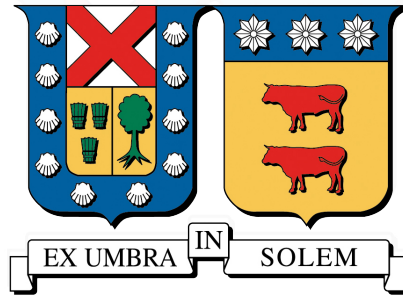


UNIVERSIDAD TÉCNICA FEDERICO SANTA MARÍA
DEPARTAMENTO DE MATEMÁTICA
VALPARAÍSO - CHILE



***Búsqueda de criterios para el dimensionamiento
óptimos de Sistemas Solares Térmicos mediante
aproximación con máquinas de aprendizaje***

***Requisito para obtener el título de Ingeniero Civil
Matemático***

***Ricardo Elías Ávalos Rojas
Profesor guía: Carlos Valle Vidal
Profesor correferente: Francisco Cuevas Pacheco***

Noviembre 2025



CONSTANCIA DE VALIDACIÓN Y CONFIDENCIALIDAD DE MONOGRAFÍA A REPOSITORIO ACADÉMICO

1.- IDENTIFICACIÓN DEL TRABAJO ACADÉMICO

Tipo de monografía (marcar una opción): ☒ Memoria o trabajo de título ☐ Tesis de Postgrado

Título del trabajo: Búsqueda de criterios para el dimensionamiento óptimos de Sistemas Solares Térmicos mediante aproximación con máquinas de aprendizaje.

Nombre del candidato(a): Ricardo Elías Ávalos rojas.

Carrera / Grado: Ingeniero civil matemático.

Campus: Casa Central **Departamento:** Matemáticas.

2.- VALIDACIÓN DEL PROFESOR GUÍA/DIRECTOR DE TESIS

Yo, Carlos Valle Vidal, en mi calidad de profesor(a) guía/director(a) del trabajo académico mencionado anteriormente

DEJO CONSTANCIA que:

- He revisado esta versión del documento y corresponde a la versión final aprobada del trabajo.
- El trabajo cumple con los requisitos académicos y de formato establecidos por la institución.

3.- EVALUACIÓN DE CONFIDENCIALIDAD POR PROPIEDAD INDUSTRIAL (marcar una opción)

☒ El trabajo **NO contiene** información que amerite confidencialidad y puede ser publicado de inmediato en repositorio con acceso abierto.

☐ El trabajo **CONTIENE** información con potenciales implicancias de propiedad industrial o intelectual y requiere un periodo de confidencialidad (**embargo**) por (**marcar una opción**):

☐ 6 meses ☐ 12 meses ☐ 2 años ☐ 3 años ☐ 5 años ☐ 10 años

Fundamentación de la necesidad de confidencialidad (obligatorio si se solicita embargo):

4.- FIRMAS

Profesor(a) guía o director(a) de memoria o tesis:

Fecha: 12-12-2025

Firma:

Estudiante o Candidato(a):

Fecha: 10-12-2025

Firma:

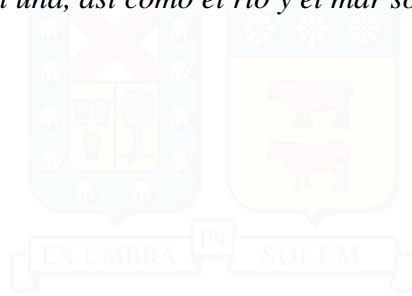
Este formulario debe ser insertado como página 2 de la memoria o tesis, completado y firmado por estudiante y profesor(a) antes de la entrega en portal PRISMA de Biblioteca USM.

Índice de Contenidos

Capítulo 1 Introducción	2
Capítulo 2 Propuesta	4
2.1 Problema	4
2.2 Propuesta de solución	6
Capítulo 3 Antecedentes	7
3.1 Modelamiento de Sistemas Solares Térmicos	7
3.2 Estado del arte	13
3.3 Objetivos del problema	14
Capítulo 4 Metodología	16
4.1 Generación de base de datos de simulaciones	16
4.2 Preprocesamiento y análisis de los datos	19
4.3 Algoritmos a utilizar	20
4.4 Entrenamiento de modelos	25
4.5 Búsqueda de grilla para hiperparámetros	26
Capítulo 5 Resultados	27
5.1 Desempeño de los algoritmos	27
5.1.1 Fracción solar	28
5.1.2 VAN	28
5.2 Búsqueda de configuraciones óptimas	29
5.2.1 Fracción solar	30
5.2.2 VAN	30
5.2.3 Expansión del mado	31
Capítulo 6 Conclusiones	33
Capítulo 7 Anexo	35
Bibliografía	45

“Porque la vida y la muerte son una, así como el río y el mar son uno también.”

— *Khalil Gibran*



A mis padres.

1 Introducción

La creciente preocupación por el cambio climático y la necesidad de transitar hacia un uso más eficiente y sostenible de la energía han impulsado el desarrollo e implementación de tecnologías basadas en energías renovables. Entre ellas, los Sistemas Solares Térmicos (SST) se han posicionado como una alternativa eficiente y versátil para la generación de energía térmica, como se puede notar en [Jaisankar, Ananth, Thulasi, Jayasuthakar, and Sheeba \[2011\]](#), donde se hace un recuento sobre los avances para aumentar las eficiencias térmicas de los SST especialmente en aplicaciones de calentamiento de agua tanto a nivel residencial como industrial. Su capacidad para reducir el consumo de combustibles fósiles y, en consecuencia, las emisiones de gases de efecto invernadero, los convierte en una herramienta clave para el cumplimiento de los objetivos medioambientales y de eficiencia energética a nivel global y nacional, ya que pueden ser una alternativa real para demandas térmicas para la minería, como se estudió en [Quiñones, Felbol, Valenzuela, Cardemil, and Escobar \[2020\]](#).

En el caso de Chile, la diversidad climática y geográfica ofrece un potencial considerable para el aprovechamiento de la energía solar. Sin embargo, la correcta implementación de un SST requiere de un dimensionamiento óptimo que considere simultáneamente criterios energéticos y económicos. Este proceso se ve condicionado por la alta variabilidad de parámetros como la radiación solar, la temperatura ambiental, el volumen de consumo y las características técnicas de los componentes del sistema. Determinar las configuraciones que maximizan la eficiencia y la rentabilidad del sistema es, por tanto, un desafío relevante. Unos casos estudiados se pueden encontrar en [Fuentes, Pailahueque, Muñoz, Escobar, and Cardemil \[2024\]](#) donde se presentan distintos tipos de colectores a partir de casos reales de implementaciones en regiones del centro y sur de Chile.

Los métodos tradicionales de modelación matemática permiten estimar con precisión el desempeño de un SST y evaluar configuraciones específicas, pero presentan como principal desventaja altos costos computacionales. Cada simulación involucra un modelo iterativo que resuelve ecuaciones de equilibrio térmico a partir de series de tiempo meteorológicas y de operación, con tiempos de ejecución que pueden superar los seis segundos por simulación. Esto dificulta la exploración exhaustiva de combinaciones posibles de parámetros, especialmente cuando se busca optimizar el diseño del SST en su rendimiento, pero no se conoce cómo hacerlo y se cuenta con un espacio de búsqueda amplio.

Ante esta limitación, se han explorado alternativas que permitan reducir el tiempo de cálculo sin sacrificar significativamente la precisión de los resultados. Una de ellas es el uso de modelos de aprendizaje automático (Machine Learning) capaces de aproximar la función latente que relaciona las variables de diseño y las métricas de desempeño (fracción solar y Valor Actual Neto (VAN)). Estos modelos pueden entrenarse con una base de datos generada a partir de simulaciones, de modo que, una vez ajustados, realicen predicciones en tiempos mucho menores.

Este trabajo propone y evalúa el uso de algoritmos de aprendizaje automático —incluyendo máquinas de vectores de soporte (SVM), árboles de regresión y redes neuronales— como herramientas para predecir la fracción solar y el VAN de SST en distintas configuraciones y condiciones de operación. El objetivo es encontrar configuraciones cercanas a las óptimas en términos energéticos y económicos, reduciendo de forma significativa el tiempo computacional respecto a la modelación matemática convencional.

Para ello, se construyó una base de datos de más de un millón de simulaciones representando diferentes combinaciones de parámetros de diseño, tipos de colectores y ubicaciones geográficas en Chile. Esta base permitió entrenar y evaluar el desempeño de los modelos propuestos, comparando sus resultados con los obtenidos mediante la modelación teórica tradicional. Adicionalmente, se estudió su capacidad para identificar configuraciones óptimas dentro y fuera del conjunto de entrenamiento, analizando sus limitaciones y potencial de mejora.

En síntesis, este trabajo busca contribuir al desarrollo de metodologías más eficientes para el dimensionamiento de SST, combinando el rigor de la modelación física con la rapidez y flexibilidad de las técnicas de aprendizaje automático, con miras a facilitar la implementación de proyectos más rentables y sostenibles en el contexto chileno.

En el Capítulo 2 se define el problema de optimización, especificando los criterios de dimensionamiento óptimo —maximización del VAN y de la fracción solar— y las restricciones del sistema. Se propone el uso de modelos de aprendizaje automático como alternativa para reducir el costo computacional del modelo iterativo tradicional.

En el Capítulo 3 se presenta el modelo matemático de un sistema solar térmico, detallando los principales componentes —colector solar, intercambiador de calor, estanque de acumulación y sistema auxiliar— y la forma en que interactúan. Asimismo, se introducen los indicadores de evaluación económica y energética, como el Valor Actual Neto (VAN) y la fracción solar.

El Capítulo 4 describe el procedimiento para la generación de la base de datos de simulaciones que sirve de insumo a los modelos de aprendizaje automático. Se incluyen los parámetros de diseño, la selección de localidades representativas a lo largo de Chile y el preprocesamiento aplicado a los datos. También se presentan los algoritmos de aprendizaje considerados y la metodología de entrenamiento empleada.

El Capítulo 5 se centra en la evaluación del desempeño de los algoritmos de aprendizaje automático seleccionados, comparando su precisión y tiempos de ejecución en la predicción de la fracción solar y del VAN. Además, se analizan sus capacidades de generalización a localidades no incluidas en el entrenamiento y a configuraciones fuera del rango simulado, con especial atención a la identificación de configuraciones óptimas.

Finalmente, el Capítulo 6 presenta las conclusiones del trabajo y las líneas de investigación futura. Se discute la relevancia de los resultados obtenidos para la planificación y diseño de proyectos solares térmicos en Chile, así como las posibles mejoras metodológicas.

2 Propuesta

El diseño de sistemas solares térmicos (SST) para aplicaciones industriales constituye un desafío técnico y económico de gran relevancia en el contexto energético chileno. La transición hacia fuentes renovables, impulsada tanto por compromisos internacionales como por la necesidad local de reducir costos y emisiones, requiere metodologías que permitan dimensionar estos sistemas de forma eficiente. En particular, los proyectos deben equilibrar criterios económicos —como la rentabilidad de la inversión— y criterios técnicos —como la cobertura de la demanda energética—, de manera que el resultado final se traduzca en soluciones factibles y sostenibles para distintos sectores productivos.

Chile presenta condiciones especialmente favorables para el uso de SST debido a su variada geografía y altos niveles de radiación solar en gran parte del territorio. Sin embargo, esta misma diversidad climática genera una complejidad adicional: el desempeño de un sistema depende fuertemente de las condiciones meteorológicas locales, de la disponibilidad de combustibles auxiliares y de los costos asociados a los equipos. En consecuencia, las estrategias de dimensionamiento deben considerar no solo las variables de diseño propias del sistema —como el área colectora o el volumen de almacenamiento—, sino también la influencia de parámetros externos que pueden variar significativamente entre regiones.

La formulación matemática del problema busca capturar esta interacción entre diseño, costos y desempeño energético. Sin embargo, el carácter iterativo de los modelos de simulación empleados, junto con la elevada variabilidad temporal de los datos meteorológicos, conduce a una complejidad computacional considerable. Este aspecto limita el uso de técnicas analíticas de optimización y obliga a explorar enfoques alternativos, como la búsqueda exhaustiva, los métodos heurísticos o, más recientemente, el empleo de modelos de aprendizaje automático capaces de aproximar el comportamiento térmico de los sistemas.

En este capítulo se plantea el problema central de investigación: definir y resolver criterios de dimensionamiento óptimo para proyectos de SST en demandas industriales a lo largo de Chile. Para ello, se consideran dos perspectivas complementarias —la maximización del retorno económico y la maximización de la eficiencia energética—, con el fin de proporcionar un marco metodológico que permita evaluar, comparar y seleccionar configuraciones de diseño robustas y adaptadas a las condiciones del país.

2.1. Problema

El problema de interés consiste en determinar dos criterios de dimensionamiento óptimo para proyectos de Sistemas Solares Térmicos (SST) destinados a demandas industriales en todo Chile,

empleando la modelación matemática presentada en la Sección 3.1.

Los criterios de dimensionamiento óptimo se evalúan de manera independiente, es decir, cada uno se optimiza por separado. Estos son:

- **Retorno de la inversión:** consiste en maximizar el Valor Actual Neto (VAN) del proyecto, definido en la ecuación (3.3). De manera equivalente, puede interpretarse como la minimización del período de retorno de la inversión, considerando que los costos de mantención se suponen constantes.
- **Eficiencia energética:** definida como el porcentaje de la demanda de agua caliente cubierta por el SST. En este caso, se busca maximizar la fracción solar (F), expresada en la ecuación (3.4).

La búsqueda de estos óptimos se realiza considerando las principales variables de diseño y dimensionamiento de un SST, que a su vez constituyen las variables de decisión del problema:

- Volumen del estanque de almacenamiento (V).
- Área colectora del colector solar (A).

Estas variables están sujetas a parámetros determinados por la ubicación del proyecto y por las opciones disponibles de colectores y estanques:

- Serie de tiempo de radiación solar correspondiente a un año meteorológico típico (glb).
- Serie de tiempo de la temperatura ambiente (T^{amb}).
- Volumen de consumo mensual promedio de agua caliente sanitaria (V^c).
- Temperatura de consumo (T^c).
- Precio del combustible auxiliar (C^c).
- Costo por metro cuadrado de área colectora (C^A).
- Costo por metro cúbico de volumen de almacenamiento (C^V).
- Serie de tiempo de sombra (S).
- Eficiencia óptica del colector (τ).
- Pérdida térmica de las placas (U_l).

Estos parámetros deben entenderse como parámetros fijos en la evaluación de un proyecto SST. Si bien pueden variar entre distintas ubicaciones —por ejemplo, debido a diferencias en la radiación solar—, no tendría sentido buscar una serie de tiempo “óptima” para un lugar en el que no existan las condiciones logísticas necesarias para consumir el agua caliente generada.

El problema considera las condiciones meteorológicas de todo el territorio chileno, disponiendo de datos de radiación solar y temperatura ambiental para un año meteorológico típico, obtenidos a partir de la latitud y longitud del emplazamiento. Esta información se consulta en las bases de datos del Ministerio de Energía de Chile, las cuales poseen una resolución espacial de un kilómetro. Dado que el país presenta una gran diversidad y variabilidad climática, esta característica es relevante para la modelación. Para más detalles sobre la base de datos meteorológica, véase [Molina, Falvey, and Rondanelli \[2017\]](#).

Es importante considerar que la simulación de los modelos matemáticos descritos en la Sección 3.1 debe realizarse mediante un modelo iterativo que aproxima las ecuaciones de equilibrio térmico involucradas a partir de las series de tiempo. La cantidad de cálculos y el nivel de interconexión entre ellos generan una alta complejidad, lo que imposibilita determinar la dependencia de la función objetivo con respecto a las variables de interés de forma analítica. En otras palabras, las variables a optimizar no presentan una relación funcional explícita con las variables de decisión; no existe una fórmula cerrada distinta al propio modelo iterativo que permita calcular esta dependencia. Por lo tanto, no es factible aplicar técnicas clásicas de optimización.

Las alternativas más evidentes corresponden a métodos de búsqueda exhaustiva o a técnicas de naturaleza heurística. Es decir, ante la ausencia de una expresión cerrada para el VAN y la fracción solar en función de las variables de decisión que permita encontrar los óptimos directamente, se debe optar por explorar todas las combinaciones posibles (método exhaustivo) o bien por utilizar estrategias heurísticas que orienten la búsqueda hacia combinaciones con mayor potencial. En ambos casos, el número de simulaciones requeridas es elevado. Considerando que el tiempo promedio de una simulación individual es de 6,0919 segundos y que existe una alta variabilidad en las series de tiempo, alcanzar una aproximación en tiempos computacionales razonables resulta inviable.

Una alternativa para mitigar esta limitación y reducir los costos computacionales consiste en reemplazar la etapa de modelación directa del SST por un modelo de aprendizaje automático capaz de estimar de forma aproximada su desempeño, evitando así el elevado costo de simulaciones repetidas. Tal como se presenta en Li, Liu, Liu, Zhang, et al. [2017], es posible emplear redes neuronales para predecir el rendimiento de una configuración de SST, en Ahmad, Reynolds, and Rezgui [2018] se realiza una comparación entre árboles de regresión, SVM y random forest, dando como resultados que podrían realizar una predicción horaria de los desempeños, lo cual da pie a pensar en la factibilidad de utilizarlos para las predicciones en este contexto. De esta manera, la optimización (ya sea mediante métodos exhaustivos o heurísticos) puede realizarse en tiempos razonables, obteniendo configuraciones cercanas a los óptimos. Posteriormente, el candidato seleccionado por el modelo de aprendizaje se evalúa mediante la modelación detallada del SST, permitiendo describir con precisión sus características.

2.2. Propuesta de solución

Se propone utilizar una o más máquinas de aprendizaje que puedan reemplazar la modelación matemática de los SST, permitiendo obtener el VAN (3.3) y la fracción solar (3.4) a partir de las configuraciones y parámetros de diseño de un SST, manteniendo un error cuadrático medio aceptable a la vez, que se consigue reducir significativamente los tiempos computacionales, considerando que la modelación matemática actual demora en promedio 6.0919 segundos. Posteriormente, se realizaría una exploración de diversas configuraciones para identificar aquellas que mejor resuelvan los problemas (3.5) y (3.6).

3 Antecedentes

En este capítulo se desarrolla un modelo matemático para Sistemas Solares Térmicos (SST), detallando su formulación, los supuestos considerados y los resultados que se esperan de su implementación. Este trabajo se fundamenta en la metodología presentada por Olavarría [2013], actualmente incorporada en el sitio web del Ministerio de Energía, Explorador Solar, en la sección “Sistemas Solares Térmicos para Demandas Industriales (ACS)”. La modelación constituye la base necesaria para abordar el problema central de esta tesis: identificar configuraciones óptimas de SST tanto en términos financieros como energéticos.

Además de describir la modelación, se presenta una revisión del estado del arte que permite situar este trabajo dentro del marco de la investigación previa, destacando coincidencias metodológicas y señalando las principales diferencias respecto de otros enfoques. Finalmente, se formulan los objetivos específicos que guiarán la resolución del problema de dimensionamiento, estableciendo así un puente entre la modelación física del sistema y las técnicas de optimización y aprendizaje automático que se aplicarán en los capítulos siguientes.

3.1. Modelamiento de Sistemas Solares Térmicos

Un sistema solar térmico es un conjunto de dispositivos y tecnologías diseñados para capturar y aprovechar la energía solar, transformándola en energía térmica utilizable en procesos de calentamiento. A diferencia de los paneles solares fotovoltaicos, que convierten la luz solar directamente en electricidad, los sistemas solares térmicos utilizan el calor del sol para elevar la temperatura de fluidos. Estos sistemas tienen aplicaciones diversas, incluyendo usos residenciales, industriales y comerciales, para el calentamiento de agua y otros propósitos.

Este Capítulo presenta un diagrama de los componentes que constituyen un Sistema Solar Térmico (SST), compuesto por cuatro módulos principales:

1. El primer módulo es el **colector solar**, que captura la energía solar y la convierte en calor. Existen diversos tipos de colectores como los colectores de placa plana, de tubos de vacío, parabólicos cilíndricos y fresnel lineal como son presentados en Fuentes et al. [2024].
2. El segundo módulo es el **intercambiador de calor**, que facilita la transferencia del calor del fluido caloportador al agua u otro fluido que se desee calentar.
3. El tercer módulo es el **estanque de almacenamiento de agua caliente sanitaria**, que almacena el agua caliente producida, manteniéndola disponible hasta su uso.
4. El cuarto módulo es el **sistema auxiliar**, que proporciona el calor adicional necesario para alcanzar la temperatura deseada del agua, usualmente mediante un calentador a gas o un

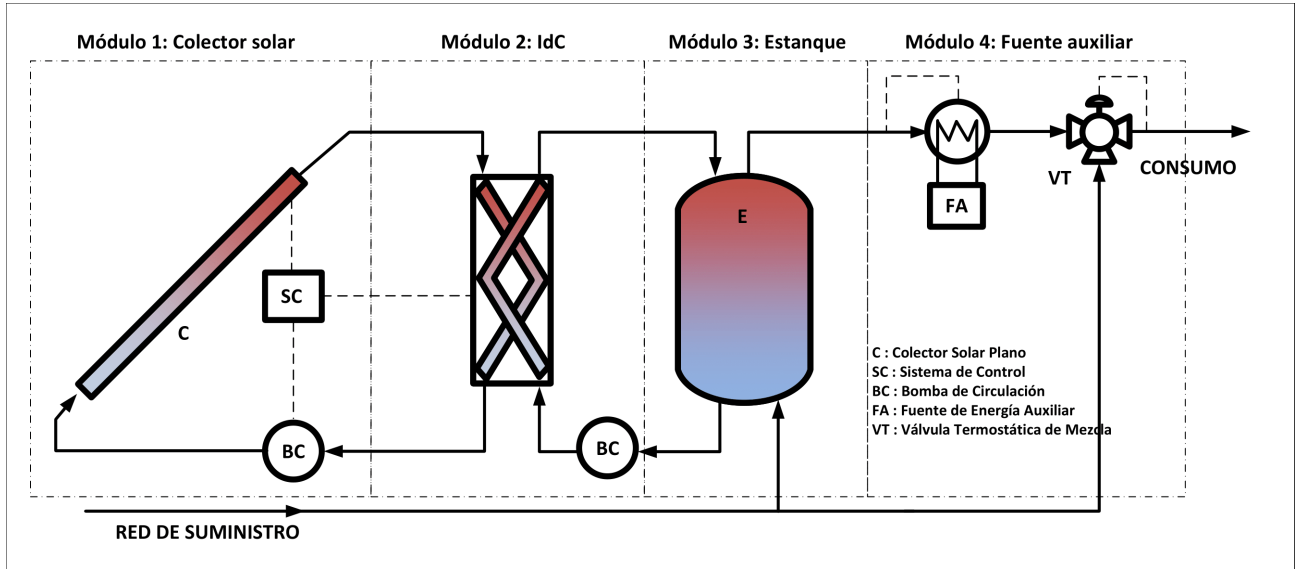


Figura 3.1: Sistema Solar Térmico industrial

calentador eléctrico.

Los cuatro módulos descritos y representados en la Figura 3.1 se han modelado matemáticamente de diversas maneras, dependiendo de las especificaciones y necesidades del sistema.

En esta sección se detallará cada uno de estos módulos, su funcionamiento y la interacción entre ellos dentro del sistema global.

1. **El colector solar:** El colector se modela como una función que calcula su temperatura de salida para una hora específica del año, a partir de tres series de tiempo: la radiación solar global normal (GLB), la temperatura ambiental y la temperatura de entrada proveniente del IDC, esta última calculada de manera horaria. Para este cálculo se utilizan los siguientes parámetros:

$A :=$ Área del colector , $A \in \mathbb{R}^+$.

$\tau :=$ Eficiencia óptica del colector , $\tau \in \mathbb{R}^+$.

$U_l :=$ Pérdida de las placas , $U_l \in \mathbb{R}^+$.

$\dot{m}_{\text{heat}} :=$ Flujo del colector , $\dot{m}_{\text{heat}} \in \mathbb{R}^+$.

$cp :=$ Calor específico del fluido a calentar , $cp \in \mathbb{R}^+$.

$T^{\text{out colector}} :=$ Temperatura de salida hacia el colector , $T^{\text{out colector}} \in (\mathbb{R}^+)^{8760}$.

$T^{\text{in colector}} :=$ Temperatura de entrada desde el colector, en el instante t , $T^{\text{in colector}} \in (\mathbb{R}^+)^{8760}$.

$T^{\text{amb}} :=$ Temperatura ambiental , $T^{\text{amb}} \in (\mathbb{R}^+)^{8760}$.

$glb :=$ Radiación global normal , $glb \in (\mathbb{R}_0^+)^{8760}$.

$S :=$ Sombra , $S \in [0, 1]$.

La ecuación que determina la temperatura de salida del colector solar es la siguiente:

$$T_t^{\text{out colector}} = \begin{cases} \frac{(1-S_t)A(glb_t \cdot \tau - U_f(T_t^{\text{in colector}} - T_t^{\text{amb}}))}{\dot{m}_{\text{heat}} \cdot c_p + T_{t-1}^{\text{in colector}}} & , \quad glb_t > 0, 1 \leq t \leq 8760 \\ T_{t-1}^{\text{in colector}} & , \quad glb_t = 0, 1 \leq t \leq 8760 \end{cases}.$$

2. **Intercambiador de calor:** El intercambiador de calor facilita la transferencia térmica entre la temperatura de salida del fluido desde colector solar y la temperatura de entrada del agua caliente sanitaria al estanque de almacenamiento. Para la modelación se introducen los siguientes parámetros. %Agregar kas nuevas.

$T^{\text{out idc}}$:= Temperatura de salida desde el intercambiador de calor hacia el estanque
 $, T^{\text{out idc}} \in (\mathbb{R}^+)^{8760}$.
 e := Eficiencia del intercambiador de calor, $e \in [0, 1]$.

Lo que da pie a las ecuaciones que modelan esta transferencia:

$$T_t^{\text{in colector}} := \max(T_t^{\text{amb}}, (1 - e)T_t^{\text{out colector}}), \quad (3.1)$$

$$T_t^{\text{out idc}} := eT_t^{\text{out colector}}, \quad (3.2)$$

donde, la ecuación (3.1) toma en cuenta la pérdida de calor del colector solar al entrar al intercambiador de calor (IDC), donde pierde un porcentaje de su temperatura, determinado por la eficiencia del IDC. La nueva temperatura de entrada al colector se establece como el mayor valor entre la temperatura reducida y la temperatura ambiente. Esta temperatura modificada es la que se reintroduce al flujo del colector solar. Posteriormente, el calor resultante de esta operación es transferido al estanque de almacenamiento, como se detalla en la ecuación (3.2).

3. **Estanque de acumulación de agua:** En un sistema solar térmico, un estanque estratificado de acumulación de agua es un dispositivo diseñado para almacenar agua caliente de manera eficiente, aprovechando las diferencias naturales de densidad originadas por las variaciones de temperatura. Este tipo de estanque resulta fundamental en muchos sistemas solares térmicos, ya que permite maximizar la eficiencia y el aprovechamiento energético.

El agua caliente, al ser menos densa que el agua fría, tiende a ascender, mientras que el agua fría desciende. En un estanque estratificado, este fenómeno genera capas o estratos de agua a distintas temperaturas: el agua más caliente se ubica en la parte superior y la más fría en el fondo. En Garnier, Currie, and Muneer [2009] se presenta un trabajo sobre la modelación de estanques estratificados y posteriormente en la Sección del Estado Del Arte 3.2 se comenta sobre los efectos que pueden tener utilizar estanques estratificados sobre estanques totalmente mezclados.

Como se muestra en la Figura 3.1, el tanque está diseñado para favorecer esta estratificación natural. Su forma es vertical y cuenta con aislamiento térmico para minimizar las pérdidas de calor. Además, las tomas de agua se ubican estratégicamente a diferentes alturas. El

agua caliente proveniente del intercambiador de calor solar ingresa por la parte superior, mientras que el agua fría que requiere calentamiento entra por la parte inferior.

La extracción de agua caliente sanitaria se realiza desde la parte superior del estanque, donde se encuentra a la temperatura más alta, lo que permite su uso a la máxima temperatura disponible.

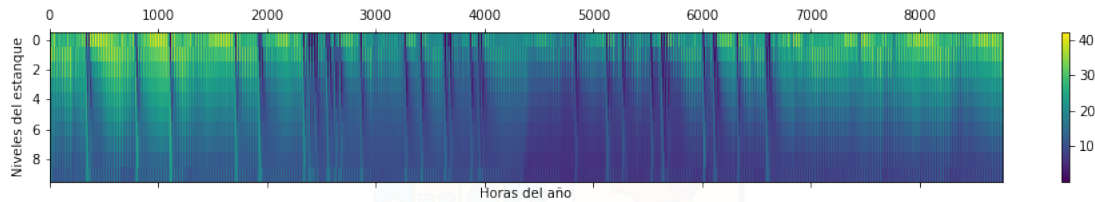


Figura 3.2: Ejemplo de estanque simulado

El estanque de acumulación se modeló siguiendo el trabajo de Kleinbach, Beckman, and Klein [1993], en el cual se propone un modelo para un estanque estratificado. Este cuenta con un volumen total y se divide en una cierta cantidad de nodos horizontales que segmentan el estanque en discos imaginarios, con el fin de realizar un balance energético en cada nodo.

Como se observa en la Figura 3.3, este esquema requiere la definición y el uso de algunos parámetros, los cuales se presentan a continuación.

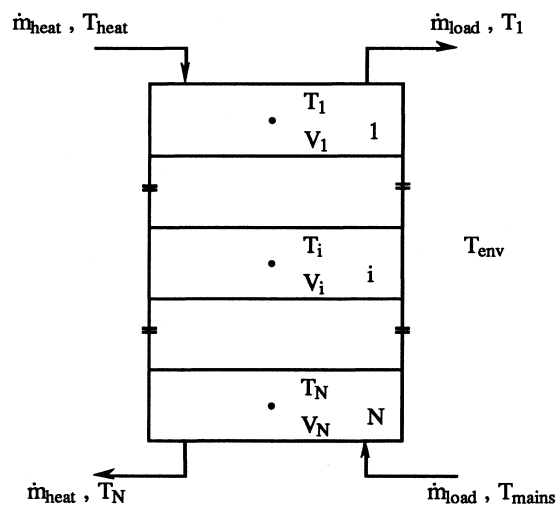


Figura 3.3: Estanque de acumulación

V := Volumen de estanque de almacenamiento.

V_i := Volumen del fluido en el nodo i .

M_i := Masa del fluido en el nodo i .

C_f := Calor específico del agua.

T_i := Temperatura en el nodo i .

$\alpha_i := \begin{cases} 1, & \text{Si fluido desde la fuente de calor entra en el nodo} \\ 0, & \text{En otro caso} \end{cases}$

T_i := Temperatura en el i -ésimo nodo.

$\beta_i := \begin{cases} 1, & \text{Si fluido que regresa de la carga entra en nodo } i \\ 0, & \text{en los otros casos} \end{cases}$

$\dot{m}_{\text{load}}$:= Caudal másico desde el estanque hacia el intercambiador de calor.

$$\gamma_i := \dot{m}_{\text{heat}} \sum_{j=1}^{i-1} \alpha_j - \dot{m}_{\text{load}} \sum_{j=i+1}^N \beta_j$$

$$\delta_i = \begin{cases} 1, & \gamma_i > 0 \\ 0, & \gamma_i \leq 0 \end{cases}$$

$$\varepsilon = \begin{cases} 1, & \text{Si el sistema auxiliar esta funcionando} \\ 0, & \text{En otro caso} \end{cases}$$

$\dot{Q}_{\text{aux},i}$:= Tasa de energía ingresada al nodo i desde el sistema auxiliar.

$UA_{\text{fl},i}$:= Coeficiente general de perdida de calor hacia la fuente auxiliar para el nodo i .

UA_i := Pérdida de energía entre el nodo i y el ambiente.

T_{env} := Temperatura ambiente.

Lo que da pie al siguiente balance energético.

$$\begin{aligned} M_i C_f \frac{dT_i}{dt} = & \alpha_i \dot{m}_{\text{heat}} C_f (T_{\text{heat}} - T_i) + \beta_i \dot{m}_{\text{load}} C_f (T_{\text{env}} - T_i) \\ & + \delta_i \gamma_i C_f (T_{i-1} - T_i) + (1 - \delta_i) \gamma_i C_f (T_i - T_{i+1}) \\ & - UA_{\text{fl},i} (T_i - T_{\text{fl}}) - UA_i (T_i - T_{\text{env}}). \end{aligned}$$

4. **El sistema auxiliar** suministra el calor necesario para que el agua extraída del estanque de almacenamiento alcance la temperatura requerida para su consumo. Para este fin, se emplean los siguientes parámetros:

Q^{demanda} := Calor consumido en la demanda de agua caliente,

Q^{estanque} := Calor entregado desde el estanque de almacenamiento.

El sistema auxiliar, da pie a una ecuación de balance energético simplificada.

$$Q_{\text{aux}} = \begin{cases} Q^{\text{demanda}} - Q^{\text{estanque}} & , \quad Q^{\text{demanda}} \geq Q^{\text{estanque}} \\ 0 & , \quad Q^{\text{demanda}} < Q^{\text{estanque}} \end{cases}.$$

Los módulos anteriores se complementan con la agregación de todos los costos asociados a las variables de cada uno, tales como el área del colector solar, el volumen del estanque de almacenamiento, los modelos empleados para ambos, la mano de obra de instalación, el tipo de sistema auxiliar y el combustible que este utiliza, además de los costos de mantención correspondientes al período de evaluación del proyecto.

La suma de estos costos permite realizar la evaluación del rendimiento económico de un proyecto SST. Para ello, se emplean dos indicadores ampliamente utilizados: el VAN y el año de retorno de la inversión (payback year) [Brealey, Myers, and Allen \[2014\]](#). Estos se consideran herramientas robustas para el análisis de rentabilidad de proyectos.

Para su formulación, es necesario considerar las siguientes variables:

M := Período de evaluación del proyecto en años.

C_0 := Inversión inicial del SST.

C_n := Flujos netos anuales producidos como resultados del ahorro en energía, $1 \leq n \leq M$.

r = Tasa de descuento.

Con estas variables se puede estimar mediante la siguiente relación.

$$\text{VAN}(C_0, \dots, C_n) = C_0 + \sum_{n=1}^M \frac{C_n}{(1+r)^n} \quad , \quad (3.3)$$

mientras que el año de retorno de la inversión es el primer $M \in \mathbb{N}$ en el cual el proyecto tiene VAN positivo.

De manera complementaria a los análisis económicos, se realiza un análisis simple sobre el rendimiento del SST que es determinar el porcentaje de la demanda cubierta, llamada fracción solar,

$$F = \frac{Q^{\text{estanque}}}{Q^{\text{demanda}}} \quad . \quad (3.4)$$

Esta variable es de interés pues se traduce tanto en una eficiencia del dimensionamiento del sistema, como en una estimación del ahorro de combustible generada, en otras palabras una medida de la reducción de la contaminación que conlleva un SST.

Algunas de las principales características de los sistemas solares térmicos son:

- a) Son eficientes energéticamente, ya que al contar con un SST bien diseñado se puede cubrir una gran parte de las necesidades de agua caliente sanitaria en un recinto, reduciendo significativamente el consumo de energía convencional.
- b) Reducen las emisiones, pues al disminuir la dependencia de combustibles fósiles para calentar el agua, se reducen las emisiones de efecto invernadero.
- c) Son atractivos económicamente, ya que aunque cuentan con una inversión inicial significativa, a largo plazo se generan ahorros en las facturas de electricidad o de gas, generando el retorno de la inversión, además de poder contar con diferentes ayudas y subvenciones para promover su instalación y utilización como una tecnología eco amigable.
- d) Son versátiles, ya que los SST se pueden integrar con otros sistemas de calentamiento para garantizar el consumo cuando el SST no pueda cubrir toda la demanda.
- e) Tienen una vida útil, ya que con un mantenimiento adecuado, un sistema solar térmico puede tener una vida útil de 20 a 25 años o más.
- f) Son autónomos, debido a que en algunos contextos donde hay dificultades logísticas para abastecer o utilizar ciertos combustibles, los SST cuentan con una ventaja al utilizar la energía renovable, inagotable, gratuita del sol. Como en el caso de la minería de alta altura.

Por ello, un sistema solar térmico representa una opción altamente eficiente, económica y sostenible para cubrir las necesidades de calentamiento de agua y otros fluidos. Su adopción contribuye de manera significativa a la lucha contra el cambio climático y la dependencia de fuentes de energía no renovables.

Con las definiciones de las ecuaciones (3.3) y (3.4) se pueden formalizar los problemas a abordar de la siguiente forma.

$$\begin{aligned} \text{maximizar} \quad & \text{VAN}(A, V) = C_0(A, V, C^A, C^V) + \sum_{n=1}^M \frac{C_n(A, V, \text{glb}, T^{\text{amb}}, T^c, C^c)}{(1+r)^n} \\ \text{sujeto a} \quad & A, V \in \mathbb{R}^+, \end{aligned} \quad (3.5)$$

$$\begin{aligned} \text{maximizar} \quad & F(A, V) = \frac{Q^{\text{estaque}}(A, V, \text{glb}, T^{\text{amb}}, T^c)}{Q^{\text{demanda}}(T^{\text{amb}}, T^c)}, \\ \text{sujeto a} \quad & A, V \in \mathbb{R}^+. \end{aligned} \quad (3.6)$$

3.2. Estado del arte

En Klein, Beckman, and Duffie [1976] se presenta un método de dimensionamiento de sistemas solares térmicos que vincula los componentes del sistema con su desempeño mediante la fracción solar, denominado F-chart. Posteriormente, Klein y Beckman generalizan este enfoque y

proponen el método ϕ -F-chart [Klein and Beckman \[1979\]](#). Ambos métodos se validan a partir de simulaciones con TRNSYS y permiten estimar la fracción solar, reduciendo así el costo computacional de las simulaciones. Estos desarrollos son utilizados en [Kulkarni, Kedare, and Bandyopadhyay \[2007\]](#) para proponer una metodología de diseño y dimensionamiento óptimo basada en relaciones entre curvas de desempeño. Esta línea de trabajo es afín a la de esta tesis; sin embargo, los métodos F-chart y ϕ -F-chart suponen un estanque de almacenamiento completamente mezclado. Aunque en [Klein and Beckman \[1979\]](#) se señala que un estanque estratificado no debería diferir en más de un 3 % respecto de uno totalmente mezclado, evidencia posterior muestra mejoras del orden del 20 % a favor de la estratificación [Kleinbach et al. \[1993\]](#). Avances recientes sobre estratificación, incluidos experimentos, diseños y factores que la influyen, pueden encontrarse en [Bouhal, Fertahi, Agrouaz, El Rhafiki, Kousksou, and Jamil \[2017\]](#).

El presente trabajo propone un método de diseño para sistemas solares térmicos con estanques estratificados; por ello, las metodologías mencionadas (aunque útiles) no pueden reproducirse directamente. La discontinuidad de temperatura entre estratos impide utilizar las curvas caracterizadas en [Klein and Beckman \[1979\]](#), como se hace en [Kulkarni et al. \[2007\]](#). En consecuencia, y siguiendo las ideas expuestas en la literatura citada, se opta por reemplazar el modelo iterativo descrito en el capítulo por un método de menor costo computacional y error aceptable al estimar el desempeño de las variables y de las funciones objetivo a optimizar en (3.5) y (3.6). Asimismo, se consideran los aportes de [Li et al. \[2017\]](#) y [Ahmad et al. \[2018\]](#) para sustituir el modelo iterativo mediante una o más máquinas de aprendizaje capaces de aproximar la función latente que relaciona el desempeño de los SST (fracción solar en los casos citados) con sus componentes de diseño.

En cuanto al VAN, no se identificó literatura que prediga directamente esta magnitud para SST de agua caliente sanitaria en demanda industrial. No obstante, existen trabajos como [AIT LAHOUSSE OUALI, TOULI, ALAMI MERROUNI, and MOUKHTAR \[2024\]](#) que, mediante redes neuronales, estiman el desempeño económico a partir de una métrica nivelada (levelized cost of heat, LCoH) en plantas de concentración solar. Si bien dichas plantas comparten el principio de generación de calor a partir de la radiación solar, emplean tecnologías y modelos distintos y responden a escenarios diferentes; con todo, tales estudios motivan el enfoque de predecir el VAN a partir de variables y parámetros de diseño. Ya que este sería el primer trabajo en esa línea con SST para agua caliente sanitaria.

Con estos antecedentes, es posible explorar el espacio de diseño para identificar variables candidatas a óptimo de las funciones objetivo y, posteriormente, validar dichas configuraciones mediante el modelo iterativo de alta fidelidad, evaluando, además, el desempeño de sus vecindarios en el espacio de soluciones.

3.3. Objetivos del problema

Los objetivos de este trabajo son:

1. Conseguir resultados similares al modelamiento presentado en la sección 3.1 para la fracción solar y el VAN, bajo la métrica de error cuadrático medio para medir la aproximación entre resultados.

2. Conseguir resultados en menores tiempos computacionales que la modelación presentada en la sección 3.1 cuyo tiempo computacional es de 6.0919 segundos de media por simulación.
3. Conseguir aproximaciones a las configuraciones óptimas que resuelven los problemas (3.5) y (3.6)



4 Metodología

En este capítulo se describe la construcción de la base de datos de simulaciones que sustenta la estrategia propuesta para el dimensionamiento óptimo de sistemas solares térmicos (SST) en Chile. La base se elaboró considerando 20 ciudades representativas de la diversidad climática y de radiación solar del país, junto con seis modelos de colectores aprobados por la Subsecretaría de Energía y Combustibles (SEC). Se generaron combinaciones de parámetros de diseño —como volumen de estanque, área colectora, temperatura de consumo y costo del combustible auxiliar— bajo restricciones de estabilidad del modelo, alcanzando un total de 1 231 150 simulaciones. Para su ejecución, se aprovechó procesamiento paralelo en un equipo de alto rendimiento, lo que permitió reducir el tiempo de construcción de 85 a 20 días.

Posteriormente, se aplicaron técnicas de preprocesamiento, destacando la estandarización mediante z-score, con el fin de garantizar un adecuado comportamiento de los algoritmos de aprendizaje automático. Asimismo, se analizaron las dependencias entre las variables objetivo (VAN y fracción solar), concluyéndose que deben ser predichas de manera independiente.

El capítulo introduce también los algoritmos seleccionados para la predicción: el método F-chart como línea base, máquinas de soporte vectorial (SVM) para regresión, árboles de regresión y redes neuronales. Cada técnica se explica desde un punto de vista conceptual y matemático, destacando sus fortalezas y limitaciones en el contexto de los SST.

Finalmente, se aborda el entrenamiento de modelos, diferenciando parámetros y hiperparámetros, y presentando las estrategias utilizadas para su ajuste, incluyendo validación cruzada y búsqueda en malla. Dado el gran número de combinaciones posibles en el ajuste de hiperparámetros, se implementó un enfoque de búsqueda aleatoria controlada, evaluando una alta cantidad de configuraciones por modelo y variable objetivo, lo que permitió balancear exhaustividad y eficiencia computacional.

En conjunto, este capítulo establece las bases metodológicas y computacionales necesarias para que los algoritmos de aprendizaje automático puedan aproximar el desempeño de los SST y servir como sustitutos del modelo de simulación directa en las etapas de optimización.

4.1. Generación de base de datos de simulaciones

Para la estrategia propuesta, es necesario construir una base de datos representativa de las diferentes configuraciones que pueden utilizarse en un sistema de colectores solares térmicos (SST), así como de lugares geográficos representativos de la radiación solar en Chile. Por esta razón, se utilizan veinte ciudades de Chile, que se encuentran en la Tabla 4.2. Estas ciudades provienen del trabajo de Olavarría [2013], en el cual se presenta un listado de ciudades representativas para el

estudio del potencial de energía solar térmica. De este listado se seleccionaron algunas ciudades y se agregaron otras nuevas en función de criterios de zonas de radiación en Chile y tras una discusión con un equipo multidisciplinario.

Para la construcción de la base de datos, se utilizan seis modelos de colectores solares aprobados por la Superintendencia de Electricidad y Combustibles (SEC), los cuales se presentan en la Tabla 4.1.

Tabla 4.1: Características de colectores solares SEC

Modelo de colector	Eficiencia óptica del colector	Precio en pesos m ²
1	81.1	333 007
2	64.4	321 867
3	61.2	310 727
4	76.3	299 587
5	76.5	288 447
6	69.1	277 307

En la base de datos se utilizan configuraciones a partir de los siguientes valores por parámetro de diseño, estos parámetros toman todas las combinaciones posibles a priori, pero deben cumplir la siguiente relación entre volumen de estanque y área colectora. Esto es para evitar inestabilidades y divergencias en el modelo de estanque presentado en la Sección 3.1.

$$40 < \frac{1000V}{A} < 180.$$

- **Volumen de estanque (m³):** 0.08, 0.20, 0.5, 0.8, 1, 3, 5, 8, 13.
- **Área colectora (m²):** 1.19, 4.76, 12, 30, 70, 100, 160, 325.
- **Temperatura de consumo (°C):** 40, 50, 60, 70.
- **Modelo de colector:** 1, 2, 3, 4, 5, 6.
- **Volumen de consumo (m³):** 5, 10, 20, 40, 80, 150, 400, 700, 1 000, 1 500, 2 000, 3 000, 4 500, 7 000, 10 000.
- **Costo de combustible auxiliar (CLP/m³):** 100, 800, 1 500.

Estas combinaciones representan las distintas configuraciones son un abanico representativos de distintos SST. Otro aspecto importante es el lugar donde se realizan las simulaciones, ya que esto añade una dimensión espacial significativa al problema. Para ello, se utilizan las 20 ciudades presentadas en la Tabla 4.2 donde se puede ver el nombre de la ciudad, la Latitud, Longitud y Radiación Solar Global anual. Con esto se genera una base de datos que consta de 1 231 150 simulaciones de sistemas solares térmicos con todas las combinaciones mencionadas que cumplan con la relación de volumen y área.

Algo destacable en la construcción de la base de datos es que el tiempo promedio de una simulación de la modelación matemática presentada en el capítulo 2.1 es de 6,0919 segundos, con una desviación estándar de 1,3097 segundos. Por lo tanto, la construcción de la base de datos no constituye una tarea de corta duración.

Tabla 4.2: Ciudades para entrenar

Zona	Ciudad	Latitud	Longitud	GSR kWh/m ² d
1	Arica	-18.5	-70.2	4.79
2	Iquique	-20.5	-70.2	4.74
3	Calama	-22.5	-68.9	6.49
4	Antofagasta	-23.4	-70.5	5.24
5	Copiapó	-27.3	-70.4	5.28
6	La Serena	-28.6	-70.8	4.20
7	Valparaíso	-33.0	-71.6	3.75
8	Santiago	-33.5	-70.7	3.90
9	Rancagua	-34.2	-70.7	5.40
10	Talca	-35.4	-71.7	5.13
11	Chillán	-36.6	-72.1	5.12
12	Concepción	-36.8	-73.1	4.13
13	Temuco	-38.8	-72.6	3.90
14	Valdivia	-39.8	-73.2	4.64
15	Puerto Montt	-41.4	-73.1	3.23
16	Castro	-42.5	-73.8	3.71
17	Coyhaique	-45.6	-72.0	3.52
18	Puerto Natales	-51.73	-72.5	3.08
19	Punta Arenas	-53.0	-71.0	2.99
20	Puerto Williams	-54.94	-67.61	3.08

Para reducir este tiempo, se utilizó la librería multiprocessing, aprovechando el 100 % de la capacidad de un procesador AMD Ryzen 5 5600X de 6 núcleos a 3,70 GHz y 32 GB de RAM. Gracias a ello, el tiempo total de construcción de la base de datos fue de aproximadamente 20 días, en contraste con los 85 días que habría requerido si se hubiera realizado únicamente de forma secuencial.

4.2. Preprocesamiento y análisis de los datos

En la Tabla 4.3 se presenta la matriz de correlación lineal entre la fracción solar y el VAN. En ella se puede observar que buscar una relación lineal entre las variables objetivo no parece una tarea viable. De manera similar, la Tabla 4.4 muestra la matriz de información mutua, donde se evidencia que la dependencia entre ambas variables objetivo es lo suficientemente baja como para justificar su predicción de forma independiente. Ambos gráficos dan información de que el trabajo debe centrarse en predecir las variables por separado.

Correlación lineal	VAN	Fracción solar
VAN	1	0.15
Fracción solar	0.15	1

Tabla 4.3: Correlación lineal del VAN y la fracción solar

Información mutua	VAN	Fracción solar
VAN	1	0.58
Fracción solar	0.58	1

Tabla 4.4: Información mutua del VAN y la fracción solar

Para el preprocesamiento de los datos se emplea una técnica denominada normalización, la cual consiste en transformar cada característica (feature) del sistema. Este procedimiento tiene como objetivo facilitar la convergencia de algoritmos que utilizan el método del gradiente, así como aquellos que incorporan técnicas de penalización L1 o L2 esto es explicado en [James, Witten, Hastie, Tibshirani, et al. \[2013\]](#). De esta forma, se evita que variables en escalas distintas dominen el proceso de aprendizaje únicamente por su magnitud.

En este trabajo, se utiliza específicamente la técnica de estandarización mediante puntuación z (z-score), definida por la siguiente ecuación:

$$x'_i = \frac{x_i - \mu}{\sigma}, \quad (4.1)$$

donde μ representa la media y σ la desviación estándar de la característica, ambas calculadas sobre el conjunto de entrenamiento.

Las características consideradas para los modelos de aprendizaje son las siguientes:

1. Costo del colector.
2. Costo del estanque de almacenamiento.
3. Radiación global máxima de la ciudad.
4. Radiación global promedio de la ciudad.
5. Temperatura ambiente máxima de la ciudad.
6. Temperatura ambiente mínima de la ciudad.

7. Temperatura ambiente promedio de la ciudad.
8. Temperatura de consumo.
9. Volumen de consumo.
10. Radiación promedio desde Enero a Diciembre desagregada por mes y ciudad correspondiente.
11. Perdidas del colector.
12. Rendimiento del colector.
13. Volumen del estanque de almacenamiento.

Para el entrenamiento de los algoritmos de aprendizaje automático, el conjunto total de datos se divide en tres subconjuntos:

- Conjunto de entrenamiento (60 %): utilizado para ajustar los parámetros de los modelos.
- Conjunto de validación (20 %): empleado para ajustar los hiperparámetros, en caso de que el algoritmo no utilice validación cruzada.
- Conjunto de prueba (20 %): se mantiene completamente independiente del proceso de entrenamiento, y se utiliza para la evaluación final del modelo, permitiendo estimar su capacidad de generalización a datos no vistos.

La partición de estos conjuntos se realiza estratificando por ciudad y volumen de consumo, con el fin de mantener proporciones representativas de estas variables en cada subconjunto y asegurar una distribución equilibrada durante el entrenamiento y evaluación del modelo.

4.3. Algoritmos a utilizar

En esta Sección se dan a conocer los algoritmos que se utilizaron para predecir el VAN y la fracción solar a partir de las configuraciones y parámetros de diseño del SST. Los Métodos Support Vector Machine, árbol de regresión y redes neuronales son explicados a partir de [James et al. \[2013\]](#).

Método F-chart

El método F-chart, introducido en el estado del arte, fue desarrollado y explicado en detalle en [Klein et al. \[1976\]](#) y [Klein and Beckman \[1979\]](#). Para efectos de este capítulo, sólo se considerará que ambos trabajos presentan al F-chart como una herramienta para predecir la fracción solar de un sistema solar térmico a partir de las características de los SST, explicando brevemente cómo se calcula y cuáles son algunas limitaciones del método. Asimismo, el método F-chart será utilizado como referencia para evaluar y comparar el desempeño de otros enfoques en la predicción de la fracción solar, constituyendo así una línea base frente a la cual se contrastarán los algoritmos entrenados. El cálculo de la fracción solar mensual mediante F-chart se realiza a través de la

siguiente ecuación.

$$f_k = 1.029Y - 0.065Z - 0.245Y^2 + 0.0018Z^2 + 0.0215Y^3. \quad (4.2)$$

Donde las variables Z e Y representan y se calculan de la siguiente manera.

$$\begin{aligned} Z &= \frac{\text{Pérdida de energía del colector durante un mes}}{\text{Total de demanda termica durante un mes}}, \\ Z &= \tau U_l \frac{F'_R}{\tau} (T^C - T^{amb}) \Delta t \frac{A}{L}, \\ Y &= \frac{\text{Total de energía absorbida por los colectores durante un mes}}{\text{Total de demanda de calor durante un mes}}, \\ Y &= \tau (\tau_\alpha) \frac{F'_R}{\tau} \left(\frac{\overline{\tau_\alpha}}{\tau_\alpha} \right) \overline{H_T} N \frac{A}{L}. \end{aligned}$$

Donde,

F'_R = Eficiencia del intercambiador de calor [%],

Δt = Tiempo total del mes en segundos [s],

L = Demanda de calor mensual,

τ = Transmitancia de la cubierta del colector,

τ_α = Transmitancia-absortancia a incidencia normal mensual promedio [%],

$\overline{\tau_\alpha}$ = Transmitancia normal angular [%],

$\overline{H_T}$ = Radiación diaria incidente a la superficie del colector promedio mensual por unidad de area [MJ/m^2],

N = Numero de días en el mes.

Como el método F-chart calcula la fracción solar promedio mensual; para obtener un resultado anual, se utiliza la expresión.

$$f = \frac{\sum_{k=1}^{12} f_k L_k}{\sum_{k=1}^{12} L_k}.$$

El método F-chart utiliza los promedios de radiaciones mensuales de un año meteorológico, entre otras variables, para realizar las predicciones. Es por ello que en este trabajo se incluirán estos promedios como entradas para los modelos de aprendizaje a utilizar, los cuales se encuentran en el anexo identificados como las Tablas 7.1 y 7.2.

Es importante recordar que el método F-chart es una metodología que se deriva de simulaciones y un trabajo de regresión de ciertas configuraciones de sistemas solares térmicos y que, al ser un modelo empírico-estadístico, tiene las siguientes limitaciones:

- Si se aplican otro tipo de colectores solares se pueden encontrar diferencias entre el método y una simulación más ad-hoc.
- El modelo F-chart asume que los estanques de almacenamientos son totalmente mezclados y no contempla la utilización de estanques estratificados.
- Los parámetros adimensionales Z e Y tienen zonas de convergencias, en la literatura en general se utilizan los rangos de convergencia en particular en Kerme and Kaneesamkandi [2015] enuncian que $0 < Z < 15$ e $0 < Y < 3$ son los parámetros de convergencia, además de esta consideración también es enunciada por Klein y Beckman en Klein and Beckman [1979] por lo que al acercarse a las fronteras se esperan diferencias significativas en los resultados, mientras que fuera de estas zonas no se espera que los resultados sean confiables ya que se podrían dar errores más que significativos, aunque los rangos de convergencia presentados .

Es por esto que es esperable que el método F-chart no sea preciso y contenga errores de estimación, pues el modelo presentado utiliza un estanque estratificado como se puede ver en 3.3, además de que en varios casos las variables Z estimadas se encuentran por encima de la zona de convergencia, pero es un método rápido y no existen alternativas en la literatura que puedan evitar estos problemas.

SVM (Support Vector Machine) para regresión: Las SVM constituyen un método de aprendizaje supervisado que aborda problemas de separación en el espacio de características mediante la aplicación de funciones de transformación sobre el espacio de entrada. Dichas transformaciones, denominadas *funciones base*, permiten representar los datos en un espacio de mayor dimensionalidad, facilitando la modelación de relaciones no lineales presentes en el dominio original.

Sea el conjunto de funciones base $h_m(x)$, $m = 1, \dots, M$, previamente seleccionadas, las cuales se emplean para generar el vector de características transformadas:

$$h(x_i) = (h_1(x_i), h_2(x_i), \dots, h_M(x_i)), \quad i = 1, \dots, N.$$

A partir de estas representaciones, el modelo SVM para regresión ajusta los parámetros $\hat{\beta}$ y β_0 , dando lugar a la función de predicción:

$$\hat{f}(x) = h(x)^T \hat{\beta} + \beta_0.$$

Las SVMs pueden ser utilizadas para problemas de clasificación y para problemas de regresión que buscan una respuesta cuantitativa. Para ello se considera la regresión lineal.

$$f(x) = x^T \beta + \beta_0, \tag{4.3}$$

y luego se busca manejar la generalización no lineal de la regresión. Para estimar β , se debe

considerar el problema de minimización

$$\begin{aligned} \text{minimizar} \quad & H(\beta, \beta_0) = \sum_{i=1}^N V(1 - y_i f(x_i)) + \frac{\lambda}{2} \|\beta\|^2, \\ & \beta_0, \beta \end{aligned} \quad (4.4)$$

donde

$$V_\epsilon(r) = \begin{cases} 0 & \text{si } |r| < \epsilon \\ |r| - \epsilon, & \text{en otro caso.} \end{cases} \quad (4.5)$$

Esta ecuación establece un criterio para ignorar los errores cuyo valor absoluto sea menor que ϵ , con el fin de reducir la influencia de los puntos cercanos a la frontera del hiperespacio generado y, de este modo, excluirlos del proceso de optimización. En el contexto de la regresión, estos puntos de “bajo error” corresponden a aquellos con residuos pequeños.

Árbol de regresión: Los árboles de regresión buscan resolver problemas de regresión, por medio de estratificar y segmentar el espacio predictor en una cantidad simple de regiones. Si la data cuenta con p variables predictoras (inputs) y una variable a predecir (target), y se cuentan con N observaciones, tenemos:

$$(x_i, y_i), \quad i \in 1, 2, \dots, N, \quad \text{con } x_i = (x_{i1}, x_{i2}, \dots, x_{ip}) \in \mathbb{R}^{1 \times p}, y_i \in \mathbb{R}. \quad (4.6)$$

Luego se divide el espacio predictor a partir de los posibles valores para las p observaciones $x_1, x_2, \dots, x_p$ en M regiones disjuntas $R_1, R_2, \dots, R_M$. Para cada observación que se encuentre en la región R_j el modelo debe responder una constante c_j , es decir:

$$f(x) = \sum_{m=1}^j c_j I(x \in R_j). \quad (4.7)$$

Con,

$$I(x \in R_j) = \begin{cases} 1 & \text{, si el vector de predictores } x \text{ pertenece a la región } R_j, \\ 0 & \text{, en caso contrario.} \end{cases}$$

Luego si se busca minimizar la suma cuadrada del residuo (SSR) $\sum_{i=1}^N (y_i - f(x_i))^2$, es fácil ver que el mejor $\hat{c}_j$ es el promedio de los $y_i \in R_j$.

Resulta evidente que encontrar las particiones que minimizan el SSR es computacionalmente inviable. Por ello, se recurre a un algoritmo greedy. Este algoritmo comienza considerando el conjunto completo de datos y busca generar una partición en dos hiperplanos, evaluando una variable de división l y un punto de división s . A partir de estos, se define el par de hiperplanos:

$$R_1(l, s) = \{X|X_l \leq s\},$$

$$R_2(l, s) = \{X|X_l > s\}.$$

Luego se buscan los puntos j que dividan el punto s , tal que resuelvan

$$\min_{(l,s)} \{ \min_{c_1} \sum_{x_i \in R_1(l,s)} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2(l,s)} (y_i - c_2)^2 \}.$$

El mínimo se encuentra precisamente en el promedio de ambas regiones, lo que permite generar la primera división en dos hiperplanos. Posteriormente, en cada hiperplano se repite el procedimiento para subdividir las nuevas regiones. Este proceso se evalúa mediante validación cruzada, con el fin de seleccionar el árbol con mejor desempeño según la métrica SSR.

Redes neuronales:

Las redes neuronales son un modelo computacional que busca aproximar la función latente entre variables de entrada y salida por medio de neuronas que buscan modelar relaciones no lineales entre los datos, reconocer patrones y realizar tareas como clasificación o, en este caso, regresión.

Una neurona artificial es la unidad básica de una red neuronal. El funcionamiento de una neurona se puede modelar de manera matemática mediante una función de activación aplicada a una combinación lineal de sus entradas.

Sea $\mathbf{x} = (x_1, x_2, \dots, x_n)$ el vector de entrada a la neurona y $\mathbf{w} = (w_1, w_2, \dots, w_n)$ el vector de los pesos correspondiente a cada entrada. La salida de la neurona está dada por una función no lineal aplicada a la suma ponderada de las entradas, más un término de sesgo $\mathbf{b} \in \mathbb{R}$:

$$z = \mathbf{w}^T \mathbf{x} + \mathbf{b}$$

$$y = \phi(z)$$

Donde,

- z es el valor de activación lineal.
- y es la salida de la neurona.
- ϕ es la función de activación, típicamente ReLU, sigmoide o tangente hiperbólica.

Una red neuronal típica consta de varias capas de neuronas que se organizan de la siguiente manera:

- Capa de entrada: Recibe los datos de entrada. Cada neurona en esta capa corresponde a una característica del conjunto de datos.
- Capas ocultas: Son las capas intermedias que procesan la información a través de una serie

de transformaciones no lineales. Cada neurona en una capa oculta recibe las salidas de la capa anterior, las pondera mediante un conjunto de pesos y las pasa a través de una función de activación.

- **Capa de salida:** Genera la predicción final de la red, según la tarea a resolver.

De este modo, la salida $y^{(L)}$ de una red neuronal con L capas se define como la composición de las funciones de activación aplicadas en cada capa:

$$y^{(L)} = \phi^{(L)}(\mathbf{W}^{(L)}y^{(L-1)} + \mathbf{b}^{(L)})$$

donde $y^{(L-1)}$ es la salida de la capa anterior, $\mathbf{W}^{(L)}$ es la matriz de pesos de la capa L y $\mathbf{b}^{(L)}$ es el vector de sesgos de esa capa. Este proceso de transmisión de información desde la capa de entrada hacia la capa de salida se denomina propagación hacia adelante.

Más información sobre los modelos de aprendizaje automático se puede encontrar en [Hastie, Tibshirani, Friedman, and Franklin \[2005\]](#) y [James et al. \[2013\]](#).

4.4. Entrenamiento de modelos

Los algoritmos de aprendizaje automático mencionados en la Sección 4.3 emplean dos tipos de parámetros:

- **Parámetros del modelo:** Son aquellos que se ajustan durante el entrenamiento mediante alguna técnica de optimización (por ejemplo, los pesos y sesgos en redes neuronales, o los coeficientes en una regresión lineal).
- **Hiperparámetros:** Son valores que se mantienen fijos durante cada proceso de entrenamiento y controlan el comportamiento del algoritmo (por ejemplo, la tasa de aprendizaje, el número de capas ocultas, la profundidad máxima de un árbol de regresión, etc.).

Para determinar la combinación óptima de hiperparámetros, se utilizan técnicas de búsqueda estructurada y validación, en particular se utilizará.

1. *Búsqueda en malla (grid search):* explora de manera exhaustiva un conjunto predefinido de valores para cada hiperparámetro.

Estas técnicas suelen ir acompañadas de un esquema de *validación cruzada*, en el cual el conjunto de datos de entrenamiento se divide en k particiones. Para cada combinación de hiperparámetros:

1. Se entrena el modelo en $k - 1$ particiones.
2. Se evalúa su desempeño en la partición restante, repitiendo el proceso para cada partición.
3. Se calcula una métrica agregada (por ejemplo, exactitud, AUC, error cuadrático medio) sobre las k ejecuciones.

Finalmente, se selecciona la configuración de hiperparámetros que maximiza (o minimiza) la métrica de interés en validación. Una vez elegido el mejor conjunto de hiperparámetros, el

modelo resultante se entrena de nuevo con todos los datos de entrenamiento y, si es necesario, se evalúa su capacidad de generalización en un conjunto de prueba independiente.

4.5. Búsqueda de grilla para hiperparámetros

A continuación se presentarán los hiperparámetros explorados para los modelos de aprendizaje, como el árbol de regresión, SVR y redes neuronales, tanto para el VAN como para la fracción solar.

- Árbol de regresión:

1. Máxima profundidad del árbol. $\{1, 2, \dots, 40\}$. La máxima profundidad del árbol determina la cantidad de niveles iterativos en los que se pueden realizar divisiones antes de que el algoritmo termine de entrenar.
2. Mínimo de datos para realizar una división. $\{2, 3, \dots, 10\}$. El mínimo de datos para realizar una división determina la cantidad mínima de datos necesarios para evaluar la división de la data, o lo que es lo mismo, la cantidad que se considera para que se entienda que es una hoja pura.

- SVM:

1. Kernel, $\{\text{linear}, \text{rbf}\}$. El kernel define la función utilizada para transformar los datos en un espacio de mayor dimensión.
2. Parámetro de penalización, $\{0.1, 10, 100, 1\,000\}$. Controla el trade-off entre maximizar el margen y minimizar el error en las predicciones.

- Redes neuronales

1. Número de capas ocultas. $\{1, 2, 3, 4, 5\}$.
2. Número de neuronas por capa oculta. $\{12, 15, 18, 21\}$.
3. Funciones de activación. $\{\text{ReLU}, \text{tanh}\}$
4. Coeficiente de aprendizaje. $\{0.001, 0.01, 0.1\}$

Dando un total de 74 752 configuraciones posibles de la combinatoria de estos factores, por lo que se hace un random search dentro de todas las combinaciones, probando 1 000 combinaciones por modelo para VAN y para fracción solar.

5 Resultados

En este capítulo se presentan las metodologías empleadas para entrenar, evaluar y comparar el desempeño de distintos algoritmos de aprendizaje automático en la predicción de la fracción solar y el VAN de sistemas solares térmicos. El objetivo central es determinar qué modelos permiten obtener predicciones precisas y eficientes, reduciendo significativamente los costos computacionales asociados al modelo matemático tradicional.

El proceso incluye la definición y ajuste de hiperparámetros mediante estrategias de búsqueda en malla y validación cruzada, diferenciando entre parámetros internos de los modelos e hiperparámetros de control. Se consideran, en particular, árboles de regresión, SVM y redes neuronales, contrastando su desempeño con el método F-chart como referencia y con el modelo teórico detallado previamente.

La evaluación de los modelos se realiza a partir de métricas de error cuadrático medio (MSE) y tiempos de ejecución, lo que permite identificar los algoritmos más adecuados para cada variable de interés: árboles de regresión para la fracción solar y redes neuronales para el VAN. Finalmente, se analiza la capacidad de generalización de los modelos mediante simulaciones en localidades no incluidas en el entrenamiento y mediante la expansión del espacio de búsqueda, con el fin de evaluar su efectividad en la identificación de configuraciones óptimas bajo condiciones no previamente observadas.

5.1. Desempeño de los algoritmos

En la presente sección, se comparan el desempeño de los algoritmos, presentados en la sección 4.3 para predecir la fracción solar y el VAN. La comparación se realiza a partir de MSE y tiempos computacionales, evaluando el rendimiento de cada modelo sobre un conjunto de prueba que consta de 246 227 simulaciones previas.

Para la predicción de la fracción solar, se analizan los métodos, destacándose el árbol de regresión por su precisión, mientras que el método F-chart presenta los menores tiempos computacionales, ya que es una fórmula directa. A pesar de que el tiempo de ejecución de algunos modelos es considerablemente más alto que el de otros, se concluye que los tiempos resultantes son aceptables en comparación con los tiempos computacionales del modelo matemático tradicional, que es más costoso en términos de procesamiento. En este contexto, el árbol de regresión se posiciona como la mejor opción para la predicción de la fracción solar, dado su bajo MSE y su tiempo de ejecución eficiente.

Por otro lado, al abordar la predicción del VAN, las redes neuronales sobresalen por su precisión, presentando un MSE notablemente inferior al de otros métodos. Sin embargo, los

tiempos de computación para las redes neuronales son más altos que los del árbol de regresión, aunque continúan siendo razonables en comparación con el modelo teórico matemático. Por tanto, se determina que, para la predicción del VAN, la red neuronal es el mejor método, pues logra un balance entre alta precisión y tiempos de ejecución que, si bien son superiores a los de otros algoritmos, siguen siendo adecuados.

A continuación, se detallan los resultados obtenidos de la comparación de los algoritmos para ambos casos, junto con los parámetros y configuraciones que permitieron alcanzar los mejores desempeños.

5.1.1. Fracción solar

En la Tabla 5.1 se comparan los desempeños de los algoritmos en la predicción de la fracción solar, junto con sus respectivos errores cuadráticos medios (MSE) y tiempos computacionales. Estos tiempos fueron medidos sobre un conjunto de prueba compuesto por 246 227 simulaciones almacenadas.

Se observa que el mejor desempeño en términos de precisión lo presenta el árbol de regresión, mientras que el menor tiempo computacional corresponde al método F-chart. Sin embargo, ambos tiempos son más que satisfactorios si se los compara con el del modelo matemático implementado en la Sección 3.1, el cual presenta tiempos computacionales considerablemente más altos.

Considerando que el MSE del árbol de regresión representa solo un 0,49 % del MSE del método F-chart, y que sus tiempos computacionales no superan en más de cuatro veces a los de dicho método —manteniéndose, además, por debajo de los 0,1 segundos—, se considera que estos tiempos son aceptables. Por lo tanto, bajo estas dos consideraciones, el mejor método para la predicción de la fracción solar es el árbol de regresión.

Método	MSE en test	Tiempo en predecir el test set [s]
Árbol de regresión	0.1788	0.0904
F-chart	3.6450	0.0260
SVR	35.2085	10 177.1
Redes neuronales	41.5672	4.1423
Modelo Teórico	0	1 499 990.261

Tabla 5.1: Mejores desempeños de algoritmos de fracción solar

El árbol de regresión con mejor desempeño cuenta con los hiperparámetros:

1. Máxima profundidad del árbol de 30.
2. Mínimo de datos para realizar una división. Igual a 3.

5.1.2. VAN

En la Tabla 5.2 se comparan los desempeños de los algoritmos en la predicción del VAN, junto con sus respectivos MSE y tiempos computacionales. Estos tiempos fueron medidos sobre un

conjunto de prueba compuesto por 246 227 simulaciones almacenadas.

Se observa que el mejor desempeño en términos de precisión lo presentan las redes neuronales, cuyo MSE es dos órdenes de magnitud menor al del árbol de regresión, que es el segundo mejor método en cuanto a error. Por otro lado, el menor tiempo computacional corresponde al árbol de regresión. No obstante, aunque el tiempo de ejecución de las redes neuronales es significativamente superior al del árbol de regresión, sigue siendo aceptable, especialmente si se lo compara con el tiempo requerido por el modelo matemático implementado en la Sección 3.1.

Considerando que el MSE de la red neuronal es dos órdenes de magnitud menor al del árbol de regresión, y que su tiempo computacional se mantiene en torno a los 5 segundos, se considera que dicho tiempo es aceptable. Por lo tanto, bajo estas dos consideraciones, el mejor método para la predicción del VAN es la red neuronal.

Método	MSE en test	Tiempo en predecir el test set [s]
Árbol de regresión	7.8186×10^{18}	0.0290
SVR	1.9268×10^{19}	5 722.6
Redes neuronales	6.3348×10^{16}	5.5889
Modelo Teórico	0	1 499 990.261

Tabla 5.2: Mejores desempeños de algoritmos de VAN

La arquitectura de la red neuronal con mejor desempeño es la que cuenta con la arquitectura presentada en la tabla 5.3 con un learning rate de 0.01.

Capa	Dimensión de salida	Parametros	Función de activación
Input layer	25	0	-
Densa 1	12	312	ReLU
Densa 2	15	195	ReLU
Densa 3	15	240	ReLU
Densa 4	1	16	linear

Tabla 5.3: Arquitectura mejor red neuronal de VAN

5.2. Búsqueda de configuraciones óptimas

Para evaluar el desempeño de los modelos de aprendizaje entrenados, se construyeron simulaciones con el mismo mallado descrito en la sección 4.1, pero aplicándolo en localidades distintas a las de entrenamiento. En particular, se eligieron Sierra Gorda y Maipú como dos casos representativos: la primera por su elevada irradiación solar y la segunda por su carácter urbano. Al no formar parte de los datos de entrenamiento, estas localidades permiten verificar la capacidad de los modelos para capturar la influencia de la radiación solar.

5.2.1. Fracción solar

Para evaluar la predicción de la fracción solar promedio anual, se emplea el árbol de regresión que obtuvo el menor error cuadrático medio (véase tabla 5.1) ya que es por órdenes de magnitud el menor error además de tener tiempos computacionales más que satisfactorios. Con el fin de identificar las configuraciones óptimas (aquellas que maximizan la fracción solar), se agrupan las simulaciones según las variables “volumen de consumo” y “temperatura de consumo”, las cuales determinan la demanda térmica y, por ende, condicionan la fracción solar alcanzable, con lo que la data está agrupada en pares de estas variables. De este modo, para cada ciudad se generan 60 grupos de configuraciones, de los cuales solo una por grupo puede ser la configuración óptima. Esta agrupación corresponde a escenarios donde estos parámetros son fijados de manera externa y son condiciones del problema, bajo ningún contexto son variables de decisión.

El árbol de regresión acierta plenamente en 57 de los 60 casos para Sierra Gorda; en Maipú ocurre lo mismo, como se aprecia en la tabla 5.4. Con lo que, en total, de las 120 configuraciones óptimas (60 por localidad), el modelo acierta en 114. Esto sugiere que el árbol de regresión fue capaz de captar con precisión la relación latente entre la radiación solar y la fracción solar, a pesar de haber sido entrenado en localidades distintas.

Aunque no acierta siempre, lo hace en el 95 % de los casos, lo cual indica un desempeño muy favorable al objetivo de identificar los óptimos por medio de esta metodología.

Localidad	Aciertos	Fallas	Total
Sierra Gorda	57	3	60
Maipú	57	3	60

Tabla 5.4: Desempeño predicción de óptimos para la fracción solar

5.2.2. VAN

Para evaluar la predicción del VAN, se emplea la red neuronal que obtuvo el menor error cuadrático medio (véase tabla 5.2). Con el objetivo de identificar las configuraciones óptimas (aquellas que maximizan el VAN), se agrupa la salida de las simulaciones según los parámetros “volumen de consumo” y “temperatura de consumo”, los cuales determinan la demanda térmica y, por ende, influyen en los costos y flujos de caja que conforman el VAN. De este modo, para cada ciudad se generan 60 grupos de configuraciones, de los cuales solo una por grupo puede considerarse óptima.

En el caso de la predicción del VAN, se observa un marcado contraste respecto a la fracción solar, ya que el modelo no logra acertar en ninguno de los óptimos. Sin embargo, las configuraciones predichas no siempre se alejan demasiado de las óptimas. Para Sierra Gorda, el error cuadrático medio de las predicciones en el mallado original es de 2.0058×10^{10} , valor que, aunque seis órdenes de magnitud inferior al error cuadrático medio de la red neuronal reportado en la tabla 5.2, sigue siendo demasiado elevado para cumplir el objetivo de identificar configuraciones óptimas.

Para Maipú, se tiene un comportamiento similar, el error cuadrático medio es de 1.2504×10^{10} y aunque en algunas configuraciones sólo se encuentra a uno o dos vecinos de algún parámetro del óptimo, no acierta en ninguno de los casos.

En todos los casos, los errores son excesivamente altos para identificar de forma fiable los óptimos. Si bien en alrededor de 40 de los 120 casos las configuraciones predichas se encuentran a solo un vecino de la óptima, en general el modelo no logra detectar de manera consistente las configuraciones óptimas como era el objetivo planteado en este trabajo.

5.2.3. Expansión del mallado

Para complementar la evaluación del árbol de regresión y la red neuronal, se amplió el espacio de búsqueda mediante una expansión del mallado en Sierra Gorda y Maipú. Se partió de la malla original descrita en la sección 4.1 y se añadieron nuevos valores en los parámetros clave: volumen del estanque (a partir de 15 m^3 y hacia arriba) y área colectora (a partir de 500 m^2 y en adelante). Este nuevo mallado es sólo utilizado para probar la capacidad de generalización de las máquinas entrenadas, no se realiza un nuevo entrenamiento a partir de la misma, sólo se quiere evaluar el desempeño fuera de los rangos entrenados para evaluar posibles consideraciones a futuro. El nuevo mallado incluye ahora las combinaciones de los siguientes parámetros:

- **Volumen de estanque (m^3):** 0.08, 0.20, 0.50, 0.80, 1, 3, 5, 8, 13, 15, 20, 30, 50, 80, 100, 150, 200, 300, 500.
- **Área colectora (m^2):** 1.19, 4.76, 12, 30, 70, 100, 160, 325, 500, 1 000, 1 500, 2 000, 3 000, 4 500, 7 000, 10 000.
- **Temperatura de consumo ($^{\circ}\text{C}$):** 40, 50, 60, 70.
- **Modelo de colector:** 1, 2, 3, 4, 5, 6.
- **Volumen de consumo (m^3):** 5, 10, 20, 40, 80, 150, 400, 700, 1 000, 1 500, 2 000, 3 000, 4 500, 7 000, 10 000.
- **Costo de combustible auxiliar (CLP/m^3):** 100, 800, 1 500.

Sierra Gorda Se comparan los valores máximos obtenidos con el nuevo mallado frente a los predichos por el árbol de regresión de menor error cuadrático medio (tabla 5.1). Estos máximos se definen agrupando “volumen de consumo” y “temperatura de consumo” para generar 60 configuraciones óptimas. De las 171 720 configuraciones simuladas para esta localidad.

Para la fracción solar el árbol acierta en 23 de las 60 configuraciones óptimas. Sin embargo, los errores crecen con la demanda térmica y las diferencias son especialmente marcadas fuera de los rangos de entrenamiento. Solo 1 de las 37 predicciones erróneas corresponde a una configuración vista durante el entrenamiento; en los 36 restantes, al menos uno de los parámetros (área colectora o volumen de estanque) resulta nuevo y queda fuera del espacio de entrenamiento.

Estas discrepancias se detallan en las tablas 7.7 y 7.8, donde se evidencia la dificultad del modelo para generalizar a nuevas áreas colectoras y volúmenes de estanque.

Para el VAN en el mallado extendido, el error cuadrático medio de la red neuronal asciende a 4.8106×10^{10} , lo que confirma la limitación de la red neuronal en identificar los óptimos de manera acertada; en particular, mayores problemas tiene en relacionar el comportamiento fuera de la zona de entrenamiento.

Maipú La misma metodología se aplica en Maipú: se generan 60 configuraciones óptimas de volumen y temperatura de consumo (entre las 171 720 posibles) y se comparan las predicciones con los valores reales (tabla 5.1). El árbol acierta en 23 casos y muestra un comportamiento análogo al de Sierra Gorda: el error aumenta con la demanda térmica y las diferencias son significativas para parámetros fuera del rango de entrenamiento. Solo 1 de las 37 predicciones erróneas corresponde a una configuración vista durante el entrenamiento; en los 36 restantes, al menos uno de los parámetros (área colectora o volumen de estanque) resulta nuevo y queda fuera del espacio de entrenamiento.

Las discrepancias en Maipú se presentan en las tablas 7.9 y 7.10.

Para el VAN en el mallado extendido, el error cuadrático medio de la red neuronal asciende a 3.5424×10^{10} , lo que confirma la limitación y dificultad de la red neuronal para identificar los óptimos, además de demostrar el aumento en los problemas para determinar la relación entre las variables y el VAN fuera de la región de entrenamiento.

Con estos resultados en Sierra Gorda y Maipú se evidencia que las máquinas de aprendizaje presentan limitaciones para predecir de buena manera los desempeños cuando deben hacerlo para configuraciones con las que no fueron entrenadas o se escapan del intervalo de entrenamiento, más no así con respecto a localidades que no fueran parte del entrenamiento, lo que da un resultado similar a los obtenidos por el método F-chart en Klein et al. [1976] y Klein and Beckman [1979].

6 Conclusiones

A partir de los resultados presentados en la sección 5.1, es posible analizar y evaluar el grado de cumplimiento de los objetivos planteados en la sección 3.3.

En primer lugar, respecto al objetivo "Conseguir resultados similares al modelamiento presentado en la sección 3.1 para la fracción solar y el VAN, utilizando como métrica el error cuadrático medio (MSE) para medir la aproximación entre resultados", se observa lo siguiente:

A partir de la tabla de fracción solar 5.1, se evidencia un MSE considerablemente bajo, incluso inferior al obtenido con el método F-chart, ampliamente utilizado en la literatura sobre dimensionamiento de SST y explicado en detalle en Klein et al. [1976] y Klein and Beckman [1979]. Este resultado representa un logro relevante y permite concluir que se cumple de manera satisfactoria el objetivo de alcanzar resultados comparables al modelamiento de 3.1.

Por otro lado, a partir de la tabla del VAN 5.2, se observa que el MSE no es controlado adecuadamente por ninguna de las máquinas de aprendizaje utilizadas. Esto sugiere que la información entregada a los modelos es insuficiente y que sería necesario incorporar más variables económicas para facilitar la identificación del flujo de caja asociado al funcionamiento del sistema. A diferencia de la fracción solar, no existen trabajos previos orientados a aproximar el VAN, por lo que este estudio constituye un primer acercamiento. Si bien los resultados se alinean de forma preliminar con lo requerido, no son suficientes y dejan un margen evidente para mejoras futuras.

En relación con el segundo objetivo, "Conseguir resultados en menores tiempos computacionales que la modelación presentada en la sección 3.1, cuyo tiempo medio por simulación es de 6,0919 segundos", se observa que, según las tablas 5.1 y 5.2, los modelos propuestos logran predecir simulaciones con tiempos equivalentes al $6.0267 \times 10^{-6} \%$ del tiempo requerido por el modelo teórico para la fracción solar, y al $3.7260 \times 10^{-4} \%$ para el VAN. En ambos casos, los tiempos computacionales son por órdenes de magnitud muy inferiores al 1 %, lo que permite concluir que este objetivo se cumple con creces, demostrando que entrenar máquinas de aprendizaje es una estrategia altamente viable para reducir los tiempos de simulación.

En cuanto al tercer objetivo, "Conseguir aproximaciones de configuraciones óptimas que resuelvan los problemas (3.5) y (3.6)", los resultados muestran que para la fracción solar el objetivo se cumple casi con total precisión dentro del mallado original del problema, como se detalla en la sección 5.2.1. No obstante, al buscar configuraciones fuera del mallado de entrenamiento, la precisión disminuye, lo que sugiere que futuros trabajos deberían enfocarse en mejorar el mallado y evaluar cómo diferentes configuraciones afectan el desempeño de los modelos.

En contraste, para el VAN y considerando únicamente el mallado de entrenamiento, el objetivo se cumple solo parcialmente. Tal como se describe en 5.2.2, bastantes de las configuraciones óptimas estimadas se ubican en posiciones adyacentes al valor óptimo real, lo que las convierte

en aproximaciones útiles, pero no lo suficientemente precisas para la mayoría de los casos. Esta limitación está estrechamente relacionada con el problema identificado en el primer objetivo para la estimación del VAN.

Finalmente, al explorar escenarios fuera del mallado de entrenamiento, se observa que los modelos presentan dificultades tanto para aproximar el valor objetivo como para identificar los óptimos, lo que evidencia una alta sensibilidad frente a las fronteras del conjunto de entrenamiento. Este comportamiento es consistente con la naturaleza de las máquinas de aprendizaje, las cuales dependen de aproximar la función latente entre entradas y salidas a partir de datos de alta calidad y en cantidad suficiente. En este contexto, la incorporación de ciertas variables puede introducir ruido y, en consecuencia, afectar negativamente el desempeño del modelo.

Como trabajo futuro, resulta fundamental avanzar en la predicción del VAN, ya que actualmente no existe literatura específica en este ámbito. Si bien este estudio constituye un esfuerzo pionero en dicha línea, los resultados muestran que aún queda un amplio camino por recorrer. En particular, la predicción de los óptimos asociados al VAN no fue satisfactoria y los errores no pudieron ser controlados de manera adecuada con las técnicas utilizadas, lo que abre la puerta a nuevas aproximaciones metodológicas, además de contar con el trabajo de [AIT LAHOUSINE OUALI et al. \[2024\]](#) como motivación al trabajar con redes neuronales para predecir una métrica económica en problemas similares.

7 Anexo

Ciudad	Enero	Febrero	Marzo	Abril	Mayo	Junio	Julio	Agosto
Arica	340.52	316.15	306.97	230.11	186.17	154.77	186.78	208.39
Iquique	344.43	330.23	307.84	200.46	173.54	172.67	161.61	151.90
Calama	340.05	329.49	291.21	215.66	169.35	165.69	181.35	193.37
Antofagasta	384.69	361.68	340.30	267.10	223.827	209.06	230.23	270.20
Copiapó	367.63	335.74	285.17	225.36	178.36	177.45	215.67	276.89
Serena	299.00	286.82	203.95	178.72	121.11	128.52	137.54	159.32
Valparaíso	336.23	299.06	233.22	173.68	108.89	101.86	99.93	134.20
Santiago	365.12	332.95	271.55	185.73	128.78	116.11	127.32	150.77
Rancagua	380.78	337.05	278.35	193.43	132.17	111.50	123.37	145.16
Talca	374.20	315.46	261.18	177.68	109.84	94.18	99.94	134.48
Chillán	364.82	298.84	234.66	164.33	108.77	92.84	93.76	128.79
Concepción	358.28	294.31	209.64	167.65	110.62	92.99	93.726	124.97
Temuco	311.01	278.14	170.87	127.03	102.40	80.65	85.74	106.08
Valdivia	318.54	272.29	167.62	123.71	80.543	70.67	68.95	93.84
Puerto Montt	248.14	240.11	152.19	100.76	76.794	60.010	60.17	93.03
Castro	241.41	230.47	143.92	110.23	74.60	63.25	65.41	86.29
Coyhaique	241.58	223.02	168.65	110.74	76.02	52.91	61.13	93.21
Puerto Natales	215.81	181.42	129.20	67.246	47.25	30.03	44.51	61.08
Punta Arenas	238.83	189.92	122.39	68.56	41.80	25.072	35.76	55.48
Puerto Williams	215.41	158.86	113.26	66.53	35.40	19.25	27.06	56.42

Tabla 7.1: Radiación promedio desde Enero a Agosto.



Ciudad	Septiembre	Octubre	Noviembre	Diciembre
Arica	235.07	304.19	346.56	359.016
Iquique	260.15	288.17	341.29	365.96
Calama	264.03	300.19	334.33	368.57
Antofagasta	326.473	369.46	402.15	405.65
Copiapó	276.89	323.20	354.01	370.53
Serena	245.59	259.29	257.44	326.86
Valparaíso	237.52	266.51	297.90	332.58
Santiago	227.79	263.25	329.39	369.91
Rancagua	245.68	252.20	346.16	368.92
Talca	225.23	237.79	330.94	359.69
Chillan	220.72	232.45	311.59	346.39
Concepción	214.05	226.56	314.17	311.73
Temuco	184.40	212.82	259.75	295.20
Valdivia	177.44	203.23	253.65	289.17
Puerto Montt	165.72	191.03	235.34	247.93
Castro	158.57	196.73	205.53	236.90
Coyhaique	132.56	201.11	264.26	299.95
Puerto Natales	103.47	153.67	213.22	232.07
Punta Arenas	104.28	162.00	231.47	268.60
Puerto Williams	94.88	139.02	199.66	219.48

Tabla 7.2: Radiación promedio desde Septiembre a Diciembre.

Tabla 7.3: Comparación fracción Sierra Gorda configuraciones entrenadas Parte 1.

Volumen predicho	Volumen optimo	Area predicha	Area optima	Diferencia optimo
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
8.00	8.00	70.00	70.00	0.00
8.00	8.00	70.00	70.00	0.00
5.00	8.00	30.00	70.00	-4.13
8.00	8.00	70.00	70.00	0.00
13.00	13.00	160.00	160.00	0.00
13.00	13.00	160.00	160.00	0.00
13.00	13.00	100.00	160.00	-1.49
13.00	13.00	100.00	160.00	-1.25
8.00	8.00	160.00	160.00	0.00
8.00	8.00	160.00	160.00	0.00
8.00	8.00	160.00	160.00	0.00
8.00	8.00	160.00	160.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00

Tabla 7.4: Comparación fracción Sierra Gorda configuraciones entrenadas Parte 2.

Volumen predicho	Volumen optimo	Area predicha	Area optima	Diferencia optimo
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00

Tabla 7.5: Comparación fracción Maipú configuraciones entrenadas Parte 1.

Volumen predicho	Volumen optimo	Area predicha	Area optima	Diferencia optimo
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
8.00	8.00	70.00	70.00	0.00
8.00	8.00	70.00	70.00	0.00
5.00	8.00	30.00	70.00	-3.31
8.00	8.00	70.00	70.00	0.00
13.00	13.00	160.00	160.00	0.00
13.00	13.00	160.00	160.00	0.00
13.00	13.00	100.00	160.00	-1.13
13.00	13.00	100.00	160.00	-0.98
8.00	8.00	160.00	160.00	0.00
8.00	8.00	160.00	160.00	0.00
8.00	8.00	160.00	160.00	0.00
8.00	8.00	160.00	160.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00

Tabla 7.6: Comparación fracción Maipú configuraciones entrenadas Parte 2.

Volumen predicho	Volumen optimo	Area predicha	Area optima	Diferencia optimo
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00
13.00	13.00	325.00	325.00	0.00

Tabla 7.7: Comparación fracción Sierra Gorda configuraciones no entrenadas Parte 1.

Volumen predicho	Volumen optimo	Area predicha	Area optima	Diferencia optimo
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
8.00	8.00	70.00	70.00	0.00
8.00	8.00	70.00	70.00	0.00
5.00	8.00	30.00	70.00	-4.13
8.00	8.00	70.00	70.00	0.00
13.00	20.00	160.00	160.00	-1.42
13.00	20.00	160.00	160.00	-0.94
13.00	20.00	100.00	160.00	-2.19
13.00	20.00	100.00	160.00	-1.81
8.00	50.00	160.00	325.00	-3.95
8.00	50.00	160.00	325.00	-2.85
8.00	50.00	160.00	325.00	-2.23
8.00	50.00	160.00	325.00	-1.84
13.00	50.00	325.00	325.00	-6.95
13.00	50.00	325.00	325.00	-4.91
13.00	50.00	325.00	325.00	-3.80
13.00	50.00	325.00	325.00	-3.10
13.00	80.00	325.00	500.00	-10.70
13.00	80.00	325.00	500.00	-7.66
13.00	80.00	325.00	500.00	-5.97
13.00	80.00	325.00	500.00	-4.90
13.00	80.00	325.00	500.00	-10.28
13.00	80.00	325.00	500.00	-7.39
13.00	80.00	325.00	500.00	-5.78
13.00	80.00	325.00	500.00	-4.75



Tabla 7.8: Comparación fracción Sierra Gorda configuraciones no entrenadas Parte 2.

Volumen predicho	Volumen óptimo	Área predicha	Área óptima	Diferencia óptimo
13.00	150.00	325.00	1000.00	-27.79
13.00	150.00	325.00	1000.00	-19.97
13.00	150.00	325.00	1000.00	-15.61
13.00	150.00	325.00	1000.00	-12.82
13.00	200.00	325.00	1500.00	-41.71
13.00	200.00	325.00	1500.00	-29.87
13.00	200.00	325.00	1500.00	-23.31
13.00	200.00	325.00	1500.00	-19.12
13.00	300.00	325.00	2000.00	-58.23
13.00	300.00	325.00	2000.00	-41.59
13.00	300.00	325.00	2000.00	-32.41
13.00	300.00	325.00	2000.00	-26.57
13.00	500.00	325.00	3000.00	-69.89
13.00	500.00	325.00	3000.00	-49.90
13.00	500.00	325.00	3000.00	-38.88
13.00	500.00	325.00	3000.00	-31.87

Tabla 7.9: Comparación fracción Maipú configuraciones no entrenadas Parte 1.

Volumen predicho	Volumen optimo	Area predicha	Area optima	Diferencia optimo
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.20	0.20	1.19	1.19	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
0.80	0.80	4.76	4.76	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
1.00	1.00	12.00	12.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
5.00	5.00	30.00	30.00	0.00
8.00	8.00	70.00	70.00	0.00
8.00	8.00	70.00	70.00	0.00
5.00	8.00	30.00	70.00	-15.76
8.00	8.00	70.00	70.00	0.00
13.00	20.00	160.00	160.00	-6.73
13.00	20.00	160.00	160.00	-5.10
13.00	20.00	100.00	160.00	-9.86
13.00	20.00	100.00	160.00	-8.17
8.00	50.00	160.00	325.00	-10.27
8.00	50.00	160.00	325.00	-7.80
8.00	50.00	160.00	325.00	-6.28
8.00	50.00	160.00	325.00	-5.26
13.00	50.00	325.00	325.00	-9.18
13.00	50.00	325.00	325.00	-6.70
13.00	50.00	325.00	325.00	-5.28
13.00	50.00	325.00	325.00	-4.37
13.00	80.00	325.00	500.00	-12.41
13.00	80.00	325.00	500.00	-9.05
13.00	80.00	325.00	500.00	-7.15
13.00	80.00	325.00	500.00	-5.91
13.00	150.00	325.00	1000.00	-14.60
13.00	150.00	325.00	1000.00	-10.95
13.00	150.00	325.00	1000.00	-8.77
13.00	150.00	325.00	1000.00	-7.32



Tabla 7.10: Comparación fracción Maipú configuraciones no entrenadas Parte 2.

Volumen predicho	Volumen optimo	Area predicha	Area optima	Diferencia optimo
13.00	150.00	325.00	1000.00	-30.24
13.00	150.00	325.00	1000.00	-21.72
13.00	150.00	325.00	1000.00	-17.01
13.00	150.00	325.00	1000.00	-13.99
13.00	200.00	325.00	1500.00	-43.55
13.00	200.00	325.00	1500.00	-31.10
13.00	200.00	325.00	1500.00	-24.27
13.00	200.00	325.00	1500.00	-19.92
13.00	500.00	325.00	3000.00	-50.92
13.00	500.00	325.00	3000.00	-36.51
13.00	500.00	325.00	3000.00	-28.55
13.00	500.00	325.00	3000.00	-23.47
13.00	500.00	325.00	3000.00	-37.94
13.00	500.00	325.00	3000.00	-26.68
13.00	500.00	325.00	3000.00	-20.63
13.00	500.00	325.00	3000.00	-16.84

Bibliografía

- Muhammad Waseem Ahmad, Jonathan Reynolds, and Yacine Rezgui. Predictive modelling for solar thermal energy systems: A comparison of support vector regression, random forest, extra trees and regression trees. *Journal of cleaner production*, 203:810–821, 2018.
- Hanane AIT LAHOUSSE OUALI, Samir TOULI, Ahmed ALAMI MERROUNI, and Ibrahim MOUKHTAR. Artificial neural network-based lcoh estimation for concentrated solar power plants for industrial process heating applications. *Applied Thermal Engineering*, 236:121810, 2024. ISSN 1359-4311. doi: <https://doi.org/10.1016/j.applthermaleng.2023.121810>. URL <https://www.sciencedirect.com/science/article/pii/S1359431123018392>.
- T Bouhal, S Fertahi, Y Agrouaz, T El Rhafiki, Tarik Kousksou, and A Jamil. Numerical modeling and optimization of thermal stratification in solar hot water storage tanks for domestic applications: Cfd study. *Solar Energy*, 157:441–455, 2017.
- Richard A Brealey, Stewart C Myers, and Franklin Allen. *Principles of corporate finance*. McGraw-hill, 2014.
- Francisco Fuentes, Nicolás Pailahueque, Iván Muñoz, Rodrigo Escobar, and José M. Cardemil. Assessing the integration of solar process heat in the dairy industry: A case study in chile. *Sustainable Energy Technologies and Assessments*, 69:103894, 2024. ISSN 2213-1388. doi: <https://doi.org/10.1016/j.seta.2024.103894>. URL <https://www.sciencedirect.com/science/article/pii/S221313882400290X>.
- C. Garnier, J. Currie, and T. Muneer. Integrated collector storage solar water heater: Temperature stratification. *Applied Energy*, 86(9):1465–1469, 2009. ISSN 0306-2619. doi: <https://doi.org/10.1016/j.apenergy.2008.12.009>. URL <https://www.sciencedirect.com/science/article/pii/S0306261908003243>.
- Trevor Hastie, Robert Tibshirani, Jerome Friedman, and James Franklin. The elements of statistical learning: data mining, inference and prediction. *The Mathematical Intelligencer*, 27(2):83–85, 2005.
- S Jaisankar, J Ananth, S Thulasi, ST Jayasuthakar, and KN Sheeba. A comprehensive review on solar water heaters. *Renewable and sustainable energy reviews*, 15(6):3045–3050, 2011.
- Gareth James, Daniela Witten, Trevor Hastie, Robert Tibshirani, et al. *An introduction to statistical learning*, volume 112. Springer, 2013.
- Esa Kerme and Zakariya Kaneesamkandi. Performance analysis and design of liquid based solar heating system. *Journal of Thermal Engineering*, 1(5):182–191, 2015.

- SA Klein and WA Beckman. A general design method for closed-loop solar energy systems. *Solar energy*, 22(3):269–282, 1979.
- Sanford Alan Klein, William A Beckman, and John A Duffie. A design procedure for solar heating systems. *Solar energy*, 18(2):113–127, 1976.
- Eberhard Markus Kleinbach, WA Beckman, and SA Klein. Performance study of one-dimensional models for stratified thermal storage tanks. *Solar energy*, 50(2):155–166, 1993.
- Govind N Kulkarni, Shireesh B Kedare, and Santanu Bandyopadhyay. Determination of design space and optimization of solar water heating systems. *Solar Energy*, 81(8):958–968, 2007.
- Hao Li, Zhijian Liu, Kejun Liu, Zhien Zhang, et al. Predictive power of machine learning for optimizing solar water heater performance: the potential application of high-throughput screening. *International journal of photoenergy*, 2017, 2017.
- Alejandra Molina, Mark Falvey, and Roberto Rondanelli. A solar radiation database for chile. *Scientific reports*, 7(1):14823, 2017.
- Fabián Andrés Bustos Olavarría. Una metodología para evaluar el potencial de energia solar termica y fotovoltaica en chile, 2013.
- Gonzalo Quiñones, Carlos Felbol, Carlos Valenzuela, Jose M. Cardemil, and Rodrigo A. Escobar. Analyzing the potential for solar thermal energy utilization in the chilean copper mining industry. *Solar Energy*, 197:292–310, 2020. ISSN 0038-092X. doi: <https://doi.org/10.1016/j.solener.2020.01.009>. URL <https://www.sciencedirect.com/science/article/pii/S0038092X20300116>.