UNIVERSIDAD DE CHILE
FACULTAD DE CIENCIAS FÍSICAS Y MATEMÁTICAS
DEPARTAMENTO DE INGENIERÍA MECÁNICA

ASSESSMENT OF DEEP LEARNING TECHNIQUES FOR DIAGNOSIS IN THERMAL SYSTEMS THROUGH ANOMALY DETECTION

MEMORIA PARA OPTAR AL TÍTULO DE
INGENIERA CIVIL MECÁNICA

CAMILA ASUNCIÓN CORREA JULLIAN

PROFESOR GUÍA:
JOSÉ MIGUEL CARDEMIL IGLESIAS

MIEMBROS DE LA COMISIÓN
ENRIQUE LÓPEZ DROGUETT
MASOUD BEZHAD

SANTIAGO DE CHILE
2019

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGERNIERA CIVIL MECÁNICA
POR: CAMILA ASUNCIÓN CORREA JULLIAN
FECHA: 2019
PROF. GUÍA: JOSÉ MIGUEL CARDEMIL IGLESIAS

Assessment of Deep Learning Techniques for Fault Detection and Diagnosis in Thermal Systems

When evaluating the performance of thermal systems, keeping a temporal record of its variables, such as temperature and flow, is useful to obtain relevant information of its operational conditions. Reliability studies of equipment and components are a fundamental step to reduce maintenance costs and extend the useful life of said systems. Identifying anomalous behavior can also be used to detect changes in consumption patterns or sudden degradation of components in the system. In recent years, various Deep Learning (DL) techniques have been successfully applied to the identification and quantification of damage in mechanical systems. Hence, it is of interest to assess their application to performance diagnosis in thermal systems.

Solar hot water systems are viable and sustainable energy sources for domestic and industrial needs, and although not complex in their operation, they require a correct integration and regular maintenance measures to effectively reduce the use of fossil fuels. Monitoring, however, elevates the costs of these systems, and thus, strategic decisions must be made to select the most relevant components in the system to be observed. Traditionally, temperature and water flow in solar collectors, pumps and heat tanks are the main variables analyzed under different meteorological conditions.

The presented thesis consists on the assessment of different DL techniques for the development of an anomaly detection model. The model is applied to a solar thermal system for water heating located at the Beauchef 851 building, which is analyzed and then simulated with TRNSYS software. This representation allows the generation of large amounts of synthetic data based on measurable variables such as temperature, flow, and environmental conditions. This is used to represent operational and fault induced states.

Deep Learning techniques are proposed for the analysis of sequential information corresponding to the data generated through the TRNSYS simulation. Different techniques are evaluated, such as Deep Recurrent Neural Networks, to assess their performance for temperature prediction under a variety of configurations depending on the previous time-window provided and the horizon of evaluation. Furthermore, the precision of these predictions is then used to develop a detection method of anomalous behavior in consumption profiles, solar collector efficiency and pump operation. The increase of the temperatures registered at the solar collector's outlet due to fault induced water-draw patterns is recognized as an anomaly with a mean accuracy of 85% for the samples studied. The decrease of the same registered temperatures due to heat underproduction in the solar field yields an accuracy of 70%. Even though the anomaly detection results obtained are highly sensitive to the simulated conditions, the assessment of different techniques shows potential to improvement if complemented with experimental measurements and validation.

RESUMEN DE LA MEMORIA PARA OPTAR
AL TÍTULO DE INGERNIERA CIVIL MECÁNICA
POR: CAMILA ASUNCIÓN CORREA JULLIAN
FECHA: 2019
PROF. GUÍA: JOSÉ MIGUEL CARDEMIL IGLESIAS

Evaluación de técnicas de Aprendizaje Profundo para la detección y diagnóstico de fallas en sistemas térmicos

A la hora de evaluar el desempeño de sistemas térmicos, mantener registros temporales de temperatura y caudal permiten obtener información sobre el rendimiento y estado de operación del sistema. Estudios de confiabilidad en equipos y componentes son un proceso fundamental para reducir costos de mantención y aumentar la vida útil de estos. La identificación de comportamientos anómalos se puede utilizar para detectar variaciones inesperadas en patrones de consumo o en la degradación de componentes en el sistema. En los últimos años, diversas técnicas de aprendizaje profundo se han aplicado de manera exitosa en la identificación y cuantificación de daño en distintos sistemas mecánicos. Por lo anterior, es de interés evaluar su uso para el análisis de desempeño en sistemas térmicos, en particular, técnicas especializadas para el análisis de series temporales.

Los sistemas solares térmicos son una fuente de energía viable y sustentable para aplicaciones de agua caliente a nivel domiciliario e industrial. Su operación requiere una correcta integración y mantención para efectivamente reducir el consumo de combustibles fósiles. Sin embargo, un sistema de monitoreo aumenta los costos del sistema, por lo que se deben tomar decisiones estratégicas para seleccionar componentes críticos a los cuales observar. Temperaturas y caudales en colectores solares, bombas y acumuladores de calor son las principales variables para analizar bajo diferentes condiciones meteorológicas.

El presente Trabajo de Título consiste en la evaluación de distintas técnicas de Aprendizaje Profundo para el desarrollo de un modelo de diagnóstico de detección de anomalías en sistemas térmicos. El caso de estudio utilizado es el sistema de agua caliente solar del edificio Beauchef 851, el cual es analizado y simulado con el software TRNSYS. A través de esta representación, es posible generar grandes cantidades de datos tales como temperatura, flujo y las condiciones ambientales para representar condiciones nominales y anómalas inducidas en el sistema. Se plantea utilizar técnicas de aprendizaje profundo para el análisis de información secuencial correspondiente a los datos generados a través de la simulación en TRNSYS.

Se evalúan diferentes técnicas para el análisis temporal como, por ejemplo, Redes Neuronales Recurrentes Profundas para predicción de temperaturas bajo variadas configuraciones y horizontes de evaluación. Esto, con el fin de desarrollar un método para la detección de anomalías en patrones de consumo, eficiencia de los colectores solares y operación de las bombas. El aumento de la temperatura registrada a la salida del campo solar causada por una alteración en la demanda de agua caliente es identificada como anomalía con una exactitud de un 86% en las muestras estudiadas. A su vez, la detección de la reducción de la misma temperatura debido a anomalías inducidas en la eficiencia del colector obtiene una exactitud de un 70%. A pesar de la sensibilidad del modelo de detección, estos resultados son prometedores ante la posibilidad de integrar mediciones y validaciones experimentales de este.

生き甲斐

- Ikigai.

Acknowledgments

En primer lugar, quiero agradecer a mi profesor guía, Dr. José Miguel Cardemil Iglesias, por el apoyo y la confianza que me ha hecho creer en mi futuro profesional y académico. Asimismo, agradecer a mi profesor co-guía, Dr. Enrique López Droguett, por el tiempo dedicado y la paciencia para responder mis dudas. Gracias por inspirarme a ir más lejos.

A mi familia que siempre ha estado ahí para mí, en especial mi tío Rodrigo quien me dio el sueño de estudiar ingeniería. A mi padre Leopoldo, por el ejemplo de superación personal y la evidencia de que las buenas intenciones son más fuertes que la adversidad. En especial y, por sobre todo, a mi madre Margarita Jullian, por haberme enseñado resiliencia y a pelear por lo que considero correcto, la curiosidad por el mundo y todos sus misterios. Solo me queda llevar a viajar por el mundo a esta Dama de Hierro, que lo dio todo por mí.

A mis amigos desde el inicio de los tiempos, Guillermo y Benjamín, gran parte de mi ilusión sobre la universidad se las debo a ustedes. A María José por ser la compañera que siempre me levanta, gracias por tu paciencia y empatía a esta ridícula. A mi gran familia de amigos de mecánica que siempre están ahí para sacarme una sonrisa y sentirme acompañada, por ñoños y por disfrutar el tiempo que tenemos. En especial, un abrazo a las *mecaniniñas*, para decirles a todos que esto no es solo un mundo de hombres. Finalmente, a Pablo, con quien seis años de compartir lado a lado no han sido suficientes para aprender de todo lo inmenso del mundo.

Quiero agradecer también al equipo de taekwondo de la FCFM. Determinación, dedicación y fortaleza para desafiarse a mejorar, todos los días, en todos los aspectos de la vida han sido conceptos fundamentales para mi crecimiento en estos últimos dos años. Eternamente agradecida, en especial del grupo de mujeres fuertes que me acogieron como una compañera más.

No me arrepentiré ningún solo día en haber entrado a la Universidad de Chile. Mi compromiso por luchar por el medio ambiente y el bienestar social nunca dejarán este pequeño cuerpo. La rebeldía se la debo a mis compañeras de infancia, Alexandra, y en especial, María Paz, de las que no importa la distancia, siempre estarán en mi corazón. No hay que saber siempre a dónde vamos, si sabemos dónde hemos estado. Las adoro.

En estos últimos años, también quiero agradecer a Guido, Piero, Rodrigo y Camila, quienes me adoptaron en su amistad sincera y pura, con sus perspectivas de mundo y conversaciones apasionadas y complejas en las cuales solemos perdernos. Y a quién me adoptó de una manera más literal que figurativa, gracias a Francia Martel, quien me abrió las puertas de su casa y corazón con total confianza.

Finalmente, a Ignacio Cofré, un sol que entró en mi vida para quedarse, al cual la distancia y el tiempo me convencen aún más de luchar por el camino que tenemos juntos por delante. Decir que me diste alas es poco, ya no tengo vuelta atrás. Te amo.

Gracias a todos por estar siempre ahí. Por hacerme lo que soy, estoy eternamente en deuda.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

Chapter 1

Introduction

Early fault detection is an important step in improving the availability of any equipment or mechanism. This allows to take appropriate maintenance measures in order to prevent further degradation and unexpected component failure. In the case of energy systems, monitoring variables which enable an in-depth analysis of the operational conditions constitute a fundamental step to improve their performance, by adjusting control systems or correcting integration schemes to increase their efficiency.

Improving the performance and reliability of renewable energy systems, seeking to decrease their investment and operational costs, is a high priority for scientists and engineers since it represents the biggest entry barrier to a market currently dominated by fossil-fuel based energy sources [1]. In the search of reliable, flexible and renewable energy sources, energy storage has been a focus of investigation in the recent years, such as lithium-ion batteries and molten-salts for solar photovoltaic and concentrating solar power plants, respectively. Nevertheless, other alternatives to increase the dispatchability of renewable energy sources reside in reliability analysis and fault diagnostics systems, which have been applied for decades in different mechanical systems in which monitoring is available. Timely identification of anomalies derived from malfunctions and failure in a system's components can increase operational time and energy outputs, for instance, in thermal plants where pressure and temperature are automatically controlled.

Fault diagnostics can be achieved through various techniques, from simulation and model-based approaches to data-driven methods. Machine Learning (ML) introduced a kind of data-driven models which rely less importantly on the analysis and simulation of the studied system. For instance, Artificial Neural Networks (ANN) have already been used to analyze thermal systems; particularly refrigeration, heat pumps, and solar thermal. Likewise, performance prediction and detection of anomalies during operation have been approached using ANN, with successful results, as described in Section 3.2 .

The growing successful cases in applying Deep Learning (DL) techniques for risk and reliability assessment in mechanical systems may offer an additional development path for reliability assessment in thermal systems. In recent years, modern architectures such as Convolutional Neural Networks (CNN) and Long Short-Term Memory (LSTM) Recurrent Neural Networks (RNN) have been successfully applied to fault detection in rotary machinery [2] and for the estimation of Remaining Useful Life (RUL) in lithium-ion batteries [3]. Thus, the use of deeper architectures designed to analyze temporal dependencies, such as RNN, could lead to achieve more precise results, help to understand periodic behavior, and evaluate relationships between the components in a thermal system.

1.1 General Background

During the last decades, the pursue for viable alternatives for a cleaner and diversified energy matrix has led to the study of new technologies, seeking to meet the growing energy demand through more efficient means and with lower environmental impact. While fossil fuel sources are traditionally burned for electricity generation and thermal demands, renewable energy accounts for only 13.4 % of the worldwide energy consumption [4]. Reducing world-wide fossil resources consumption is not only important from an environment-friendly point of view, but also from an economic standpoint. As its prices are highly sensitive to political struggles, investing in local energy generation is an opportunity for countries without natural access to fossil fuels to further develop.

In 2016, electricity generation from fossil fuels accounted for more than 65% of total world gross electricity production [5]. During the same year, almost 40% of global energy-related CO_2 emissions came from water and space heating in buildings and for industrial processes. In this context, the scientific community has devoted significant efforts in promoting research on alternative energy sources, aiming to replace the utilization of fossil fuels and decarbonizing energy generation [6]. These efforts are applied to meet either domestic, industrial or transportation energy demands. On one hand, wind and solar photovoltaic energy systems have grown significantly during the past two decades and as the technologies' cost continues to lower, they have become cost-effective when compared to traditional electricity generation methods. On the other hand, solar thermal systems are a viable alternative to supply thermal demands in locations where solar radiation is available. There are many widespread solar thermal solutions with potential to deliver energy loads for small and large-scale applications. In such systems, solar collectors are used to absorb solar radiation and convert it into useful heat, transported by a working fluid. This heat can then be used either to generate electricity, deliver heat to industrial processes, or for sanitary water demands.

Solar Hot Water (SHW) systems can be used to directly supply hot water for residential or commercial needs. They are frequently coupled with thermal storage, allowing to match the solar radiation availability and different demand profiles, which are commonly higher during non-insolation hours. As SHW systems are usually combined with auxiliary heating, such as gas or electrical boilers, an additional difficulty exists to correctly identify faults in a particular component. For example, if the hot water coming from the solar thermal system is not reaching the desired temperatures, the auxiliary system will provide the required energy. Even if the systems' costs are frequently inspected, no trustworthy information can be obtained of the systems' performance and energy production unless temperature and flow values are monitored. Therefore, faults may go unnoticed until the components are too degraded for simple maintenance procedures, elevating the overall costs and complexity of the installation. This can also have a negative effect on consumers' choices, distancing them from self-supplied energy sources and technologies in favor of more reliable and traditional alternatives.

Fault Diagnosis Systems (FDS) have been designed to detect, localize and assess damage present in a system. These methods have different approaches: qualitative or quantitative, depending on the knowledge there is of the system. These may also have a model-based or data-based approach. Model-based diagnosis systems heavily rely on expert knowledge of the installation and the underlying physics of failure that can exist in it. Thus, it is a high-cost approach which could deter investment if security is not a prime issue. For instance, thermo-economic methods combine concepts of exergy and economy to model the system, contrasting the predicted state of it with monitored data through complex thermodynamic analysis and economic assessment. Instead, data-driven models rely on specific monitored data and the operational history of the system.

Nowadays, together with the development of less expensive monitoring technologies, the availability of data such as temperature, pressure, vibration, rotational speed and other measured variables are used to study the health state of a system. As the computational power to analyze this information has grown, new techniques have been studied and developed aiming to get a better insight about the provided data. Machine Learning, as mentioned above, has been applied to a wide variety of problems and several different approaches have been developed depending on the quantity and quality of the available information of a system.

1.2 Motivation

Solar Hot Water systems, as well as other renewable energy sources, have been commercially available for decades, their popularity varying through the years depending on the economic scenario of fossil fuels and the influence of energy policies around the world. Since 2010, a tax franchise implemented in Chile grants a tax reduction for constructor companies who include SHW systems in new residential buildings, which value equals the installation and maintenance costs of the system. Thus, up to December 2017, the number of houses with solar collectors had risen over 112,000 under this policy, additional to the growth of the private sector [7].

However, the real impact of the increasing number of solar installations has not been quantified directly. Indirect estimations, such as the reduction of fossil-fuel consumption, do not contain enough information to assess if the SHW systems are working correctly. This is particularly important as malfunctions in the solar collectors can be hidden by the auxiliary heating system and thus can go unnoticed until complete failure. SHW systems are a commercially proven, efficient alternative for water heating for residential and commercial applications, reducing energy consumption and their emissions [8], but face performance issues when not correctly designed or integrated appropriately. Continuous inspection, monitoring and fault detection are of great importance for a good long-term performance of solar thermal systems [9].

Prognosis and Health Management (PHM) is an important component in modern engineering systems, in which algorithms are designed and used to detect anomalies, diagnose faults and predict future states of the system, for instance, the RUL. In turn, these can increase the system availability

and performance by managing maintenance issues [10]. In the past decades, there has been a proliferation of different approaches for this purpose, coupled with the development of Machine Learning applications in engineering systems to assist decision-making analysis. Data-driven approaches (DDA) have made data collection and analytics within diagnostic technologies essential components and high priority research topics [11].

This dependency on monitored data is not always beneficial. The lack of historical operational and failure data limits the wide range of techniques that can be applied for the system's health management. However, if the design conditions are known but historical operational data is scarce, model-based approaches are a viable alternative to study the system. Although, instead of basing the performance diagnosis system on a physical model of the system, it can be used as a method to construct artificial operational data as an initial approach to health management, as done in [12], [13]. This approach potentially reduces the time expended to obtain useful information of the system. Hence, even if failure data is not available, early anomaly detection is possible to implement if the system is continuously monitored.

Deep Learning techniques for PHM have been successfully applied to many engineering systems. Notable examples of rotary mechanical components and electrical batteries have been analyzed through measurable variables such as vibrations, acoustic emissions, temperature, state of charge and state of health, among others [3], [14], [15]. However, there is a limited amount of research which approaches performance diagnosis in thermal systems with these new techniques.

The SHW system installed on the roof of the western tower at the Beauchef 851 building at the Faculty of Physical and Mathematical Sciences (Universidad de Chile) is studied under this perspective. As part of the automated climatization system of the building, the components of the renewable subsystem interact with other traditional equipment to supply different heating and cooling loads present at the university building. However, there is little information about the performance of this subsystem. Therefore, the development of a diagnosis framework is a useful tool to study and assess this system, with the goal of increasing its availability and effect on the building's energy consumption.

1.3 Objectives and Scope

From the general context surrounding SHW systems and Deep Learning applications in reliability studies described above, the objectives and scope of this thesis are presented in this section.

1.3.1 General Objective

The main objective of this thesis is to assess Deep Learning techniques for the development of an anomaly detection and performance diagnosis framework applied to a thermal system.

1.3.2 Specific Objectives

To achieve the general objective, the following tasks must be completed:

- Characterize the performance of the solar hot water system at the Beauchef 851 building.
- Simulate the SHW system in TRNSYS software [16] to generate fault-free and fault-induced operational synthetic data.
- Evaluate Deep Learning applications for performance prediction of solar thermal systems. These include: ANN, RNN, and LSTM.
- Develop a prediction-based framework for anomaly detection in solar thermal systems.

1.3.3 Scope

The scope of this work, developed during the spring semester of 2018, is the assessment of three Deep Learning techniques for an anomaly detection framework, based on time series prediction in thermal systems. This is applied to a case study of a solar thermal hot water subsystem of campus using TRNSYS software to generate data that is not limited to time restrictions. Due to malfunctioning in the system and lack of reliable experimental measurements, the calibration and validation of the TRNSYS model is left outside of the scope of this work, although the model is constructed based on the original design of the system. For this reason, the TRNSYS model is a simplified version of the installed solar system and in the future requires experimental validation for an implementation of the anomaly detection framework. From TRNSYS software, temperature and flow values are extracted under meteorological conditions from April 7th to September 22nd, 2018.

From the generated databases, different architectures (ANN, RNN, LSTM) are implemented to assess their precision and accuracy to predict time-series using a many-to-one approach and extended prediction horizons. The construction and training of the models are implemented in TensorFlow [17]. Fine-tuning of certain hyperparameters, such as learning rates and the application of regularization techniques are not included in the analysis, emphasizing on principal characteristics such as number of units, layer and activation functions of the models. This work is focused on the prediction of two relevant temperatures in the solar thermal system and two anomalous behaviors due to induced faults developed in the TRNSYS model of the system. The anomalies are detected through the study of the prediction errors of the models trained with healthy data when presented with fault-induced measurements.

Results are presented in tables and figures for the three main themes of this work: construction of the TRNSYS simulation, training and test for different DL framework configurations and the implementation of anomaly detection based on time-series reconstruction errors.

1.4 Outline

The remaining of this document is structured as follows: Chapter 2 contains a theoretical background review covering crucial aspects of SHW systems and DL applications in systems' health management. Chapter 3 is a review of related research, focusing on different FDS and their applicability to health management. Following the theoretical background of this work, the methodology is presented in Chapter 4. Chapter 5 details the SHW system, available information and simplifications applied to the construction of the TRNSYS model. Chapter 6 refers to the design and selection of DL techniques applied to the studied case. Chapter 7 contains main results and analysis. Finally, comments, conclusions and future work are stated in Chapter 8.

Chapter 2

Background

In this section, the specific background is presented and discussed, concerning components and possible failures during operation of solar thermal systems. This is followed by a brief description of Deep Learning theory, definitions, techniques and the applications of modern architectures.

2.1 Solar Thermal Systems

SHW systems can be applied for residential, commercial and even for industrial use to supply thermal demands. At least 20% of consumed energy in an average household is used for water heating [18], and thus, the study and optimization of this technology is a significant step in the quest to reduce dependency on fossil-fuels. A general scheme of an indirect heating system is presented in Figure 1 [19]. The main component is the solar collector, which acts as a heat exchanger between the incident solar radiation and a working fluid. To supply the thermal demand, the hot water may be directly dispatched from the solar collectors to the plumbing system or indirectly through a heat exchanger. In this figure, an indirect heating circuit is shown. Here, the primary circuit transports heat absorbed in the solar collector towards the hot water tank, through a coil heat exchanger immersed in the tank. From this tank, the hot water is dispatched as required. Additionally, an auxiliary heater in the form of an immersion heater inside the hot water tank is installed, allowing the secondary circuit exiting the tank to meet the hot water demand when the solar resource is low or not available. A brief description of the main components in a SHW system are discussed in this section.

2.1.1 Solar Collectors

Solar collectors' basic operating principle consists on the heating of a fluid with the incident solar radiation on a surface. To meet low temperature demands required by the target system, non-concentrating flat plate or evacuated tube collectors are commonly used. To increase the heat gained by absorbed radiation while reducing heat loss towards the environment, specific materials are selected by their optical properties for the construction of these collectors. Non-concentrating solar collectors can use both beam and diffuse radiation and thus, do not require tracking mechanisms to follow the sun's path [20]. Thus, the solar geometry of the installation is determined by the tilt angle and collector's orientation (azimuth).

Applications range from solar water heating, building heating, air conditioning and industrial process heat. These can be installed as passive systems or as forced-convection systems, such as the one presented in Figure 1. For annual hot water production, the optimum tilt angle is equal to the latitude of the location, with angle variations of ±10° to ±15° depending on the application [21].

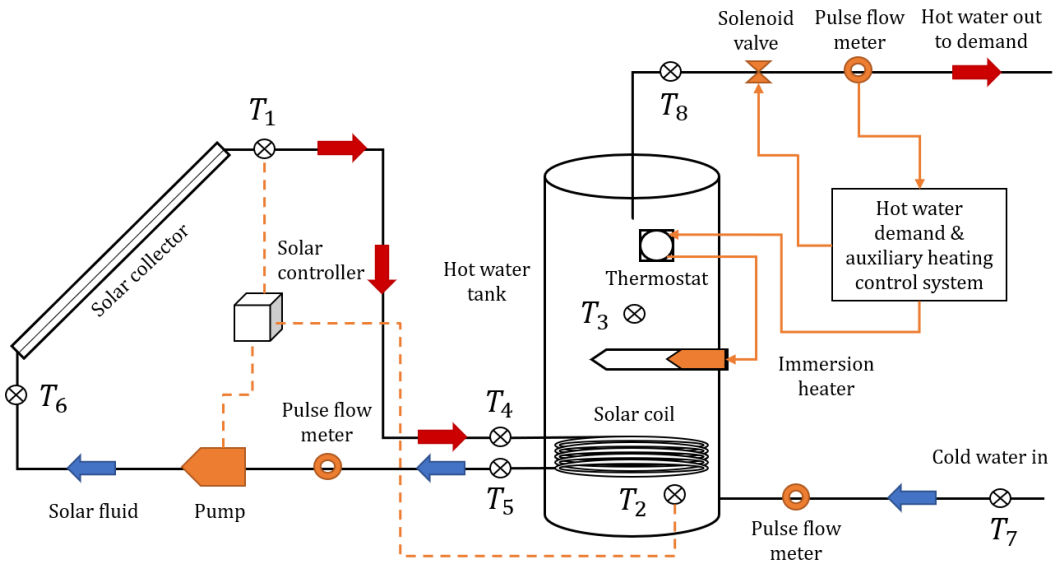


Figure 1: Solar Hot Water system diagram. Adapted from [19].

These stationary technologies generally operate at maximum temperatures around 90°C but can reach temperatures above 150°C when increasing the solar collector system's pressure or using other thermal fluids instead of water. At a given pressure, the minimum and maximum operating temperatures are limited by the freezing and boiling points of the working fluid, respectively. The heat transfer mechanism between the heated work fluid and the target system can be of two types:

- Direct heat transfer: the same working fluid that is transported in the interior of the tubes is the one delivered as a source of heat to the system
- Indirect heat transfer: if the ambient temperatures can reach the freezing temperature of the working fluid at the nominal pressure of the system (for example, under 0°C for water), an indirect heat transfer configuration is employed.

Both flat plate collectors (FPC) and evacuated tube (ETC) collectors are constructed with a transparent cover and an absorber surface, along with thermal insultation and fluid tubes designed for their specific geometry depending if the installation is of natural or forced convection. Solar radiation passes through the cover, which is transparent for shortwave radiation, towards the absorber which present high absorptivity and low emittance. The cover is opaque to the longwave thermal radiation reflected and emitted from the absorber, effectively capturing heat within the collector. Heat is then transmitted to the fluid by the absorber through different geometries as shown in Figure 2 and Figure 3 for FPC and ETC collectors, respectively.

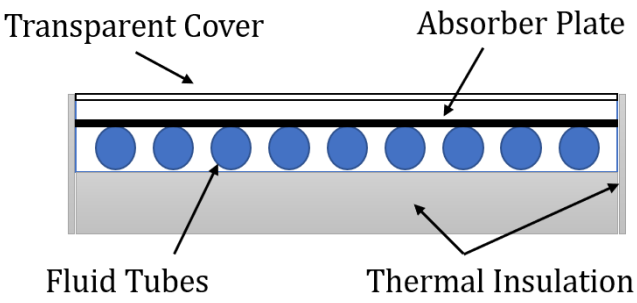


Figure 2: Flat Plate Collector General Geometry.

To effectively compare the efficiency of different solar collectors, boundary conditions such as ambient temperature and incident radiation must be taken into account. Thus, test and performance indicators, such as the efficiency curves, are defined by ISO 9806 and EN 12975 standards, among others [22]. Parameters such as the total area of absorption, rated flow, efficiency coefficients and Incidence Angle Modifiers (IAM) are fundamental to characterize a solar collector and its expected performance under different weather conditions.

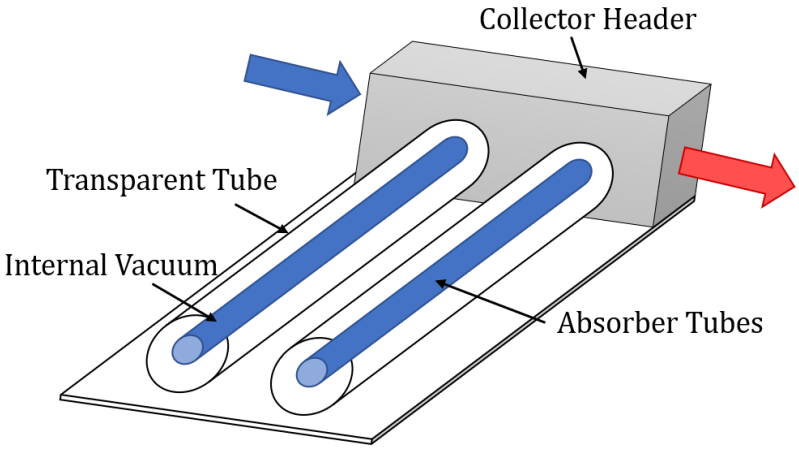


Figure 3: Evacuated Tube Collector General Geometry.

The analysis of the collector's energy balance shows distribution of incident solar energy into useful energy gains, thermal losses and optical losses. The equation which describes this balance in a FP is shown as follows:

$$Q_u = A_c F_R [S - U_L(T_i - T_a)] \quad (1)$$

The initial available solar radiation is reduced by optical losses, which is determined by the physical properties of the collector. The effectively absorbed solar radiation per unit area S is further reduced by the thermal losses to the surroundings through conduction, convection and infrared radiation

mechanisms, represented in Figure 4. These losses, combined in the overall thermal loss coefficient U_L, also depend on the difference between the fluid inlet temperature T_i and ambient temperature T_a. This implies the existence of a maximum useful energy gain between the balance of available energy and thermal losses. Assuming steady state, the useful heat gains through an area A_c is the difference between the absorbed radiation and thermal losses, Q_u.

The heat removal factor F_R is used to express the relation between the actual useful energy gain to the ideal situation in which the whole collector was at a temperature of T_i,

$$F_R = \frac{\dot{m}\, C_p\left(T_{f_o} - T_{f_i}\right)}{A_c\left[S - U_L\left(T_{f_i} - T_a\right)\right]} \tag{2}$$

Where T_{f_i} and T_{f_o} are the inlet and outlet fluid temperature, $\dot{m}$ is the total collector flow rate and C_p is the specific heat capacity of the fluid.

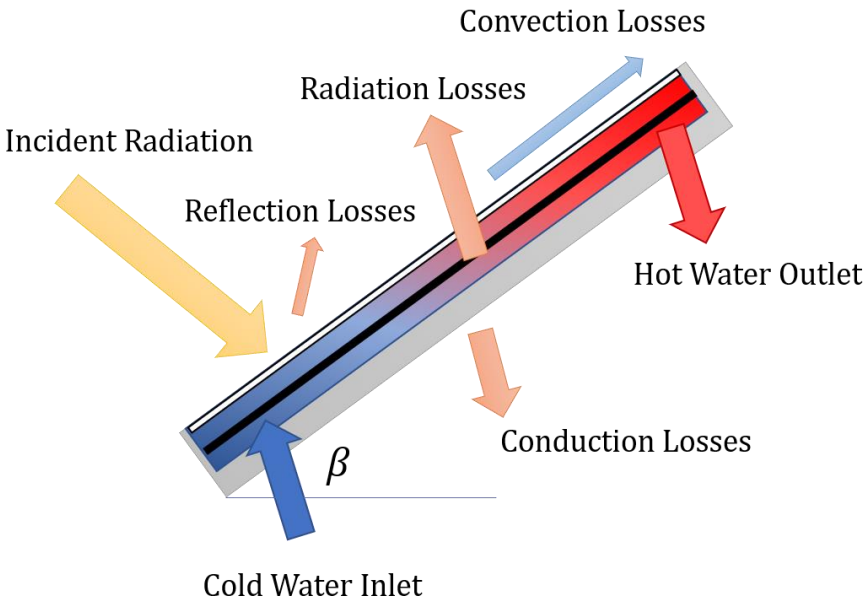


Figure 4: Solar thermal collector heat flow diagram.

From the development and analysis of Equation (1), the efficiency of a solar collector can be described by the following expression.

$$\eta = F_R \tau\alpha - F_R U_L \left[\frac{T_i - T_a}{G}\right] \tag{3}$$

Here, the first term expresses the heat gain of the system, which depends on the heat removal factor F_R of the collector and its optical properties: transmittance τ and absorptivity α. The second term refers to the heat losses in the system. This also depends on F_R and the overall loss coefficient U_L depending on the geometry and material choice, the temperature difference between two points in the collector and the incident global radiation G.

The collectors' efficiency curves expressed in Equation (3) may be presented in another format by the manufacturer, as shown below. Both equations represent the difference between the heat gains and losses in the system. In this case, a_0 represents the optical efficiency, while a_1 and a_2 are the first and second order loss coefficients, respectively.

$$\eta = a_0 + a_1 \cdot \frac{(T_i - T_a)}{G} + a_2 \cdot \frac{(T_i + T_a)^2}{G} \quad (4)$$

2.1.2 Evacuated Tube Solar Collector

The Evacuated Tube Collectors (ETC) consist of a series of double glass tubes with an internal vacuum cavity between them, as shown in Figure 3. Inside the inner tube or absorber, the working fluid is heated by the incident radiation, while the vacuum and the optical properties of the outer tube reduce heat loss to the environment. Thus, even though the collectors' cost is higher and, as the FPC, its efficiency diminishes with larger temperature differences between the collector and the environment, ETC perform better than flat plate collectors under cold ambient temperatures [23]. This enables the system to reach higher temperatures and supply a wider range of thermal demands.

As mentioned above, ETC can also be used for direct or indirect heat transfer. Heat extraction can be through a U-pipe, heat pipe or direct liquid contact [24]. An alternative for indirect heat transfer in ETC is the use of Heat Pipes, small copper tubes placed in the interior of ETC. Inside the Heat Pipe a mixture of water and refrigerant is used as working fluid, increasing the boiling point and reducing melting points. In this case, the incident solar radiation can evaporate the fluid within the pipe and then its latent heat is transferred to the solar fluid, as presented in Figure 5 [25]. There are many different configurations for the position of the Heat Pipe, its attachment to the outer tube and the entry of the phase-changing fluid to it.

However, the highest temperatures that can be reached in a solar collector are limited by its stagnation temperature, a point where the heat gains are equal to the environmental losses. Its value varies with the design and materials used, and for ETC it is generally found between 180°C and 250°C. To avoid reaching this point, heat dampening mechanisms can be integrated to the system, avoiding overpressure during the warmest months of the year.

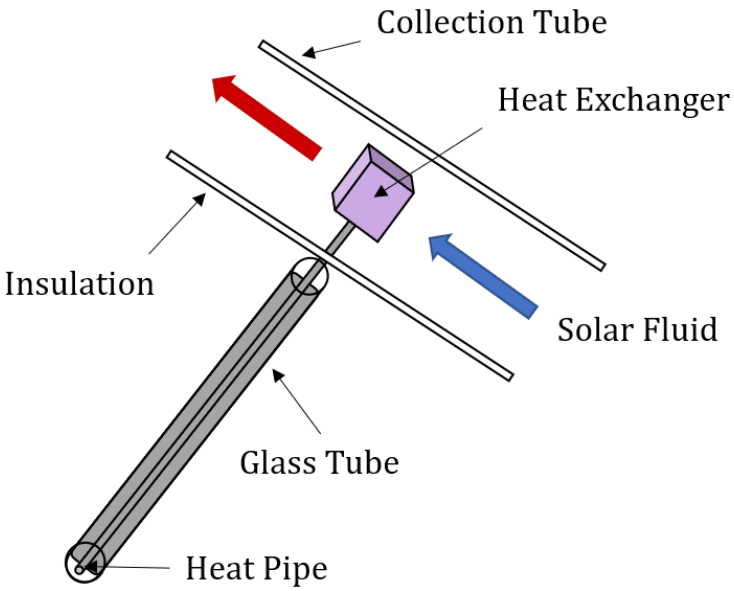


Figure 5: Heat Pipe Evacuated Tube Collector. Adapted from [25].

2.1.3 Other Components

As shown in Figure 1, another main component of a SHW system is a thermal storage unit, represented by a hot water tank or a heat accumulator. This device has the objective of dampening temperature changes in the system and allowing to dispatch energy demands during the hours of the day when incident radiation decreases.

Also, due to the variability and uncertainty associated with the solar resource, it is of interest that a SHW is coupled to an auxiliary heat generation system, ensuring that the energy demands are met even in cloudy days. In Figure 1, this is represented by a submerged heater in the storage tank, however, the auxiliary heat source can be a natural gas boiler or heat pumps, among other conventional devices.

For systems of higher complexity, the use of heat exchangers is required. A frequently used configuration are plate heat exchangers, which are compact devices presenting high heat transfer areas in relation to their size. Other required components are centrifugal pumps and different valves used to control the flow and pressure differences which naturally occur within the system.

Finally, it is of interest to integrate an automatic control system that regulates the operation of the solar thermal system based on solar radiation availability, working hours, temperatures and pressures measured in its components.

2.1.4 Faults and Failures During Operation

In solar thermal systems, a significant difficulty is to identify the existence of alterations or anomalies in their operation, since they are usually coupled to auxiliary heat-generating systems. Heat underproduction may be associated with unexpected meteorological conditions, degradation, or failure in one or more of its components [9], [26]. For example:

- Pump and valve failures. The most common sensors employed are temperature and differential pressure sensors. Therefore, due to the limited availability of flow meters, no other relevant information about the flow can be extracted.

- Heat loss. Degradation of thermal insulation components can only be determined by strict monitoring of surrounding temperatures and heat fluxes. This applies to heat exchangers, hot water tanks, piping, among others.

- Collector failure. Depending on the type of solar collector, many different failure modes can be observed. In the case of evacuated tube heat pipe collectors, pressure loss or over-pressure may result in fracture and consequent evaporation and leakage of the internal working fluid. Heat loss from insulation degradation is also possible.

- Issues with the control system may also arise. Faulty temperature sensors or actuators, sensors placed in inadequate positions, or problems with their internal configurations may obstruct heat production and effective energy transfer to supply the hot water demand.

These examples of failures during operation can occur unnoticed if the auxiliary circuit is not monitored. Several authors have analyzed different approaches for identifying and diagnosing the health of thermal systems, some of which are reviewed in section 3.2

2.2 Machine Learning and Deep Learning

Since the 1990s and early 2000s, Machine Learning (ML) techniques became a popular method for data analysis and decision-making tools. Based on the available information, a ML model can *learn* from the data, extracting abstract relationships within the studied variables to classify or predict future values. Many Prognosis and Health Management (PHM) studies have been addressed with ML applications. As monitoring grew as a widespread tool in security management and computational hardware advances allowed massive and complex data analysis, engineers began to study and implement such algorithms in different areas. As mentioned before, ML addresses challenges such as classification and regression, seeking to recognize and replicate hidden patterns within available data. For instance, such tasks can be to classify the health state of a system or to obtain estimated future values of a desired variable.

At present, a variety of ML techniques exist, designed for specific purposes and datasets. To *learn*, a parametric algorithm must adjust its parameters to the training data provided. This is done through the minimization or maximization of a cost function, which depends on the task at hand. The performance of a model will be affected by the architecture or hyperparameters selected beforehand, as well as the quality of the training data. The learning process or *training* depends on the previous knowledge of the analyzed data: while Unsupervised ML has no previous information on how the data is related, Supervised ML methods require input data and corresponding labels to complete the desired task.

2.2.1 Neural Networks

Artificial Neural Networks (ANN) are a common technique used in both Supervised and Unsupervised ML for both Regression and Classification. These are roughly inspired in how biological neurons function, where an external stimulus is propagated through the neural network, from sensory to response units. A basic structure is shown in Figure 6 [27], from where the weighted sum of input signals is used to determine the desired output of each neuron.

The basic mathematical representation of an ANN considers a series of matrices and vectors (weights and biases) representing the transitions between the input, hidden and output layers of a network. In this case, the training of the model consists of an optimization process which seeks to minimize the error of the predicted output when compared to the label of the original data. This is usually done by propagating error from the output to the input layers, adjusting the weights and biases within each layer to correctly represent the given data (W and b in Figure 6, respectively). This is known as backpropagation, by which chain rule and gradient descent adjust these parameters as presented by Rumelhart et. al. [28]. Then, after the NN is trained, it can interpret new unseen data input and replicate the learned behavior. The accuracy of the model will depend on the chosen architecture hyperparameters, the nature of the data and learning process.

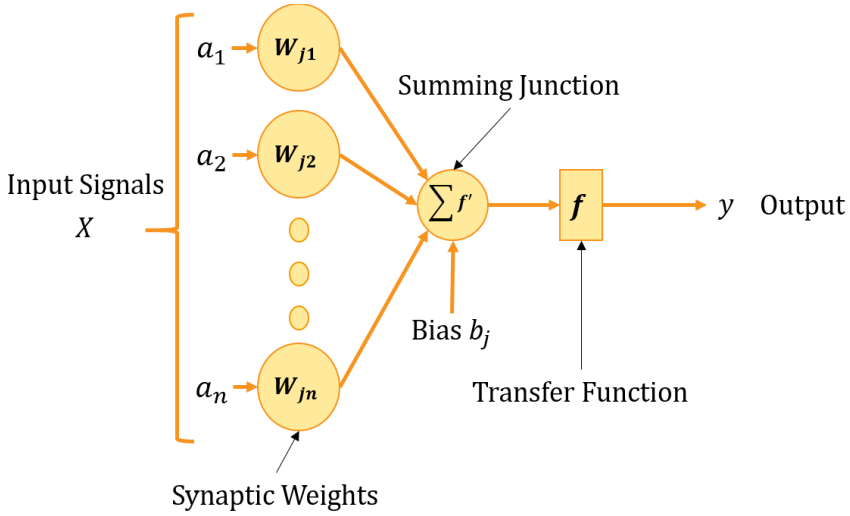


Figure 6: Basic structure of a Feed Forward Neural Network. Adapted from [27].

This can be summarized in the following equations where y is a vector of known labels and y_n the predicted output. Frequently, the mean square error (MSE) is used as the target function to be minimized in regression tasks. The output of the network y_n depends on the input data $\boldsymbol{X}$ and the weights and biases of each layer. A non-linear activation function σ is applied between layers to transmit information according to its nature, for example a sigmoid, tanh or ReLU function.

$$min(f) = MSE(y_n, y) \tag{5}$$

$$y_n = \sigma(W^t \boldsymbol{X} + b) \tag{6}$$

2.2.2 Deep Neural Networks

Since the design of basic ANN, several architectures have been developed to tackle a variety of tasks. In this context, Deep Learning approaches have become a widely-used tool in several applications of ML tasks, such as object detection, speech recognition and replication [29]. These architectures have matured to approach highly complex and non-linear tasks.

For instance, in Multi-Layered Perceptron (MLP), depth is added to the basic ANN network by stacking several hidden layers, as shown in Figure 7, where consecutive non-linear mathematical operations are applied to each layer. Equation (6) is then rewritten considering the output of each layer h_i depends on the output of the previous layer h_{i-1}.

$$h_i = \sigma(h_{i-1}^T W_i + b_i) \tag{7}$$

Other complex deep architectures, such as Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN), have been designed to focus on specific feature extraction techniques for spatial and temporal relationships. Due to the large number of parameters the deeper models require to represent the data, big datasets need to be analyzed to correctly train the model. This, as well as the complex mathematical operations involved, such as convolutions or internal cycles, increase training and testing time. For this reason, the use of DL algorithms has become popular during the last decade, since the increased development and accessibility to GPU and other computational components at lower costs.

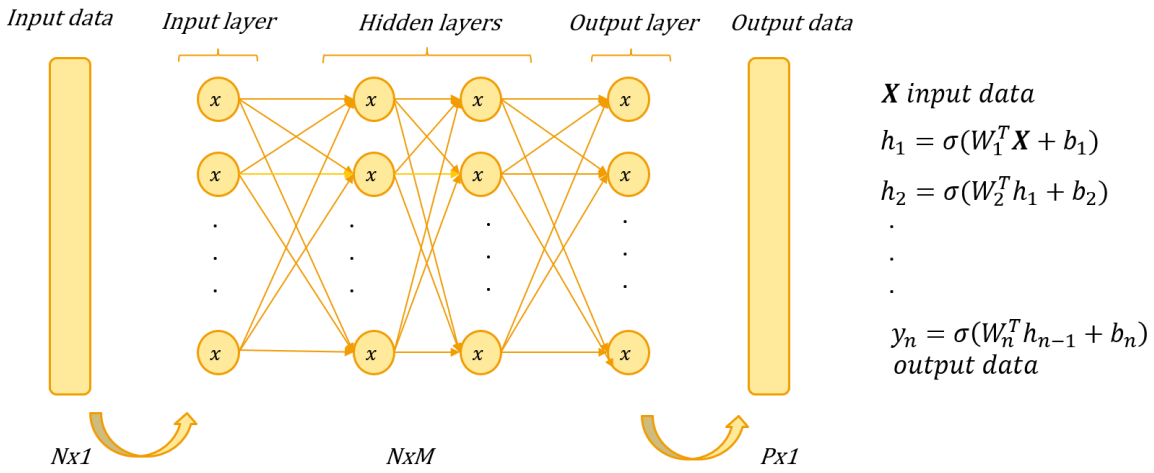


Figure 7: Example of a NN structure.

The success of deep architectures resides in their ability to obtain higher levels of abstraction from the input data. This property enhances its feature extraction capabilities and, due to its hierarchical architecture, it also helps the model converge faster to a good solution. Thus, Deep Neural Networks (DNN) have proven to be more powerful to address specific ML tasks than single-layered ANN.

2.2.3 Recurrent Neural Network

Deep Learning frameworks have been used to model and predict different physical phenomena, such as weather patterns, health management and medical diagnosis, among others [11], [30], [31]. Within the most common applications, RNN are particularly well adapted to analyze time series [29]. These models are, in principle, able to create and process memories of arbitrary sequences of input data, mixing sequential and parallel information processing with efficiency [32].

A major drawback of standard neural networks is the underlying assumption that training and test examples are independent. Thus, even though DNN have reached outstanding sequencing results without directly modeling time, it has limited power over long-range dependencies [33]. However,

in an RNN, each unrolled layer will receive the current input data $x(t)$, as well as the previous hidden state of the node $h(t-1)$. The output $y_{pred}(t)$ will depend on the hidden value of the current time step $h(t)$ and an activation function g. This is depicted in the following equations, while the unfolded temporal structure is presented in Figure 8.

$$h(t) = \sigma(W_x \cdot x(t) + W_r \cdot h(t-1) + b_h \tag{8}$$

$$y_{pred(t)} = g\left(W_y^t \cdot h(t) + b_y\right) \tag{9}$$

It is observed in Figure 8 that each layer in the unfolded network will represent a time-step with shared weights for input, hidden and output data (W_x, W_r, W_y) and biases in hidden and output layers (b_h, b_y) (See equations (8) and (9)). Thus, the learning algorithm must be modified to a backpropagation through time (BPTT), which will operate similarly to the gradient descent to adjust the network's variables [34].

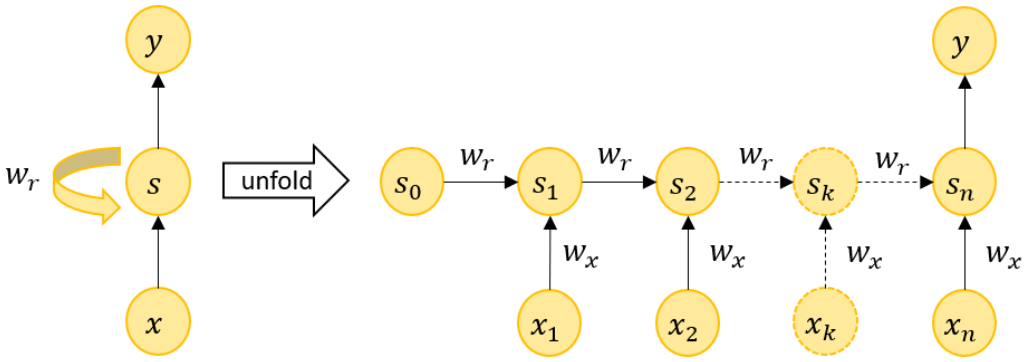


Figure 8: Unfolded architecture of a recurrent neural network.

However, for long-range dependencies *vanishing* or *exploding* gradients may appear when propagating errors towards the inner layers [35]. These optimization problems increase the computational demand of using these techniques. While truncated backpropagation through time (TBPTT) offers a partial solution, setting a maximum number of steps for the propagation, it is still limited for long-range dependencies. Thus, other architectures, as well as the use of regularization techniques, have been proposed to address this challenge.

For instance, both Long Short-Term Memory (LSTM) and Gated Recurrent Units (GRU) introduce a memory cell designed to overcome the issue of vanishing gradient, replacing the nodes in the hidden layers [36]. These have also been combined in different applications, such as natural language processing and estimation of RUL in mechanical components [37], [38].

2.2.4 Long Short-Term Memory RNN

The classical architecture of Long Short-Term Memory RNN consists on the interaction of *gates* or processes within a memory cell, each of which are independent RNN. Long-term memory is stored in what is called the *cell state* and is selectively updated with new input data stored in the short-term memory [39]. LSTM architecture has been widely and successfully applied to a variety of ML problems such as Natural Language Processing (NLP). In industry, the most commonly used variant known as Vanilla LSTM has been used to estimate the RUL of different components, which is one of the most challenging and significant decision-making indicators for security and maintenance issues [40].

In Figure 9 a schematic representation of Vanilla LSTM is shown. Here, three *gates* control the memory update from a previous state at timestep $t-1$ and the new input data at timestep t. Each gate has its own set of weights (W, U) and biases (b) of a regular RNN and use a sigmoid activation function. From left to right in the center *cell* of Figure 9 there are the forget gate, input gate and output gate. Each cell has three inputs: the previous cell-state c_{t-1}, the previous output h_{t-1} and the new data x_t. There are two outputs from the cell: the present cell state c_t and the present output of the hidden state h_t.

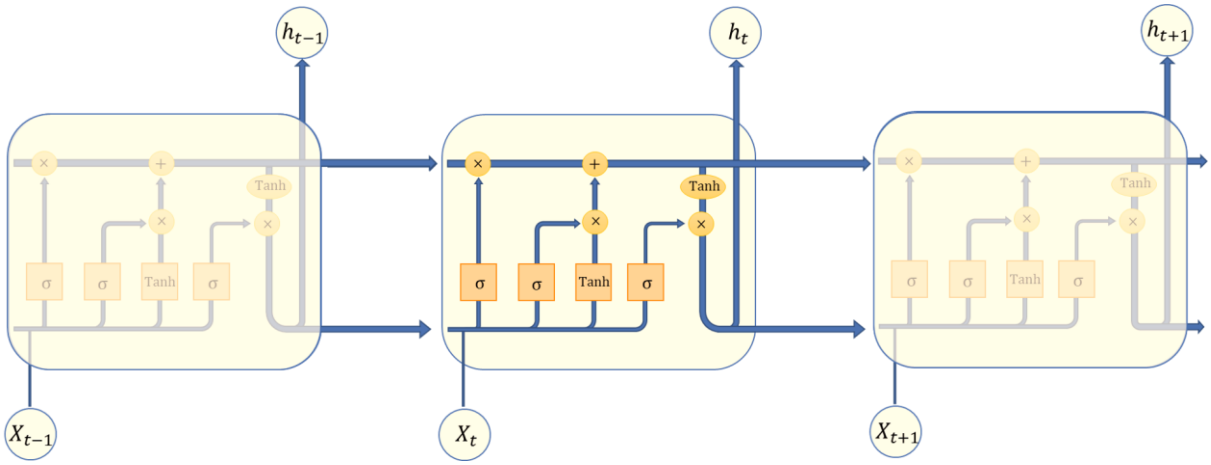


Figure 9: Vanilla LSTM architecture.

When a new input x_t enters the memory cell, the forget gate will select the information to be thrown away from the previous cell-state c_{t-1}, represented as follows.

$$f^t = \sigma\left(W_f x^t + U_f h^{t-1} + b_f\right) \tag{10}$$

The input gate, on the other hand, selects the information that will be updated while a new cell-state candidate is built from the input data. This is represented by the following equations.

$$i^t = \sigma(W_i x^t + U_i h^{t-1} + b_i) \quad (11)$$

$$a^t = tanh(W_c x^t + U_c h^{t-1} + b_c) \quad (12)$$

A combination from the forget gate (Equation (10)) with the previous cell-state c_{t-1} and from the input gate (Equation (11)) and the new candidate (Equation (12)) will create the new cell state through Equation (13).

$$c^t = f^t \cdot c^{t-1} + i^t \cdot a^t \quad (13)$$

Finally, the output gate selects which part of the cell-state obtained with Equation (13) will be used as a useful output h_t.

$$o^t = \sigma(W_o x^t + U_o h^{t-1} + b_o) \quad (14)$$

$$h^t = o^t \cdot tanh(c^t) \quad (15)$$

Other variants to LSTM have been designed, such as the GRU. This simplified version merges the internal state with the hidden state and the forget gate and input gate [29]. In [40] it was shown that although LSTM performs better for short-term prediction and GRU for long-term prediction, the difference is not significant, and both outperform the standard RNN. Additionally, Bidirectional Recurrent Neural Networks (BRNN) introduces inputs from previous and future timesteps to predict outputs [41].

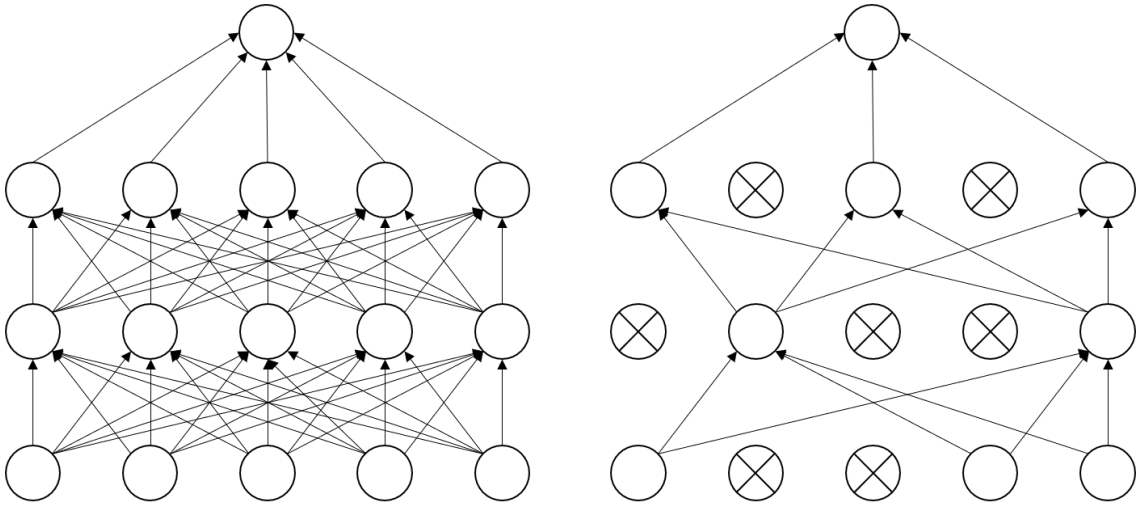
Finally, even greater complexity can be reached when stacking several memory cells in sequence, forming Deep RNN cells. Deep LSTM (DLSTM) have shown successful results, even when raw sensory data is directly fed into the model. In [42], Zhao et al. argues that when analyzing sequential data, even though the use of multi-domain feature extractions such as statistical, frequency and time-frequency features have reached satisfying results with other models (such image-processing ones as CNN), they do not model the intrinsic sequential nature of it. Additionally, they require intensive expert knowledge or feature engineering to extract input data.

Alternatively, to increase the analytic power of sequence-based models, stacking memory cells using DLSTM allows for a better abstraction of temporal relationships within the data. By enabling the model to learn characteristics of the raw signal at different time scales and distributing parameters in layers, this model achieves better results than Vanilla LSTM when predicting tool wear from raw input data.

2.2.5 Regularization Techniques

Deep architectures usually display high degrees of freedom and during the training process, the parameters can be over-adjusted to replicate the initial training set. Overall, this reduces the model's performance with unseen test data.

Dropout is a popular technique used to prevent overfitting of deep neural networks, improving its performance on supervised learning tasks. In Figure 10 [43] a representation of a stochastic regularization process for weight training is shown. Dropout uses a Bernoulli distribution with *p* probability of a node in a layer being removed, reducing the inputs for the next layer. This probability is also a hyperparameter that must be selected appropriately, optimizing between lower performance due to overfitting and increasing the training time.



(a) Standard Neural Net (b) After applying dropout

Figure 10: Dropout Neural Net Model. Adapted from [43].

Chapter 3

Literature Review

In this section, a selection of bibliographic works is revised and discussed, concerning the general context and methods used for fault detection in thermal systems. Studies on Deep Learning applications to thermal systems and fault detection models applied to thermal systems are revised.

3.1 Performance Prediction Models in Solar Systems

Machine Learning algorithms have been applied to describe and predict the performance of thermal systems and solar energy applications during the last decade. Their use is mostly focused on models seeking to predict values of solar radiation under different weather conditions and performance indicators of different components, such as solar collectors, heat pumps and hot-water systems.

Ghritlahre et. al. [27] presented a literature review related to performance predictions of solar collector systems, in which ANN techniques have been implemented. The work describes references for over 30 different representations which seek to model and predict thermal performance of various types of thermal systems. Some notable examples are the following:

- Flat plate and evacuated tube solar collector performance [44]–[46].
- Performance prediction of solar hot water systems [47]–[49].
- Performance prediction of other thermal components, for example solar assisted heat pumps [50] and thermal energy storage [51].

The authors highlight the flexibility of the ANN models to interpret complex and incomplete data, while significantly reducing the calculation time required to predict performance and obtain intrinsic characteristics of the system. Different types of neural models such as Multi-Layered Perceptron (MLP), Radial Basis Functions (RBF) and Adaptive-Network-Based Fuzzy Inference System (ANFIS) are used for performance prediction in the aforementioned works. They also present a brief list of topics which could be developed in future works regarding optimization and comparison between proposed models. However, the architectures used in these works (up to 2016) are far from the current state-of-the-art on ANN and DL as mentioned in section 2.2.3 before.

Additionally, ANN has been used in combination with Genetic Algorithms (GA) to optimize design variables in ETC U-pipe solar collectors. In [46], parameters such as the collector tilt angle, absorber plate and glass cover emissivity were optimized. Ideal meteorological conditions were also identified, such as wind velocity, ambient temperature and solar radiation, reaching a thermal

performance of 83.5 %. In [52] an extensive database of commercially available water-in-glass evacuated tube solar collectors (WGET-SWH) was used to train a model to predict heat collection rate and heat loss coefficients. Model inputs include tube length, number of tubes, tank volume, collector area, tilt angle and final output temperature. This model was then used to combine said parameters and optimize the design of solar energy systems.

Several reasons may explain why complex DL architectures, such as LSTM, have not been explored for solar thermal applications. For example, in 2018 a comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data was presented [53]. This work arises from the analysis that LSTM had not been applied to forecasting irradiance, whereas shallow ML algorithms (Support Vector Machine (SVM), Random Forest, among others) and standard ANN (Feed-Forward Neural Networks (FFNN), RBF) have. However, the results of this work aim to obtain accurate forecast for photovoltaic (PV) based energy plants. This is due to the fact that PV power production, stability and storage dimensioning is strongly characterized by fluctuating outputs influenced by instantaneous meteorological conditions. Predictive features are a critical issue for the electrical grid management; and, as shown recently, also for smart grid applications: planning, storage system sizing and market participation of variable renewable energy sources [54].

Similarly, in solar thermal systems power production and heat generation are directly affected by the variation of meteorological conditions. However, as solar thermal energy presents a natural inertia, lower precision is required to obtain acceptable results compared to PV plants. For power production, Rodat et. al in [55] presents a model combining meteorological forecasts and thermal performances of Fresnel solar collectors for heat storage management. Additionally, there are also many model-based meteorological forecast methods, including estimations and predictions for global, direct and diffuse radiation used for solar thermal energy production.

3.2 Fault Diagnostic Systems

Fault Diagnostics Systems (FDS) mainly include fault or failure detection and localization through the evaluation of the operational conditions of a system. Several methods, architectures and applications have been developed and are classified into three broad categories: quantitative model-based methods, qualitative model-based methods, and process history-based methods [8]. The method used depends on the available information of the studied system.

Traditional methods are usually model-based. For example, in [56] a *thermoeconomic* analysis is applied by combining economic and thermodynamic analysis based on the Second Law of Thermodynamics, through the concept of exergy cost. Performance degradation diagnosis of thermal plants based on advanced exergy analysis is used for detection and localization of developing faults [57]. In [26] a review of long-term fault detection approaches in solar medium and large thermal systems is presented, where several model-based methods of manual and

automatic fault detection from monitored data are compared, along with their estimated cost of implementation. Generally, as the size of the thermal system grows, so does the cost of monitoring and processing the results from model-based diagnosis methods.

The use of ML and DL algorithms have reduced the dependence on model-based systems and do not require a full comprehension or simulation of the studied system. Based on monitoring data, detection of faulty states may be solved by regression, where fault-free states are predicted, or by classification. These methods need data from healthy, deteriorated and faulty states correctly labeled, or alternatively require a method to classify faulty states. Many ANN-based systems are reviewed in [58].

3.2.1 Anomaly Detection and Diagnostics

An anomaly is defined as any point in time in which the behavior of a system is unusual compared to regular past behavior. This does not necessarily imply a fault or failure, as fluctuations naturally occur under varying operational conditions, but are regularly monitored for safety reasons in complex systems. Anomalous behaviors can be classified to known fault conditions through diagnostics modules, however generic algorithms require previous knowledge of the connection between anomalies and failure modes. Anomaly detection is then defined as the decision on whether the monitored behavior is experiencing unexpected performance of an intended function [11]. Further information is required to link the anomalous behavior with a known fault and whether the fault may lead to a failure state in the system.

One of the first works which integrated ANN with FDS in solar thermal installations was developed by Kalogirou et al. [12], whom developed an automatic solar water heater (SWH) fault diagnosis system consisting on a prediction module, residual calculator and diagnosis module. A diagram of the studied system is presented in Figure 11, where all monitored temperatures are indicated. Predictions of fault-free temperatures under different environmental conditions are obtained with ANN and compared with monitored data to identify faults in the solar collector and pipe insulation. Free-fault data is simulated in TRNSYS with weather data from two locations in Cyprus and France and used to train an MLP of three hidden layers of 19 neurons each. Data inputs and the predicted output for the prediction module is shown schematically in Figure 12. The residual module compares measured data to outputs predicted by ANN, using the relative error as a metric between temperature differences.

Finally, in the diagnosis module, the residuals are compared to constant thresholds values and classifying the operation into four categories from the magnitude of the residual: normal, low probability of failure, high probability of failure, and failure. The thresholds are determined manually, compromising between false alarms and early detection of the failure. There are two detection criteria: when five consecutive states are identified as the same type of fault, and when the average value of ten consecutive states overcomes a threshold (M10a in Figure 13). The FDS is then tested with two scenarios: a 10 % decrease of the collector efficiency factor, and a 10 %

increase of U value of pipe insulation. In both cases, the system is sensible to 5 % variance and the M10a procedure is more stable and can detect earlier stages of deterioration. It must be noted that flow availability is a binary [0,1] type of data and that the diagnosis module cannot identify the cause of the failure, just the general location given by the monitored values.

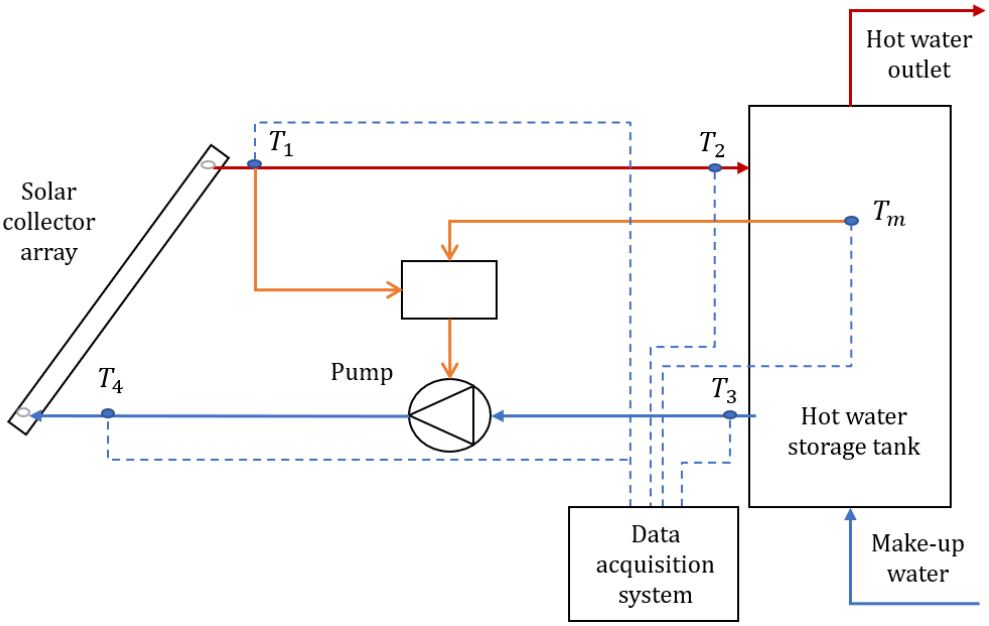


Figure 11: Schematic diagram of the simple solar system. Adapted from [12].

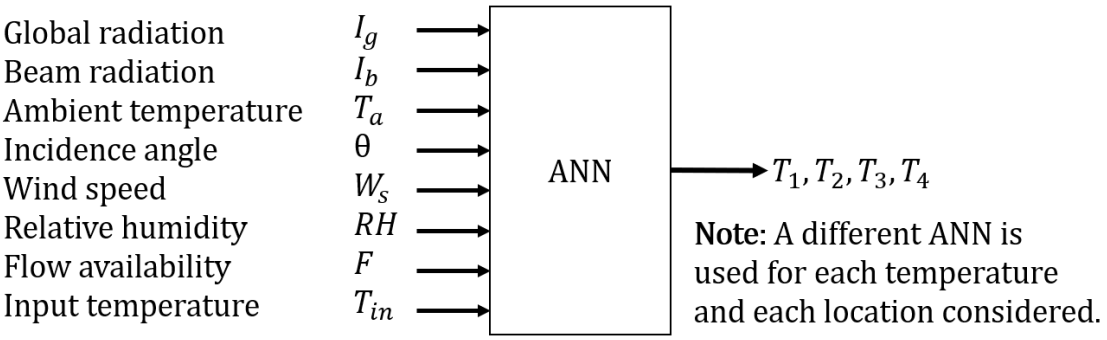


Figure 12: ANN design concept. Adapted from [12].

Another example is the application of Adaptive Resonance Theory Neural Networks (ART) to monitor SHW systems and detect existing or developing faults [8]. ART architecture also has a type of memory cell, used to recognize patterns. The detection of anomalies or classification process consists on identifying if the input data corresponds to a learned category or if a new one

must be created. Therefore, fault detection is given by new categories recognized by the algorithm depending on how strict the internal vigilance system is. As done in [12], TRNSYS is used to simulate the experimental setup data for long periods of time. The ART is able to recognize pump failure, pump degradation, shading and thermosiphoning (reverse flow). Nevertheless, expert knowledge of the system is still required to oversee, recognize and correlate the creation of new categories, to real faults and failures in the system.

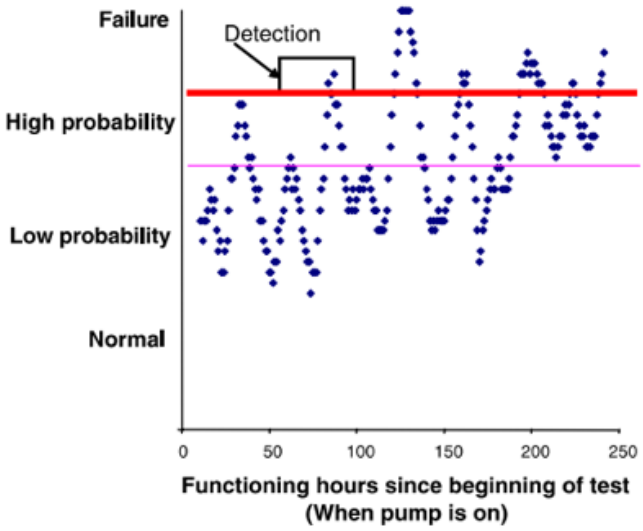


Figure 13: Averaged sequence against the functionating hour of the whole system monitoring and location of the detection using the M10a method (Cypriot system) [12].

As it had been mentioned before, most ANN fault detection algorithms are based on regression models where the fault-free conditions are predicted and compared to the measured data, unless information is available from the faulty conditions. Using simulation software, such as TRNSYS, has proven to be a viable alternative for data collection, however its application is limited and must be validated with experimental data. Another approach was proposed in [13], in which an online fault detection and diagnosis under uncertainty module is developed for a District Heating (DH) system using Bayesian Networks. Here, the pressure sensors of the DH network from the locality of Skultuna in Sweden are modeled in OpenModelica. This software is used to generate healthy data and thus, by comparing the residual values calculated from the model's result and the online monitoring system, three health states are identified. The residual values are then processed by a Bayesian Network to assess the probability of a real anomaly being detected by analyzing the cause-effect relationships from the sequence of data. If ten consecutive or more timesteps are identified as anomalies, the diagnosis module recognizes a leak in the system.

Decision tree (DT)-based data-driven methods have also been successfully applied to energy demand modeling, fault energy consumption pattern recognition and diagnosis. In [59] faults in a Variable Refrigerant Flow (VRF) systems are analyzed with a classification and regression

decision tree-based fault diagnosis module (CART) for supervised fault detection. Four health states are identified as follows: normal, fouling, refrigerant undercharge and refrigerant overcharge. The overall fault diagnosis accuracies obtained are 94.44% for an experimental dataset, and 80.55% for the on-line testing datasets.

Despite the wide range of techniques and models developed for anomaly and fault detection, there are challenges in system diagnostics which have not been overcome yet. Insensitivity to different operational conditions, false alarms and high uncertainty present in real-time processing have been identified as some of the more pressing issues [11]. In this context, time series modeling applications through more complex Machine Learning algorithms have been applied for anomaly detection in fault diagnosis. For instance, in [60] a deep LSTM-based framework was proposed for anomaly detection in superconductor magnets at the CERN laboratory. Time series modeling of the voltage registered in the superconductor magnets was assessed. Not only immediate voltage predictions were studied, but also the ability to forecast several timesteps ahead in time. Figure 14 shows how the precision of said predictions is reduced when evaluating more than one timesteps ahead in time.

There are also examples for unsupervised anomaly detection techniques, which reduces the cost of labeling great quantities of data for analysis and do not require previous fault data. Auto encoders (AE) are neural networks whose objective is to reconstruct and simulate the input data by learning a latent representation of features from it. This is similar to a Principal Component Analysis (PCA) used to reduce the dimensionality of the input data for later analysis. Thus, they have become an efficient tool de-noising signals and for anomaly detection. The use of RNN-based encoders to analyze timeseries addresses one of the greatest challenges for accurate future predictions which is the presence of noise inherent in sensor data. Additionally, their ability to transform a varying-length sequence as input data to a fixed output make it a more flexible model than regular sequence architectures. Figure 15 shows an example of original data a reconstructed time series from an RNN-based encoder-decoder, which can be seen that accurately imitates the behavior of the input signal.

By using the reconstruction errors between the next period data and from the output RNN-based denoising AE, anomalous conditions can be detected and further classified from these outputs. In [61] the fault diagnosis framework for rolling bearings with GRU-RNN-based AE was proposed. In Figure 16 the reconstruction errors are shown for four health states of rolling bearings, which prove useful to indirectly classify the health state through the AE. In [62] an anomaly detection model based on encoder-decoder framework is also applied for enhancing the reliability of smart grids. The framework of this work is shown in Figure 17, in which online measurements are compared to the output of the model. As before, an unexpected high reconstruction error indicates the probability of an anomalous behavior.

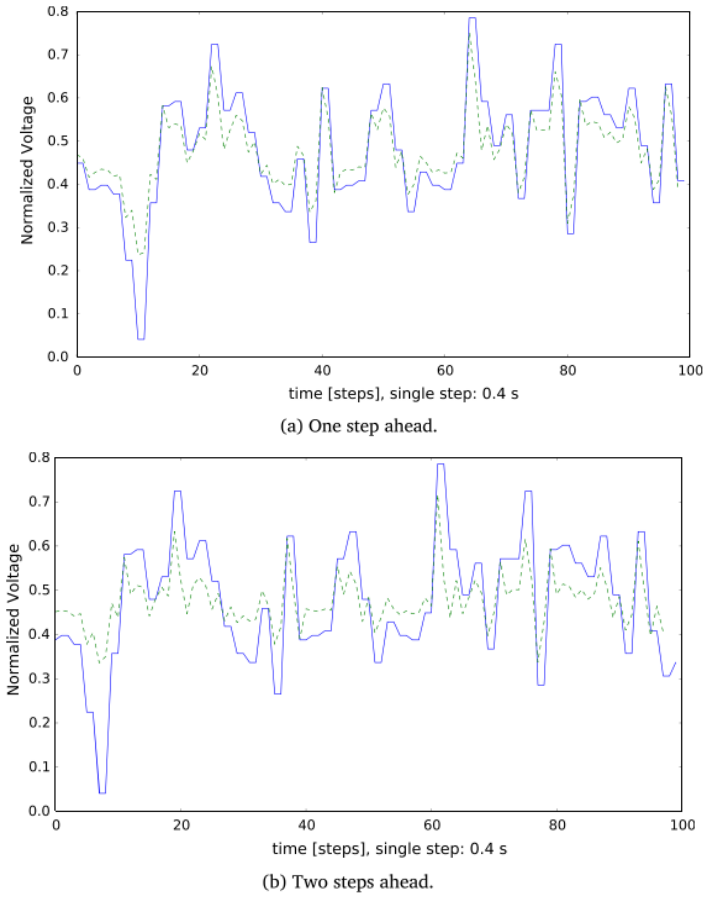


Figure 14: Two examples of prediction for one and two steps ahead in time. Predicted signal is plotted in a green broken line [60].

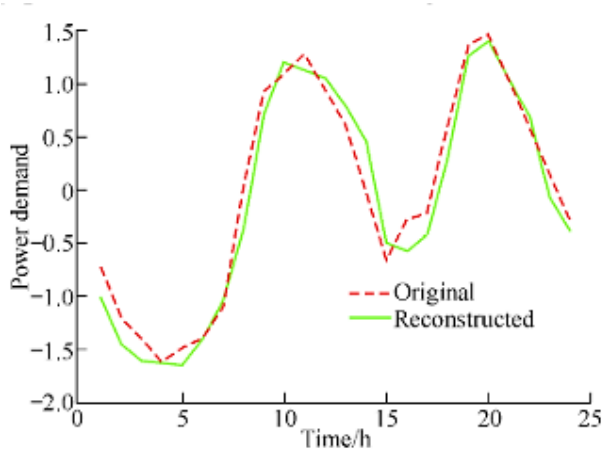


Figure 15: Original series and reconstructed series by an RNN encoder-decoder for smart grid data from [62].

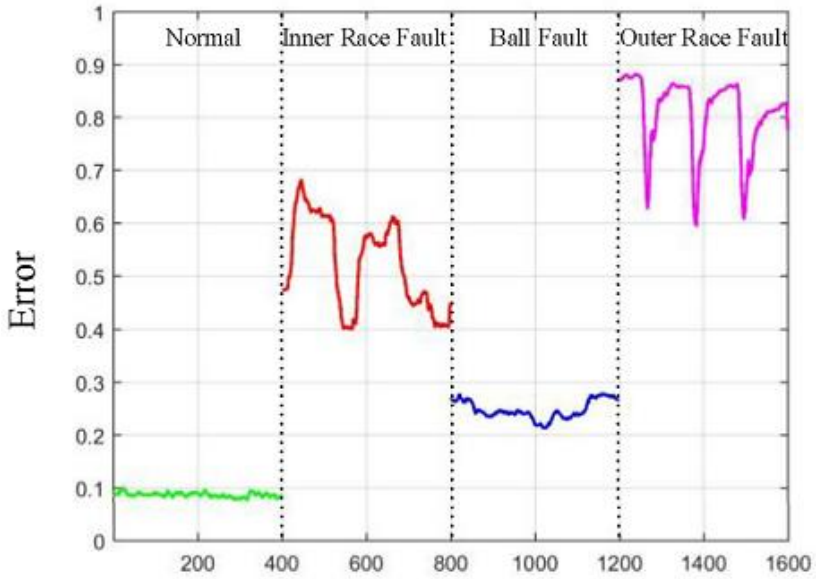


Figure 16: Reconstruction error generated by the GRU-NP-DAE proposed in [61].

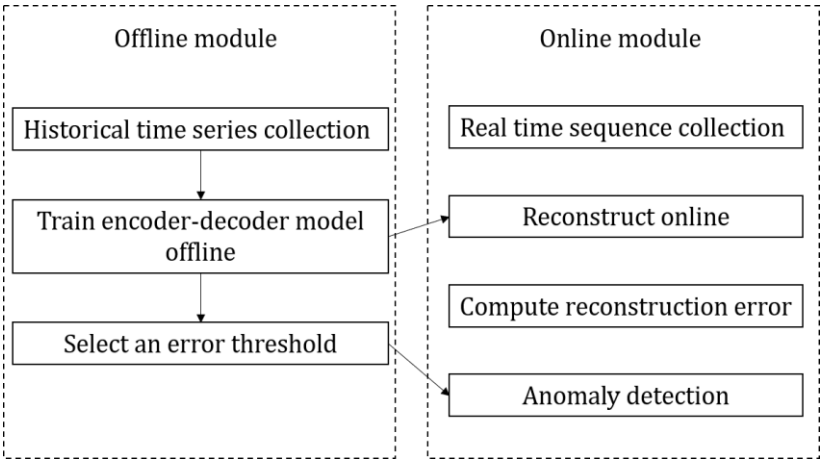


Figure 17: Framework of an RNN-based encoder-decoder. Adapted from in [62].

However, in this case, the detection phase is based on the selection of a threshold τ which directly depends on the input data and the distance d of the output prediction. This distance includes the effect of the difference of values in the same timestep as well as the trend of the data, while the threshold is selected by maximizing equation below.

$$\mathrm{F} = \frac{2PR}{P+R} \tag{16}$$

Where $P = \frac{N_{TP}}{N_{TP}+N_{FP}}$ represents the precision and $R = \frac{N_{TP}}{N_{TP}+N_{FN}}$ the number of recalls based on Table 1.

Table 1: Definitions of precision and recall. Adapted from [62].

Category	Anomaly samples	Normal samples
Predicted to be anomaly	N_{TP}: Number of anomaly samples which are predicted to be anomaly correctly	N_{FP}: Number of normal samples which are predicted to be anomaly wrongly
Predicted to be normal	N_{FN}: Number of anomaly samples which are predicted to be anomaly wrongly	N_{TN}: Number of normal samples which are predicted to be normal correctly

Finally, the function used to decide whether an anomalous behavior has been detected is shown in Equation (17). Results obtained from the analysis of this work demonstrate that the encoder-decoder framework was able to successfully detect over 95% of the anomalies.

$$f = \begin{cases} 0, d \geq \tau \\ 1, d < \tau \end{cases} \quad (17)$$

3.2.2 Uncertainty Estimation

When predicting future values of a timeseries with ML techniques, the output of any regression model is subject to various sources of uncertainty. This may arise from the recollected sensory data and the inherent bias of the trained model. Additionally, as ML algorithms have stochastic behavior, whether from parameter initialization or from the applied dropout, results carry uncertainty that must be quantified when studying systems in which security is an important issue.

When using residual-based fault detection frameworks, the quantification of the prediction's uncertainty is of great importance to reduce the number of false classifications obtained with the use of fixed-thresholds. Natural-occurring disturbances, as well as the model's uncertainty, require wider threshold values but suffer from a decrease of sensitivity towards faults. In Figure 18, a residual-based fault detection model with fixed thresholds is shown. The value of the threshold is selected based on a compromise between false alarms and correct detection, however this can prove difficult when the data is subjected to noise. Thus, Figure 19 shows time-variant adaptive threshold values depending on the input data used to counter noise and the uncertainty of the model.

In [63] an MLP model parameters and resulting parameter uncertainty is estimated through the application of the Outer Bounding Ellipsoid (OBE) algorithm. By obtaining the mathematical description of the model's uncertainty, the parameters can be estimated from a previously defined feasible parameter set. This was implemented in a robust fault-detection model with the aim of providing support in diagnostic decisions. However, this method, which increases complexity as the number of parameters increase, can only give an estimate solution which is not guaranteed to be unique. Also, there was a low sensitivity to smaller faults.

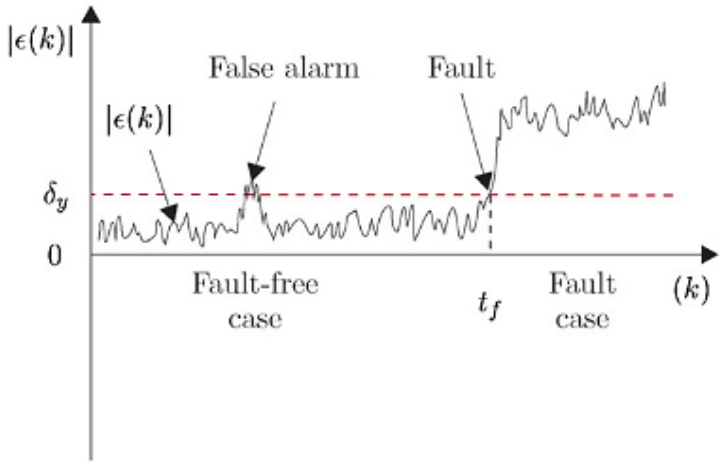


Figure 18: Residual-based fault detection [63].

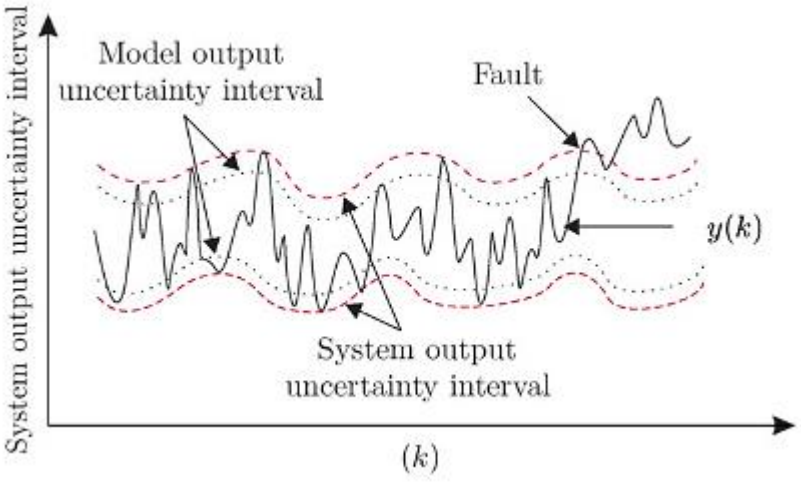


Figure 19: Application of system output uncertainty interval to robust fault detection [63].

Unlike Bayesian models, which offer a mathematical framework to analyze uncertainty at a high computational cost, it has been proposed that the use of a Bernoulli dropout applied between each weighted layer of a non-linear network can estimate the uncertainty of the model. Indeed, Yarin Gal in [64] proved that the use of the dropout is mathematically equivalent to a Monte Carlo approximation of a Deep Gaussian Process (GP). Thus, it is possible to obtain the network's uncertainty from the trained model without any additional assumptions of the architecture. This allows an analysis of uncertainty as a Gaussian distribution without sacrificing complexity or accuracy of the model. This is applied by averaging a high number of forward passes of the data through a trained network, applying a 10% of dropout probability to map the stochastic nature of the model.

3.3 Summary

A number of investigations have been revised thus far to contextualize the motivation of the present work. From the titles shown in Table 2 it has been discussed that the applications of Neural Networks to solar thermal systems mainly seek to obtain performance predictions models under different conditions. However, with the exception of [53] in which LSTM-based models are used for solar irradiance prediction, these works do not use specialized architectures for temporal or spatial analysis, such as RNN or CNN, or rely heavily on heuristic logic. This leaves an untapped potential to develop more flexible data-driven fault detection systems applied to thermal systems than the ones stated in Table 3. For instance, while [8] and [12] both reached high accuracy for fault detection in SHW systems, both rely on selected thresholds based on expert knowledge to identify anomalous behavior in different components.

In Table 4, other supervised fault detection methods are presented. Successful applications for District Heating [13] and Variable Refrigerant Flow [59] systems explore the probabilistic nature of fault detection. In [60] a LSTM-based model is used to predict future values of sensor measurements to determine the health state of LHC superconducting magnets. The sequence errors between the predicted sequence and the measured values are used to determine anomalous behavior for each timestep in the future. However, this idea can be developed with unsupervised or semi-supervised methods, such as Auto-Encoders. A fault detection framework specialized for temporal sequence analysis for rolling bearings presented in [61], in which the input sequence is reconstructed from the latent space trained with. fault-free data. Reconstruction errors are then used to classify the health state of the component. Finally, in [62] the same principle is explored applied to smart grid management. Nevertheless, in this work the challenge of the prediction's uncertainty is addressed and a method to obtain an optimized threshold for fault detection is presented. This approach deals with the natural fluctuations of input values, which under a fixed threshold, increase the number of false positive or false negative classifications.

Table 2: Summary of Performance Prediction Models in Literary Review.

Performance Prediction Models	Ref. Num.	Technique used	Authors	Year	Title
Review of performance predictions models of solar collector systems using ANN.	40	ANN	H. K. Ghritlahre and R. K. Prasad,	2018	"Application of ANN technique to predict the performance of solar collector systems - A review,"
Optimization of solar collector design parameters through performance prediction	49	ANN	H. Li, Z. Liu, K. Liu, and Z. Zhang,	2017	"Predictive Power of Machine Learning for Optimizing Solar Water Heater Performance: The Potential Application of High-Throughput Screening,"
Day-ahead forecast of solar irradiance for PV system management	50	LSTM	S. Srivastava and S. Lessmann,	2018	"A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data,"
Day-ahead PV output power production	51	ANN	S. Leva, A. Dolara, F. Grimaccia, M. Mussetta, and E. Ogliari,	2017	"Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power,"

* Source: Own elaboration.

Table 3: Summary of FDS for Thermal Systems in Literary Review.

Fault Diagnostics Systems	Ref. Num.	Technique used	Authors	Year	Title
Review of model-based, long-term fault detection solar thermal systems	23	Model-based	A. C. de Keizer, K. Vajen, and U. Jordan,	2011	"Review of long-term fault detection approaches in solar thermal systems,"
Model-based economic and exergy analysis for fault detection	53	Thermo-economic	A. Picallo-Perez, J. M. Sala-Lizarraga, E. Iribar-Solabarrieta, M. Odriozola-Maritorena, and L. Portillo-Valdés,	2017	"Application of the malfunction thermoeconomic diagnosis to a dynamic heating and DHW facility for fault detection,"
Detection and localization of developing faults	54	Exergy analysis	P. Fu, N. Wang, L. Wang, T. Morosuk, Y. Yang, and G. Tsatsaronis,	2016	"Performance degradation diagnosis of thermal power plants: A method based on advanced exergy analysis,"
SHW diagnosis system for faults in solar collectors and insulation pipes	56	ANN	S. Kalogirou, S. Lalot, G. Florides, and B. Desmet,	2008	"Development of a neural network-based fault diagnostic system for solar thermal applications,"
Monitor SHW systems for pump failure, collector shading and thermosyphoning detection	8	ART	H. He, T. P. Caudell, D. F. Menicucci, and A. A. Mammoli,	2012	"Application of Adaptive Resonance Theory neural networks to monitor solar hot water systems and detect existing or developing faults,"

* Source: Own elaboration.

Table 4: Summary of DL-based Fault Detection Applications in Literary Review.

Deep Learning Applications	Ref. Num.	Technique used	Authors	Year	Title
Review of ANN-based fault detection systems	55	ANN	J. Korbicz, J. M. Koscienlny, and Z. Kowalczuk,	2004	Fault Diagnosis Models, Artificial Intelligence, Application Vo. 1
Online fault detection under uncertainty for District Heating	57	Bayesian Networks	N. Zimmerman, E. Dahlquist, and K. Kyprianidis	2017	, "Towards On-line Fault Detection and Diagnostics in District Heating Systems,"
Refrigerant level related fault detection in Variable Refrigerant Flow systems	58	CART	G. Li et al.,	2018	"An improved decision tree-based fault diagnosis method for practical variable refrigerant flow system using virtual sensor-based fault indicators,"
Time-series modeling for CERN superconductor magnets fault detection	59	LSTM	M. Wielgosz, A. Skoczeń, and M. Mertik,	2017	"Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets,"
Unsupervised Auto-Encoder based fault detection for rolling bearings	60	GRU-RNN-AE	H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang,	2018	"Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders,"
Unsupervised anomaly detection in Smart Grids	61	AE	Z. Fengming, L. Shufang, G. Zhimin, W. Bo, T. Shiming, and P. Mingming,	2017	"Anomaly detection in smart grid based on encoder-decoder framework with recurrent neural network,"

* Source: Own elaboration.

Chapter 4

Methodology

To achieve the main objective of this work, a diagnosis framework must be developed to identify anomalies during operation of a solar thermal system. Thus, the following steps shown schematically in Figure 20 are implemented.

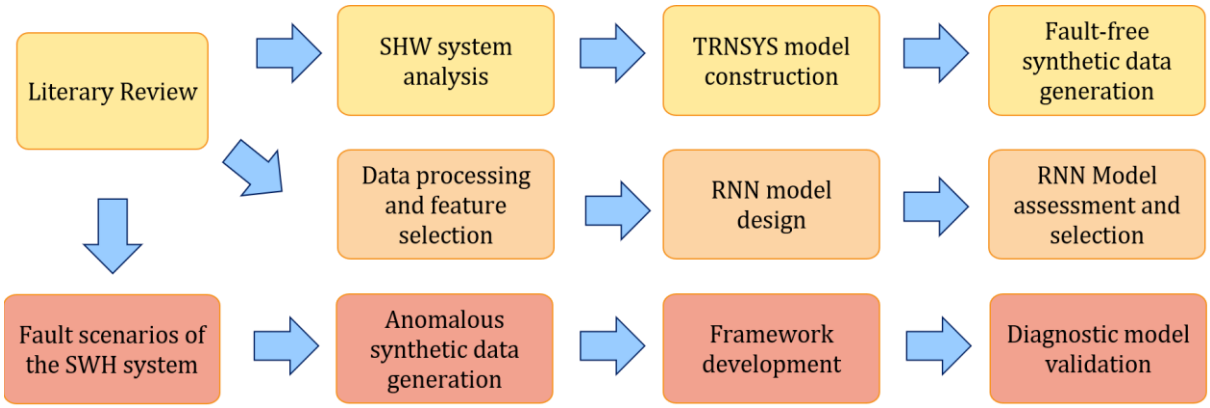


Figure 20: Schematic representation of the proposed methodology.

4.1 Literature Review

For this work, three main topics were analyzed: Solar Hot Water systems (SHW), Fault Diagnosis Systems (FDS) and data-driven approaches based on Deep Learning (DL). The purpose of this literature review is to establish connections between these topics. In particular, to explore the application of DL, which is widely used in Reliability analysis and time-series predictions in other mechanical systems, and asses its applicability for anomaly detection and diagnosis in thermal systems.

4.2 SHW TRNSYS Model Construction

To correctly carry out an analysis of the SHW system's behavior and the health states of its components, a large time window of data is required to analyze the seasonality of the monitored variables. In this case, environmental conditions affect the heat output of the solar field and thus, of the rest of the system. For this reason, a method was required to gather data in a short period of time.

Considering that the monitoring system in the building only begun recording data since the start of 2018, the only way to obtain the SHW behavior was through a simulation of it. Based on the technical information of the components installed in the building and the real-time measurements of environmental conditions, thermal outputs in critical components in the system are generated and subsequently analyzed.

4.2.1 SHW System Data and Available Information

Data recollection from different sources is a critical step for the developed study. To correctly simulate the system in TRSNSYS the following data is considered:

- **Meteorological data:** Solar radiation, ambient temperature, wind speed and wind direction are required to properly characterize the environmental conditions in which the SHW system operates. A complete meteorological station is placed next to the solar collectors' array at the Beauchef 851 building, from which data is extracted and processed. Radiation and other meteorological values are then synchronized to later be fed into the TRNSYS software.

- **SHW system information:** Design conditions and the main features of the system are defined, according to the information provided by the manufacturer. Also, the information about secondary equipment such as heat pumps, heat exchangers, heat accumulators, centrifugal pumps and the control system was also considered. Nevertheless, appropriate simplifications were introduced for the purpose of reducing the computational expenses in TRNSYS software.

- **Operational data:** Temperature and differential pressure sensors are installed in various locations in the system. This information is used to compare and validate the physical model of the system developed in TRNSYS.

4.2.2 TRNSYS Simulation

A simplified version of the SWH system installed in the western tower at the Beauchef 851 building is built in the simulation software. TRNSYS operates with types or component blocks which are integrated to experimentally validated equations and calculation modes. A set of parameters defined by the user are applied to personalize the simulation, including thermal performance datasheets for certain components. Here, relevant technical information of the system, such as heating and cooling capacities of chiller and heat pump equipment, centrifugal pump power and rated flow, as well as area and optical characteristics of the solar field, is used to select the types and define their parameters. Control systems, demand profiles and operation schedules are also used to represent the SHW system in the simulation software. Details of these values are found in Chapter 5.

As the system's current situation does not allow for a full experimental validation of the TRNSYS model, the generated temperatures may vary from the sample of measured values. Malfunctions have also been detected in the actual installation. For this reason, as series of simplifications are applied to the model's design, prioritizing the internal logic of the model rather than reaching the expected values. Tuning of parameters, connections between components and the control system are considered. This also accounts for the integration issues identified during the study of the system and the interactions with other subsystems of the buildings climatization system.

4.2.3 Nominal Synthetic Data Generation

TRNSYS software allows the extraction of the simulated data in user-defined format documents. For each subsystem defined in the simulated model, a set of variables is monitored and extracted for analysis with a sampling frequency of a minute. These values consist on temperatures and control signals, leaving aside flow and energy consumption values due to the lack of experimental validation options for these variables. Thus, these values are just an approximation of the design conditions of the system. The length of the simulation window depends on the availability of radiation and meteorological data which is fed into the TRNSYS model. The software then generates a minute-based simulation of the behavior of the monitored variables during the selected time window. The TRNSYS simulation is the primary source of input data and a critical step in the development of this study.

4.3 Deep Learning Framework

The purpose of the framework is to predict future values of target variables, based on the generated data from the TRNSYS simulation. The training of the model is done only with data under nominal conditions of the thermal system. For this, the databases generated are preprocessed and features are selected for the prediction of each target variable. The performance of ANN, RNN and LSTM architectures are compared under similar conditions and selected hyperparameters to identify strengths and weaknesses of each model. The objective is to select a model which can accurately predict values one timestep in the future, as well as reproducing the behavior during short sequences of several timesteps in the future.

4.3.1 Data Preprocessing

The output data from the TRNSYS model, such as radiation, temperature and operational conditions are preprocessed before selecting relevant variables for analysis. First, variables are selected and organized in sets of characteristics which are used to predict target temperatures in the system. Each of these datasets is averaged over a time period of 6 minutes to reduce the effect of the instantaneous variability of solar radiation. Due to the nature of the diffuse component, minute-based measurements may suffer from larger uncertainties caused by the passing of clouds. As

thermal systems have a natural inertia, this average will not significantly impact the values of measured temperatures.

4.3.2 Framework Development

From the simulated data, a selection of features is established for the construction of the supervised regression framework. This model is later used for the anomaly detection in the system. Decisions about the architecture and DL techniques that are applied must be based on the nature of the input data and on the literary review done concerning the fault detection criteria to be selected. Due to their characteristics, RNN models are particularly well-adapted for time-series analysis. Thus, variations of this architecture are analyzed for temperature prediction under different scenarios of available data. Variations in the architecture rise from the following decisions:

1. **Predicted values:** The objective of the DL framework is to accurately predict selected variables of the system. A priori, three temperatures are chosen: solar field outlet, preheating section heat accumulators' output, and hot water temperature supplied to the demand. The latter is excluded due to the similarities to the heat accumulators' temperature behavior.

2. **Prediction horizon:** The number of timesteps in the future the model is required to predict. This horizon can be different during the training and the test session. However, the accuracy of the predicted value may decrease if the test is significantly longer than the training data presented.

3. **Length of the datasets:** The number of timesteps prior to the prediction fed into the network will have different effects depending on the model used. This may vary between a day or a week.

4. **Content of the datasets:** The initial database of temperatures, weather data and operational conditions, are separated into training, validation, and test sets. For each of the desired predicted temperatures, features will be selected to train the models.

For each target temperature, hyperparameter tuning is carried out to fit each of the corresponding datasets. Shallow and deep configurations are considered. Dropout is used to prevent overfitting the model with the training data. The models which are assessed for the supervised regression task are the following: RNN, LSTM and ANN. The latter is used to compare the proposed models with previous investigations.

4.3.3 Assessment of Proposed Models

The assessment of the models is divided into two phases. First, the performance of each model is evaluated through the Root Mean Square Error (RMSE) metric for single future temperature values. Additionally, the percentage of the relative error is analyzed. Each dataset is divided into training, validation and testing sets. The training process is carried out only on the training set, and the RMSE is evaluated over both the training and the validation subsets in each iteration epoch. Finally, the test RMSE is calculated over the test subset. Additionally, other metrics are used to evaluate the models' predictions, such as variance, explained variance score and R^2. Secondly, samples of daily temperature sequence samples are extracted. These samples are used to evaluate the model's capacity to predict further timesteps in the future as a second testing process.

4.4 Model Implementation

The last section of the methodology is dedicated to the anomaly detection process. This is carried out by using the previously trained models and new input data generated with the TRNSYS software. Faults are induced in the simulation and the new data is fed into the selected DL model, which was trained with fault-free values. The prediction errors obtained from the fault-induced data are then used to identify the anomalous behavior. For each fault-induced scenario, three cases under different meteorological conditions are selected. The prediction errors are then compared when the trained model is fed with healthy and with anomalous data during the same period. Finally a detection threshold is selected to separate healthy from anomalous data.

4.4.1 Anomalous Synthetic Data Generation

From the analysis of the SHW system and the resulting simulation model in TRNSYS, a selection of anomalous scenarios with a distinguishable impact on the thermal performance of the SHW components is made. Frequent issues of SHW systems are evaluated in the simulation and the anomalous synthetic data is generated and extracted as daily samples for each set of characteristics defined in the previous steps.

4.4.2 Anomaly Detection Framework

As temperature is the main studied variable, unexpected variations of these values are monitored throughout the system. These may be a result of a change in meteorological conditions, faulty sensor equipment, or from unidentified faults in the system. The lack of historical data does not allow an in-depth study of which faults and failures develop in the system, and thus, the focus will be on identifying abnormal conditions for a further performance and energy consumption analysis.

The proposed diagnosis model is based on quantifying and interpreting the relative errors between input data from the system and the predicted values of the prognosis model. The prediction errors are used to select a threshold which will then classify the healthy from the anomalous data sequences. To increase the flexibility of the anomaly detection method, the value of the threshold is not fixed beforehand. Instead, the value of the threshold is calculated based on the performance of the trained model when tested with each specific sample, incorporating the uncertainty of the prediction into the classification process. From this initial analysis, it is possible to construct performance indicators based on different consumption profiles, solar collector efficiency and meteorological conditions.

4.5 Resources

For the development of this work, the following resources have been identified.

- Technical information of the air conditioning system of the Beauchef 851 building. This is available as the company in charge of the engineering design of the system, CINTEC S.A. has shared relevant technical documents.

- Software: TRNSYS 18 and Autodesk Inventor DWG True View. All these are available for use through academic or free licenses. For programming, Python 3.6 language with Tensorflow 1.4 was used in an Ubuntu 64 bits 16.04 LTS operating system.

- Measured data from the meteorological station installed at Beauchef 851. Radiation information is directly available for use obtained from the following instrumentation: Pyranometer Kipp&Zonnen CMP21, Pyrgeometer Kipp&Zonnen CGR4, Pyrheliometer Kipp&Zonnen CHP1, Sun Tracker Kipp&Zonnen SOLYS2 and Datalogger Scientific Campbell CR1000. Other meteorological variables are obtained from a R.M. Young Meteorological Instrumentation Wind Monitor Model 05103-5, Campbell Scientific HMP60 Temperature and Relative Humidity Probe and Datalogger Scientific Campebell CR6 Measurement and Control Center.

- Access to the Smart Reliability and Maintenance Integration Laboratory (SMRILab) and the use of the computers there installed. An Intel®Core™ i7-6700K CPU @ 4.00 GHz x8 with a TITAN Xp/PCIe/SSE2 Graphics card is used.

- Measured data from the SHW. Temperatures and operational status of different equipment of the fully-functioning system must be available for comparison purposes.

Chapter 5

System Description

The solar-driven climatization subsystem at the western building of Beauchef 851 is analyzed from the technical documents and blue-prints delivered by Cintec S.A[1], company in charge of the engineering design. Information from the system such as design temperatures, equipment characteristics and the control system design is extracted to build the TRNSYS model of the subsystem. A schematic representation of the principal components of the subsystem is illustrated in Figure 21.

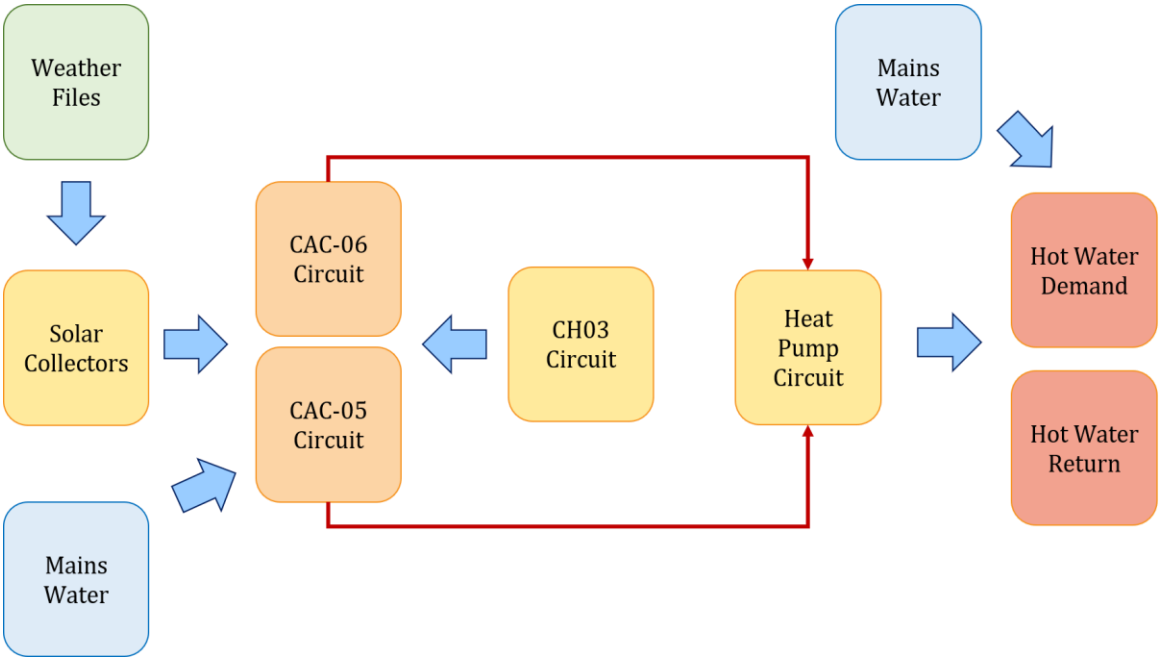


Figure 21: Schematic representation of the solar thermal system at Beauchef 851. Source: Own elaboration based on documents delivered by Cintec S.A.

From the solar collectors and the condenser output of the CH-03 chiller, heat is driven to the mid-temperature tanks CAC-05 and CAC-06. A second heating section is provided by heat pumps to reach the desired temperature for the demand. Mains water is used to regulate the output temperature and as make-up water in the preheating section. These components, along with heat exchangers and centrifugal pumps are monitored for temperature and operation status. In Figure 22, a process flow diagram of the system is shown. Red lines represent the hot water flowing in the system, while the blue lines correspond to the cool water. The output of the mid-temperature tanks

[1] Engineering firm dedicated to HVAC&R systems. Reference: http://www.cintec.cl/index.html.

is then heated again by a series of heat pumps and finally delivered to the hot water system of the building, as shown in Figure 23.

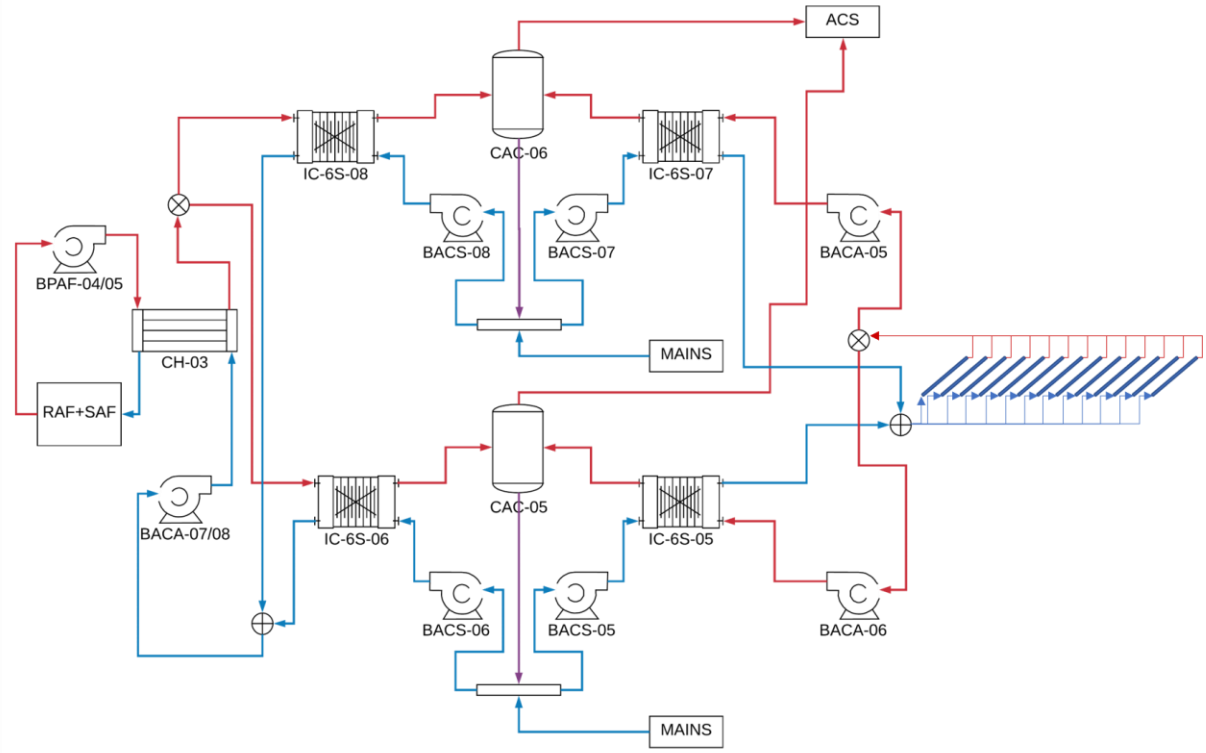


Figure 22: Process Flow Diagram of the Pre-heating section in SWH System. Source: Own elaboration based on documents delivered by Cintec S.A.

5.1 Nominal Design Conditions

The western building at Beauchef 851 (Faculty of Physical and Mathematical Sciences, FCFM, for its acronym in Spanish) is located at:

- Latitude 33° South
- Longitude 70° West
- Altitude above sea level 550 m.

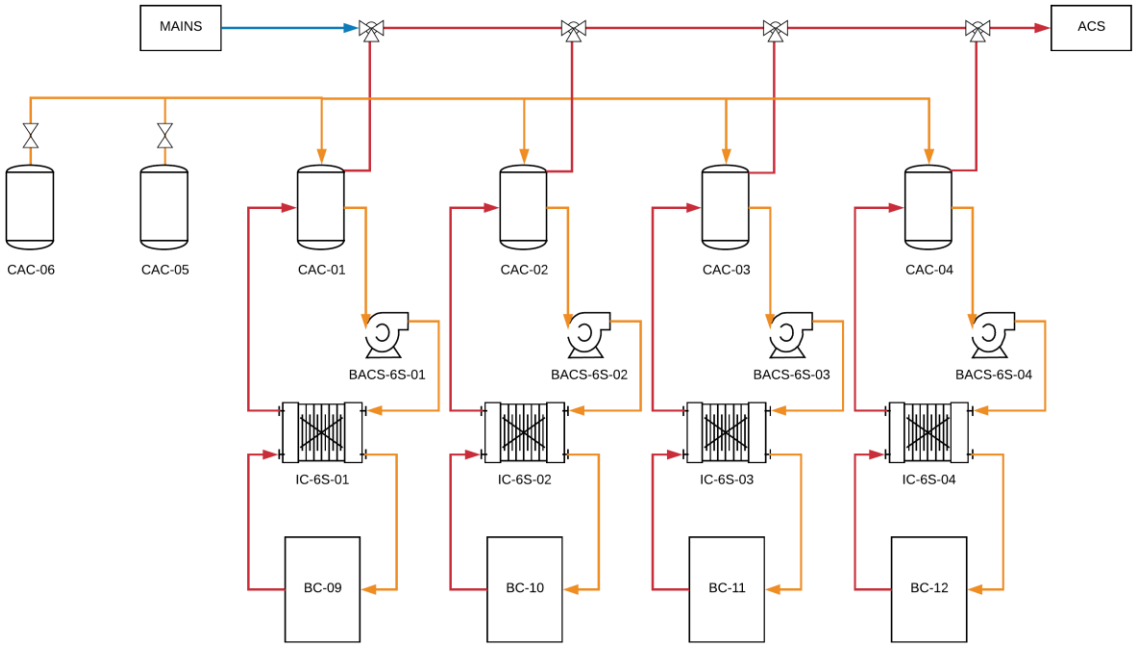


Figure 23: Process Flow Diagram of the Heating section in SWH System. Source: Own elaboration based on documents delivered by Cintec S.A.

There are two design conditions established for the climatization system of the building, in summer and winter with the following characteristics:

- Summer
  - Dry Bulb Temperature: 33°C
  - Relative Humidity: 40%

- Winter
  - Dry Bulb Temperature: 2°C
  - Relative Humidity 90%

Therefore, design conditions for offices, classrooms, sport infrastructure and other interior spaces are defined by temperatures, humidity, air renovation, heat rejection from equipment, lightning and use profiles of each space. As mentioned before, the SHW subsystem supplies sanitary hot water for the showers in the dressing rooms in the -3 floor. The nominal hot water demand is estimated at 24,000 L at 40°C for 12 hours daily. For such purpose, the following temperatures have been determined:

- Hot water tank temperature: 60°C

- Mains water temperature: 10°C
- Preheated water temperature: 35°C

The SHW is designed to preheat the sanitary water from 10°C to 40°C during the summer and 30-35°C during the winter. When the available radiation is not enough to reach these temperatures, auxiliary heat is provided by the heat-recovery chiller (CH-03). This preheated water is to be stored in two hot water tanks (CAC-05/06); then it is heated up to 60°C by heat pumps (BC-09/10/11/12) and subsequently stored in other four hot water tanks (CAC-01/02/03/04). Finally, it is dispatched by demand from these tanks, and the unused hot water reentries the system at the two first water tanks. A brief description of each component of the solar-driven subsystem is presented below.

5.1.1 Solar Collector Field

The solar field is composed by evacuated tube heat pipe collectors, amounting to a total absorption area of 105.6 m^2 and whose main characteristics are listed in Table 5. Given the available space on the roof of the building, the spatial distribution of the 44 solar collectors is shown in Figure 24.

Table 5: Design solar collector main characteristics.

Tube and Heat Pipe	
Number of tubes	30
Tube diameter	80 mm
Tube thickness	1.6 mm
Tube Material	Borosilicate
Heat Pipe Material	Copper
Dimensions	
Total Area	4.94 m^2
Absorption Area	2.40 m^2
Fluid capacity	1.80 L
Length	1,990 mm
Width	2,487 mm
Thickness	180 mm
Inclination angle	15°
Optical properties	
Absorptance	92%
Emittance	10%
Operational conditions	
Pressure	8 bar
Stagnation temperature	250°C
Refrigerant	Propylene glycol, 20%

*Source: Own elaboration based on documents delivered by Cintec S.A.

From the common inlet of the solar field, the flow is divided into thirteen collector strings connected in series, eight rows of three collectors and five rows of four collectors each. Each row is supported on metallic bases which maintain an inclination angle of 15° and a north orientation. The datasheet describing the selected model Hitek Solar NSC 58-30 is shown in Appendix B, tested under the EN 12975 regulations. In Table 6 the Incidence Angle Modifiers (IAM) are presented and the efficiency parameters related to aperture area and thermal properties are listed in Table 7.

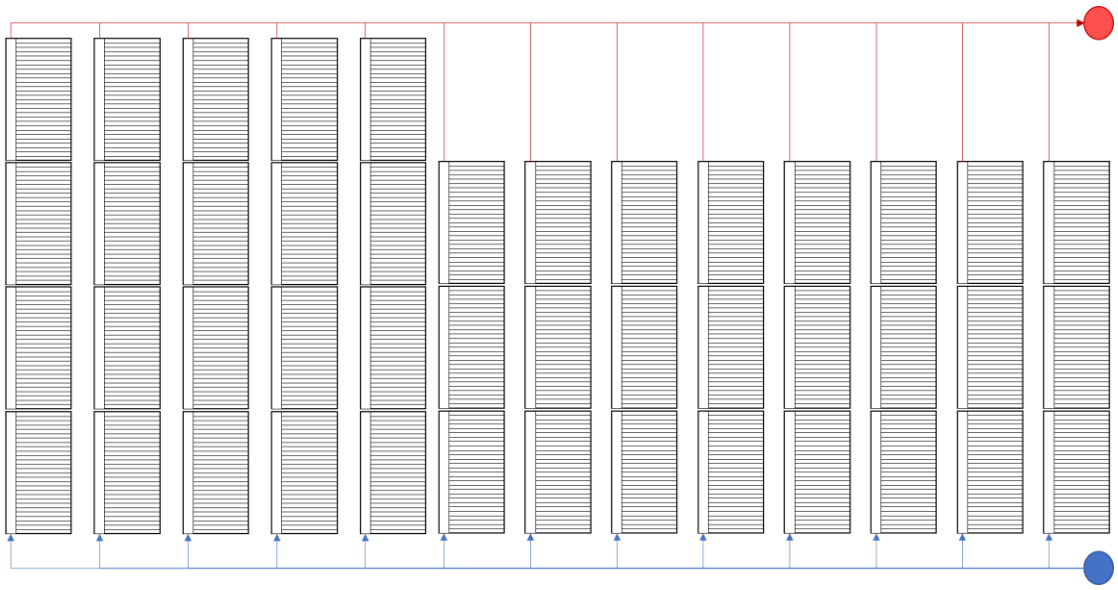


Figure 24: Spatial distribution of the solar field at Beauchef 851. Source: Own elaboration based on documents delivered by Cintec S.A.

Table 6: Incidence Angle Modifiers for selected solar collector.

Incidence Angle Modifiers	10°	20°	30°	40°	50°	60°	70°
$K_\theta(\theta_T)$	1.010	1.019	1.056	1.151	1.452	1.462	1.261
$K_\theta(\theta_L)$	0.999	0.994	1.018	0.974	0.952	0.913	0.833

*Source: Summary of EN12975 Test Results for Hitek Solar NSC-58-30 in Appendix B.

Table 7: Thermal Properties of selected solar collector.

Collector Efficiency Parameters		
η_0	0.618	-
a_{1a}	1.377	$W/(m^2K)$
a_{2a}	0.018	$W/(m^2K^2)$
Effective Thermal Capacity	5.684	$kJ/(m^2K)$

*Source: Summary of EN12975 Test Results for Hitek Solar NSC-58-30 in Appendix B.

5.1.2 Heat Recovery Chiller

In the climatization system, the cooling and heating needed for air conditioning are supplied by the heat pumps, cooling towers and chiller. Additionally, a third chiller with heat recovery must supply the water and space heating required for the indoor swimming pool, and the auxiliary preheating of the sanitary hot water, coupled with the SHW system. Vapor-compression chillers are machines which are used to remove heat from a liquid through a refrigeration cycle. A schematic diagram of the cycle is shown in Figure 25, where the cooled and chilled water stream interact with the refrigerant within through the condenser and evaporator heat exchangers. In the evaporator, the liquid refrigerant absorbs heat from the entering chilled water stream and then is passed through a compressor, further elevating its temperature until reaching an overheated steam. In the condenser, heat is ejected to the cooling water stream, and liquid refrigerant is then passed through an expansion valve, lowering its temperature to reenter the evaporator.

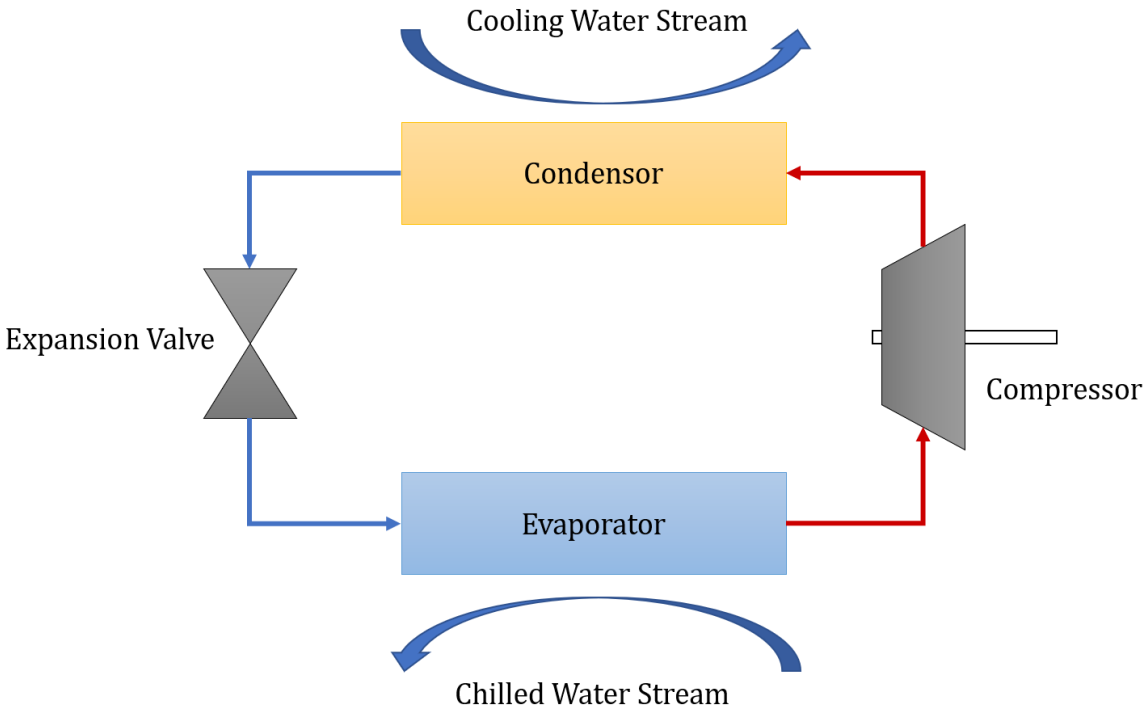


Figure 25: Schematic Diagram of Single Stage Water Cooled Chiller.

Table 8 presents the design conditions for the installed equipment, a Thermocold CWC Prozone 1320 Z C model, along with the model's performance matrix in Table 9[2]. Additionally, the operation range for this model is shown in Figure 26, where the aforementioned temperatures are within this range.

[2] See Appendix 8 and Appendix 9.

Table 8: Design conditions for the heat-recovery chiller CH-03.

Evaporator side	
Water flow $[m^3/h]$	32.5
Inlet temperature [°C]	12.7
Outlet temperature [°C]	6
Condenser side	
Water flow $[m^3/h]$	54.7
Inlet temperature [°C]	44.5
Outlet temperature [°C]	50
EER	3.0
Capacity [kW]	254
Refrigerant	R-134a

* Source: Own elaboration based on documents delivered by Cintec S.A.

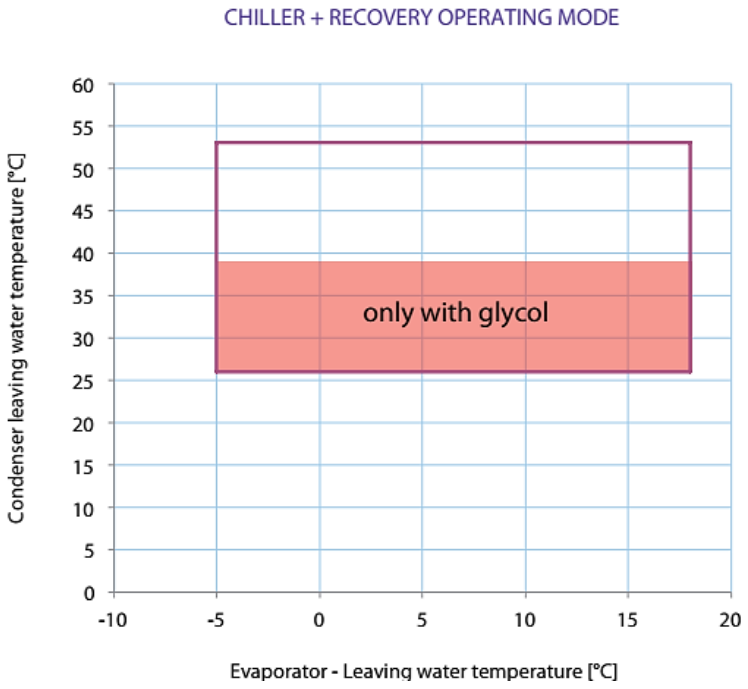


Figure 26: Water Cooled Refrigeration Cycle. Source: Thermocold CWC Prozone Data Sheets, Appendix C.

Table 9: Cooling Capacity and COP. Adapted from Thermocold CWC Prozone Data Sheet, Appendix C.

Evaporator Outlet T°C	Property	Condenser Inlet T°C						
		30	35	40	45	50	55	
5	Capacity	309.30	293.50	276.70	258.90	240.80	221.30	kW
	COP	5.42	4.68	4.00	3.39	2.84	2.35	
6	Capacity	319.00	302.80	285.50	267.10	248.40	228.40	kW
	COP	5.57	4.82	4.13	3.49	2.93	2.42	
7	Capacity	328.90	312.20	294.40	275.50	256.30	235.70	kW
	COP	5.72	4.96	4.25	3.60	3.02	2.50	
8	Capacity	339.10	321.90	303.50	284.10	264.40	243.20	kW
	COP	5.89	5.09	4.37	3.70	3.11	2.57	
9	Capacity	349.40	331.70	312.90	292.90	272.60	250.90	kW
	COP	6.04	5.24	4.50	3.81	3.20	2.65	
10	Capacity	360.00	341.90	322.50	302.00	281.10	258.70	kW
	COP	6.21	5.38	4.62	3.92	3.30	2.73	

In Appendix 4 a diagram of the chiller circuit is presented. The cooling water flow which exits the CH-03 chiller is divided to two heat loads: the heating section of the SHW system and to other sanitary hot water uses in the building. The returning flows are joined through a mixing valve before reentering the chiller.

5.1.3 Heat Exchangers

Throughout the system there are several flat plate type heat exchangers. An example of this configuration is shown in Figure 27, where the hot and cold flow of the system are represented in red and blue, respectively. Different designs of plate heat exchangers exist, combining materials and corrugated surfaces to increase the effectiveness of the heat exchange between fluids. These were selected to facilitate maintenance activities. The properties stated in Table 10 belong to the heat exchangers involved in the solar-driven subsystem:

- IC-6S-05/07: Heat exchanger between the solar collector field (primary circuit) and the heat accumulator CAC-05/06 (secondary circuit).

- IC-6S-06/08: Heat exchanger between the heat-recovery chiller (primary circuit) and the heat accumulator CAC-05/06 (secondary circuit).

In Table 11 properties for the heat exchangers involved in the heat pump-driven subsystem are presented.

- IC-6S-01/02/03/04: Heat exchanger between the heat pumps BC-09/10/11/12 (primary circuit) and the heat accumulator CAC-01/02/03/04 (secondary circuit).

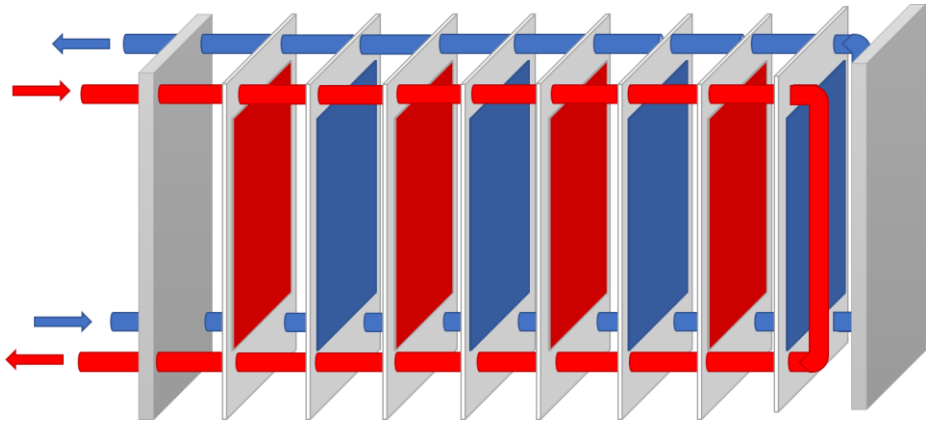


Figure 27: Schematic Representation of a Plate Heat Exchanger.

Table 10: Design conditions for Solar Heat Exchangers.

	IC-6S-05/07	IC-6S-06/08
Primary Circuit		
Inlet Temperature [°C]	45.0	45.0
Outlet Temperature [°C]	37.0	39.5
Secondary Circuit		
Inlet Temperature [°C]	27.0	29.5
Outlet Temperature [°C]	35.0	35.0
Heat Exchanger		
Area [m^2]	0.78	1.76
Water Flow [m^3/h]	5.3	13.6
Pressure Loss [bar]	0.31	0.46
Number of plates	21	45
Plate Thickness [mm]	0.7	0.7
Plate Material	Stainless Steel	Stainless Steel
Pressure [bar]	8/10	8/10

* Source: Own elaboration based on documents delivered by Cintec S.A.

Table 11: Design conditions for Heat Pump Heat Exchangers.

IC-6S-01/02/03/04	
Primary Circuit	
Inlet Temperature [°C]	65
Outlet Temperature [°C]	61
Secondary Circuit	
Inlet Temperature [°C]	54
Outlet Temperature [°C]	58
Area [m^2]	1.93
Water Flow [m^3/h]	5.3
Pressure Loss [bar]	0.4
Number of plates	49
Plate Thickness [mm]	0.7
Plate Material	Stainless Steel
Pressure [bar]	8/10

* Source: Own elaboration based on documents delivered by Cintec S.A.

5.1.4 Hot Water Tanks

The hot water accumulators or storage tanks are the principal components which enable the system to dispatch hot water out of the peak radiation hours during the day. The CAC-05/06 heat accumulators for the preheating section dispatch hot water to the CAC-01/02/03/04 heat accumulators, coupled with heat pumps to raise the temperature until 45°C is reached. The main dimensions are described below in Table 12. As the system is pressurized, for CAC-05/06, these are constant volume tanks with the following inlets from:

- Secondary circuit from the IC-6S-05/07 heat exchangers, constant flow.
- Secondary circuit from the IC-6S-06/08 heat exchangers, constant flow.
- Recirculating hot water not supplied to the user.

The returning flows may cause overpressure, in which case the excess water is recirculated into the secondary circuit, or experience low pressure, for which water is injected from the mains system, depending on the required volume. Similarly, the CAC-01/02/03/04 heat accumulators have two inlets:

- Secondary circuit from the IC-6S-01/02/03/04 heat exchangers, constant flow which connect the primary circuit from the heat pumps BC-09/10/11/12.
- Preheating section from the CAC-05/06 heat accumulators.

To further control the dispatched hot water temperature, mains water is extracted and mixed at the outlet of the CAC-01/02/03/04 heat accumulators.

Table 12: Heat Tank Dimensions and Operating Pressure.

CAC-05/06	
Capacity [L]	4,000
Diameter [mm]	1,600
Height [mm]	2,100
Pressure [bar]	8

* Source: Own elaboration based on documents delivered by Cintec S.A.

The heat tanks are manufactured from black steel 3724-ES with an 8 mm thickness. It is reinforced with fiber glass and with a 40 mm thick foam isolation to reduce heat loss.

5.1.5 Heat Pumps

The heating section of the SHW system is comprised by four heat pumps, which rise the temperature from the preheating section from 40°C to 45°C. Each heat pump BC-09/10/11/12 is connected to a heat accumulator through heat exchangers. In Table 13, the main design conditions for this equipment is presented. The schematic representation of a heat pump is similar to the water-cooled chiller shown in Figure 25.

Table 13: Design conditions for Heat Pumps.

BC-09/10/11/12	Hot Water
Water flow [m^3/h]	5.3
Inlet temperature [°C]	61
Outlet temperature [°C]	65
Pressure loss [bar]	0.45
Capacity [kW]	18.5
Refrigerant	R-410a

* Source: Own elaboration based on documents delivered by Cintec S.A.

5.1.6 Mains Water Source

To correctly estimate the value and seasonal variation of the mains water temperature, a numeric correlation is presented in [65], influenced by annual ambient temperature and the depth of the mains system, as shown in Figure 28. These variations in the mains water significantly affect the performance of water heating systems, such as SWH and can lead to a 10% error in heat gains calculations. For instance, the solar collector's efficiency is described in the equation below.

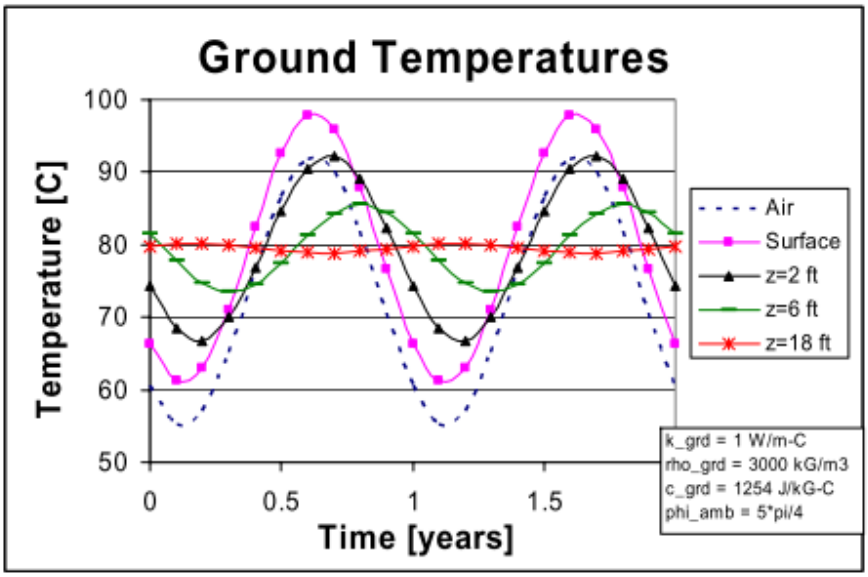


Figure 28: Plot of ground temperature annual variation, from [65].

$$\frac{\delta\eta_{coll}}{\eta_{coll}} = -F_R U_L \cdot \frac{\delta T_{mains}}{\eta_{coll} I} \tag{18}$$

Where the temporal dependence of the solar collector's efficiency η_{coll} is determined by the temperature differences δT_{mains} and the incident radiation I, as well as F_R and U_L described in Section 2.1.1 . The sensitivity of a SWH system is estimated at a 9.3% loss of annual energy savings[3]. This also considers that mains water temperature value varies in $\delta T_{mains} = \pm 3°C$.

[3] System parameters: $F_R U_L = 5W/m^2K; I = 400W/m^2; \eta_{coll} = 0.4$.

Equation (19) is used to determine the mains water temperature for an hourly basis for a year. Each term is described in the following Equations (20)-(23), based on the parameters presented in Table 14 and reference values for Santiago, Chile in Table 15.

$$T_{mains} = T_{mains,avg} + \Delta T_{mains}\, sin(\omega_{ann} t - \phi_{amb} - \phi_{mains}) \quad (19)$$

$$\mathrm{T_{mains,avg}} = \mathrm{T_{amb,avg}} + \Delta \mathrm{T_{offset}} \quad (20)$$

$$\Delta T_{mains} = [K_1 + K_2(T_{amb,ann} - T_{ref})]\Delta T_{amb} \quad (21)$$

$$\Delta \mathrm{T_{amb}} = \frac{\mathrm{T_{mon,max}} - \mathrm{T_{mon,min}}}{2} \quad (22)$$

$$\phi_{\mathrm{mains}} = \mathrm{K_3} + \mathrm{K_4}(\mathrm{T_{amb,ann}} - \mathrm{T_{ref}}) \quad (23)$$

$$\phi_{\mathrm{amb}} = \omega_{\mathrm{ann}} \cdot \mathrm{N} \cdot 24 + 90 \quad (24)$$

with:

- T_{mains}: The estimated value of the mains water temperature during an hour t during a year of 8760 hours.
- $T_{mains,avg}$: Average mains water temperature.
- ΔT_{mains}: Annual mains water temperature variation.
- ω_{ann}: Annual angular frequency of temperature variation.
- ϕ_{amb}: Angle offset representing the ambient temperature annual variation.
- ϕ_{mains}: Angle offset representing the mains temperature annual variation.
- ΔT_{offset}: Temperature difference from $T_{amb,ann}$ in T_{mains} correlation. This value is presented in Table 14.
- K_1, K_2, K_3, K_4: Numeric parameters defined in Table 14 representing reference temperatures.
- $T_{amb,ann}$: Annual average ambient temperature.

- T_{ref}: Reference minimum average monthly temperature registered in the selected location. Value selected is shown in Table 15.

- ΔT_{amb}: Average annual ambient temperature variation.

- $T_{mon,max}$, $T_{mon,min}$: Maximum and minimum average ambient monthly temperature.

Table 14: Parameters from Mains Temperature correlation, from [65].

Parameter	Value
ΔT_{offset}	0.6 °F
K_1	0.4
K_2	0.010 $°F^{-1}$
K_3	35 °
K_4	-0.01°/ °F

Table 15: Reference values for T Mains correlation.

Variable	Value
$T_{amb,ann}$	14.95 °C
$T_{mon,max}$	21.58 °C
$T_{mon,min}$	8.40 °C
T_{ref}	6.67 °C

The coefficients were obtained from nine locations in the U.S. and do not consider locations with freezing ground temperatures and or where snow cover exists during the winter. As the highest temperature is expected to be on January N=15, the angular frequency of the ambient temperature is estimated with Equation (24).

5.1.7 Centrifugal Pumps

Centrifugal pumps are installed between each mayor component of the system. This is required to maintain constant pressure in the system. The pumps are the main actuators of the control system and operate at constant speed, as detailed in Table 16. The main pumps in the system are:

- BACA-05/06: Solar hot water pumps. Connect the solar field with the primary circuit of the IC-6S-05/07 heat exchangers.
- BACS-05/07: Connect the mains water and excess water from the CAC-05/06 heat accumulators and the secondary circuit of the IC-6S-05/07 heat exchangers.
- BACS-06/08: Connect the mains water and excess water from the CAC-05/06 heat accumulators and the secondary circuit of the IC-6S-06/08 heat exchangers.
- BACA-07/08: Connect the outlet of the primary circuit of the IC-6S-06/08 heat exchangers to the CH-03 heat-recovery chiller. The system operates with one pump in stand-by mode.
- BPAF-04/05: Operate in the cooling circuit of the CH-03 chiller, connecting the discharged water used in the air conditioning system with the inlet of the evaporator's side. The system operates with one pump in stand-by mode.
- BACS-01/02/03/04: Operate the heating circuit between the BC-09/10/11/12 heat pumps and the corresponding heat accumulators through the IC-6S-01/02/03/04 heat exchangers.

Table 16: Centrifugal Pumps Design Conditions

	BACA-05/06	BACS-05/07	BACS-06/08	BACA-07/08	BPAF-04/05	BACS-01/02/03/04
Water flow [m^3/h]	5.3	5.3	13.5	52.5	30	5.3
Power [kW]	1.1	0.7	0.7	3.75	2.25	0.4
Static pressure [bar]	1.0	0.4	0.6	1.36	1.2	0.5
Speed [RPM]	2900	2700	2700	1450	1450	2900

* Source: Own elaboration based on documents delivered by Cintec S.A.

5.1.8 Control System

The system is equipped with temperature and differential pressure instrumentation in each main component. The operation of the solar-driven subsystem is defined by the following:

1. If the temperature registered at the outlet of the hot side of the CH-03 is less than 25°C, the internal pumps of the CH-03 stop and its operation is paralyzed.

2. Both heat accumulators CAC-05/06 have a relief valve which releases hot water if the average temperature is over 70°C. This continues until 50°C is reached.

3. If the temperature registered at the outlet of the solar collector field is greater than the average temperature in the hot water accumulators, the BACA-05/06 pumps are activated.

4. Hot water is dispatched from the CAC-05/06 tanks to the CAC-01/02/03/04 if the average temperature is over 55°C and there is an active demand in the system.

5. Mixing valves are activated if the outlet of the CAC-01/02/03/04 heat accumulators is greater than 45°C, in which case mains water is injected before being dispatched.

5.2 Weather Files and Available Meteorological Data

The weather information required for the TRNSYS simulation is obtained from two CR1000 dataloggers installed next to the solar field at the Beauchef 851 western building. One receives solar radiation measurements and the other registers meteorological variables shown in Table 17[4]. These data files are preprocessed and synchronized to the same timestamp, to then be fed into the solar collector ETC modules. The following values approximately describe this location:

- Time Zone: -3 GMT
- Longitude: 70.66 W
- Latitude: 33.45 S

[4] Radiation and other meteorological values are registered with a different time zone reference. While the CR1000 datalogger is set to GMT 0 Greenwich Time for global horizontal, diffuse horizontal and direct normal irradiance, values of ambient temperature, relative humidity, wind speed and wind direction are registered at GMT -4. From both dataloggers the data can be obtained with a one-minute interval. To correctly integrate these values to the TRNSYS simulation, these must be shifted to GMT -3 for Santiago, Chile.

Table 17: Available Meteorological Data.

CR1000	Units	CR1000 GEOF	Units
Global Horizontal Irradiance (GHI)	$[W/m^2]$	Ambient Temperature	$[°C]$
Diffuse Horizontal Irradiance (DFI)	$[W/m^2]$	Relative Humidity	[%]
Direct Normal Irradiance (DNI)	$[W/m^2]$	Wind Speed	$[m/s]$
		Wind Direction	[°]

One additional correction must be applied to the measured meteorological data. Solar time is used instead of the local clock time for computing the actual position of the sun, based on the apparent angular motion of the sun across the sky. For solar time, the solar noon corresponds to the moment in which the sun crosses the meridian of the local observer. The following equation is used to shift the local standard time (GMT -3) to the corresponding solar time.

$$Solar\ time - standard\ time = 4(L_{st} - L_{loc}) + E \tag{25}$$

where L_{st} is the standard meridian for the local time zone, L_{loc} is the longitude of the specified location and E is the equation of time given by Equation (26)[5].

$$E = 229.2(0.000075 + 0.001868 \cdot \cos B - 0.032077 \cdot \sin B - 0.014615 \cdot \cos 2B - 0.04089 \cdot \sin 2B) \tag{26}$$

where:

$$B = \frac{360(N-1)}{365} \tag{27}$$

and N is the number of the day of the year. Details on the path of the sun across the sky are found in Appendix D. With the use of solar angles, the quantity of incident radiation on a specific surface can be estimated.

In Table 18, an extract from the meteorological database is shown with the original and corrected timestamp obtained from the method described above.

[5] From [20], $L_{st} = 45° W$ and $L_{loc} = 70.66° W$.

Additionally, an example of a daily profile for global (GHI), diffuse (DFI) and direct (DNI) components of solar radiation is shown in Figure 29. A week profile of radiation values is also presented in Appendix F. An example profile of wind speed and direction is presented in Figure 30, as well as ambient temperature and relative humidity in Figure 31.

Table 18: Extract of Radiation Values and Corrected Timestamp for Santiago, Chile.

Date	Timestamp	Corrected Timestamp	GHI [W/m^2]	DNI [W/m^2]	DFI [W/m^2]
2018-05-17	14:38:00	13:25:00	356.062	494.077	112.992
2018-05-17	14:39:00	13:26:00	354.851	492.447	111.508
2018-05-17	14:40:00	13:27:00	363.792	508.183	111.835
2018-05-17	14:41:00	13:28:00	391.994	559.647	113.833
2018-05-17	14:42:00	13:29:00	373.827	525.763	111.582
2018-05-17	14:43:00	13:30:00	372.495	522.006	110.873
2018-05-17	14:44:00	13:31:00	377.372	529.307	111.235
2018-05-17	14:45:00	13:32:00	376.661	525.976	111.334
2018-05-17	14:46:00	13:33:00	397.283	566.664	110.358
2018-05-17	14:47:00	13:34:00	379.395	534.340	107.873
2018-05-17	14:48:00	13:35:00	389.749	553.692	107.084

*Source: Own elaboration based on measurements from meteorological station.

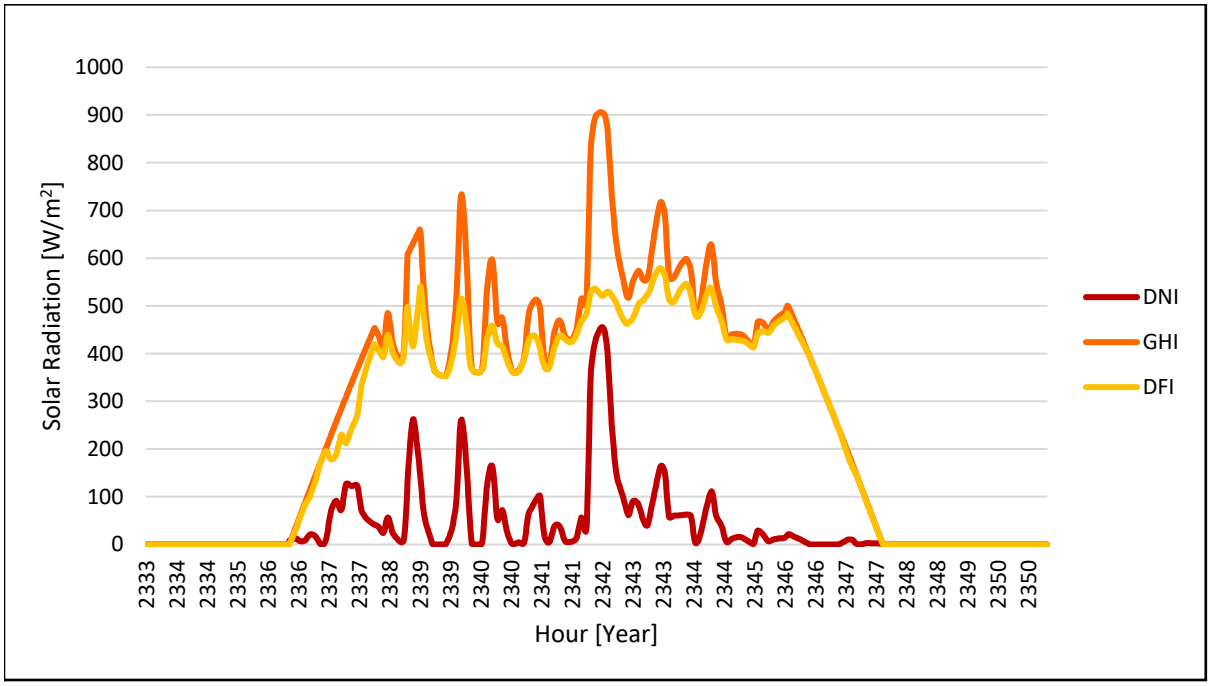


Figure 29: Extract of Solar Radiation Values. Source: Own elaboration based on measurements from meteorological station.

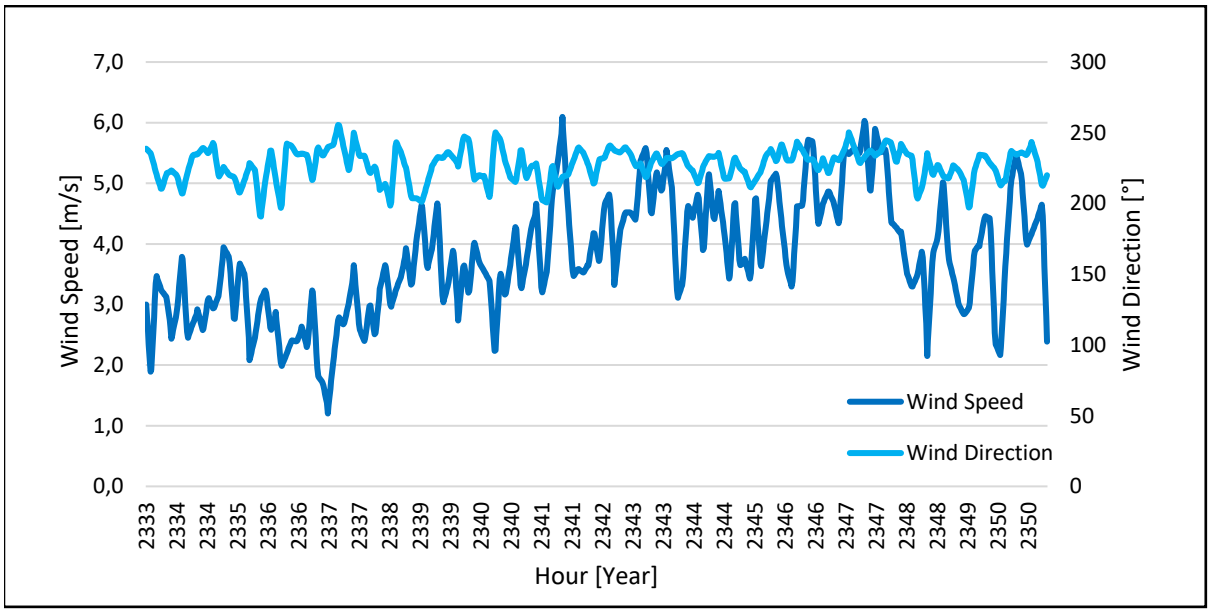


Figure 30: Extract from Wind Speed and Direction Values. Source: Own elaboration based on measurements from meteorological station.

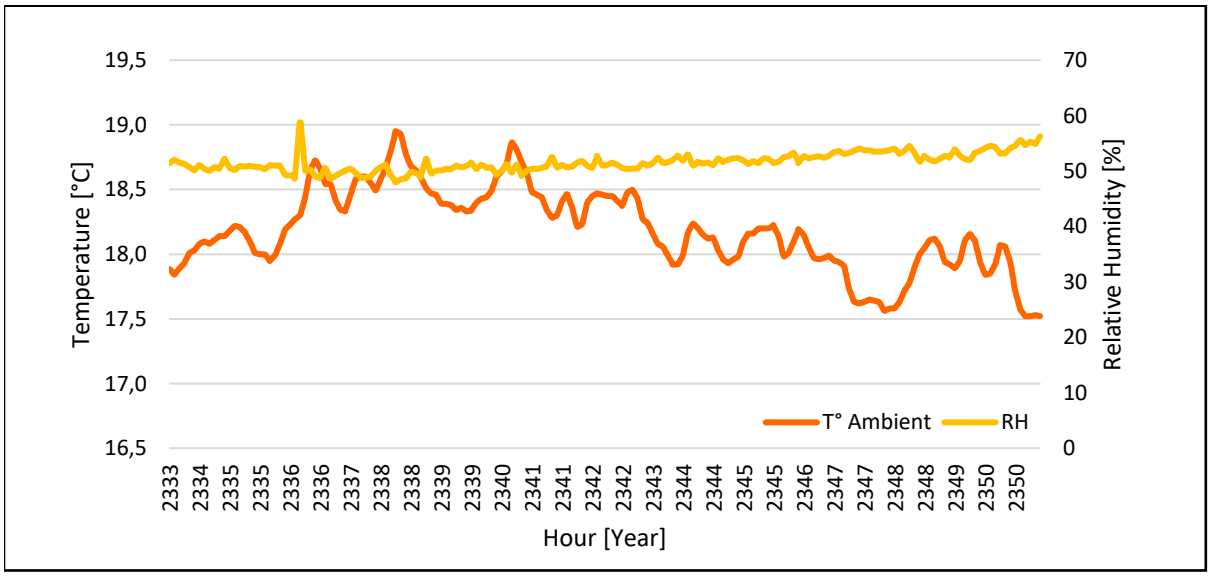


Figure 31: Extract from Ambient Temperature and Relative Humidity Values. Source: Own elaboration based on measurements from meteorological station.

5.3 Model Simplifications

For the SHW system implementation in TRNSYS, a series of assumptions are made aiming to reduce the complexity of the system and increase its flexibility to simulate different scenarios. The most important simplifications are the following:

- Heat Pumps are replaced by auxiliary water heaters with the same thermal capacities as the model described in Table 13.

- Design conditions are maintained for water flows returning to the CH-03 chiller which have not been simulated. These are the returning flow temperature to the evaporator, expected to be at 12.7°C, and the returning temperature from the secondary heat load designed at 44.5°C.

- The temperature control systems are simplified as the following:

    - If the temperature registered at the outlet of the solar collector field is greater than the average CAC-05/06 temperatures, the BACA-05/06 pumps are activated.
    - Hot water is dispatched from the CAC-05/06 tanks to the CAC-01/02/03/04 if the there is an active demand in the system.
    - Mixing valves are activated if the outlet of the CAC-01/02/03/04 heat accumulators is higher than 45°C, in which case mains water is introduced before being dispatched.

- As it is unlikely that the heat accumulators CAC-05/06 average temperature exceeds 70°C, the control subroutine which released hot water to the drain system is suppressed. Thus, the heat accumulators are also simplified.

- Operation of the system is maintained during 14 hours in weekdays from 7 AM to 9 PM. During weekends the operation hours are reduced to 11, from 8 AM to 7 PM.

- As no real-time measurements of the hot water demand profile were available, a standard weekly profile was drawn from estimations of user experience and the design conditions. This profile is shown in Figure 32.

For the construction of the TRNSYS model, the unit types and description are shown in Table 19 to Table 22. Finally, a schematic representation of the principal diagram blocks is shown in Figure 33, while the details for each sub-block are shown in Appendix A. From this model, operational fault-free data is generated to train the selected Neural Network architecture.

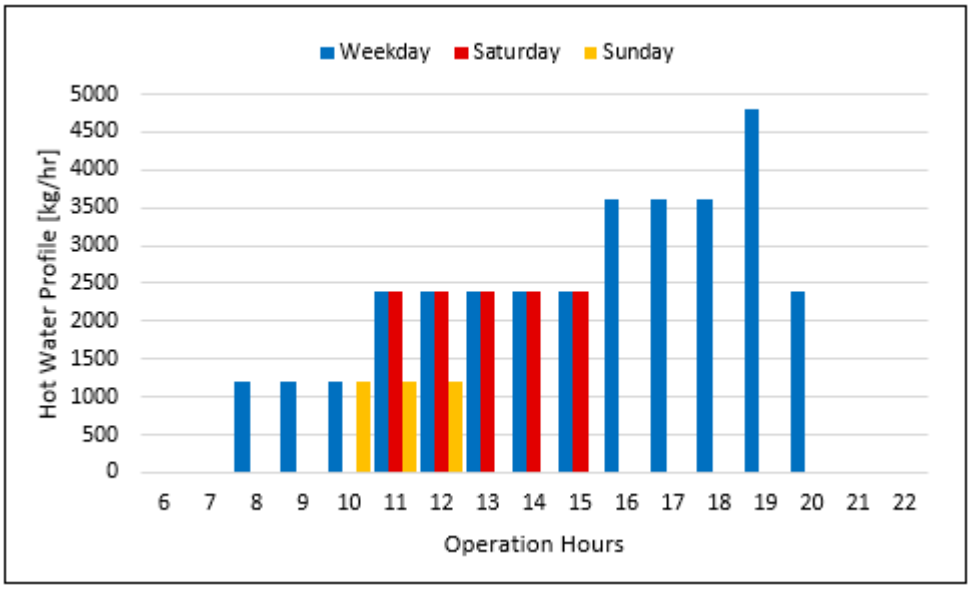


Figure 32: Weekly Hot Water Demand Profile. Source: Own elaboration.

Table 19: TRNSYS Models used for SWH System, Solar Section.

Name	TRNSYS Model	Type Name
Weather Files		
Santiago TMY	Type 99	Weather Data Reading and Processing – User Format
Solar Field		
Solar Collector 1-2	Type 71	Solar Collector; Evacuated Tube
Field-Collector Diverter	Type 647	Fluid Diverting Valve w/Up to 100 Outlets
Field-Collector Mixer	Type 649	Mixing Valve for Fluids w/Up to 100 Inlets
General Circuit		
BACA-05/06	Type 654	Single-Speed Pump
Temp. Control	Type 2	ON/OFF Differential Controller; Successive Substitution
Preheat Mixer	Type 649	Mixing Valve for Fluids w/Up to 100 Inlets
Return Diverter	Type 647	Fluid Diverting Valve w/Up to 100 Outlets
Operation	Type 14	Time Dependent Forcing Function
Weekly Load	Type 14	Time Dependent Forcing Function

*Source: Own elaboration from TRNSYS documentation [16].

Table 20: TRNSYS Models used for SHW System, Operation Schedule.

Name	TRNSYS Model	Type Name
Schedules		
Operation	Type 41	Load Profile Sequencer - Unique Days of the Week
OpWeek	Type 14	Time Dependent Forcing Function
OpSat	Type 14	Time Dependent Forcing Function
OpSun	Type 14	Time Dependent Forcing Function
Weekly Load	Type 41	Load Profile Sequencer - Unique Days of the Week
WeeklyLoad	Type 14	Time Dependent Forcing Function
SatLoad	Type 14	Time Dependent Forcing Function
SunLoad	Type 14	Time Dependent Forcing Function

*Source: Own elaboration from TRNSYS documentation [16].

Table 21: TRNSYS Models used for SWH System, Preheating Section.

Name	TRNSYS Model	Type Name
Circuit CAC-05/06		
IC-6S-05/06/07/08	Type 91	Heat Exchanger with User-Provided Effectiveness
BACS-05/06/07/08	Type 654	Single-Speed Pump
Manifold BACS OUT ½	Type 647	Fluid Diverting Valve w/Up to 100 Outlets
Manifold BACS IN ½	Type 649	Mixing Valve for Fluids w/Up to 100 Inlets
OutletCAC05/06	Type 649	Mixing Valve for Fluids w/Up to 100 Inlets
CAC-05/06	Type 534	Cylindrical Storage Tank with Immersed Heat Exchangers
Circuit CH-03		
Diverter CH	Type 647	Fluid Diverting Valve w/Up to 100 Outlets
Mixer CH	Type 649	Mixing Valve for Fluids w/Up to 100 Inlets
CH-03 RC	Type 666	Water-Cooled Chiller
BPAF-04/05	Type 654	Single-Speed Pump
BACA-07/08	Type 654	Single-Speed Pump

*Source: Own elaboration from TRNSYS documentation [16].

Table 22: TRNSYS Models used for SWH System, Heating Section.

Name	TRNSYS Model	Type Name
Circuit Heat Pumps		
HP Diverter	Type 647	Fluid Diverting Valve w/Up to 100 Outlets
HP Mixer	Type 649	Mixing Valve for Fluids w/Up to 100 Inlets
CAC-01/02/03/04	Type 534	Cylindrical Storage Tank with Immersed Heat Exchangers
BACS-6S-01/02/03/04	Type 654	Single-Speed Pump
IC-6S-01/02/03/04	Type 91	Heat Exchanger with User-Provided Effectiveness
BC-09/10/11/12	Type 138	Auxiliary Fluid Heater
TempValve	Type 11	Controlled Flow Mixer

*Source: Own elaboration from TRNSYS documentation [16].

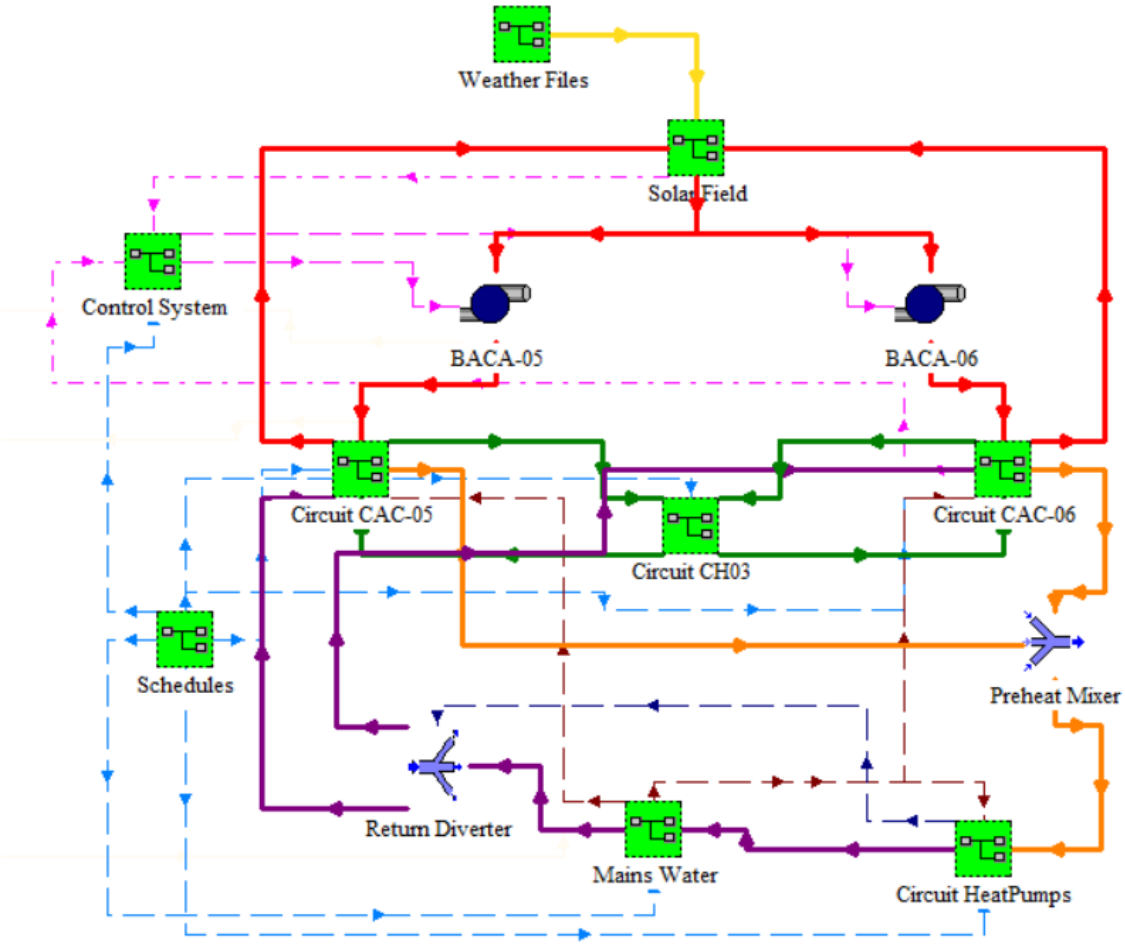


Figure 33: TRNSYS model schematic representation. Source: Own elaboration from TRNSYS documentation [16].

5.4 Monitored Variables

Due to the availability of weather data detailed in Table 17, the TRNSYS model was run for a period of 168 days, from April 7th to September 22th, 2018. The outputs of the simulation are registered each minute, giving a total of 241,922 data entries. The principal monitored values presented in Table 23, which describe the performance of the solar preheating subsystem. These variables are used to generate the data inputs for the DL models after a preprocess phase.

Although the monitored variables are not directly used as input for the prediction models, Table 24 summarizes those registered for the auxiliary heating system and the final output temperature to the hot water demand. The obtained values are approximations of the real operational conditions. In particular, the monitored variables for the heating section correspond to outputs from units which were not completely modeled, such as the heat recovery chiller CH-03 and the heat pumps BC-09 to 12. An extract from the generated timeseries for the monitored variables presented in Table 23 are shown in Figure 34 to Figure 38. As both heat accumulators CAC-05 and CAC-06 behave identically in the simulation, only one of the preheating circuits is further analyzed. The monitored components are then from the CAC-05 heat accumulator and the IC-6S-05/06 heat exchangers circuit.

Table 23: Monitored variables for preheating system.

Name	Description	Units	Subsystem
T Coll in	Inlet temperature to the solar collector field.	°C	Solar
T Coll out	Outlet temperature of the solar collector field.	°C	Solar
T Pipe out	Outlet temperature of the solar collector field considering ambient heat loss.	°C	Solar
Pump signal-1	Control signal to solar pumps BACA-05/06	-	Solar
Pump signal-2	Control signal to pumps BACS-05/06	-	CAC-05
IC05-Load-In	BACS-05 outlet temperature	°C	CAC-05
IC05-Load-Out	IC05 return temperature to solar collectors	°C	CAC-05
IC05-Source-Out	IC05 outlet temperature to CAC-05	°C	CAC-05
IC06-Load-In	BACS-06 outlet temperature	°C	CAC-05
IC06-Load-Out	IC06 outlet temperature to CAC-05	°C	CAC-05
IC06-Source-Out	IC06 return temperature to CH-03	°C	CAC-05
CAC-05	Average temperature in CAC-05 tank	°C	CAC-05
Pump signal-3	Control signal to pumps BACS-07/08	-	CAC-06
IC07-Load-In	BACS-07 outlet temperature	°C	CAC-06
IC07-Load-Out	IC07 return temperature to solar collectors	°C	CAC-06
IC07-Source-Out	IC07 outlet temperature to CAC-06	°C	CAC-06
IC08-Load-In	BACS-08 outlet temperature	°C	CAC-06
IC08-Load-Out	IC08 outlet temperature to CAC-06	°C	CAC-06
IC08-Source-Out	IC08 return temperature to CH-03	°C	CAC-06
CAC-05	Average temperature in CAC-06 tank	°C	CAC-06

*Source: Own elaboration from TRNSYS simulation results.

Table 24: Monitored variables for auxiliary heating system.

Name	Description	Units	Subsystem
T Chilled	Outlet chilled water temperature from CH-03	°C	CH-03
T Cooling	Outlet cooling water temperature from CH-03	°C	CH-03
COP	Coefficient of Performance of CH-03	-	CH-03
Pump signal-4	Control signal to pump BPAF-04/05	-	CH-03
Pump signal-5	Control signal to pump BACA-07/08	-	CH-03
Preheat outlet	Inlet temperature to the heating system	°C	HP
Heating outlet	Outlet temperature from the heating system	°C	HP
Pump signal-6	Control signal to pump BACS-01 to 04	°C	HP
IC-Source-Out	IC-01 to 04 return temperature to heat pumps	°C	HP
IC-Load-Out	IC-01 to 04 outlet temperature to CAC-01 to 04	°C	HP
Temp. Valve	Tempering valve outlet temperature to demand	°C	HP
Mains Temp.	Mains water temperature inlet for tempering valve	°C	HP
Return Temp.	Return hot water to CAC-05/06	°C	HP

*Source: Own elaboration from TRNSYS simulation results.

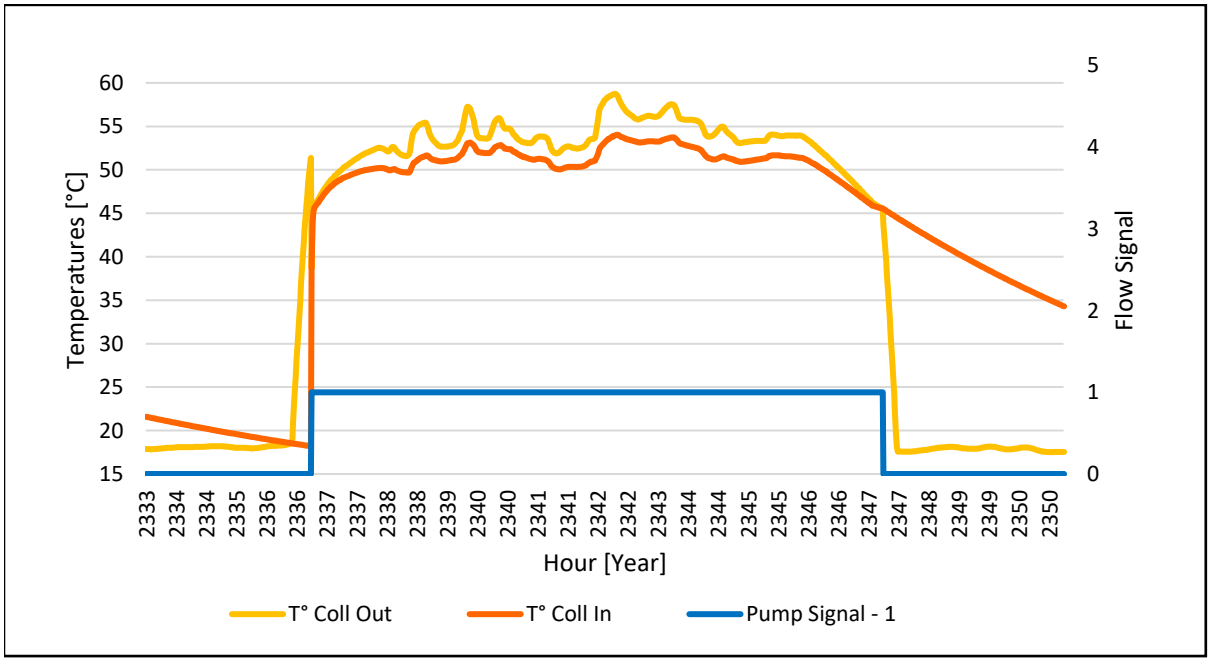


Figure 34: Extract from Solar Collector Field Monitored Values. Source: Own elaboration from TRNSYS simulation results.

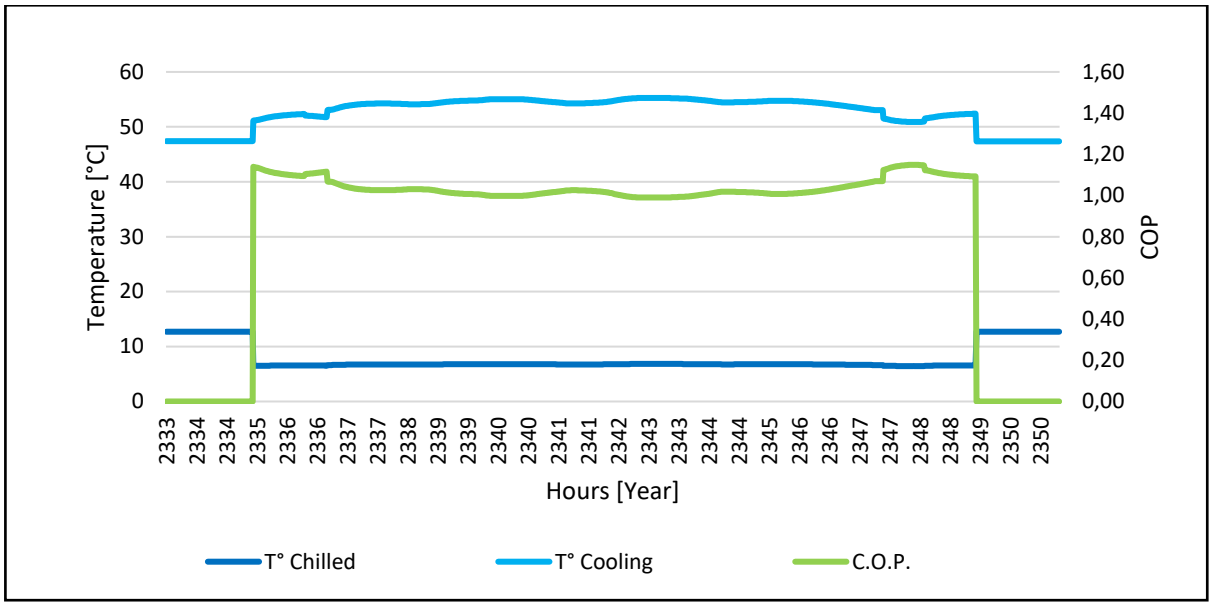


Figure 35: Extract from Heat Recovery Chiller Monitored Values. Source: Own elaboration from TRNSYS simulation results.

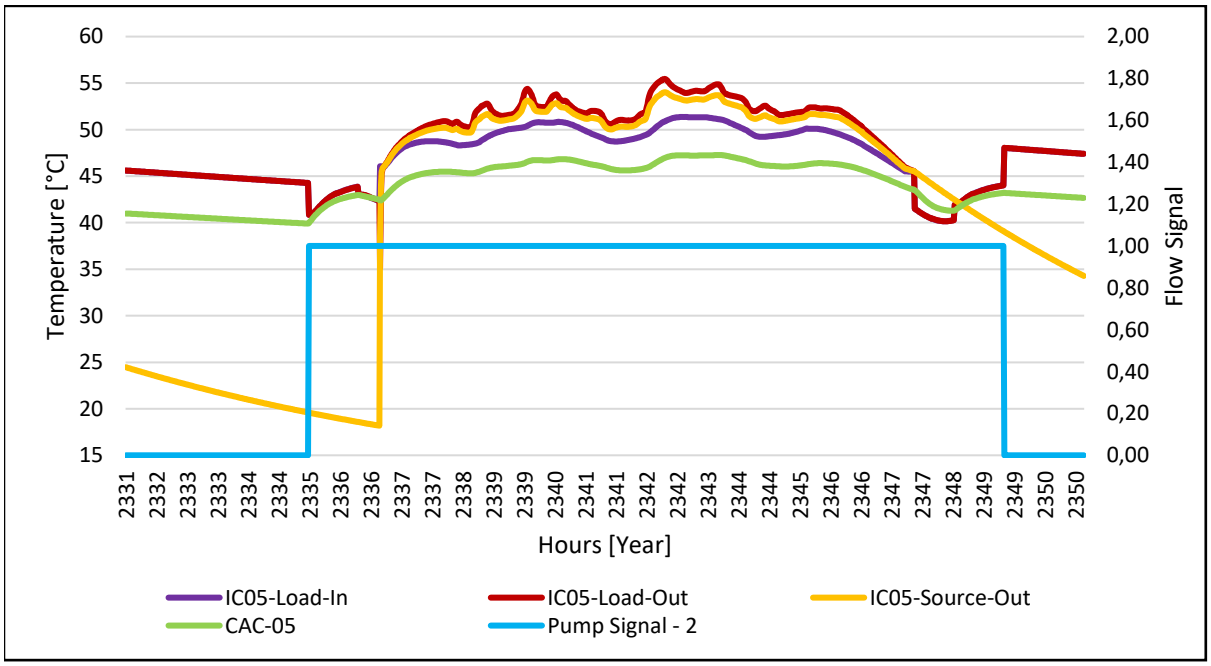


Figure 36: Extract from CAC-05 and IC-6S-05 Circuit Monitored Values. Source: Own elaboration from TRNSYS simulation results.

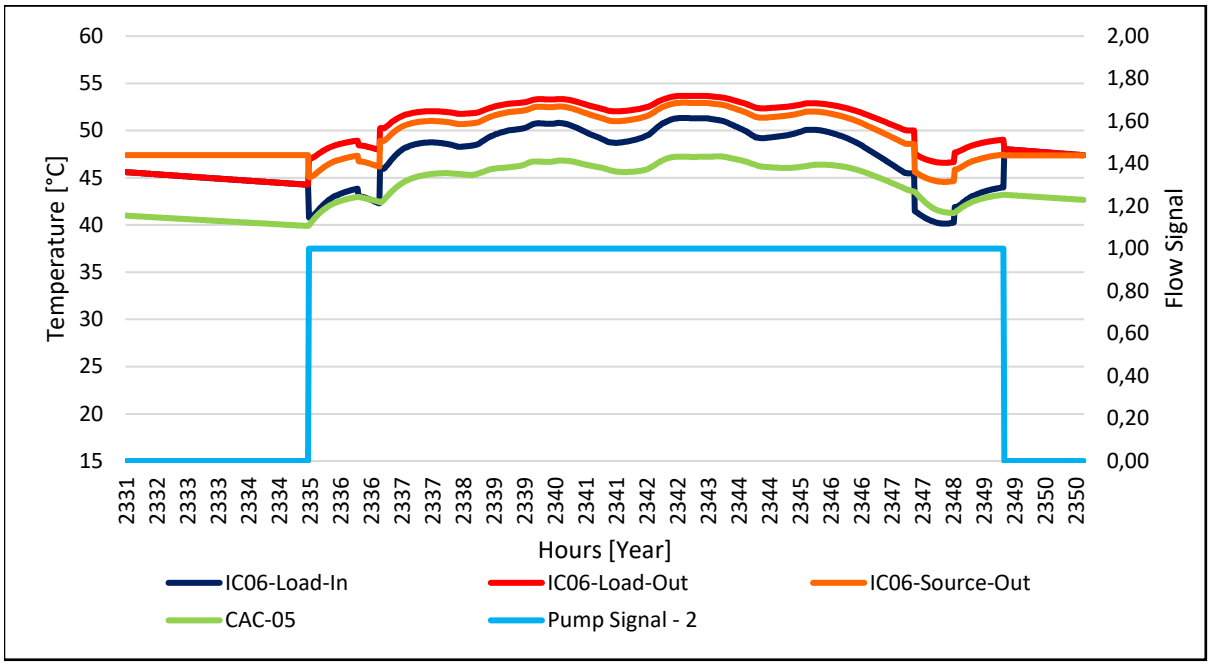


Figure 37: Extract from CAC-05 and IC-6S-06 Circuit Monitored Values. Source: Own elaboration from TRNSYS simulation results.

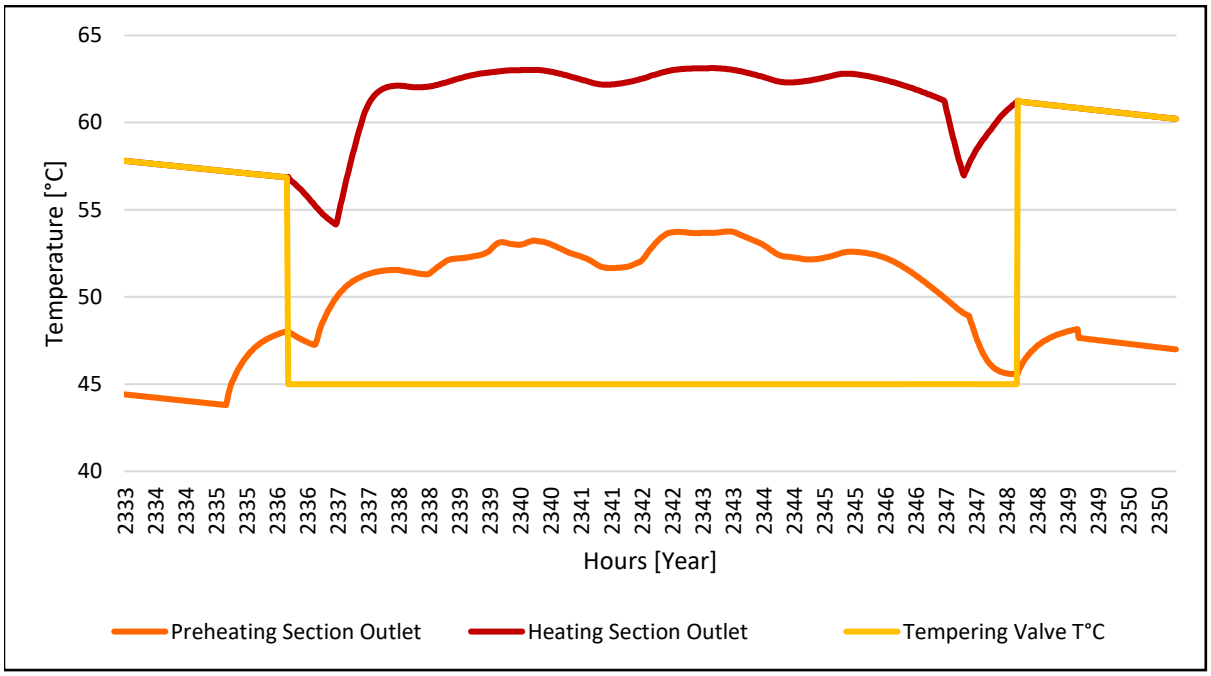


Figure 38: Extract from Heating Section Monitored Values. Source: Own elaboration from TRNSYS simulation results.

As shown in Table 25, by comparing the temperatures obtained by the TRNSYS model and a sample from the monitored data, the difference amounts to 22%. The highest difference is found for the IC-6S-05/06 heat exchangers, while the measured values from the IC-6S-07 are overlooked as faulty.

Table 25: Difference between design, simulated and measured temperatures in the system.

Temperature	Design	Simulated	Measured	Simulation vs Design	Simulation vs Measured	Measured vs Design
IC05-CS-IN	45.0	53.2	35.9	18.2%	48.3%	20.3%
IC05-CS-OUT	37.0	45.0	35.7	21.7%	26.3%	3.6%
IC05-CAC-IN	27.0	48.4	31.9	79.4%	51.7%	-18.3%
IC05-CAC-OUT	35.0	50.4	35.9	43.9%	40.5%	-2.4%
IC06-CH-IN	45.0	54.5	44.6	21.1%	22.3%	1.0%
IC06-CH-OUT	39.5	50.8	41.5	28.5%	22.2%	-5.2%
IC06-CAC-OUT	35.0	51.6	40.1	47.4%	28.8%	-14.4%
IC06-CAC-IN	29.5	48.4	26.8	64.1%	80.8%	9.2%
IC07-CS-IN	45.0	53.2	24.9	18.2%	113.8%	44.7%
IC07-CS-OUT	37.0	45.0	23.7	21.7%	89.9%	35.9%
IC07-CAC-IN	27.0	48.4	23.7	79.4%	104.3%	12.2%
IC07-CAC-OUT	35.0	50.4	24.6	43.9%	104.7%	29.7%
IC08-CH-IN	45.0	54.5	44.0	21.1%	23.8%	2.2%
IC08-CH-OUT	39.5	50.8	42.9	28.5%	18.2%	-8.7%
IC08-CAC-OUT	35.0	51.6	41.7	47.4%	23.6%	-19.2%
IC08-CAC-IN	29.5	48.4	40.5	64.1%	19.7%	-37.2%
CAC-05	35.0	45.2	40.3	29.3%	12.4%	-15.0%
CAC-06	35.0	45.2	38.0	29.3%	19.0%	-8.6%
CH-CH-IN	12.7	12.7	13.6	0.2%	-6.2%	-6.8%
CH-CH-OUT	6.0	6.8	12.0	12.7%	-43.4%	-99.2%
CH-CL-IN	44.5	47.6	42.4	7.0%	12.4%	4.8%
CH-CL-OUT	50.0	54.5	44.6	9.0%	22.2%	10.8%
COL-OUT	45.0	53.2	53.2	18.2%	-0.1%	-18.2%
		Average		32.8%	36.3%	-3.4%
		Corrected Average		31.1%	22.2%	-10.6%

*Source: Own elaboration from TRNSYS simulation results.

5.5 Fault Scenarios

To study the effect of faulty states in the SHW system with no prior recollected historic data, faults must be induced and studied in the TRNSYS simulation. From operational experience, the biggest issues arise from overheating due to lower demand profiles than expected, especially during summer. Also, during winter it is common to draw lower temperatures than required. As the latter is more related to the auxiliary heating system, the anomaly detection process is centered in the preheating solar section. Two main faults are explored: overheating caused by low hot water demand or pump failure, and the reduction of the solar energy yielded to the system caused by a drop of the collector's efficiency. The units that are directly affected are the following:

- BACA-05/06.
- Solar Collectors 1-2.
- Weekly Load.

It is important to monitor these components over the auxiliary heating system, since the heat input from the CH-03 chiller and the BC-09/12 heat pumps can hide possible damages in the SHW system. These components have not been modeled completely and thus only behave under design conditions with limited variability. For anomaly detection purposes, only the solar collector outlet temperature was analyzed, and not the average tank temperature due to the limits of the physical model to represent actual operation of the system.

5.5.1 Anomalous Water-Draw Profiles

As mentioned before, the solar field regularly presents overheating periods. This is mainly due to the logic with which the control system activates the circulation pumps connecting the solar field and the heat exchangers, as well as the lack of other heat-dampening methods to prevent this behavior and reducing the heat delivered to the load. This can be seen in Figure 39, in which even though the temperatures registered in the solar collectors reaches over 130°C, the solar field's outlet temperature on average is barely 32°C. This behavior is repeated over long periods of time, as shown in Appendix E. Even if the energy output from the solar field is low, the hot water discharged to demand is within operational range. This is due to the action of heat pumps and the heat-recovering chiller, and thus the system, from this point of view, operates correctly.

For the collector model installed in the system, the reference stagnation temperature is of 250°C, as shown in Table 5. However, while the solar collectors are designed to operate until their stagnation temperature, the overheating events may induce faults in auxiliary components, leading to possible failures in the system. Although no formal fault history has been recorded, malfunctions derived from overheating have occurred and concern has been expressed by the campus' administration team on several occasions. For instance, the centrifugal pumps have already been

replaced once before due to corrosion failure, affected by temperature rises. As a rule of thumb, a 10°C temperature increase causes the chemical reaction rate to double, however it strictly depends on the operation conditions of each system [66].

Hence, the outlet solar field temperature is monitored for unexpected changes. During normal operation, as the system operates *pulling* hot water to supply the demand from the heat accumulators, these variations arise from sudden changes in the water-draw profile perceived by the system. An example is shown in Figure 40, in which an induced fault is able to increase up to 10°C the normal operating temperatures for over three hours due to a drastic fall of hot water demand. This event may occur due to an effective reduction of demand profiles or other failures upstream of the solar preheating section of the system, as for example, pump failures. This behavior continues until the control system shuts down the BACA-05/06 pumps, leading to normal temperature decay during non-operating hours. The detection of rising temperatures outside expected values and normal patterns, can be useful to interrupt the operation of the system or to implement other temperature dampening measures.

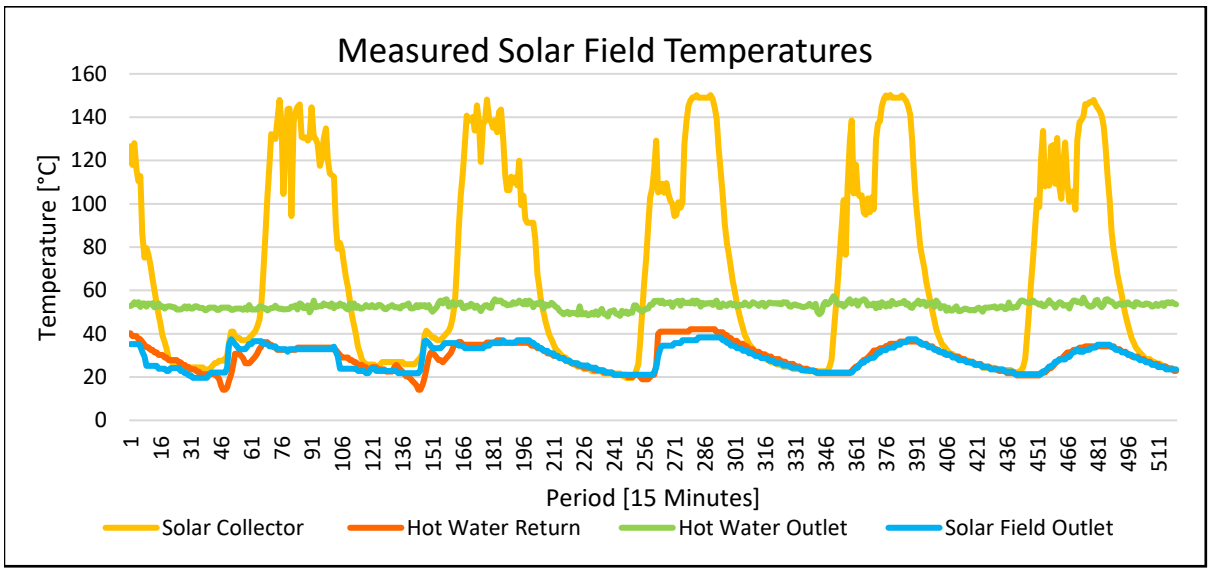


Figure 39: Extract from measured solar field temperatures. Source: Own elaboration measured values in the SHW system.

5.5.2 Solar Collector Efficiency

A sudden reduction of the solar collector's energy input to the system, which is not coherent with the input pump signal, can be a sign of a fault in the solar field. This may be caused by faults in the circulation system or in the collectors itself, reducing the useful heat extracted from the absorbed solar radiation as the heat loss to the environment increases. However, faults which develop over time such as dust accumulation, are not considered due to the limited operational history available for analysis.

From a thermal perspective, due to the integration strategy of the components installed, the auxiliary systems represented by both the heat-recovery chiller and the heat-pumps can absorb anomalous behavior of the solar field. For this reason, a lower solar field input is less recognizable as an anomalous behavior than a sudden increase of temperature, as modeled in the previous section. Through TRNSYS software, the drop of the solar collector's efficiency was modeled as a reduction of the incident radiation, as design parameters of said component are not variable during the simulation.

In Figure 40, an example of an altered temperature profile is shown. The effect of the induced fault on the solar collector's inlet and outlet temperatures drastically reduces the heat gains of the system and the temperatures experience up to 10°C decrease. However, this local fault does not extend from the solar field through the heat exchangers to the rest of the system, and thus after the fault has initially developed, the temperature drop is not significant. The prediction of the solar collector outlet temperature is also useful to quantify the expected renewable energy contribution to the system's gains and to recognize integration issues which lower the participation of the solar field in the SHW system.

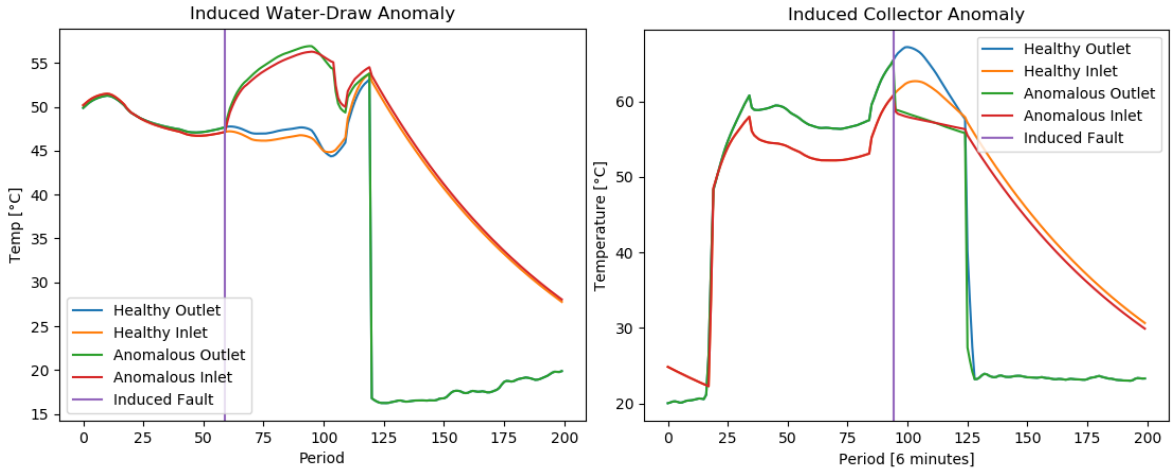


Figure 40: Examples of induced anomaly in Solar Collector's Temperatures.

Chapter 6

Proposed Diagnosis Framework

The proposed diagnosis framework consists of two principal tasks: the prediction of the target temperatures' timeseries, and the anomaly detection based on RMSE values when fed with new data from the TRNSYS simulation. The first focuses on the design and assessment of Deep Learning techniques to accurately represent the timeseries from the data delivered by the TRNSYS model, and the latter on the selection and evaluation of sensible thresholds based on the incoming data.

6.1 Data Preprocessing

From the SHW system's monitored values shown in Table 23, a selection of features relevant to each target temperature is made. These selected features are then preprocessed to generate the input data matrixes for the DL framework, as shown in Figure 41. This initial selection of variables is an important step to reduce the complexity of the model and focus on local variations influencing each target temperature.

For instance, even if radiation measurements intuitively seem crucial for the prediction of the solar collector outlet temperature, these values and the other meteorological conditions are dropped, since their effect is already considered by other simulated variables. This is the case of the control signal of the solar pumps, which depends not only of the available radiation but also on the heat accumulator's temperature. For the average heat tank temperature, only the immediate component's temperatures shown in Appendix 3 are selected as the effect of other components are already accounted by the interactions of the heat exchangers. The selected monitored variables for each target temperature are shown in Table 26.

Table 26: Target Temperatures and Selected Monitored Variables.

Solar collector outlet	CAC-05 Average
Ambient temperature	Pump signal-1
Solar collector inlet	Pump signal-2
Pump signal-1	IC05-Load-Out
Pipe outlet temperature	IC06-Load-Out
	Mains Temp.
	Return Temp.

*See Table 23 and Table 24 for reference.

As it was mentioned before, the output data extracted as shown in Table 18 must be averaged over a period of six minutes to reduce variability of solar radiation measurements. As such, a day now consists of 240 periods of six minutes, significantly reducing the length of the sequences to be analyzed.

6.2 Training and Test Sample Generation

From the target temperatures and related variables defined in Table 26, the training, validation and test datasets are prepared to be fed into the network. The timeseries data is reshaped into sequence matrixes by a sliding window. The number of samples depends on the length of the sequence's time window t_w and the slide step t_s. Figure 41 depicts the methodology which from an original dataset of N_t features during a T_p period of time, three matrixes will be extracted.

A sliding window of size $t_w \times N_t$ passes over the original dataset, extracting the values within its boundaries for S iterations. As only the following timestep is predicted, the sliding step used is of one to extract continuous timeseries data. These values are then reshaped into matrixes where each row represents a sequence of a whole time-window. Three different length sequence are generated to assess the performance of the different architectures: one, three, and seven days, respectively. The seven-day sequence is included to consider the complete effect of the weekly hot water demand profile presented in Figure 32.

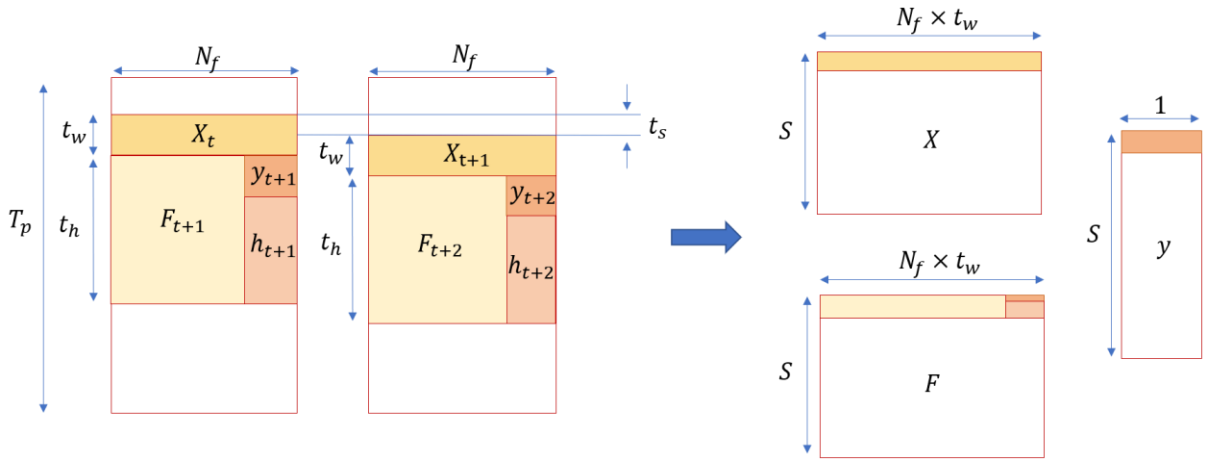


Figure 41: Preprocessing of the data.

The generated X matrix contains S number of samples of length t_w for N_f features. The labels y are shaped into a vector of length S for each of the sequences in the X matrix. Additionally, a *future* state matrix F is extracted, containing the t_h following timesteps for both the X (F_{t+1}) and the y (h_{t+1}) data.

This data is used to evaluate the ability of the model to predict several timesteps in the future, using the previous prediction as a new row in the input data. The selected number of future timesteps is ten, which corresponds to an hour in the future from the original prediction. Furthermore, the matrixes and labels are divided into the training, validation and test set. The test set consists of 20% of the original data, while the 80% remaining is divided into training and validation data. Another 20% of the data is taken for validation purposes. For both datasets analyzed, the size of the training, validations and test sets are presented in Table 27.

Finally, each set of data matrixes must be normalized by a scaler operation, so as the magnitude of any specific feature does not bias the model. This is done through a Min-Max normalization as shown in Equation (28). From a vector X, the minimum (X_{min}) and maximum (X_{max}) values are extracted to scale the data. This operation is only fit to the X training data and then is applied to the validation and test sets. The labels, in this case the output temperatures, are untouched by this step.

$$\mathrm{X_{scaled}} = \frac{\mathrm{X} - \mathrm{X_{min}}}{\mathrm{X_{max}} - \mathrm{X_{min}}} \tag{28}$$

Table 27: Dataset division and number of samples.

Datasets	N° Previous Days	Series Size	Train Set	Validation Set	Test Set
Solar Collector Outlet T°	1	240	5836	1460	1824
	3	720	17510	4378	5472
	7	1680	23654	5919	7393
Average Tank T°	1	240	5836	1460	1824
	3	720	17510	4378	5472
	7	1680	23654	5919	7393

6.3 Hyperparameter Selection

Parametric models, such as ANN, are defined by a set of trainable and non-trainable parameters. The latter, also known as hyperparameters, define the basis of the architecture and consist on a set of values which must be chosen based on the input data, the task the model is required to do, and its performance. In RNN models, the following decisions must be made regarding the architecture and configuration of the network:

- **Unit type:** ANN, Vanilla RNN and LSTM are the main models used.

- **Number of layers:** Depending on the relationships within the analyzed data, a deeper network can allow the mapping of more complex functions, as each layer will be able to extract different temporal features from the original data.

- **Number of units:** For each layer, the number of units must be selected. Deeper architectures with a smaller number of units per layer are generally preferred to wide shallow layers, to allow a more complex understanding of the data.

- **Dropout probability:** Dropout can be used as a regularization technique, with the probability of dropping a unit is of [0,1]. This inherent randomness allows the network to train without relying on specific weights.

- **Activation function:** Each hidden layer has an activation function. This are non-linear function such as Tanh or ReLU for hidden layers. These are shown in the following equations.

$$\mathrm{Tanh(x)} = \frac{2}{1 + e^{-2x}} - 1 \tag{29}$$

$$ReLU = \begin{cases} 0 \; for \; x < 0 \\ x \; for \; x \; \geq 0 \end{cases} \tag{30}$$

- **Optimizer:** The selected optimizer defines the method by which the weights and biases of the network are updated through backpropagation. Various techniques exist, responsible of the minimization of the objective function through gradient descent. Adaptive methods, such as Adam [67] and RMSprop [68] obtain more stable results than other optimization techniques [69].

- **Learning rate:** Defines the pace at which the optimizer affects the gradient descent used during the backpropagation process. Each optimizer has its own set of internal parameters other than the learning rate, but only the effect of the learning rate variations is studied.

- **Batch size:** The batch size determines the number of samples that are fed into the network per training iteration. A small batch size may increase the prediction's precision but lose the ability to generalize results, however larger batch sizes require more memory space to train.

- **Number of epochs:** An epoch is defined as a complete, forward and backward, pass of all the batches of the dataset through the network. The performance of the models depends on how many training epochs it reaches. The optimizer and learning rate, as well as the train and validation error are used to diagnose the number of required epochs for the model to converge.

- **Sequence length:** The length of the input data sequence also affects the performance of the framework for future predictions. Shorter sequences are more easily captured with Vanilla RNN, while LSTM and GRU can process more complex data [70].

- **Weights and bias initializers:** Variable initialization can have an impact on the time the model takes to converge during and its ability to locate a convenient local minimum. Usually biases are initialized at zero. However, weights can be initialized by different means using random set values generated by scaled and truncated normal distributions.

- **Loss functions:** Mean-Square Error (MSE) and Root Mean-Squared Error (RMSE) are frequent metrics used to evaluate the precision of predicted values in supervised ML tasks. The minimization of the RMSE, shown in Equation (31), tends to penalize further the difference between the real and the predicted values than the MSE.

$$\text{RMSE} = \sqrt{\sum_{i=1}^{n} \frac{\left(y_{real} - y_{pred}\right)^2}{n}} \tag{31}$$

The number of hyperparameters imply a complex process of tuning each of these values, besides the relationships and effect each hyperparameter has on the others and the overall performance of the model for a specific input data. Through trial-and-error, acceptable results may be achieved, although other techniques, such as a grid search, have been proposed to methodically reach better results. In the following sections the different model configurations that were trained are presented, as well as a schematic representation of the architecture used.

6.4 Proposed RNN and LSTM Models

The analyzed RNN and LSTM-based models trained configurations are shown in Table 28 and Table 29. The length of the temporal image, units per layer of the RNN-LSTM model, units per layer of the output MLP, activations functions used, batch size and number of training epochs are presented. These values were proposed, tested and selected by comparing the RMSE obtained over the training, validation and test sets. Though trial-and-error process, it has been noted that some hyperparameters have a higher impact than others on the metrics utilized, and thus various alternatives have been explored for the number of units, layers and batch size.

A schematic representation of the RNN-based architecture is shown in Figure 42, in which the input data of size $N_f \times t_w$ is fed into the RNN unit. The values from the last temporal unit is extracted in a many-to-one configuration, and later flattened and passed through an MLP of N_h units to obtain the predicted values. To visualize a deeper architecture of an RNN, in Figure 43, all

the temporal outputs from the first RNN layer are then connected as input values to the second layer. This is repeated for every deep layer, leading to a similar output MLP as in the RNN.

Similarly, a visualization of the LSTM-based model is presented in Figure 44, in which the cell and hidden state are propagated through the temporal units to produce the outputs. Just as before, the last temporal output is passed through an MLP structure to predict the temperatures. A deep version is presented in Figure 45, in which the high number of trainable parameters significantly increases the epochs needed for the training and validation errors to converge.

Table 28: Hyperparameter values used for RNN-based model.

Model	**RNN**	**DRNN**		
Data Length	1 Day	1 Day	3 Days	7 Days
Number of RNN units	64 128 256	64-32 32-16	64-32 128-64-32	64-32
Number of MLP units	128 512	128	128	128
Activation Function	ReLU	ReLU	ReLU	ReLU
Epochs	100	100	100	100
Batch Size	32 64 128	32	32	32

Table 29: Hyperparameter values used for LSTM-based model

Model	**LSTM**		**DLSTM**		
Data Length	1 Day	3 Days	1 Day	3 Days	7 Days
Number of LSTM units	64	64 128	64, 32 128,64,32 256, 128,64	64, 32 128,64 128,64,32	64,32
Number of MLP units	256 512	512	256 512 1024	256 512	512
Activation Function	Tanh ReLU	Tanh	Tanh	Tanh	Tanh
Epochs	50 100	50 75	150	75 150 200	75
Batch Size	64	64	64	64	64

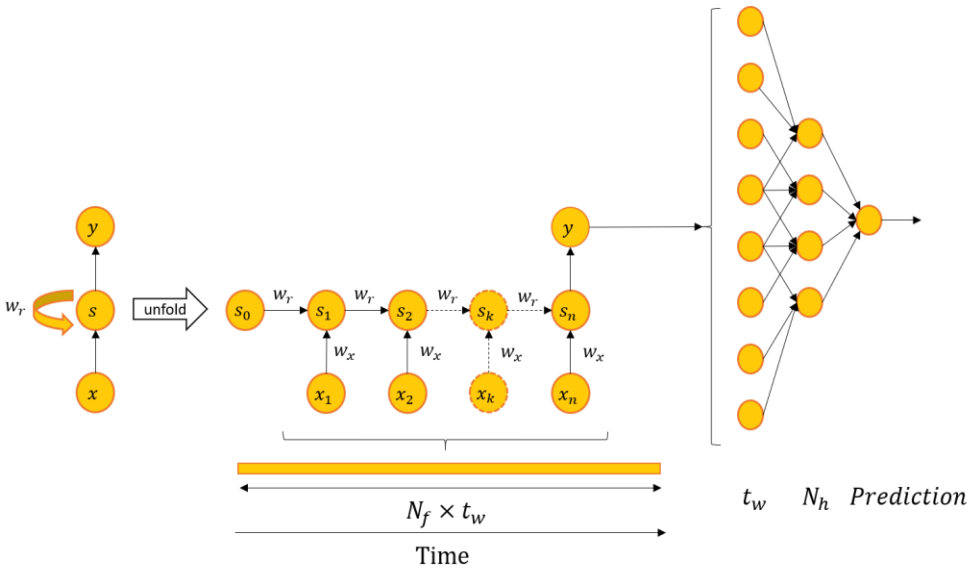


Figure 42: Schematic representation of proposed RNN architecture.

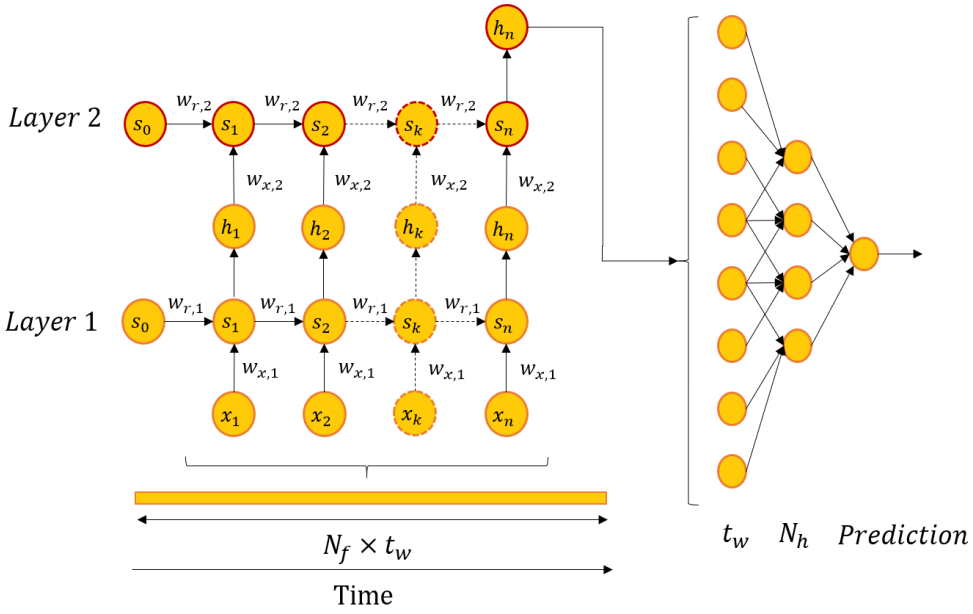


Figure 43: Schematic representation of proposed DRNN architecture.

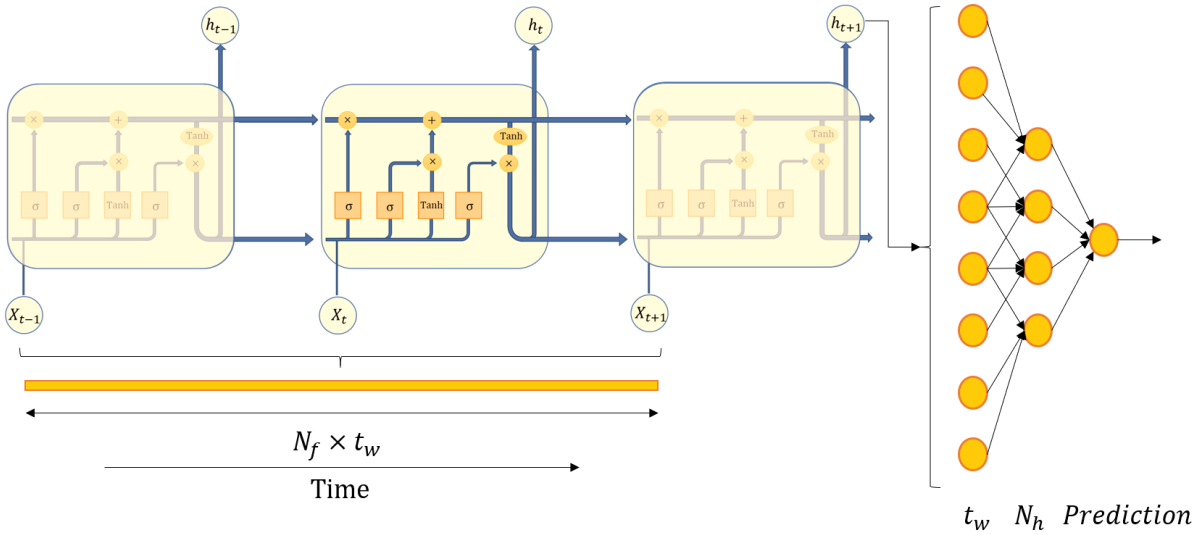


Figure 44: Schematic representation of proposed LSTM architecture.

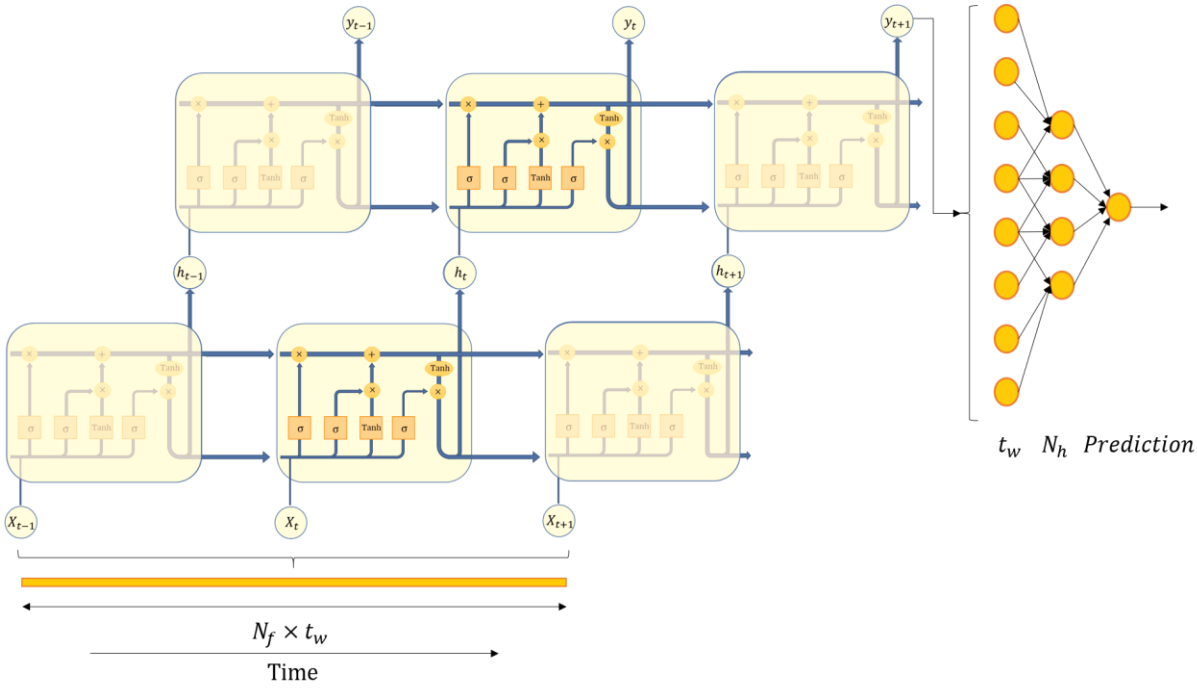


Figure 45: Schematic representation of proposed DLSTM architecture.

6.5 Anomaly Detection

From the frameworks presented in the previous section, the one which displays a better performance under the stated criteria is used to design an anomaly detector. The fundamental concept of this is to take advantage of a well-trained model which would respond to a new healthy data input similarly to the test set. However, when presented with anomalous data, the selected framework is expected to wrongly predict future values. The discrepancy between the predicted and real temperature values is interpreted as a sign of anomalous behavior in the system.

For this, anomalous data profiles are generated by TRNSYS through the use of user-defined data readers. A faulty water-draw forcing function and the reduction of incident radiation are introduced in the simulation showed in Figure 33. An anomalous water-draw is considered as the drastic reduction of demanded hot water from the system. This implies a larger return flow, which elevates the temperatures of the system. The original forcing function of 2400 L/h mean flow demanded is reduced to 200 L/h, maintaining the same weekly profile shown in Figure 32. This difference is reflected on registered temperatures almost 10°C higher in the solar field's temperature outlet. Similarly, for a temperature reduction of similar proportions, the effective incident solar radiation on the collector is blocked 80%. For the anomalous water-draw, the input of the Return Diverter shown in Table 19 is modified, while a reduction of the incident radiation is manually incorporated into the model based on the output of the Weather File used (refer to Table 18 and Figure 34 for TRNSYS model's intervened components). Both anomalous scenarios are modeled as sudden changes in the system, to take into consideration the high thermal inertia that it presents, due to the auxiliary heat sources. To effectively assess the ability of the fault detection model, it is desired that the effect of the anomaly causes temperature discrepancies for over at least 10 timesteps, or an hour. With the available information, fluctuations in the system with a cycle life shorter than an hour and lower magnitudes cannot be categorically defined as anomalous.

Three cases are analyzed for each of the anomalous profiles under different meteorological conditions during the period of April 7th and September 22nd. The available time-windows are filtered to operational hours of the solar field, as it is the main source of variability in the system. For each case, a healthy and an anomalous temporal image is used to test the predictions of the selected model. Then, the following methods are used to assess the model's performance:

- **Prediction errors:** The model is used for the prediction of 20 timesteps in the future since the induced fault in the system. Each prediction is compared separately to the original timeseries, for both the anomalous and the healthy profile. As mentioned before, the reconstruction of the anomalous profile is expected to present a higher RMSE.

- **Accuracy of anomaly detection:** Through the selection of a threshold sensible to the input data based on the variance of the model's prediction, the accuracy of the classification of health states are analyzed with the following metrics: accuracy, precision, recall and F1.

Chapter 7

Results and Discussion

The frameworks results are divided in the selection of the DL model to accurately predict the SHW system's selected temperatures, and in the ability to detect anomalies with these predictions. The accuracy of the models' results is compared through the RMSE metric during training and testing sessions, as well as the relative error of these values compared to the original sequences. From the tested values presented in Table 28 and Table 29, the following observations are made:

1. As a general result, the longer the temporal image and the deeper the network, the longer is the time required for the training process, amount up to 9 hours in the case of the DLSTM model trained with 7-day data. Refer to Section 4.5 for details of the computer used.

2. For the same length of temporal image provided, for example three days, LSTM models require an average of 5.24 hours to train, while RNN train in 3.25 hours. This is due to the larger number of epochs required for the LTSM training error to converge steadily.

3. For deep LSTM architectures, the use of Tanh as the activation function between layers yield more stable results than ReLU, which yields better results with the DRNN models.

4. While experimenting with different configurations, it is found that even though the optimizer Adam converges in a lower number of epochs, the obtained RMSE values are higher when compared to the RMSProp optimizer. This can be explained by how the exponentially decaying gradients are introduced in both optimizers to regulate the effective learning rate. On one hand, RMSprop uses an adaptive learning rate divided by an exponentially decaying average of the squared gradients to reduce the effect of vanishing gradients while updating the weights of the network. On the other hand, Adam additionally keeps an exponential decaying average of the moments of the gradients, increasing the convergence speed [71].

5. The choice between single values or a full sequence prediction depends on the availability and frequency of the training and test data. While a many-to-many approach is a bigger challenge for training and fine-tuning of the models' parameters, a many-to-one allows analysis for each timestep individually and simplifies an eventual implementation with real-time input data.

7.1 Temperature Prediction

The first criteria to assess the performance of the trained ANN, RNN and LSTM models is to obtain low RMSE values with small dispersion for the predicted temperature values for both models. From the models' point of view, significant efforts can be made to reduce the RMSE values lower than 1°C. However, from a practical perspective, obtaining a RMSE value below the measuring tools' uncertainty bounds is not necessary. Hence, with the present information available, a punctual prediction with an RMSE under 1°C is not directly differentiable from the thermal systems' natural variations.

7.1.1 Solar Collector Outlet Temperature

Low values of RMSE under 2°C are obtained for most of the trained models when trained to predict the solar collector outlet's temperature. In Figure 46, an example is shown for the test set belonging to the 3-day temporal image dataset for a DRNN model with three layers of [64, 32, 16] units each. The difference between the predicted y_{pred} and real y_{real} values is also presented, obtaining a relative error of 4.45%. Train and test errors for the other tested models are presented in Table 30, in which the model with the best results for each temporal image length is highlighted. This comparison is also shown in Figure 47, where the lower RMSE values are obtained for the RNN-based models for punctual temperature predictions.

In Table 31 it is shown that the best results are yielded by the DLSTM Ver 2 model of [128, 64] units for the 3-days temporal image. This model displays the lowest test error of 3.45%, the lowest variance (0.520) and the highest R^2 and explained variance score (0.997). Thus, even though the mean test RMSE of 1.47°C is slightly greater than the RNN counterpart, the predicted values of the DLSTM-Ver 2 model are more precise, reducing the uncertainty of the regression.

From both Table 30 and Figure 47, it is observed that RMSE values are significantly reduced when the length of the temporal image is increased, except for the DNN models in which the difference is less notorious. These errors are reduced by half when increasing the length from 1 day to 3 days of previous data for the RNN and LSTM models. A longer time-window can include the effects of weather and water-draw profiles to predict the future values with a higher accuracy.

Thus, from the standard demand profile presented in Figure 32 above, a 7-day temporal image may increase the likeliness between the real and predicted values if is able to incorporate better the sequence of weekdays and weekends. Nevertheless, Figure 48 shows the behavior of the training and validation errors throughout the training session, and the difference between 3-day and 7-day long temporal images is not significant and even tends to slightly overfit the training data.

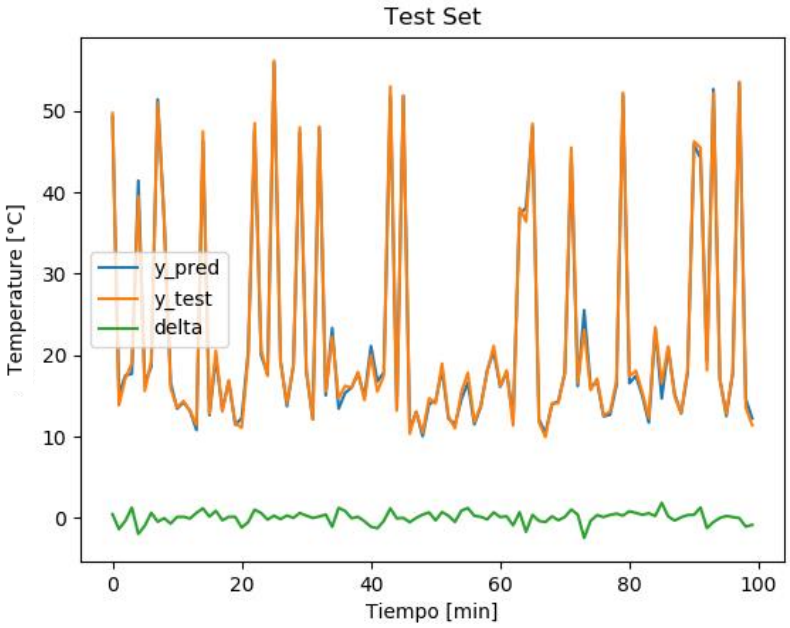


Figure 46: Test RMSE values for 3-day DRNN Ver 2 Model for Solar Collector Outlet Temperature.

Table 30: Train and test errors for selected models for Solar Collector Outlet Temperature.

Model	**Num. Days**	**Units**	**Train Error**	**Test Error**
DNN	1	64-32	0.53	1.74
RNN	1	64	0.71	1.68
DRNN	**1**	**64-32**	**0.78**	**1.28**
LSTM	1	64	2.18	2.5
DNN Ver 1	3	64-32	0.91	1.43
DNN Ver 2	3	64-32-16	0.54	1.29
DNN Ver 3	3	128-64	1.56	1.66
DRNN Ver 1	3	64-32	0.62	0.92
DRNN Ver 2	**3**	**64-32-16**	**0.55**	**0.89**
DLSTM Ver 1	3	64-32	0.67	1.42
DLSTM Ver 2	3	128-64	0.83	1.47
DLSTM Ver 3	3	128-64-32	0.43	1.66
DNN	7	64-32-16	0.39	1.26
DRNN	**7**	**64-32**	**0.55**	**0.94**
DLSTM	7	64-32	0.47	1.38

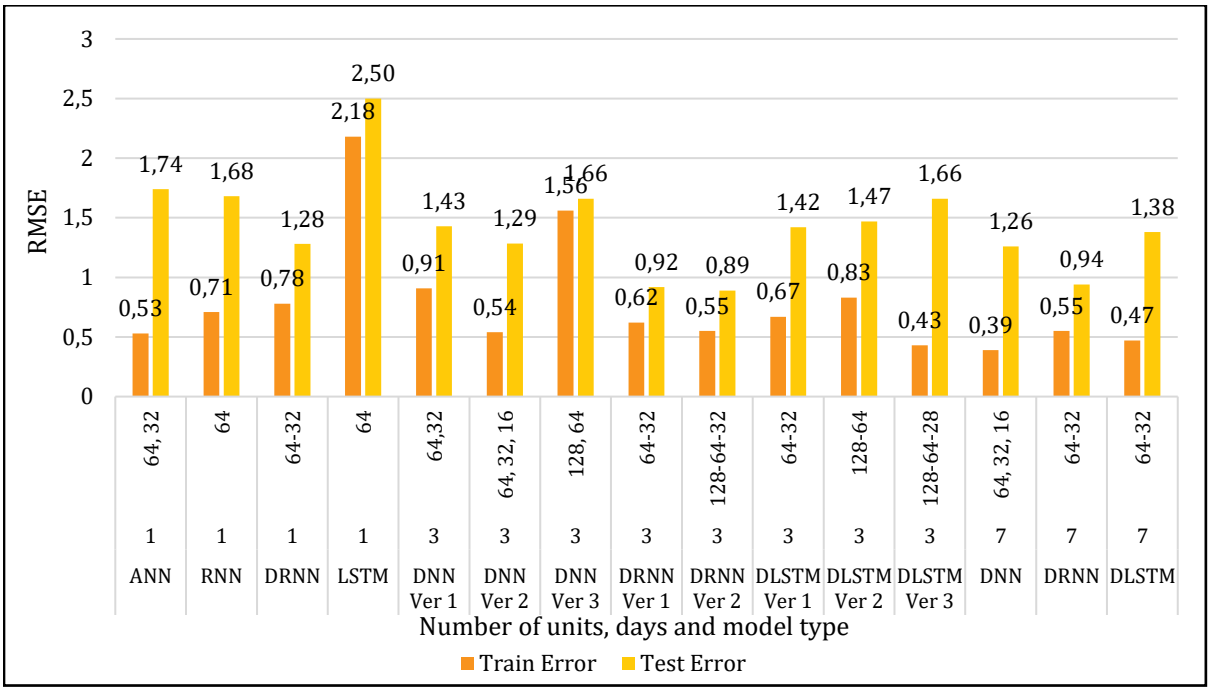


Figure 47: Comparison of train and test errors for selected models for Solar Collector Outlet Temperature Prediction [°C].

Table 31: Statistical Metrics for selected models for Solar Collector Outlet Temperature.

Model	**Num. Days**	**Units**	**Test Error %**	**Variance**	**R2**	**Explained Variance**
DNN	1	64-32	7.82	2.910	0.986	0.986
RNN	1	64	6.20	2.460	0.987	0.988
DRNN	**1**	**64-32**	**5.50**	**1.221**	**0.993**	**0.993**
LSTM	1	64	7.70	3.410	0.982	0.984
DNN Ver 1	3	64-32	6.97	1.587	0.988	0.991
DNN Ver 2	3	64-32-16	6.38	1.502	0.990	0.991
DNN Ver 3	3	128-64	9.08	2.425	0.984	0.986
DRNN Ver 1	3	64-32	7.10	1.516	0.991	0.991
DRNN Ver 2	**3**	**64-32-16**	**4.45**	**0.619**	**0.996**	**0.996**
DLSTM Ver 1	3	64-32	4.33	0.618	0.996	0.996
DLSTM Ver 2	**3**	**128-64**	**3.45**	**0.520**	**0.997**	**0.997**
DLSTM Ver 3	3	128-64-32	5.99	0.898	0.993	0.995
DNN	7	64-32-16	5.38	1.550	0.991	0.991
DRNN	**7**	**64-32**	**4.43**	**0.644**	**0.996**	**0.996**
DLSTM	7	64-32	6.10	1.510	0.991	0.991

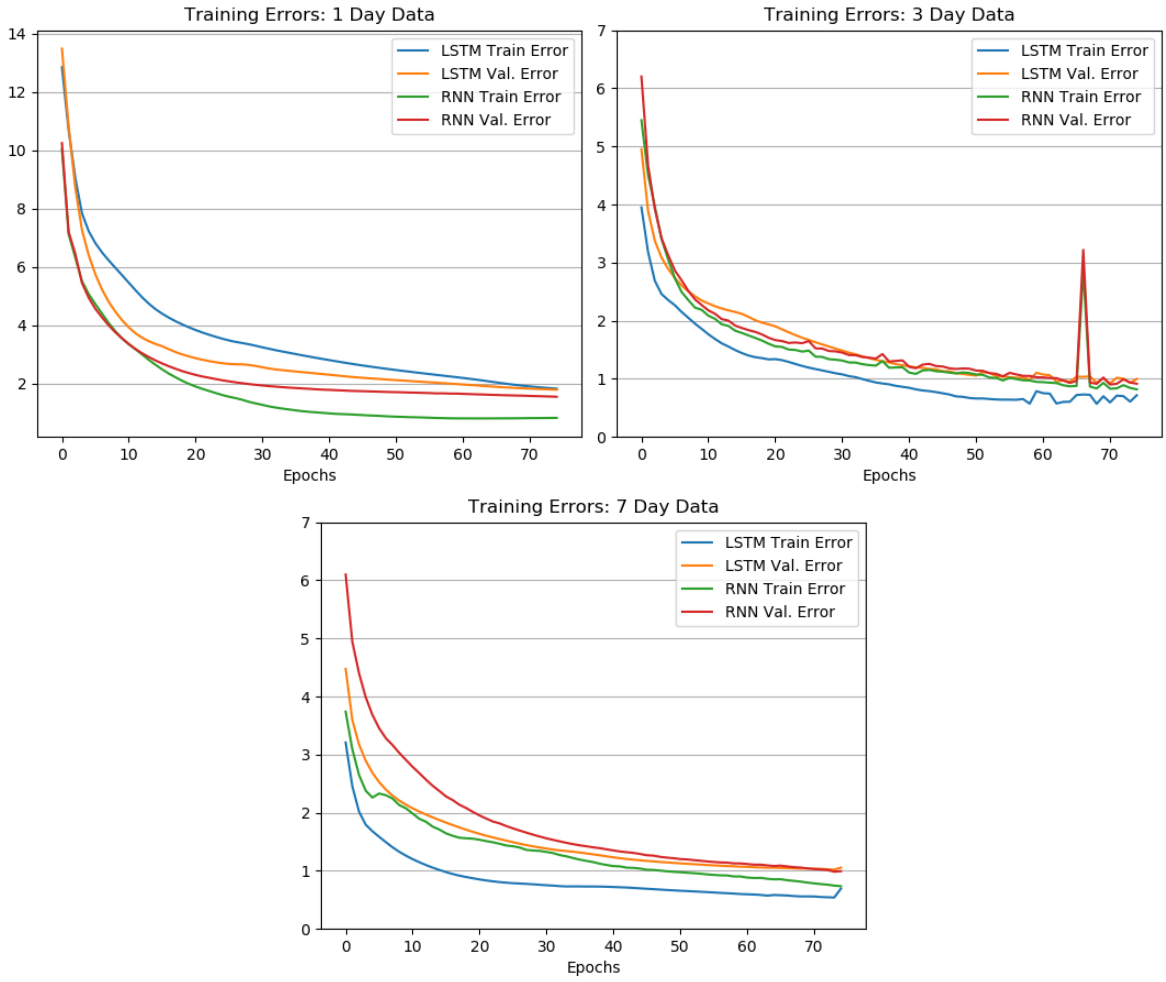


Figure 48: Comparison of training RMSE for different lengths of temporal images given.

7.1.2 Average Hot Water Tank Temperature

The prediction of the average temperature for the hot water tanks CAC-05/06 poses a lesser challenge for the models due to the stationarity caused by two main factors: the fluid's natural thermal inertia and the input of a stable heat source from the heat-recovery chiller. Thus, training and test errors are below the smallest registered for the solar collector outlet temperature and it is not as direct to identify the model which performs better. Additionally, the variances are lower and the R^2 and explained variance score are consistently higher.

Figure 49 shows an example of the difference of the predicted and real values for the average temperature of CAC-05. As the variability of the temperature decreases due to external factor, the causes of unstable predictions also are reduced. For the best performing models for the solar collector outlet temperature identified in the previous section, training and test error, as well as other statistic metrics are presented in Table 32 and Table 33. As shown in Figure 50, the model with lowest training and test error is yielded by the DLSTM Ver 2 model of only 0.18 and 0.20 °C respectively. Such low values of RMSE errors can only be verified with precise instrumentation and the quantification of the uncertainty the measurements carry. This model also presents the lowest variance of 0.033.

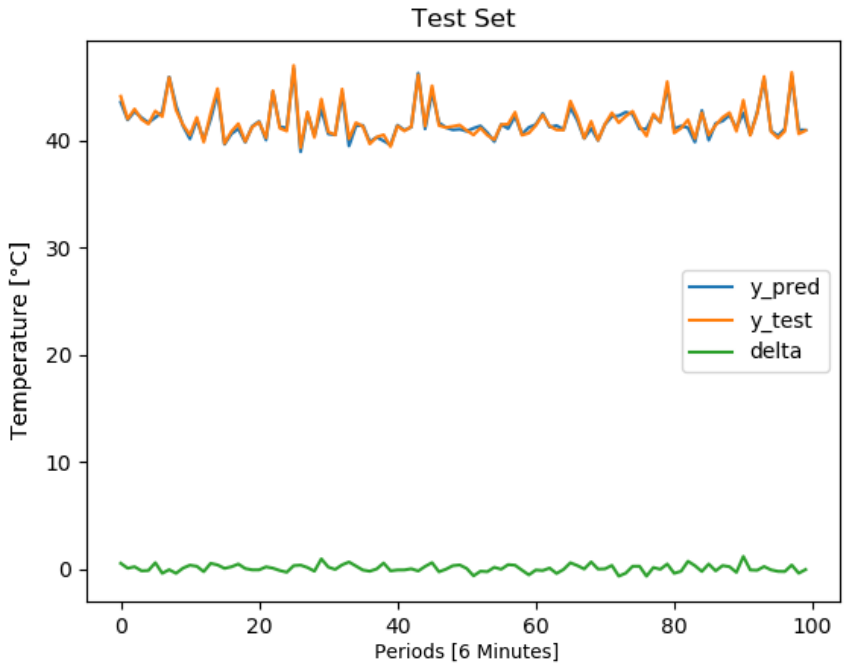


Figure 49: Test RMSE values DRNN Ver 2 Model for Average Tank Temperature.

Table 32: Train and test errors for selected models for Average Tank Temperature.

Model	Num. Days	Units	Train Error	Test Error
DNN Ver 2	3	64-32-16	0.40	0.37
DRNN Ver 2	3	128-64-32	0.25	0.36
DLSTM Ver 1	3	64-32	0.28	0.30
DLSTM Ver 2	**3**	**128-64**	**0.18**	**0.20**

Table 33: Statistical Metrics for selected models for Average Tank Temperature.

Model	Num. Days	Units	Test Error %	Variance	R2	Explained Variance
DNN Ver 2	3	64-32-16	0.88	0.069	0.96	0.979
DRNN Ver 2	3	128-64-32	0.87	0.129	0.96	0.961
DLSTM Ver 1	3	64-32	0.71	0.036	0.97	0.989
DLSTM Ver 2	**3**	**128-64**	**0.48**	**0.033**	**0.99**	**0.990**

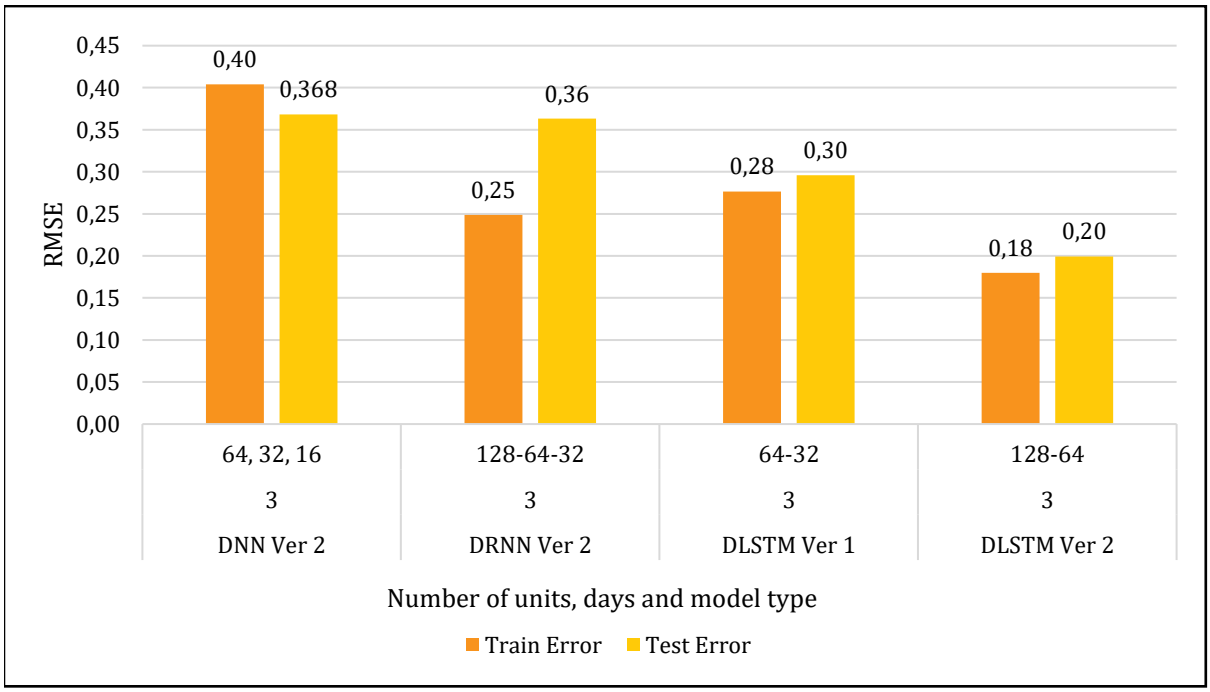


Figure 50: Comparison of train and test errors for selected models for Average Tank Temperature.

Finally, in Figure 51, the evolution of training and validation errors for the four models used to predict the average hot water tank temperatures are presented. Even if the training process is irregular, low RMSE values are reached, which reflects the small variability and thus high predictability of this monitored temperature. It can be noted that the ANN and LSTM models converge to a more usual behavior, whereas the RNN does not improve significantly during training.

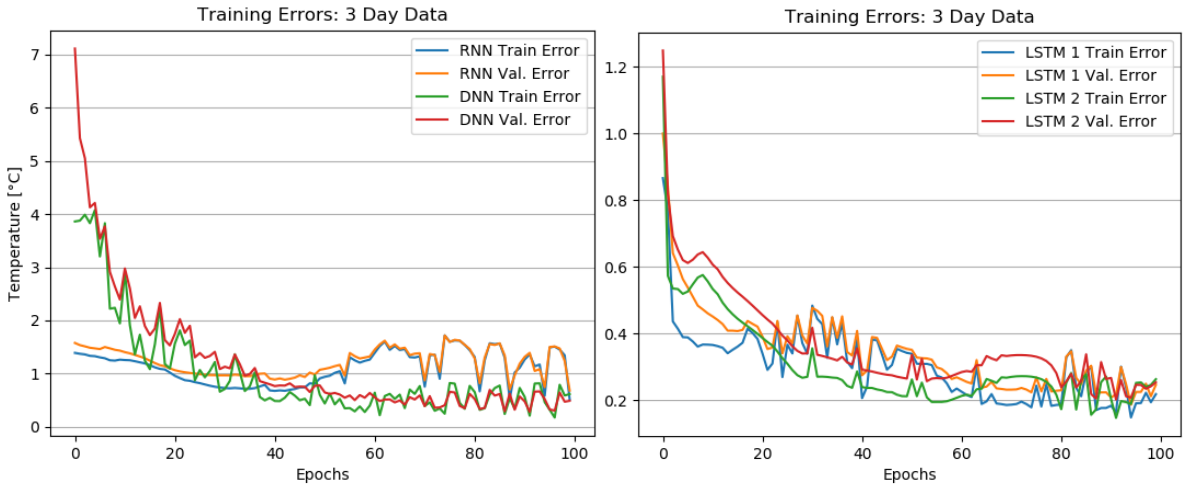


Figure 51: Comparison of training errors for 3-day length of temporal images given for average tank temperature.

7.2 Sequence Prediction

Although the trained models yield low RMSE values, isolated temperature predictions do not yield enough information to determine if a significant anomaly has occurred during operation in the system. Thus, the ability to accurately predict longer sequences is also an interesting result. Due to the frequency of the incoming measurements, it is possible to rely on single predictions for each timestep and asses the model's ability to interpret and regenerate the following sequence, rather than otherwise relying on sparsely-timed measurements which would require full future sequence prediction.

When generating a sequence, the model will update the sliding temporal window over the data and thus will incorporate previous predictions' errors in future values. Intuitively, if the single datapoint prediction error is small, the output sequence could maintain the real behavior of the data within a narrow uncertainty gap. Although the RNN-based models yielded lower test RMSE for single temperature predicted values, in Figure 52 it can be seen that these are not able to reproduce the data sequence correctly, and even though the temperature difference is low (at most 3°C), the behavior is inverted for Sample 2. For Sample 10, however, the three-layered RNN model of [64, 32, 16] units outperforms the two layered version of [64, 32] units, implying the existence of a more abstract relationship within the data which deeper models can extract.

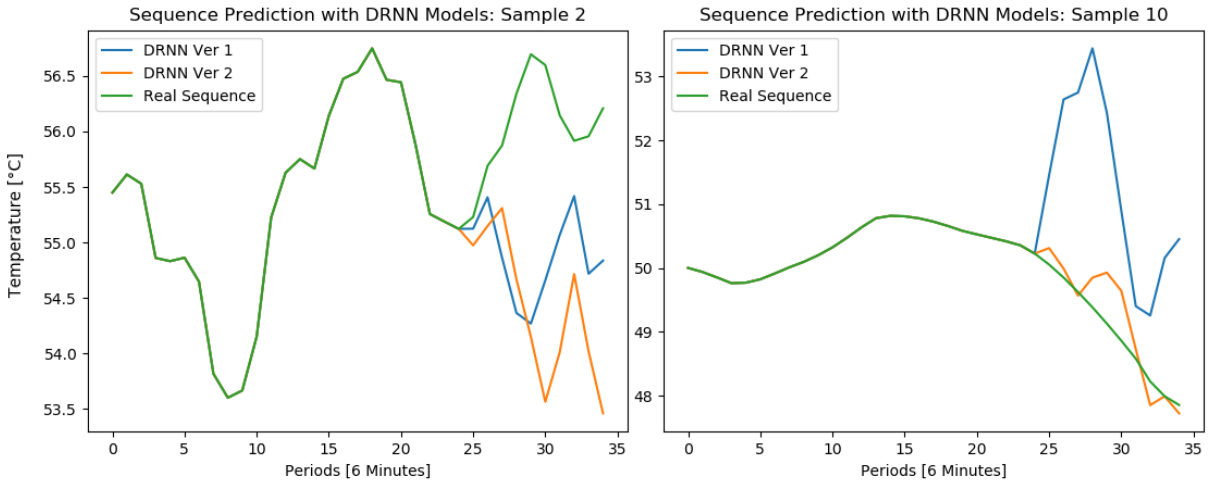


Figure 52: DRNN Model Sequence Prediction.

Less surprisingly, due to the fact then DNN architectures are not specialized in timeseries analysis, in Figure 53 it can be seen that the these models do not adequately replicate the sequence, exhibiting a similar behavior than the RNN-based models with an error rate as high as 6°C.

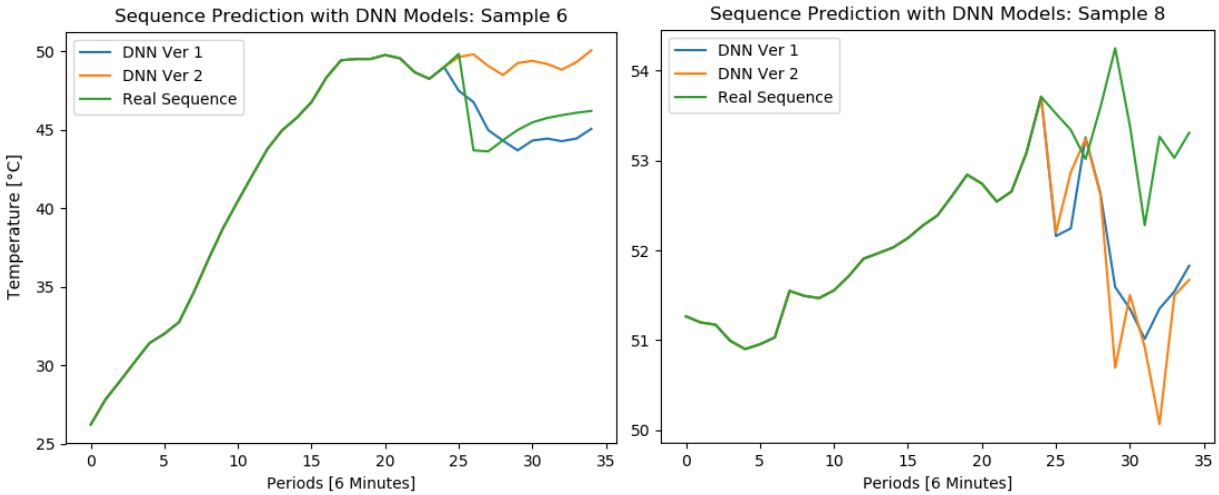


Figure 53: DNN Model Sequence Prediction.

While the LSTM-based models had achieved worse results than the RNN-based models predicted single temperature values, in Figure 54 it can be seen that it has better capabilities of following the temperature sequence with lower differences between the predicted and real values. The DLSTM-Ver 2 model, with [128, 64] units, achieves a maximum difference of 0.5°C on Sample 2. Even though the model exhibits the same problems of desynchronizing the predicted peak values for Sample 8, the error is reduced to 1°C.

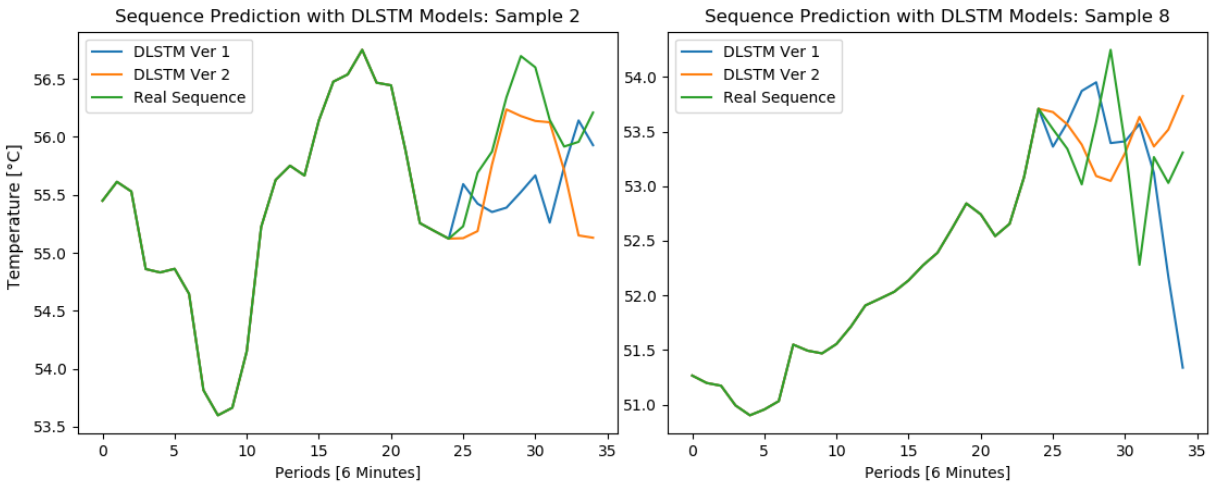


Figure 54: DLSTM Model Sequence Prediction.

Additional to these general observations, the sequence prediction accuracy was assessed for the solar collector outlet temperature as well the average tank temperature for the 3-day dataset.

7.2.1 Solar Collector Outlet Temperature

The outlet temperature from the solar field is bound to natural variations which make the prediction of future values a useful tool to assess the performance of the system. Sudden changes in weather, control signals and other physical mechanisms within the system have been modeled using TRNSYS software and from these, internal relationships have been extracted by the trained models to reproduce similar time sequences. In Table 34, the mean and standard deviation of the obtained RMSE values are presented for models DNN, RNN and LSTM-based for a 3-day length temporal images for a selection of 10 random samples. While the DNN-based model is outperformed by both the other models in terms of the mean RMSE obtained, it has a smaller standard deviation than the RNN model, which also possesses the lowest RMSE score. These results also reaffirm the idea that the LSTM-based model has a better ability to replicate the temporal sequence at hand, compromising the accuracy with the precision of the prediction.

A selection of these samples and the performance of the models is shown in Figure 55, while the remaining samples can be found in Appendix G. The samples have been filtered for operational hours of the solar collector field, depending on the control system of the solar pumps BACA-05/06, in which anomalous behavior is of higher risk due to the high temperatures reached. During these operating hours, the DNN model tends to underestimate the outlet temperatures, while the RNN version tends to overestimate it. Although the LSTM predictions can replicate the general temperature tendencies in Samples 6, 8 and 10, they also present discrepancies with the simulated values, such as in Sample 2. Even if the magnitude of the error is small, this is preferable than if the variations of the predicted sequences are unsynchronized with the tendencies of the real data.

Table 34: Sequence Reconstruction Error by Model for Solar Collector Outlet Temperature.

RMSE	**DNN Ver. 2**		**DRNN Ver. 2**		**DLSTM Ver 2.**	
Sample	**Mean**	**STD**	**Mean**	**STD**	**Mean**	**STD**
1	3.39	5.35	2.11	4.28	0.57	3.99
2	2.55	0.92	1.66	0.94	0.46	0.49
3	0.74	0.50	0.10	0.59	0.75	0.36
4	0.90	0.31	0.28	1.53	0.08	0.43
5	0.49	0.67	0.13	0.40	0.03	0.55
6	0.62	1.57	-1.36	2.59	1.09	2.13
7	0.86	0.40	0.49	0.69	1.40	0.36
8	1.40	0.73	0.32	1.06	0.12	0.88
9	1.17	0.56	0.16	0.41	0.80	0.43
10	0.62	0.69	0.20	0.36	0.17	0.50
Average	**1.27**	**1.17**	**0.41**	**1.28**	**0.55**	**1.01**

The mean values of RMSE and variance of the models' for a prediction horizon of 10 timesteps (corresponding to an hour) are also presented in Figure 56 to Figure 58 for each of the samples selected. These three models share some common defects, such as the apparent challenge to replicate rising temperature levels compared to decreasing temperature sequences[6]. This might be explained by the fact that the model for a 3-day length temporal image is not able to completely interpret the weekly water-draw profile, and thus struggles to identify the initial response of the control system that activates the circulation pump of the solar system.

As mentioned before, while the RNN and LSTM models consistently perform better than the ANN architecture, the latter yields lower variance for more than half of the samples tested. This result may suggest why ANN models have had a more widespread use than the more complex models which require significant efforts to select adequate hyperparameters and training conditions, as well as longer training times. Still, Figure 58 shows that the LSTM yields superior results for sequence prediction, obtaining the smallest variance between the samples. In Figure 59, the mean RMSE errors for the three models are compared, showing that the RNN model has the most unstable prediction tendencies. Unlike the ANN and LSTM models which overestimate the predicted temperatures except for one sample, RNN predictions overestimate and underestimate almost equally. This limits the integration of, for example, additional parameters in the cost functions to penalize more one tendency of inaccuracy or another.

The results yielded by the LSTM model suggest that sequence predictions can effectively benefit from the use of this architectures, given the availability of long enough previous data sequences.

[6] Sample 1. Refer to Appendix 12 and Appendix 13.

However, the desynchronized predicted values seen in Sample 6 suggest that a finer selection of input variables is required to effectively represent the relationship between the physical quantities.

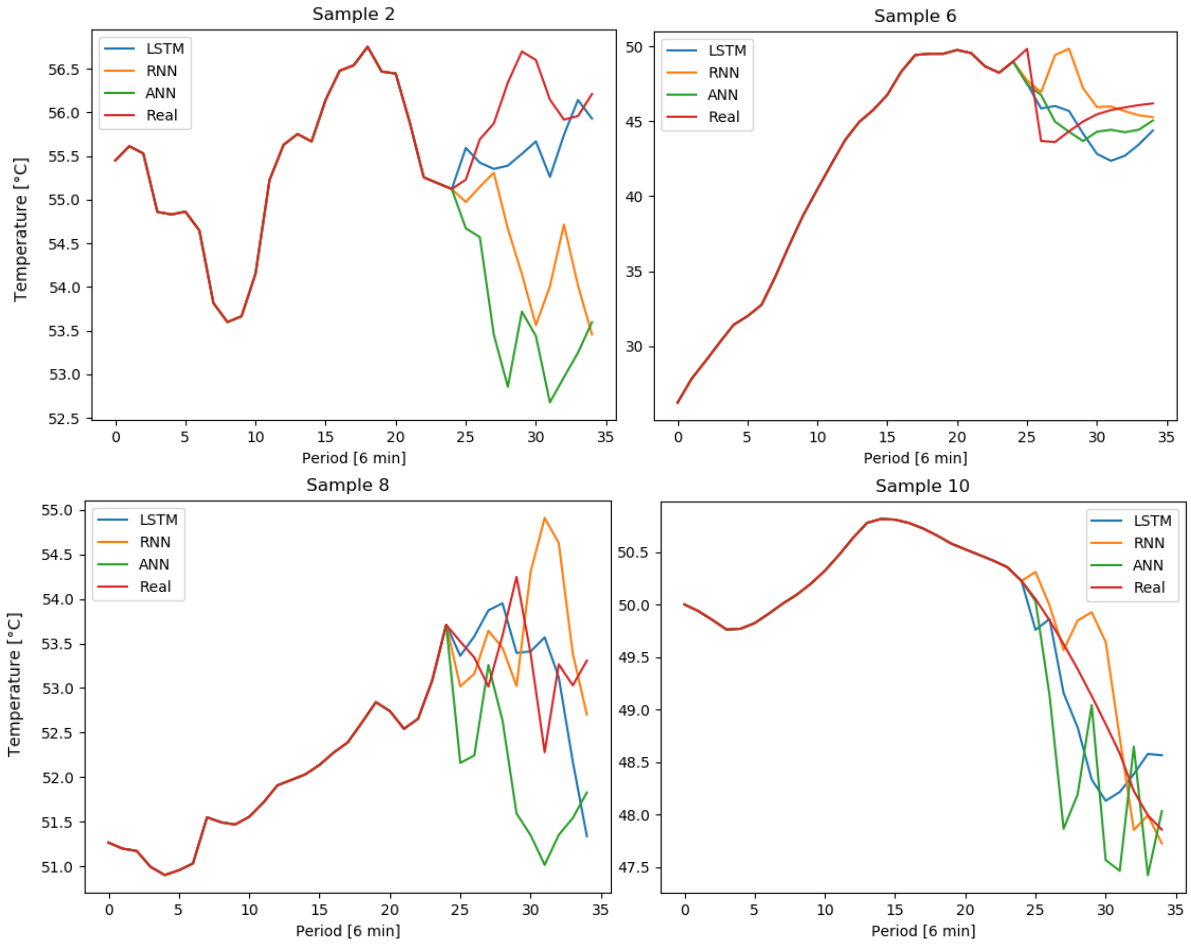


Figure 55: Reconstruction Samples for Solar Collector Outlet Temperature.

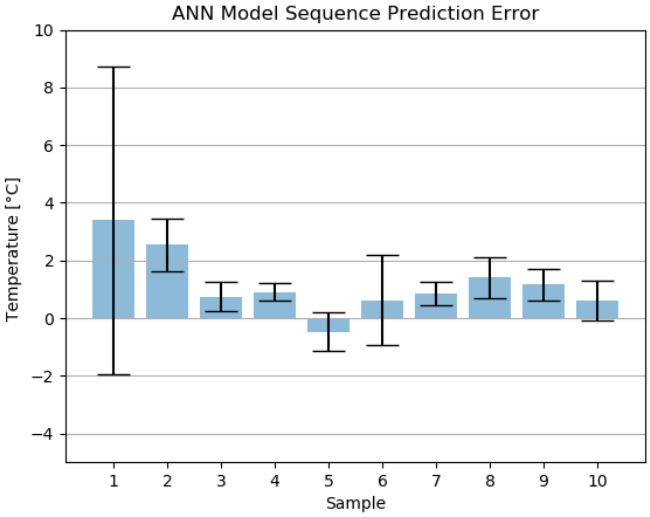


Figure 56: Mean Sequence Prediction Error for ANN Model for Solar Collector Outlet Temperature.

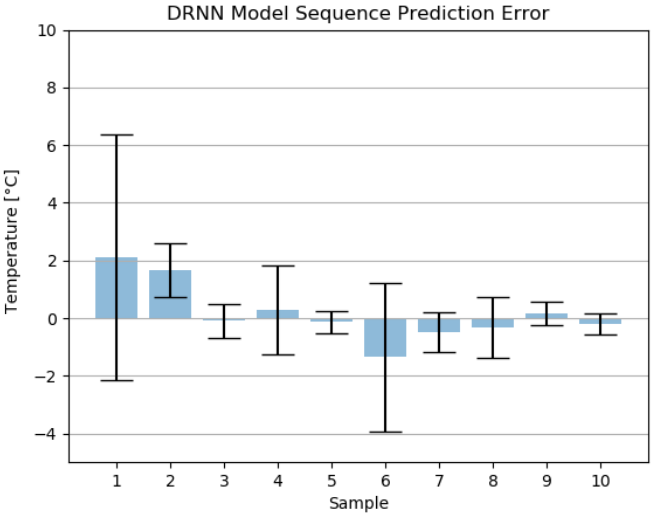


Figure 57: Mean Sequence Prediction Error for RNN Model for Solar Collector Outlet Temperature.

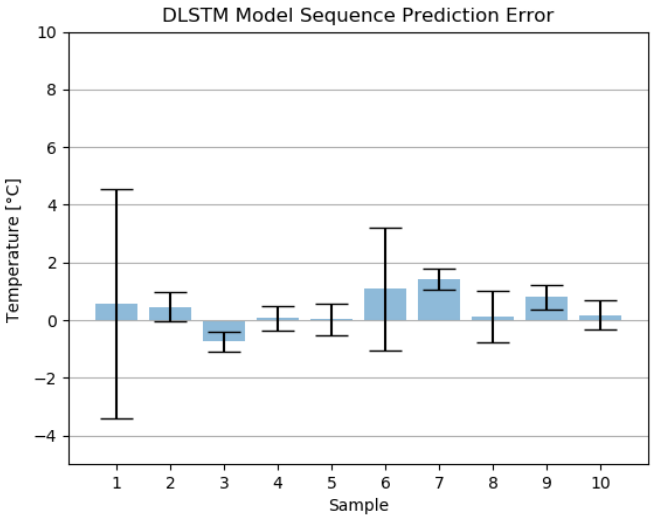


Figure 58: Mean Sequence Prediction Error for LSTM Model for Solar Collector Outlet Temperature.

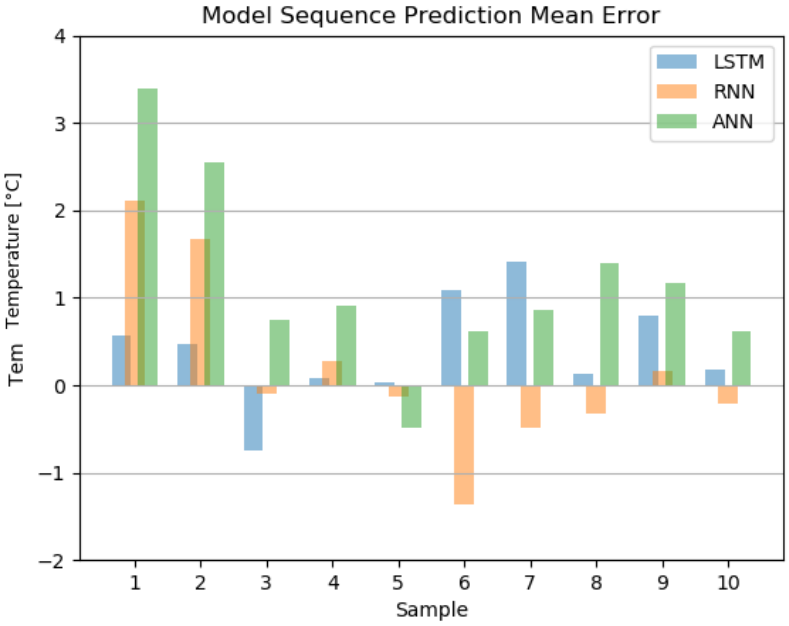


Figure 59: Comparison of relative Mean Prediction Errors for selected models for Solar Collector Outlet Temperature.

7.2.2 Average Tank Temperature

As stated before, the prediction of the average tank temperatures yields low RMSE error values and small variance, due to the natural inertia of a body of water of 4000 L. However, such small variability of the results is rather a symptom of the limitations of the physical model. This is due to the simplifications incurred during the simulation of the heat-recovery chiller. The lack of operational data to contrast and validate the simulation renders the representation of this component as a constant heat source rather than presenting the natural fluctuations occurring in the climatization system. This gives the thermal input from the IC-6S-06/08 heat exchanger a larger weight than the solar input transported by the IC-6S-05/07 components, as well as doubling the design water flow, as stated in Table 16.

In Table 35 the mean and standard deviation of the RMSE values are presented for the DNN, RNN and LSTM models for a 3-day length of temporal images. The stated behavior regarding the models is repeated, with the LSTM model vastly outperforming DNN and RNN models, in that increasing order.

Table 35: Sequence Reconstruction Error by Model for Average Tank Temperature.

RMSE	**ANN Ver. 2**		**DRNN Ver. 2**		**DLSTM Ver. 2**	
Sample	**Mean**	**STD**	**Mean**	**STD**	**Mean**	**STD**
1	0.19	0.11	0.22	0.28	0.01	0.03
2	0.31	0.09	0.21	0.30	0.09	0.08
3	0.51	0.13	0.21	0.10	0.08	0.11
4	0.22	0.10	0.04	0.07	0.11	0.15
5	0.16	0.18	0.41	0.31	0.13	0.06
6	0.29	0.12	0.58	0.74	0.19	0.28
7	0.24	0.22	0.11	0.16	0.03	0.05
8	0.05	0.25	0.53	0.22	0.04	0.04
9	0.06	0.16	0.51	0.22	0.07	0.09
10	0.12	0.20	0.12	0.20	0.28	0.15
Average	**0.22**	**0.16**	**0.29**	**0.26**	**0.10**	**0.10**

A selection of these samples and the performance of the models is shown in Figure 60, while the remaining samples can be found in Appendix H. Similar to the previous samples for the solar collector output temperature, these have been filtered for operational hours of the solar collector field risk due to the higher temperatures reached. For these predictions the RNN model is the least adaptable to abrupt temperature changes, tending to underestimate the real values. Nevertheless, even if the results enhance the LSTM model's ability to accurately predict temperature sequences, all three models yield acceptable results.

The mean values of RMSE and variance of the models' for a prediction horizon of 10 timesteps are also presented in Figure 61 to Figure 63 for each of the samples selected. In contrast to the tendency obtained for the other predicted temperature, here both DNN and LSTM tend to underestimate the sequence's values, while the RNN tends to do the contrary. Again, this model yields the higher variance for its predicted values.

In Figure 64, the mean RMSE errors for the three models are compared. In this case, both RNN and LSTM models do not have a clear tendency, while the ANN repeatedly underestimates the temperature values. Either way, each model can accurately predict temperature sequences within less than 0.5°C of uncertainty.

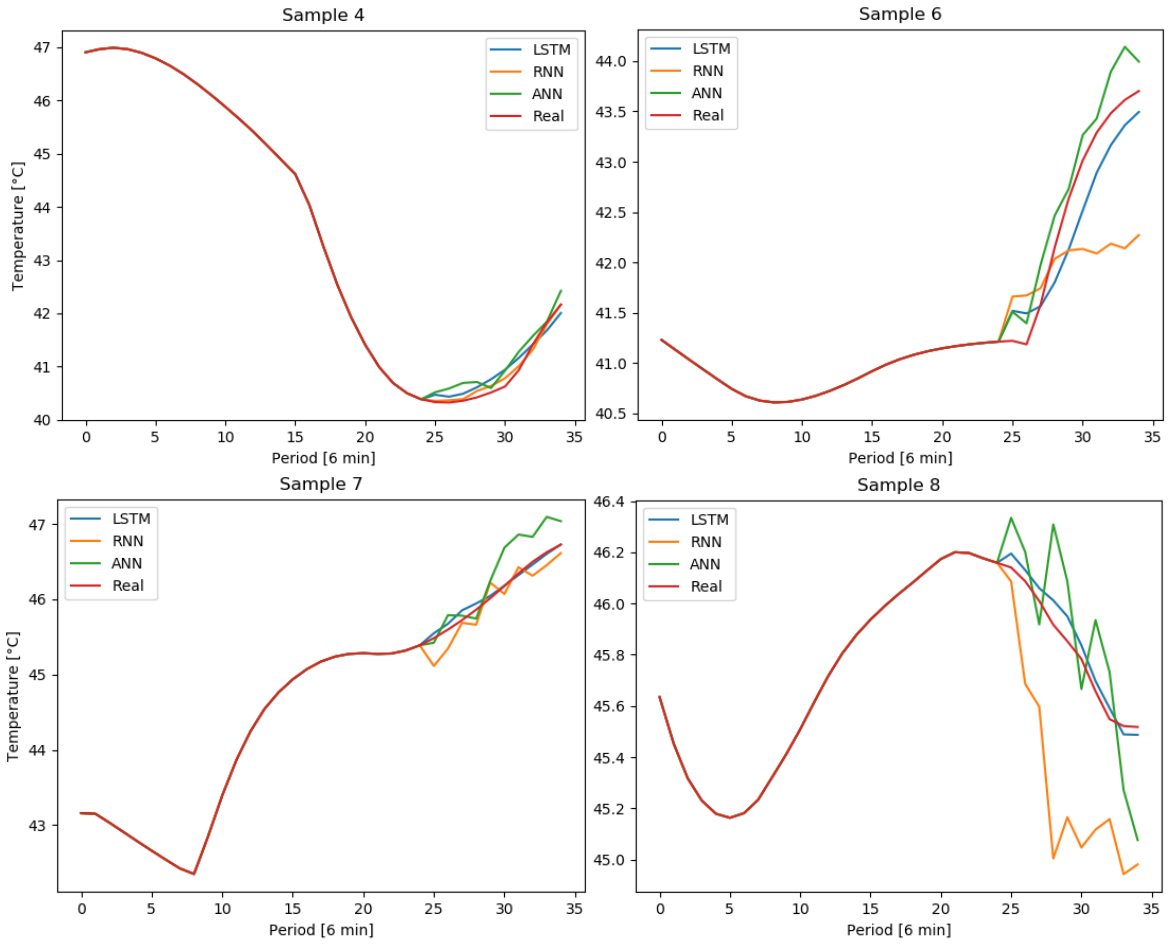


Figure 60: Reconstruction Samples for Average Tank Temperature.

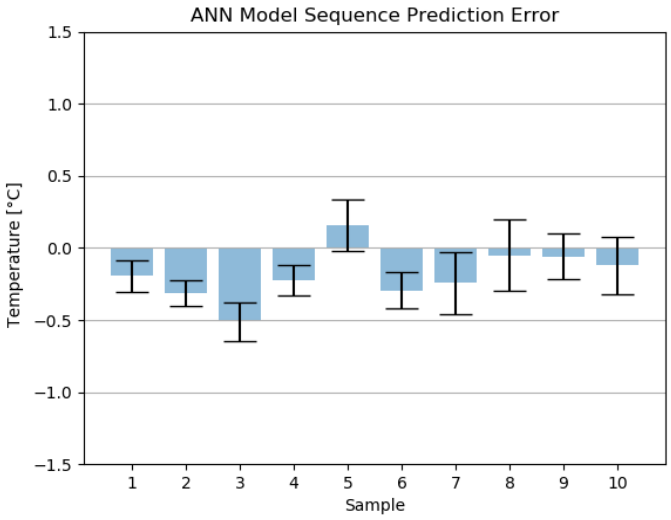


Figure 61: ANN Model Mean Sequence Prediction Error for Average Tank Temperature.

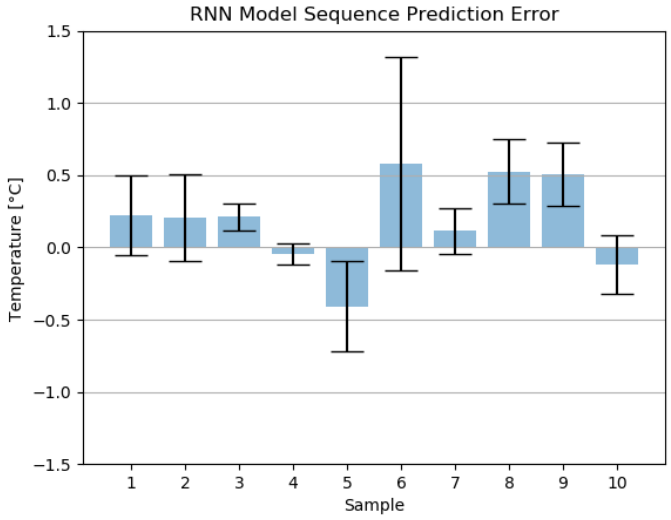


Figure 62: RNN Model Mean Sequence Prediction Error for Average Tank Temperature.

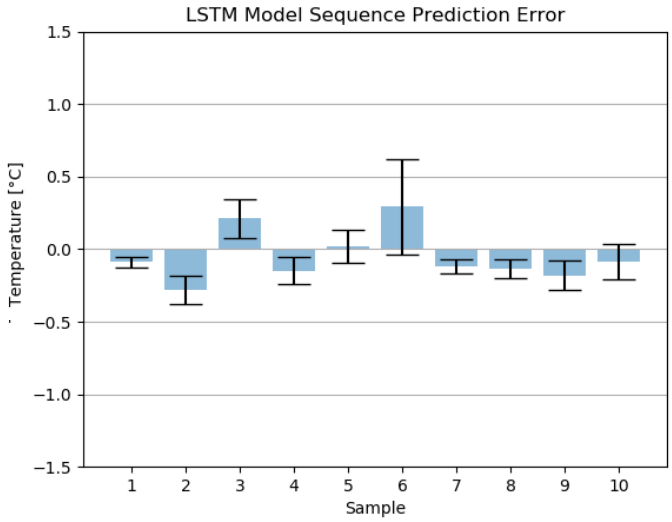


Figure 63: LSTM Model Mean Sequence Prediction Error for Average Tank Temperature.

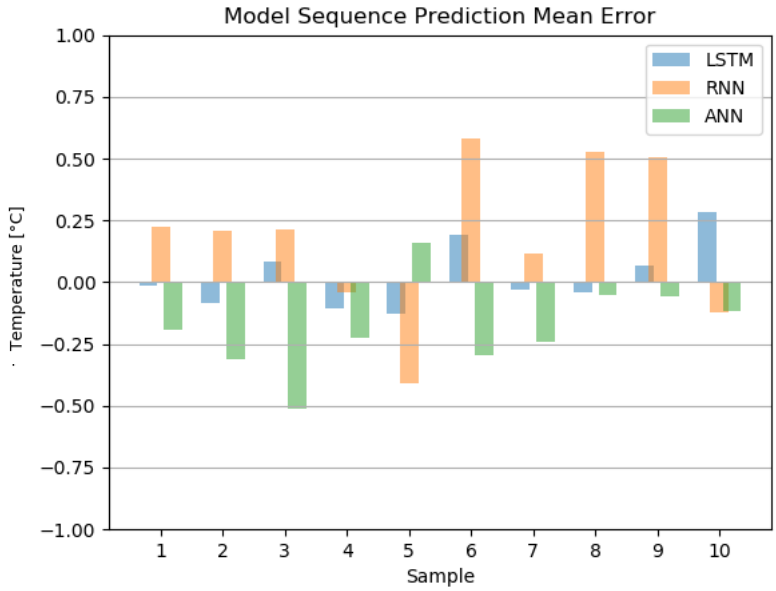


Figure 64: Comparison of relative Mean Prediction Errors for selected models for Average Tank Temperature.

7.3 Diagnosis Performance

Due to the natural variability of the solar collector outlet temperature, it is of more interest to detect anomalies in its behavior than with the average tank temperature. In the heat accumulators, the effect of the anomalous water-draw patterns and the collectors' efficiency is lessened by the constant heat input from the auxiliary sources. The DLSTM Ver 2 model, first shown in Table 30, is selected as it yielded the best results considering individual temperature prediction and sequence LSTM used.

As expressed above, not only a low value of the RMSE is used to compare the performance of the trained models, but also a small variance in relation to the mean error obtained. Under these criteria, the DLSTM Ver. 2 model is selected for further use in the anomaly detection framework. In Table 36 the selected hyperparameters are presented. A similar architecture is shown in Figure 45, in which two layers of LSTM cells are used together with an MLP to obtain the predicted value.

Table 36: Selected Deep Learning Architecture.

Hyperparameter	**Value**
Architecture	LSTM
Units & Layers	128-64
Activation Function	Tanh
Optimizer	RMSProp
Learning Rate	10^{-3}
Batch Size	64
Epochs	150
Length of time-window	3 days
MLP Hidden Units	512

A period of seven days must be simulated for the use of a 3 day-long model, such as the LSTM model selected. This is to ensure each time-window extracted is the same length, as shown in Figure 41. From the extraction of the last time-window before the induced fault, a total of 20 periods (or two hours) are predicted. Three cases are selected for analysis, during which the solar collector field was operational. These are shown in Table 37 and selected to represent different meteorological conditions from the available weather data.

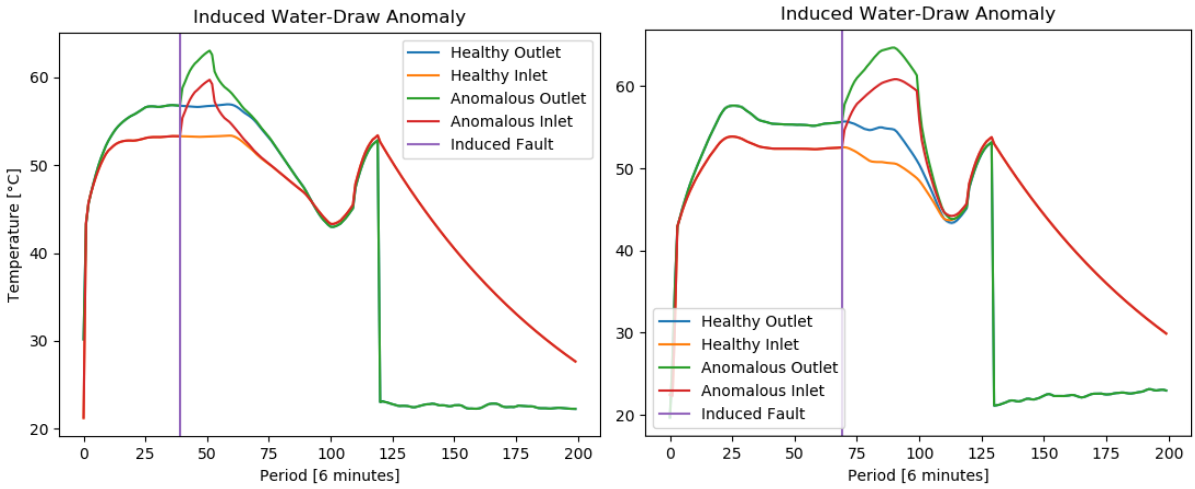
Table 37: Studied Weekly Profiles for Induced Anomalies.

Case	Start Hour	End Hour	Start Date	End Date	Induced Fault
1	2,377	2,545	April 9th	April 16th	2,440
2	4,321	4,489	June 29th	July 4th	4,405
3	5,832	6,000	August 30th	September 7th	5,919

7.3.1 Anomalous Water-Draw Profiles

The effect of the anomalous water-draw profiles for each of the three cases studied are shown in Figure 40 and Figure 65. These show the difference between healthy and anomalous temperature values caused by the induced fault. While the effect on temperature is almost instant, it endures for more than an hour until the solar control system suspends the operation of the pumps. The full sequences extracted for analysis are shown in Appendix I .

The generated healthy and anomalous sequences are fed to the LSTM model and then the prediction errors are analyzed. From the behavior of the reconstruction errors, the anomaly detection threshold is selected and finally the confusion matrix for each case are presented. A summary for the water-draw anomaly is presented in Table 38. The sequence reconstruction error is a direct comparison between the healthy prediction with the original healthy time-window. The same is done for the anomalous profile. For Figure 66, Figure 68 and Figure 70, the anomalous profile generated higher prediction errors, indicating that the model's latent representation of the healthy data is different enough to be recognized as an anomaly. Although both predictions obtain similar results with different threshold values, Case 1 presents the lowest overall errors, the threshold which yields best results is Case 3. Case 2, on the other hand occurs during mid-winter. As the energy delivered by the solar field low due to the limited radiation availability, the systems' temperatures are lower. Thus, the anomaly has a reduced effect on the system (as Figure 65 suggests) and so is the ability to identify anomalies.



(a) Case 2 (b) Case 3

Figure 65: Faulty Water-Draw Profiles: Solar Collector's Temperatures, Case 2 and 3.

The threshold is selected as the value of the anomalous predictions error's variance, as a representation of the dispersion within the data, considering that the healthy profile should have a smaller dispersion than the anomalous data. This allows the selection to be adjusted to each new input data and how the LSTM model recognized and replicates the sequences, rather than external analysis over the data. This, however, limits an instant response from the maintenance team, since

at least 5 measurements are needed to discriminate anomalous behavior. On the other hand, even though a two-hour window generated only 20 datapoints to detect anomalies, it is a practical measure considering that the online data recollection system is updated every minute. The details for each case are presented in the following sections.

Table 38: Summary of Anomaly Detection Results for Faulty Water-Draw Profiles.

Case	**1**	**2**	**3**
Threshold	1.817	3.662	3.866
True Positive	19	12	20
False Positive	1	8	0
True Negative	18	14	19
False Negative	2	6	1
Accuracy	0.925	0.650	0.975
Precision	0.940	0.630	1.000
Recall	0.900	0.700	0.950
F1 Score	0.923	0.667	0.974

I. Case 1

As the LSTM model uses previous predictions to construct the whole temperature sequence, it is expected that the reconstructions error increase with a longer horizon predicted. Figure 66 shows that the prediction error increases during the 20 future timesteps for the anomalous profiles, meanwhile the healthy reconstruction errors are consistently lower.

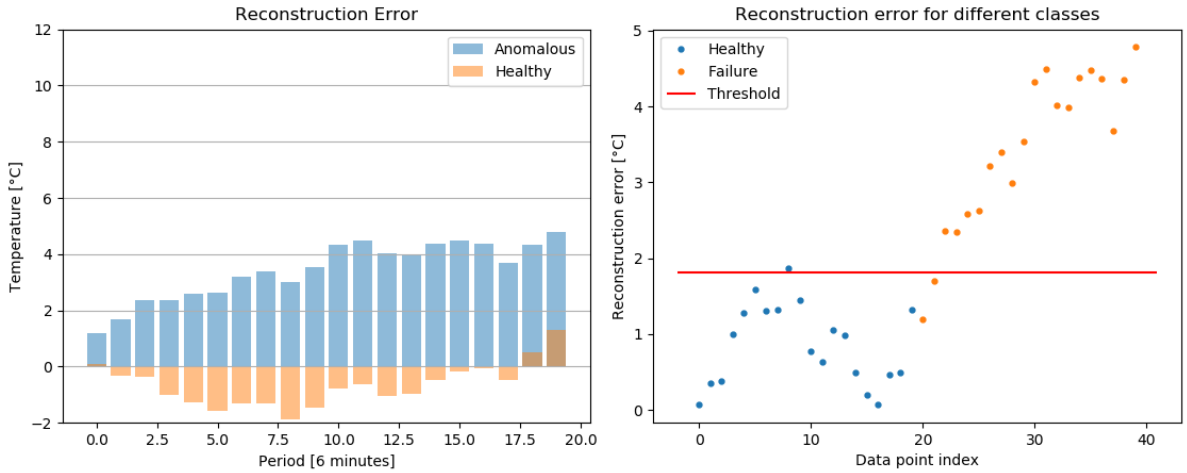


Figure 66: Faulty Water-Draw Profiles, Case 1: Prediction Error and Anomaly Detection Threshold Selection.

Additionally, the anomalous predictions tend to constantly overestimate the anomalous-induced temperatures. Healthy scenario yield errors with a peak of almost 2°C, meanwhile further future

anomalous values are more dispersed. This effect allows a defined separation of the healthy and anomalous data by selecting a threshold based on the variance of the reconstruction errors of the anomalous data. The results of the classification for this case are summarized in Figure 67, in which only 7.5% cases are misclassified.

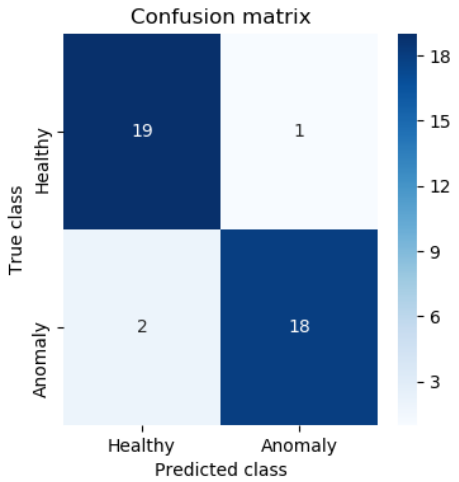


Figure 67: Faulty Water-Draw Profiles, Case 1: Confusion Matrix.

II. Case 2

In this case, the detection threshold does not generate a clear division as before. The prediction errors do not follow a trend as in case 1, and even the healthy errors surpass the anomalous generated during the last periods, in which the accumulated uncertainty of the prediction is higher. However, the magnitude difference remains for most part of the sequence, with the anomaly being overestimated in up to 10°C as shown in Figure 68.

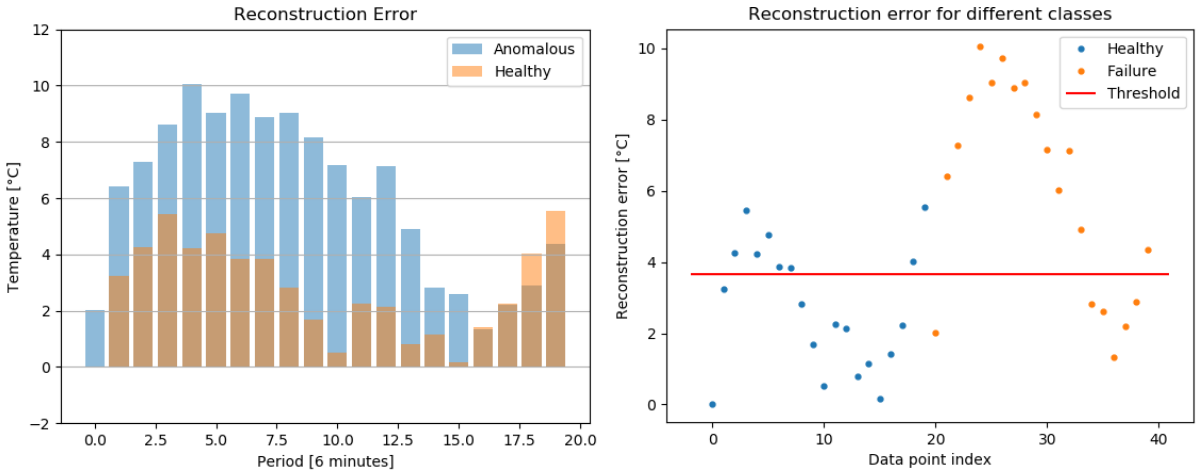


Figure 68: Faulty Water-Draw Profiles, Case 2: Prediction Error and Anomaly Detection Threshold Selection.

This leads to a higher threshold than before and a larger number of misclassified datapoints, however, the balance between correct and incorrect classifications is still favorable. The greater dispersion of healthy values and the model's incorporation of the anomalous data towards the end of the series (reducing the predictions errors) results in a lower performance for this case. Finally, the confusion matrix is presented in Figure 69.

The number of false positives is larger than the false negatives, an undesired increase of the number of false alarms. However, if the temporal order of the errors is analyzed, the selected threshold is able to detect the initial sequence (roughly 1:20 hours) of the anomalous data, while the healthy data significantly misclassified at the beginning of the series for 6 periods (36 minutes).

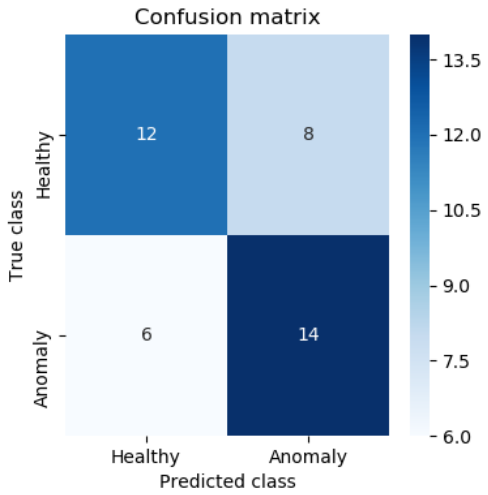


Figure 69: Faulty Water-Draw Profiles, Case 2: Confusion Matrix.

III. Case 3

The last case presented for the anomalous water-draw profile presents similar results to case 1, in which the threshold separates the healthy from the anomalous class as shown in Figure 70. However, the threshold is also larger, as in case 2, and both errors of the predicted values are over 4°C higher.

The confusion matrix obtained is shown in Figure 71, in which only one anomalous datapoint is misclassified. When comparing with Figure 70, it is observed that the value corresponds to the first anomalous output of the sequence, which again suggests that another criteria can be introduced, concerning the temporal behavior of the reconstruction errors can be useful for further anomaly detection.

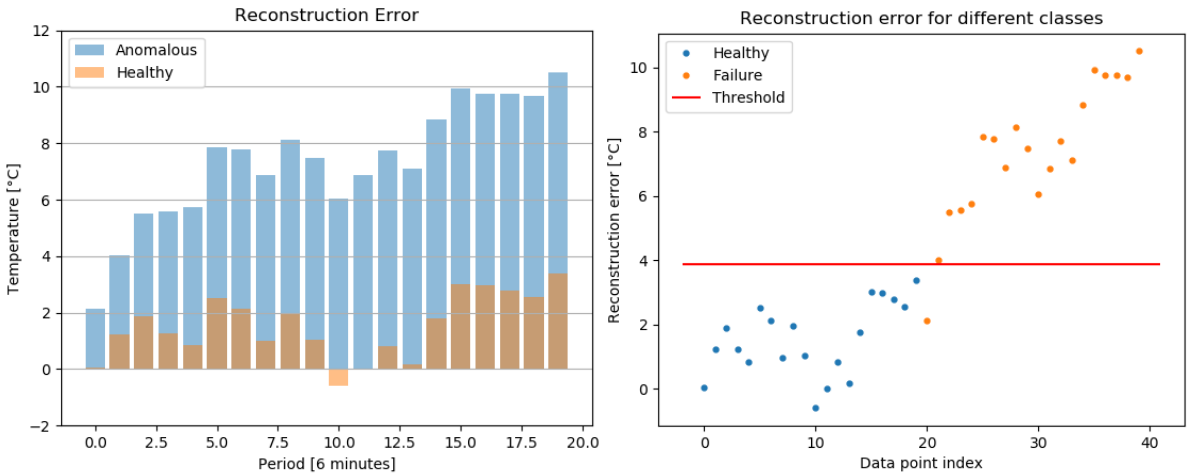


Figure 70: Faulty Water-Draw Profiles, Case 3: Prediction Error and Anomaly Detection Threshold Selection.

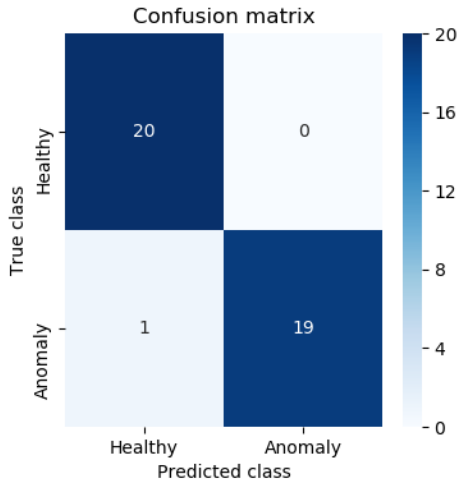


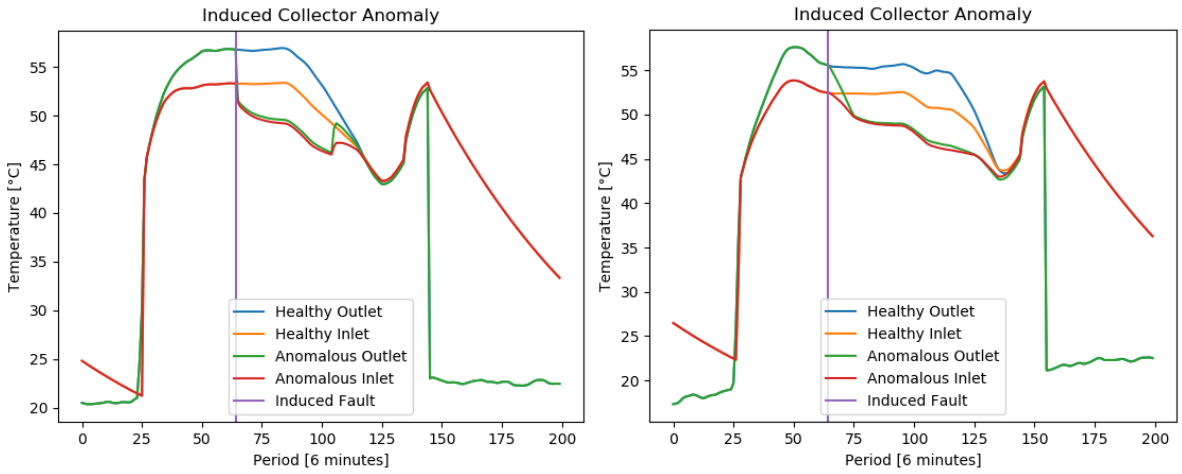
Figure 71: Faulty Water-Draw Profiles, Case 3: Confusion Matrix.

7.3.2 Solar Collector Efficiency Drop

Similarly, Figure 40 and Figure 72 show the anomalous temperature profiles generated by the reduction of the collector's efficiency for the three cases shown in Table 37. In this scenario, as the useful heat gains from the solar collectors is reduced, temperatures drop around 5°C within since the moment the fault is induced on the system. As aforementioned, the effect of the reduced absorbed radiation is effective until the solar control system suspends the circulation pumps. The full sequences extracted for this scenario are shown in Appendix J.

In Table 39 a summary of the performance of the LSTM-based model for anomaly detection is presented. For this anomaly scenario, the overall performance is lower, reaching a maximum accuracy of 0.85, and yields a high number of false negatives. The sequence reconstruction errors shown in Figure 73, Figure 75 and Figure 77 all present higher prediction errors for the healthy input data than the anomalous profile, overestimating the sequence of temperatures.

For the reason stated above, while the threshold is still based on the variance of the prediction's errors, the detection logic is inversed as the anomaly corresponds to lower temperatures than expected. Case 1 shows a better performance than the others, which are similar. However, Case 2 again yields the lowest score for every metric used. The results obtained also indicate that the reduction of the energy delivered by the solar system is an anomaly harder to capture and recognize from the characteristics learned by the model. This might be explained by the periodical oscillation of the temperature sequence, in which lower values are natural during certain hours of operation. From the perspective of the thermal model, the effect of the lower radiation absorbed is reduced by the constant heat influx from the heat-recovery chiller and the returning flow from the heat pump circuit. The details of each case are presented in the following sections.



(a) Case 2 (b) Case 3

Figure 72: Collector Efficiency Drop: Solar Collector's Temperatures, Case 2 and 3.

Table 39: Summary of Anomaly Detection Results for Collector Efficiency Drop Faults.

Case	**1**	**2**	**3**
Threshold	2.907	1.975	2.319
True Positive	20	14	15
False Positive	0	6	5
True Negative	13	10	12
False Negative	7	14	8
Accuracy	0.825	0.600	0.675
Precision	1.000	0.625	0.705
Recall	0.650	0.500	0.600
F1 Score	0.788	0.555	0.648

I. Case 1

As mentioned before, the behavior of the prediction errors is inverse when compared to the water-draw anomaly which elevates the system's temperature. This anomaly, however, causes the solar collector's outlet temperature to decrease. As the prediction errors shown in Figure 73 are higher for the healthy dataset than the anomalous one, it is useful to invert the logic for the selection of the detection threshold. If any datapoint is within the dispersion of the anomalous predictions, then it is recognized as an anomalous state. This again limits the time-window for early detection, requiring at least 4-6 periods to effectively obtain a representative threshold value.

Case 1 yields the best results for the detection of a drop in the collector's efficiency, although presenting a high number of false negatives as shown in Figure 74. This reduces the overall confidence of the detector, as the time required to correctly identify the fault is over 10 periods, an hour, since it manifested.

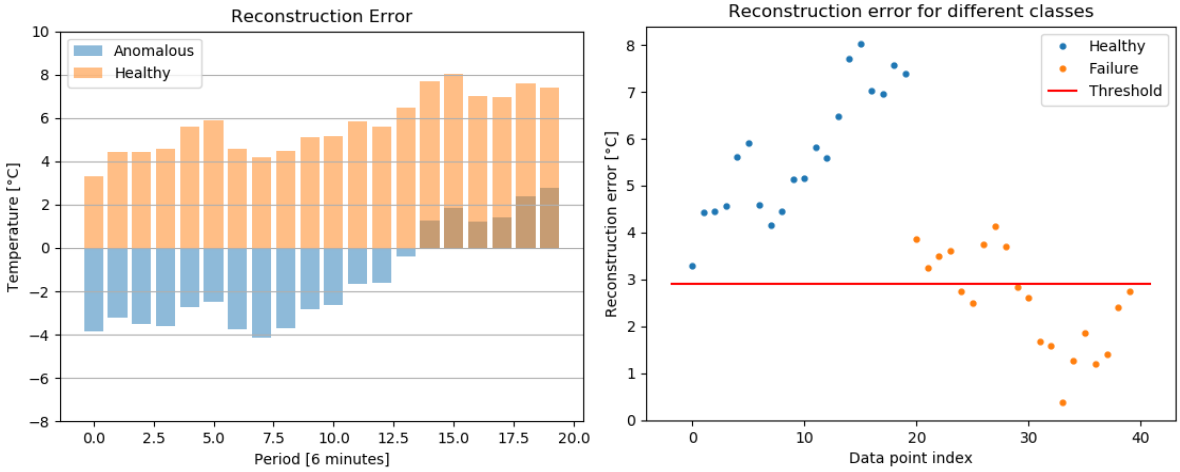


Figure 73: Collector Efficiency Drop, Case 1: Prediction Error and Anomaly Detection Threshold Selection.

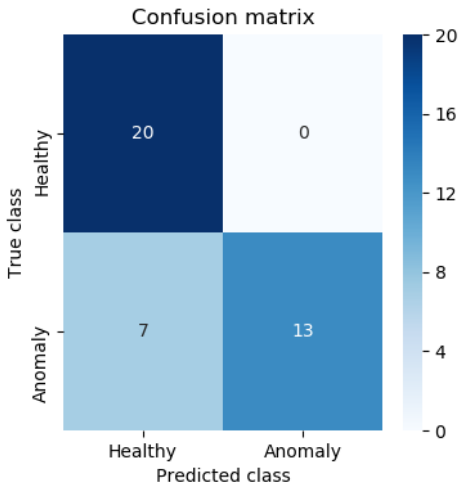


Figure 74: Collector Efficiency Drop, Case 1: Confusion Matrix.

II. Case 2

Figure 75 shows the predictions errors and the selected threshold for anomaly detection for the second case analyzed. An interesting behavior presented by the reconstructions errors is how the healthy and anomaly trends are inverted, overestimating and underestimating the values respectively with a maximum of ± 5°C. However, as a result, the dispersion of both datasets is higher, and the selected threshold is not able to differentiate between them adequately.

As shown in Figure 76, the number of false negatives and true negatives are equal, implying that identifying lower temperatures during winter period may require other features to accurately represent the internal behavior of the system.

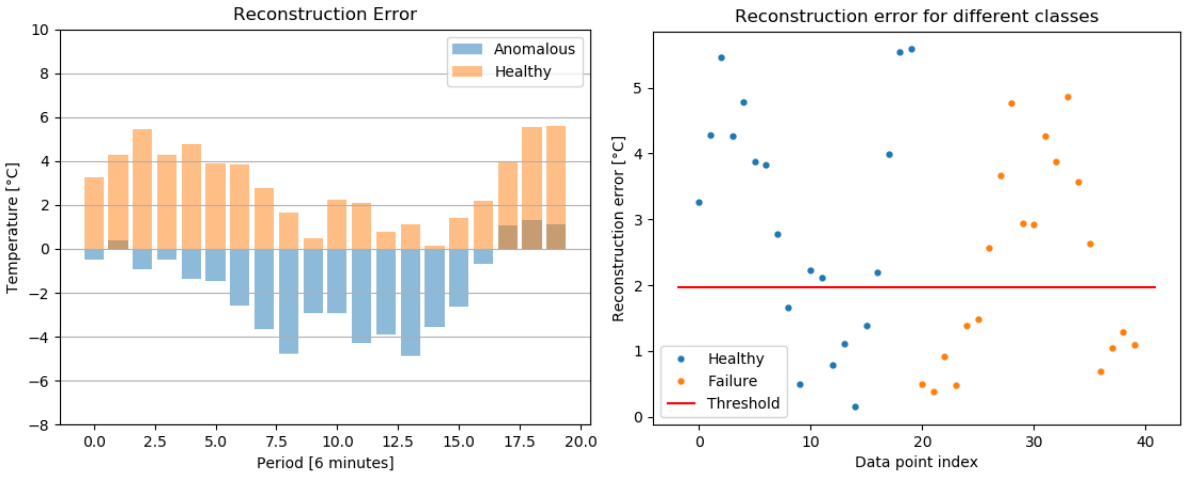


Figure 75: Collector Efficiency Drop, Case 2: Prediction Error and Anomaly Detection Threshold Selection.

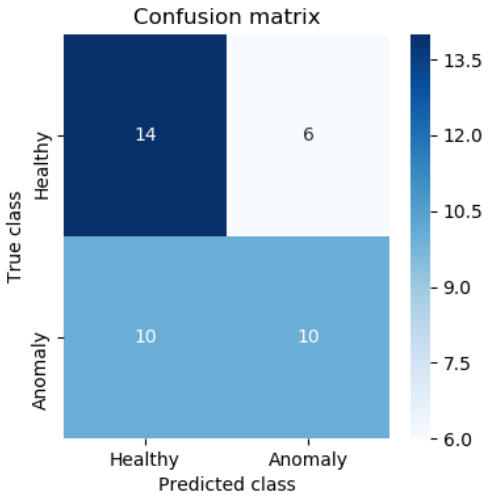


Figure 76: Collector Efficiency Drop, Case 2: Confusion Matrix.

III. Case 3

The third case presented also behaves differently than both cases presented before, as the prediction errors for the anomalous temperature profile are higher than the healthy prediction. Thus, from the samples presented, no common pattern may be identified, as the in the case of the water-draw anomaly. In fact, in this case the anomalous prediction error in Figure 77 increases with the number of timesteps, similar to the observed for the water-draw anomaly.

However, just as in Case 2, the dispersion of both errors surpasses the threshold, resulting in a high number of false negatives as shown in Figure 78. By analyzing the sequence of the reconstruction errors, this also suggests that the anomaly detection logic must include a temporal factor, as from an hour since the fault was induced, the health states are accurately classified.

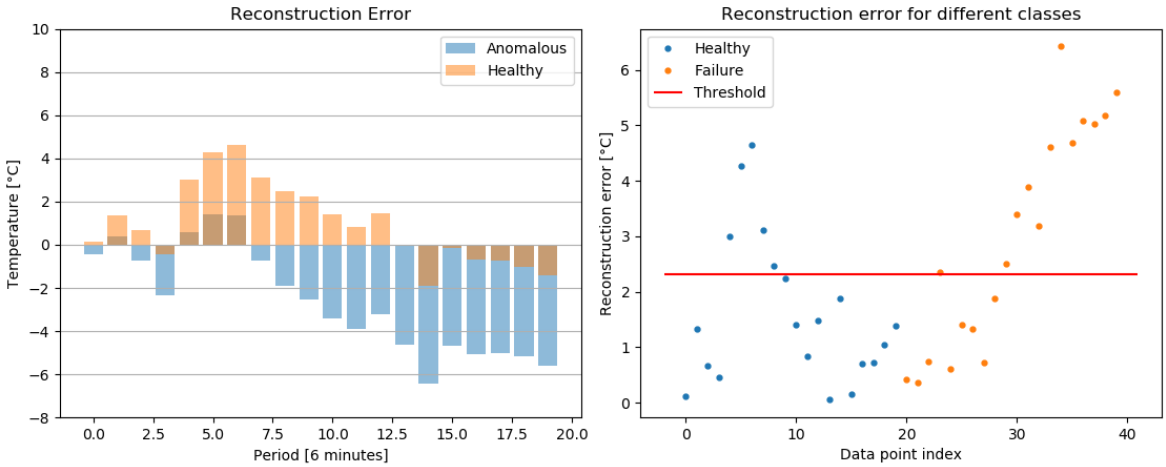


Figure 77: Collector Efficiency Drop, Case 3: Prediction Error and Anomaly Detection Threshold Selection.

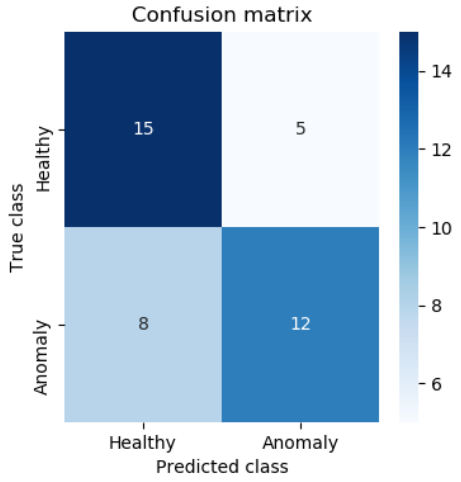


Figure 78: Collector Efficiency Drop, Case 3: Confusion Matrix.

7.3.3 Summary

Summarizing the results obtained for the anomaly detection under two different scenarios, Figure 79 shows the total predictions for the health state of the system. These results yield an overall precision of 83% and an accuracy of 76% as shown in Table 40.

However, as the detection model yields such different results for the anomaly caused by the faulty water-draw profile and the reduction of the collector's efficiency, a deeper analysis is required to interpret these differences. This is crucial to develop a robust framework which can be implemented for real-time monitoring and prognosis of the SHW system. The biggest setback for this framework is the fact that it is built on simulated data which has not been experimentally validated. This, coupled with the restrictions on the control system and the ideal conditions of the auxiliary heat inputs, reduces the number of hours in which the solar field is operational. Consequently, the LSTM is unnecessarily biased towards lower temperatures, which could explain the unfavorable results obtained for this scenario.

The lower accuracy achieved for the reduced temperature scenario also limits the identification of faults within the collector, while the detection of high temperatures allows the detection of anomalies occurring upstream of the solar field. Considering the presence of auxiliary heat sources, the second case is of greater importance for the analysis of the thermal system as a whole, since the occurence of high temperature events can lead to damage in other components, particularly in the circulation pumps.

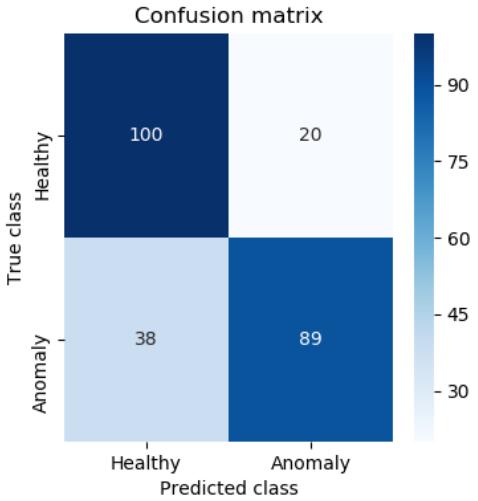


Figure 79: Summary of Anomaly Detection results for both scenarios.

Table 40: Summary of Anomaly Detection classification metrics.

Precision	Recall	Accuracy	F1
0.83	0.72	0.76	0.77

Chapter 8

Conclusions and Comments

The presented work seeks to develop an anomaly detection framework for a solar thermal system, based on synthetic data generated from a specialized software and deep learning techniques. As such, the development of the thesis has been divided into three main tasks: the construction of a thermal model in TRNSYS software based on design conditions, the training of a deep learning model capable of extracting abstract temporal relationships from the data, and finally, the use of this trained model to predict temperature values and subsequently classify the health state of the system.

First, the model is built in TRNSYS software. It incorporates all the mayor thermal and hydraulic components of the system, based on the information available of the design conditions. However, integration and maintenance issues do not allow the use of experimental measurements to validate the results yield by the software. Still, from the analysis carried out of the system and the available measured values, malfunctions in instrumentation and the control system are identified. A series of simplifications and suppositions about the internal logic and certain components are required. This limits the representability of the thermal model for components related specifically to the heat-recovery chiller and the heat pumps.

Secondly, for the assessment of deep learning techniques to predict future values based on the generated temperature and control signal profiles, different models and hyperparameters are explored. The performance of three main models under similar conditions is compared, namely: ANN, RNN and LSTM. These models are tested for similar hyperparameters. For ANN and RNN the use of the whole previous day as input data is enough to predict the following timesteps' output accurately. However, LSTM requires longer sequences to reach comparable results. The improvement of the LSTM's performance with a three day-length temporal matrix reaches an RMSE below 1°C, but the difference is not significant to when a seven-day sequence is analyzed. Other lengths of time-windows can yield better interpretations of the temporal sequences, such as a five-day time-window representing the labor week period which matches the higher demand for hot water in the system. The use of a time-window coherent with the demand pattern may yield a better latent representation of the data, and subsequently, more accurate and precise predictions.

The precision and accuracy of the predictions for single and sequence of temperature values is carried out in two points of the system: the solar field's outlet temperature and the average hot water tank temperature. Although RNN models obtain a lower RMSE for single temperature predictions, its performance decreases when faced with the sequence prediction. Opposed to this behavior, the LSTM's performance is better at replicating sequences than both other models, obtaining the lowest variance of the predicted values as well.

The results obtained under the 2°C error threshold can only be corroborated with experimental measurements and the quantification of the instrumentation's uncertainty, as the TRNSYS simulation is not stochastic in nature. The RMSE values obtained for the average tank temperature are lower for all the three trained models when compared to the solar collector outlet temperature predictions. As mentioned before, this is expected due to the small variability experimented by the tank temperature during operation, due to the constant input of heat from the auxiliary sources. As the natural variations cannot be accounted for, this temperature is not considered for the anomaly detection process. The DLSTM Ver 2 design is selected as the best performing model for sequence predictions and is used in the last task.

Only operational hours of the solar pumps are used for sample selection. The major drawback of the induced faults simulated in TRNSYS software is the fact that their effect must be significative on the system's temperatures to be classified as anomalous. This means that faults developed over time, such as dust accumulation or changes in consumption patterns, may be overlooked, as well as faults developed during non-operating hours. With more information available, anomaly detection methods could be more adequate to identify subtle changes in heat production.

The proposed method to identify anomalies consists on using the model's prediction values variance as threshold to classify health states of the system. This yields accurate results for the water-draw anomalous behavior, in which the system's temperatures rise over expected trends. This scenario is recognized with a mean accuracy of 85%. Meanwhile, the detection of the anomalous heat underproduction by the solar field reaches a 70% of accuracy. The restrictions of the physical representation of the TRNSYS model limits the applicability of this framework. Decreasing the water-draw results in temperatures rising over 10°C, while the reduction of the useful heat gained by the collectors only lowers expected temperatures by 5°C. Additionally, lower temperatures are more frequent within the simulated values due to the daily trend of the solar resource, naturally decreasing when the radiation is reduced. This makes the second scenario not as easily recognizable as anomalous behavior. As mentioned before, this is due to the representation of the data in the latent space obtained from the feature and time-window initial selection. An overall precision of 83% is obtained for the anomaly detection framework. From this previous analysis, the idea to incorporate temporal criteria to identify anomalous behavior arises. It is also noted that at least five predictions are required to effectively classify developing faults in the system. The obtained results for the anomaly detection framework heavily depend on the representativity of the thermal model, and for further developments, experimental measurements must be carried out.

In previous literature, the frequent use of ANN frameworks for performance prediction and anomaly detection in thermal systems, raises the question as to why other more complex architectures have not been experimented with. By comparing ANN with RNN and LSTM under similar conditions, each model presents strengths and weaknesses regarding the required data, training process and performance. In this work, the trained LSTM model yields both more accurate predictions than ANN and more precise values than RNN for temperature sequence predictions. Ultimately, this leads to believe that excellent results can be achieved with further experimentation and calibration of the presented framework.

8.1 Future Work

The obtained results show the strengths and shortcomings of an initial approach for anomaly detection in a SHW system. The lack of experimental validation weighs heavily on the results and methods chosen for this framework, and thus it is a pending challenge to validate the TRNSYS simulation to gain greater representativity of the thermal system. Additionally, the quantification and identification of the measurements' and simulation's uncertainty thresholds is fundamental to set a target precision to be reached by the trained Deep Learning models.

Another aspect to be improved is the preprocessing of the temperature sequences and the generation of a representative latent space of the selected features. As other variables in the system become monitored with the installation of better equipment, such as the flowmeter recently installed at the outlet of the solar field, other methods can be applied to obtain a better multisensorial representation of the data, such as CNN. As mentioned above, the use of a time-window coherent with the demand profile might also help achieve a better interpretation of the relationships within the measured data. To seek even greater flexibility, unsupervised methods, such as the use of auto-encoders might yield more useful results for the detection of abnormal behavior in the analyzed systems.

Finally, other data-based metrics can be explored for anomaly and fault detection in systems. By quantifying the uncertainty of the experimental measurements, the simulation's results and the model's predictions, other methods can be implemented to determine the health state of the system.

Glossary

AE	-	Auto-Encoders.
ANFIS	-	Adaptive-Network-Based Fuzzy Inference System.
ANN	-	Artificial Neural Network.
ART	-	Adaptive Resonance Theory Neural Network.
BPTT	-	Backpropagation Through Time.
BRNN	-	Bidirectional Recurrent Neural Network.
CART	-	Classification and Regression Decision Tree.
CERN	-	Conseil Européen pour la Recherche Nucléaire.
CNN	-	Convolutional Neural Network.
DDA	-	Data-Driven Approaches.
DH	-	District Heating.
DL	-	Deep Learning.
DLSTM	-	Deep Long Short-Term Memory Recurrent Neural Network.
DNN	-	Deep Neural Network.
DT	-	Decision Tree.
ETC	-	Evacuated Tube Collector.
FDS	-	Fault Diagnostic System.
FFNN	-	Feed-Forward Neural Network.
FPC	-	Flat Plate Collector.
GA	-	Genetic Algorithm.
GP	-	Gaussian Process.
GPU	-	Graphics Processing Unit.
IAM	-	Incidence Angle Modifier.
LSTM	-	Long-Short Term Memory Recurrent Neural Network.
ML	-	Machine Learning.
MLP	-	Multi-Layer Perceptron.
MSE	-	Mean Square Error.
NLP	-	Natural Language Processing.
NN	-	Neural Network.
OBE	-	Outer Bounding Ellipsoid algorithm.
PCA	-	Principal Component Analysis.
PHM	-	Prognostics and Health Management.
PV	-	Solar Photovoltaic technology.
SVM	-	Support Vector Machine.
RMSE	-	Root Mean Squared Error.
RBF	-	Radial Basis Function.
RNN	-	Recurrent Neural Network.
RUL	-	Remaining Useful Life.

SHW	-	Solar Hot Water system.
TBPTT	-	Truncated Backpropagation Through Time.
TRNSYS	-	Transient System Simulation program.
VRF	-	Variable Refrigerant Flow systems.
WGET-SHW	-	Water-In-Glass Evacuated Tube Solar Hot Water system.

Bibliography

[1] I. Renewable and E. Agency, *Opportunities to accelerate national energy transitions through advanced deployment of renewables*, no. November. 2018.

[2] S. V. H. Olivier Janssens, Viktor Slavkovikj, Bram Vervisch, Kurt Stockman, Mia Loccufier, Steven Verstockt, Rik Van de Walle, "Convolutional Neural Network Based Fault Detection for Rotating Machinery," *Journal of Sound and Vibration*, vol. 377. pp. 331–345, 2016.

[3] Y. Liu, G. Zhao, X. Peng, and C. Hu, "Lithium-ion Battery Remaining Useful Life Prediction with Long Short-term Memory Recurrent Neural Network," no. 1, pp. 1–7, 2016.

[4] International Energy Agency, "Statistics Key Renewables Trends," 2016.

[5] International Energy Agency, "Electricity Information 2018: Overview," 2018.

[6] International Energy Agency, "Market Report Series Renewables 2O17," 2017.

[7] Ministerio de Energía and Gobierno de Chile, "Fomento Sistemas Solares Térmicos," 2018. [Online]. Available: http://www.minenergia.cl/sst. [Accessed: 12-Oct-2018].

[8] H. He, T. P. Caudell, D. F. Menicucci, and A. A. Mammoli, "Application of Adaptive Resonance Theory neural networks to monitor solar hot water systems and detect existing or developing faults," *Sol. Energy*, vol. 86, no. 9, pp. 2318–2333, 2012.

[9] C. De Keizer, S. Kuethe, U. Jordan, and K. Vajen, "Simulation-based long-term fault detection for solar thermal systems," *Sol. Energy*, vol. 93, pp. 109–120, 2013.

[10] G. Niu, *Data-Driven Technology for Engineering Systems Health Management*. 2017.

[11] S. Khan and T. Yairi, "A review on the application of deep learning in system health management," *Mech. Syst. Signal Process.*, vol. 107, pp. 241–265, 2018.

[12] S. Kalogirou, S. Lalot, G. Florides, and B. Desmet, "Development of a neural network-based fault diagnostic system for solar thermal applications," *Sol. Energy*, vol. 82, no. 2, 2008.

[13] N. Zimmerman, E. Dahlquist, and K. Kyprianidis, "Towards On-line Fault Detection and Diagnostics in District Heating Systems," in *Energy Procedia*, 2017, vol. 105, pp. 1960–1966.

[14] Y. Lei, N. Li, L. Guo, N. Li, T. Yan, and J. Lin, "Machinery health prognostics: A systematic review from data acquisition to RUL prediction," *Mech. Syst. Signal Process.*, vol. 104, pp. 799–834, 2018.

[15] R. Zhao, R. Yan, J. Wang, and K. Mao, "Learning to monitor machine health with convolutional Bi-directional LSTM networks," *Sensors (Switzerland)*, vol. 17, no. 2, 2017.

[16] University of Wisconsin--Madison. Solar Energy Laboratory., "TRNSYS, a transient simulation program." Madison, Wis. :The Laboratory, 1975.

[17] M. Abadi *et al.*, "TensorFlow : A System for Large-Scale Machine Learning This paper is included in the Proceedings of the TensorFlow : A system for large-scale machine

learning," *Proc 12th USENIX Conf. Oper. Syst. Des. Implement.*, 2016.

[18] A. Jamar, Z. A. A. Majid, W. H. Azmi, M. Norhafana, and A. A. Razak, "A review of water heating system for solar energy applications," *Int. Commun. Heat Mass Transf.*, vol. 76, pp. 178–187, 2016.

[19] L. M. Ayompe, A. Duffy, M. Mc Keever, M. Conlon, and S. J. McCormack, "Comparative field performance study of flat plate and heat pipe evacuated tube collectors (ETCs) for domestic water heating systems in a temperate climate," *Energy*, vol. 36, no. 5, pp. 3370–3378, May 2011.

[20] John A Duffie and W. A. Beckman, *Solar Engineering of Thermal Processes Solar Engineering*, Fourth. Hoboken, New Jersey: John Wiley & Sons, Inc., 2013.

[21] S. Kalogirou, "The potential of solar industrial process heat applications," *Appl. Energy*, vol. 76, no. 4, pp. 337–361, Dec. 2003.

[22] European Solar Thermal Industry Federation, "QAiST: Quality Assurance in Solar Heating and Cooling Technology A guide to the standard EN 12975," *Intelligent Energy Europe*. [Online]. Available: http://qaist.org. [Accessed: 20-Aug-2004].

[23] S. Kalogirou, *Solar Energy Engineering Processes and Systems*, Second. Elsevier Inc., 2009.

[24] R. K. Mishra, V. Garg, and G. N. Tiwari, "Thermal modeling and development of characteristic equations of evacuated tubular collector (ETC)," *Sol. Energy*, vol. 116, pp. 165–176, 2015.

[25] M. A. Sabiha, R. Saidur, S. Mekhilef, and O. Mahian, "Progress and latest developments of evacuated tube solar collectors," *Renew. Sustain. Energy Rev.*, vol. 51, pp. 1038–1054, 2015.

[26] A. C. de Keizer, K. Vajen, and U. Jordan, "Review of long-term fault detection approaches in solar thermal systems," *Sol. Energy*, vol. 85, no. 7, pp. 1430–1439, 2011.

[27] H. K. Ghritlahre and R. K. Prasad, "Application of ANN technique to predict the performance of solar collector systems - A review," *Renew. Sustain. Energy Rev.*, vol. 84, no. January, pp. 75–88, 2018.

[28] D. E. RUMELHART, G. E. HINTON, R. J. WILLIAMS, G. E. HINTON, and R. J. WILLIAMS, "Learning Internal Representations by Error Propagation," in *Readings in Cognitive Science*, Elsevier, 1988, pp. 399–421.

[29] H. Wang and B. Raj, "On the Origin of Deep Learning," pp. 1–72, 2017.

[30] A. Alzahrani, P. Shamsi, C. Dagli, and M. Ferdowsi, "Solar Irradiance Forecasting Using Deep Neural Networks," *Procedia Comput. Sci.*, vol. 114, pp. 304–313, 2017.

[31] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Medical Image Analysis*, vol. 42. Elsevier, pp. 60–88, 01-Dec-2017.

[32] J. Schmidhuber, "Deep Learning in neural networks: An overview," *Neural Networks*, vol. 61, pp. 85–117, 2015.

[33] Z. C. Lipton, J. Berkowitz, and C. Elkan, "A Critical Review of Recurrent Neural

Networks for Sequence Learning," pp. 1–38, 2015.

[34] P. J. Werbos, "Backpropagation Through Time : What It Does and How to Do It," vol. 78, no. October, pp. 1550–1560, 1990.

[35] Y. Bengio, P. Simard, and P. Frasconi, "Learning Long Term Dependencies with Gradient Descent is Difficult," *IEEE Trans. Neural Networks*, vol. 5, no. 2, pp. 157–166, 1994.

[36] S. Hochreiter and J. Urgen Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1735–1780, 1997.

[37] A. Ogawa and T. Hori, "Error detection and accuracy estimation in automatic speech recognition using deep bidirectional recurrent neural networks," *Speech Commun.*, vol. 89, pp. 70–83, May 2017.

[38] S. Zheng, K. Ristovski, A. Farahat, and C. Gupta, "Long Short-Term Memory Network for Remaining Useful Life estimation," *2017 IEEE Int. Conf. Progn. Heal. Manag.*, pp. 88–95, 2017.

[39] S. Hochreiter and J. J. Schmidhuber, "Long Short-Term Memory," *Neural Comput.*, vol. 9, no. 8, pp. 1–32, 1997.

[40] Y. Wu, M. Yuan, S. Dong, L. Lin, and Y. Liu, "Remaining useful life estimation of engineered systems using vanilla LSTM neural networks," *Neurocomputing*, vol. 275, pp. 167–179, 2018.

[41] M. Schuster and K. K. Paliwal, "Bidirectional recurrent neural networks," *IEEE Trans. Signal Process.*, vol. 45, no. 11, pp. 2673–2681, 1997.

[42] R. Zhao, J. Wang, R. Yan, and K. Mao, "Machine health monitoring with LSTM networks," *Proc. Int. Conf. Sens. Technol. ICST*, 2016.

[43] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A Simple Way to Prevent Neural Networks from Overfitting," *J. Mach. Learn. Res.*, vol. 15, pp. 1929–1958, 2014.

[44] S. A. Kalogirou, S. Panteliou, and A. Dentsoras, "Artificial neural networks used for the performance prediction of a thermosiphon solar water heater," *Renew. Energy*, vol. 18, no. 1, pp. 87–99, 1999.

[45] W. Yaïci and E. Entchev, "Performance prediction of a solar thermal energy system using artificial neural networks," *Appl. Therm. Eng.*, vol. 73, no. 1, pp. 1348–1359, Dec. 2014.

[46] E. Dikmen, M. Ayaz, H. H. Ezen, E. U. Küçüksille, and A. Ş. Şahin, "Estimation and optimization of thermal performance of evacuated tube solar collector system," *Heat Mass Transf. und Stoffuebertragung*, vol. 50, no. 5, pp. 711–719, 2014.

[47] S. A. Kalogirou, "Prediction of flat-plate collector performance parameters using artificial neural networks," *Sol. Energy*, vol. 80, no. 3, pp. 248–259, Mar. 2006.

[48] C. Cetiner, F. Halici, H. Cacur, and I. Taymaz, "Generating hot water by solar energy and application of neural network," *Appl. Therm. Eng.*, vol. 25, no. 8–9, pp. 1337–1348, Jun. 2005.

[49] M. Souliotis, S. Kalogirou, and Y. Tripanagnostopoulos, "Modelling of an ICS solar water

heater using artificial neural networks and TRNSYS," *Renew. Energy*, vol. 34, no. 5, pp. 1333–1339, May 2009.

[50] M. Mohanraj, S. Jayaraj, and C. Muraleedharan, "Modeling of a direct expansion solar assisted heat pump using artificial neural networks," *Int. J. Green Energy*, vol. 5, no. 6, pp. 520–532, 2008.

[51] Y. Varol, A. Koca, H. F. Oztop, and E. Avci, "Forecasting of thermal energy storage performance of Phase Change Material in a solar collector using soft computing techniques," *Expert Syst. Appl.*, vol. 37, no. 4, pp. 2724–2732, Apr. 2010.

[52] H. Li, Z. Liu, K. Liu, and Z. Zhang, "Predictive Power of Machine Learning for Optimizing Solar Water Heater Performance: The Potential Application of High-Throughput Screening," *Int. J. Photoenergy*, vol. 2017, 2017.

[53] S. Srivastava and S. Lessmann, "A comparative study of LSTM neural networks in forecasting day-ahead global horizontal irradiance with satellite data," *Sol. Energy*, vol. 162, no. August 2017, pp. 232–247, 2018.

[54] S. Leva, A. Dolara, F. Grimaccia, M. Mussetta, and E. Ogliari, "Analysis and validation of 24 hours ahead neural network forecasting of photovoltaic output power," *Math. Comput. Simul.*, vol. 131, pp. 88–100, 2017.

[55] S. Rodat, C. Tantolin, X. Le Pivert, and S. Lespinats, "Daily forecast of solar thermal energy production for heat storage management," *Journal of Cleaner Production*, vol. 139. pp. 86–98, 2016.

[56] A. Picallo-Perez, J. M. Sala-Lizarraga, E. Iribar-Solabarrieta, M. Odriozola-Maritorena, and L. Portillo-Valdés, "Application of the malfunction thermoeconomic diagnosis to a dynamic heating and DHW facility for fault detection," *Energy Build.*, vol. 135, pp. 385–397, 2017.

[57] P. Fu, N. Wang, L. Wang, T. Morosuk, Y. Yang, and G. Tsatsaronis, "Performance degradation diagnosis of thermal power plants: A method based on advanced exergy analysis," *Energy Convers. Manag.*, vol. 130, pp. 219–229, 2016.

[58] J. Korbicz, J. M. Koscienlny, and Z. Kowalczuk, *Fault Diagnosis Models, Artificial Intelligence, Appliction Vo. 1*, 1st ed. Berlin: Springer-Verlag Berlin Heidelberg GmbH, 2004.

[59] G. Li *et al.*, "An improved decision tree-based fault diagnosis method for practical variable refrigerant flow system using virtual sensor-based fault indicators," *Appl. Therm. Eng.*, vol. 129, pp. 1292–1303, 2018.

[60] M. Wielgosz, A. Skoczeń, and M. Mertik, "Using LSTM recurrent neural networks for monitoring the LHC superconducting magnets," *Nucl. Instruments Methods Phys. Res. Sect. A Accel. Spectrometers, Detect. Assoc. Equip.*, vol. 867, pp. 40–50, 2017.

[61] H. Liu, J. Zhou, Y. Zheng, W. Jiang, and Y. Zhang, "Fault diagnosis of rolling bearings with recurrent neural network-based autoencoders," *ISA Trans.*, pp. 1–12, 2018.

[62] Z. Fengming, L. Shufang, G. Zhimin, W. Bo, T. Shiming, and P. Mingming, "Anomaly detection in smart grid based on encoder-decoder framework with recurrent neural network," *J. China Univ. Posts Telecommun.*, vol. 24, no. 6, pp. 67–73, 2017.

[63] M. Mrugalski, M. Witczak, and J. Korbicz, “Confidence estimation of the multi-layer perceptron and its application in fault detection systems,” *Eng. Appl. Artif. Intell.*, vol. 21, no. 6, pp. 895–906, Sep. 2008.

[64] Y. Gal and Z. Ghahramani, “Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning,” vol. 14, no. 1, pp. 145–56, Jun. 2015.

[65] J. Burch and C. Christensen, “Towards Development of an Algorithm for Mains Water Temperature,” *InterSolar 2007 Conf.*, pp. 5–10, 2007.

[66] Sulzer Pumps, *Centrifugal Pump Handbook - Index*. 2010.

[67] D. P. Kingma and J. Ba, “Adam: A Method for Stochastic Optimization,” *AIP Conf. Proc.*, vol. 1631, pp. 58–62, Dec. 2014.

[68] T. Tijmen and H. Geoffrey E., “Lecture 6.5- RMSprop: Divide the gradient by a running average of its recent magnitude,” *COURSERA: Neural Netw Mach Learn 4:2*, 2012. .

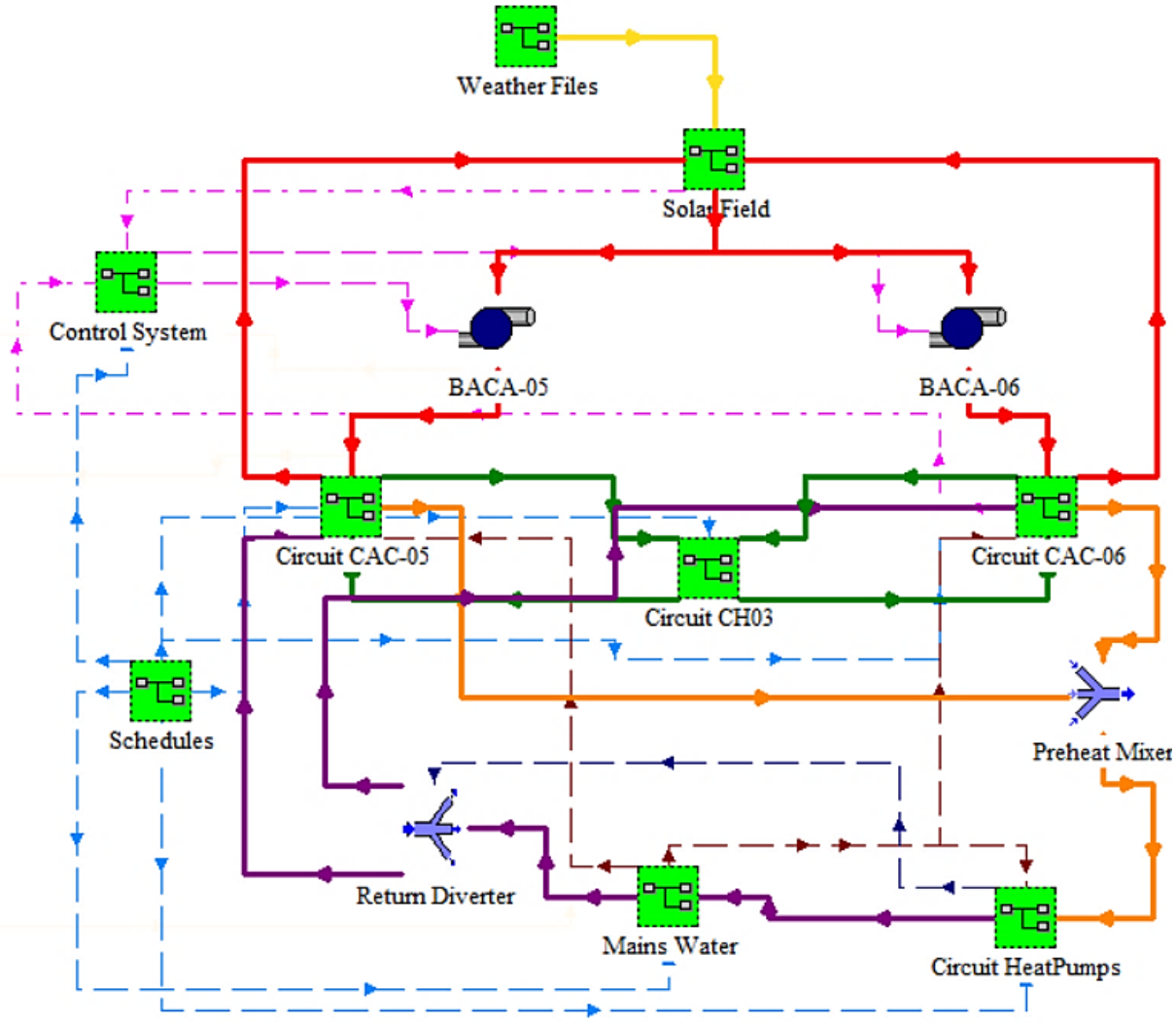
[69] N. Reimers and I. Gurevych, “Optimal Hyperparameters for Deep LSTM-Networks for Sequence Labeling Tasks,” 2017.

[70] F. M. Bianchi, E. Maiorino, M. C. Kampffmeyer, A. Rizzi, and R. Jenssen, *An overview and comparative analysis of Recurrent Neural Networks for Short Term Load Forecasting*. 2017.

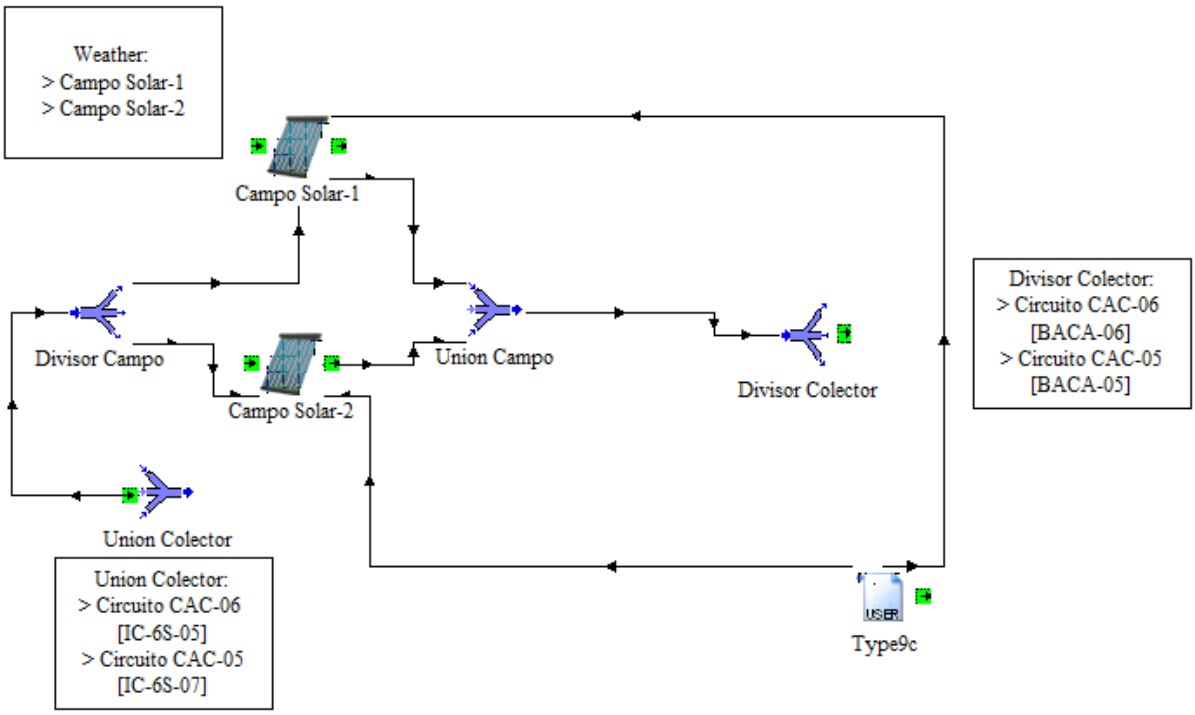
[71] S. Ruder, “An overview of gradient descent optimization algorithms,” pp. 1–14, 2016.

Appendix

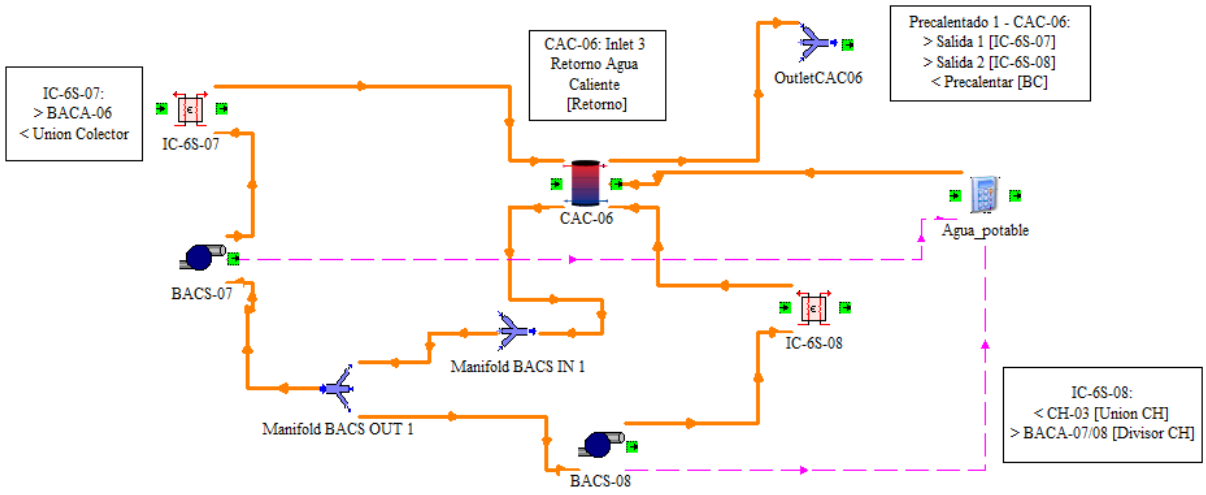
A. TRNSYS Model



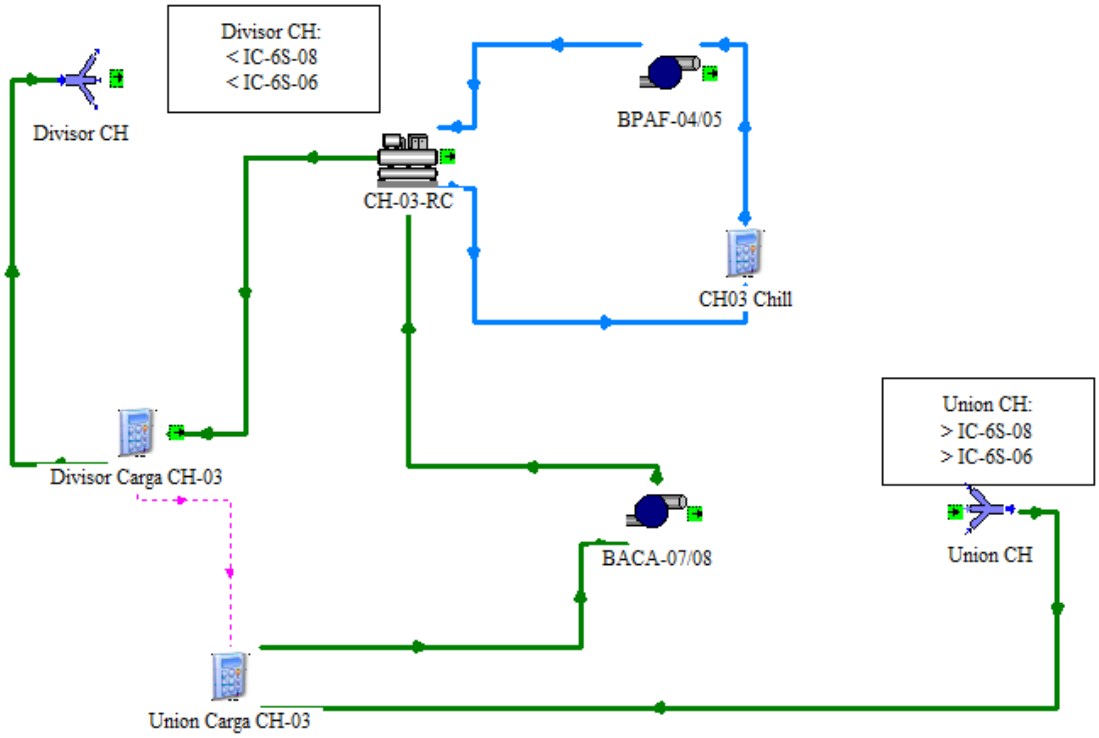
Appendix 1: Simple TRNSYS model sample. Source: Own elaboration from TRNSYS documentation [16].



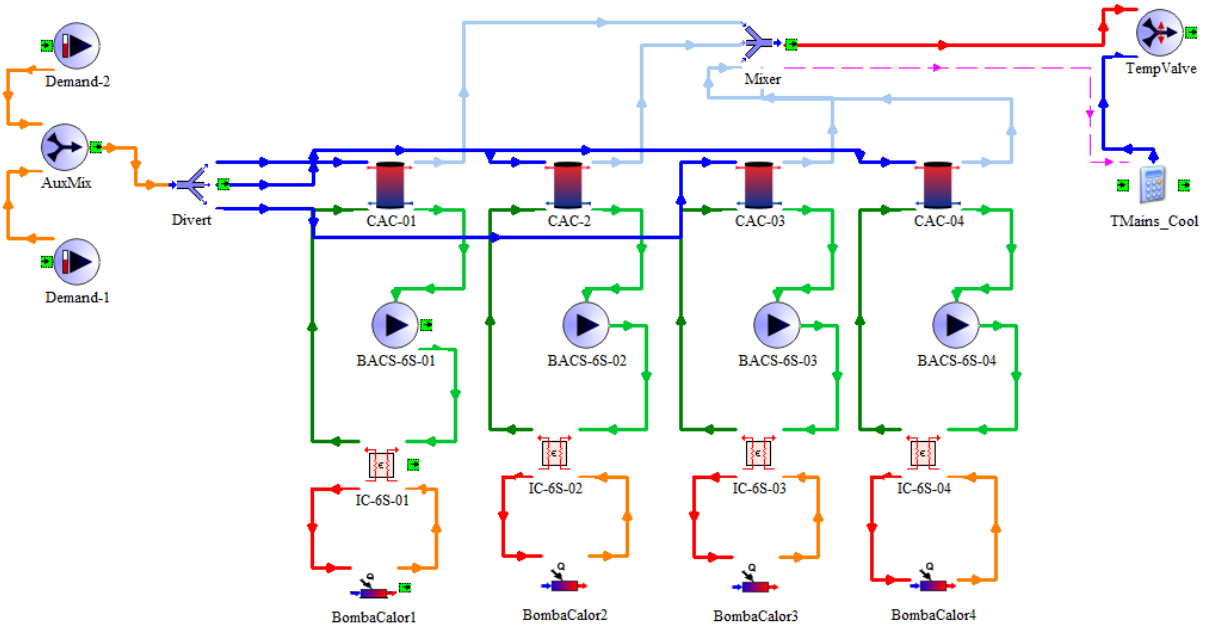
Appendix 2: TRNSYS Solar Field Subsystem. Source: Own elaboration from TRNSYS documentation [16].



Appendix 3: TRNSYS Heat Storage CAC-06 Subsystem. Source: Own elaboration from TRNSYS documentation [16].



Appendix 4: TRNSYS Chiller Subsystem. Source: Own elaboration from TRNSYS documentation [16].



Appendix 5: TRNSYS Heat Pump Subsystem. Source: Own elaboration from TRNSYS documentation [16].

B. Solar Collector Datasheet

TÜVRheinland®
DIN CERTCO

Page 1/2

Summary of EN 12975 Test Results, annex to Solar KEYMARK Certificate		Licence number	011 - 7S504 R
		Date of issue	

Company holding the licence	China Hitek Solar Co., Ltd.	Country	China		
Brand (optional)		Website	www.hitek-solar.com		
Street, number	No. 25 Zhujiang Road	E-mail	info@hitek-solar.com		
Postal Code	214028	Tel.	+86	51 085 229 209	
City	New District, Wuxi	Fax			

Collector Type (flat plate / evacuate tubular / un-glazed)	Evacuated tubular collector
Integration in the roof possible ?	No

Collector name	Aperture area (Aa) [m²]	Gross length [mm]	Gross width [mm]	Gross height [mm]	Gross area (Ag) [m²]	Power output per collector unit G = 1000 W/m² Tm-Ta : 0 K [W]	10 K [W]	30 K [W]	50 K [W]	70 K [W]
NSC-58-10	0.950	1990	887	180	1.765	587	572	532	478	409
NSC-58-12	1.120	1990	1047	180	2.084	692	675	627	563	483
NSC-58-15	1.400	1990	1287	180	2.561	865	843	784	704	604
NSC-58-16	1.493	1990	1367	180	2.720	922	899	836	751	644
NSC-58-18	1.680	1990	1527	180	3.039	1038	1012	941	845	724
NSC-58-20	1.866	1990	1687	180	3.357	1153	1124	1045	938	804
NSC-58-22	2.053	1990	1847	180	3.676	1268	1236	1150	1032	885
NSC-58-24	2.240	1990	2007	180	3.994	1384	1349	1254	1127	966
NSC-58-25	2.333	1990	2087	180	4.153	1441	1405	1306	1173	1006
NSC-58-30	2.800	1990	2487	180	4.949	1730	1686	1568	1408	1207

Collector efficiency parameters related to aperture area (Aa) Type of fluid and flow rate see note 1	η_{0a}	0.618	-
	a_{1a}	1.3767	W/(m²K)
	a_{2a}	0.0184	W/(m²K²)
Stagnation temperature - Weather conditions see note 2	t_{stg}	187	°C
Effective thermal capacity	$c_{eff} = C/A_a$	5.684	kJ/(m²K)
Max. operation pressure - see note 3	p_{max}	600	kPa

Incidence angle modifiers $K_\theta(\theta)$	G_{DF}/G_{TOT} min	G_{DF}/G_{TOT} max	θ_T / θ_L	50°	10°	20°	30°	40°	60°	70°
	0.08	0.14	$K_\theta(\theta_T)$	1.452	1.010	1.019	1.056	1.151	1.462	1.261
			$K_\theta(\theta_L)$	0.952	0.999	0.994	1.018	0.974	0.913	0.833
G_{DF}/G_{TOT}: min&max - while measuring					Optional values					

Testing Laboratory	IZES gGmbH, TZSB an der HTW
Website	www.izes.de/tzsb/
Test report id. number	KT08_15_5, C917LPEN-A1, C917QPEN-A1
Date of test report	18-06-2013, 08-04-2013, 08-04-2013
Perf. test method	EN 12975-2 6.1.4 (outdoor)

Comments of testing laboratory :
1) Kollektoren/collectors NSC-58-12, 15, 16, 18, 20, 22, 24, 25: Maße aus den technischen Zeichnungen/measure out of drawings 2) Kollektoren/collectors NSC-58-12, 15, 16, 18, 20, 22, 24, 25, 30: Leistung je Kollektormodul (0 K - 70 K) mit Kollektorleistungsparametern des Kollektors NSC-58-10 berechnet/ the power output per collector unit (0 K - 70 K) is calculated based on the performance parameters of collector NSC-58-10 3) Sämtliche Prüfungen am Kollektor NSC-58-30 wurden am SPF Institut für Solartechnik in Rapperswil, Schweiz durchgeführt/All tests of the collector NSC-58-30 were performed at the SPF Institut für Solartechnik in Rapperswil, Switzerland

Note 1	Fluid	Water	Flow rate	0.0190	kg/s per m²
Note 2	Irradiance, Gs=1000 W/m²; Ambient temperature , Ta=30 °C				
Note 3	Given by manufacturer				

tzsb

VERSION 3.6, 2012.01.20

DIN CERTCO • Alboinstraße 56 • 12103 Berlin
Tel: +49 30 7562-1131 • Fax: +49 30 7562-1141 • E-Mail: info@dincertco.de • www.dincertco.de.

Appendix 6: Hitek Solar Datasheet, page 1.

TÜVRheinland®
DIN CERTCO

Annual collector output based on EN 12975 Test Results, annex to Solar KEYMARK Certificate	Licence number	011 - 7S504 R
	Issued	0

Annual collector output kWh

Collector name	Location and collector temperature (Tm)											
	Athens			Davos			Stockholm			Würzburg		
	25°C	50°C	75°C	25°C	50°C	75°C	25°C	50°C	75°C	25°C	50°C	75°C
NSC-58-10	1105	925	705	989	792	569	669	519	364	723	563	393
NSC-58-12	1303	1091	831	1166	934	671	789	611	429	853	663	463
NSC-58-15	1628	1363	1039	1457	1167	839	987	764	536	1066	829	578
NSC-58-16	1737	1454	1108	1554	1245	894	1052	815	572	1137	884	617
NSC-58-18	1954	1636	1247	1748	1401	1006	1184	917	644	1279	995	694
NSC-58-20	2170	1817	1385	1942	1556	1118	1315	1018	715	1421	1105	771
NSC-58-22	2388	1999	1523	2136	1712	1230	1447	1121	787	1563	1216	848
NSC-58-24	2605	2181	1662	2331	1868	1342	1578	1223	858	1705	1327	925
NSC-58-25	2714	2272	1731	2428	1945	1398	1644	1273	894	1776	1382	964
NSC-58-30	3257	2726	2078	2914	2335	1677	1973	1528	1073	2132	1658	1157

Collector mounting: Fixed or tracking	Fixed; slope = latitude - 15° (rounded to nearest 5°)

Overview of locations

Location	Latitude °	Gtot kWh/m²	Ta °C	Collector orientation or tracking mode
Athens	38	1 765	18.5	South, 25°
Davos	47	1 714	3.2	South, 30°
Stockholm	59	1 166	7.5	South, 45°
Würzburg	50	1 244	9.0	South, 35°

Gtot	Annual total irradiation on collector plane	kWh/m²
Ta	Mean annual ambient air temperature	°C
Tm	Constant collector operating temperature (mean of in- and outlet temperatures)	°C

Calculation of the annual collector performance is done by the official Solar Keymark spreadsheet tool. Hour by hour the collector output is calculated according to the efficiency parameters from the Keymark test using constant collector operating temperature (Tm). Detailed description with all equations used is available from the Solar Keymark web site (direct link: http://www.estif.org/solarkeymark/annexb1.php)

DIN CERTCO • Alboinstraße 56 • 12103 Berlin
Tel: +49 30 7562-1131 • Fax: +49 30 7562-1141 • E-Mail: info@dincertco.de • www.dincertco.de.

Datasheet version:
VERSION 3.6, 2012.01.20
Calculation program version:
3.07, October 2011 (SP)

Appendix 7: Hitek Solar Datasheet, page 2.

C. Heat-Recovery Chiller Datasheet

3. Tecnichal data

GENERAL TECHNICAL DATA — **CWC PROZONE C-H**

MODEL		155 Z	170 Z	190 Z	1120 Z	1130 Z	1170 Z	1190 Z	1210 Z
COOLING									
Cooling capacity	kW	52,8	65,6	86,2	113,0	129,4	168,1	184,5	208,1
Compressors power input	kW	10,8	13,4	17,6	23,0	26,1	34,1	37,2	42,0
Total EER		4,89	4,90	4,90	4,91	4,96	4,93	4,96	4,95
ESEER		6,23	6,29	6,28	6,09	6,03	6,03	6,14	6,18
IPLV (1)		6,71	6,73	6,72	6,52	6,52	6,45	6,56	6,60
HEATING									
Heating capacity	kW	59,3	73,6	96,8	126,8	145,0	188,6	206,7	233,3
Compressors power input in heating mode	kW	13,4	16,7	21,9	28,6	32,5	42,5	46,4	52,4
Total COP		4,43	4,41	4,42	4,43	4,46	4,44	4,45	4,45
COMPRESSORS									
Number of compressors	n	2	2	2	2	2	2	2	2
Refrigerant circuits	n	1	1	1	1	1	1	1	1
Unloading steps	n	2	2	2	2	2	2	2	2
Refrigerant charge	kg	7,1	8,9	11,6	15,3	17,5	22,7	24,9	28,1
Oil charge	kg	4,1	5,1	6,7	8,7	10,0	13,0	14,3	16,1
EVAPORATOR									
Water flow	m^3/h	9,08	11,28	14,83	19,44	22,26	28,91	31,73	35,79
Pressure drops	kPa	52,1	61,0	98,5	89,3	102,4	59,7	63,1	63,3
Water flow (Heat Pump)	m^3/h	7,9	9,8	12,9	16,9	19,4	25,1	27,6	31,1
Pressure drops (Heat Pump)	kPa	39,4	45,9	74,4	67,5	77,4	45,1	47,6	47,8
CONDENSER									
Water flow	m^3/h	10,94	13,59	17,85	23,39	26,75	34,78	38,13	43,02
Pressure drops	kPa	53,0	76,0	78,0	50,0	55,0	60,0	57,0	59,0
Water flow (Heat Pump)	m^3/h	10,2	12,7	16,6	21,8	24,9	32,4	35,6	40,1
Pressure drops (Heat Pump)	kPa	46,1	66,0	67,8	43,5	47,8	52,2	49,5	51,3
SOUND DATA									
Sound power level (ISO 3744)	dB	77,0	77,7	79,2	83,3	85,6	86,8	88,0	88,0
Sound pressure level at 5 mt (ISO 3744)	dB	50,8	51,5	53,0	57,1	59,4	60,6	61,8	61,8
Sound pressure level at 10 mt (ISO 3744)	dB	45,4	46,1	47,6	51,7	54,0	55,2	56,4	56,4
DIMENSIONS AND WEIGHT									
Lenght	mm	1535	1535	1535	1535	1535	1535	1535	1535
Width	mm	685	685	685	685	685	685	685	685
Height	mm	1660	1660	1660	1660	1660	1660	1660	1660
Shipping weight	kg	456	492	528	547	642	671	808	869

Cooling: Condenser water In/Out 30/35°C; Evaporator water temperature In/Out 12/7°C.
Heating: Condenser water In/Out 40/45°C; Evaporator water temperature In/Out 15/10°C.
(1) Calculated according to ARI standard 550/590.
Water flow rate and sound pressure levels refer to summer period.

Appendix 8: CWC Prozone 1320Z C Datasheet, page 1.

3. Tecnichal data

COOLING CAPACITY PERFORMANCES **CWC PROZONE C**

Twout			1250 Z						1320 Z					
			Water temperature						Water temperature					
			30	35	40	45	50	55	30	35	40	45	50	55
5	Pf	kW	238,3	226,8	214,5	201,0	186,6	170,2	309,3	293,5	276,7	258,9	240,8	221,3
	Pa	kW	44,1	48,5	53,5	59,2	65,6	72,8	57,1	62,7	69,1	76,4	84,8	94,3
	qwe	m³/h	41,0	39,0	36,9	34,6	32,1	29,3	53,2	50,5	47,6	44,5	41,4	38,1
	dpwe	kPa	56,9	51,5	46,1	40,5	34,9	29,0	59,2	53,3	47,4	41,5	35,9	30,3
	qwc	m³/h	48,6	47,4	46,1	44,8	43,4	41,8	63,0	61,3	59,5	57,7	56,0	54,3
	dpwc	kPa	56,0	53,2	50,4	47,5	44,7	41,5	59,1	55,9	52,7	49,5	46,7	43,9
6	Pf	kW	245,7	233,9	221,3	207,5	192,8	176,1	319,0	302,8	285,5	267,1	248,4	228,4
	Pa	kW	44,2	48,6	53,6	59,3	65,7	72,9	57,3	62,8	69,2	76,5	84,8	94,4
	qwe	m³/h	42,3	40,2	38,1	35,7	33,2	30,3	54,9	52,1	49,1	45,9	42,7	39,3
	dpwe	kPa	60,5	54,8	49,1	43,1	37,3	31,1	63,0	56,7	50,4	44,1	38,2	32,3
	qwc	m³/h	49,9	48,6	47,3	45,9	44,5	42,8	64,7	62,9	61,0	59,1	57,3	55,5
	dpwc	kPa	59,0	56,0	53,1	50,0	46,9	43,5	62,4	58,9	55,4	52,0	48,9	45,9
7	Pf	kW	253,3	**241,2**	228,3	214,1	199,2	182,1	328,9	**312,2**	294,4	275,5	256,3	235,7
	Pa	kW	44,3	**48,7**	53,7	59,3	65,7	72,9	57,5	**63,0**	69,3	76,6	84,9	94,4
	qwe	m³/h	43,6	**41,39**	39,3	36,8	34,3	31,3	56,6	**53,57**	50,6	47,4	44,1	40,5
	dpwe	kPa	64,3	**58,0**	52,2	45,9	39,7	33,2	66,9	**60,0**	53,6	47,0	40,6	34,4
	qwc	m³/h	51,2	**49,86**	48,5	47,0	45,6	43,9	66,5	**64,53**	62,6	60,6	58,7	56,8
	dpwc	kPa	62,2	**59,0**	55,8	52,5	49,3	45,7	65,8	**62,0**	58,3	54,6	51,3	48,0
8	Pf	kW	261,1	248,6	235,4	220,9	205,6	188,2	339,1	321,9	303,5	284,1	264,4	243,2
	Pa	kW	44,4	48,8	53,8	59,4	65,8	73,0	57,6	63,2	69,5	76,8	85,1	94,5
	qwe	m³/h	44,9	42,8	40,5	38,0	35,4	32,4	58,3	55,4	52,2	48,9	45,5	41,8
	dpwe	kPa	68,3	61,9	55,5	48,9	42,4	35,5	71,1	64,1	57,0	49,9	43,2	36,6
	qwc	m³/h	52,5	51,2	49,7	48,2	46,7	44,9	68,2	66,2	64,2	62,1	60,1	58,1
	dpwc	kPa	65,5	62,1	58,7	55,2	51,7	47,9	69,3	65,3	61,3	57,4	53,8	50,2
9	Pf	kW	269,0	256,3	242,6	227,8	212,2	194,5	349,4	331,7	312,9	292,9	272,6	250,9
	Pa	kW	44,6	48,9	53,8	59,5	65,8	73,0	57,8	63,3	69,6	76,9	85,2	94,6
	qwe	m³/h	46,3	44,1	41,7	39,2	36,5	33,4	60,1	57,1	53,8	50,4	46,9	43,2
	dpwe	kPa	72,5	65,8	59,0	52,0	45,1	37,9	75,5	68,1	60,6	53,1	46,0	38,9
	qwc	m³/h	53,9	52,5	51,0	49,4	47,8	46,0	70,0	68,0	65,8	63,6	61,5	59,4
	dpwc	kPa	69,0	65,4	61,7	57,9	54,3	50,2	73,0	68,7	64,5	60,2	56,4	52,6
10	Pf	kW	277,1	264,0	250,0	234,9	218,9	200,8	360,0	341,9	322,5	302,0	281,1	258,7
	Pa	kW	44,7	49,0	53,9	59,5	65,9	73,1	58,0	63,5	69,8	77,0	85,3	94,8
	qwe	m³/h	47,7	45,4	43,0	40,4	37,7	34,5	61,9	58,8	55,5	51,9	48,4	44,5
	dpwe	kPa	77,0	69,8	62,6	55,3	48,0	40,4	80,2	72,3	64,3	56,4	48,9	41,4
	qwc	m³/h	55,4	53,8	52,3	50,6	49,0	47,1	71,9	69,7	67,5	65,2	63,0	60,8
	dpwc	kPa	72,7	68,8	64,9	60,9	57,0	52,7	77,0	72,4	67,8	63,3	59,1	55,0

Twout Leaving water temperature (°C)
Pf Cooling capacity (kW)
Pa Compressors power input (kW)
qwe Evaporator water flow (m³/h)
dpwe Evaporator pressure drop (kPa)
qwc Condenser water flow (m³/h)
dpwc Condenser pressure drop (kPa)

Appendix 9: CWC Prozone 1320Z C Datasheet, page 2.

D. Solar Angles

The daily and yearly path across the sky of the sun can be described by solar angles. These variables allow the calculation of the quantity of incidence radiation on a particular surface, described by the incidence angle shown in Equation (39) [20]. TRNSYS is able to obtain these angles based on the weather data, coordinates and time zone presented.

The solar collector field is designed with the following parameters:

- Inclination angle:

$$\beta = 15° \tag{32}$$

- Surface azimuth angle, facing north:

$$\gamma = 0° \tag{33}$$

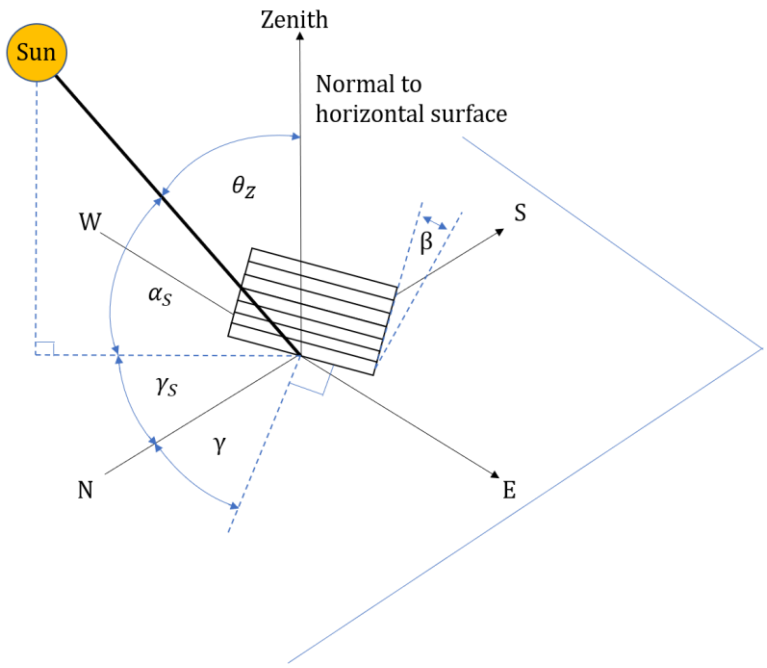


Figure 80: Zenith angle, slope, surface azimuth angle and solar azimuth angle for a tilted surface. Adapted from [20].

The zenith angle is described by the following equation, which in turn depends on the solar declination angle (Eq.(35)) and the hour angle (Eq. (36)).

$$\cos\theta_z = \cos L \cdot \cos\delta \cdot \cos\omega + \sin L \cdot \sin\delta \tag{34}$$

$$\delta = 0.006918 - 0.399912\,cos\,B + 0.070257\,sin\,B - 0.006758\,cos\,2B + 0.000907\,sin\,2B - 0.002697\,cos\,3B + 0.00148\,sin\,3B \tag{35}$$

$$\omega = (\mathrm{h} - 12) \cdot 15 \tag{36}$$

Where h is the solar time in hour units.

From Figure 80, with the zenith angle, the elevation angle is described by:

$$\alpha + \theta_z = \frac{\pi}{2} \tag{37}$$

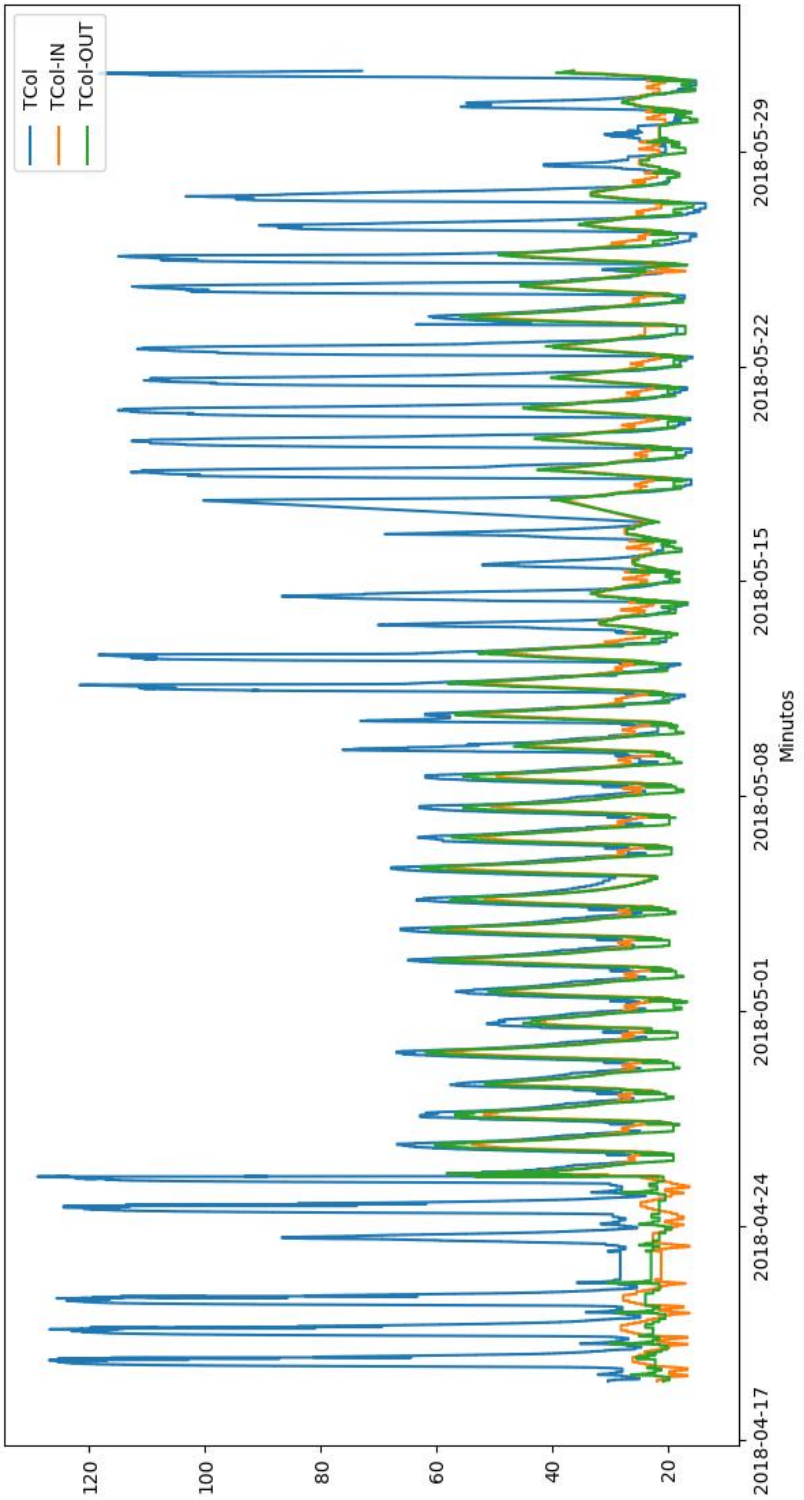
Thus, the solar azimuth can be found with the following expression

$$sin\,\gamma_s = -\frac{cos\,\delta \cdot sin\,\omega}{cos\,\alpha} \tag{38}$$

Finally, the incidence angle on a tilted surface is described by Equation (39). This angle is used to calculate the incident global and diffuse irradiance, as well as the direct horizontal irradiance from the weather data provided.

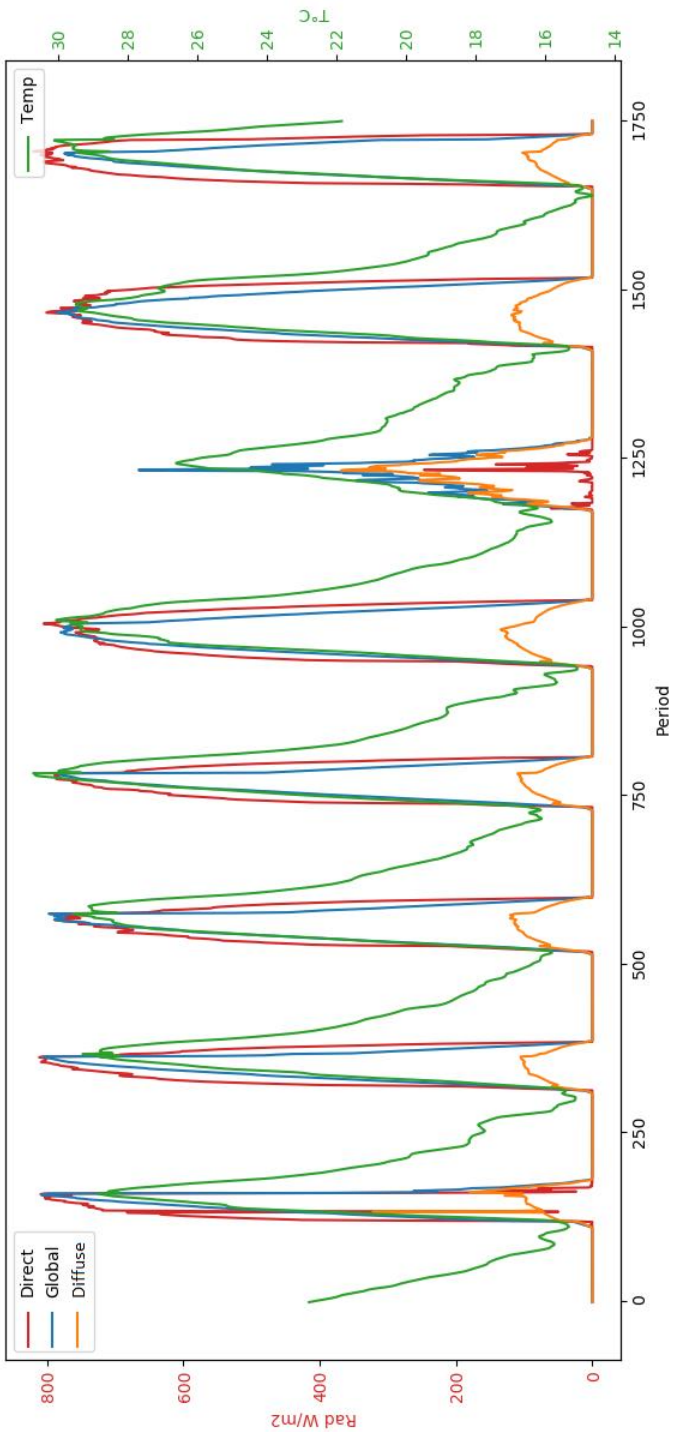
$$\cos\theta = \cos\theta_z \cdot \cos\beta + \sin\theta_z \cdot \sin\beta \cdot \cos(\gamma_s - \gamma) \tag{39}$$

E. Solar Collector Temperatures



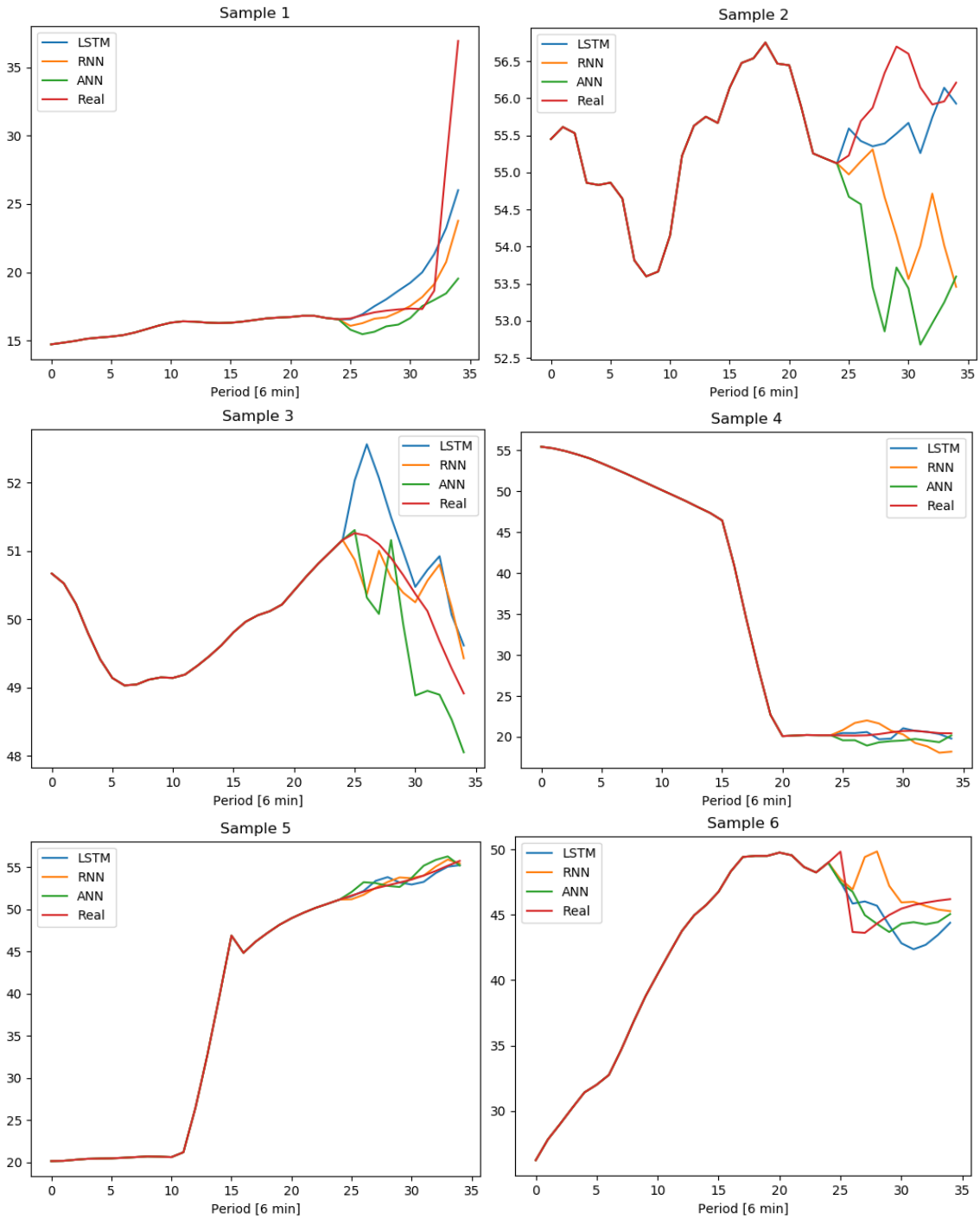
Appendix 10: Measured temperatures from the solar collector. field.

F. Measured Radiation and Ambient Temperature Values

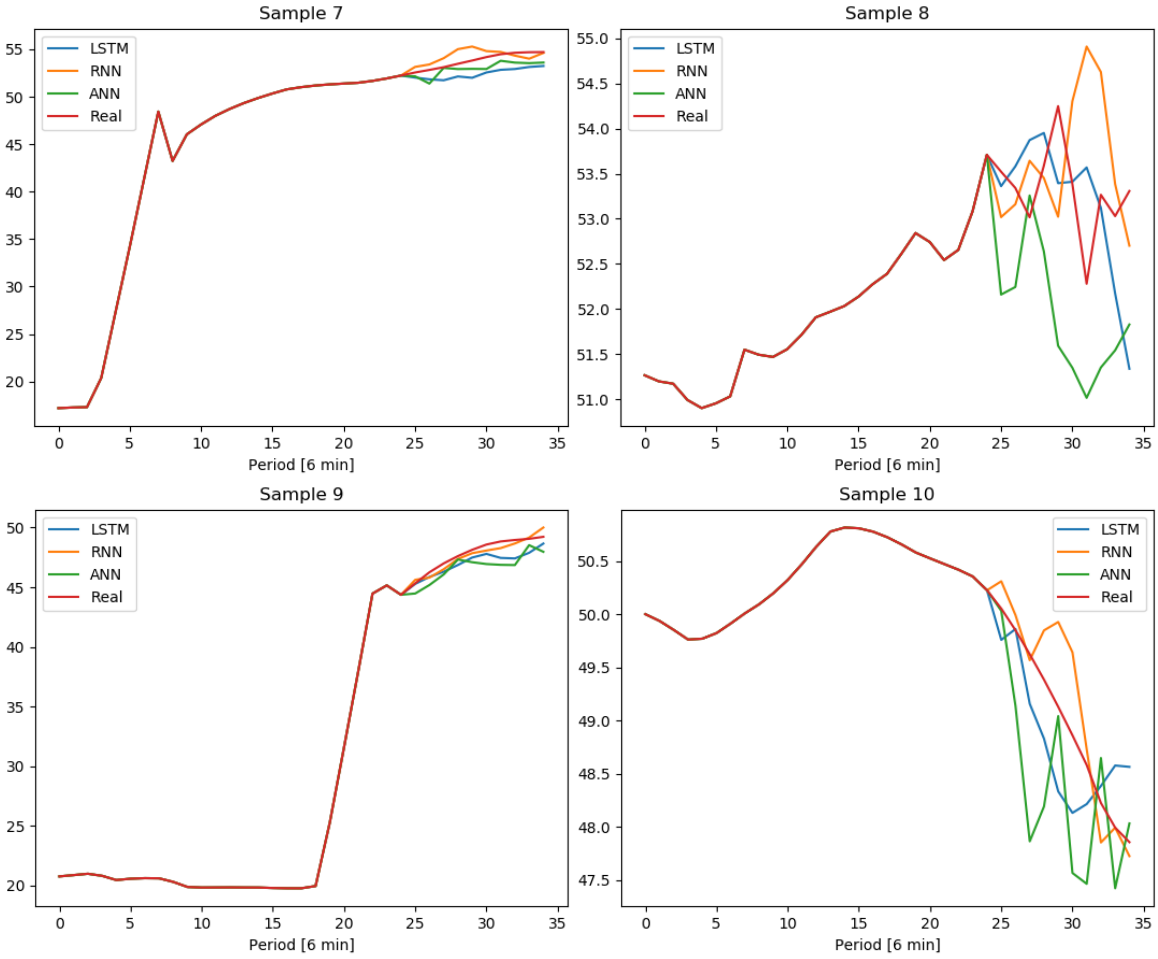


Appendix 11: Measured Radiation and Ambient Temperature values for 13-04 to 25-05.

G. Sequence Prediction for Solar Collector Outlet Temperature Samples

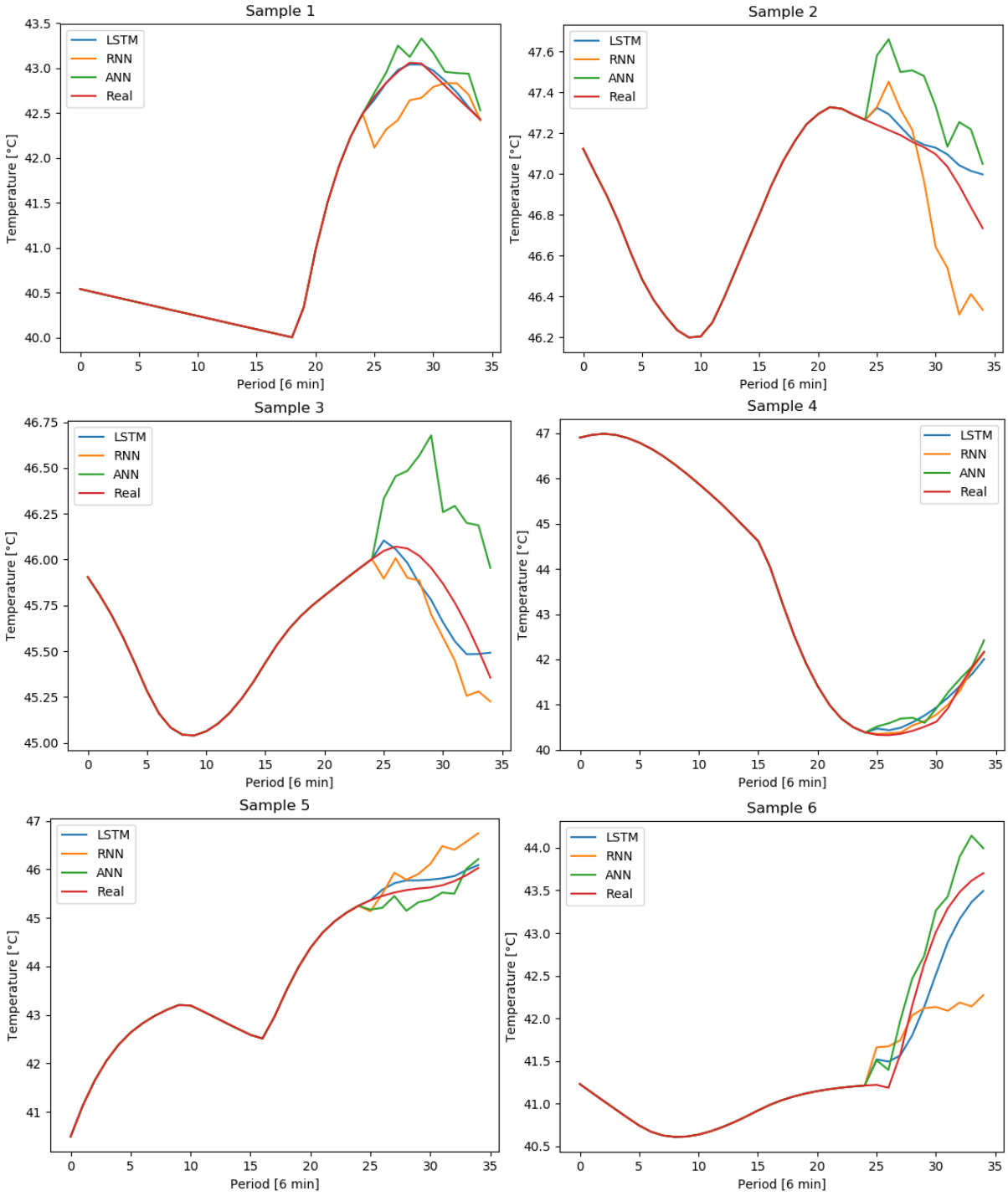


Appendix 12: Sequence Prediction for Solar Collector Output Temperature, Samples 1-6.

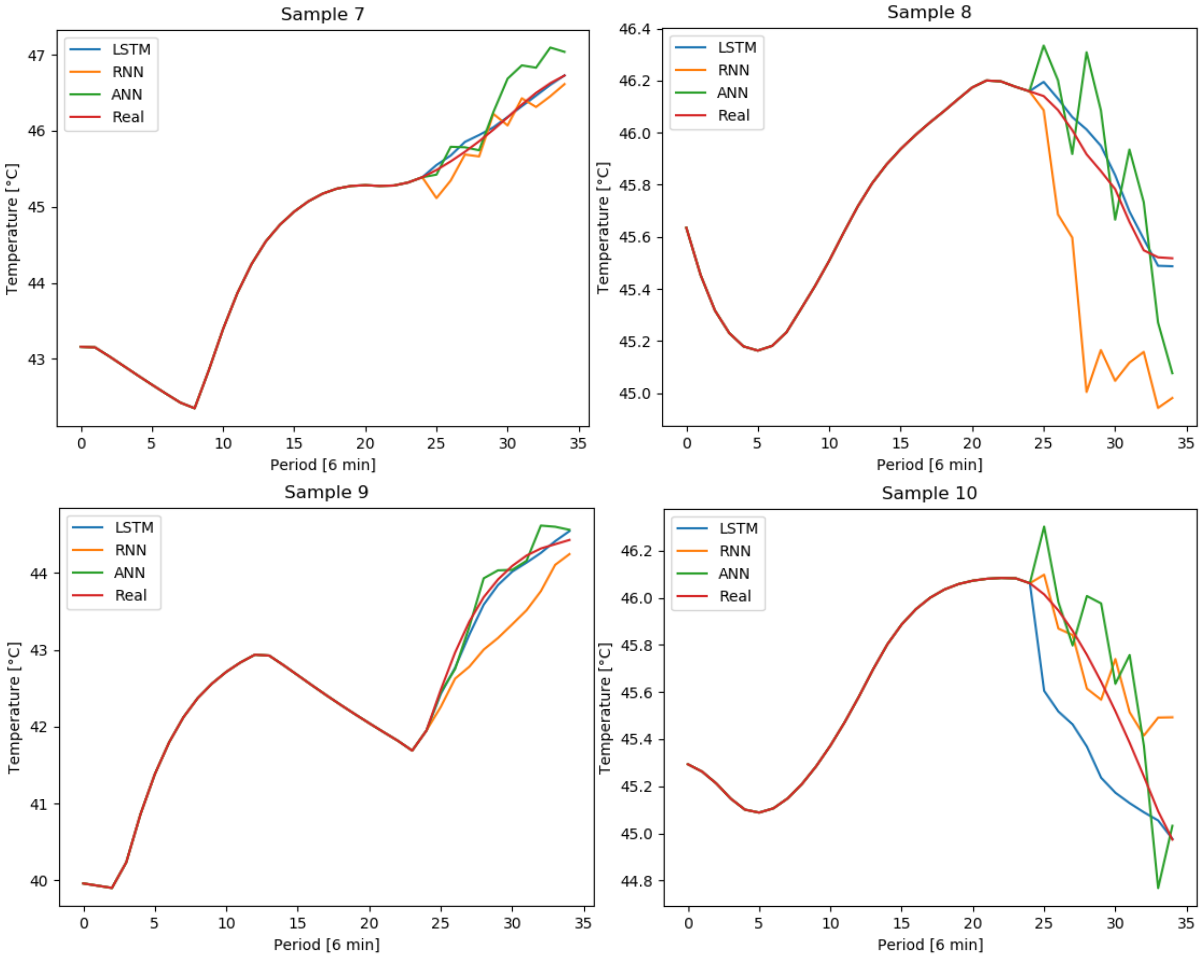


Appendix 13: Sequence Prediction for Solar Collector Output Temperature, Samples 7-10.

H. Sequence Prediction for Average Tank Temperature Samples

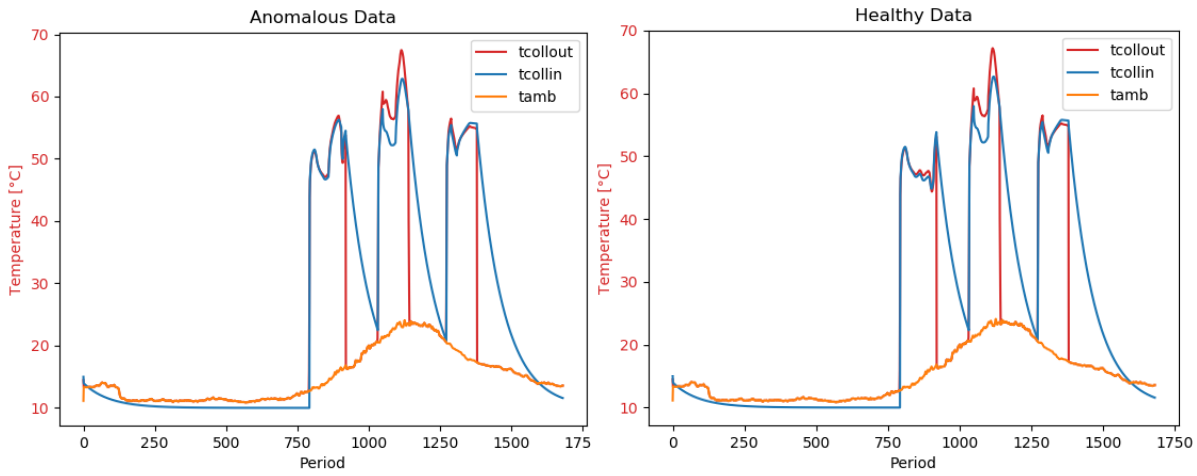


Appendix 14: Sequence Prediction for Average Tank Temperature, Samples 1-6.

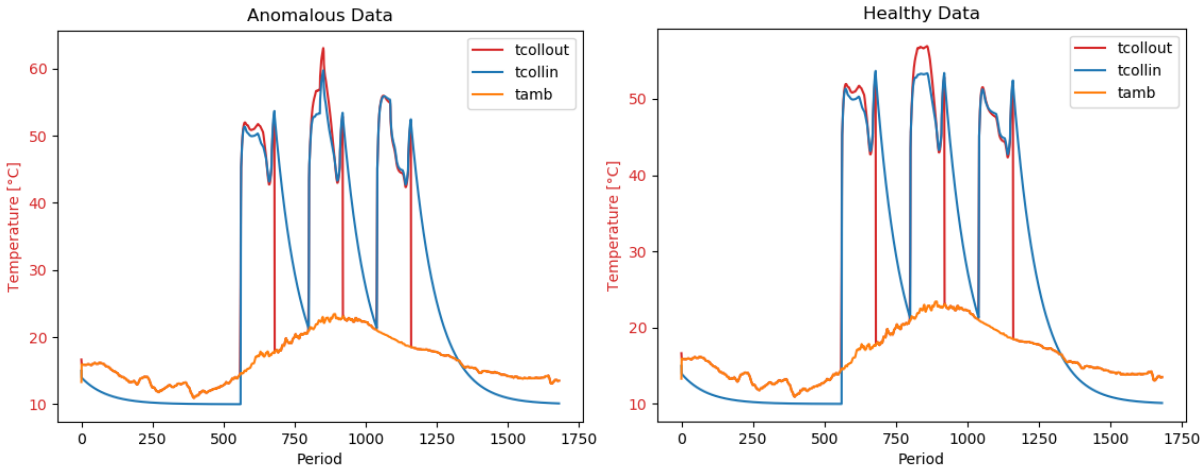


Appendix 15: Sequence Prediction for Average Tank Temperature, Samples 7-10.

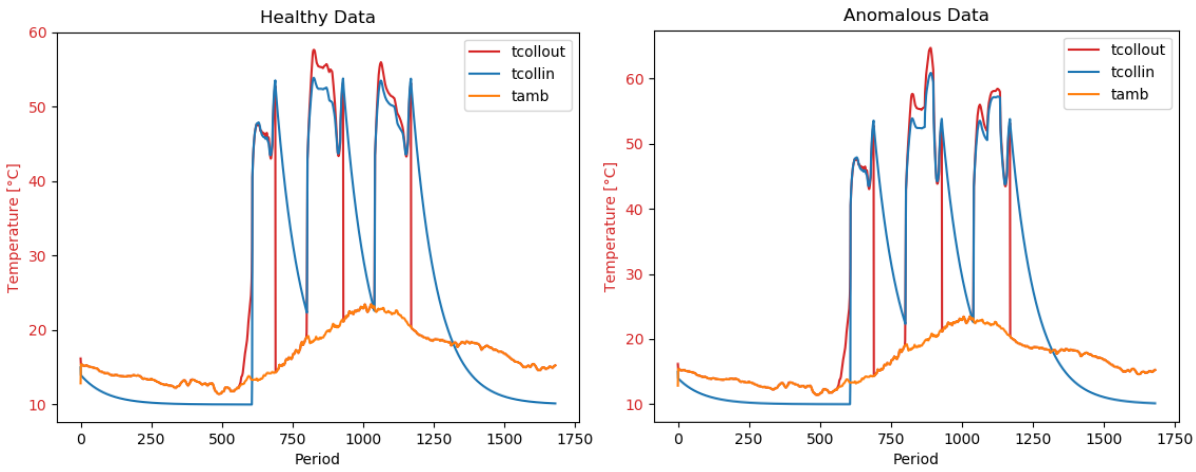
I. Anomalous Behavior for Water-Draw Induced Fault



Appendix 16: Faulty Water-Draw Profiles, Case 1: Comparison of Healthy and Anomalous Solar Collector Temperatures.

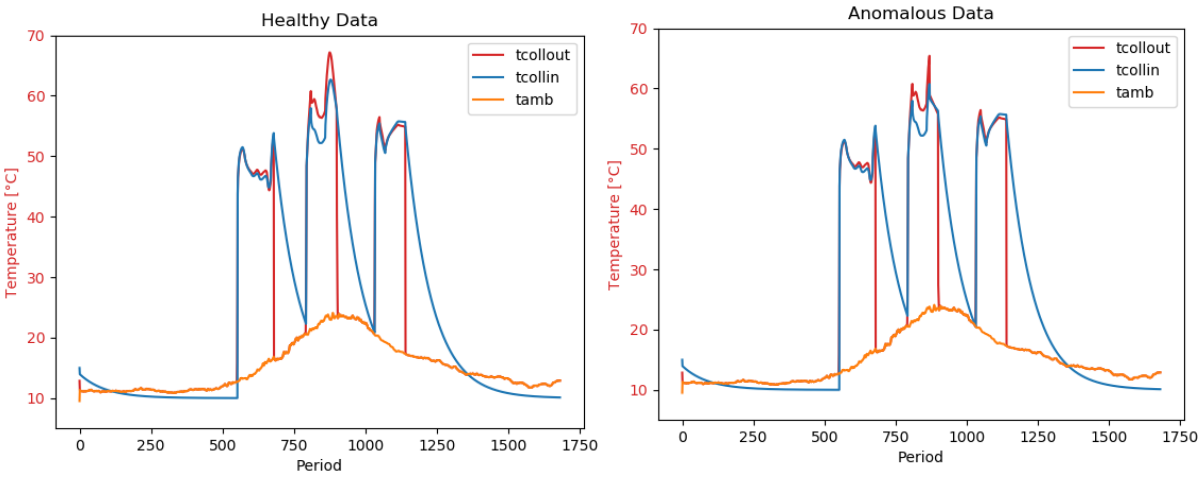


Appendix 17: Faulty Water-Draw Profiles, Case 2: Comparison of Comparison of Healthy and Anomalous Solar Collector Temperatures.

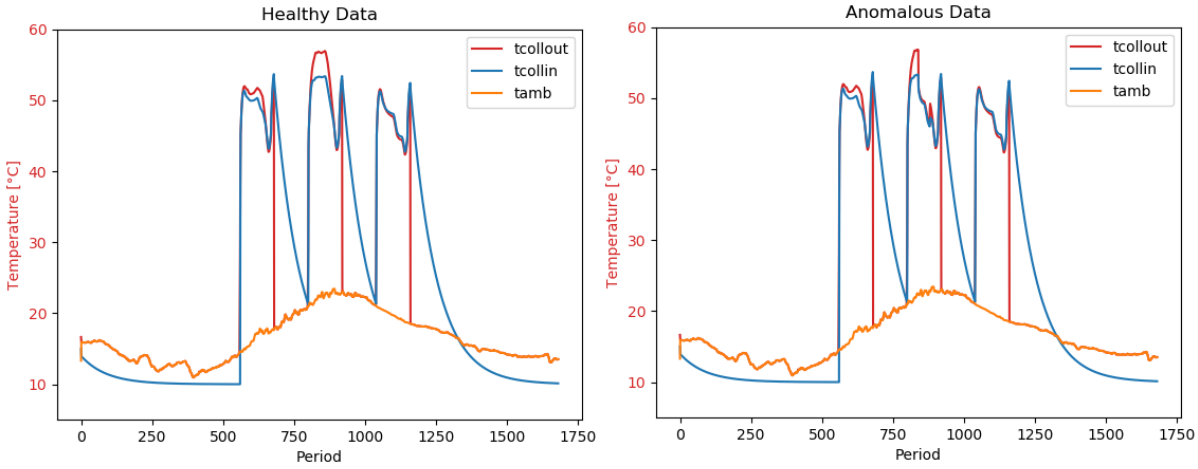


Appendix 18: Faulty Water-Draw Profiles, Case 3: Comparison of Comparison of Healthy and Anomalous Solar Collector Temperatures.

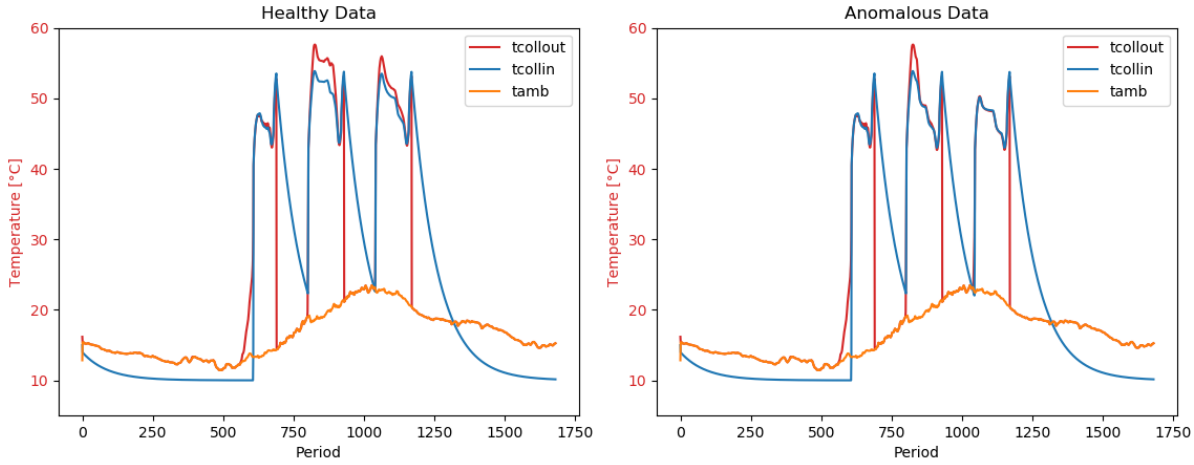
J. Anomalous Behavior for Collector Efficiency Induced Fault



Appendix 19: Faulty Collector Efficiency, Case 1: Comparison of Healthy and Anomalous Solar Collector Temperatures.



Appendix 20: Faulty Collector Efficiency, Case 2: Comparison of Healthy and Anomalous Solar Collector Temperatures.



Appendix 21: Faulty Collector Efficiency, Case 3: Comparison of Healthy and Anomalous Solar Collector Temperatures.