati sono presentati in forma sintetica. In particolare, in Figura 43 è riportata l'errore percentuale di T_{av} al variare del livello della Mesh rispetto alla soluzione di riferimento (livello 6). Per ogni livello sono inoltre indicati i valori di errore minimo e massimo riscontrati tra le configurazioni analizzate.

L'errore è definito nel modo seguente:

$$\text{Errore}(L) (\%) = 100 \cdot \frac{|T_{av}(L6) - T_{av}(L)|}{T_{av}(L6)} \quad (25)$$

La presenza dei casi estremi viene mostrata separatamente rispetto al solo insieme random, così da chiarire il contributo delle geometrie più gravose sull'errore e sui tempi.

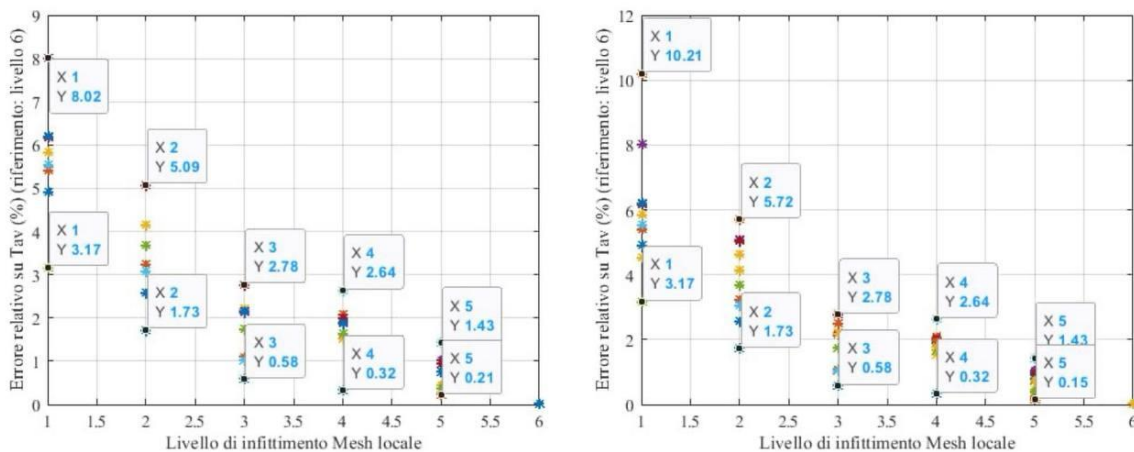


Figura 43: Errore percentuale su T_{av} al variare del livello di infittimento della Mesh locale, calcolato rispetto alla soluzione di riferimento (livello 6). A sinistra sono riportati i risultati ottenuti sulle sole configurazioni random; a destra sono inclusi anche i due casi estremi (configurazione completamente popolata e configurazione poco popolata)

Dall'analisi in Figura 43 si osserva che, nella maggior parte dei livelli di infittimento, l'inclusione delle due configurazioni estreme non altera in modo significativo l'intervallo di errore ottenuto sulle sole configurazioni random: gli estremi rientrano nel range già coperto dai casi casuali. Fanno eccezione i livelli 1, 2 e 5.

In particolare, la configurazione "21 pin solo in zona 5" (centrale) presenta errori più elevati ai livelli 1 e 2, innalzando il valore massimo osservato; al livello 5, invece, porta a uno scostamento più contenuto, abbassando il valore minimo. Questo comportamento è coerente con la natura della configurazione: concentrando tutti i pin nell'area centrale, l'effetto su T_{av} è più marcato, rendendo la risposta più sensibile alla discretizzazione rispetto a configurazioni con pin distribuiti su più zone.

In Figura 44 è mostrato l'andamento dei tempi di simulazione, evidenziando i valori massimi per ciascun livello. Per le configurazioni random, al livello 6 è inoltre indicato il valore minimo, così da evidenziare che anche nel caso più favorevole all'interno del set random il tempo di calcolo risulta comunque dell'ordine di circa un'ora per singola simulazione.

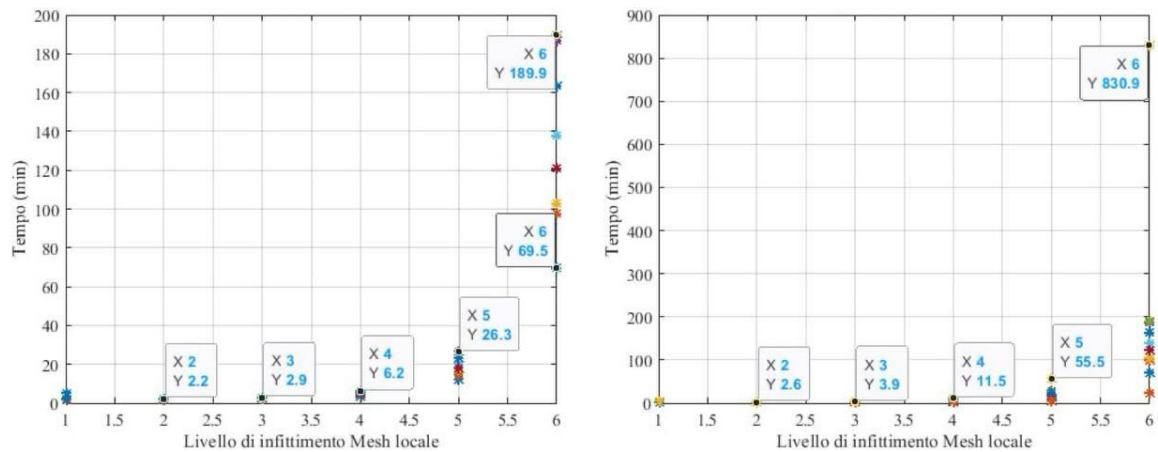


Figura 44: Tempi di simulazione al variare del livello di infittimento della Mesh locale. A sinistra sono riportati i risultati ottenuti sulle sole configurazioni random; a destra sono inclusi anche i due casi estremi (configurazione completamente popolata e configurazione poco popolata)

Dalla Figura 44 e dal relativo zoom in Figura 45 si osserva che, a differenza di quanto accade per gli errori su T_{av} , l'inclusione delle configurazioni estreme influenza in modo sistematico anche la distribuzione dei tempi di simulazione. In particolare, la configurazione completamente popolata tende ad aumentare i tempi massimi a tutti i livelli di infittimento, mentre la configurazione "21 pin solo in zona 5" abbassa i tempi minimi. Questo comportamento è atteso: al crescere del numero di cilindri aumenta il numero di regioni soggette a Mesh locale e, di conseguenza, la discretizzazione complessiva risulta più fitta e il costo computazionale cresce.

Considerando congiuntamente i grafici di tempo e gli errori percentuali su T_{av} , i livelli di infittimento 3 e 4 risultano i più promettenti in termini di compromesso tra affidabilità e costo computazionale. Dallo zoom in Figura 45 si sceglie infine di adottare il livello 3: a fronte di errori percentuali comparabili rispetto al livello 4, il tempo medio per simulazione risulta decisamente inferiore. Anche differenze dell'ordine del minuto, infatti, diventano significative su una campagna di 1024 simulazioni, traducendosi in diverse ore aggiuntive di calcolo complessivo.

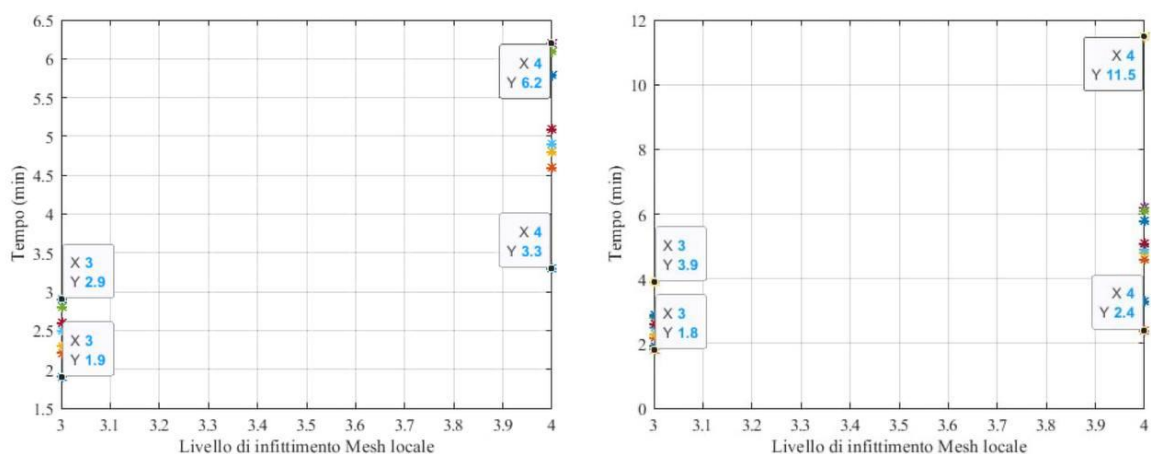


Figura 45: Zoom dei tempi di simulazione al variare del livello di infittimento della Mesh locale, focalizzato sui livelli più rilevanti per il compromesso tra costo computazionale e affidabilità. A sinistra sono riportati i risultati ottenuti sulle sole configurazioni random; a destra sono inclusi anche i due casi estremi (configurazione completamente popolata e configurazione poco popolata)

In conclusione, adottando un livello di infittimento pari a 3 si attende di ottenere risultati sufficientemente affidabili ad un costo computazionale ragionevole:

- lo scostamento di T_{av} risulta compreso indicativamente tra compreso tra 0.58% e 2.78% rispetto alla soluzione ottenuta con il livello 6, assunto come riferimento più accurato;
- il livello 6, tuttavia, richiede tempi di calcolo dell'ordine delle ore per una singola simulazione. Con il livello 3, il tempo di simulazione per configurazione si mantiene invece nell'intervallo di circa 2–4 minuti, risultando compatibile con la generazione del dataset su larga scala.

Va osservato inoltre che le configurazioni considerate nelle analisi successive risultano, in media, meno dense rispetto a quelle adottate in questa fase preliminare, poiché il numero massimo di cilindri per zona viene ridotto da 21 a 14. È quindi ragionevole attendersi che, a parità di livello di Mesh locale, i tempi di simulazione possano risultare leggermente inferiori a quelli qui stimati.

5.2 Generazione del dataset mediante sequenze di Sobol

Poiché l'addestramento della rete neurale richiede un dataset rappresentativo della relazione tra configurazione geometrica e risposta termica, si è reso necessario utilizzare una procedura per selezionare in modo efficace le configurazioni da simulare per creare il dataset. A tale scopo, è stato adottato un campionamento generato in MATLAB mediante la funzione “*sobolset*”^[26], disponibile all'interno dell'Add-On Statistics and Machine Learning Toolbox. Sobolset produce una sequenza di Sobol nell'iper-cubo unitario $[0,1]^d$, dove d rappresenta la dimensione dello spazio di progetto, cioè il numero di variabili di input. Le sequenze di Sobol sono qui introdotte non con l'obiettivo di fornirne una trattazione esaustiva, ma per richiamare i principi che ne motivano l'impiego nella definizione del piano di campionamento. Esse appartengono ai metodi quasi-Monte Carlo (QMC) e sono progettate per ottenere punti ben distribuiti nello spazio degli input^[27], risultando adatte alla costruzione di DoE *space-filling* per esperimenti numerici costosi. I generatori QMC producono campioni altamente uniformi nell'iper-cubo e tendono a riempire sistematicamente eventuali “buchi” nei primi segmenti della sequenza, migliorando la copertura del dominio rispetto a estrazioni puramente casuali (Monte Carlo).

Nel presente caso, l'obiettivo del DoE non è esplorare in modo esaustivo l'insieme delle configurazioni possibili (che sarebbe computazionalmente inaccessibile), ma costruire un dataset rappresentativo dell'intero dominio delle variabili a zona, così da addestrare un modello surrogato robusto. Per questo scopo, l'uso di Sobol presenta i seguenti vantaggi:

1. distribuire i campioni in modo regolare sull'iper-cubo $[0,1]^d$, garantendo una copertura *space-filling* del dominio degli input;
2. essere riproducibile, grazie alla natura deterministica della sequenza (e alla possibilità di fissare le opzioni di generazione/scrambling);
3. ridurre correlazioni spurie tra variabili introdotte dal piano di campionamento: la copertura *space-filling* favorisce l'esplorazione di combinazioni più varie delle *feature*, rendendo l'apprendimento della rete più stabile e limitando la ridondanza informativa tra input;
4. incrementare progressivamente la numerosità del dataset senza rigenerare l'intero piano di campionamento, permettendo di estendere la campagna “a blocchi” (ad esempio 256→512→1024) senza perdere le simulazioni già eseguite.

Quest'ultimo aspetto è particolarmente utile in ambito CFD, dove la campagna può essere spezzata in più *run* e i risultati intermedi possono essere già utilizzati per valutare la qualità del surrogato.

Di seguito viene descritto come è stato utilizzato Sobolset per generare le configurazioni del dataset, specificando la mappatura dalle variabili continue ai conteggi interi per zona e l'integrazione con la pipeline automatizzata MATLAB–SolidWorks. Il frammento principale di codice relativo alla generazione della sequenza di campionamento è riportato nel Codice 8.

```
% Creazione generatore di sequenza di Sobol
sob = sobolset(d, 'Skip', 1024);
% Applicazione scrambling
sob = scramble(sob, 'MatousekAffineOwen');
% Generazione punti della sequenza
U = net(sob, Nsamples);
% Clippaggio valori per evitare esattamente 0 o 1
U = min(max(U, 1e-12), 1-1e-12);
```

Codice 8: MATLAB, impostazione e utilizzo di Sobolset

La sequenza di Sobol è stata generata in MATLAB definendo innanzitutto la dimensione d dello spazio degli input ($d = 9$, una variabile per ciascuna zona). Il point set è stato quindi creato tramite:

- Definizione del generatore Sobol e salto iniziale (“*skip*”) dei primi punti della sequenza. Tale scelta è adottata per evitare che il campionamento inizi con un segmento iniziale potenzialmente meno “equilibrato” e per ottenere, fin dalle prime configurazioni utilizzate nel dataset, una copertura più regolare dell’iper-cubo degli input.
- Applicazione di uno *scrambling* (randomizzazione controllata) alla sequenza. Lo *scrambling* mantiene le proprietà di uniformità (*low-discrepancy*) tipiche del campionamento Sobol, ma riduce la presenza di pattern deterministici che potrebbero riflettersi in correlazioni spurie o in una copertura meno efficace di alcune proiezioni dello spazio. Per garantire la riproducibilità dello scrambling, è stato fissato il Seed del generatore pseudo-casuale in MATLAB, `rng(42)`, in modo da ottenere sempre la stessa sequenza campionata a parità di impostazioni. Il valore del Seed è stato scelto arbitrariamente, avendo il solo scopo di rendere riproducibile la generazione della sequenza senza attribuire al valore numerico selezionato alcun significato particolare.
- Estrazione dei primi N_{samples} punti della sequenza: grazie alla struttura delle sequenze di Sobol, ogni prefisso (ad esempio i primi 256 punti) costituisce già un campionamento *space-filling* dell’intero iper-cubo degli input e, aumentando progressivamente il numero di punti preferibilmente in potenze di due ($N = 2^k$, ad esempio $256 \rightarrow 512 \rightarrow 1024 \rightarrow 2048 \rightarrow 4096$), si ottiene un infittimento della copertura senza dover rigenerare i campioni precedenti. Per garantire coerenza tra i diversi step, la sequenza è stata generata una sola volta fissando un massimo $N_{\text{samples}} = 4096$; le campagne successive possono quindi utilizzare porzioni consecutive della stessa sequenza (ad esempio: prima campagna punti 1–256, seconda 257–512, terza 513–1024, e così via). In questo modo il numero di punti effettivamente simulati N viene scelto progressivamente in funzione del trade-off tra costo computazionale della campagna CFD e qualità del surrogato.
- Dopo la generazione, si ottiene una matrice $U \in \mathbb{R}^{4096 \times 9}$, dove le 4096 righe corrispondono alle configurazioni campionate e le 9 colonne alle nove variabili di input (una per ciascuna zona). In particolare, per la configurazione i -esima la riga:

$$U(i,:) = [u_1, \dots, u_9]_i$$

contiene i nove valori $u_j \in [0,1]$ che verranno successivamente mappati nei conteggi interi di cilindri per le rispettive zone. I campioni U vengono quindi limitati nell’intervallo $(\varepsilon, 1 - \varepsilon)$, con $\varepsilon = 10^{-12}$, per evitare casi limite in cui u_j assuma valori estremi (0 o 1) che, in seguito alla trasformazione discreta verso il numero di cilindri per zona (interi nell’intervallo 0–14), potrebbero generare valori al di fuori dei limiti imposti.

I punti della sequenza di Sobol, generati in $[0,1]^d$, devono essere convertiti nel vettore decisionale discreto: $x = [n_1, n_2, \dots, n_9]$. Tale mappatura avviene tramite:

$$n_j = \lfloor u_j (c + 1) \rfloor \quad (26)$$

con $c = 14$, pari alla capacità massima definita in precedenza. In questo modo ogni n_j assume valori interi compresi tra 0 e 14 inclusi.

Una volta ottenuti i vettori decisionali discreti, è quindi possibile costruire il dataset vero e proprio. In particolare, l'insieme delle configurazioni campionate costituisce la matrice delle *feature*:

$$X \in \mathbb{R}^{N \times 9}$$

in cui ogni riga contiene il vettore $[n_1, \dots, n_9]_i$ associato a una specifica configurazione. Per ciascuna riga di X , il workflow automatizzato (descritto in precedenza), aggiorna la geometria CAD del dissipatore, esegue la corrispondente simulazione CFD in SolidWorks Flow Simulation ed estrae il valore di temperatura media $T_{av,i}$. Tali valori vengono raccolti nel vettore *target*:

$$y \in \mathbb{R}^{N \times 1},$$

dove $y_i = T_{av,i}$ rappresenta la risposta termica associata alla configurazione i -esima. Il dataset finale risulta pertanto formato da coppie input-output (X, y) .

Per verificare la corretta corrispondenza tra la configurazione generata in ambiente MATLAB e la geometria effettivamente analizzata in SolidWorks, è utile riportare un esempio rappresentativo. In Figura 46 è mostrato il confronto tra la disposizione dei cilindri ottenuta mediante lo script MATLAB e il corrispondente modello CAD generato in SolidWorks, così da evidenziare la coerenza tra la codifica discreta del vettore decisionale e la sua effettiva traduzione geometrica. Nella stessa figura, il modello SolidWorks è riportato con la Mesh, in modo da mostrare come, nell'ambito della procedura automatizzata, il raffinamento locale venga mantenuto unicamente in corrispondenza dei cilindri effettivamente presenti nella configurazione considerata.

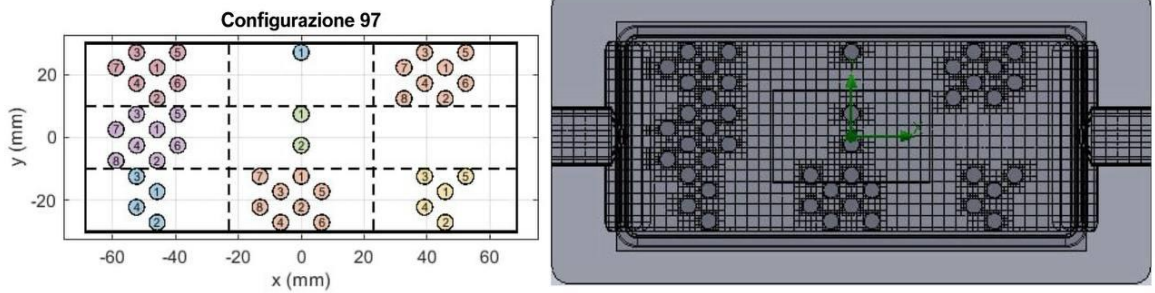


Figura 46: Confronto tra la configurazione ottenuta in MATLAB e il corrispondente modello CAD in SolidWorks per la 97-esima configurazione, associata a $x = [4, 8, 5, 8, 2, 0, 7, 1, 8]$. Per agevolare il confronto visivo, l'immagine di SolidWorks è stata specchiata

In Figura 47 è riportata la mappa di temperatura della stessa configurazione, inserita a completamento dell'esempio. La distribuzione di temperatura mostra in modo chiaro il contributo dei cilindri al comportamento termico della configurazione, fornendo un riscontro qualitativo della bontà della Mesh adottata.

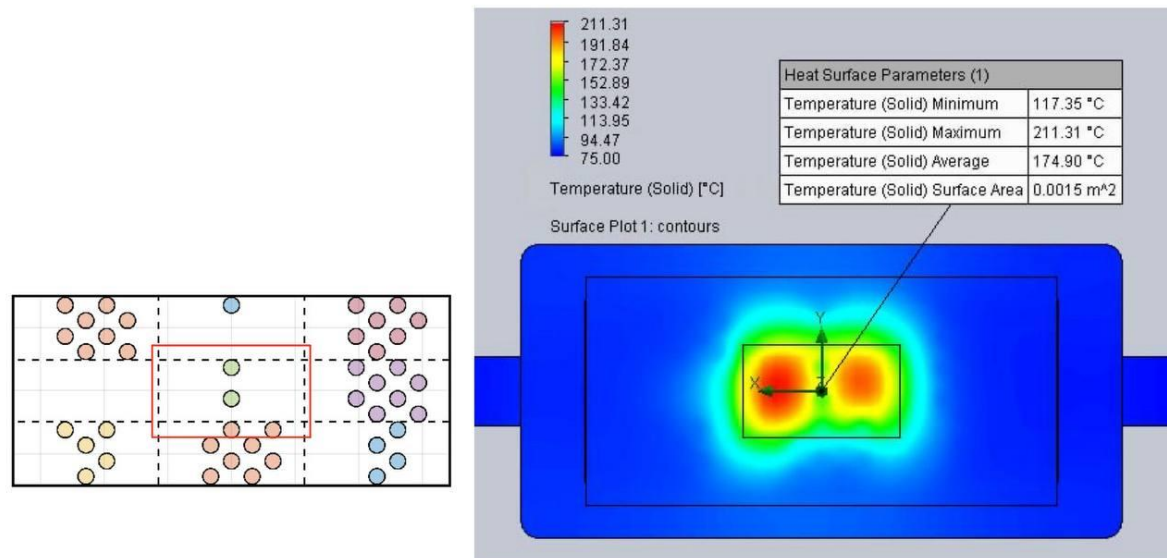


Figura 47: Mappa di temperatura ottenuta mediante simulazione CFD in SolidWorks Flow Simulation per la 97-esima configurazione del dataset. Per agevolare il confronto visivo, la rappresentazione della stessa configurazione in MATLAB è stata specchiata rispetto all'orientamento originale; inoltre, sono stati rimossi i numeri identificativi dei cilindri ed è stata aggiunta l'area calda

Per completare la descrizione della fase di costruzione del dataset, si riporta infine una stima del costo computazionale complessivo della pipeline automatizzata, che costituirà un utile termine di confronto con i tempi di valutazione del modello surrogato nelle fasi successive.

Tale stima è stata ottenuta misurando il tempo richiesto per un piccolo lotto di 10 configurazioni, comprensivo sia della fase di aggiornamento geometrico sia dell'esecuzione della simulazione CFD. Il tempo totale misurato è risultato pari a circa 1930,2 s, corrispondenti in media a circa 193 s per configurazione, ovvero poco più di 3 minuti.

Assumendo questo valore medio come riferimento indicativo, la generazione di un dataset di 1024 configurazioni richiede un tempo complessivo di circa 55 ore di calcolo. Tale stima deve essere interpretata come approssimata, poiché la campagna è stata in pratica suddivisa in più blocchi consecutivi da 256, 256 e 512 simulazioni. Questa scelta non è stata casuale: la costruzione incrementale del dataset è stata adottata per valutare in modo progressivo l'effetto della numerosità dei dati sull'accuratezza del modello surrogato e, soprattutto, sulla sua efficacia all'interno della procedura di ottimizzazione. A tal fine, per i dataset da 256 e 512 configurazioni sono state testate diverse architetture di rete neurale, così da verificarne il comportamento al variare della complessità del modello. Il confronto complessivo tra errori predittivi e fronti di Pareto ottenuti ha mostrato che il dataset da 1024 simulazioni costituiva il miglior compromesso tra qualità dei risultati e costo computazionale, risultando più convincente dei casi più piccoli e, al tempo stesso, sufficientemente accurato da non giustificare un ulteriore ampliamento della campagna CFD.

5.3 Modello surrogato basato su rete neurale artificiale feed-forward

Come anticipato nel Capitolo 4, il principale limite dell'approccio basato su simulazioni CFD direttamente inserite nel loop di ottimizzazione risiede nell'elevato costo computazionale richiesto per valutare un numero sufficientemente ampio di configurazioni. Per superare tale criticità, si introduce in questa sezione un modello surrogato basato su rete neurale artificiale, con l'obiettivo di approssimare la relazione tra configurazione geometrica del dissipatore e corrispondente

risposta termica. In particolare, il modello ha il compito di predire la temperatura media ovvero calcolare $\hat{T}_{av}$ a partire dal vettore decisionale $x = [n_1, n_2, \dots, n_9]$. Tale impostazione consente di sostituire, nelle successive fasi iterative di esplorazione del dominio di progetto, le ripetute valutazioni CFD con una previsione numerica estremamente più rapida.

Nel presente lavoro, il modello neurale è stato implementato in MATLAB mediante la funzione “*feedforwardnet*” [28], disponibile all’interno dell’Add-On Deep Learning Toolbox, che consente di costruire reti neurali *feed-forward* con uno o più *layer* nascosti e di addestrarle su problemi generali di mappatura input–output. Una rete *feed-forward* è una rete neurale artificiale in cui l’informazione si propaga in un’unica direzione, dall’input verso l’output, senza retroazioni tra i *layer*. La struttura della rete è organizzata in livelli successivi di nodi di calcolo, detti neuroni: il primo livello riceve le variabili di input, i livelli intermedi, detti *layer* nascosti, elaborano progressivamente l’informazione mediante combinazioni pesate e funzioni di attivazione non lineari, mentre il livello finale produce la stima della grandezza di interesse. I neuroni dei *layer* nascosti non corrispondono a variabili fisiche direttamente osservabili, ma costituiscono rappresentazioni interne apprese durante l’addestramento, utili a catturare relazioni anche non lineari tra input e output.

La struttura concettuale del modello neurale è schematizzata in Figura 48. In essa sono evidenziati i nove ingressi associati alle variabili di progetto (*feature*), i *layer* nascosti impiegati dalla rete e il nodo di uscita corrispondente alla previsione della temperatura media sulla superficie di interesse (*target*).

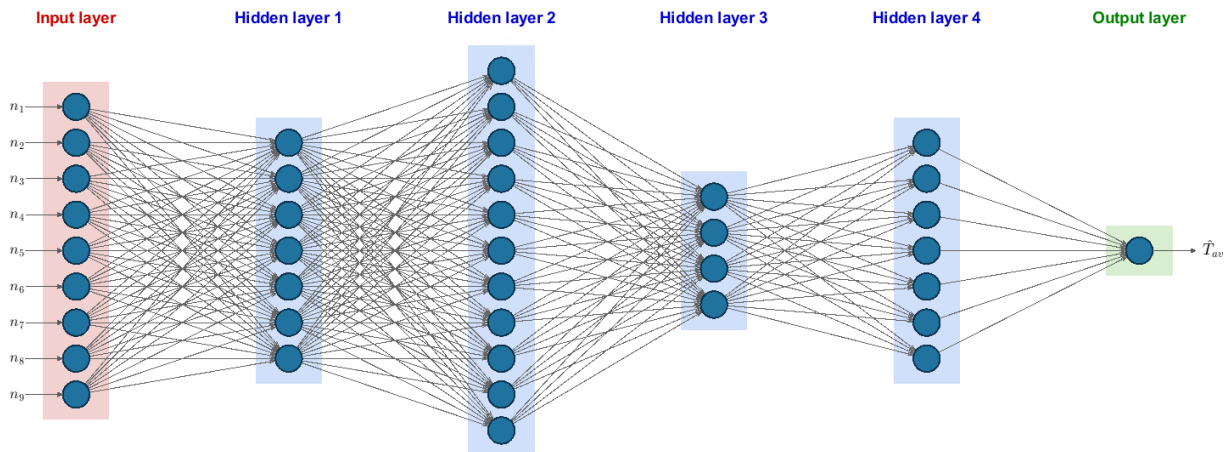


Figura 48: Schema concettuale della rete neurale *feed-forward* utilizzata come modello surrogato nel presente lavoro

Le reti neurali *feed-forward* sono ampiamente impiegate come modelli surrogati in problemi di regressione e approssimazione funzionale, in particolare quando sia necessario apprendere relazioni input–output non lineari a partire da dati numerici e disporre poi di predizioni rapide a basso costo computazionale [29]. Tale scelta risulta particolarmente adatta al problema in esame, poiché la temperatura media del dissipatore è il risultato di fenomeni termofluidodinamici accoppiati, comprendenti conduzione nel solido e convezione nel liquido, la cui intensità dipende in modo non lineare dalla geometria. Non si intende qui proporre una trattazione generale ed esaustiva delle reti neurali, ma descrivere gli elementi del modello adottato che risultano rilevanti ai fini del presente studio.

I risultati e i grafici riportati nei paragrafi seguenti si riferiscono al modello addestrato sul dataset completo di 1024 configurazioni.

5.3.1 Analisi preliminari di correlazione

Prima dell'addestramento della rete è stata svolta un'analisi esplorativa delle relazioni tra le variabili di input e il *target* termico, mediante il calcolo dei coefficienti di correlazione di Pearson e di Spearman, riportati in Figura 49. Tale analisi permette di cogliere, rispettivamente, l'intensità delle associazioni lineari e delle dipendenze monotone tra ciascun ingresso e la temperatura media, fornendo così una prima indicazione quantitativa del ruolo esercitato dalle diverse regioni del dissipatore sulla risposta termica.

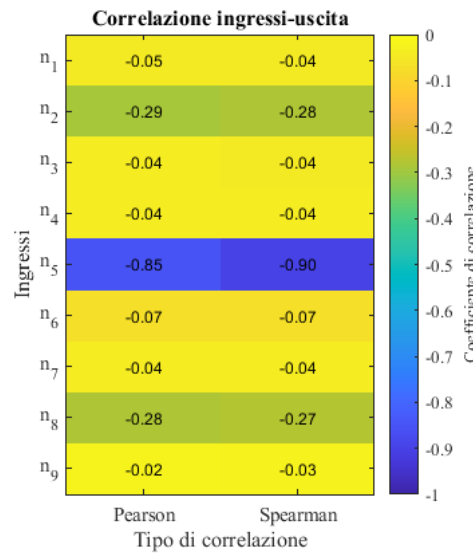


Figura 49: Correlazione input-output tra le nove variabili di riempimento regionale e il target termico

Poiché un aumento del numero di pin in una determinata zona tende, in generale, a intensificare lo scambio termico e quindi a ridurre la temperatura media del dissipatore, è naturale che la correlazione tra le variabili di riempimento delle zone e l'output termico sia negativa. Come atteso sulla base della geometria del problema (Figura 37 paragrafo 5.1.1), la variabile n_5 mostra l'associazione negativa più marcata con l'output (Pearson ≈ -0.85 , Spearman ≈ -0.90); correlazioni di entità moderata si osservano anche per n_2 e n_8 (intorno a $-0.28/-0.29$), mentre le restanti variabili presentano correlazioni deboli, prossime a zero. Il valore leggermente più elevato in modulo del coefficiente di Spearman per n_5 suggerisce che il suo effetto sia prevalentemente monotono, con potenziali non linearità. Sebbene queste tendenze, pienamente coerenti con l'intuizione fisica del problema, evidenzino che solo alcune variabili presentano un legame marcato con l'output, tutte e nove vengono mantenute come ingressi del modello surrogato. Infatti, l'analisi di correlazione considera ciascuna variabile singolarmente rispetto all'output e fornisce quindi una misura soltanto bivariata del legame ingresso-uscita. Un coefficiente di correlazione debole non implica necessariamente che una variabile sia irrilevante: il suo contributo potrebbe emergere in combinazione con altre variabili, oppure attraverso effetti non lineari che una misura *pairwise* non è in grado di rappresentare pienamente.

Accanto all'analisi del legame tra ingressi e *target* termico, è stata inoltre verificata la struttura di correlazione tra le stesse variabili di input, al fine di accertare che il dataset fornisse alla rete neurale ingressi statisticamente ben distribuiti e non caratterizzati da dipendenze marcate reciproche. Questo aspetto è rilevante in fase di addestramento, poiché la presenza di forti correlazioni tra le *feature* può introdurre ridondanza informativa e rendere meno chiaro il contributo effettivo delle singole variabili nel processo di apprendimento. Tale verifica risulta particolarmente opportuna nel presente lavoro, poiché le configurazioni sono state inizialmente generate mediante sequenza di Sobol nello spazio continuo normalizzato $[0,1]^d$, ma successivamente trasformate in conteggi interi

di pin per ciascuna zona. Sebbene il campionamento Sobol sia progettato per distribuire i punti in modo regolare e *space-filling* nel dominio di progetto, la successiva discretizzazione potrebbe, almeno in linea di principio, alterare tale proprietà e introdurre correlazioni spurie tra le variabili di input. A questo scopo sono quindi state calcolate le matrici di correlazione di Pearson e di Spearman riportate in Figura 50.

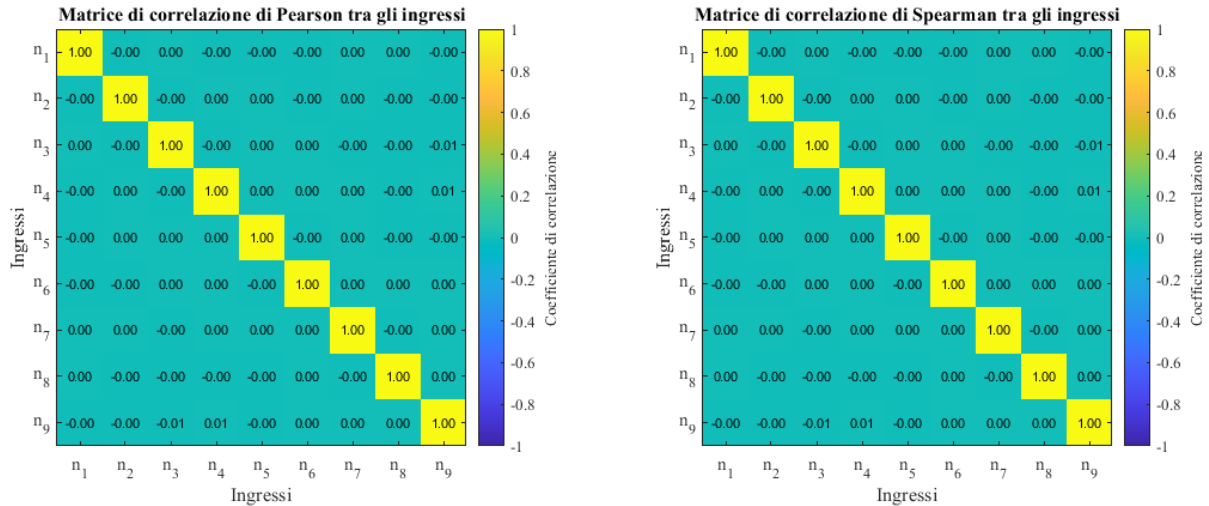


Figura 50: Matrici di correlazione tra le nove variabili di input

In entrambe le matrici, i coefficienti fuori diagonale risultano molto prossimi a zero, mentre sulla diagonale principale compare, come atteso, correlazione unitaria, in quanto ogni variabile è perfettamente correlata con sé stessa. Inoltre, essendo la matrice di correlazione simmetrica rispetto alla diagonale principale, l'osservazione di una sola delle due metà triangolari è sufficiente, poiché l'altra contiene le medesime informazioni replicate. Proprio il fatto che i termini fuori diagonale sono tutti molto piccoli mostra che, nel dataset ottenuto, non emergono né relazioni lineari significative tra le variabili di input, come evidenziato dai coefficienti di Pearson, né dipendenze monotone marcate, come suggerito dai coefficienti di Spearman. In altri termini, il numero di pin assegnato a una determinata zona non tende sistematicamente a crescere o diminuire al variare del numero di pin nelle altre zone.

Nel complesso, questa evidenza supporta la bontà della procedura adottata: la sequenza di Sobol, anche dopo la trasformazione in variabili intere, mantiene una struttura degli ingressi priva di correlazioni forti. Va comunque precisato che questa analisi non consente di dimostrare in senso rigoroso l'assenza di qualunque forma di dipendenza tra le variabili di input. Non rappresenta quindi una prova definitiva, ma costituisce comunque una buona indicazione del fatto che gli input non siano correlati in modo significativo.

5.3.2 Suddivisione del dataset e pre-processing dei dati

Il dataset è stato suddiviso in modo riproducibile in tre sottoinsiemi: *training*, *validation* e *test*, secondo una ripartizione pari rispettivamente al 60%, 20% e 20% dei campioni. Tale scelta rappresenta un compromesso ragionevole, in quanto consente di destinare una quota sufficientemente ampia dei dati all'addestramento della rete, preservando al tempo stesso due insiemi distinti per il monitoraggio della generalizzazione e per la successiva valutazione finale del modello. I tre set vengono infatti utilizzati nel modo seguente: Il *training* set viene utilizzato per l'aggiornamento dei pesi, il *validation* set per monitorare la capacità di generalizzazione durante l'addestramento e arrestare il *training* prima di incorrere in *overfitting*, ossia nella condizione in cui

il modello si adatta eccessivamente ai dati di *training*, perdendo efficacia predittiva su dati non visti; il *test* set viene lasciato completamente non visto fino alla valutazione finale del modello. Il ruolo operativo del *validation* set all'interno della procedura di *training* sarà chiarito più nel dettaglio nel paragrafo successivo. Per rendere la procedura replicabile, la suddivisione è stata effettuata fissando il generatore pseudo-casuale di MATLAB mediante il comando `rng(42, 'twister')`. In questo contesto, il valore 42 non possiede alcun significato specifico, ma rappresenta semplicemente un Seed scelto in modo arbitrario, il cui scopo è rendere ripetibile la partizione del dataset. Il termine *twister* indica invece l'algoritmo pseudo-casuale utilizzato da MATLAB, ossia il Mersenne Twister, ampiamente impiegato per applicazioni numeriche generali grazie alla sua buona qualità statistica e alla capacità di generare sequenze riproducibili.

Prima del *training*, input e output sono stati standardizzati mediante normalizzazione z-score:

$$z = \frac{x - \mu}{\sigma} \quad (27)$$

dove μ e σ rappresentano, rispettivamente, la media e la deviazione standard della variabile considerata x .

In particolare, media e deviazione standard sono state calcolate solo sul *training* set, e poi applicate anche ai set di validazione e *test*. Questa scelta è importante perché evita fenomeni di *data leakage*: se le statistiche di normalizzazione venissero stimate sull'intero dataset, il modello riceverebbe indirettamente informazioni anche sui dati che dovrebbero invece restare non visti. Dopo l'addestramento, le previsioni della rete sono state ricondotte alle unità fisiche originali mediante la trasformazione inversa della normalizzazione.

5.3.3 Architettura della rete neurale

Nel seguito si entra più nel dettaglio dell'architettura effettivamente adottata e delle principali scelte implementative utilizzate in MATLAB per l'addestramento del modello surrogato, seguendo sostanzialmente la sequenza logica con cui tali elementi compaiono nel codice sviluppato: Codice 9.

```
% Definizione della rete neurale feed-forward:
net = feedforwardnet([7, 11, 4, 7], 'trainscg');
net.layers{1}.transferFcn = 'tansig';
net.layers{2}.transferFcn = 'tansig';
net.layers{3}.transferFcn = 'tansig';
net.layers{4}.transferFcn = 'tansig';
net.layers{5}.transferFcn = 'purelin';

% Obiettivo & regolarizzazione
net.performFcn = 'mse';
net.performParam.regularization = 0.01;
net.trainParam.goal = 1e-5;
net.trainParam.min_grad = 1e-6;
net.trainParam.max_fail = 100;
net.trainParam.epochs = 1e4;
```

Codice 9: MATLAB, definizione dell'architettura della rete neurale e dei principali parametri di addestramento

- Come già richiamato nell'introduzione della Sezione 5.3, la rete, implementata con `feedforwardnet`, appartiene alla famiglia delle reti *feed-forward* ed è, più precisamente, un *multilayer perceptron* completamente connesso. Con tale espressione si intende una rete organizzata in *layer* successivi di neuroni artificiali, nella quale ciascun neurone di un dato livello riceve in ingresso l'uscita di tutti i neuroni del livello precedente, come rappresentato

schematicamente in Figura 48 (Sezione 5.3). Ogni neurone realizza una trasformazione del tipo:

$$a_j = \phi \left(\sum_{i=1}^n w_{ji} x_i + b_j \right) \quad (28)$$

dove: n è il numero di ingressi del neurone, x_i sono gli ingressi del neurone, w_{ji} sono i pesi associati agli ingressi, b_j è il *bias* del neurone, $\phi(\cdot)$ è la funzione di attivazione e a_j è l'uscita del neurone.

L'Eq. 28 è una combinazione lineare degli ingressi, corretta da un termine di *bias*, seguita dall'applicazione di una funzione di attivazione non lineare. L'intera rete può quindi essere interpretata come una composizione di trasformazioni lineari e non lineari, capace di approssimare relazioni input-output anche complesse.

- L'architettura adottata comprende quattro *layer* nascosti, con numero di neuroni pari a [7, 11, 4, 7], anch'essa rappresentata schematicamente in Figura 48 (Sezione 5.3). Tale configurazione è stata individuata empiricamente, sulla base delle prove svolte, come compromesso tra capacità espressiva e complessità del modello: una rete troppo piccola rischierebbe infatti di non rappresentare adeguatamente le non linearità del problema, mentre una rete eccessivamente ampia introdurrebbe un numero elevato di parametri liberi, aumentando il rischio di *overfitting*. L'andamento non monotono del numero di neuroni nei *layer* nascosti riflette una successione di fasi di espansione e compressione della rappresentazione interna, utile a rielaborare in modo non lineare le variabili di ingresso prima della previsione finale.
- L'addestramento è stato eseguito mediante l'algoritmo "*trainscg*", acronimo di *scaled conjugate gradient*. Si tratta di un algoritmo di ottimizzazione numerica basato sul gradiente, utilizzato per minimizzare iterativamente la funzione costo della rete aggiornando pesi e bias. Rispetto alla discesa del gradiente standard, che si muove localmente nella direzione di massima diminuzione dell'errore, il metodo del gradiente coniugato cerca invece di costruire, a ogni iterazione, direzioni di ricerca più efficaci, tenendo conto anche dell'andamento seguito nei passi precedenti. In questo modo, la discesa verso il minimo della funzione costo risulta in genere più rapida e regolare.
- Nei quattro *layer* nascosti è stata adottata la funzione di attivazione iperbolica tangente sigmoide, indicata in MATLAB con "*tansig*". Si tratta di una scelta classica nelle reti *feed-forward* implementate nell'ambiente MATLAB. Essa produce, per ciascun neurone, un'uscita limitata nell'intervallo $(-1, 1)$, introducendo la non linearità necessaria affinché la rete non si riduca a una semplice combinazione lineare degli ingressi. In termini pratici, l'impiego di una funzione non lineare nei *layer* nascosti consente al modello di catturare relazioni complesse tra configurazione geometrica e risposta termica, che non sarebbero rappresentabili mediante sole trasformazioni lineari. La scelta di *tansig* risulta inoltre coerente con la preventiva standardizzazione z-score degli input e dell'output, poiché i dati normalizzati tendono a distribuirsi attorno a zero e risultano quindi ben compatibili con una funzione di attivazione centrata e limitata.
- Nel *layer* di uscita è stata invece impiegata una funzione lineare, indicata in MATLAB con "*purelin*". Questa scelta è tipica nei problemi di regressione, poiché consente alla rete di produrre un valore continuo non vincolato a un intervallo predefinito. Nel presente caso, l'uscita del modello deve rappresentare una stima della temperatura media, cioè una grandezza scalare continua; l'utilizzo di una funzione lineare nel *layer* finale evita quindi di introdurre saturazioni artificiali o limitazioni non fisicamente motivate sul valore predetto.

- Come funzione costo è stato adottato il *mean squared error* (MSE), cioè l'errore quadratico medio tra output previsto e valore *target*:

$$\text{MSE} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (29)$$

dove N è il numero di campioni considerati, y_i è il valore *target* normalizzato dell'output, $\hat{y}_i$ è la predizione normalizzata della rete.

Si tratta della scelta più comune nei problemi di regressione, poiché fornisce una misura sintetica dello scostamento complessivo tra predizioni e valori reali, penalizzando maggiormente gli errori più elevati. Nel presente lavoro, l'MSE è stato utilizzato con un termine di regolarizzazione pari a 0.01, introdotto per limitare la crescita eccessiva dei pesi e ridurre così il rischio di *overfitting*. Nel pratico, il termine di regolarizzazione entra direttamente nella funzione costo minimizzata durante l'addestramento. Ciò significa che l'algoritmo non ottimizza soltanto l'errore di previsione sui dati, ma anche una penalizzazione associata all'ampiezza di pesi e *bias*, favorendo così soluzioni meno estreme e con migliore capacità di generalizzazione. L'addestramento è stato eseguito fissando un massimo di 10^4 epoche, dove per epoca si intende una passata completa su tutti i campioni del training set, al termine della quale la rete aggiorna i propri parametri sulla base dell'errore osservato. Sono stati inoltre impostati i parametri $\text{goal} = 10^{-5}$ e $\text{min_grad} = 10^{-6}$, che consentono di interrompere anticipatamente il *training* qualora l'errore raggiunga un valore sufficientemente basso oppure il gradiente diventi trascurabile. Un ruolo particolarmente importante è svolto dal parametro $\text{max_fail} = 100$, che implementa il criterio di *early stopping* basato sul *validation* set, esplicitandone operativamente il ruolo già anticipato nel paragrafo precedente. Il set di validazione, infatti, non viene utilizzato per aggiornare i pesi della rete, ma serve proprio a monitorarne la capacità di generalizzazione durante l'addestramento. In particolare, il *training* viene arrestato se la prestazione sui dati di validazione, valutata a ogni epoca, non mostra miglioramenti per 100 controlli consecutivi. La scelta di un valore relativamente elevato è stata adottata per evitare un'interruzione troppo precoce dovuta a oscillazioni locali o a piccoli peggioramenti temporanei della prestazione sul *validation* set, lasciando alla rete un margine sufficiente per proseguire l'ottimizzazione qualora fossero ancora possibili miglioramenti successivi. Al tempo stesso, questo criterio consente di interrompere l'addestramento quando il modello smette di migliorare in termini di capacità di generalizzazione, limitando così il rischio di *overfitting*. Nel caso in esame, il criterio di validazione è risultato anche il motivo effettivo dell'arresto del *training*.

Il comportamento effettivo della procedura di addestramento è illustrato in Figura 51, che riporta l'andamento del MSE sui set di *training*, *validation* e *test* al variare delle epoche.

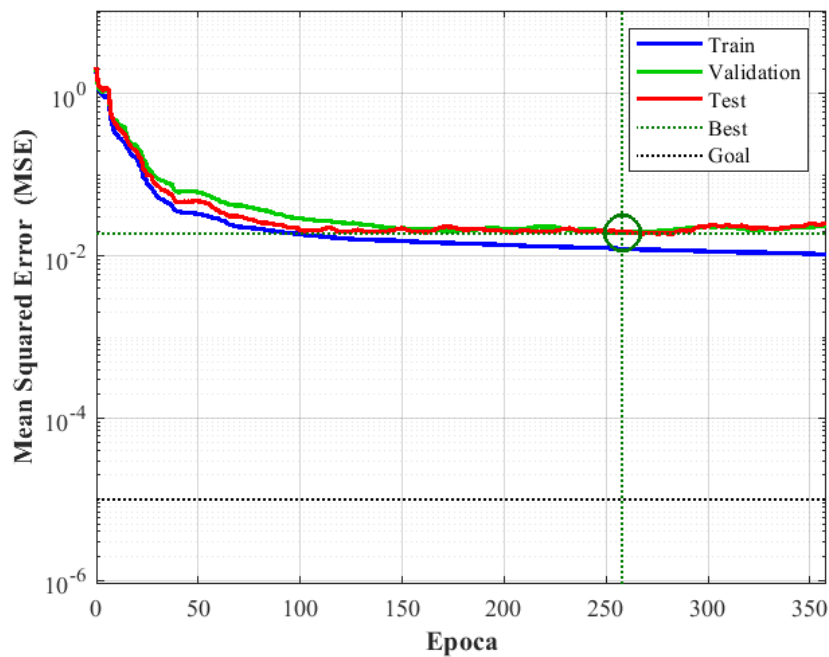


Figura 51: Andamento del MSE sui set di training, validation e test durante l'addestramento della rete neurale

Si osserva che la miglior prestazione sul set di validazione si ottiene all'epoca 258, con un valore di MSE pari a 0,019, mentre l'arresto del *training* avviene all'epoca 358. La rete assunta come modello finale è quella corrispondente alla miglior epoca, cioè all'epoca 258, associata alla migliore capacità di generalizzazione stimata sul *validation* set. Dal grafico si osserva inoltre che, in corrispondenza della regione di minimo, le curve di *validation* e *test* assumono valori molto vicini tra loro. Questo comportamento costituisce un'indicazione favorevole della bontà della scelta effettuata, poiché suggerisce che la prestazione della rete sui dati di validazione sia rappresentativa anche del comportamento su dati non visti. Va inoltre precisato che i valori di MSE riportati in Figura 51 sono riferiti ai dati normalizzati, poiché il *training* della rete viene eseguito sugli input e sugli output previamente standardizzati mediante normalizzazione z-score. Di conseguenza, tali valori non sono direttamente espressi nelle unità fisiche originali della temperatura, ma risultano comunque pienamente significativi ai fini del monitoraggio della convergenza e del confronto tra le prestazioni relative dei tre sottoinsiemi durante l'addestramento.

Per una valutazione più immediata dell'accuratezza del modello nella scala fisica del problema, sono stati inoltre calcolati gli errori sui dati denormalizzati. In termini di MSE, essi risultano pari a 2,54 sul *training* set, 5,04 sul *validation* set e 5,45 sul *test* set, corrispondenti a valori di RMSE (Root Mean Squared Error) rispettivamente pari a circa 1,59 °C, 2,25 °C e 2,34 °C. Il fatto che l'errore sul *training* set risulti più basso rispetto a quelli di *validation* e *test* è del tutto atteso, poiché la rete è ottimizzata direttamente sui dati di addestramento. Risulta tuttavia importante osservare che la differenza rispetto ai valori ottenuti su *validation* e *test* rimane contenuta. Anche questo rappresenta un segnale positivo, poiché suggerisce che il modello mantenga una buona capacità di generalizzazione senza mostrare un degrado marcato delle prestazioni sui dati non visti.

Infine, l'andamento delle curve nelle epoche successive alla migliore mostra che, mentre l'errore sul *training* set continua a diminuire, le curve di *validation* e *test* tendono invece a peggiorare. Questo comportamento è coerente con la dinamica tipica dell'*overfitting*. In questo senso, la scelta di arrestare il *training* sulla base del *validation* set risulta appropriata, poiché consente di individuare un punto di arresto plausibile prima che il sovradattamento diventi più marcato.

Il tempo complessivo di addestramento è stato di circa 2 secondi.

5.3.4 Valutazione del modello surrogato

Una volta completato l'addestramento, il modello surrogato è stato valutato sul test set, costituito da configurazioni mai viste dalla rete durante le fasi di aggiornamento dei pesi e di selezione del punto di arresto. Tale verifica è essenziale, poiché consente di stimare la capacità del modello di generalizzare a nuove configurazioni geometriche, ossia proprio la proprietà richiesta per il suo successivo impiego come sostituto della CFD nelle fasi iterative di esplorazione del dominio di progetto.

Il primo riscontro è fornito dal grafico di parità riportato in Figura 52, nel quale le predizioni della rete sul test set vengono confrontate con i corrispondenti valori ottenuti dalle simulazioni CFD.

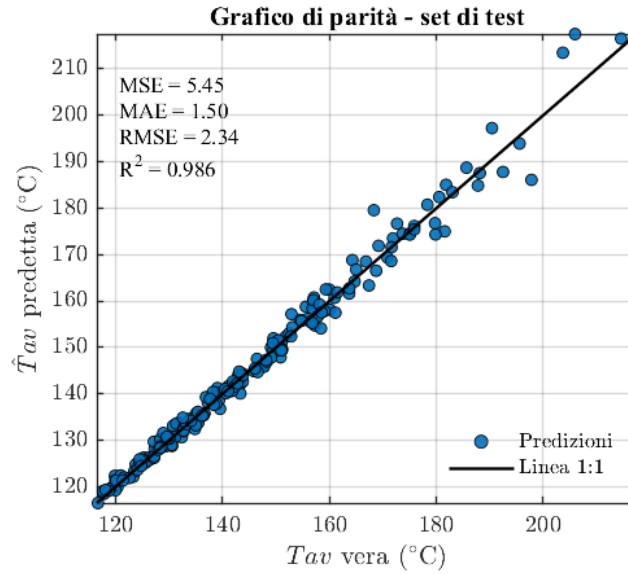


Figura 52: Grafico di parità sul set di test tra i valori di temperatura media predetti dal modello surrogato e quelli ottenuti dalle simulazioni CFD

I punti risultano distribuiti in prossimità della retta ideale 1:1 lungo l'intero intervallo dei valori di temperatura considerati, senza che emergano scostamenti sistematici evidenti. Questo comportamento suggerisce che il modello sia in grado di riprodurre con buona fedeltà la relazione input-output appresa dal dataset, mantenendo prestazioni soddisfacenti anche su campioni non utilizzati nel *training*. Per una valutazione quantitativa delle prestazioni del modello sono state considerate le metriche MSE, MAE (*Mean Absolute Error*), RMSE e R^2 , ciascuna utile a descrivere un aspetto diverso dell'accuratezza predittiva. In particolare, il valore di MSE viene riportato poiché rappresenta la funzione obiettivo minimizzata durante l'addestramento della rete, e costituisce quindi la metrica più direttamente collegata al processo di *training*. Tuttavia, essendo espresso nel quadrato dell'unità di misura della variabile di uscita, esso risulta meno immediato da interpretare dal punto di vista fisico. Per questo motivo risultano particolarmente significativi il MAE e l'RMSE in quanto entrambi espressi nella stessa unità di misura della grandezza predetta. In particolare, il MAE rappresenta lo scostamento assoluto medio tra il valore predetto dalla rete e quello ottenuto dalla simulazione CFD:

$$MAE = \frac{1}{N} \sum_{i=1}^N |T_{av,i} - \hat{T}_{av,i}| \quad (30)$$

Nel caso in esame, il valore MAE=1.50 °C indica che, mediamente, la rete neurale commette un errore di circa 1.5 °C nella stima di T_{av} . L'RMSE, pari a 2.34 °C, fornisce invece una misura dell'errore che attribuisce maggior peso agli scostamenti più elevati, risultando quindi utile per cogliere l'eventuale presenza di predizioni peggiori, pur rimanendo anch'esso su valori contenuti. Infine, il

coefficiente di determinazione $R^2 = 0.986$ conferma l'elevato grado di accordo tra i valori predetti dal modello e quelli ottenuti tramite simulazione CFD.

Per integrare questa valutazione con una misura più direttamente interpretabile in termini ingegneristici, è stato inoltre analizzato l'errore relativo (in valore assoluto) sul test set, definito come:

$$\text{Errore}_i (\%) = 100 \cdot \frac{|T_{av,i} - \hat{T}_{av,i}|}{T_{av,i}} \quad (31)$$

La distribuzione di tale errore è riportata in Figura 53.

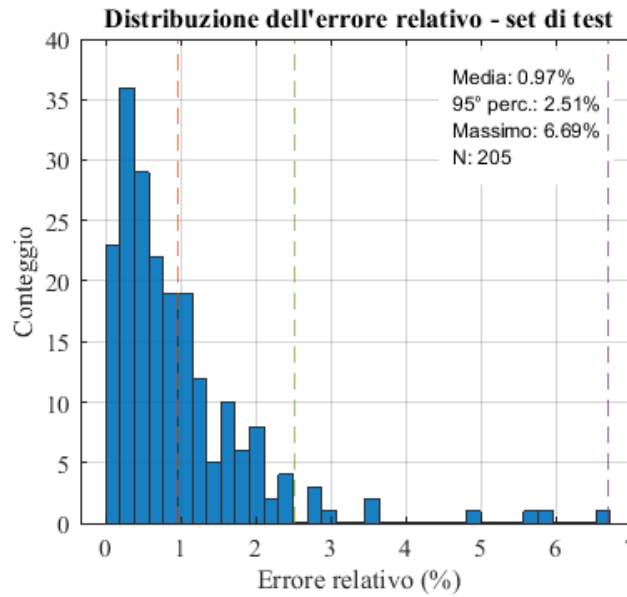


Figura 53: Distribuzione dell'errore relativo sul set di test del modello surrogato

Si osserva che la maggior parte dei campioni presenta errori contenuti e concentrati su valori bassi, mentre solo un numero limitato di configurazioni si colloca nella coda a errore più elevato. In particolare, l'errore relativo medio sul test set risulta pari a 0.97%, il 95° percentile è pari a 2.51%, mentre il valore massimo osservato è pari a 6.69%, su un totale di 205 campioni di test. Nel complesso, tali risultati indicano che il modello surrogato fornisce una stima di T_{av} sufficientemente accurata per essere impiegata come strumento di *screening* rapido delle configurazioni e quindi guidare l'esplorazione multi-obiettivo del dominio di progetto.

In termini indicativi, si può inoltre osservare che l'ordine di grandezza dell'errore del surrogato rispetto alla CFD (medio 0.97%, 95° percentile 2.51%) è comparabile a quello dello scostamento già riscontrato tra diversi livelli di Mesh nell'analisi di sensitività numerica (livello 3 vs livello 6, Paragrafo 5.1.3, compreso tra 0.58% e 2.78%). Tale confronto, pur da interpretare con cautela, aiuta a collocare l'accuratezza del modello all'interno del margine di variabilità già presente nella catena simulativa. A ciò si aggiunge il vantaggio in termini di costo computazionale: il tempo medio di valutazione di una singola configurazione è stato stimato a partire dal tempo complessivo richiesto per la valutazione di 10^6 configurazioni mediante il modello surrogato. Il tempo totale così ottenuto è pari a circa 0,7 s, da cui risulta un tempo medio di circa $0,7 \times 10^{-6}$ s per configurazione. Tale stima ha il solo scopo di fornire un ordine di grandezza dei tempi di valutazione e non intende costituire un'analisi prestazionale esaustiva; nondimeno, essa evidenzia con chiarezza come l'impiego del modello surrogato renda praticabile un'esplorazione estensiva dello spazio di progetto, che risulterebbe del tutto proibitiva mediante sole simulazioni CFD.

Va comunque precisato che il modello non sostituisce in senso assoluto la simulazione ad alta fedeltà. La presenza di alcuni casi con errore più elevato suggerisce infatti che la rete, pur generalizzando bene nel complesso, possa mostrare una minore accuratezza in particolari regioni dello spazio di progetto o in corrispondenza di configurazioni meno rappresentate dal dataset. Per questa ragione, come verrà mostrato nella sezione seguente, il ruolo del surrogato non è quello di fornire una validazione finale delle soluzioni, bensì di rendere computazionalmente sostenibile una fase di ricerca e selezione preliminare su larga scala, riservando alla CFD la successiva verifica delle configurazioni ottenute.

5.4 Ottimizzazione multi-obiettivo assistita da surrogato basata sull'algoritmo genetico NSGA-II

Nel presente lavoro il problema viene formulato in termini di ottimizzazione multi-obiettivo, considerando come variabili decisionali il numero di pin assegnati a ciascuna delle 9 zone della superficie $[n_1, n_2, \dots, n_9]$, e come obiettivo due funzioni di tali variabili tra loro concorrenti:

- la minimizzazione della temperatura media sulla superficie di interesse T_{av} .
- la minimizzazione del numero complessivo di pin $n_{pin,tot} = \sum_{j=1}^9 n_j$.

Il primo obiettivo mira a migliorare le prestazioni termiche del dissipatore, mentre il secondo consente di limitare il numero di cilindri impiegati, con possibili benefici in termini di peso, quantità di materiale e semplicità realizzativa. In un contesto di questo tipo non esiste un'unica soluzione ottima in senso assoluto, poiché una riduzione di T_{av} richiede in genere un aumento del numero di pin, mentre configurazioni più leggere o meno dense tendono a penalizzare le prestazioni termiche. Risulta quindi più appropriato ricercare un insieme di configurazioni di compromesso, ciascuna caratterizzata dal fatto che il miglioramento di un obiettivo comporta inevitabilmente il peggioramento dell'altro. In ambito multi-obiettivo, tali configurazioni prendono il nome di soluzioni non dominate. Una soluzione si dice non dominata se non esiste un'altra configurazione capace di migliorare almeno uno degli obiettivi senza peggiorare l'altro. L'insieme delle soluzioni non dominate costituisce il fronte di Pareto, che rappresenta quindi il luogo dei compromessi efficienti tra prestazioni termiche e numero di pin.

Per l'esplorazione di tale fronte è stato adottato un algoritmo genetico multi-obiettivo. Questa scelta è giustificata sia dalla natura discreta delle variabili decisionali, sia dalla struttura non lineare e irregolare del problema, per la quale metodi basati su gradienti risultano poco adatti. Inoltre, anche se il modello surrogato consente valutazioni estremamente rapide, lo spazio combinatorio rimane molto ampio: assumendo per ciascuna delle 9 variabili i 15 valori interi compresi tra 0 e 14, il numero teorico di configurazioni possibili è pari a $15^9 \approx 3.84 \times 10^{10}$; sebbene, in linea puramente teorica, la sola valutazione del surrogato su tutte le configurazioni richiederebbe un tempo dell'ordine di circa 7 ore (usando il tempo stimato nel paragrafo precedente), un'esplorazione esaustiva risulterebbe comunque poco praticabile, poiché imporrebbe la gestione, l'ordinamento e il filtraggio di una mole enorme di dati per l'individuazione delle soluzioni migliori e del corrispondente fronte di Pareto. In questo senso, l'impiego di un algoritmo genetico consente di concentrare la ricerca su regioni promettenti dello spazio di progetto, evitando una scansione completa dell'intero insieme combinatorio.

Nel presente lavoro è stata utilizzata l'implementazione MATLAB “*gamultiobj*”^[30], disponibile all'interno dell'Add-On Global Optimization Toolbox. Gamultiobj adotta un algoritmo genetico elitista controllato, descritto dalla documentazione ufficiale come una variante di NSGA-II^[31] (*Nondominated Sorting Genetic Algorithm*). La logica di funzionamento, pur senza voler entrare in

una trattazione esaustiva, può essere compresa con riferimento alle Figure 54, 55 e 56. In Figura 54 si può osservare lo schema generale di NSGA-II.

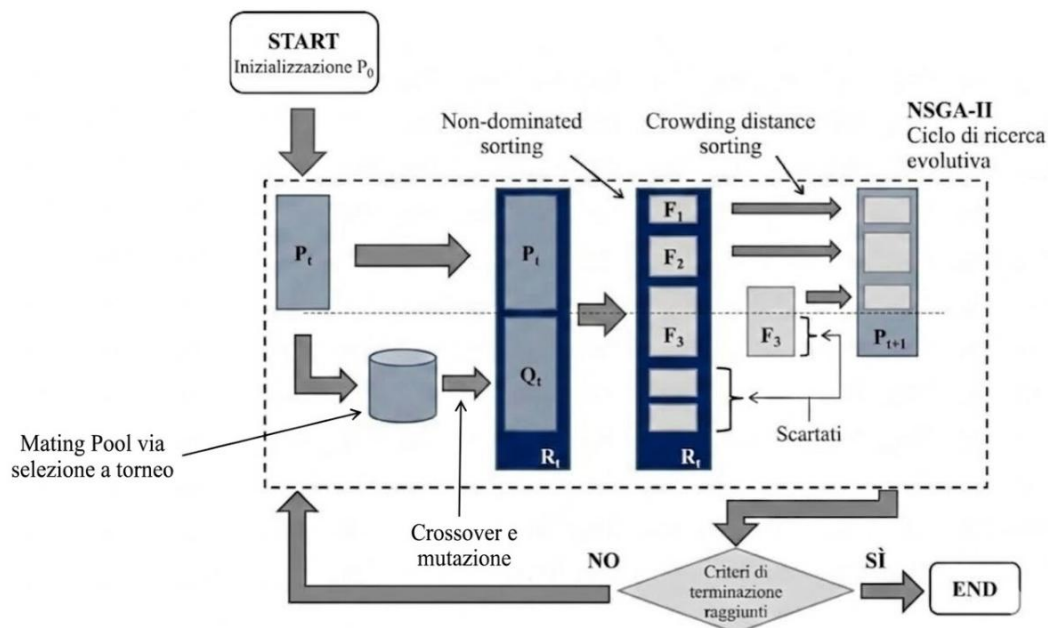


Figura 54 ^[32]: Struttura generale di NSGA-II, schema riadattato

La popolazione iniziale viene generata casualmente all'interno dei limiti ammissibili delle variabili decisionali. A ogni generazione, l'algoritmo mantiene una popolazione di soluzioni candidate P_t per la quale i valori delle due funzioni obiettivo sono già noti dalla generazione precedente, ad eccezione della popolazione iniziale. A partire dalla popolazione corrente P_t , vengono generate nuove soluzioni Q_t mediante i classici operatori genetici di selezione, crossover e mutazione (Figura 55).

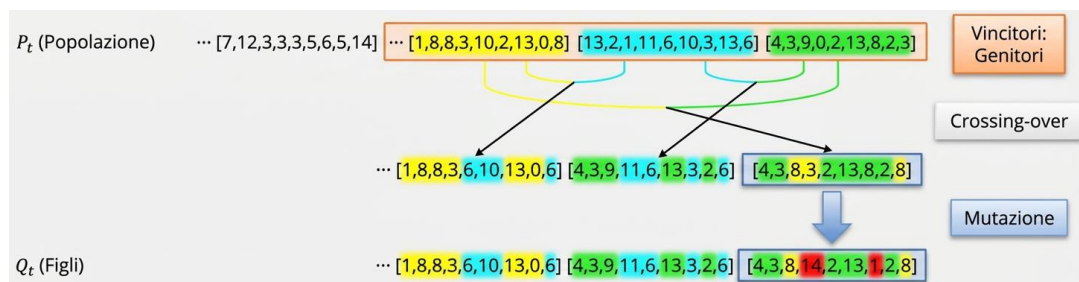


Figura 55: Schema illustrativo semplificato degli operatori genetici di selezione, crossover e mutazione. I vettori riportati non rappresentano configurazioni significative, ma sono utilizzati a scopo puramente illustrativo per evidenziare il funzionamento degli operatori genetici nel caso in esame

In particolare, la selezione individua gli individui da riprodurre mediante una selezione a torneo, nella quale piccoli gruppi casuali di candidati della popolazione corrente vengono messi a confronto e, in ciascun torneo, viene scelto il più promettente. Tale giudizio si basa sugli stessi criteri che verranno esplicitati nel seguito per la costruzione della nuova popolazione P_{t+1} , ossia il rango di dominanza (*non-dominated sorting*) e, a parità di rango, la diversità lungo il fronte (*crowding distance sorting*). I due momenti dell'algoritmo, tuttavia, hanno finalità differenti. Nel primo caso, tali criteri servono a definire il *mating pool*, cioè l'insieme dei genitori destinati alla riproduzione; nel secondo, vengono impiegati per selezionare gli individui che sopravvivono e compongono la popolazione della generazione successiva P_{t+1} . Il crossover combina parti di due genitori per

generare nuovi candidati. La mutazione introduce piccole variazioni casuali, utili per mantenere capacità esplorativa ed evitare una convergenza prematura verso regioni troppo limitate dello spazio di progetto.

L'insieme della popolazione P_t e dei figli Q_t viene quindi unito nella popolazione estesa R_t . Su tale insieme, dopo aver calcolato le funzioni obiettivo per Q_t , viene applicato il *non-dominated sorting*, cioè l'ordinamento in fronti successivi $F_1, F_2, F_3, \dots$, dove il primo fronte F_1 contiene le soluzioni non dominate (rango 1), il secondo contiene quelle dominate solo dagli elementi di F_1 (rango 2), e così via. In questo modo a ogni individuo viene assegnato un rango di dominanza, che misura la sua qualità rispetto al criterio di Pareto: rango più basso significa soluzione preferibile. La nuova popolazione della generazione successiva P_{t+1} viene quindi costruita riempiendola progressivamente con i fronti migliori, fino a raggiungere la dimensione prefissata. Quando l'ultimo fronte da inserire non entra completamente nella nuova popolazione, interviene il criterio della *crowding distance*, rappresentato schematicamente in Figura 56.

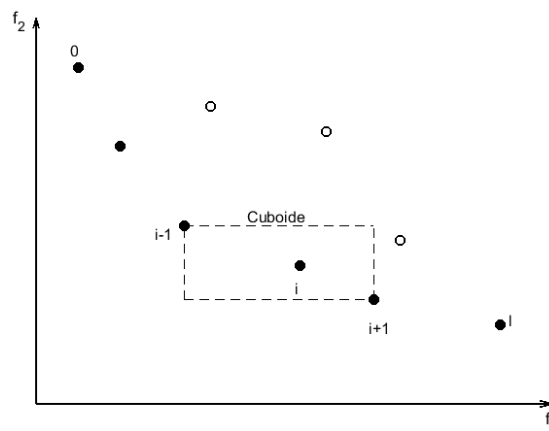


Figura 56 ^[31]: Rappresentazione schematica della *crowding-distance*, i punti contrassegnati da cerchi pieni sono soluzioni dello stesso fronte non dominato, grafico riadattato

Per ciascun punto del medesimo fronte non dominato si stima una distanza locale rispetto ai vicini nello spazio degli obiettivi; i punti con *crowding distance* maggiore si trovano in regioni meno dense e vengono preferiti, perché aiutano a mantenere una distribuzione più uniforme delle soluzioni lungo il fronte di Pareto. In questo modo NSGA-II non cerca solo soluzioni buone, ma anche un fronte ben distribuito.

La procedura viene iterata fino al soddisfacimento dei criteri di terminazione.

In sintesi, il *non-dominated sorting* guida la convergenza verso il fronte di Pareto, mentre la *crowding distance* contribuisce a preservarne la diversità. Il metodo è detto elitista controllato perché le soluzioni migliori non vengono perse da una generazione alla successiva, ma al tempo stesso il numero di individui appartenenti al fronte di Pareto viene controllato per evitare che la popolazione collassi troppo presto su una regione ristretta del fronte.

5.4.1 Implementazione della procedura di ottimizzazione

L'ottimizzazione, come anticipato, è stata implementata in MATLAB mediante la funzione Gamultiobj. I parametri significativi riportati nel Codice 10 vengono introdotti progressivamente nei punti in cui risultano rilevanti.

```

% Impostazione opzioni Gamultiobj
options = optimoptions('gamultiobj', ...
    'PopulationSize', 200, ...
    'MaxGenerations', 300, ...
    'FunctionTolerance', 1e-6, ...
    'UseParallel', true, ...
    'Display', 'iter', ...
    'PlotFcn', {@gaplotpareto});
% Utilizzo di Gamultiobj
[Xpareto, Fpareto, exitflag, output] = gamultiobj( ...
    fitnessFcn, nvars, [], [], [], [], lb, ub, [], intcon, options);

```

Codice 10: MATLAB, impostazione e utilizzo di Gamultiobj

- La procedura è stata formulata assumendo come variabili decisionali i 9 valori interi

$$x = [n_1, n_2, \dots, n_9]$$

L'argomento `fitnessFcn` passato a `gamultiobj` definisce le due funzioni obiettivo da minimizzare, entrambe calcolate a partire dal vettore decisionale x . In particolare, il primo obiettivo è la temperatura media predetta dal modello surrogato:

$$f_1(x) = \hat{T}_{av}(x),$$

mentre il secondo è il numero totale di pin:

$$f_2(x) = \sum_{j=1}^9 n_j$$

L'ottimizzazione consiste quindi nel determinare come ripartire i pin tra le 9 zone, in modo da ottenere il miglior compromesso tra prestazioni termiche e numero complessivo di cilindri.

- Il parametro `nvars` specifica il numero di variabili decisionali del problema e, nel caso in esame, assume quindi valore pari a 9. I vettori `lb` e `ub` hanno dimensione del vettore decisionale e definiscono rispettivamente i limiti inferiori e superiori delle variabili. Poiché tutte le zone hanno la stessa capacità massima, i limiti risultano:

$$\begin{aligned} lb &= [0, 0, \dots, 0] \\ ub &= [14, 14, \dots, 14]. \end{aligned}$$

Ne consegue che ciascuna variabile durante l'ottimizzazione può assumere esclusivamente valori compresi tra 0 e 14. La fattibilità intera è inoltre imposta mediante il vettore `intcon` = 1:9, che indica al solver che tutte e 9 le variabili devono essere trattate come intere.

- Le opzioni principali del solver sono state impostate come segue: `PopulationSize` = 200, `MaxGenerations` = 300 e `FunctionTolerance` = 10^{-6} . La popolazione di 200 individui è stata scelta per garantire, a ogni generazione, una sufficiente varietà di configurazioni candidate nello spazio discreto delle 9 variabili, evitando al tempo stesso un costo computazionale eccessivo per singolo *run*. Inoltre, poiché il numero totale di pin può assumere valori interi compresi tra 0 e 126, nel caso plausibile in cui a ciascun incremento di $n_{pin,tot}$ corrisponda almeno una configurazione capace di migliorare la temperatura media rispetto ai valori ottenibili con un numero inferiore di pin, il fronte di Pareto nel piano $(n_{pin,tot}, \hat{T}_{av})$ potrebbe contenere fino a 127 punti distinti, uno per ciascun valore di $n_{pin,tot}$; a meno di eventuali

configurazioni equivalenti che condividano lo stesso numero di pin e il medesimo valore predetto di temperatura. In tale prospettiva, una popolazione di 200 individui risulta sufficientemente ampia da rappresentare in ogni generazione una porzione significativa del fronte potenzialmente ottenibile. Il numero massimo di 300 generazioni fornisce invece un budget evolutivo sufficientemente ampio per consentire al fronte di Pareto di stabilizzarsi progressivamente; ciascuna generazione corrisponde infatti a un ciclo evolutivo dell'algoritmo, come illustrato in precedenza. Il parametro `FunctionTolerance` controlla il criterio di arresto basato sulla stabilizzazione del fronte di Pareto. Nel caso di `Gamultiobj`, esso non viene applicato direttamente ai valori delle funzioni obiettivo, ma a una quantità interna denominata *spread*, che misura sia quanto il fronte si stia ancora spostando da una generazione alla successiva, sia quanto i punti risultino distribuiti lungo di esso. Quando, per un certo numero di generazioni consecutive, la variazione relativa media di tale quantità diventa sufficientemente piccola, il solver considera il fronte sostanzialmente stabilizzato e può arrestare l'ottimizzazione. Nel presente lavoro è stato scelto un valore particolarmente restrittivo, pari a 10^{-6} , inferiore al valore di default di `Gamultiobj` (10^{-4}). Tale impostazione è giustificata dal ridotto costo computazionale delle valutazioni del modello surrogato: essendo il singolo *run* comunque poco oneroso, si è ritenuto preferibile adottare una tolleranza severa, in modo da lasciare all'algoritmo un margine evolutivo più ampio e limitare gli arresti anticipati ai soli casi effettivamente necessari.

- L'opzione `UseParallel = true` è stata attivata per valutare in parallelo le funzioni obiettivo sui diversi individui della popolazione, sfruttando l'ambiente parallelo disponibile. Tale scelta ha consentito di ridurre il tempo di esecuzione complessivo dei singoli *run*. Va però osservato che, in presenza di parallelizzazione, la perfetta ripetibilità della traiettoria evolutiva non è sempre rigorosamente garantita, poiché l'ordine effettivo di esecuzione delle valutazioni può dipendere dall'ambiente di calcolo e dal numero di *worker* disponibili. I *worker* rappresentano le unità di esecuzione parallela utilizzate da MATLAB per distribuire il calcolo tra più processi simultanei. Per questo motivo, pur fissando esplicitamente il `Seed` del generatore pseudo-casuale, che determina la specifica realizzazione stocastica dell'ottimizzazione, inclusa la popolazione iniziale casuale e le successive operazioni di selezione a torneo, crossover e mutazione, si è preferito basare l'analisi finale su più *run* indipendenti, così da ridurre la dipendenza dei risultati da una singola traiettoria evolutiva. Nel caso in esame, le esecuzioni sono state effettuate utilizzando 16 worker.
- In uscita, `Gamultiobj` restituisce anzitutto la matrice `Xpareto`, le cui righe rappresentano le configurazioni decisionali appartenenti al set Pareto trovato dal solver. A ciascuna di esse corrisponde una riga della matrice `Fpareto`, che contiene i valori delle due funzioni obiettivo. Tali output costituiscono il risultato diretto di una singola ottimizzazione surrogata.

L'ottimizzazione, come appena descritto, non è stata eseguita una sola volta, ma ripetuta per 10 *run* indipendenti, associate a `Seed` differenti compresi tra 42 e 51, `rng(42-51, 'twister')`. I valori numerici dei `Seed` sono stati scelti in modo arbitrario, senza seguire un criterio specifico, e servono unicamente a definire realizzazioni differenti della componente pseudo-casuale dell'algoritmo. Le soluzioni non dominate ottenute nelle singole *run* sono state quindi aggregate e filtrate, così da costruire un fronte di Pareto globale, successivamente sottoposto a validazione. I risultati di tale procedura vengono discussi nel paragrafo successivo.

5.4.2 Risultati dell'ottimizzazione e validazione CFD ad alta fedeltà

Si inizia commentando le 10 *run* di ottimizzazione. Ciascuna delle esecuzioni indipendenti di `Gamultiobj` ha richiesto mediamente 72 s per un totale di circa 12 minuti, restituendo per ogni *run* un insieme Pareto costituito da 70 configurazioni non dominate. In tutti i casi, l'arresto del solver è

avvenuto per raggiungimento del numero massimo di generazioni imposto ($\text{MaxGenerations} = 300$), corrispondente a 60 000 valutazioni del modello surrogato per *run*.

A titolo esemplificativo, in Figura 57 è riportato l'andamento del Pareto *spread* medio relativo alla prima *run*.

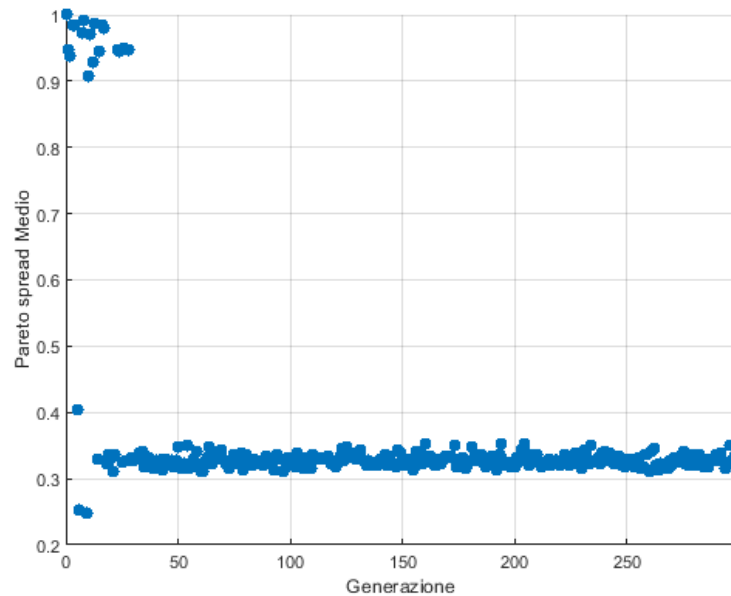


Figura 57: Andamento del Pareto *spread* Medio in funzione della generazione per la prima *run* di ottimizzazione

Si osserva come, dopo una fase iniziale di assestamento, il valore dello *spread* medio si stabilizzi rapidamente e mostri soltanto oscillazioni contenute, suggerendo che il fronte di Pareto abbia raggiunto una condizione prossima alla convergenza ben prima della generazione 300. Tuttavia, considerato il costo computazionale molto ridotto di ciascuna ottimizzazione grazie all'impiego del modello surrogato, si è preferito mantenere il limite massimo di 300 generazioni, non tanto per necessità di ulteriore convergenza, quanto come verifica di robustezza della stabilità raggiunta, così da ridurre il rischio di arrestare prematuramente l'algoritmo in presenza di una stabilizzazione solo apparente.

Dalle soluzioni non dominate ottenute nelle dieci ottimizzazioni è stato costruito il fronte di Pareto globale del modello surrogato. In Figura 58 esso è mostrato insieme all'intero dataset, al Pareto del dataset stesso e ai punti ottenuti dalla validazione CFD delle configurazioni ottimizzate.

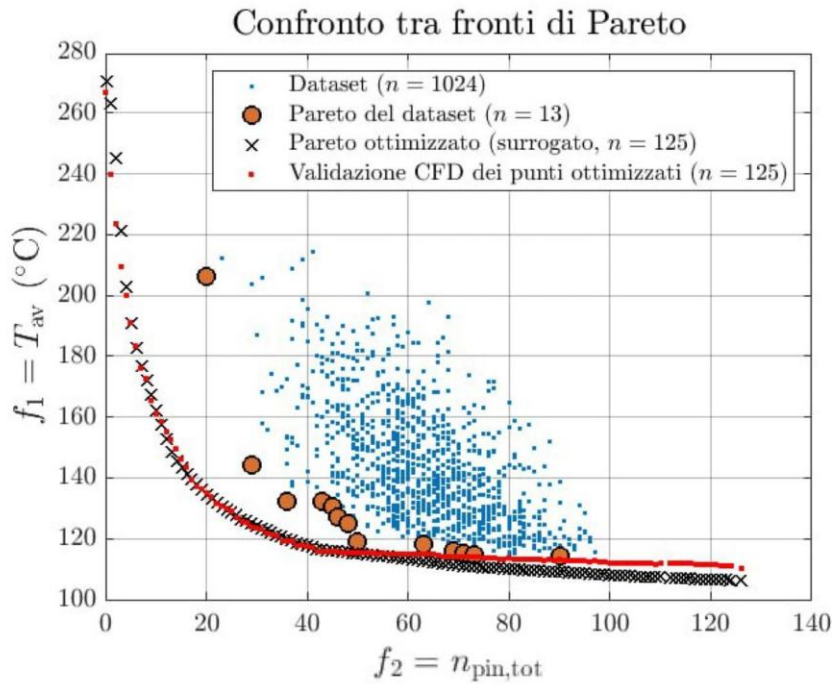


Figura 58: Confronto nel piano obiettivi (f_2, f_1) tra l'intero dataset CFD, il relativo fronte di Pareto, il fronte di Pareto globale ottenuto mediante ottimizzazione sul modello surrogato e la successiva validazione CFD dei punti ottimizzati

La procedura ha portato all'individuazione di un insieme complessivo di 125 configurazioni non dominate, distribuite lungo un ampio intervallo di valori di $n_{\text{pin,tot}}$. In particolare, risultano rappresentati quasi tutti i numeri totali di pin possibili, con la sola assenza delle configurazioni corrispondenti a $n_{\text{pin,tot}} = 111$ e $n_{\text{pin,tot}} = 125$.

Il confronto con il dataset originario mostra che il fronte ottenuto mediante ottimizzazione non si limita a riprodurre i compromessi già presenti nei dati iniziali, ma individua una famiglia più ricca e continua di soluzioni candidate (125 configurazioni non dominate contro le 13 del dataset iniziale), collocate in regioni del piano obiettivi scarsamente rappresentate dal campionamento iniziale e migliorative rispetto ai migliori compromessi già contenuti nel dataset. In questo senso, il modello surrogato non è stato utilizzato soltanto come strumento di interpolazione, ma come vero motore di esplorazione del dominio di progetto.

Per verificare l'attendibilità fisica delle configurazioni ottimizzate, tutte le 125 soluzioni generate dal surrogato sono state sottoposte a simulazione CFD. Questa fase costituisce il passaggio essenziale di validazione finale dell'approccio e ha richiesto complessivamente poco più di 6 ore di calcolo. Già dal confronto riportato in Figura 58 si osserva come gli scostamenti più marcati tra predizione surrogata e validazione CFD tendano a concentrarsi nelle code del fronte, mentre nella regione centrale l'accordo risulta generalmente molto buono. Per approfondire tale aspetto, in Figura 59 sono riportati sia la distribuzione dell'errore relativo sui punti validati, sia l'andamento dell'errore stesso in funzione del numero totale di pin, così da evidenziare non solo l'entità complessiva dello scostamento, ma anche le regioni del fronte in cui esso si manifesta con maggiore intensità.

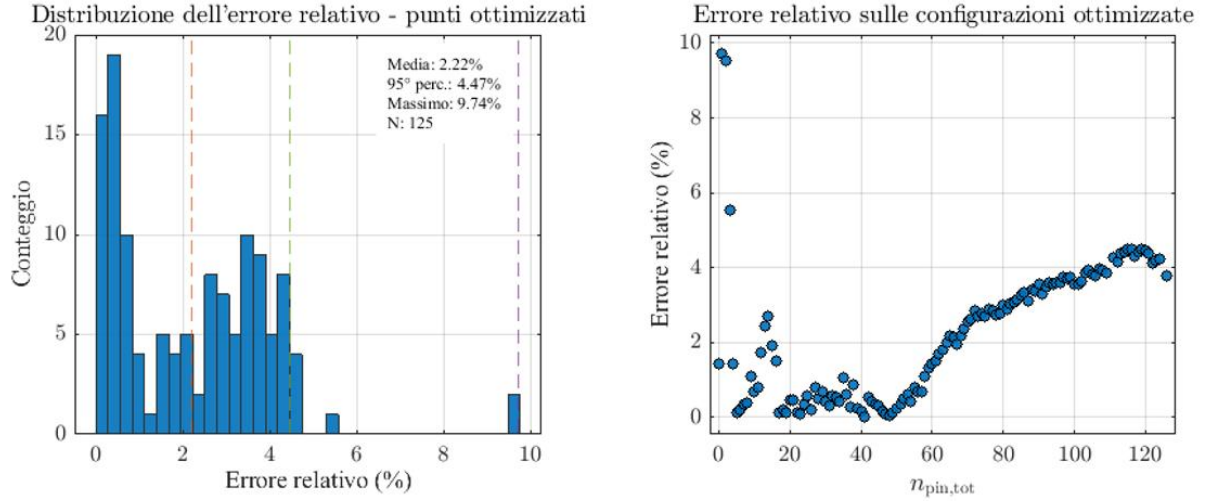


Figura 59: Distribuzione dell'errore relativo sui 125 punti ottimizzati validati mediante CFD (a sinistra) e andamento dello stesso errore in funzione del numero totale di pin $n_{pin,tot}$ (a destra)

Il confronto tra predizioni surrogate e valori CFD sui punti ottimizzati evidenzia un buon accordo complessivo, pur con errori leggermente più elevati rispetto a quelli osservati sul *test set* casuale (paragrafo 5.3.4). Tale risultato è del tutto atteso, dal momento che l'ottimizzazione porta inevitabilmente a esplorare anche regioni del dominio poco rappresentate nel dataset, nelle quali la capacità di generalizzazione del surrogato può risultare meno accurata rispetto a quanto osservato su un *test set* casuale. In particolare, l'errore relativo (come precedentemente definito nell'Eq 31, paragrafo 5.3.4) sui 125 punti validati risulta pari in media al 2.22%, con 95° percentile pari a 4.47% e massimo pari a 9.74%. L'andamento dell'errore relativo in funzione di $n_{pin,tot}$, riportato nella parte destra di Figura 59, conferma in modo più dettagliato che gli scostamenti maggiori si concentrano nelle code del fronte. In particolare, gli errori più elevati si osservano per configurazioni con numero di pin molto basso, mentre nella prima parte della fascia intermedia l'accordo tra surrogato e CFD risulta ottimo. Per valori più elevati di $n_{pin,tot}$, a partire indicativamente da circa 60 pin, si osserva tuttavia una nuova crescita progressiva dell'errore relativo. L'andamento complessivo può essere interpretato alla luce dei grafici riportati in Figura 60, nei quali sono mostrati, rispettivamente, il numero teorico di combinazioni possibili per ciascun valore di $n_{pin,tot}$ e la distribuzione effettiva dei valori di $n_{pin,tot}$ nel dataset utilizzato per l'addestramento, la validazione e il test del modello surrogato.

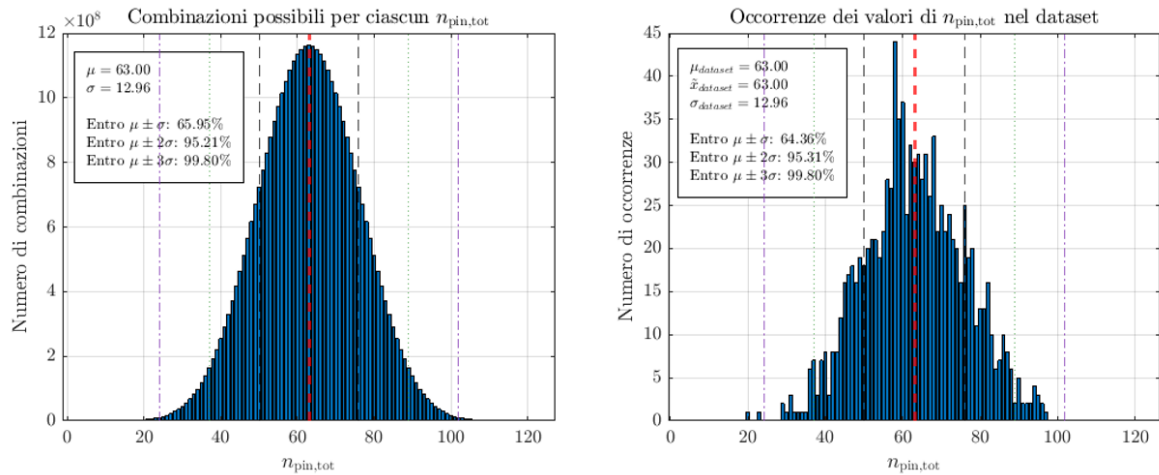


Figura 60: Numero di configurazioni possibili in funzione del numero totale di pin (a sinistra). Distribuzione dei valori di $n_{pin,tot}$ nel dataset di 1024 configurazioni (a destra)

Occorre infatti ricordare che il campionamento non è stato effettuato rispetto a $n_{\text{pin,tot}}$, bensì nello spazio delle 9 variabili di input $[n_1, \dots, n_9]$, che descrivono la distribuzione dei pin nelle diverse zone. Di conseguenza, è naturale che risultino maggiormente rappresentati i valori di $n_{\text{pin,tot}}$ associati a un numero più elevato di configurazioni possibili. Ad esempio, nei casi estremi $n_{\text{pin,tot}} = 0$ e $n_{\text{pin,tot}} = 126$ esiste una sola configurazione possibile, mentre il numero di combinazioni cresce fino a raggiungere un massimo in corrispondenza dei valori centrali, in particolare attorno a $n_{\text{pin,tot}} = 63$ (ordine di 10^9 configurazioni possibili), per poi decrescere nuovamente in modo simmetrico.

Come mostrato in Figura 60, la distribuzione teorica del numero di combinazioni possibili al variare di $n_{\text{pin,tot}}$ presenta un andamento pressoché gaussiano. In modo coerente, anche il campione effettivamente estratto mediante Sobol sampling presenta una maggiore concentrazione di configurazioni nei valori centrali di $n_{\text{pin,tot}}$, mentre le code risultano inevitabilmente meno rappresentate. I valori di media e deviazione standard riportati in figura rafforzano ulteriormente quanto appena esposto: l'intervallo $\mu \pm 2\sigma$ è pari a circa $[37, 89]$, cioè circa il 42% dell'intervallo totale $[0, 126]$, e contiene già circa il 95% sia delle configurazioni teoricamente possibili sia di quelle effettivamente campionate.

Tale struttura del dataset contribuisce a spiegare perché, per valori molto bassi di $n_{\text{pin,tot}}$, il surrogato possa mostrare errori più marcati: in tali regioni il numero di configurazioni disponibili è ridotto, il campionamento è meno denso e, al tempo stesso, la temperatura varia in modo relativamente rapido al variare della configurazione. Per contro, nella regione centrale, maggiormente popolata e meglio rappresentata nel dataset, il modello riesce a cogliere con maggiore accuratezza l'effetto della distribuzione zonale dei pin sulla risposta termica.

Analogamente al caso di $n_{\text{pin,tot}}$ molto bassi, anche la regione ad alto numero di pin risulta poco rappresentata nel dataset. In questo caso, tuttavia, la temperatura varia in modo molto più contenuto e il problema non è tanto un'elevata sensibilità locale, quanto piuttosto la difficoltà del surrogato nel cogliere correttamente il comportamento asintotico del fronte: la scarsità di campioni nelle configurazioni più dense tende, infatti, a far sovrastimare il margine residuo di miglioramento di T_{av} , ciò contribuisce a spiegare l'andamento crescente degli errori a partire da circa $n_{\text{pin,tot}} = 60$ pin.

Nel complesso, questa lettura fornisce una chiave interpretativa plausibile degli andamenti osservati, pur senza voler attribuire a tale analisi un significato rigorosamente esaustivo.

Al di là dell'analisi puntuale degli errori, l'aspetto più rilevante dal punto di vista progettuale è il fronte di Pareto ottenuto dopo validazione CFD, riportato in Figura 61.

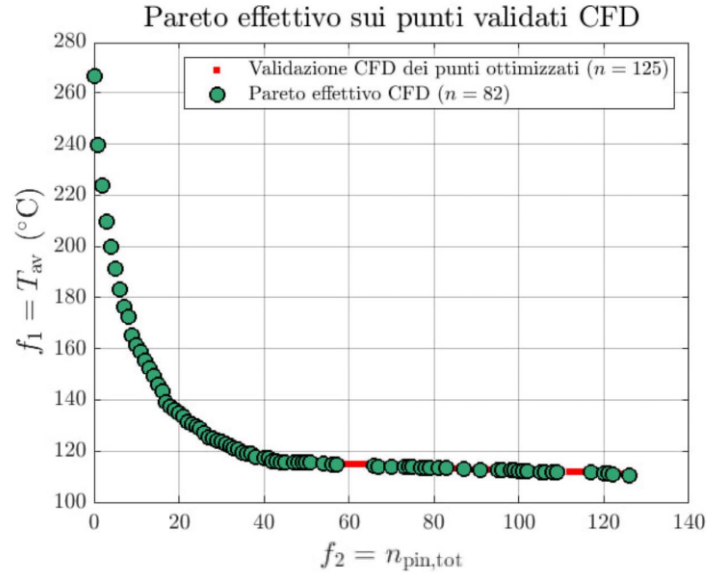


Figura 61: Fronte di Pareto effettivo ottenuto sui risultati validati mediante simulazione CFD

Va precisato che esso non coincide necessariamente con il fronte di Pareto assoluto del problema, poiché non vi è garanzia che l'insieme contenga in senso stretto tutte le migliori soluzioni possibili. Tuttavia, la forma regolare del fronte e la buona densità dei punti ottenuti (82 configurazioni contro le 13 ottenute dal dataset) suggeriscono che la procedura abbia individuato un insieme di compromessi fisicamente credibili e progettualmente significativi.

Inoltre, nonostante le limitazioni discusse nell'analisi degli errori, è opportuno ricordare che l'obiettivo ultimo del *workflow* non era tanto prevedere con accuratezza assoluta ogni singolo valore di T_{av} , quanto piuttosto guidare efficacemente la ricerca delle configurazioni migliori. Sotto questo profilo, l'ottimizzazione risulta convincente: le configurazioni trovate riflettono una logica di distribuzione fisicamente coerente, come discusso nel seguito; inoltre, dopo validazione CFD, 82 delle 125 configurazioni analizzate appartengono ancora al fronte di Pareto effettivo, ciò indica che il surrogato ha preservato in larga misura il corretto ordinamento relativo tra configurazioni concorrenti.

L'analisi delle configurazioni lungo il fronte mostra che l'algoritmo non solo coglie correttamente la logica fisica del problema, ma la preserva anche in modo coerente al variare del numero totale di pin. In particolare, per valori di $n_{pin,tot}$ compresi tra 0 e 14, i cilindri vengono progressivamente collocati nella sola zona 5, ossia quella centrale in corrispondenza dell'area calda. Per valori compresi tra 15 e 42, la zona 5 rimane saturata a 14 cilindri, mentre i pin aggiuntivi vengono distribuiti in modo pressoché alternato tra le zone 2 e 8. Ad esempio, una configurazione con 18 pin complessivi corrisponde a una distribuzione 14 + 2 + 2. Questo schema prosegue fino alla configurazione a 42 pin, nella quale risulta sostanzialmente occupata l'intera regione maggiormente influente dal punto di vista termico. Tale progressione è illustrata in modo esemplificativo in Figura 62, nella quale sono riportate alcune configurazioni appartenenti al fronte di Pareto, così da rendere più evidente la logica di riempimento appena descritta.

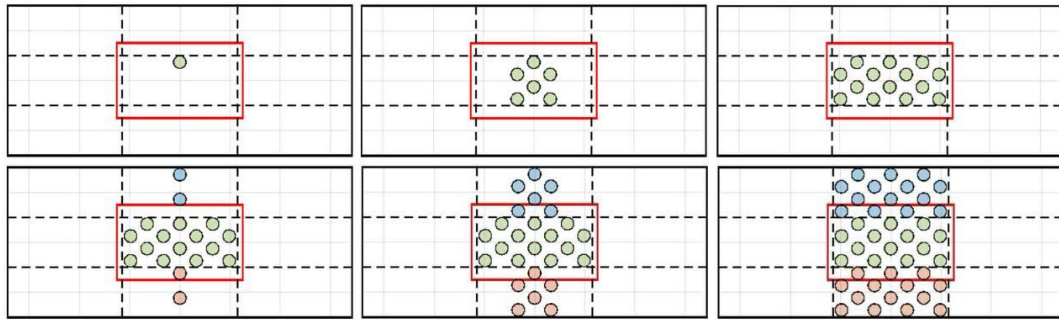


Figura 62: Alcune configurazioni rappresentative appartenenti al fronte di Pareto validato, riportate per evidenziare la progressiva saturazione delle zone termicamente più influenti al crescere del numero totale di pin: dapprima la zona 5 centrale, quindi le zone 2 e 8 in modo pressoché alternato

Questo risultato non era affatto scontato. Al crescere di $n_{\text{pin,tot}}$, infatti, cambia completamente l'insieme delle configurazioni e cresce rapidamente il numero di distribuzioni possibili dei pin tra le 9 zone. Nulla garantiva quindi, a priori, che passando da una configurazione ad un'altra, ad esempio da 14 a 15 pin, il modello mantenesse la stessa logica di riempimento, limitandosi ad aggiungere il nuovo cilindro nella zona più opportuna; avrebbe invece potuto alterare in modo meno ordinato anche la distribuzione dei 14 pin precedenti. Il fatto che ciò non avvenga, e che le configurazioni ottimizzate mostrino una progressione regolare e fisicamente interpretabile, costituisce un'indicazione importante della capacità del *workflow* di selezionare soluzioni coerenti tra le numerosissime combinazioni possibili.

Per numeri di pin superiori a 42 le configurazioni risultano progressivamente meno rappresentative sul piano progettuale, poiché le variazioni di T_{av} diventano più contenute; ciò significa che, una volta saturate le regioni termicamente più efficaci, il posizionamento dei pin residui in una zona piuttosto che in un'altra diventa relativamente indifferente ai fini della prestazione globale. Tale comportamento è coerente con il fatto che ci si colloca oltre il *knee point* del fronte, ossia in una regione in cui ulteriori incrementi del numero di pin producono benefici termici via via marginali e quindi meno interessanti dal punto di vista dell'ottimizzazione.

A completamento della discussione, nelle Figure 63 e 64 sono riportati due casi rappresentativi del fronte validato, corrispondenti rispettivamente a una configurazione a 14 pin e a una configurazione a 42 pin, insieme alle relative mappe termiche ottenute mediante simulazione CFD.

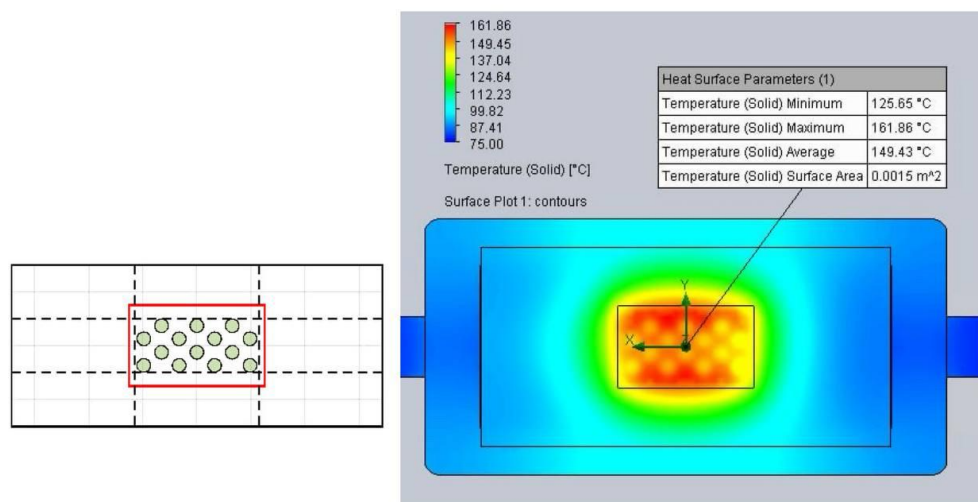


Figura 63: Configurazione ottimizzata a 14 pin appartenente al fronte di Pareto validato e corrispondente mappa termica ottenuta mediante simulazione CFD

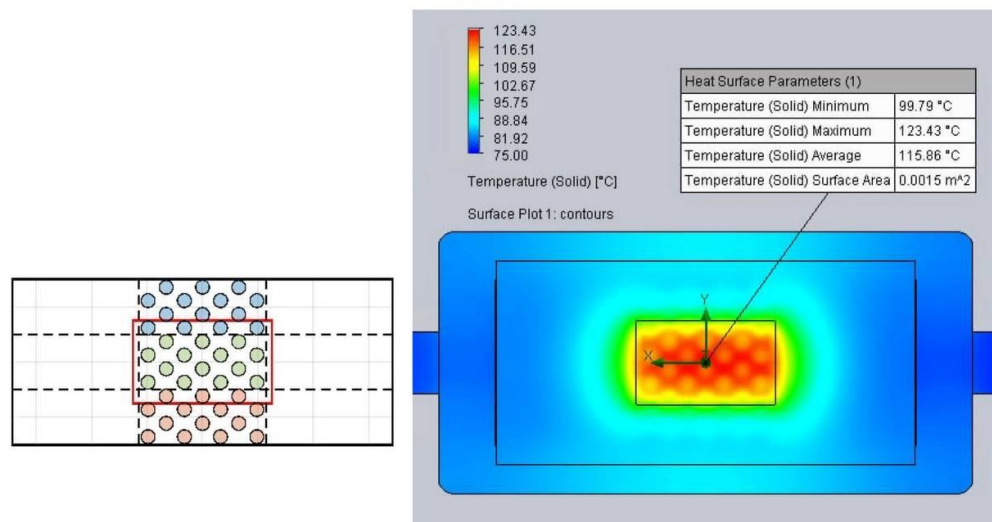


Figura 64: Configurazione ottimizzata a 42 pin appartenente al fronte di Pareto validato e corrispondente mappa termica ottenuta mediante simulazione CFD

Nel complesso, i risultati ottenuti possono quindi essere considerati promettenti per due ragioni principali: in primo luogo, l'ottimizzazione surrogata genera rapidamente un fronte molto più ricco di quello contenuto nel dataset iniziale; in secondo luogo, la validazione CFD conferma la bontà della maggior parte delle soluzioni individuate.

5.5 Conclusioni

In questo capitolo è stato sviluppato e validato un *workflow* completo per l'ottimizzazione multi-obiettivo assistita da surrogato di una configurazione pin-fin. Il metodo integra in un'unica procedura la generazione automatizzata della geometria CAD, l'esecuzione automatica delle simulazioni CFD e la costruzione di un dataset di configurazioni, successivamente utilizzato per l'addestramento di un modello surrogato basato su rete neurale. Il surrogato così ottenuto è stato infine impiegato all'interno di una procedura di ottimizzazione genetica multi-obiettivo, consentendo l'individuazione di configurazioni Pareto poi validate mediante simulazioni CFD ad alta fedeltà.

Il principale risultato del capitolo riguarda il guadagno in termini di costo computazionale. Se si ipotizzasse infatti di utilizzare direttamente l'algoritmo genetico sulla CFD, anche assumendo una convergenza già dopo 50 generazioni (Figura 57, paragrafo 5.4.2), una singola *run* con popolazione pari a 200 individui richiederebbe $200 \times 50 = 10\,000$ simulazioni. Considerando il tempo stimato medio di circa 193 s per aggiornamento CAD e simulazione CFD, ciò corrisponderebbe a circa 22 giorni di calcolo per una sola *run*, che nel caso del surrogato restituisce peraltro un fronte di Pareto costituito da 70 configurazioni non dominate. Nel presente lavoro, al contrario, il modello surrogato ha consentito di eseguire 10 *run* complete da 300 generazioni ciascuna (assicurando convergenza in modo robusto), per un totale di $200 \times 300 \times 10 = 600\,000$ valutazioni, costruendo infine un fronte globale di 125 configurazioni non dominate (82 se si considera il fronte validato), quindi più ricco e più affidabile sotto il profilo statistico rispetto a quello ottenibile da una singola esecuzione. Un numero di simulazioni di tale ordine di grandezza, se affrontato direttamente con CFD, richiederebbe più di 3 anni e mezzo di calcolo, risultando del tutto impraticabile.

L'approccio sviluppato consente invece di concentrare l'impiego della CFD soltanto nei passaggi realmente necessari: circa 55 ore per la generazione iniziale del dataset, pochi minuti per l'addestramento del modello neurale e per l'ottimizzazione surrogata, e circa 6 ore per la validazione finale CFD. Il tempo complessivo dell'intera procedura risulta così poco più di 61 ore.

Il vantaggio del metodo non risiede quindi nell'azzeramento del costo della CFD, bensì nella sua riallocazione strategica: la CFD viene impiegata per costruire un dataset informativo e solo in chiusura per validare le soluzioni candidate, evitando di investirla in modo inefficiente durante l'esplorazione iterativa dell'algoritmo genetico.

Dal punto di vista qualitativo, i risultati ottenuti sono promettenti. Sebbene il caso studio considerato sia relativamente semplice dal punto di vista dell'interpretazione fisica, esso non lo è affatto dal punto di vista combinatorio, poiché lo spazio di progetto associato alla distribuzione dei pin nelle 9 zone è estremamente ampio ($15^9 \approx 3.84 \times 10^{10}$ configurazioni possibili). In questo contesto, il fatto che il metodo riesca a individuare configurazioni fisicamente coerenti, a costruire un fronte di Pareto ben strutturato e a preservare, dopo validazione CFD, una parte significativa dell'ordinamento relativo tra le soluzioni rappresenta un'indicazione importante della validità dell'approccio.

Nel complesso, il capitolo ha quindi mostrato che l'impiego combinato di automazione CAD-CFD, *surrogate modeling* e ottimizzazione genetica multi-obiettivo costituisce una strategia efficace per affrontare problemi di progetto altrimenti proibitivi dal punto di vista computazionale. Pur essendo stato applicato a un caso ancora semplificato, il metodo appare sufficientemente robusto e flessibile da risultare promettente anche per problemi più realistici e meno intuitivi. Nel capitolo successivo, pur mantenendo ancora alcune semplificazioni, il metodo sviluppato verrà applicato a un caso di maggiore interesse applicativo, affrontando un problema di ottimizzazione relativo alla forma dei pin-fin.

6. Applicazione del metodo a un caso più realistico: ottimizzazione della forma dei pin

Dopo aver sviluppato e validato nel Capitolo 5 un metodo di ottimizzazione multi-obiettivo assistita da surrogato su un caso semplificato di distribuzione spaziale dei pin, in questo capitolo la stessa logica metodologica viene applicata a un problema di maggiore interesse progettuale. L'attenzione non è più rivolta alla disposizione dei cilindri su zone discrete della superficie riscaldata, bensì alla forma dei pin.

Anche in questo caso il problema viene formulato come ottimizzazione multi-obiettivo. Il primo obiettivo è la minimizzazione della temperatura media T_{av} sulla superficie riscaldata, assunta come indicatore sintetico delle prestazioni termiche del dissipatore. Il secondo obiettivo è la minimizzazione del volume del pin, assunto come misura indiretta della quantità di materiale impiegato, e quindi della massa del sistema.

L'interesse del problema non risiede soltanto nell'individuazione di forme efficienti in senso assoluto, ma anche nel confronto con una geometria di riferimento nota e ampiamente utilizzata, quale il cilindro a diametro costante. Tale confronto consente di verificare se la redistribuzione del materiale lungo l'altezza del pin possa tradursi in un vantaggio concreto rispetto a una soluzione tradizionale e di ricavare elementi interpretativi utili a comprendere le ragioni di tale eventuale miglioramento.

Il problema analizzato si colloca in una posizione intermedia tra il caso puramente metodologico discusso nel capitolo precedente e un'applicazione pienamente industriale. Pur mantenendo alcune semplificazioni del problema progettuale complessivo, esso introduce un grado di libertà geometrico significativamente più ricco, legato alla variazione del profilo del pin lungo l'altezza. Il capitolo ha quindi una duplice finalità: da un lato, verificare la trasferibilità del *workflow* sviluppato nel Capitolo 5 a un problema nuovo; dall'altro, comprendere quali caratteristiche geometriche consentano alle forme ottimizzate di superare le prestazioni dei pin cilindrici. Anche la scelta di concentrare l'analisi esclusivamente sulla forma del pin rappresenta, di per sé, una semplificazione, poiché in un contesto più generale potrebbero essere inclusi ulteriori gradi di libertà progettuali, come lo spessore della *baseplate* o altri parametri geometrici del dissipatore. Tale restrizione è tuttavia intenzionale, in quanto consente di isolare l'effetto della forma dei pin e di valutarne più chiaramente il contributo. I risultati ottenuti devono però essere interpretati con riferimento allo specifico assetto considerato, nel quale portata, numero di pin, disposizione spaziale, altezza e materiali sono mantenuti fissati.

L'analisi consente quindi di valutare se, nelle condizioni considerate, l'adozione di forme diverse dal cilindro permetta una distribuzione del materiale più efficiente, offrendo al contempo indicazioni progettuali utili sull'effettiva convenienza di esplorare geometrie alternative, senza però pretendere di trarre conclusioni generali indipendenti dal caso studiato.

6.1 Setup e dataset

Il modello geometrico di simulazione adottato in questo capitolo deriva direttamente da quello sviluppato e discusso nella Sezione 4.1 (Figura 23), usato anche nel Capitolo 5. Rispetto al capitolo precedente, sono state introdotte alcune modifiche, rese necessarie dalla natura del nuovo problema.

In primo luogo, è stata modificata la forma dell'area riscaldata, passata da $50\text{ mm} \times 30\text{ mm}$ a $28\text{ mm} \times 55\text{ mm}$. Pur variando di poco l'area complessiva, questa nuova geometria si sviluppa

maggiormente lungo l'asse y e occupa quasi interamente la larghezza disponibile in tale direzione. Tale scelta è stata adottata per ottenere una configurazione in cui i pin, distribuiti in modo da limitare i corridoi laterali di *bypass* del fluido, risultassero comunque tutti collocati sulla zona direttamente riscaldata. La maggiore estensione dell'area calda lungo la direzione y consente infatti di soddisfare simultaneamente entrambe le esigenze: ridurre il *bypass* laterale, ossia il passaggio del flusso ai lati della schiera con interazione limitata con i pin, e fare in modo che tutti gli elementi agiscano direttamente sulla superficie di interesse. Ciò rende più significativo l'effetto della forma dei pin sulla prestazione termica complessiva.

Per quanto riguarda il numero e la disposizione dei pin, è stata adottata una configurazione composta da 11 elementi, distribuiti secondo uno schema a scacchiera su tre colonne, rispettivamente da 4, 3 e 4 pin. La scelta del numero di pin e della loro collocazione è stata effettuata considerando la geometria più ingombrante prevista dalla parametrizzazione della forma, che verrà descritta nel paragrafo successivo. In particolare, la configurazione più piena corrisponde al caso di un pin cilindrico a diametro costante pari a 7 mm, mentre la più snella corrisponde al cilindro di diametro costante pari a 1 mm. Assumendo quindi come caso limite il diametro massimo di 7 mm, il posizionamento dei pin è stato definito in modo da garantire la compatibilità geometrica anche per la configurazione più ingombrante. Lungo la direzione x , la larghezza disponibile di 28 mm consente l'inserimento di tre pin da 7 mm, con una distanza uniforme tra le circonferenze pari a 1.75 mm, ottenuta distribuendo lo spazio residuo nei quattro intervalli tra bordo e pin e tra pin adiacenti. Analogamente, lungo la direzione y , la lunghezza di 55 mm consentirebbe teoricamente l'inserimento di sette pin da 7 mm, con una distanza uniforme tra le circonferenze pari a 0.75 mm, ottenuta distribuendo lo spazio residuo negli otto intervalli disponibili. A partire da questa griglia teorica 3×7 , corrispondente a 21 possibili posizioni, sono stati rimossi 10 pin in modo da ottenere una disposizione a scacchiera tipica dei pin fin, costituita appunto da 11 pin distribuiti secondo lo schema 4-3-4. Tale procedimento, schematizzato in Figura 65, definisce in modo univoco il posizionamento degli assi dei pin, che restano fissati per tutte le configurazioni esaminate nel capitolo.

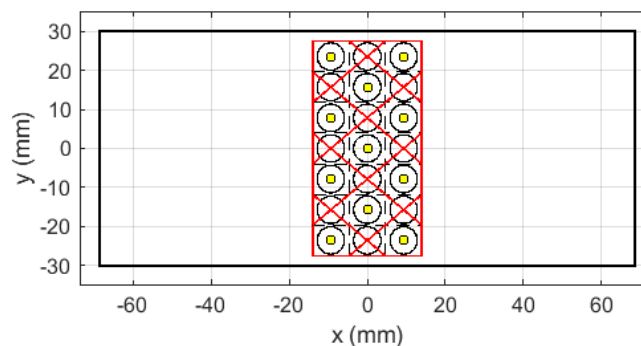


Figura 65: Schema della procedura adottata per definire il posizionamento degli assi dei pin sulla superficie riscaldata

In Tabella 10 sono inoltre riportate le quote geometriche utilizzate per il loro posizionamento, al fine di definire in modo completo la configurazione di riferimento e renderne riproducibile la generazione in MATLAB.

Posizionamento assi Pin-Fin	
Zona posizionamento Pin (Area calda)	28 mm x 55 mm
Zona effettiva posizionamento Pin	24,5 mm x 53,5 mm
Distanza assi Pin esterni - bordi Area calda (sinistro e destro)	3,5 mm (raggio) + 1,75 mm (spazio) = 5,25 mm
Distanza assi Pin esterni - bordi Area calda (superiore e inferiore)	3,5 mm (raggio) + 0,75 mm (spazio) = 4,25 mm
Distanza tra assi Pin colonne adiacenti lungo x	3,5 mm (raggio) + 1,75 mm (spazio) + 3,5 mm (raggio) = 8,75 mm
Distanza tra assi Pin colonne adiacenti lungo y	3,5 mm (raggio) + 0,75 mm (spazio) + 3,5 mm (raggio) = 7,75 mm
Distanza tra assi Pin righe adiacenti su stessa colonna	2 x Distanza tra assi Pin colonne adiacenti lungo y = 15,5 mm
Colonne totali di Pin	3
Righe di Pin per colonna	4-3-4

Tabella 10: Quote geometriche utilizzate per il posizionamento degli assi dei pin sulla superficie riscaldata

Un'ulteriore modifica rilevante rispetto ai capitoli precedenti riguarda i materiali assegnati ai pin e alla piastra. In questo capitolo, infatti, entrambi sono assunti in alluminio. Tale scelta è legata al fatto che l'analisi si concentra su forme dei pin variabili e potenzialmente complesse, la cui realizzazione risulta più coerente con tecnologie di *additive manufacturing* rispetto a soluzioni tradizionali ottenute per lavorazioni convenzionali. In questo contesto, l'alluminio rappresenta un materiale più plausibile del rame per un'eventuale applicazione reale. Va tuttavia osservato che le proprietà del materiale ottenuto mediante *additive manufacturing* possono differire da quelle del materiale massivo tradizionale; di conseguenza, anche sotto questo profilo, il caso analizzato mantiene carattere preliminare e semplificato. L'obiettivo non è infatti rappresentare in modo esaustivo il comportamento di un componente realmente producibile, quanto piuttosto valutare se la variazione di forma possa mostrare un potenziale interesse progettuale.

Le modifiche introdotte nel setup geometrico e nei materiali hanno reso inoltre necessario ricalibrare il carico termico applicato al sistema. La potenza dissipata è stata pertanto ridotta da 870 W a 600 W, così da ottenere livelli di temperatura ragionevoli anche nei casi limite della parametrizzazione, corrispondenti rispettivamente al cilindro a diametro costante di 1 mm e a quello di 7 mm. Tale scelta è stata effettuata con l'obiettivo di mantenere il problema entro un intervallo termico fisicamente plausibile. Va infatti osservato che, nel contesto applicativo reale, il vincolo di interesse riguarda tipicamente la temperatura massima di giunzione, mentre nel presente studio si considera come output la temperatura media sulla superficie riscaldata. Tuttavia, il modello adottato costituisce una semplificazione del sistema reale, poiché la potenza termica viene applicata direttamente alla superficie del dissipatore. Di conseguenza, i valori di temperatura ottenuti non vanno interpretati come una stima diretta della temperatura di giunzione, ma come un indicatore termico coerente con il modello e utile al confronto tra configurazioni differenti. In quest'ottica, la riduzione della potenza ha consentito di collocare i casi estremi della parametrizzazione in un intervallo di risposta ragionevole, evitando condizioni termiche eccessivamente lontane da quelle di possibile interesse applicativo. Con la potenza così ridefinita, i due casi limite forniscono temperature medie indicative comprese tra circa 128 °C e 194 °C. Sebbene tali valori non siano direttamente confrontabili con un limite di temperatura di giunzione, risultano comunque utili per collocare il problema in un intervallo termico sufficientemente ampio e informativo ai fini dell'addestramento del modello surrogato.

Successivamente, si è ritenuto opportuno ridurre anche l'altezza dei pin da 6 mm a 5.7 mm. Tale modifica è stata introdotta per evitare che i pin risultassero a contatto con il *cooler*. In presenza dello stesso materiale assegnato ai diversi componenti, si temeva infatti che SolidWorks potesse interpretare tale contatto come una continuità di materiale, introducendo un ponte di conduzione termica che non rappresenta correttamente il sistema reale. Nella configurazione fisica di interesse, infatti, pin e case non costituiscono un componente monolitico, e la loro interfaccia non può essere assimilata a una continuità perfetta di materiale. Diverso è invece il caso del contatto tra pin e

piastra, che rappresenta proprio il carattere tipico della soluzione *direct cooling* considerata, nella quale i pin costituiscono un'estensione geometrica della piastra stessa.

Infine, la Mesh locale applicata ai pin è stata incrementata da livello 3 a livello 4. A differenza di quanto fatto nel capitolo precedente, in questo caso non è stata svolta un'analisi di approfondimento altrettanto estesa sul livello di discretizzazione locale. La scelta è stata piuttosto guidata da una considerazione pratica: la sensibile riduzione del numero di pin rispetto al caso precedente rendeva plausibile valutare un incremento del raffinamento locale, particolarmente opportuno in un contesto in cui l'attenzione è rivolta proprio alla variazione della forma dei pin. Un infittimento maggiore della Mesh consente infatti di descrivere con più accuratezza geometrie che variano lungo l'altezza. Sono quindi state effettuate alcune prove preliminari per verificare se tale incremento fosse computazionalmente sostenibile. In particolare, su un assieme di simulazione nuovo, e quindi non ancora affetto dai rallentamenti dovuti alle numerose esecuzioni successive, il solo tempo di calcolo CFD è risultato pari a circa 42 s per il cilindro di diametro costante pari a 1 mm e 80 s per quello di 7 mm con Mesh locale di livello 4. Passando al livello 5, i tempi aumentano invece fino a circa 1 minuto per il cilindro da 1 mm e 3 minuti per quello da 7 mm. Considerando che tali valori si riferiscono esclusivamente alla simulazione CFD, e non includono né l'aggiornamento CAD né i rallentamenti progressivi osservati nel corso di campagne numerose, si è ritenuto opportuno adottare il livello 4 come compromesso ragionevole tra accuratezza e sostenibilità computazionale, senza spingersi fino al livello 5, che, pur mantenendo tempi ancora ragionevoli, avrebbe introdotto un aumento del costo computazionale ritenuto non conveniente per una campagna estesa di simulazioni.

In Tabella 11 sono riassunte le principali modifiche del setup CFD adottato in questo capitolo rispetto a quello introdotto in Tabella 2, Sezione 4.1.

Modifiche Setup CFD	
Materiali	Piastra e Pin-Fin: Alluminio
ϕ	600 W
Definizione dell'area riscaldata	Superficie ridotta sulla piastra (28 mm x 55 mm)
Mesh Locale sui Pin	Livello 4

Tabella 11: Modifiche setup CFD adottato nel presente capitolo

6.1.1 Parametrizzazione compatta

Per descrivere la forma dei pin è stata adottata la parametrizzazione proposta da Nguyen et al. ^[21] nell'articolo: "*Shape optimization of pin fin array in a cooling channel using genetic algorithm and machine learning*", basata sulla discretizzazione del profilo radiale lungo l'altezza. Tale parametrizzazione consente di descrivere forme assialsimmetriche anche molto diverse tra loro mediante un numero limitato di variabili, mantenendo al contempo un adeguato controllo sulla regolarità geometrica.

In particolare, la forma del pin è descritta assegnando il diametro in 7 sezioni di controllo equispaziate lungo l'altezza. Il vettore decisionale risulta quindi:

$$x = [d_1, d_2, \dots, d_7]$$

dove d_j rappresenta il diametro del pin nella j -esima sezione. Ciascuna variabile è discreta e può assumere valori interi compresi tra 1 mm e 7 mm:

$$1\text{ mm} \leq d_j \leq 7\text{ mm}$$

La discretizzazione dei diametri è stata introdotta per limitare la dimensione dello spazio progettuale: una formulazione continua avrebbe infatti ampliato il numero di configurazioni ammissibili, rendendo meno controllabile il problema. L'intervallo scelto consente di esplorare una famiglia sufficientemente ampia di geometrie, comprendente sia configurazioni molto snelle sia forme prossime al massimo riempimento compatibile con il numero e la disposizione fissata dei pin. In questo modo è possibile confrontare, all'interno della stessa parametrizzazione, un insieme piuttosto esteso di soluzioni, includendo anche cilindri a diametro costante con livelli di occupazione dello spazio sensibilmente differenti.

A partire dai sette diametri assegnati nelle sezioni di controllo, il profilo laterale del pin viene ricostruito mediante un'interpolazione cubica a tratti, in cui ciascun segmento della curva è definito dai valori del diametro agli estremi e dalle pendenze del profilo in tali punti (interpolazione di tipo Hermite). In particolare, è stata sviluppata in MATLAB una funzione dedicata che, dati i punti rappresentativi del profilo, costruisce una curva continua e regolare passante per tutti i punti assegnati.

La procedura si articola in tre passaggi principali. In primo luogo, i punti vengono ordinati rispetto alla coordinata verticale. Successivamente, viene utilizzata la funzione MATLAB “pchip”, che fornisce un'interpolazione cubica *shape-preserving*, utile per stimare le derivate nei punti di controllo limitando fenomeni di *overshoot*. Le derivate così ottenute vengono poi modificate imponendo esplicitamente derivata nulla al primo e all'ultimo punto del profilo, in modo da ottenere una tangente orizzontale sia alla base sia all'estremità libera del pin. Nel terzo passaggio, a partire dai valori del diametro e dalle derivate assegnate nei nodi, viene ricostruita una nuova interpolazione cubica Hermite a tratti mediante la funzione “mkpp”, che genera la rappresentazione polinomiale a tratti del profilo.

Questa costruzione consente di ottenere un profilo che soddisfa simultaneamente alcune proprietà desiderate:

- il passaggio esatto per i sette diametri assegnati;
- la regolarità C^1 del profilo interpolato;
- derivata nulla al primo e all'ultimo punto;
- un andamento localmente controllato, coerente con la logica *shape-preserving* di Pchip.

L'imposizione di tangente nulla agli estremi è stata adottata come criterio di regolarizzazione geometrica della famiglia di forme considerate: alla base del pin consente una transizione più dolce con la piastra, mentre all'estremità libera evita terminazioni con pendenze eccessivamente accentuate o arbitrarie, che avrebbero potuto generare forme meno regolari.

Tramite questa procedura il profilo ottenuto risulta sufficientemente liscio e geometricamente realistico, pur mantenendo la capacità di rappresentare forme anche molto differenti tra loro; inoltre stabilisce una corrispondenza univoca tra vettore decisionale e geometria del pin.

In Figura 66 è riportata una rappresentazione grafica della procedura di costruzione del profilo del pin, che comprende sia la parametrizzazione mediante i sette diametri di controllo sia la successiva interpolazione della curva. Si osservi inoltre che il diametro d_1 corrisponde alla base del pin, in corrispondenza del collegamento con la piastra, mentre d_7 rappresenta la sezione terminale libera superiore.

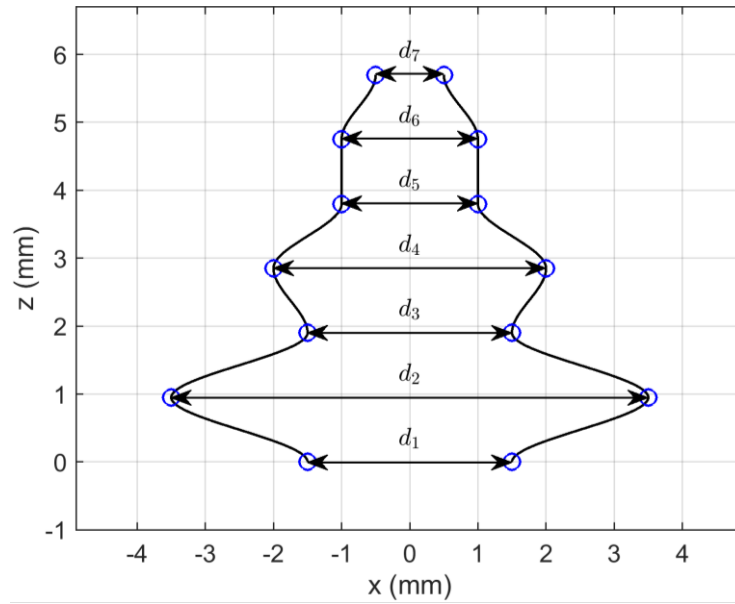


Figura 66: Rappresentazione della procedura di costruzione del profilo del pin

La forma tridimensionale finale del pin si ottiene quindi come superficie di rivoluzione del profilo così definito attorno al proprio asse verticale.

Infine, in Figura 67 sono riportati alcuni esempi di forme ottenibili mediante la procedura descritta, con lo scopo di rendere più concreto il contenuto del presente paragrafo e fornire una rappresentazione visiva della varietà geometrica. La figura non ha pretesa esaustiva, anche perché il numero complessivo di configurazioni generabili risulta estremamente elevato ($7^7 = 823\,543$, ognuna delle 7 variabili può assumere 7 valori discreti).

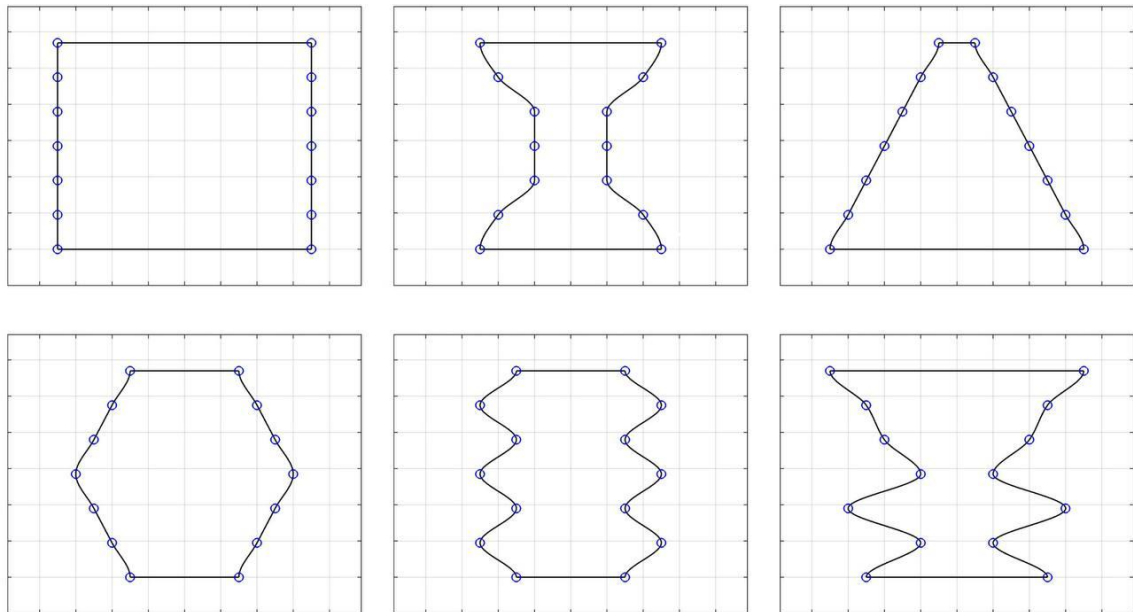


Figura 67: Esempi di forme dei pin ottenibili

6.1.2 Preparazione del modello CAD per la generazione del dataset

Il passaggio dall'ottimizzazione della distribuzione spaziale dei pin, affrontata nel Capitolo 5, all'ottimizzazione della loro forma ha reso necessario lo sviluppo di nuove procedure automatiche per la generazione CAD.

Il punto di partenza è un assieme SolidWorks contenente pin cilindrici standard posizionati nelle sedi prefissate sulla griglia mediante la stessa automazione di inserimento già adottata nei capitoli precedenti, che garantisce un posizionamento preciso e riproducibile dei componenti, indipendente da errori manuali. Su tali componenti sono state definite preventivamente le Mesh locali, secondo lo stesso principio discusso nello scorso capitolo.

La scelta di utilizzare sempre lo stesso assieme base ha permesso di sfruttare una caratteristica utile dell'ambiente CAD: sovrascrivendo con lo stesso nome il file di un componente già presente nell'assieme, SolidWorks ne aggiorna automaticamente la geometria mantenendo le relazioni di assemblaggio e, soprattutto, le impostazioni di Mesh locale associate. In questo modo è stato possibile sostituire automaticamente i pin cilindrici con nuovi pin a profilo variabile senza dover ridefinire ogni volta l'intero modello.

Per generare automaticamente i pin in SolidWorks è stata sviluppata una procedura dedicata di trasferimento della geometria da MATLAB all'ambiente CAD. Una volta costruita in MATLAB la curva continua del profilo del pin secondo la procedura descritta nel paragrafo precedente, essa viene campionata mediante 1000 punti equispaziati lungo l'altezza. Il numero di punti è stato scelto in modo empirico, verificando che risultasse sufficiente a ricostruire in SolidWorks un profilo visivamente liscio e privo di approssimazioni evidenti. Questa scelta è legata al metodo utilizzato per il trasferimento della geometria: SolidWorks consente infatti di ricostruire una curva a partire da una sequenza di punti tridimensionali (x, y, z) . Le coordinate dei punti campionati vengono quindi esportate da MATLAB e importate in SolidWorks, dove sono utilizzate per generare la curva guida del profilo. A partire da tale curva, il software costruisce poi il corrispondente solido di rivoluzione attorno all'asse assegnato. L'intera procedura viene eseguita automaticamente per tutte le configurazioni appartenenti alla campagna numerica.

In Figura 68 è mostrato un esempio di ricostruzione del profilo del pin in SolidWorks a partire dalla curva generata in MATLAB, mentre in Figura 69 è illustrata la successiva sostituzione automatica del componente all'interno dell'assieme di simulazione. In Figura 69, nella vista in alto a destra, i pin sono stati rappresentati in trasparenza per evidenziare la variazione di sezione lungo l'altezza; per questo motivo risultano visibili più circonferenze sovrapposte.

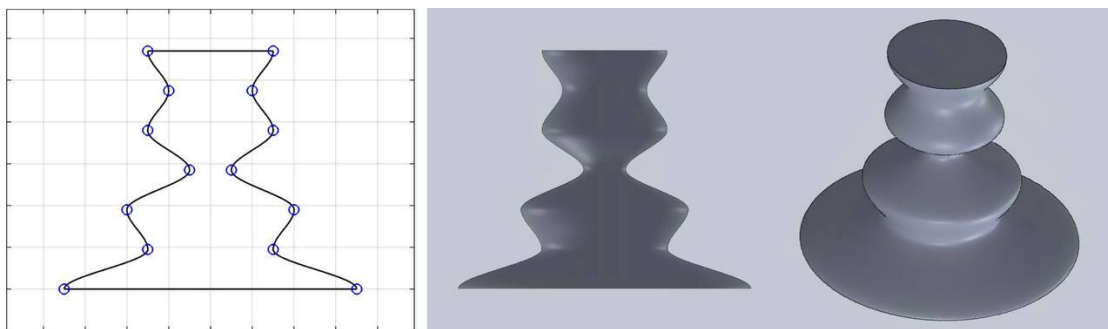


Figura 68: Esempio di ricostruzione automatica in SolidWorks del profilo del pin generato in MATLAB

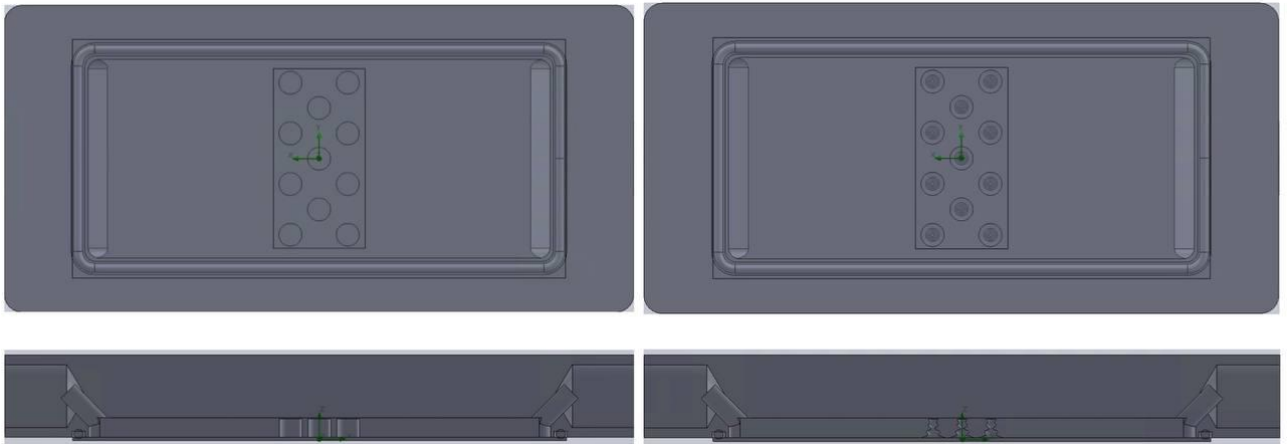


Figura 69: Esempio di sostituzione automatica del pin nell'assieme di simulazione. Il componente generato in MATLAB e ricostruito in SolidWorks sostituisce il corrispondente pin cilindrico standard mantenendo invariati posizionamento, relazioni di assemblaggio e impostazioni di Mesh locale

L'utilizzo ripetuto dello stesso assieme per un numero elevato di simulazioni ha però mostrato un progressivo decadimento delle prestazioni, con un aumento della pesantezza del modello in termini di memoria e tempi di gestione. Tale comportamento è stato osservato soprattutto dopo numerosi salvataggi e successive esecuzioni di simulazioni, ed è verosimilmente riconducibile all'accumulo di informazioni interne associate all'assieme e agli studi CFD. Per limitare questo effetto, tra una campagna e l'altra l'assieme operativo veniva sostituito con una copia "pulita", preparata prima dell'avvio delle simulazioni. Inoltre, al termine di ogni simulazione, è stata implementata la cancellazione automatica dei file generati da Flow Simulation, così da evitare l'accumulo di grandi quantità di dati sul disco. Tale accorgimento si è rivelato importante sia per contenere l'occupazione di memoria, sia per preservare la fluidità dell'intero *workflow* nelle campagne più lunghe.

6.1.3 Generazione del dataset

La procedura di campionamento adottata è la medesima già descritta nella Sezione 5.2 e non viene quindi qui ripetuta nel dettaglio. Le sole differenze riguardano il numero di variabili decisionali, che in questo caso è pari a 7 anziché 9, e i relativi limiti di variazione, compresi tra 1 e 7 anziché tra 0 e 14. Il codice è stato pertanto adattato a queste modifiche, lasciando invariata l'impostazione generale della procedura. Naturalmente, rispetto al capitolo precedente cambia invece la procedura di aggiornamento CAD associata a ciascuna configurazione campionata: in questo caso non viene modificata la distribuzione spaziale dei pin, ma la forma di tutti i pin presenti nell'assieme, come descritto nei paragrafi precedenti.

Anche in questo capitolo il dataset finale utilizzato per l'addestramento del modello surrogato è costituito da 1024 simulazioni CFD. La scelta è stata effettuata in modo progressivo, confrontando dataset di diversa numerosità (256, 512 e 1024) e valutandone l'effetto sia sull'accuratezza predittiva della rete sia sulla qualità dei fronti di Pareto ottenuti. Da tale confronto è emerso che il dataset da 1024 configurazioni rappresentava il miglior compromesso tra qualità dei risultati e costo computazionale, senza rendere necessario un ulteriore ampliamento della campagna CFD. Anche in questo caso, il tempo complessivo richiesto per la generazione del dataset è stato stimato a partire da un piccolo lotto di 10 configurazioni, considerando congiuntamente sia l'aggiornamento geometrico sia l'esecuzione della simulazione CFD. Il tempo totale misurato per tale lotto è risultato pari a circa 965,9 s, corrispondenti in media a circa 96,6 s per configurazione. Assumendo questo valore medio come riferimento indicativo, la generazione del dataset finale di 1024 configurazioni richiede un tempo complessivo stimato pari a circa 27 ore.

6.2 Modello surrogato

Anche per il presente problema è stato adottato il medesimo modello surrogato già descritto nella Sezione 5.3. La differenza principale rispetto al caso precedente riguarda la natura delle variabili di ingresso: in questo capitolo, la rete neurale riceve infatti in ingresso i sette diametri corrispondenti alle sezioni di controllo del pin e restituisce in uscita la temperatura media predetta $\hat{T}_{av}$, mentre nel problema trattato nel Capitolo 5 gli input erano costituiti dai nove valori associati al numero di pin nelle diverse zone della superficie riscaldata. Per tutti gli altri aspetti, l'impostazione è stata mantenuta invariata. In particolare, sono stati lasciati immutati la suddivisione del dataset e l'architettura della rete, poiché anche nel presente caso tali scelte hanno fornito prestazioni soddisfacenti.

6.2.1 Analisi preliminari di correlazione

Seguendo la stessa linea descrittiva già adottata nel Capitolo 5, prima dell'addestramento della rete neurale è stata svolta un'analisi preliminare delle relazioni tra variabili di input e output. In particolare, sono stati calcolati i coefficienti di correlazione di Pearson e di Spearman sia tra i sette diametri di controllo e la temperatura media T_{av} , sia tra gli stessi diametri e il volume del pin. I risultati sono riportati in Figura 70.

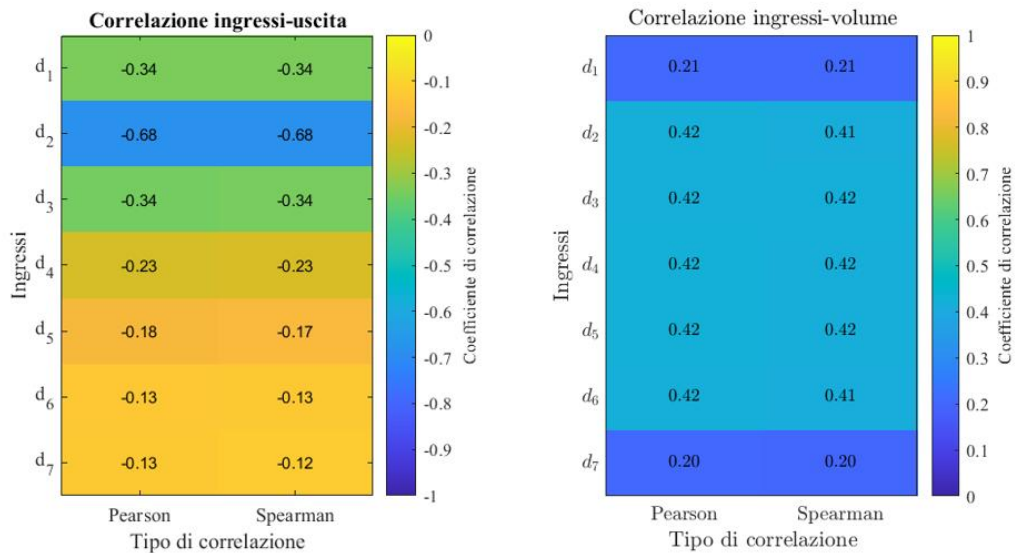


Figura 70: Correlazione input–output (a sinistra) e input–volume (a destra) tra i sette diametri di controllo del pin, il target termico e il volume del pin

Per quanto riguarda il legame tra ingressi e target termico, tutti i coefficienti risultano negativi, in accordo con l'intuizione fisica del problema: un aumento dei diametri tende infatti, in generale, ad accrescere il volume del pin e quindi la quantità di materiale disponibile per la conduzione, con conseguente riduzione della temperatura media sulla superficie riscaldata. L'associazione negativa più marcata si osserva per d_2 (Pearson e Spearman circa pari a -0.68), seguito da d_1 e d_3 , che mostrano correlazioni ancora apprezzabili ma più moderate (circa -0.34). Per i diametri successivi l'influenza tende progressivamente a ridursi, fino a valori più deboli per d_6 e d_7 .

Diversamente dal caso del Capitolo 5, tuttavia, l'interpretazione di queste correlazioni non può essere condotta considerando il solo legame input– T_{av} . In questo problema, infatti, ciascun diametro non contribuisce nello stesso modo alla quantità totale di materiale: i sette parametri definiscono sezioni poste a quote differenti lungo l'altezza del pin e la loro variazione non incide

uniformemente sul volume complessivo. Per questo motivo, la lettura delle correlazioni con il target termico va affiancata a quella delle correlazioni input-volume, riportate anch'esse in Figura 70.

Si osserva infatti che d_2 , d_3 , d_4 , d_5 e d_6 mostrano tutti una correlazione positiva con il volume prossima a 0.42, mentre d_1 e d_7 risultano sensibilmente meno influenti, con coefficienti intorno a 0.20–0.21. Questo comportamento è coerente con la struttura della parametrizzazione: i diametri estremi agiscono infatti su porzioni più ridotte della geometria, poiché definiscono rispettivamente la base e la sommità del pin, mentre i diametri intermedi influenzano tratti più estesi del profilo e quindi incidono maggiormente sul volume totale.

Alla luce di ciò, le correlazioni input- T_{av} devono essere interpretate in modo duplice. Da un lato, esse riflettono l'effetto che ciascun diametro esercita sulla forma del pin; dall'altro, incorporano anche il diverso peso con cui ciascuna variabile contribuisce alla variazione del volume, e quindi della quantità di materiale disponibile. In altri termini, una maggiore influenza di una certa variabile sull'output termico non implica necessariamente che quella sezione sia "più importante" in senso puramente geometrico, ma può dipendere anche dal fatto che la sua variazione comporti un incremento più marcato del volume complessivo.

In questa prospettiva, le correlazioni osservate appaiono nel complesso coerenti con l'intuizione fisica del problema. In generale, è infatti ragionevole attendersi che la presenza di materiale nelle sezioni più vicine alla base del pin risulti particolarmente favorevole, poiché contribuisce a costruire un percorso di conduzione del calore dalla superficie riscaldata verso le porzioni superiori del pin. Da questo punto di vista, il fatto che i primi diametri risultino i più influenti appare pienamente plausibile. Il confronto tra d_1 e d_2 è particolarmente istruttivo: sebbene d_1 , essendo alla base, sia intuitivamente una variabile molto rilevante, esso mostra una correlazione con T_{av} meno marcata di d_2 . Tale risultato può essere letto proprio alla luce del diverso contributo al volume complessivo: d_1 rimane certamente influente, ma la sua azione risulta in parte attenuata dal fatto che, nella parametrizzazione adottata, esso modifica una porzione geometrica più ridotta rispetto a d_2 . Più interessante è invece il comportamento di d_7 , che mostra una correlazione con T_{av} comparabile a quella di d_6 , pur essendo associato alla sezione più distante dalla base riscaldata e incidendo meno sul volume totale. Ci si potrebbe infatti attendere che, procedendo verso l'estremità del pin, l'influenza dei diametri si attenui, sia per la maggiore distanza dalla sorgente termica sia per il minore contributo volumetrico associato alla parametrizzazione. Il fatto che ciò non avvenga in modo evidente nel passaggio da d_6 a d_7 suggerisce quindi che, nella zona terminale del pin, entri in gioco anche un ulteriore effetto. Una possibile interpretazione è che d_7 , agendo direttamente sulla sezione terminale libera, influenzi in misura relativamente maggiore la superficie complessiva disponibile per lo scambio termico convettivo del pin, costituita dalla superficie laterale e da quella superiore. Tale effetto è plausibilmente accentuato anche dal fatto che, nel presente problema, l'altezza del pin è dello stesso ordine di grandezza dei diametri considerati; in queste condizioni, la superficie terminale superiore non risulta trascurabile rispetto a quella laterale, e un aumento di d_7 può quindi incidere significativamente sull'area totale di scambio. In questo modo, l'aumento di tale area potrebbe compensare almeno in parte la minore efficacia attesa dovuta alle considerazioni appena discusse.

Per approfondire questo aspetto, si è ritenuto opportuno valutare anche la correlazione tra i sette diametri di controllo e la superficie di scambio termico del pin, riportata in Figura 71. Questa grandezza è stata ricavata numericamente a partire dal profilo del pin, mediante una procedura del tutto analoga a quella adottata per il volume, il cui calcolo verrà descritto nel seguito.

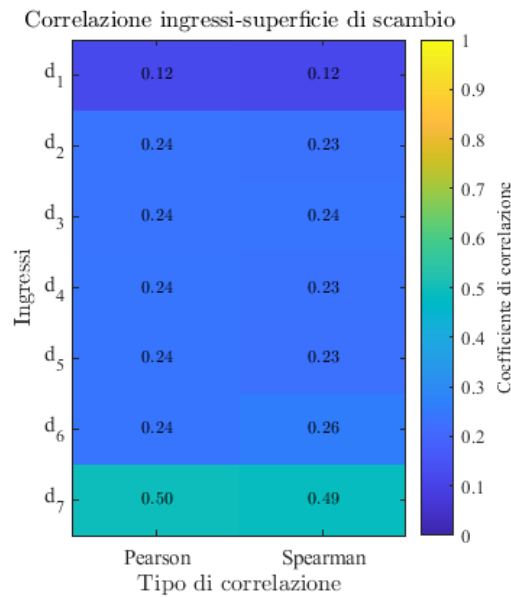


Figura 71: Correlazione input-superficie di scambio tra i sette diametri di controllo del pin e la superficie di scambio termico convettivo

L'analisi conferma che d_7 è la variabile che mostra la correlazione più elevata con la superficie di scambio, con coefficienti prossimi a 0.50, nettamente superiori a quelli degli altri diametri. Questo risultato rafforza l'interpretazione proposta in precedenza: pur incidendo meno sul volume totale, d_7 esercita un effetto particolarmente marcato sulla superficie disponibile per la convezione, proprio perché agisce direttamente sulla sezione terminale libera e quindi anche sulla superficie superiore del pin. Al contrario, d_1 mostra un'influenza più contenuta anche da questo punto di vista. Pur interessando, come d_7 , una porzione geometrica relativamente ridotta del profilo, esso agisce sulla sezione di base del pin, la cui superficie non contribuisce alla superficie di scambio termico convettivo perché integrata nella piastra. Di conseguenza, il suo effetto si manifesta unicamente attraverso la variazione della superficie laterale in prossimità della base.

Come nel capitolo precedente, va comunque sottolineato che si tratta di un'analisi soltanto bivariata, utile per una prima lettura qualitativa del problema ma non sufficiente a esaurire il ruolo delle variabili nel modello. I coefficienti di correlazione mettono infatti in relazione ciascun diametro singolarmente con l'output, senza tenere conto delle interazioni tra le variabili. Questo limite è particolarmente importante nel presente caso, poiché la forma del pin è determinata congiuntamente da tutti e sette i diametri e il contributo di ciascuno di essi può essere compreso pienamente solo nel contesto della geometria complessiva. A ciò si aggiunge un ulteriore aspetto fondamentale, difficilmente discutibile in termini puramente qualitativi: l'influenza della forma sul moto del fluido. La geometria del pin, infatti, non incide soltanto sulla conduzione interna e sulla superficie disponibile per lo scambio, ma anche sulla distribuzione locale delle velocità e quindi sulle modalità con cui il fluido interagisce con le diverse porzioni della superficie. Tale aspetto è reso ancora più complesso dal fatto che, nelle simulazioni considerate, il pin non è analizzato come elemento isolato, ma come parte di una schiera: di conseguenza, le modifiche geometriche introdotte sui singoli diametri non influenzano solo il comportamento fluidodinamico attorno al singolo pin, ma anche l'interazione tra pin adiacenti e la distribuzione complessiva del flusso. Ne consegue che l'effetto dei singoli diametri non può essere interpretato in modo isolato, ma dipende anche dalle modifiche che essi inducono nel comportamento fluidodinamico complessivo. Per questi motivi, tutte e sette le variabili vengono mantenute come ingressi della rete neurale.

Accanto all'analisi appena svolta, è stata inoltre verificata la struttura di correlazione tra le stesse variabili di input, così da controllare che il dataset fornisse alla rete neurale ingressi non caratterizzati da dipendenze reciproche marcate. Anche in questo caso, il controllo è stato effettuato mediante le matrici di correlazione di Pearson e di Spearman, riportate in Figura 72.

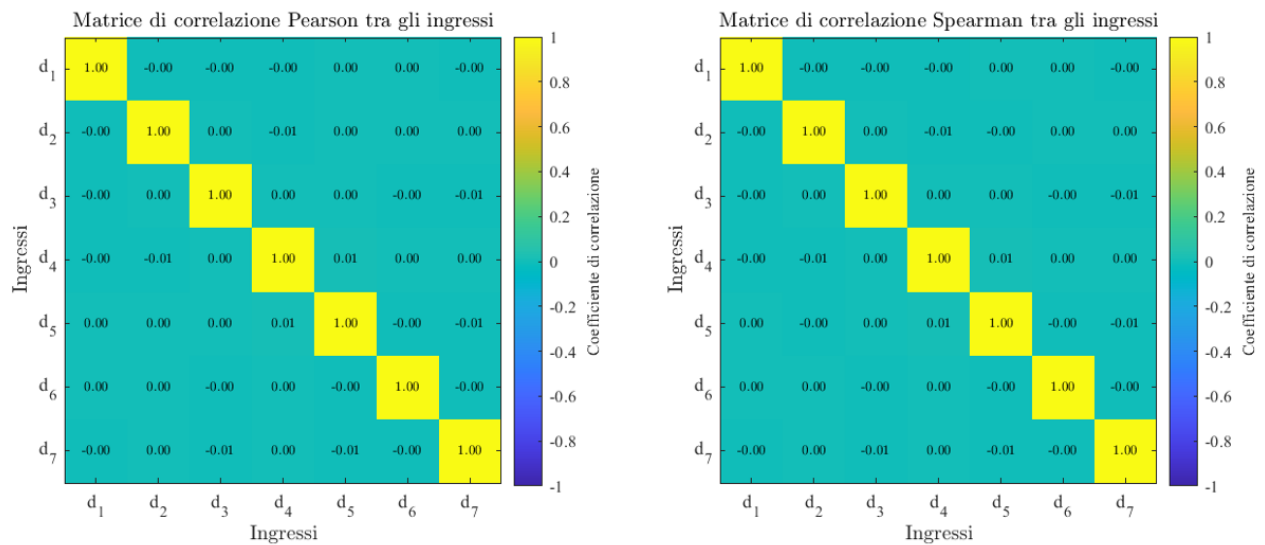


Figura 72: Matrici di correlazione tra le sette variabili di input

Come già osservato nel Capitolo 5, l'interesse principale non risiede naturalmente nei termini diagonali, che valgono unitariamente per costruzione, bensì nei coefficienti fuori diagonale. In entrambe le matrici, tali coefficienti risultano molto prossimi a zero, evidenziando l'assenza di correlazioni lineari o monotone significative tra i diversi diametri di controllo.

6.2.2 Valutazione del modello surrogato

Il comportamento della procedura di addestramento è illustrato in Figura 73, che riporta l'andamento del MSE sui set di *training*, *validation* e *test* al variare delle epoche.

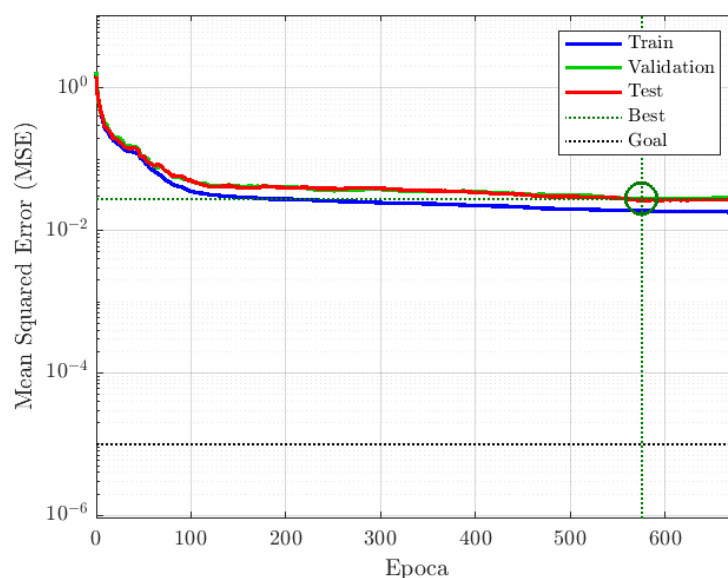


Figura 73: Andamento del MSE sui set di training, validation e test durante l'addestramento della rete neurale

Si osserva che la miglior prestazione sul set di validazione si ottiene all'epoca 575, con un valore di MSE normalizzato di circa 0,028, mentre l'arresto del *training* avviene in un'epoca successiva (675). Come nel caso precedente, la rete assunta come modello finale è quella corrispondente alla miglior epoca, associata alla migliore capacità di generalizzazione stimata sul *validation* set. Anche in questo caso, nella regione di minimo le curve di *validation* e *test* risultano molto vicine tra loro. L'andamento delle curve nelle epoche successive alla migliore mostra inoltre che, mentre l'errore sul *training* set continua a diminuire, quelli di *validation* e *test* tendono a peggiorare leggermente. Anche in questo caso il comportamento è coerente con l'insorgenza di un moderato fenomeno di *overfitting*, che viene efficacemente contenuto adottando come criterio di selezione del modello la miglior prestazione sul *validation* set. Come già osservato nel capitolo precedente, i valori di MSE riportati in Figura 73 sono riferiti ai dati normalizzati. Per una lettura più immediata dell'accuratezza del modello nella scala fisica del problema, sono stati inoltre calcolati gli errori sui dati denormalizzati. In termini di MSE, essi risultano pari a 1.89 sul *training* set, 3.07 sul *validation* set e 2.90 sul *test* set, corrispondenti a valori di RMSE rispettivamente pari a circa 1.37 °C, 1.75 °C e 1.70 °C. Per i motivi già discussi nel Capitolo 5, il fatto che le differenze tra questi numeri siano molto contenute, costituisce un'indicazione favorevole della buona capacità di generalizzazione del modello.

Il tempo complessivo di addestramento è stato infine pari a circa 3 s.

Completato l'addestramento, il modello surrogato è stato valutato sul *test* set. Il primo riscontro è fornito dal grafico di parità riportato in Figura 74, nel quale le predizioni della rete vengono confrontate con i corrispondenti valori ottenuti dalle simulazioni CFD.

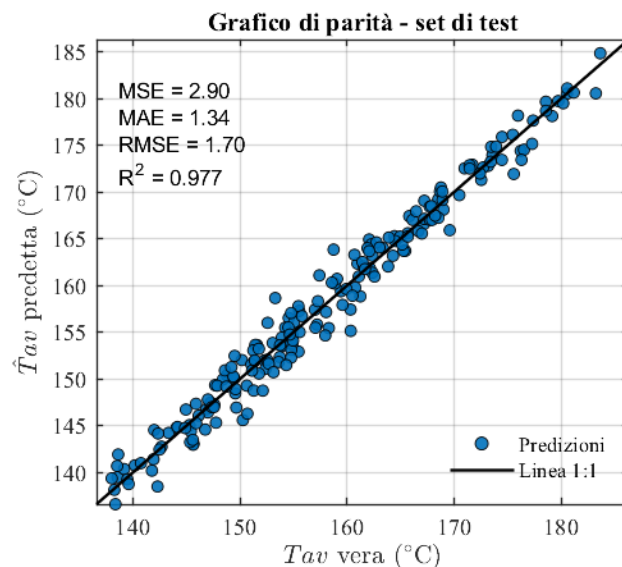


Figura 74: Grafico di parità sul set di test tra i valori di temperatura media predetti dal modello surrogato e quelli ottenuti dalle simulazioni CFD

I punti risultano distribuiti in prossimità della retta ideale 1:1 lungo l'intero intervallo dei valori considerati, senza che emergano scostamenti sistematici evidenti. Questo comportamento indica che il modello è in grado di riprodurre con buona fedeltà la relazione ingresso-uscita appresa dal dataset, mantenendo prestazioni soddisfacenti anche su campioni non utilizzati nella fase di *training*. Le metriche riportate in figura confermano tale impressione qualitativa. In particolare, sul *test* set si ottengono $MSE = 2.90$, $MAE = 1.34$ °C, $RMSE = 1.70$ °C e $R^2 = 0.977$. Il valore di MAE indica che, mediamente, il modello commette un errore di poco superiore a 1.3 °C nella stima di T_{av} , mentre l'RMSE, solo moderatamente più elevato, suggerisce l'assenza di una coda particolarmente

pesante di errori elevati. Anche il coefficiente di determinazione, prossimo a 0.98, conferma l'elevato grado di accordo tra i valori predetti e quelli ottenuti mediante CFD.

Anche nel presente capitolo per integrare questa valutazione con una misura più direttamente interpretabile in termini ingegneristici, è stato analizzato l'errore relativo sul *test set*, la cui distribuzione è riportata in Figura 75.

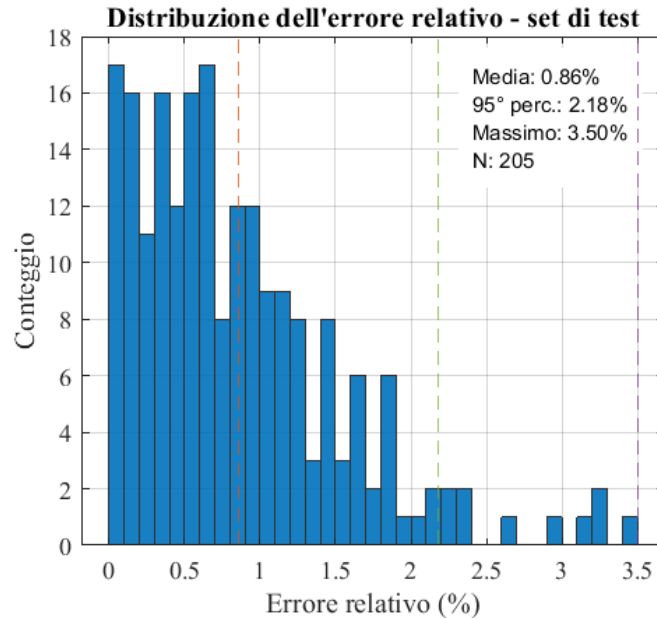


Figura 75: Distribuzione dell'errore relativo sul set di test del modello surrogato

Si osserva che la maggior parte dei campioni presenta errori contenuti e concentrati su valori bassi, mentre solo un numero limitato di configurazioni si colloca nella coda a errore più elevato. In particolare, l'errore relativo medio sul *test set* risulta pari a 0.86%, il 95° percentile è pari a 2.18%, mentre il valore massimo osservato è pari a 3.50%, su un totale di 205 campioni di *test*. Nel complesso, tali risultati indicano che il modello surrogato fornisce una stima di T_{av} sufficientemente accurata per guidare l'esplorazione del dominio di progetto. Rispetto al Capitolo 5 (paragrafo 5.3.4), gli errori risultano inoltre leggermente più contenuti; tale risultato appare ragionevole, poiché in entrambi i casi il modello è stato addestrato su un dataset di 1024 simulazioni, ma nel presente problema lo spazio di progetto è di dimensione inferiore.

Anche in questo capitolo, il modello surrogato non è naturalmente inteso come sostituto definitivo delle simulazioni CFD.

6.3 Ottimizzazione multi-obiettivo assistita da surrogato

Come anticipato nell'introduzione del capitolo, nella presente sezione il problema viene formulato come ottimizzazione multi-obiettivo, assumendo come variabili decisionali i sette diametri che definiscono la forma del pin: $x = [d_1, d_2, \dots, d_7]$.

Le due funzioni obiettivo considerate sono:

- la minimizzazione della temperatura media sulla superficie riscaldata predetta dal surrogato

$$f_1(x) = \hat{T}_{av}(x)$$

- la minimizzazione del volume del singolo pin

$$f_2(x) = V_{pin}(x)$$

Il primo obiettivo mira a migliorare le prestazioni termiche del dissipatore, mentre il secondo consente di limitare la quantità di materiale impiegato e quindi, indirettamente, il peso del sistema. Anche in questo caso non esiste un'unica soluzione ottima in senso assoluto, poiché una riduzione di T_{av} richiede in genere un aumento di materiale, mentre configurazioni più leggere tendono a penalizzare le prestazioni termiche.

Una volta addestrato il modello surrogato, esso viene impiegato per l'esplorazione dell'intero spazio progettuale. In questo caso il numero totale di combinazioni, pur essendo elevato, è sufficientemente contenuto da consentire una valutazione completa mediante il modello neurale. Le $7^7 = 823\,543$ configurazioni possibili vengono quindi analizzate integralmente per quanto riguarda la previsione della temperatura media $\hat{T}_{av}$.

Il secondo obiettivo, rappresentato dal volume del singolo pin, non richiede simulazioni CFD né una stima della rete neurale, ma viene calcolato direttamente in MATLAB a partire dalla geometria del profilo. Poiché ciascun pin è descritto come superficie di rivoluzione, il volume può essere ottenuto mediante integrazione numerica del profilo radiale lungo l'altezza. A tale scopo, la curva del profilo viene campionata in 10 000 punti equispaziati, e il volume viene calcolato con il metodo dei trapezi.

Anche questa operazione risulta sufficientemente rapida da poter essere eseguita in modo esaustivo su tutte le configurazioni ammissibili. Per questo motivo, a differenza di quanto fatto nel Capitolo 5, in questo caso è stato possibile valutare direttamente l'intero spazio discreto delle soluzioni, senza ricorrere a un algoritmo genetico per guidarne l'esplorazione.

L'individuazione del fronte di Pareto è stata effettuata ordinando le configurazioni in funzione del volume e selezionando, per ciascun valore di V_{pin} , la soluzione con temperatura predetta minima. Poiché la parametrizzazione adottata consente di ottenere un numero molto elevato di valori di volume distinti, compresi tra il minimo corrispondente al cilindro di diametro costante pari a 1 mm ($V_{pin} \approx 4.48 \text{ mm}^3$) e il massimo associato al cilindro di diametro costante pari a 7 mm ($V_{pin} \approx 219.36 \text{ mm}^3$), è stata introdotta una tolleranza pari a 10^{-6} mm^3 , così da considerare distinti due volumi solo se differiscono almeno di tale quantità. Una volta effettuato questo raggruppamento, il fronte è stato costruito mantenendo soltanto le configurazioni che, all'aumentare del volume, introducono una riduzione di $\hat{T}_{av}$ rispetto a tutte le soluzioni a volume minore. In questo modo si ottiene il corrispondente insieme di compromessi non dominati nel piano obiettivi $(V_{pin}, \hat{T}_{av})$. L'intera fase di ottimizzazione ha richiesto complessivamente circa 187 s, corrispondenti a poco più di 3 minuti. I risultati di tale procedura vengono discussi nel paragrafo successivo.

6.3.1 Risultati dell'ottimizzazione e validazione CFD ad alta fedeltà

Il confronto tra l'intero dataset CFD, il relativo fronte di Pareto, il fronte ottenuto dall'esplorazione esaustiva mediante surrogato e la successiva validazione CFD delle configurazioni ottimizzate è riportato in Figura 76.

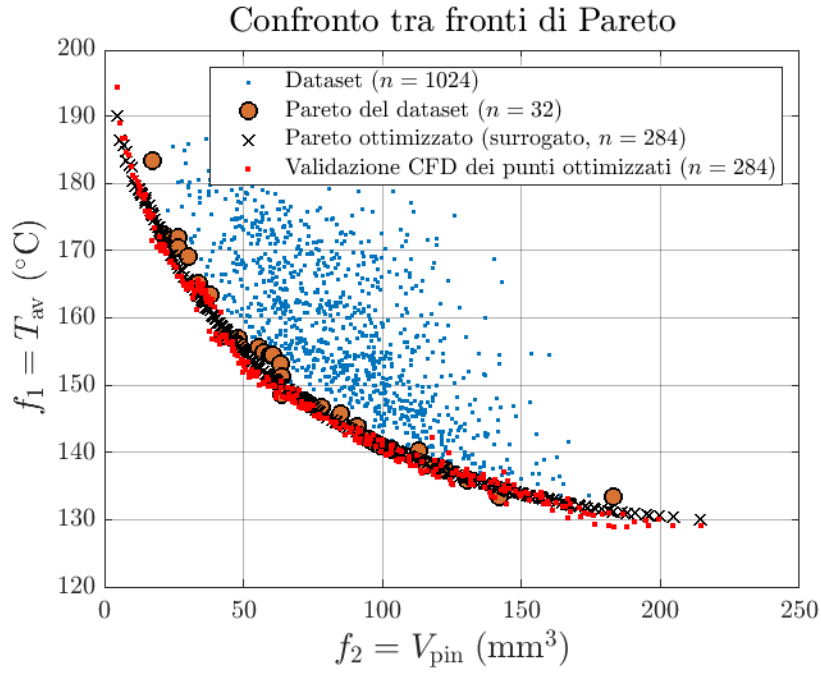


Figura 76: Confronto nel piano obiettivi (f_2, f_1) tra l'intero dataset CFD, il relativo fronte di Pareto, il fronte di Pareto globale ottenuto mediante ottimizzazione sul modello surrogato e la successiva validazione CFD dei punti ottimizzati

Dal confronto emerge che il fronte ottenuto mediante il modello surrogato non si discosta drasticamente dai migliori compromessi già presenti nel dataset iniziale, ma ne fornisce una rappresentazione molto più fitta e continua. In particolare, il dataset di 1024 simulazioni contiene 32 configurazioni non dominate, mentre l'esplorazione completa dello spazio mediante surrogato porta all'individuazione di 284 configurazioni ottimizzate (previa validazione CFD). Questo risultato suggerisce che, in uno spazio progettuale relativamente contenuto e in presenza di una buona procedura di campionamento, il dataset iniziale sia già in grado di intercettare configurazioni prossime al fronte efficiente, mentre il surrogato consente soprattutto di infittire il fronte, descriverne con maggiore continuità l'andamento e aumentare la confidenza nella sua forma complessiva, grazie alla valutazione dell'intero spazio discreto delle soluzioni ammissibili. Si osserva inoltre che alcune configurazioni del dataset sembrano, a prima vista, collocarsi in posizione leggermente migliore rispetto a parte del fronte predetto dal surrogato. Tale aspetto è tuttavia riconducibile agli errori di predizione del modello neurale e non contraddice la bontà complessiva della procedura: dopo validazione CFD, infatti, le configurazioni ottimizzate risultano coerentemente collocate al di sotto del Pareto del dataset, confermando che il fronte ottenuto mediante surrogato costituisce un affinamento fisicamente plausibile dei migliori compromessi già individuati nel campione iniziale.

Tutte le 284 soluzioni del fronte predetto sono state sottoposte a simulazione CFD. Il confronto tra predizioni surrogate e validazione finale è riassunto in Figura 77, che riporta sia la distribuzione dell'errore relativo sui punti ottimizzati sia l'andamento dello stesso errore in funzione del volume V_{pin} . Questa fase ha richiesto complessivamente un costo computazionale di circa 9 ore di calcolo.

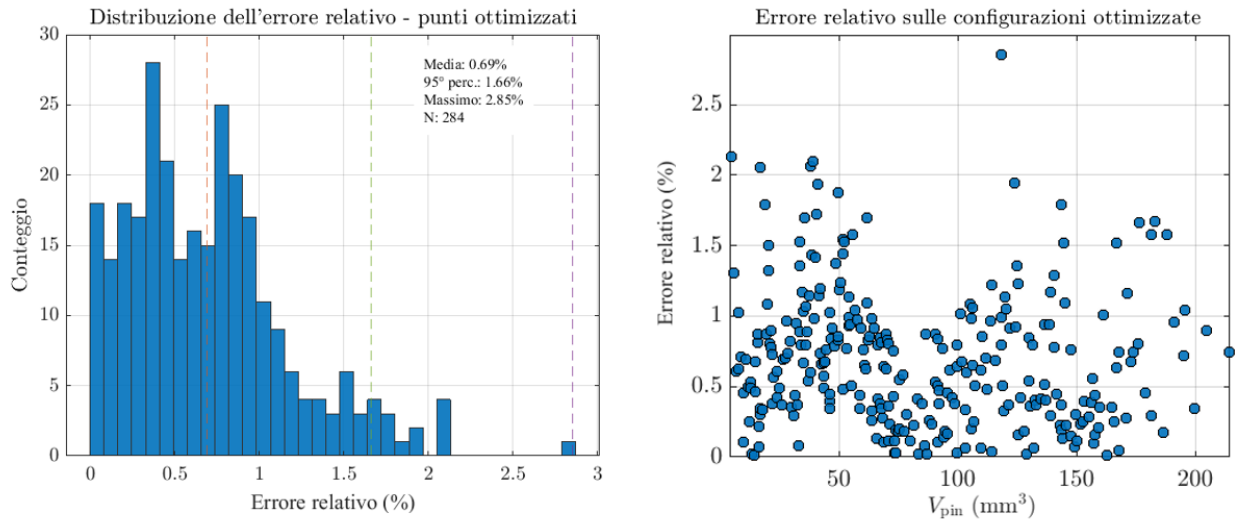


Figura 77: Distribuzione dell'errore relativo sui 284 punti ottimizzati validati mediante CFD (a sinistra) e andamento dello stesso errore in funzione del volume V_{pin} (a destra)

Il confronto tra predizioni e valori CFD evidenzia un buon accordo complessivo. In particolare, l'errore relativo sui punti ottimizzati risulta pari in media a 0.69%, con 95° percentile pari a 1.66% e massimo pari a 2.85%, su un totale di 284 configurazioni validate. Tali valori risultano non solo contenuti in senso assoluto, ma anche leggermente inferiori a quelli osservati sul *test set* casuale del modello (paragrafo 6.2.2). Sebbene la differenza sia modesta, essa risulta coerente con quanto già osservato in precedenza, ossia con il fatto che il dataset iniziale appaia già intercettare configurazioni prossime al fronte efficiente. Ciò contribuisce infatti a spiegare perché, nella regione occupata dalle soluzioni ottimizzate, l'accuratezza del modello risulti leggermente migliore rispetto a quella osservata su un insieme casuale di *test*.

Un ulteriore elemento di interesse emerge confrontando l'andamento dell'errore relativo nel presente capitolo con quello osservato nel Capitolo 5 (paragrafo 5.4.2). In quel caso l'errore mostrava una struttura ben riconoscibile lungo il fronte; nel problema attuale, invece, non si osserva un andamento sistematico altrettanto marcato in funzione di V_{pin} , e l'errore risulta distribuito in modo più omogeneo lungo l'intervallo dei volumi analizzati.

Per interpretare tale aspetto si può fare riferimento alla Figura 78, nella quale sono riportate, per diversi livelli di raggruppamento, sia la distribuzione del numero di configurazioni possibili in funzione del volume, sia quella effettivamente osservata nel dataset. Poiché V_{pin} non è una variabile discreta elementare ma una grandezza geometrica derivata, è necessario introdurre classi di volume: in assenza di tale raggruppamento, molti valori esatti di volume risulterebbero associati a pochissime configurazioni, rendendo la distribuzione poco leggibile. Le classi più fini (ad esempio 1 mm^3 lato sinistro di Figura 78) permettono di cogliere meglio la struttura locale, mentre quelle più ampie (ad esempio 5 mm^3 lato destro Figura 78) restituiscono in modo più chiaro l'andamento globale. Va inoltre ricordato che il campionamento non è stato effettuato in funzione del volume V_{pin} , bensì nello spazio delle sette variabili di input $[d_1, \dots, d_7]$. Analogamente, anche il modello surrogato non predice T_{av} a partire dal volume, ma dall'intero vettore dei diametri di controllo. Le considerazioni svolte rispetto a V_{pin} non vanno quindi intese come un giudizio sulla bontà del campionamento, ma soltanto come una chiave interpretativa utile per leggere l'andamento degli errori lungo il fronte.

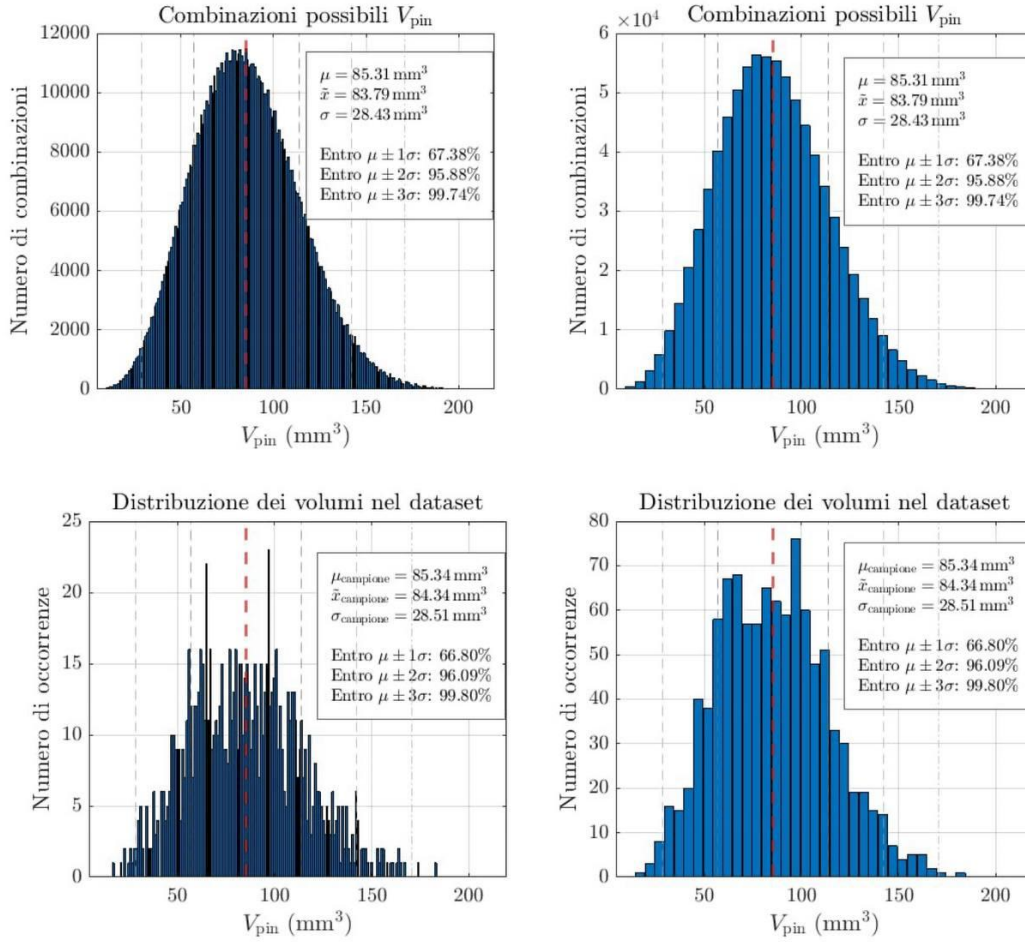


Figura 78: Numero di configurazioni possibili in funzione del volume V_{pin} nell'intero spazio di progetto (in alto) e distribuzione dei volumi nel dataset di 1024 configurazioni (in basso). A sinistra i volumi sono raggruppati in classi di 1 mm^3 , mentre a destra in classi di 5 mm^3

Anche nel presente caso la distribuzione teorica delle configurazioni possibili in funzione di V_{pin} presenta un andamento complessivamente campanulare; tuttavia, diversamente da quanto osservato nel Capitolo 5, essa non può essere interpretata come prossima a una gaussiana in senso stretto. La distribuzione appare infatti meno simmetrica e caratterizzata da una coda destra più estesa. Sebbene media e mediana risultino tra loro abbastanza vicine, entrambe si collocano al di sotto del punto medio dell'intervallo complessivo dei volumi $[4.48, 219.36] \text{ mm}^3$, che vale circa 111.92 mm^3 . In particolare, una mediana pari a circa 83.79 mm^3 implica che circa il 50% delle configurazioni si concentri nell'intervallo $[4.48, 83.79] \text{ mm}^3$, che occupa una porzione più ridotta del dominio rispetto all'intervallo complementare $[83.79, 219.36] \text{ mm}^3$. Ciò indica che, a sinistra della mediana, la densità di configurazioni è maggiore rispetto alla parte destra dell'intervallo, poiché i due sottointervalli contengono lo stesso numero di configurazioni ma il primo si estende su un range di volumi più ristretto.

In questo contesto, media e deviazione standard non vanno quindi lette con la stessa naturalezza con cui, nel Capitolo 5, descrivevano una distribuzione quasi simmetrica. Esse restano comunque utili come indicatori sintetici di posizione e dispersione, e consentono un confronto significativo tra i due problemi. In particolare, nel presente caso si ottiene $\frac{\sigma}{\mu} \approx 33\%$ mentre nel caso precedente risultava $\frac{\sigma}{\mu} \approx 21\%$. Ciò indica che, rispetto al Capitolo 5, la distribuzione dei volumi è relativamente più dispersa, ossia meno concentrata attorno al valore medio e più estesa lungo il dominio del

secondo obiettivo. In effetti, l'intervallo $\mu \pm 2\sigma$ risulta qui pari a circa $[28.45, 142.17] \text{ mm}^3$ e contiene circa il 96% delle configurazioni teoricamente possibili ed effettivamente campionate. Tale intervallo copre circa il 53% dell'intervallo complessivo dei volumi, mentre nel Capitolo 5 l'intervallo $\mu \pm 2\sigma$ contenente circa il 95% delle configurazioni copriva una porzione più ridotta del dominio del secondo obiettivo, pari a circa il 42% dell'intervallo totale (ragionamenti analoghi si possono fare con $\mu \pm \sigma$ e $\mu \pm 3\sigma$).

Questa maggiore dispersione, unita al fatto che nel presente capitolo le 1024 simulazioni del dataset rappresentano una quota molto più significativa dello spazio complessivo delle configurazioni ammissibili rispetto al caso precedente, costituisce una chiave interpretativa plausibile dell'andamento dell'errore osservato. In questo senso, il modello surrogato sembra poter contare, nel presente caso, su una rappresentazione più diffusa del dominio, il che contribuisce a spiegare perché l'errore non si concentri in modo evidente soltanto in alcune regioni del fronte, ma risulti distribuito in modo più omogeneo lungo l'intervallo dei volumi validati. Naturalmente, non si tratta di un'analisi rigorosa in senso dimostrativo, ma di una lettura utile a fornire una spiegazione ragionevole dei comportamenti osservati.

Una volta discusso il comportamento degli errori, l'attenzione può essere rivolta al risultato di maggiore interesse progettuale, ossia il fronte di Pareto ottenuto dopo validazione CFD e riportato in Figura 79.

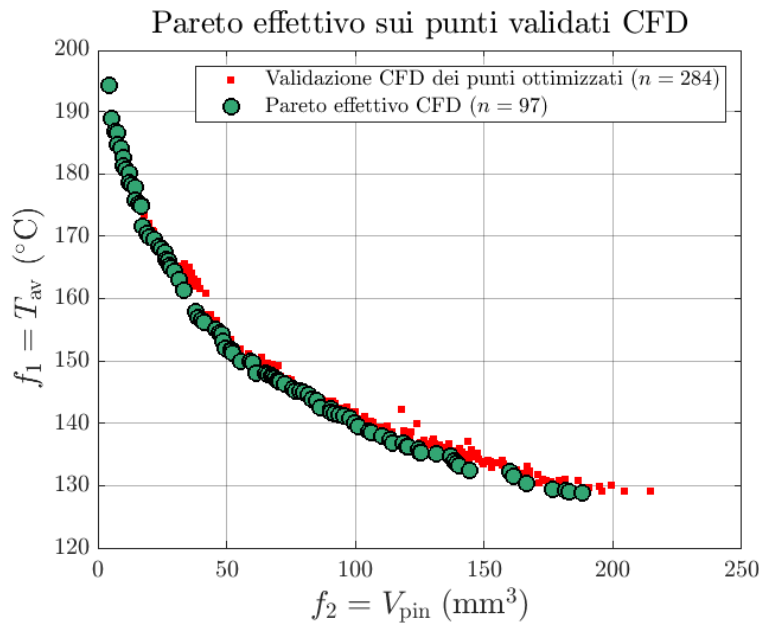


Figura 79: Fronte di Pareto effettivo ottenuto sui risultati validati mediante simulazione CFD

Anche in questo caso, va precisato che esso non coincide necessariamente con il fronte di Pareto assoluto del problema, poiché non vi è garanzia che l'insieme validato contenga in senso stretto tutte le migliori soluzioni possibili. Tuttavia, la forma regolare del fronte e la buona densità dei punti ottenuti (97 configurazioni non dominate) suggeriscono che la procedura abbia individuato un insieme di compromessi fisicamente credibili e progettualmente significativi. Rispetto al dataset iniziale, che conteneva 32 configurazioni non dominate, il fronte validato risulta quindi sensibilmente più ricco, confermando l'utilità del *workflow* nel raffinare e infittire i migliori compromessi del problema.

Anche nel presente capitolo, inoltre, l'obiettivo ultimo del *workflow* non era tanto prevedere con accuratezza assoluta ogni singolo valore di T_{av} , quanto piuttosto guidare efficacemente la ricerca

delle configurazioni migliori. Sotto questo profilo, il risultato appare comunque convincente, pur con una conservazione dell'ordinamento relativo meno marcata rispetto a quanto osservato nel Capitolo 5. Dopo validazione CFD, infatti, 97 delle 284 configurazioni individuate dal surrogato appartengono ancora al fronte effettivo, contro le 82 su 125 del caso precedente (Paragrafo 5.4.2). Tale differenza non va però interpretata come una criticità del *workflow*: nel presente problema le configurazioni migliori risultano infatti molto più dense e ravvicinate nel piano obiettivi, poiché il volume varia in modo molto più fitto rispetto al numero discreto di pin del Capitolo 5. In queste condizioni, anche scostamenti predittivi modesti possono modificare più facilmente il dettaglio dell'ordinamento locale tra soluzioni molto vicine tra loro, senza compromettere la qualità complessiva del risultato. Da questo punto di vista, l'individuazione di 97 configurazioni non dominate, ben distribuite lungo il fronte validato, costituisce comunque un esito molto positivo.

Per rendere più concreto il contenuto geometrico del fronte, in Figura 80 sono riportate nove forme rappresentative, ordinate per volume crescente e scelte tra le 97 configurazioni non dominate validate. La figura non ha naturalmente pretesa esaustiva, ma ha il solo scopo di fornire una lettura qualitativa delle tendenze geometriche che emergono lungo il fronte di Pareto.

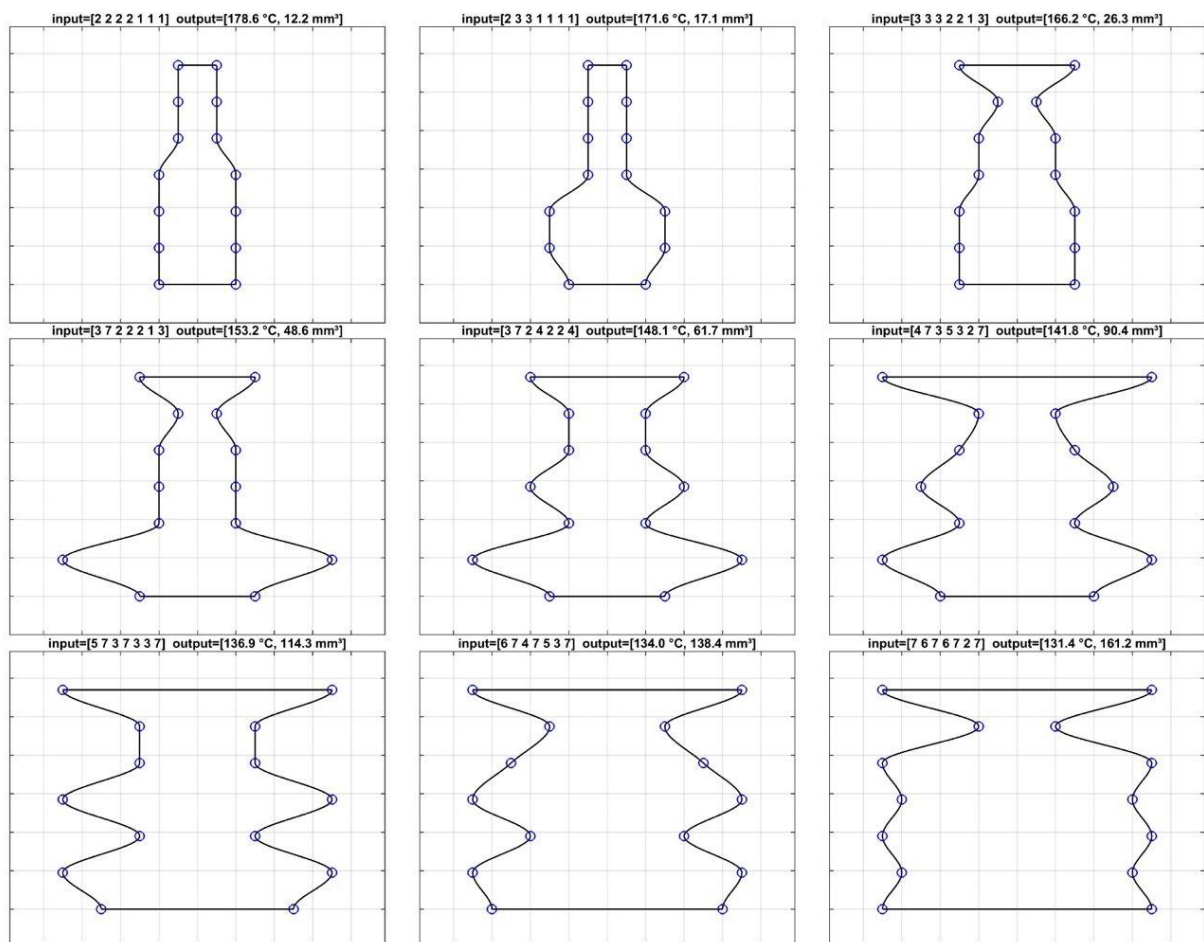


Figura 80: Alcune forme rappresentative appartenenti al fronte di Pareto validato, riportate in ordine di volume crescente per evidenziare le principali tendenze geometriche emerse lungo il fronte

L'osservazione più immediata è che, ai bassi volumi, il materiale tende a essere collocato prioritariamente nelle regioni più vicine alla base del pin. Tale comportamento è coerente con quanto discusso nell'analisi preliminare delle correlazioni: le sezioni inferiori risultano infatti particolarmente influenti sulla risposta termica, poiché contribuiscono a costruire il percorso di conduzione del calore dalla superficie riscaldata verso il resto del pin. Ciò non significa tuttavia che le configurazioni migliori concentrino semplicemente tutto il materiale alla base lasciando la parte

superiore fortemente assottigliata. Al contrario, lungo il fronte emerge una certa continuità geometrica del profilo, segno che le soluzioni ottime non dipendono soltanto dall'accumulo locale di materiale, ma da una distribuzione complessiva più equilibrata lungo l'altezza.

Passando a volumi intermedi, iniziano a manifestarsi in modo più evidente alcuni aspetti già suggeriti dall'analisi di correlazione, in particolare il ruolo di d_2 e, in misura diversa ma comunque significativa, quello di d_7 . Da un lato, le sezioni inferiori continuano a mantenere un peso rilevante. Dall'altro, cresce progressivamente anche l'importanza della zona terminale, che in molte configurazioni tende ad allargarsi. Già in questa regione del fronte iniziano inoltre a comparire profili non monotoni, con restringimenti e allargamenti alternati lungo l'altezza.

Procedendo verso i volumi più elevati, tale tendenza diventa ancora più marcata. Parallelamente, l'importanza relativa di alcuni diametri inferiori tende a ridursi. In particolare, la sezione associata a d_2 , non risulta necessariamente sempre ampia nelle configurazioni prossime all'estremo destro del fronte.

Per interpretare la comparsa dei restringimenti ed allargamenti, ci si è chiesti se essa potesse essere ricondotta almeno in parte alla necessità di aumentare la superficie disponibile per lo scambio convettivo. La presenza di profili ondulati può essere letta come un possibile tentativo di aumentare la superficie laterale senza un corrispondente incremento uniforme del volume.

Per verificare se questa interpretazione fosse effettivamente plausibile, è stato considerato separatamente il problema della massimizzazione dell'area di scambio termico del pin in funzione del volume. Sono state quindi individuate, per ciascun livello di volume, le configurazioni caratterizzate dall'area più elevata, mantenendo soltanto quelle per cui l'aumento di volume risultasse giustificato da un effettivo incremento di superficie rispetto a tutte le configurazioni di volume inferiore. Alcuni esempi di tali configurazioni sono riportati in Figura 81.

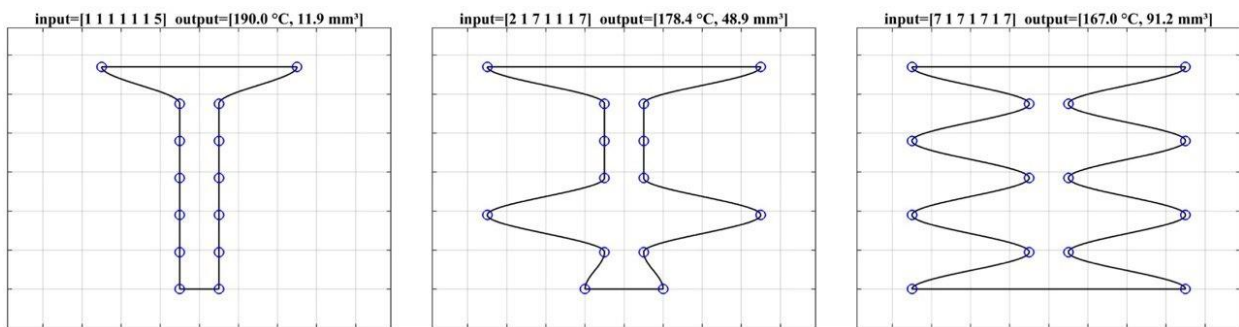


Figura 81: Esempi di configurazioni del pin appartenenti al fronte di compromesso tra volume e superficie di scambio termico

Le forme così ottenute mostrano effettivamente alcuni pattern simili a quelli osservati lungo il fronte ottimizzato termicamente. In ogni caso, il confronto con le configurazioni ottenute mediante l'ottimizzazione multi-obiettivo temperatura-volume, riportato in Figura 82, evidenzia chiaramente l'inferiorità delle prestazioni termiche associate a queste soluzioni.

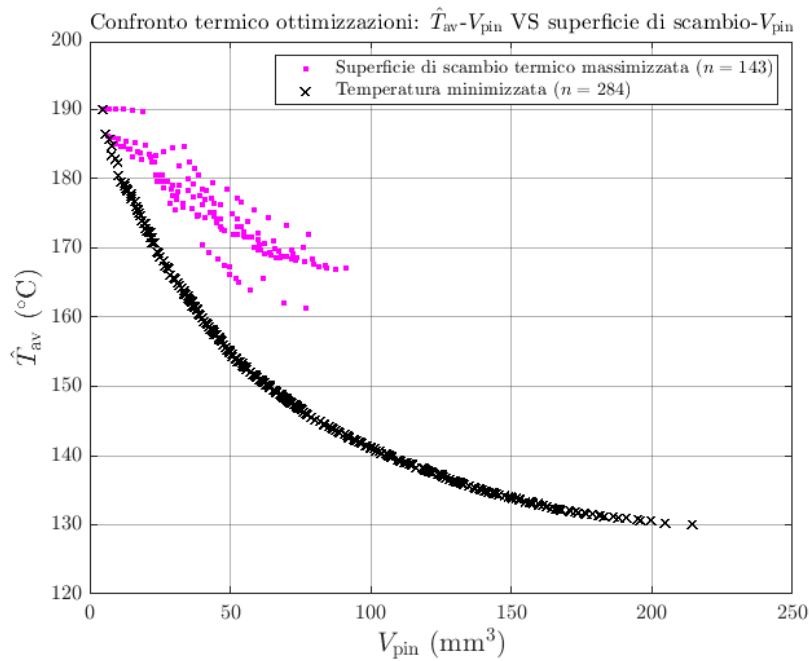


Figura 82: Confronto nel piano $(V_{pin}, \hat{T}_{av})$ tra le configurazioni ottenute mediante ottimizzazione multi-obiettivo temperatura-volume e quelle ottenute massimizzando la superficie di scambio termico del pin

Le temperature mostrate sono quelle stimate dal modello surrogato e non validate mediante CFD; nondimeno, il divario resta evidente. Si osserva inoltre che le configurazioni ottenute dal fronte volume-superficie di scambio termico si arrestano a circa 91 mm^3 : oltre tale soglia, non esistono configurazioni per cui a un aumento di volume corrisponda anche un aumento della superficie, e non compaiono quindi ulteriori configurazioni non dominate.

Il fatto che, a parità di volume, le forme selezionate secondo la massimizzazione dell'area mostrino prestazioni inferiori indica che questo criterio è in grado di spiegare solo parzialmente le forme osservate lungo il fronte di Pareto termico. Esso coglie alcuni tratti geometrici ricorrenti, ma non è sufficiente a riprodurre le configurazioni realmente migliori.

Una possibile lettura complessiva è allora che lungo il fronte esista una sorta di gerarchia nell'impiego del materiale. Ai bassi volumi, esso verrebbe utilizzato prioritariamente per rafforzare il percorso di conduzione del calore, soprattutto nelle regioni inferiori del pin; solo a volumi più elevati diventerebbe progressivamente possibile destinarne una parte anche all'incremento della superficie di scambio.

Nel complesso, i risultati ottenuti mostrano che il problema, pur non essendo banale, conserva un livello di complessità tale da permettere una lettura qualitativa almeno dei principali tratti geometrici emersi lungo il fronte. Questo aspetto è particolarmente utile, perché consente di attribuire un significato fisico alle configurazioni individuate e fornisce un primo riscontro del fatto che le forme ottimizzate non siano il prodotto di una ricerca puramente numerica priva di interpretazione, ma riflettano effettivamente alcuni meccanismi coerenti con il problema termofluidodinamico considerato.

Al tempo stesso, tali considerazioni restano necessariamente parziali e non sarebbero sufficienti, da sole, a guidare la progettazione in modo affidabile. I diversi effetti coinvolti quali: distribuzione del materiale, conduzione del calore, estensione della superficie libera e interazione tra forma e fluido, non agiscono infatti separatamente, ma si combinano in modo complesso in uno spazio di progetto di 823 543 configurazioni possibili.

Proprio in questo risiede il principale interesse del metodo adottato. Se nel Capitolo 5 il *workflow* aveva già mostrato di saper affrontare efficacemente un problema molto esteso ma concettualmente più semplice, il caso analizzato in questo capitolo costituisce un banco di prova più impegnativo, poiché introduce una maggiore complessità geometrica e fisica pur restando ancora sufficientemente leggibile dal punto di vista interpretativo. In un contesto di questo tipo, difficilmente il solo ragionamento qualitativo avrebbe consentito di ricostruire a priori un fronte di soluzioni credibile. Il fatto che il metodo riesca invece a individuare un insieme ampio e coerente di configurazioni non dominate rafforza l'idea, già emersa al termine del capitolo precedente, che il *workflow* sviluppato possa costituire uno strumento efficace anche per problemi progettuali più complessi.

Rimane tuttavia un aspetto che nel presente capitolo è stato discusso soltanto in modo indiretto, ossia il ruolo della forma nel modificare lo scambio con il fluido e, più in generale, l'interazione tra il fluido e la schiera di pin. Sebbene sia plausibile che una parte del vantaggio delle configurazioni ottimizzate dipenda proprio dalle modifiche indotte nei campi di moto e, di conseguenza, nella distribuzione locale del coefficiente di scambio termico, una valutazione più chiara di tali effetti richiede un'analisi CFD dedicata delle singole configurazioni.

Nel capitolo successivo verranno quindi analizzate in maggiore dettaglio due configurazioni ottimizzate ritenute particolarmente significative, confrontandole con pin cilindrici tradizionali mediante simulazioni CFD dedicate. Il riferimento al pin cilindrico è utile non soltanto perché esso rappresenta una geometria tradizionale ampiamente utilizzata, ma anche perché rende più leggibili i risultati del confronto, offrendo un termine noto rispetto al quale interpretare i vantaggi delle forme ottimizzate. Questo consentirà di osservare più chiaramente non solo gli aspetti geometrici già discussi, ma anche il modo in cui tali forme modificano i campi di temperatura, di velocità e, più in generale, le modalità di interazione con il fluido, fornendo così una lettura fisica più concreta dei vantaggi emersi dall'ottimizzazione.

6.3.2 Esempio di due soluzioni significative e confronto con cilindri tradizionali

Le forme ottimizzate sono state confrontate con una famiglia di pin cilindrici a diametro costante. In particolare, sono state valutate mediante CFD configurazioni cilindriche ottenute variando il diametro da 1 a 7 mm con passo pari a 0.2 mm. Tale scelta genera anche configurazioni non appartenenti allo spazio discreto effettivamente esplorato durante l'ottimizzazione, ma risulta utile per disporre di confronti a circa parità di volume o a circa parità di prestazione termica; in assenza di questa discretizzazione più fitta, i confronti disponibili sarebbero troppo pochi e troppo distanti tra loro per consentire un'analisi sufficientemente chiara.

In Figura 83 è riportato il confronto tra il fronte di Pareto validato CFD ottenuto nel paragrafo precedente e le prestazioni delle configurazioni cilindriche.

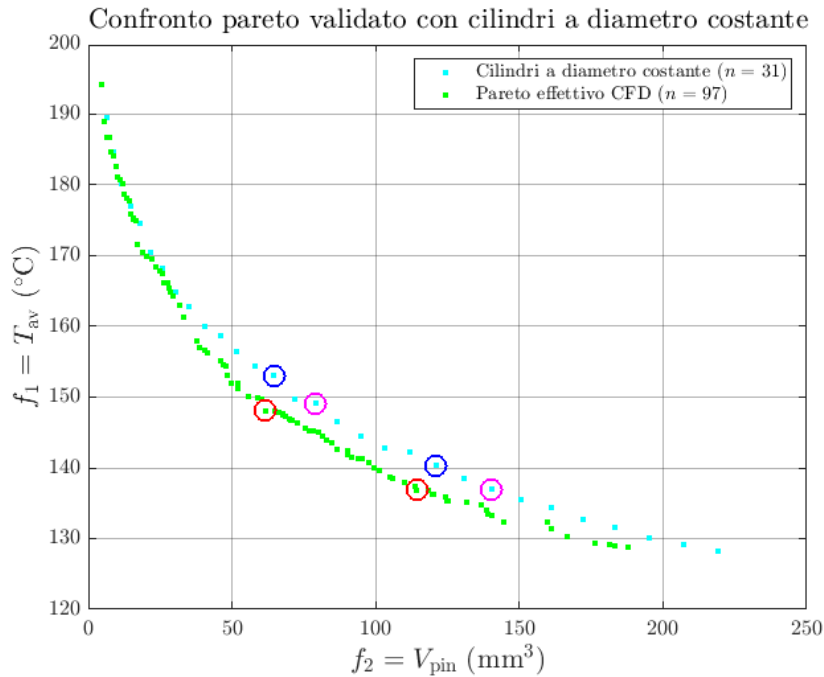


Figura 83: Confronto tra il fronte di Pareto validato CFD e le prestazioni dei pin cilindrici a diametro costante. Le due configurazioni ottimizzate selezionate per il confronto sono evidenziate in rosso; i cilindri a circa pari volume sono evidenziati in blu, mentre quelli a circa pari temperatura media sono evidenziati in magenta

Dal grafico emerge che i vantaggi delle forme ottimizzate risultano particolarmente evidenti oltre un certo livello di volume. Questo comportamento è atteso: quando il volume del pin è molto ridotto e ci si avvicina a configurazioni prossime ai cilindri di diametro 1-2 mm, vi è poco margine per redistribuire il materiale in modo realmente efficace. A ciò si aggiunge il fatto che, in tale regione, anche il numero di configurazioni ottenibili risulta più limitato, per cui diminuisce la possibilità di individuare forme sostanzialmente diverse da quelle cilindriche. Al crescere del volume disponibile, invece, la forma può agire in modo più significativo sul compromesso tra conduzione interna, superficie di scambio e interazione con il fluido, facendo emergere con maggiore chiarezza il vantaggio delle geometrie ottimizzate.

Il confronto in Figura 83 mette in evidenza due vantaggi complementari: da un lato, a parità circa di volume, alcune forme ottimizzate consentono di ridurre la temperatura media della superficie calda; dall'altro, a parità circa di temperatura media, esse richiedono un volume inferiore rispetto ai cilindri tradizionali. Per discutere più concretamente tali vantaggi, sono state selezionate due configurazioni ottimizzate significative, evidenziate in rosso in Figura 83. Ciascuna di esse viene confrontata con due configurazioni cilindriche: una a circa pari volume, evidenziata in blu, e una a circa pari temperatura media, evidenziata in magenta.

La prima configurazione ottimizzata selezionata è il pin ottenuto dal vettore [3 7 2 4 2 2 4]. Essa viene confrontata, da un lato, con il cilindro a diametro 3.8 mm, che presenta volume simile, e, dall'altro, con il cilindro a diametro 4.2 mm, che presenta prestazione termica comparabile. Nel seguito, per evitare ambiguità, la prestazione termica verrà indicata con T_{av} , mantenendo quindi la stessa notazione adottata nel resto della tesi per la temperatura media della superficie calda. Le Figure 84 e 85 riportano rispettivamente il confronto termico e quello fluidodinamico con il cilindro a circa pari volume.

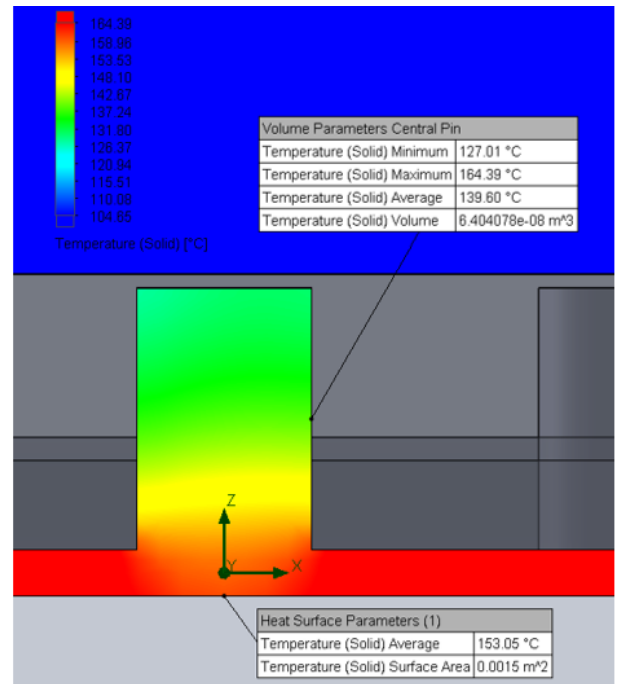
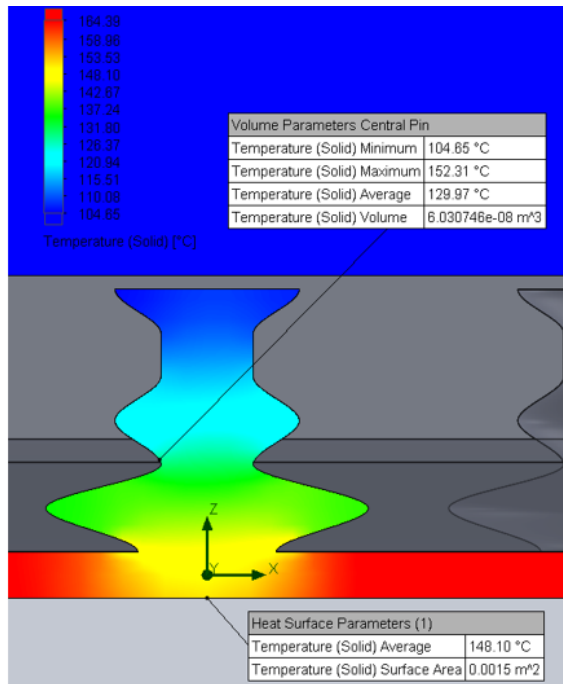


Figura 84: Confronto termico, in sezione sul pin centrale, tra la prima configurazione ottimizzata [3 7 2 4 2 2 4] e il cilindro a circa pari volume (diametro 3.8 mm). I valori riportati nelle immagini si riferiscono al solo pin centrale rappresentato

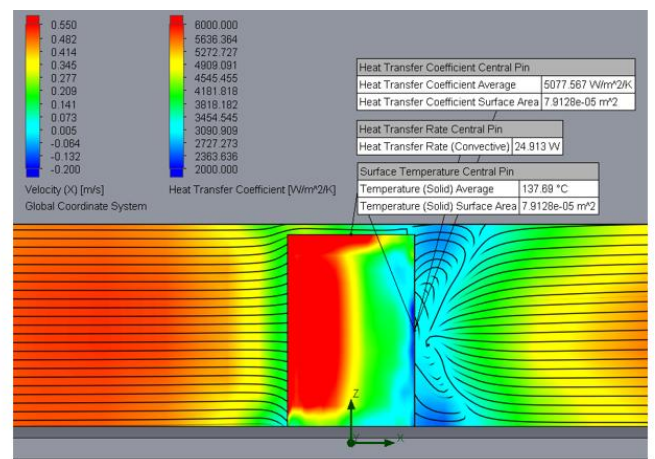
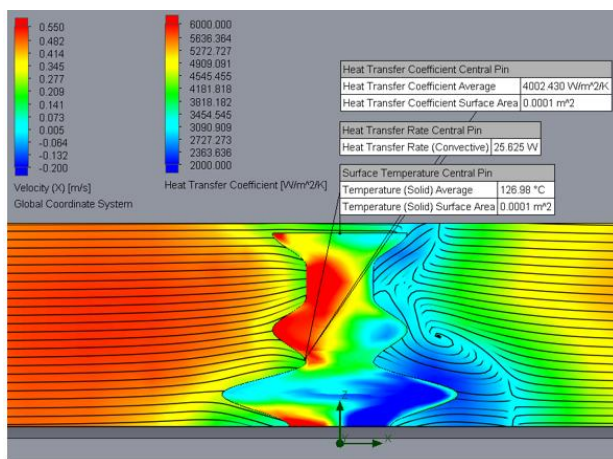


Figura 85: Confronto fluidodinamico, in sezione sul pin centrale, tra la prima configurazione ottimizzata [3 7 2 4 2 2 4] e il cilindro a circa pari volume (diametro 3.8 mm). Sono mostrate le linee di corrente nella sezione e la distribuzione superficiale del coefficiente di scambio termico convettivo sul pin centrale

Dalle mappe termiche si osserva chiaramente che il pin ottimizzato rimane a temperatura più bassa ($T_{av,pin} = 129.97 \text{ °C}$ riferita all'intero volume del pin) rispetto al cilindro ($T_{av,pin} = 139.60 \text{ °C}$), e che a tale comportamento si associa anche una riduzione della temperatura media della base calda T_{av} (148.10 °C contro 153.05 °C). La distribuzione termica suggerisce inoltre che, nel caso ottimizzato, il trasferimento di calore dalla base verso la superficie del pin avvenga con un gradiente più marcato.

I valori riportati nelle immagini permettono una prima lettura del fenomeno, ma si riferiscono al solo pin centrale, scelto come elemento rappresentativo e mostrato in sezione per ragioni di leggibilità grafica; le considerazioni quantitative complessive verranno invece discusse successivamente utilizzando valori mediati sull'intera schiera. Si è inoltre osservato che i volumi e le superfici di scambio restituiti da SolidWorks risultano leggermente diversi da quelli calcolati in MATLAB: lo scostamento è comunque modesto e non altera le conclusioni del confronto.

La Figura 85 consente invece di leggere qualitativamente il diverso comportamento fluidodinamico delle due forme. Nella sezione passante per il pin centrale, il campo di moto mostra che la geometria ottimizzata induce una struttura del flusso differente rispetto al cilindro, con ricircoli più evidenti nella regione a valle del pin. Per quanto riguarda invece la mappa del coefficiente di scambio termico convettivo h , essa appare in parte coerente con il campo di moto. Tuttavia, va ricordato che si tratta di una grandezza superficiale; ciò che si osserva in figura è quindi la distribuzione di h sulla superficie del pin. In questo caso, la configurazione cilindrica mostra localmente e in media valori di h più elevati ($h_{av,pin}$: 5078 W/(m² °C) contro 4002 W/(m² °C)). Questo aspetto, tuttavia, non è in contraddizione con la migliore prestazione della forma ottimizzata: come verrà discusso nelle tabelle successive, il vantaggio delle configurazioni ottimizzate non dipende da un semplice aumento di h , ma da un compromesso più favorevole tra coefficiente di scambio, superficie disponibile e conduzione interna nel solido.

Un secondo esempio è fornito dalla configurazione ottimizzata [5 7 3 7 3 3 7], confrontata con il cilindro a diametro 5.2 mm, a circa pari volume, e con il cilindro a diametro 5.6 mm, a circa pari T_{av} . Le Figure 86 e 87 mostrano i confronti termico e fluidodinamico con il cilindro a circa pari T_{av} .

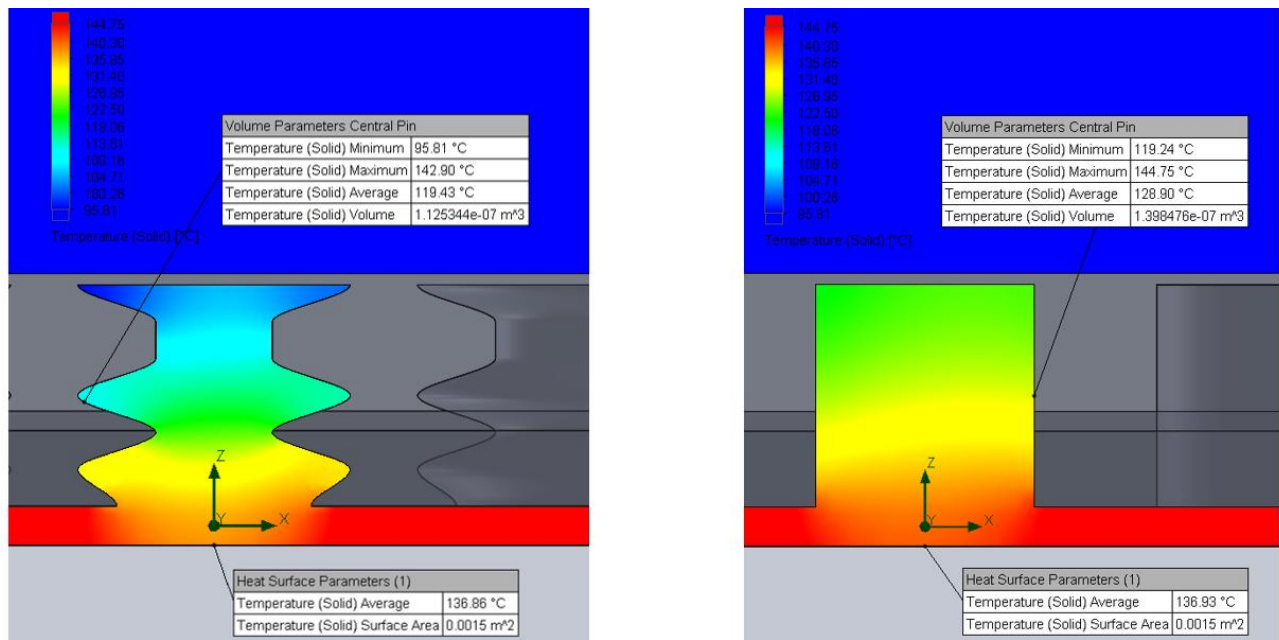


Figura 86: Confronto termico, in sezione sul pin centrale, tra la seconda configurazione ottimizzata [5 7 3 7 3 3 7] e il cilindro a circa pari T_{av} (diametro 5.6 mm). I valori riportati nelle immagini si riferiscono al solo pin centrale rappresentato

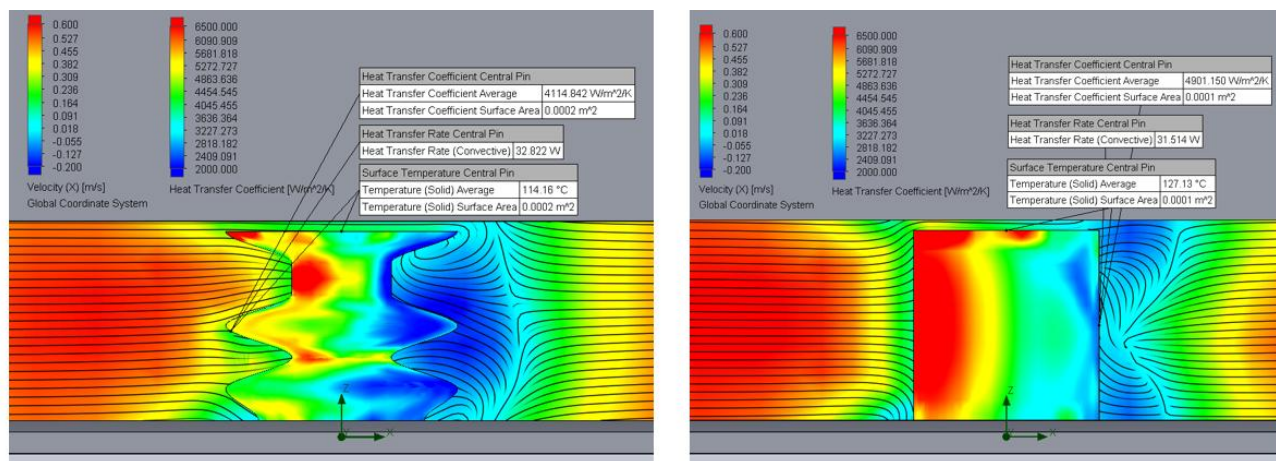


Figura 87: Confronto fluidodinamico, in sezione sul pin centrale, tra la seconda configurazione ottimizzata [5 7 3 7 3 3 7] e il cilindro a circa pari T_{av} (diametro 5.6 mm). Sono mostrate le linee di corrente nella sezione e la distribuzione superficiale del coefficiente di scambio termico convettivo sul pin centrale

Anche in questo caso, il confronto fluidodinamico porta a osservazioni in larga parte analoghe a quelle già discusse: la forma ottimizzata modifica il campo di moto locale rispetto al cilindro e presenta valori di h inferiori sulla superficie del pin. Diversa è invece la lettura del campo termico. Pur restando valide osservazioni analoghe a quelle del caso precedente sul comportamento del singolo pin, in questo caso la minore $T_{av,pin}$ della configurazione ottimizzata rispetto al cilindro non si traduce in una riduzione apprezzabile della prestazione termica globale T_{av} , che rimane infatti pressoché invariata. Ciò indica che la sola diminuzione della temperatura del pin non è sufficiente, di per sé, a spiegare il comportamento complessivo della configurazione.

L'interpretazione delle singole immagini resta tuttavia locale; per una visione più ordinata e completa risulta più utile riassumere i risultati in forma tabellare, utilizzando non i parametri del solo pin centrale, ma valori mediati sull'intera schiera. Tali valori risultano naturalmente meno favorevoli rispetto a quelli del solo pin centrale, che è collocato in corrispondenza della zona più calda ed è investito dal fluido in condizioni particolarmente efficaci.

Per iniziare, in Tabella 12 è riportato un riepilogo sintetico dei sei casi considerati.

Configurazioni	$V_{pin} (mm^3)$ (Singolo Pin)	$T_{av} (^{\circ}C)$	Miglioramento $V_{pin} \%$	Miglioramento $T_{av} \%$
Cilindro Diametro = 3.8 mm	64.64	153.05	/	/
Pin ottimizzato [3 7 2 4 2 2 4]	61.69	148.10	VS Diametro 3.8: -4.56% VS Diametro 4.2: -21.88%	VS Diametro 3.8: -3.23% VS Diametro 4.2: -0.67%
Cilindro Diametro = 4.2 mm	78.97	149.10	/	/
Cilindro Diametro = 5.2 mm	121.05	140.37	/	/
Pin ottimizzato [5 7 3 7 3 3 7]	114.35	136.90	VS Diametro 5.2: -5.54% VS Diametro 5.6: -18.55%	VS Diametro 5.2: -2.47% VS Diametro 5.6: -0.03%
Cilindro Diametro = 5.6 mm	140.39	136.94	/	/

Tabella 12: Confronto sintetico tra le due configurazioni ottimizzate selezionate e i corrispondenti cilindri di riferimento a circa pari volume e a circa pari temperatura media

Le configurazioni sono ordinate in modo da facilitare il confronto: per ciascuna delle due soluzioni ottimizzate, il cilindro riportato sopra è quello a circa pari volume, mentre quello sotto è il cilindro a circa pari prestazione termica. Questa organizzazione rende immediato osservare che entrambe le configurazioni ottimizzate risultano vantaggiose in entrambi i sensi già discussi: la prima fornisce una T_{av} inferiore rispetto al cilindro a pari volume e, allo stesso tempo, richiede un volume inferiore rispetto al cilindro a pari T_{av} ; la seconda mostra lo stesso comportamento nella fascia di volume più elevata. In termini quantitativi, la riduzione di volume a circa pari T_{av} è dell'ordine del 20%, mentre la riduzione di T_{av} a circa pari volume è dell'ordine del 3%. Va però sottolineato che i confronti sono stati impostati in modo conservativo: a circa pari volume, la configurazione ottimizzata viene comunque scelta con volume leggermente inferiore rispetto al cilindro, mentre a circa pari T_{av} viene scelta con temperatura leggermente inferiore. Le percentuali riportate rappresentano quindi una stima prudente, e non enfattizzata, del vantaggio delle forme ottimizzate.

La Tabella 13 riporta, per tutti e sei i casi considerati, un insieme di grandezze mediate sull'intera schiera e introduce in modo ordinato alcuni parametri non ancora discussi esplicitamente nella trattazione. Per distinguere tali valori dai corrispondenti parametri già mostrati nelle figure sul solo pin centrale, si adotta una barretta sovrascritta. Si mantiene invece invariata la notazione di T_{av} . In particolare, sono riportati $\bar{T}_{av, pin}$, temperatura media sul volume dei pin mediata sull'intera schiera; $\bar{T}_{av, sup\ pin}$, temperatura media della sola superficie di scambio dei pin, anch'essa mediata sull'intera schiera; $\bar{h}_{av, pin}$, coefficiente medio di scambio convettivo sui pin, mediato sull'intera schiera; e infine $S_{scambio, pin}$, superficie di scambio termico del singolo pin.

Configurazioni	$V_{pin} (mm^3)$ (Singolo Pin)	$T_{av} (^{\circ}C)$	$\bar{T}_{av, pin} (^{\circ}C)$ (Mediato su tutti i pin)	$\bar{T}_{av, sup\ pin} (^{\circ}C)$ (Mediato su tutti i pin)	$\bar{h}_{av, pin} (W/(m^2 \cdot ^{\circ}C))$ (Mediato su tutti i pin)	$S_{scambio, pin} (mm^2)$ (Singolo Pin)
Cilindro Diametro = 3.8 mm	64.64	153.05	126.60	125.11	4982.0	79.39
Pin ottimizzato [3 7 2 4 2 2 4]	61.69	148.10	118.73	116.37	3930.1	133.04
Cilindro Diametro = 4.2 mm	78.97	149.10	125.48	124.02	4834.9	89.06
Cilindro Diametro = 5.2 mm	121.05	140.37	121.90	120.48	4614.9	114.35
Pin ottimizzato [5 7 3 7 3 3 7]	114.35	136.90	112.18	107.88	3906.3	216.96
Cilindro Diametro = 5.6 mm	140.39	136.94	119.90	118.48	4658.9	124.91

Tabella 13: Grandezze termiche e fluidodinamiche mediate sull'intera schiera di pin per i sei casi considerati

I dati mostrano alcuni comportamenti comuni alle configurazioni ottimizzate, in parte già osservati nelle mappe termiche e fluidodinamiche. In primo luogo, sia la temperatura media sul volume dei pin $\bar{T}_{av, pin}$ sia la temperatura media sulla loro superficie di scambio $\bar{T}_{av, sup\ pin}$ risultano sistematicamente inferiori rispetto ai corrispondenti valori dei cilindri di riferimento. Questa prima osservazione suggerisce che, nelle configurazioni ottimizzate, si instaurino gradienti termici più elevati tra la base calda e i pin e, più in generale, all'interno del solido. In altri termini, il fatto che i pin ottimizzati rimangano mediamente più freddi, pur essendo collegati a una piastra con temperatura globale dello stesso ordine di grandezza, è indicativo di una maggiore resistenza al trasferimento termico per conduzione dalla base verso il pin.

In secondo luogo, il coefficiente medio di scambio convettivo $\bar{h}_{av, pin}$ risulta inferiore nelle configurazioni ottimizzate, mentre la superficie di scambio del singolo pin $S_{scambio, pin}$ è maggiore. Queste due osservazioni denotano chiaramente un *trade-off* di natura convettiva: le forme ottimizzate tendono infatti a ridurre il coefficiente medio di scambio locale, ma compensano tale effetto con un aumento della superficie disponibile per lo scambio termico.

Ne risulta che il vantaggio delle configurazioni ottimizzate non può essere attribuito né alla sola conduzione né al solo coefficiente h , ma va letto come il risultato di un compromesso più complesso tra conduzione interna, superficie di scambio e interazione convettiva col fluido. Proprio alla luce di questi segnali, verrà introdotta nel seguito una lettura in termini di resistenze termiche equivalenti, così da distinguere in modo più chiaro il contributo della conduzione dalla base alla superficie dei pin da quello della convezione tra schiera e fluido.

Prima di introdurre tali resistenze risulta utile osservare come la potenza termica si ripartisca tra i principali percorsi di smaltimento dal dominio caldo al fluido.

A questo scopo, la Tabella 14 distingue tre contributi: la potenza termica scambiata dalla schiera di pin con il fluido, $\phi_{schiera}$, quella scambiata direttamente dalla piastra, $\phi_{piastra}$, e quella che attraversa cooler e tubi, $\phi_{cooler+tubi}$. Si noti che la somma dei 3 contributi (ϕ_{tot}) è sempre circa 600 W ovvero la potenza termica immessa nel sistema, questo è corretto in quanto in regime stazionario la potenza che entra nel sistema deve essere uguale a quella che ne esce.

Configurazioni	$V_{pin} (mm^3)$ (Singolo Pin)	$T_{av} (^\circ C)$	$\phi_{schiera} (W)$	$\phi_{piastra} (W)$	$\phi_{cooler+tubi} (W)$	$\phi_{tot} (W)$
Cilindro Diametro = 3.8 mm	64.64	153.05	217.27	272.53	110.32	600.11
Pin ottimizzato [3 7 2 4 2 2 4]	61.69	148.10	222.29	273.15	104.56	600.00
Cilindro Diametro = 4.2 mm	78.97	149.10	231.06	263.11	105.45	599.62
Cilindro Diametro = 5.2 mm	121.05	140.37	262.41	243.08	94.47	599.96
Pin ottimizzato [5 7 3 7 3 3 7]	114.35	136.90	288.79	220.36	91.00	600.15
Cilindro Diametro = 5.6 mm	140.39	136.94	276.18	233.57	90.25	600.00

Tabella 14: Ripartizione della potenza termica complessivamente smaltita dal sistema nei tre principali percorsi di trasferimento verso il fluido

Si osserva che le configurazioni termicamente meno performanti sono associate a valori più elevati di $\phi_{cooler+tubi}$: ciò suggerisce che il percorso attraverso cooler e tubi costituisca, nel complesso, una via meno efficace di smaltimento del calore. In altre parole, quando una quota rilevante della potenza termica è costretta a percorrere questa via, la temperatura della superficie calda tende inevitabilmente ad aumentare. Questo comportamento risulta particolarmente evidente nei confronti a circa pari volume.

Diverso è invece il discorso nei confronti a circa pari prestazione termica, nei quali la quota di potenza termica che attraversa cooler e tubi rimane circa la stessa tra configurazione ottimizzata e cilindro di riferimento. In questi casi, il confronto si gioca soprattutto sul diverso modo in cui la potenza viene ripartita tra schiera e piastra.

Nel caso della soluzione [3 7 2 4 2 2 4], la potenza termica scambiata dalla schiera, risulta inferiore rispetto al cilindro a circa pari prestazione, mentre il contributo della piastra, è più elevato. Ciò indica che la forma ottimizzata non mantiene la stessa prestazione termica massimizzando lo scambio della schiera, ma aumentando lo scambio della piastra.

Complementare al caso precedente è la soluzione [5 7 3 7 3 3 7], per la quale la potenza termica ceduta dalla schiera risulta invece maggiore rispetto al cilindro a circa pari prestazione, mentre il contributo della piastra è inferiore. Se nel primo caso le prestazioni della configurazione ottimizzata

erano associate a una ripartizione più favorevole verso la piastra, qui si osserva il comportamento opposto.

I vari casi mostrano quindi che, per avere circa pari prestazione termica, il parametro che deve rimanere contenuto è soprattutto il contributo del ramo cooler+tubi, mentre la ripartizione tra schiera e piastra può variare anche in modo significativo. Questo aspetto consente di interpretare in modo più chiaro il significato fisico del miglioramento: non si tratta semplicemente di “far lavorare di più i pin”, ma di far sì che una quota maggiore possibile del calore venga smaltita nelle regioni centrali, cioè attraverso schiera e piastra, evitando per quanto possibile il ricorso a percorsi complessivamente meno efficienti. La forma dei pin ottimizzati modifica quindi non solo il comportamento della schiera in sé, ma anche il modo in cui l'intero sistema ripartisce la potenza termica tra i diversi percorsi disponibili. In sostanza, il sistema può essere interpretato in modo semplificato immaginando tre principali percorsi di smaltimento del calore tra superficie calda e fluido, ciascuno formato da una resistenza di conduzione e una di convezione in serie, e concorrenti tra loro come rami in parallelo. Questa rappresentazione è plausibile soprattutto per il percorso attraverso cooler e tubi, che può essere considerato in prima approssimazione come una via separata. Meno appropriata è invece una lettura strettamente “in parallelo” per piastra e schiera di pin, poiché i due contributi non sono indipendenti: la forma dei pin influenza in modo marcato la resistenza della schiera, ma almeno in parte anche quella della piastra, sia per il modo in cui modifica localmente l'interazione del fluido con la superficie di base, sia per la diversa porzione di piastra che viene lasciata direttamente esposta al fluido. In tale lettura qualitativa, quanto più si riescono a rendere favorevoli i percorsi attraverso i pin e, in parte, la piastra, tanto più si riduce la resistenza termica equivalente complessiva e quindi la temperatura della superficie calda, poiché la potenza totale dissipata (600 W) e la temperatura del fluido ($\approx 75^\circ\text{C}$) restano sostanzialmente invariate.

Rimane tuttavia ancora aperto il problema di comprendere come, nello specifico, le forme ottimizzate riescano a ridurre (a pari volume), o mantenere circa costante (a pari T_{av}), la resistenza termica del percorso attraverso la schiera. Per chiarire questo aspetto, i risultati vengono reinterpretati in Tabella 15 anche in termini di resistenze termiche equivalenti, così da separare più chiaramente il contributo della conduzione da quello della convezione.

Configurazioni	$V_{pin} (mm^3)$ (Singolo Pin)	$T_{av} (^\circ\text{C})$	$R_{th, sup\ calda-schiera} (^\circ\text{C}/\text{W})$	$R_{th, schiera-fluido} (^\circ\text{C}/\text{W})$	$R_{th, somma} (^\circ\text{C}/\text{W})$
Cilindro Diametro = 3.8 mm	64.64	153.05	0.129	0.231	0.359
Pin ottimizzato [3 7 2 4 2 2 4]	61.69	148.10	0.143	0.186	0.329
Cilindro Diametro = 4.2 mm	78.97	149.10	0.109	0.212	0.321
Cilindro Diametro = 5.2 mm	121.05	140.37	0.076	0.173	0.249
Pin ottimizzato [5 7 3 7 3 3 7]	114.35	136.90	0.101	0.114	0.214
Cilindro Diametro = 5.6 mm	140.39	136.94	0.067	0.157	0.224

Tabella 15: Resistenze termiche equivalenti stimate per conduzione tra superficie calda e schiera e per convezione tra schiera e fluido

Le resistenze sono state stimate nelle forme:

$$R_{th, sup\ calda-schiera} = \frac{\Delta T_{sup\ calda-schiera}}{\phi_{schiera}}$$

$$R_{th, schiera-fluido} = \frac{\Delta T_{schiera-fluido}}{\phi_{schiera}}$$

dove $\Delta T_{sup\ calda-schiera}$ rappresenta la differenza tra la temperatura media della superficie calda e la temperatura media della superficie di scambio dei pin, e $\Delta T_{schiera-fluido}$ la differenza tra la temperatura media della superficie dei pin e la temperatura del fluido indisturbato. Essendo il problema trattato in regime stazionario, la potenza termica che entra nella schiera per conduzione deve uguagliare quella che la schiera cede al fluido per convezione; per questo motivo $\phi_{schiera}$ costituisce il riferimento naturale per entrambe le resistenze. Per la temperatura del fluido si assume il valore di ingresso, pari a 75 °C, approssimazione accettabile in quanto il fluido si riscalda di poco lungo il percorso (≈ 1 °C).

In questa lettura, $R_{th, sup\ calda-schiera}$ può essere interpretata come una resistenza termica equivalente associata al tratto conduttivo tra superficie calda e schiera, mentre $R_{th, schiera-fluido}$ rappresenta in modo analogo il contributo prevalentemente convettivo tra schiera e fluido. Quest'ultima è infatti riconducibile, alla classica espressione:

$$R_{conv} = \frac{1}{hA}$$

in particolare, utilizzando $\bar{h}_{av,pin}$ e $S_{scambio,pin}$, i valori ottenuti risultano molto simili a quelli stimati per $R_{th, schiera-fluido}$.

La Tabella 15 mette in evidenza una tendenza chiara. In entrambe le configurazioni ottimizzate, la resistenza $R_{th, sup\ calda-schiera}$ risulta maggiore rispetto ai cilindri di riferimento. Questo conferma quantitativamente quanto già suggerito dalle mappe termiche e dai valori di $\bar{T}_{av,pin}$ e $\bar{T}_{av,sup\ pin}$: le forme ottimizzate tendono a penalizzare il trasferimento termico lato conduzione, coerentemente con la presenza di gradienti termici più elevati tra base e pin.

Al tempo stesso, però, la resistenza $R_{th, schiera-fluido}$ diminuisce in modo netto. Questo risultato è particolarmente significativo, perché mostra che il peggioramento del coefficiente medio di scambio $\bar{h}_{av,pin}$, osservato sia nelle immagini sia nella Tabella 13, non solo viene compensato dall'aumento della superficie di scambio $S_{scambio,pin}$, ma dà luogo a un miglioramento effettivo del comportamento convettivo complessivo della schiera. In altri termini, le forme ottimizzate sembrano peggiorare la conduzione ma migliorare in misura ancora maggiore la convezione.

Il confronto tra le resistenze suggerisce quindi che il vantaggio delle forme ottimizzate derivi da un compromesso favorevole tra questi due effetti. Nel confronto a circa pari volume, questo meccanismo è sufficiente a spiegare anche la riduzione di T_{av} : la penalizzazione conduttiva è più che controbilanciata dal miglioramento convettivo, e la resistenza equivalente del percorso attraverso la schiera si riduce nel complesso. In questo caso, senza neppure chiamare in causa il contributo della piastra, la lettura della sola schiera è già sufficiente a spiegare qualitativamente la diminuzione della temperatura della superficie calda.

Il caso a circa pari prestazione termica richiede invece una lettura più attenta. Qui la resistenza totale del percorso della schiera ottimizzata, non necessariamente è esattamente uguale a quella del cilindro di confronto, e in un caso risulta addirittura leggermente più alta. Questo, tuttavia, non contraddice il fatto che T_{av} rimanga pressoché invariata: significa piuttosto che, nel complesso, il sistema continua a smaltire il calore con efficacia simile grazie alla combinazione tra schiera e piastra, come già suggerito dalla ripartizione dei flussi in Tabella 14. In altre parole, a circa pari prestazione, la sola lettura della schiera non basta più da sola a spiegare il risultato, ma deve essere

integrata con il contributo della piastra e con il fatto che la quota di potenza termica che attraversa cooler e tubi rimane dello stesso ordine di grandezza.

Si può quindi concludere che l'ottimizzazione individua configurazioni che sfruttano il volume disponibile non per massimizzare un singolo parametro, ma per cercare compromessi favorevoli tra conduzione e convezione. Questo compromesso non è strettamente confinato al solo percorso della schiera, ma si estende anche agli effetti che la forma dei pin esercita sulla piastra e quindi, più in generale, sulla ripartizione complessiva della potenza termica nel sistema. In questo senso, il vantaggio delle forme ottimizzate non va ricondotto a un unico meccanismo dominante, bensì a una gestione più efficiente del bilancio termo-fluidodinamico complessivo.

Rispetto a quanto discusso nel paragrafo precedente, i risultati qui ottenuti rafforzano inoltre due interpretazioni già emerse. Da un lato, risulta ulteriormente supportata l'ipotesi che una maggiore concentrazione di materiale verso la base del pin sia favorevole perché consente di non penalizzare eccessivamente la conduzione nella regione più direttamente coinvolta nell'alimentazione termica della superficie estesa. Dall'altro, le forme ottimizzate suggeriscono che la redistribuzione del volume possa determinare, nel complesso, un *trade-off* favorevole tra il possibile peggioramento locale del coefficiente di scambio convettivo e l'aumento della superficie disponibile per lo scambio, contribuendo così, nel complesso, a uno scambio convettivo più efficace.

Proprio in questo ambito viene ripresa un'osservazione già anticipata nel paragrafo precedente. Si era infatti ipotizzato che l'aumento della sezione terminale potesse risultare vantaggioso per il maggiore contributo alla superficie di scambio termico. Questo è certamente vero dal punto di vista geometrico; tuttavia, il beneficio sarebbe limitato se tale superficie aggiuntiva fosse associata a uno scambio convettivo debole. La distribuzione del coefficiente di scambio termico convettivo riportata in Figura 88 mostra invece che, anche in corrispondenza della parte terminale della configurazione ottimizzata 2, i valori di h rimangono complessivamente buoni. Ciò suggerisce quindi che l'aumento di superficie nella zona finale del pin non costituisca soltanto un vantaggio geometrico, ma contribuisca effettivamente allo scambio termico, in modo coerente con le ipotesi formulate in precedenza.

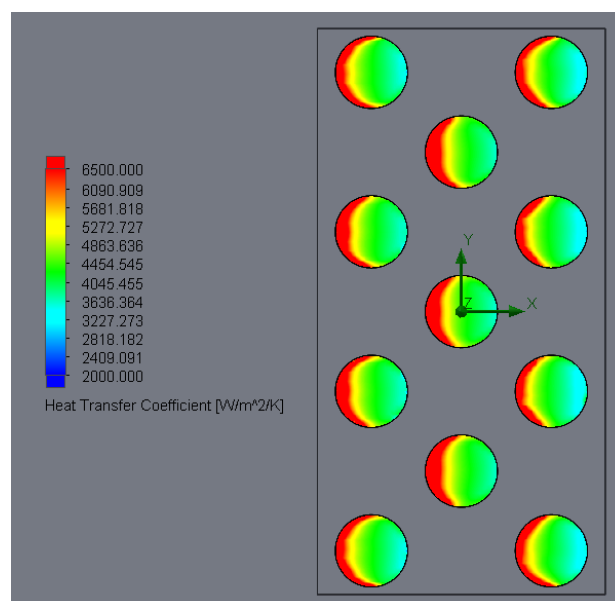


Figura 88: Distribuzione del coefficiente di scambio termico convettivo sulla superficie terminale dei pin della configurazione ottimizzata [5 7 3 7 3 3 7]

6.4 Conclusioni

Il presente capitolo ha avuto come primo obiettivo quello di verificare la trasferibilità del *workflow* sviluppato nel Capitolo 5 a un problema progettuale differente, in cui l'ottimizzazione non riguarda più la distribuzione spaziale dei pin, ma la forma del singolo elemento. Anche in questo caso, il metodo si è dimostrato soddisfacente: la procedura ha consentito di affrontare in modo sistematico il problema e di valutarne l'intero spazio discreto delle configurazioni ammissibili tramite modello surrogato, mantenendo il ricorso alla CFD nei soli passaggi più onerosi ma realmente informativi. In particolare, la generazione del dataset finale di 1024 configurazioni ha richiesto un tempo stimato di circa 27 ore, l'addestramento del modello neurale e l'intera fase di ottimizzazione sul surrogato hanno richiesto poco più di 3 minuti, e la successiva validazione CFD delle 284 configurazioni ottimizzate ha richiesto complessivamente circa 9 ore, portando il tempo totale dell'intera procedura a circa 36 ore.

Dal punto di vista dei risultati progettuali, nelle condizioni considerate il metodo ha portato all'individuazione di un fronte validato CFD costituito da 97 configurazioni non dominate. Pur non potendo identificare con certezza tale insieme con il fronte di Pareto assoluto del problema, la regolarità del fronte ottenuto e la buona densità dei punti suggeriscono che la procedura abbia individuato un insieme di compromessi progettualmente significativi.

Per verificare se tali configurazioni fossero effettivamente vantaggiose da un punto di vista progettuale, le forme ottimizzate sono state confrontate con una famiglia di pin cilindrici tradizionali a diametro costante. Il confronto ha mostrato che, soprattutto ai volumi medio-alti, le forme ottimizzate risultano effettivamente competitive rispetto ai cilindri, evidenziando due vantaggi complementari: da un lato, a circa pari volume, consentono di ridurre la temperatura media della superficie calda; dall'altro, a circa pari prestazione termica, richiedono un volume inferiore. Nei casi selezionati come più rappresentativi, la riduzione di volume a circa pari T_{av} risulta dell'ordine del 20%, mentre la riduzione di T_{av} a circa pari volume è dell'ordine del 3%. Si tratta inoltre di una stima prudente, poiché i confronti sono stati impostati in modo conservativo.

Infine, il capitolo ha cercato di fornire anche una lettura fisica dei risultati ottenuti, così da comprendere in che modo l'ottimizzatore tenda a distribuire il materiale lungo l'altezza del pin. L'analisi delle forme appartenenti al fronte ha mostrato che, ai bassi volumi, il materiale tende a concentrarsi prioritariamente nelle sezioni più vicine alla base, in accordo con l'idea che tali regioni siano particolarmente rilevanti per il percorso di conduzione del calore dalla superficie riscaldata verso il pin. All'aumentare del volume, emerge però una distribuzione più articolata, che non si riduce a un semplice ispessimento della base, ma coinvolge anche la regione terminale e porta alla comparsa di profili con alternanza di restringimenti e allargamenti. Nel complesso, si è giunti a una spiegazione ragionevole secondo cui le forme ottimizzate non massimizzano un singolo meccanismo, ma sfruttano il volume disponibile per costruire un compromesso favorevole tra conduzione interna, superficie di scambio e interazione con il fluido. Tale compromesso non riguarda esclusivamente la schiera, ma anche, almeno in parte, la piastra, sia perché la forma dei pin modifica l'interazione del fluido con la superficie di base, sia perché ne altera la porzione direttamente esposta al fluido.

7. Conclusioni generali e sviluppi futuri

Le conclusioni specifiche relative ai singoli passaggi del lavoro sono già state discusse al termine dei capitoli precedenti. In questa sede si intende quindi fornire una lettura complessiva della tesi, mettendo in evidenza il significato generale del percorso svolto, le principali criticità emerse e le possibili direzioni di sviluppo futuro.

Nel complesso, il lavoro ha consentito di sviluppare e verificare un *workflow* integrato per l'esplorazione e l'ottimizzazione di dissipatori a pin-fin raffreddati a liquido, basato sull'interazione tra modellazione CAD parametrica, simulazioni CFD automatizzate, costruzione di dataset numerici, addestramento di modelli surrogati e successiva ottimizzazione multi-obiettivo. Più che nell'individuazione di una singola geometria "ottima" in senso assoluto, il contributo principale della tesi risiede quindi nella costruzione di un metodo generale, pensato per rendere affrontabili problemi progettuali che, se trattati esclusivamente con simulazioni CFD nel *loop* di ottimizzazione, risulterebbero rapidamente troppo onerosi dal punto di vista computazionale. Tale impostazione rappresenta il filo conduttore dell'intero elaborato.

Da questo punto di vista, uno degli aspetti più significativi emersi nel corso del lavoro è proprio il valore metodologico della riformulazione. Il tentativo iniziale di utilizzare direttamente la CFD all'interno dell'ottimizzazione ha infatti mostrato in modo chiaro i limiti pratici di un approccio "*CFD in the loop*" applicato a uno spazio di progetto ampio e poco strutturato. Questa fase ha quindi avuto il ruolo di evidenziare in maniera concreta perché fosse necessario passare a una strategia diversa. In tal senso, il lavoro non si limita a mostrare che un *workflow* surrogato può funzionare, ma documenta anche in modo argomentato le ragioni per cui esso diventa preferibile rispetto all'ottimizzazione diretta in un contesto di questo tipo.

Un secondo elemento di interesse riguarda la trasferibilità dell'approccio. Il metodo sviluppato non è rimasto confinato a un unico problema progettuale, ma è stato applicato a due formulazioni differenti: una prima, volutamente semplificata, utile per verificare in modo chiaro l'efficacia della procedura end-to-end, e una seconda, più vicina a un problema applicativo reale, nella quale l'ottimizzazione ha riguardato la forma del singolo pin-fin. Questo passaggio da un caso di studio controllato a uno più realistico costituisce uno degli elementi più significativi dell'intero lavoro. Esso mostra infatti non solo che il *workflow* mantiene coerenza e trasferibilità al variare della parametrizzazione geometrica, del dataset e del contenuto fisico del problema, ma anche che la sua applicazione può produrre risultati progettualmente rilevanti. In particolare, il caso dei pin modificati ha evidenziato che l'approccio sviluppato non si limita a essere metodologicamente consistente, ma può anche portare all'individuazione di soluzioni non convenzionali con vantaggi concreti rispetto ai riferimenti tradizionali. Ne emerge quindi un duplice risultato: da un lato la solidità del metodo, dall'altro la sua effettiva capacità di generare esiti di interesse ingegneristico.

Naturalmente, tale valutazione positiva non elimina i limiti del lavoro, che è opportuno richiamare in modo esplicito. Le principali semplificazioni introdotte non riguardano tanto la logica generale del *workflow*, quanto piuttosto la sua applicazione ai casi studio considerati. In entrambi i problemi affrontati, infatti, è stato necessario circoscrivere lo spazio di progetto attraverso ipotesi geometriche e operative ben definite, così da rendere sostenibile la generazione del dataset e la successiva ottimizzazione. Questo significa che i risultati ottenuti vanno interpretati come validi all'interno delle condizioni adottate, e non come conclusioni immediatamente generalizzabili a configurazioni, vincoli o punti di funzionamento differenti. In tale prospettiva, la tesi dimostra soprattutto la fattibilità e la coerenza del metodo, mentre lascia aperta, come sviluppo naturale, un'estensione verso spazi di progetto più ampi e verso una trattazione più vicina alla complessità dei sistemi reali.

Tra le criticità emerse durante lo svolgimento del lavoro, una riguarda lo strumento software utilizzato per le simulazioni. SolidWorks Flow Simulation si è rivelato adeguato per costruire una pipeline funzionante e per condurre analisi coerenti con gli obiettivi della tesi, ma, sulla base dell'esperienza maturata, è apparso meno adatto quando il problema è passato da poche simulazioni mirate a campagne numeriche più estese e fortemente automatizzate. In particolare, la necessità di gestire macro, interfacce COM, scambio di file e apertura dell'interfaccia grafica ha introdotto un *overhead* implementativo non trascurabile, rendendo la procedura più laboriosa di quanto sarebbe auspicabile in un contesto di esplorazione progettuale su larga scala. Ciò non implica che la scelta di SolidWorks sia stata inappropriata: nel contesto di questa tesi essa ha consentito di mantenere un controllo diretto sulla geometria CAD e di sviluppare in modo chiaro il *workflow* MATLAB–SolidWorks. Tuttavia, proprio l'esperienza acquisita suggerisce, come possibile sviluppo futuro, la valutazione di ambienti di simulazione maggiormente orientati all'automazione e alla gestione di campagne parametriche estese, così da ridurre la complessità implementativa e rendere più immediata l'applicazione del metodo.

Per quanto riguarda gli sviluppi futuri sul piano modellistico e progettuale, una prima estensione naturale consiste nell'ampliamento delle variabili di progetto. Nel secondo caso studio, ad esempio, l'ottimizzazione ha riguardato la forma del singolo pin-fin, ma altre grandezze geometriche potrebbero essere incluse nel vettore decisionale, come lo spessore della piastra, l'altezza complessiva del pin, il numero totale di pin, il loro passo, oppure parametri capaci di modificare simultaneamente sia la geometria del pin sia la configurazione dell'array. Un'estensione di questo tipo renderebbe il problema più vicino a un reale scenario di progettazione integrata, in cui non si ottimizza un singolo dettaglio geometrico mantenendo tutto il resto fisso, ma si ricerca un compromesso complessivo tra più gradi di libertà.

Un secondo sviluppo importante riguarda le condizioni operative e le funzioni obiettivo. Nel presente lavoro l'attenzione è stata rivolta soprattutto alle prestazioni termiche, mentre una naturale evoluzione consisterebbe nell'includere in modo sistematico anche gli aspetti fluidodinamici. In particolare, la portata potrebbe essere trattata non più come parametro fisso ma come variabile di progetto, mentre tra gli obiettivi si potrebbe introdurre esplicitamente anche la perdita di carico, oppure una grandezza ad essa collegata come la potenza di pompaggio. Ciò permetterebbe di affrontare il problema in termini ancora più coerenti con la reale progettazione termo-idraulica, nella quale la riduzione della temperatura non può essere considerata indipendentemente dal costo energetico richiesto per ottenerla.

Dal punto di vista metodologico, sarebbe inoltre interessante approfondire il confronto tra approccio surrogato e approccio diretto a parità di budget computazionale complessivo. Nel presente lavoro, il passaggio al modello surrogato è stato motivato in modo convincente dalla difficoltà pratica dell'ottimizzazione diretta; tuttavia, in uno sviluppo futuro si potrebbe progettare un confronto ancora più sistematico, imponendo ad esempio lo stesso costo computazionale totale alle due strategie e verificando se, entro tale budget, la procedura assistita da surrogato riesca effettivamente a individuare soluzioni migliori o più robuste rispetto a una strategia CFD *in the loop* opportunamente riformulata. Un'analisi di questo tipo permetterebbe di quantificare in modo più diretto il beneficio reale introdotto dal surrogato.

Un ulteriore sviluppo di particolare interesse sarebbe la validazione sperimentale delle configurazioni più significative ottenute mediante ottimizzazione, così da verificare in modo diretto la corrispondenza tra i risultati numerici e il comportamento reale del sistema.

In conclusione, il valore principale del lavoro non risiede soltanto nei risultati ottenuti nei casi studio considerati, ma soprattutto nell'aver costruito, verificato e discusso criticamente un metodo di

lavoro trasferibile, che può costituire una base concreta per future estensioni verso problemi più ricchi, più realistici e più direttamente rilevanti dal punto di vista progettuale.

Riferimenti

- [1] Andresen, M., Ma, K., Buticchi, G., Falck, J., Blaabjerg, F., & Liserre, M. (2018). Junction Temperature Control for More Reliable Power Electronics. *IEEE Transactions on Power Electronics*, 33(1), 765-776. Article 7892031.
<https://doi.org/10.1109/TPEL.2017.2665697>
- [2] Ravanji A, Lee A, Mohammadpour J, Cheng S. Critical review on thermohydraulic performance enhancement in channel flows: a comparative study of pin fins. *Renewable and Sustainable Energy Reviews* 2023;188:113793.
<https://doi.org/10.1016/j.rser.2023.113793>
- [3] J.S. Lee, S.Y. Yoon, B. Kim, H. Lee, M.Y. Ha, J.K. Min, A topology optimization based design of a liquid-cooled heat sink with cylindrical pin fins having varying pitch, *International Journal of Heat and Mass Transfer* 172 (2021) 121172.
<https://doi.org/10.1016/j.ijheatmasstransfer.2021.121172>
- [4] A. Šikić, L. Grbčić, L. Kranjčević, Machine learning based surrogate models for microchannel heat sink optimization, *Applied Thermal Engineering* 222 (2023) 119917.
<https://doi.org/10.1016/j.applthermaleng.2022.119917>
- [5] Z. Zhang, X. Wang, Y. Yan, A review of the state-of-the-art in electronic cooling, *e- Prime - Advances in Electrical Engineering, Electronics and Energy* 1 (2021) 100009.
<https://doi.org/10.1016/j.prime.2021.100009>
- [6] Teertstra, P., Yovanovich, M. M., & Culham, J. R. (2003). *Conduction shape factor models for 3-D enclosures* (AIAA Paper No. AIAA-2003-0159). American Institute of Aeronautics and Astronautics.
<https://doi.org/10.2514/6.2003-159>
- [7] M. Bahrami, “Forced convection heat transfer” 2009.
<https://www.sfu.ca/~mbahrami/ENSC%20388/Notes/Forced%20Convection.pdf>
- [8] Dittus, F. W., & Boelter, L. M. K. (1985). Heat transfer in automobile radiators of the tubular type. *International Communications in Heat and Mass Transfer*, 12(1), 3–22.
[https://doi.org/10.1016/0735-1933\(85\)90003-X](https://doi.org/10.1016/0735-1933(85)90003-X)
- [9] Zorzetto, M., et al. (2024). *Reduced Order Modeling for Thermal Simulations of Electric Components With Surface-to-Surface Radiation* (IEEE Access / Volume 12, 2024).
<https://doi.org/10.1109/ACCESS.2024.3507367>
- [10] Zu, H., et al. (2021). Analysis of enhanced heat transfer on a passive heat sink with high-emissivity coating. *International Journal of Thermal Sciences*.
<https://doi.org/10.1016/j.ijthermalsci.2021.106971>

- [11] Ma, K., He, N., Liserre, M., & Blaabjerg, F. (2016). Frequency-Domain Thermal Modelling and Characterization of Power Semiconductor Devices. *IEEE Transactions on Power Electronics*, 31(10), 7183-7193.
<https://doi.org/10.1109/TPEL.2015.2509506>
- [12] Bar-Cohen, A., Iyengar, M., & Kraus, A. D. (2003). *Design of optimum plate-fin natural convective heat sinks*. *Journal of Electronic Packaging*, 125(2), 208–216.
<https://doi.org/10.1115/1.1568361>
- [13] Joo, Y., & Kim, S. J. (2015). *Comparison of thermal performance between plate-fin and pin-fin heat sinks in natural convection*. *International Journal of Heat and Mass Transfer*, 83, 345–356.
<https://doi.org/10.1016/j.ijheatmasstransfer.2014.12.023>
- [14] Jonsson, H., & Moshfegh, B. (2001). *Modeling of the thermal and hydraulic performance of plate fin, strip fin, and pin fin heat sinks—Influence of flow bypass*. *IEEE Transactions on Components and Packaging Technologies*, 24(2), 142–149.
<https://doi.org/10.1109/6144.926376>
- [15] Abramushkina, E., Zhaksylyk, A., Geury, T., El Baghdadi, M., & Hegazy, O. (2021). *A thorough review of cooling concepts and thermal management techniques for automotive WBG inverters: Topology, technology and integration level*. *Energies*, 14(16), 4981.
<https://doi.org/10.3390/en14164981>
- [16] Li, Y., Roux, S., Castelain, C., Fan, Y., & Luo, L. (2023). *Design and Optimization of Heat Sinks for the Liquid Cooling of Electronics with Multiple Heat Sources: A Literature Review*. *Energies*, 16(22), 7468.
<https://doi.org/10.3390/en16227468>
- [17] Li, Y., Su, Q., Sheng, G., Elmi Hosseini, S. R., Badran, B. E., Gong, P., Xin, C., & Wang, H. (2025). *Recent advances in artificial-intelligence enhanced additive manufacturing of heat exchangers for thermal management: A review*. *Materials & Design*, 256, 114339.
<https://doi.org/10.1016/j.matdes.2025.114339>
- [18] Alexandersen, J., Sigmund, O., & Aage, N. (2016). Large scale three-dimensional topology optimisation of heat sinks cooled by natural convection. *International Journal of Heat and Mass Transfer*, 100, 876-891.
<https://doi.org/10.1016/j.ijheatmasstransfer.2016.05.013>
- [19] Zeng, S., Kanargi, B., & Lee, P. S. (2018). Experimental and numerical investigation of a mini channel forced air heat sink designed by topology optimization. *International Journal of Heat and Mass Transfer*, 121, 663–679.
<https://doi.org/10.1016/j.ijheatmasstransfer.2018.01.039>

- [20] Chen, F., Wang, J., & Yang, X. (2022). Topology optimization design and numerical analysis on cold plates for lithium-ion battery thermal management. *International Journal of Heat and Mass Transfer*, 183, 122087.
<https://doi.org/10.1016/j.ijheatmasstransfer.2021.122087>
- [21] Nguyen, N. P., Maghsoudi, E., Roberts, S. N., & Kwon, B. (2023). Shape optimization of pin fin array in a cooling channel using genetic algorithm and machine learning. *International Journal of Heat and Mass Transfer*, 202, 123769.
<https://doi.org/10.1016/j.ijheatmasstransfer.2022.123769>
- [22] Holman, J. P. (2002). Heat transfer (9th ed., pp. 600–606). McGraw-Hill.
- [23] Incropera, F. P., DeWitt, D. P., Bergman, T. L., & Lavine, A. S. (2007). Fundamentals of heat and mass transfer (6th ed., pp. 941–950). John Wiley & Sons.
- [24] MathWorks. (n.d.). *patternsearch* — Find minimum of function using pattern search. MATLAB Documentation (Global Optimization Toolbox).
<https://www.mathworks.com/help/gads/patternsearch.html>
- [25] Kolda, T. G., Lewis, R. M., & Torczon, V. (2003). Optimization by direct search: New perspectives on some classical and modern methods. *SIAM Review*, 45(3), 385–482.
<https://doi.org/10.1137/S003614450242889>
- [26] MathWorks. (n.d.). *sobolset* — Create a Sobol sequence set. MATLAB Documentation (Statistics and Machine Learning Toolbox).
<https://www.mathworks.com/help/stats/sobolset.html>
- [27] Sobol', I. M. (1967). On the distribution of points in a cube and the approximate evaluation of integrals. *U.S.S.R. Computational Mathematics and Mathematical Physics*, 7(4), 86–112.
[https://doi.org/10.1016/0041-5553\(67\)90144-9](https://doi.org/10.1016/0041-5553(67)90144-9)
- [28] MathWorks. (n.d.). *feedforwardnet* — Generate feedforward neural network. MATLAB Documentation (Deep Learning Toolbox).
<https://www.mathworks.com/help/deeplearning/ref/feedforwardnet.html>
- [29] Sun, G., & Wang, S. (2019). A review of the artificial neural network surrogate modeling in aerodynamic design. *Proceedings of the Institution of Mechanical Engineers, Part G: Journal of Aerospace Engineering*, 233(16), 5863–5872.
<https://doi.org/10.1177/0954410019864485>
- [30] MathWorks. (n.d.). *gamultiobj* — Find Pareto front of multiple fitness functions using genetic algorithm. MATLAB Documentation (Global Optimization Toolbox).
<https://www.mathworks.com/help/gads/gamultiobj.html>

[31] Deb, K., Pratap, A., Agarwal, S., & Meyarivan, T. (2002). A fast and elitist multi-objective genetic algorithm: NSGA-II. *IEEE Transactions on Evolutionary Computation*, 6(2), 182–197.

<https://doi.org/10.1109/4235.996017>

[32] Acampora, G., Kaymak, U., Loia, V., & Vitiello, A. (2013). Applying NSGA-II for solving the ontology alignment problem. In *2013 IEEE International Conference on Systems, Man, and Cybernetics* (pp. 1098–1103). IEEE.

<https://doi.org/10.1109/SMC.2013.191>