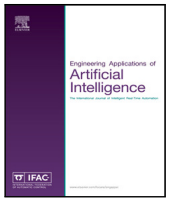
Engineering Applications of Artificial Intelligence 133 (2024) 108387

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai



Research paper

Artificial Neural Network-based digital twin for a flat plate solar collector field

M. Castilla, J.L. Redondo, A. Martínez, J.D. Álvarez *

Department of Informatics, CIESOL—CIAMBITAL—ceiA3, University of Almería, Ctra. Sacramento s/n, La Cañada de San Urbano, Almería, 04120, Spain

ARTICLE INFO

Keywords:
Digital twin
Artificial neural network
Forecasting model
Digital integration
Flat plate solar collector field

ABSTRACT

In this study, a digital twin for a flat plate solar collector field is proposed. This kind of system is used to reduce carbon dioxide emissions in bioclimatic buildings to convert them into Zero Energy Buildings. The core of the digital twin is an Artificial Neural Network prediction model, which is a good alternative to models based on physical equations for modeling systems with strong non-linearities, such as the ones found in flat plate solar collectors. The Artificial Neural Network prediction model is calibrated and validated with data saved during one year of operation comprising sunny days, cloudy days, partially cloudy days and non-operation days. Validation shows good results using several statistical metrics, suggesting that the Artificial Neural Network model is suitable for operation and control purposes. With a highly accurate virtual representation, the Artificial Neural Network model allows data analysis of the plant operator, prediction of behavior, and offers recommendations for optimizing system performance. In addition, the digital twin presented as part of this work is not just limited to the model, but is also enriched by the integration of data acquisition technologies and a user interface into a web page. This innovative integration establishes a robust framework for proactive, real-time decision-making and efficient management of the plant, ensuring enhanced system operation and sustainability.

1. Introduction

According to recent studies, the building sector is responsible for almost 40% of the total energy consumption of the EU and 36% of CO_2 emissions (European Climate Foundation and the European Alliance to Save Energy, 2022). A photovoltaic field is an ideal solution for supplying electricity to any building, but the use of solar energy transcends this narrow scope. Solar energy can be used in a variety of ways, for example, in flat plate solar collectors, by coupling them with an absorption machine, to feed the Heating, Ventilation and Air Conditioning (HVAC) system of the building, which is the main source of the building's CO_2 emissions. Moreover, this kind of solar collector can heat water for domestic purposes.

However, all methods of obtaining renewable solar energy share a common vulnerability, as their performance is intrinsically linked to meteorological conditions. Consequently, their efficiency experiences fluctuations that follow prevailing weather conditions. Therefore, due to the changing and non-controllable nature of solar irradiation, energy systems based on solar energy, like any other renewable energy system, need an energy buffer subsystem to keep the plant operating evenly during the absence of the primary renewable energy source. In this regard, solar collectors employ thermal tanks to store surplus energy, which can be taken advantage of during periods of cloud cover, as opposed to photovoltaic plants which typically have to rely on batteries to store energy for use on cloudy days.

Thus, the plant has several modes of operation depending on weather conditions and energy demand, meaning a solar collector model is necessary for the proper operation of the plant. Apart from that, the model can be used to analyze the plant's performance or for training purposes. However, nowadays, by implementing the Internet of Things (IoT) and Industry 4.0 paradigm, models can evolve into digital twins. A digital twin embeds a "virtual" image of reality constantly synchronized with the real operating scenario to provide reliable information (knowledge model) to the actual interpretation model to make sound decisions (Semeraro et al., 2021). From this definition, it becomes evident that real-time bidirectional communication is a fundamental requirement between the digital twin and the physical plant to realize the full potential of a digital twin. This real-time interplay between the physical entity and its digital counterpart stands as the chief distinction between a traditional model and a digital twin. Accordingly, digital twin technology has become an emerging and vital field in engineering for digital transformation and intelligent upgrades.

* Corresponding author.
E-mail addresses: mcastilla@ual.es (M. Castilla), jlredondo@ual.es (J.L. Redondo), ams929@inlumine.ual.es (A. Martínez), jhervas@ual.es (J.D. Álvarez).

https://doi.org/10.1016/j.engappai.2024.108387
Received 27 December 2023; Received in revised form 7 March 2024; Accepted 3 April 2024
Available online 9 April 2024

This work deals with the development and implementation of a digital twin for a flat plate solar collector field housed in a bioclimatic building, the CIESOL Research Center, at the University of Almería (UAL), Spain. The flat solar collector field comprises ten loops, each containing eight solar collectors. It is connected to an absorption machine with a refrigeration tower circuit, which feeds the HVAC system of the building with hot or cold water on demand. In addition, the flat plate solar collector field can supply hot water for domestic usage, meaning the CO_2 emissions of the CIESOL centre can be reduced, thus fulfilling several energy and bioclimatic objectives. The proposed digital twin is implemented through an Artificial Neural Network (ANN) prediction model calculated with real data from the solar collector field, which is gathered via a network of sensors strategically placed throughout the facilities. Subsequently, the ANN model is integrated into a web platform, establishing a connection with the actual solar collector field through the Open Platform Communications-Unified Architecture (OPC-UA) standard. This ensures that the digital twin continuously receives real-time data from these physical sensors. Within this digital twin, the ANN prediction model operates by forecasting the output temperature of the collector array based on the sensor inputs. As such, the digital twin can be used for operator training purposes, online fault detection or system analysis.

During the last few years, many papers dealing with digital twins have been published in several engineering fields (Tao et al., 2022). In the industrial sector, digital twins can be employed at various points in the process: during the design phase, in which they facilitate iterative optimization, virtual evaluation, and verification; in the manufacturing phase, in which they enable real-time monitoring, production control, process assessment, and optimization (Kumar et al., 2023); and finally in the service phase, in which they support predictive maintenance, fault detection and diagnosis (Rachmawati et al., 2023), and performance prediction (Liu et al., 2021). Digital twins have not been applied to the industrial sector alone; they have extended their reach into various other domains, including the aerospace industry (Shafto et al., 2010), robotics (Yan et al., 2018; Kuts et al., 2020; Li et al., 2020; Soliman et al., 2023), and the energy sector (Ghenai et al., 2022; Semeraro et al., 2023), among others.

To our knowledge, there is only one paper in literature which presents some similarities with the work developed in this paper. Specifically, in Machado et al. (2023) a digital twin for a Fresnel solar collector connected to an absorption machine is introduced. It is worth mentioning that in Fresnel solar collectors, solar irradiation is focused on a pipe located at the top of the structure. Moreover, the authors of that paper establish a comparison between an Adaptive Neuro-Fuzzy Inference Systems (ANFIS) model and a phenomenological model based on Partial Differential Equations (PDEs) and parameter identification. The results obtained show comparable results in terms of accuracy for both models, although the ANFIS model requires less time for execution.

Although the Fresnel collector employs different technology than our system, both systems share a common objective, namely, using solar energy to heat a heat-transfer fluid. Additionally, both Fresnel and flat plate solar collectors are mathematically described using similar PDEs due to the shared energy transport phenomena occurring within the pipe. However, despite these similarities, there are significant differences between the two works. First, our work makes use of much larger datasets for training, validation, and testing, spanning an entire year and encompassing all plant operation modes. In contrast, the other work only has on twenty-five days' worth of data. Second, our approach does not involve the comparison of the ANN model with a different type of model. Instead, we validate it using data from another collector loop within the solar field. This represents a novel contribution and distinction, allowing us to highlight the versatility and applicability of our ANN prediction model to several collector loops within the same field. This not only demonstrates its robustness, but also its generalizability. Lastly, another important distinction from the previous work is the integration of the digital twin into the web and its real-time communication with the system, which offers a clear view of its implementation and functionality.

The rest of the paper is organized as follows: in Section 2 the methodology used to develop the ANN is described. In addition, this section also includes a description of the CIESOL Research Center and the flat solar collector field. After that, Section 3 is devoted to the digital twin and lastly, the conclusions are presented in Section 4.

2. Development of the Artificial Neural Network prediction model

2.1. CIESOL Research Center

The flat plate solar collector field housed in the CIESOL Research Center is used as a testbed in this work (see Fig. 1(a)). Located on the campus of the University of Almería, CIESOL is a solar energy research center jointly run with the *Plataforma Solar de Almería* (PSA) and falls under the purview of the *Centro de Investigaciones Energéticas, Medioambientales y Tecnológicas* (CIEMAT) (Castilla et al., 2014). The building is equipped with a photovoltaic panel array to fulfill its electrical requirements, along with a flat plate solar collector field for sanitary hot water (see Fig. 1(b)).

The flat plate solar collector field is located on the rooftop of the building and it is composed by ten loops, each with eight solar collectors. In addition, the flat plate solar collector field is tilted at a 30-degree angle to optimize solar resource utilization and ensure uniform water flow distribution throughout all loops. Moreover, the output water temperature must be in the range of [−20,110] °C to avoid damage. The output water from the flat plate solar collector field is used to feed an absorption machine with a total cooling power of 70 kW. The HVAC system of the CIESOL building has two operating modes: summer and winter. For example, during summer operation mode, the HVAC system exhibits the following behavior (Pasamontes et al., 2011): first, water flows through the flat plate solar collector field, increasing its temperature. Subsequently, this hot water is used to feed the absorption machine. Afterwards, the absorption machine provides chilled water that is injected into the fan coil system. The fan coil system is distributed throughout the building and allows the indoor temperature to be maintained within the desired level in any room of the building. In this way, the energy demands of the HVAC system can be covered with solar energy, thus decreasing the CO_2 emissions of the building.

Fig. 2 provides a visual representation of the field layout, showing the exact locations of the integrated sensors, these being for the following:

- Outdoor air temperature, Tout (°C): This variable measures the outdoor air temperature constantly for subsequent recording.
- Solar irradiation, Irr (W/m^2): It signifies direct solar radiation, i.e., the solar radiation incident on the surface, measured at each time instant.
- Flow rates, Q1 and Q2 (m^3/h): Q1 constitutes the water that flows to the flat plate solar collector field, whereas Q2 is the water flow that the valve V1 diverts and, therefore, does not enter the solar collector field. Thus, with both flow rates, it is possible to calculate the total inlet flow to the solar collector field and, therefore, to each of the flat collector loops.
- Temperature for the inlet of a loop, SP_i with i ranging from 11 to 20 (°C): It indicates the inlet temperature of the heat transfer fluid at the beginning of a loop.
- Temperature for the outlet of a loop, SP_i with i ranging from 1 to 10 (°C): It indicates the outlet temperature of the heat transfer fluid from loop i of the flat collector field.

(a) Front view of the CIESOL Research Center (b) The CIESOL flat-plate solar collector field

Fig. 1. The CIESOL Research Center.

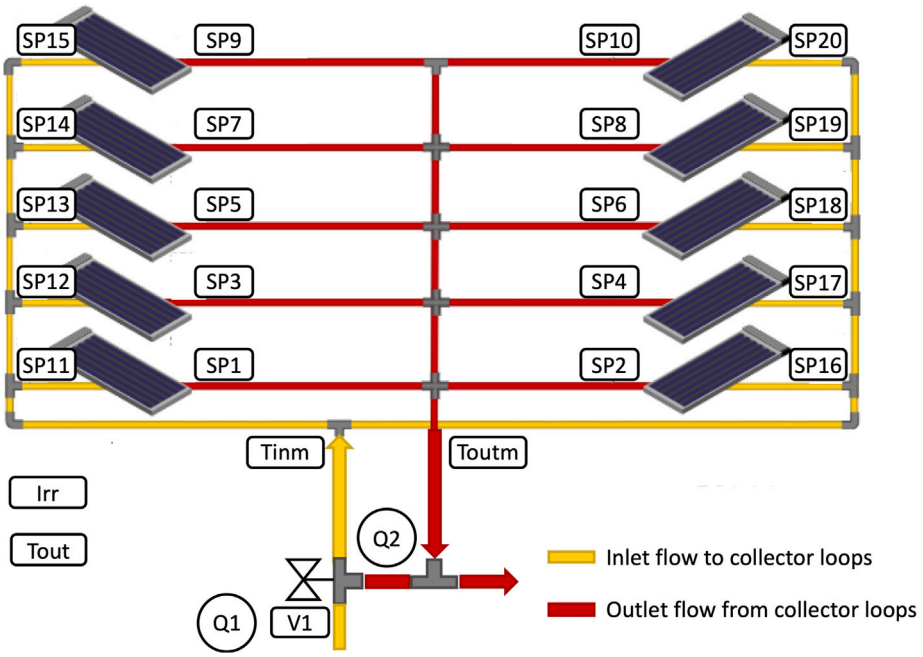


Fig. 2. Schematic diagram of the flat plate solar collector field together with its sensors.

The sensor data is stored in text files with a ".txt" extension. Each day is represented by a separate file containing the sensor readings for every minute of the day. For the study at hand, we have collected historical data for two collector loops from one year of operation. It should be noted that the data underwent preprocessing, as the sensors can malfunction. It happens that they may produce values that are outside their ranges, indicating an anomaly. In such cases, if the number of anomalous samples is limited, we use autoregressive models calculated from the data before and after the incorrect data to reconstruct such values. Conversely, if the number of outliers is large, the entire day's data is discarded. Similarly, a sensor can be disconnected, and a noticeable jump in time stamp can appear, leading to gaps in the data. In these cases, we follow a similar approach: if the data gap is significant, the entire day is excluded; if the gap is minor, an attempt is made to reconstruct the missing values by following the same procedure previously described, that is, by using autoregressive models. This preprocessing step was vital and ensured that the model received high-quality data for training, validation, and testing, thus preventing issues arising from invalid data.

As pointed out previously, a flat plate solar collector is a system ruled by PDEs as a result of the transport energy phenomena that take place inside the pipe. As such, it is considered a non-linear distributed system where partial derivatives with respect to time and space appear. This kind of system has strong non-linearities that must be considered when modeling the system (Pasamontes et al., 2011), which is done by applying finite differences discrete approximation where the pipe is discretized into a finite number of pieces and the partial spatial derivative is approximated. Although this model delivers good results, it has a lengthy computational time. An ANN prediction model fits well with forecasting and modeling the behavior of this kind of solar energy system, at the same time, providing results in a short computational time. This is the primary reason for its consideration in this paper.

Broadly speaking, an ANN is a computational model inspired by the functioning of the human brain. It is designed to analyze data, learn complex patterns, and discern intricate relationships within these data. This makes Artificial Neural Networks highly versatile, meaning it can be used for a wide spectrum of tasks, ranging from recognizing patterns to making classifications, predictions, and optimizing processes (Sharma et al., 2021).

In general, an ANN is composed of a series of interconnected layers of artificial neurons, which are designed to imitate the behavior of the human brain. There are three basic types of layers: the input layer, one or more hidden layers and the output layer. Its operating principles can be summarized as follows: information flows through the ANN from the input layer, where it receives external data from sensors and other input variables, to the output layer which provides a prediction or decision. Apart from that, connections among neurons have a weight assigned to them, which is used to modulate the strength of the signal. In

addition, hidden layers, which can vary in number and size (number of neurons), are responsible for extracting and learning complex patterns and behaviors as a function of the input data.

For the current work, we have employed the NTS (Neural Network Time Series) tool available in the Matlab environment. It is important to note that the efficiency and performance of such a network will depend on careful data preparation and architectural selection. Accordingly, these critical steps must be followed by the training and evaluation phases, which will be vital in ensuring that we can fully leverage the network's effectiveness.

In the subsequent subsections, we will delve into the specifics of the process, which include: (i) data preparation, which involves data cleaning, filtering, and handling missing or invalid values; (ii) variable selection, where we choose input and output variables based on prior system knowledge; (iii) data-set splitting, which is a vital step in the creation of training, validation, and testing data-sets and; (iv) determining the network's architecture, including the configuration of layers, neurons, and activation functions. The last step is to finish the training process by using the training data-set and the subsequent evaluation of the network's performance.

2.2. Data preprocessing and dataset construction

To create the ANN prediction model, we used a dataset that includes historical data from the entire year of 2014. Following the preprocessing procedure, which involved the removal of incorrect data (refer to Section 2.1 for details), the dataset consisted of temperature records for 277 days, sampled at one-minute intervals. Fig. 3 visually represents these temperature samples. The samples maintain their chronological order, as they occurred consecutively throughout the year. Then, we adopt a division ratio of $70-15-15$ for the training, validation, and testing segments, allocating 70% of the data for training, 15% for validation, and another 15% for testing. Our selection process is visually represented in Fig. 3, where the color blue represents the data used for training, green for validation, and red for model testing. Notably, our approach involves selecting complete days of data for training the model, which is highly beneficial when the model's structure relies on past output or input values. In contrast to random data selection, this method ensures that the model learns from complete information days, enhancing its ability to capture temporal dependencies and provide more accurate predictions. Furthermore, this division accounts for changing data patterns throughout the year, particularly during its last months when the flow in the field becomes unnecessary due to the low energy demands of the HVAC system, as it no longer has to supply the absorption machine. Consequently, this approach ensures that each of the three subsets (training, validation, and testing) captures the diverse dynamics encountered in the solar collector field covering non-operating, sunny, partly cloudy and cloudy days.

2.3. Selection of ANN input and output variables

In terms of input variables, we focused on sensor data that significantly affect the loop's output temperature. Specifically, these include the "Outdoor air Temperature", "Solar Irradiation" and "Loop Flow". The calculation of "Loop Flow" is derived from the difference between the water flowing into the flat plate solar collector field, denoted as $Q1$, and the water flow directed by the valve $V1$, represented as $Q2$. This difference is then divided by the number of loops within the system, namely, $(Q1-Q2)/10$. Additionally, we included the "Loop Inlet Temperature", labeled as SP_i, with i ranging from 11 to 20. The specific value of i depends on the loop in question, as illustrated in Fig. 2. Furthermore, we incorporated the date and time of each sensor reading to be used as supplementary input variables, collectively referred to as the "Julian Date". This addition is of paramount importance as it unveils distinct seasonal trends, while the times of the readings provide valuable insights into daily temperature variations.

Conversely, the output variables represent the information we seek to predict. In this project, our primary emphasis is on forecasting the loop's output temperature, specifically referred to as the "Loop Outlet Temperature", denoted as SP_i with i ranging from 1 to 10, as illustrated in Fig. 2. In addition, both input and output variables have been normalized within the range $[-1\ 1]$ by using the *mapminmax* function available in the Matlab environment.

In this study, our initial focus centers on predicting the outlet temperature for a specific loop, namely 'loop 1', whose outlet temperature is measured by the SP_2 sensor. After that, we explore the potential of extending the model's ability to forecast the temperature for a second loop, 'loop 2', associated with the outlet temperature measured by the SP_4 sensor, illustrated in Fig. 2. Finally, we delve into a comprehensive statistical analysis to assess the reliability and accuracy of temperature predictions across the entire system, especially under diverse weather conditions.

2.4. Selection of the ANN architecture and structure using data from loop 1

In this work, the structure of a recurrent neural network known as Nonlinear AutoRegressive with eXternal input (NARX) has been used. This structure allows to capture the dynamics of a system where the prediction of the output values depends on the current and past values of the output and also on past and current values of external input variables, see Eq. (1). Hence, it is characterized by having as input a tapped delay line for the inputs and another one for the output signals. Therefore, within the ANN in question, there are two critical parameters which must be set in the general configuration for forecasting: (i) *numberNeurons* (nn), which is related to the number of neurons in the hidden layer that the neural network should have and which are responsible for learning the characteristics of the time series data, and, (ii) *timeDelay* (td), which represents the number of steps or previous values supplied to the model to predict a future value in a single forward step. The architecture of the proposed NARX network can be observed in Fig. 4.

$$\hat{y}(n+1) = f(u(n), u(n-1), \ldots, u(n-td+1); y(n), y(n-1), \ldots, y(n-td+1)) \quad (1)$$

where $u(n)$ and $y(n)$ represent the input and output variables at time step n, and td is the number of past values assigned for both input and output variables. Finally, f is a non-linear mapping function which is approximated by means of a multiplayer perceptron.

We systematically conducted a series of experiments to adjust both parameters. These experiments were evaluated using the Mean Squared Error (MSE) to identify the best-performing cases. Our tests involved combinations of the *numberNeurons* parameter with values ranging from 2 to 14 (specifically, $nn = \{2, 4, 6, 8, 10, 12, 14\}$), and the *timeDelay* parameter with values ranging from 2 to 14 (also $td = \{2, 4, 6, 8, 10, 12, 14\}$). It is worth to mentioning that the computer used to calculate and evaluate the ANN models and run the code to data preprocessing has been a MacBook Pro with an Apple M1 PRo chip, 16 GB RAM, and 1 TB solid-state drive (SSD). On the other hand, the software used has been the MATLAB 2022a version.

Table 1 summarizes the MSE values obtained for the different configurations (nn, td). Additionally, to facilitate a more thorough understanding, the results from the table have also been visually depicted in Fig. 5. As can be seen, by evaluating the figures, it is clear that the best performing model is characterized by $(nn, td) = \{8, 10\}$. We will name it as *BestModel* throughout. However, there are other models that also demonstrate great performance, specifically, the duos $(nn, td) = \{6, 8\}, \{6, 10\}, \{12, 10\}, \{12, 14\}$. All of them have been highlighted in bold font in Table 1. In the following, these four models will be called *opponentModels*.

Finally, it is noticeable that for $(nn, td) = \{14, 14\}$, there is no improvement in data results compared to previous configurations. Accordingly, further experimentation with higher nn and td values is unnecessary, as it does not lead to enhanced model performance. On the contrary, continuing with additional runs would risk overfitting the model.

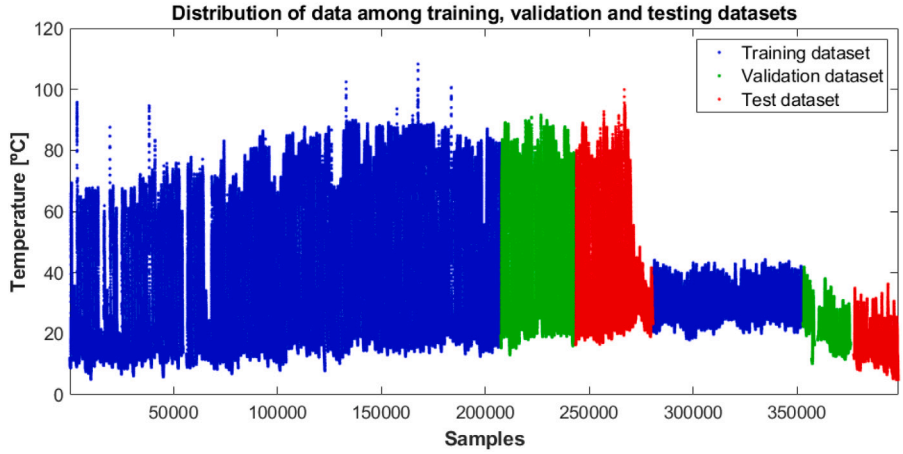


Fig. 3. Data splitting method for the training process.

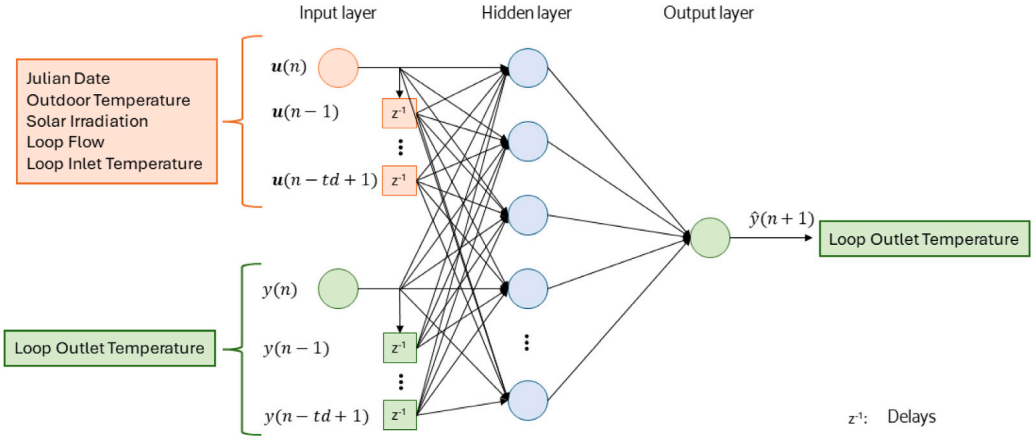


Fig. 4. Proposed neural network structure.

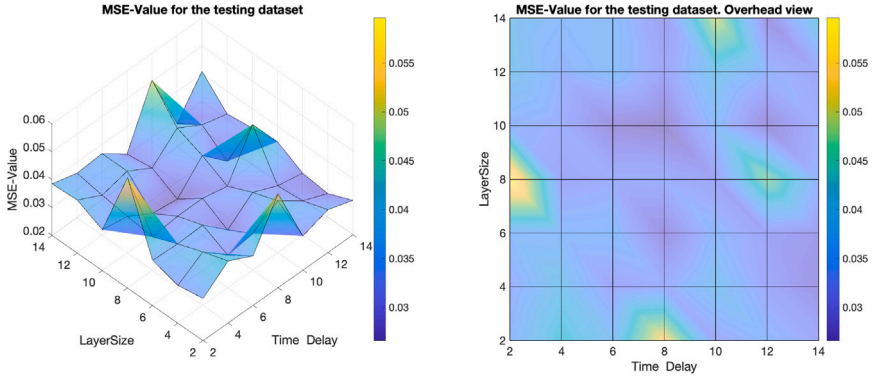


Fig. 5. Graphical representation of the MSE (left) and aerial view (right).

2.5. Performance evaluation of the preselected models with loop 2

This section is dedicated to testing whether the *BestModel* prediction model is suitable for the remaining loops in the flat plate collector field. However, in our search, to achieve a balance between model complexity and computational efficiency, we will not focus exclusively on analyzing the top-performing model. Instead, our attention turns to a closer examination of the *opponentModels*, each showing promise for further evaluation. Specifically, we will test all five models using data from another loop within the flat plate collector field, which we will call 'loop 2'. The inlet temperature of this loop is measured by the SP_{17} sensor, whereas its outlet temperature is measured by the SP_4 sensor, see Fig. 2. After that, a comprehensive statistical analysis to identify the best-implemented model will be performed.

The study involves the calculation of metrics such as MAE (Mean Absolute Error), MSE (Mean Squared Error), MAPE (Mean Absolute Percentage Error), MSPE (Mean Squared Percentage Error), RMSE (Root

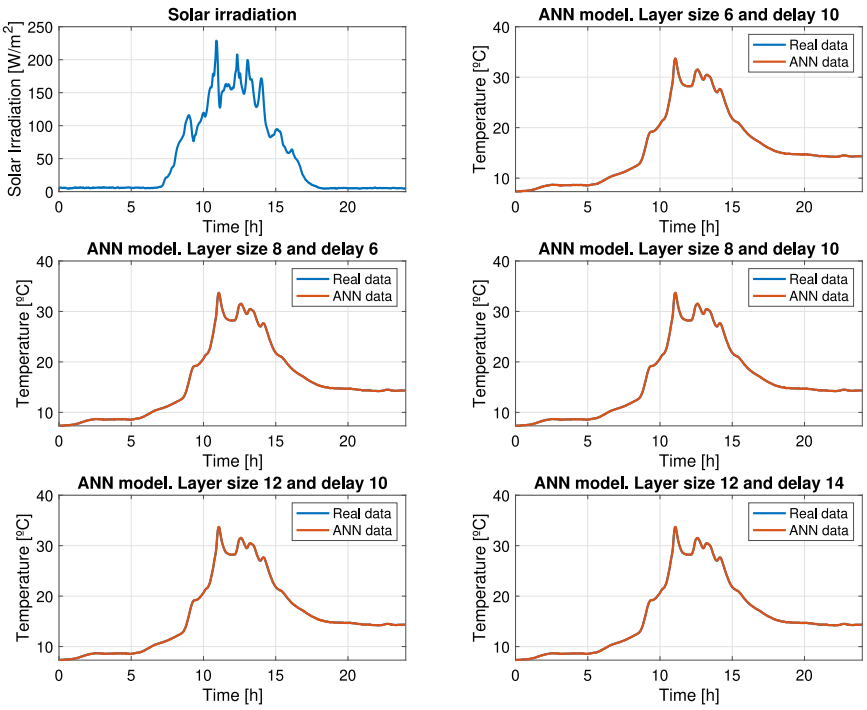


Fig. 6. Solar radiation for a cloudy day and results obtained by each prediction model versus the real solution.

Table 1
Mean Squared Error (MSE) for the testing dataset.

td	*nn*						
	2	4	6	8	10	12	14
2	0.0351	0.0422	0.0377	0.0537	0.0322	0.0353	0.0323
4	0.0335	0.0392	0.0317	0.0331	0.0385	0.0322	0.0285
6	0.0362	0.0358	0.0362	**0.0281**	0.0353	0.0313	0.0295
8	0.0596	0.0320	0.0331	0.0332	0.0312	0.0470	0.0345
10	0.0346	0.0331	**0.0278**	**0.0266**	0.0359	**0.0275**	0.0331
12	0.0390	0.0344	0.0367	0.0315	0.0395	0.0343	0.0319
14	0.0385	0.0360	0.0367	0.0302	0.0504	**0.0284**	0.0409

Mean Squared Error), NRMSE (Normalized Root Mean Squared Error), and SD (Standard Deviation). These metrics will be calculated for different weather conditions, expressly for randomly selected days characterized as cloudy, partially cloudy, and sunny, and finally for the complete dataset explained in Section 2.2.

In Figs. 6, 7, and 8, the graphs positioned in the upper left corner illustrate the solar radiation patterns for the three days selected. These three figures also show the predicted temperatures generated by the five models alongside the actual temperature values. Remarkably, the predictions exhibit an exceptionally high accuracy, closely mirroring the real results with negligible error, even though these models have been calculated with data from another loop.

The summary of the statistical analysis for these three days is displayed in Table 2. In general, the resulting errors are minimal for the subset of selected models, indicating that the predictions closely align with the actual temperature values. Besides that, the difference in the several statistical indexes among the models is negligible. This fact underlines the effectiveness of all models in predicting patterns and delivering accurate temperature forecasts, regardless of current weather conditions. The good performance of the models under different weather conditions implies that these models can capture the non-linear behavior of this kind of system whenever the operation point in which the system is working. It is worth mentioning that during cloudy days, when there is low solar irradiation, low flow is necessary to reach high temperatures at the output of the solar collector loop. On the other hand, during sunny days, the operation conditions are totally different, and a high flow rate is used to reach a loop output temperature close to the desired one. Therefore, these changes in the operational conditions make the solar collector loop have totally different dynamics.

However, it is important to note that determining an outright superior model is not straightforward. Model performance varies depending on the specific day in question. For instance, on a partially cloudy day, it appears that the smallest errors are obtained with the model parameters $(nn, td) = \{12, 10\}$. Nevertheless, for the other two days, the optimal model choice can differ depending on the specific metric used. In fact, there may be multiple viable choices for the best model, emphasizing the complexity of model selection.

In contrast, when we look at the entire dataset, we can clearly see that one model stands out as the best performer. The detailed results can be seen in Table 3, which gives an overview of the statistical analysis for the entire dataset for loop 2. For a more visual representation, see Fig. 9, where we compare the model's predictions to the actual values collected from loop 2.

What is interesting is that all five models selected provide quite similar predictions with very small errors. This suggests they are all relatively good at forecasting outlet temperature data for collector loop 2. However, the *BestModel* with parameters $(nn, td) = \{8, 10\}$ clearly outperforms the others. In fact, it performs the best across all the metrics considered. As such, it can be considered the best one for our forecasting purposes.

2.6. Prediction for multiple steps ahead

As previously mentioned, one of our main goals was fault detection in real-time. For this reason, we chose the best ANN model for predictions a single step ahead. However, it should also be possible to be interested in long-term resource management, process optimization, or risk management tasks. In such cases, adopting a predictive modeling perspective suggests that forecasting the system's behavior for multiple steps ahead is more advantageous since it offers us a more comprehensive understanding of the system's dynamics.

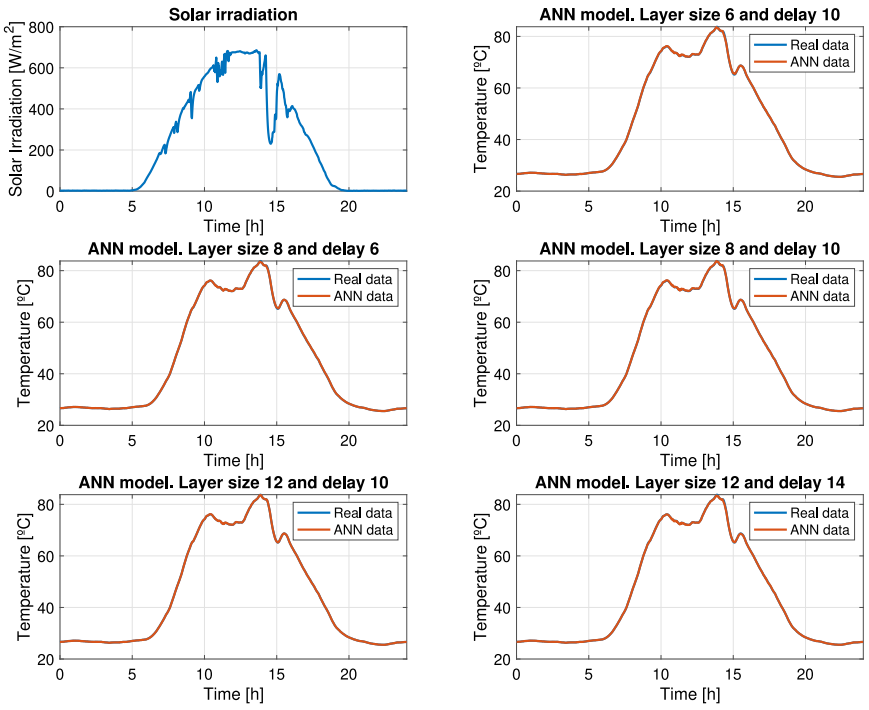


Fig. 7. Solar radiation for a partially cloudy day and results obtained by each prediction model versus the real solution.

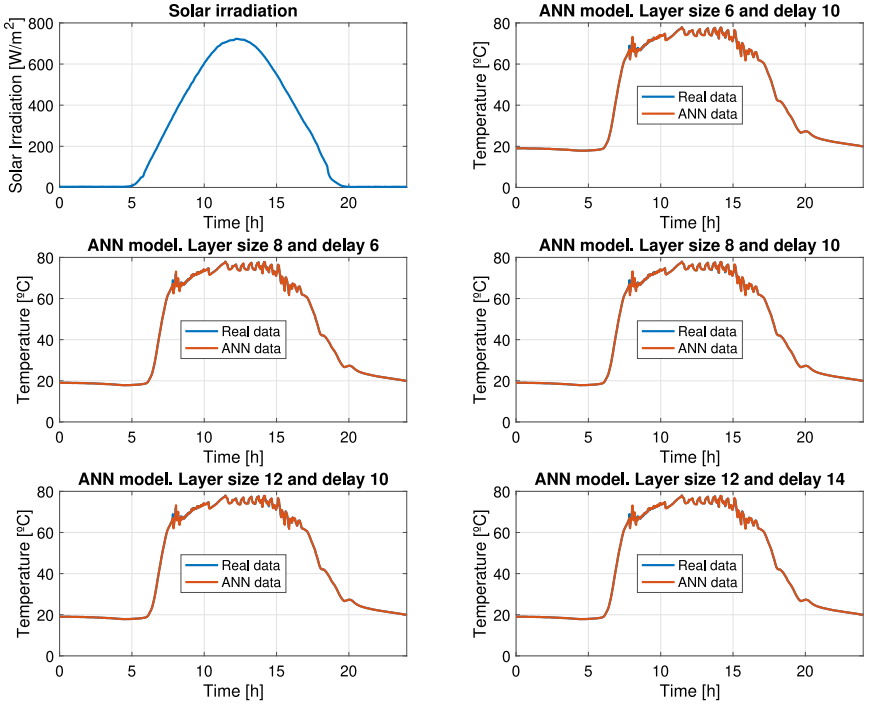


Fig. 8. Solar radiation for a sunny day and results obtained by each prediction model versus the real solution.

To illustrate this fact, we carried out an analysis where only the *BestModel* as well as the *opponetModels* were evaluated but considering multiple steps ahead, specifically five. In this case, the obtained MSE values for the models (nn, td) = {6,10}, {8,6}, {8,10}, {12,10}, and {12,14}, were 0.1058, 0.2025, 0.0408, 0.0157, and 0.3247, respectively. Notably, the model with $(nn, td) = \{12, 10\}$ emerged as the new *BestModel*, making it the preferred choice for inclusion in the digital twin for this particular analysis.

This showcases that while the shell of the digital twin remains constant, its core, that is, the ANN model, can vary depending on the ultimate objective.

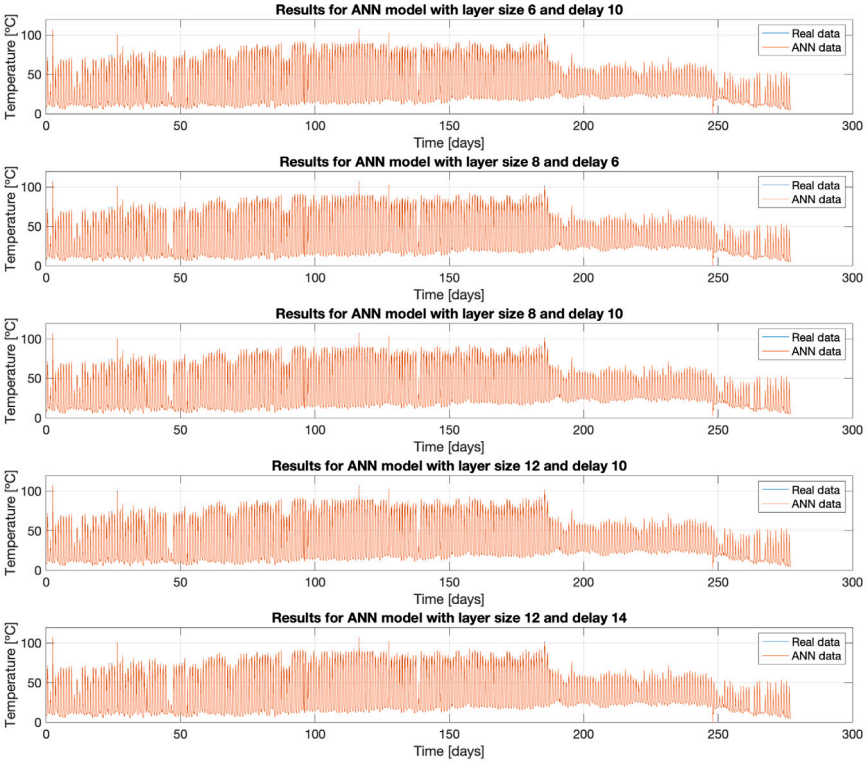


Fig. 9. Graphical representation of the results of the data from collector loop 2.

Table 2
Statistical analysis for various numberNeurons and TimeDelay configurations under different weather conditions.

	numberNeurons - TimeDelay				
	{6,10}	{8,6}	{8,10}	{12,10}	{12,14}
Metric	Cloudy day				
MSE	0.000	0.000	0.000	0.001	0.000
MAE	0.012	0.011	0.014	0.019	0.014
MSPE	0.025%	0.025%	0.033%	0.059%	0.040%
MAPE	1.204%	1.144%	1.379%	1.938%	1.362%
RMSE	0.016	0.016	0.018	0.024	0.020
NRMSE	0.099%	0.098%	0.113%	0.152%	0.125%
SD	0.015	0.016	0.018	0.024	0.018
Metric	Partially cloudy day				
MSE	0.008	0.006	0.005	0.001	0.008
MAE	0.053	0.049	0.045	0.025	0.044
MSPE	0.835%	0.567%	0.531%	0.138%	0.754%
MAPE	5.304%	4.866%	4.515%	2.476%	4.446%
RMSE	0.091	0.075	0.073	0.037	0.087
NRMSE	0.203%	0.167%	0.162%	0.082%	0.193%
SD	0.086	0.068	0.069	0.037	0.086
Metric	Sunny day				
MSE	0.072	0.073	0.073	0.072	0.072
MAE	0.112	0.114	0.103	0.108	0.102
MSPE	7.167%	7.302%	7.309%	7.187%	7.220%
MAPE	11.238%	11.436%	10.327%	10.819%	10.169%
RMSE	0.268	0.270	0.270	0.268	0.269
NRMSE	0.606%	0.612%	0.612%	0.607%	0.609%
SD	0.267	0.270	0.270	0.268	0.269

Table 3
Statistical metrics obtained for loop 2 using the entire dataset and different pairs of (nn, td).

Metric	{6,10}	{8,6}	{8,10}	{12,10}	{12,14}
MSE	0.062	0.060	0.058	0.060	0.062
MAE	0.091	0.089	0.083	0.085	0.087
MSPE	6.214%	6.011%	5.755%	5.969%	6.203%
MAPE	9.099%	8.944%	8.270%	8.516%	8.701%
RMSE	0.249	0.245	0.240	0.244	0.249
NRMSE	0.723%	0.711%	0.696%	0.709%	0.723%
SD	0.249	0.245	0.240	0.244	0.249

3. The digital twin

The core of the digital twin is the ANN prediction model as previously explained. However, our digital twin is not just limited to the model, as it is also enriched by the integration of data acquisition technologies and a user interface through a web page. This combination enables online simulation and both effective and real-time interaction with the physical system. Fig. 10 summarizes the main components of the digital twin proposed.

In the digital twin, data acquisition plays a crucial role by allowing us to collect real-time information from the physical system. The captured information is used to feed our ANN prediction model and keep it up to date, ensuring that simulations are accurate and reflect the current operating conditions of the flat plate collector field. The data from the real system are obtained through Pt100 and other IoT sensors. These data are then transmitted via the Industrial Ethernet protocol and mapped by the OPC-UA protocol. The web page is also an essential component for interacting with the digital twin. Fig. 10 summarizes the technologies used to develop the web page and the database where the data is saved. These technologies have had the following purposes: HyperText Markup Language (HTML) is a standard markup language used to create and structure the content of a web page, defining and organizing it within a web document. In this work, the last version of this language, HTML5, has been used to make the visual part belong to this web page, focussing on the frontend part of it.

On the server side, Hypertext Preprocessor (PHP) has been chosen programming language because it is suitable for web development. As a server-side language and, therefore, responsible for performing many of

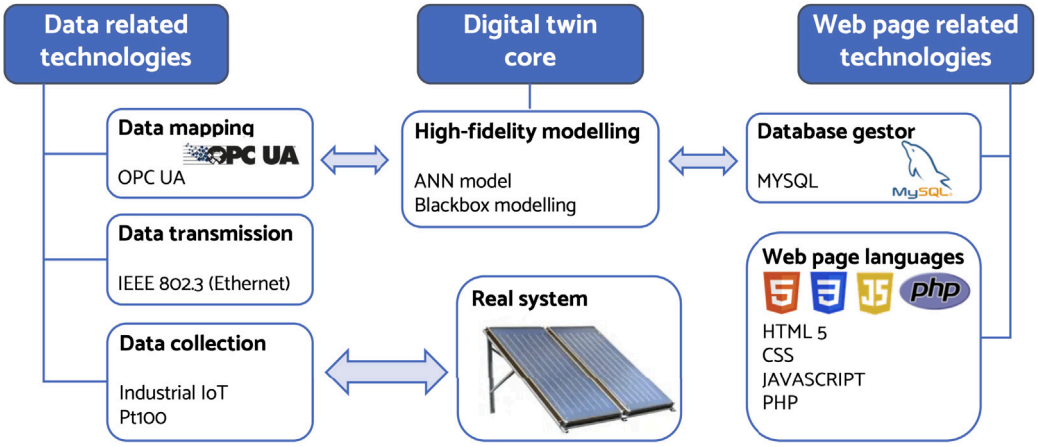


Fig. 10. Graphical representation of the digital twin.

the functionalities implemented on the web page, it has been used to establish a connection between the website and the CIESOL database and perform SQL queries to be later used in structures defined and created using HTML5.

On the other hand, the graphic design language chosen to define the presentation and style of a marked document has been Cascading Style Sheets (CSS) that works together with HTML. Therefore, CSS has been used in the development of this project to define the styles of the different pages that make up the website, giving it character and personality.

JavaScript is considered a high-level programming language with which you can add certain functionalities to a web page. In this case, it is a client-side language and, therefore, allows the client to interact with the content of the web. This language has been used to perform mainly two functions, the first being to show the value of a certain sensor visually when selected, the second being to perform the '*Show Password*' function on the login screen of the website.

Finally, as a database gestor, MySQL has been chosen. This language, which is based on the Structured Query Language (SQL), is one of the most popular database managers today. It has been used to connect to the database and, in addition, to be able to perform the respective SQL queries necessary to perform certain functionalities and even to be able to perform other operations such as creating or deleting users.

Additionally, the web page serves as a user interface, with its homepage depicted in Fig. 11. From this homepage is where the user can log into the system. In this work, we have set up three user profiles to ensure efficient data management and control over the web platform access. The first user profile corresponds to those who use the website as an informational source and do not belong to the CIESOL center. We will name them *external users*. These users can access and read data but with limitations. This feature allows for the dissemination of essential information about the flat plate solar collectors field and its operations. The second user profile, called *internal users*, is aimed at researchers or individuals who belong to the CIESOL center and require full access to the data for in-depth study and analysis purposes. These users need credentials provided by an administrator. In this regard, the *administrator*, who represents the third user profile, has complete control over the platform. In addition to managing authentication and overseeing access, the administrator can grant or revoke privileges to other users as needed. This administrative capability ensures data integrity and security within the system while also adapting to changing user requirements. The inclusion of these three user profiles creates a structure that reflects the functioning of the digital twin and ensures that data and operations of the flat plate solar collectors field are shared and managed properly, providing each type of user with the appropriate level of access and control. When having user data, it is important to be able to encrypt it and, thus, guarantee its protection. Therefore, if passwords are available, it is of great importance to carry out a data encryption process that prevents their readability, avoid obtaining them through brute force attacks, gain trust with the user and, in addition, guarantee the confidentiality of the data of such users. In this project, an encryption process was carried out using the PHP language. PHP allows us to generate a hash for passwords when creating a user. It is important to mention that a hash function is a mathematical algorithm that transforms the input data set into an alphanumeric expression that has a predetermined length. In this way, we will have a database with encrypted passwords that are irreversible or one-way.

As examples of the practicality of the web page, some pictures pertaining to the *external user* are shown in Fig. 12. As shown, in the upper left picture, Fig. 12(a), the main page offers three options: 'Solar Collector Field', 'Primary Circuit', and 'Laboratory 6'. When any of these options is selected, the user will be redirected to a specific page that allows the visual representation of the sensor scheme corresponding to that option, see Figs. 12(b)–12(d), respectively. Furthermore, users can view the data collected by each sensor once they select the one they wish to check. This is achieved through image mapping that enables these practicality.

In this way, the user can observe a graph that visually represents the data, allowing them to examine how various sensors behave at different times and on different days. To enhance the interactivity between the website and the user, a calendar field has been implemented, enabling the user to select the desired day for data visualization. Finally, it is important to mention that, as this user has the lowest level of privileges and their role is primarily to gain insight into the work done at CIESOL, it is essential to restrict the data displayed. Therefore, a simplified data representation graph has been incorporated, showing data at half-hour intervals instead of the minute-level data stored in the database.

One of the key applications of the digital twin is aiding operators in the startup and shutdown processes of the flat plate solar collectors field. In addition, it can provide insights and guidance based on real-time data, enabling smoother and more efficient transitions. Furthermore, the digital twin can play a critical role in fault detection, as it continuously monitors the system's health and can alert operators to potential issues. Fig. 13 illustrates the operation of the digital twin. It demonstrates how, after several hours of operation, the digital twin is able to replicate and capture the system behavior even during different operating conditions. Consequently, it helps the operator to detect faults in the system in real-time.

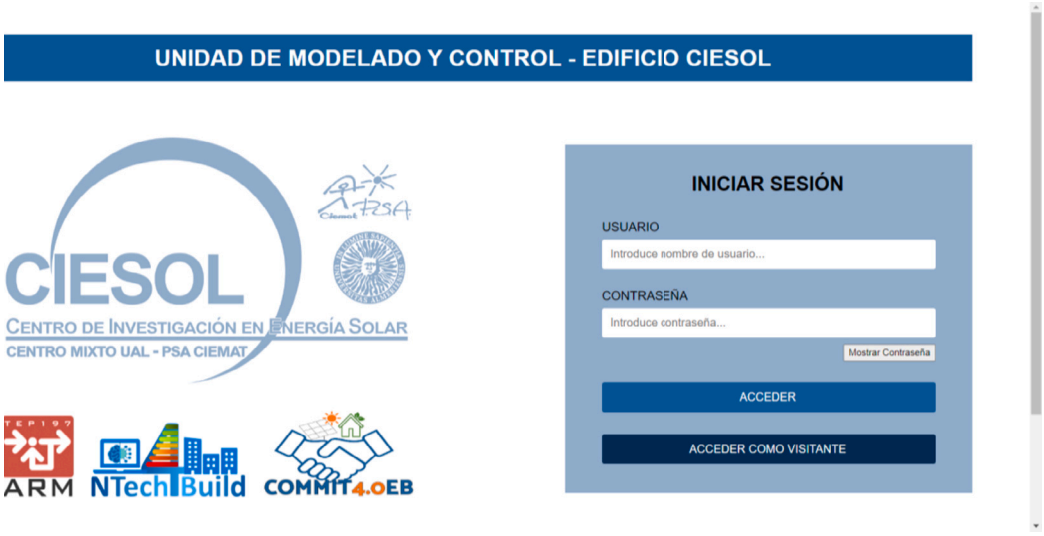
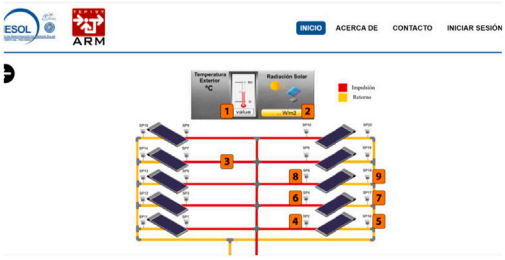
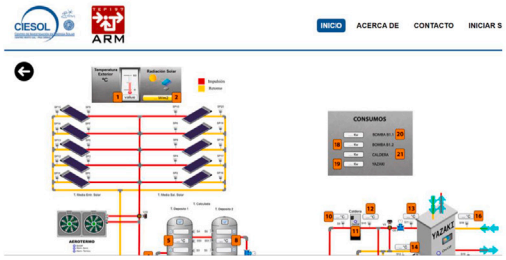


Fig. 11. Digital twin homepage.

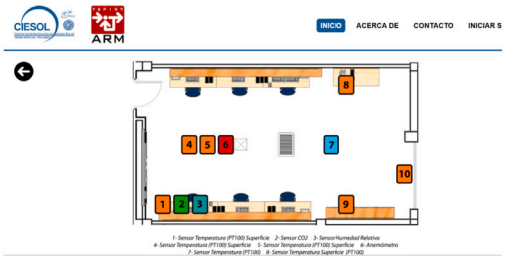
(a) Index / Home Page for External User



(b) Sensor Layout of the 'Solar Collector Field'



(c) Sensor Layout of the 'Primary Circuit'



(d) Sensor Layout of 'Laboratory 6'

Fig. 12. Pages associated with the visiting user.

4. Conclusions

The importance of net-zero energy buildings relies on two principal factors: to mitigate climate change by reducing CO_2 emissions and to reduce energy dependence. Therefore, through the integration of renewable energy sources, like solar-based ones, net-zero energy buildings can reduce greenhouse gas emissions and also promote the adoption of clean and sustainable energy technologies. Nevertheless, to guarantee adequate operation and to maximize the performance of these systems, it is necessary to apply resource optimization techniques, to identify possible malfunctions and to develop fault-tolerant strategies. To that end, digital twins play a vital role since they can provide a virtual replica of systems allowing real-time monitoring, analysis, fault-detection and simulation.

The work presented in this paper has been developed within the context of a bioclimatic building, the CIESOL research center. As mentioned previously, a crucial part of this is an HVAC system based on solar cooling which makes use of, among other elements, a flat plate solar collector field and its digital twin based on an ANN-based prediction model. The main contributions of this paper can be summarized as the development of an ANN prediction model for the outlet temperature of a flat plate solar collector field, and the integration of this ANN prediction model into a web-based digital twin framework, which provides real-time communication, monitoring and an interactive tool to develop future optimal and fault-tolerant management strategies.

To develop the ANN prediction model, different configurations for the structure of the ANN were tested using one loop from the solar collector field. From the obtained results, a pre-selection of the five most promising models was carried out and their performance evaluated for various weather conditions using a different loop from the solar collector field. The results obtained were good, with errors of less than 10% for all pre-selected models. Ultimately, the model with 8 neurons in the hidden layer and a delay time parameter of 10 samples was selected and integrated into the digital twin.

In future work, the digital twin will be extended to cover the entire solar cooling HVAC system by including other prediction models of the main components, such as the absorption machine. In addition, it will be used to perform optimal management of the available resources

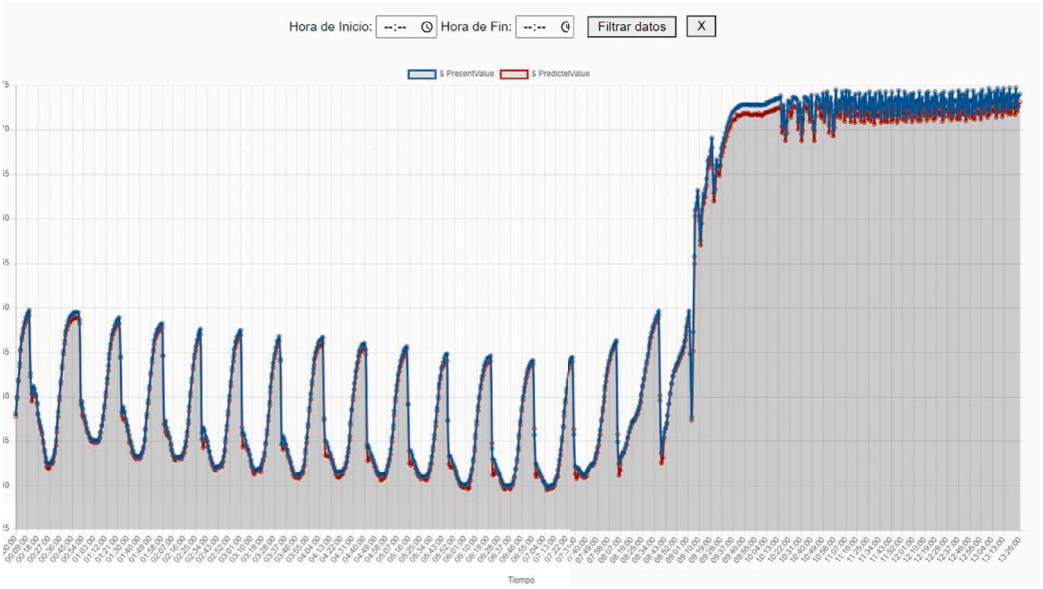


Fig. 13. Example of visualizing the present and predicted values of a temperature outlet sensor from the solar collector field.

in order to provide a comfortable environment for the users of the building, ensuring, at the same time, reduced energy consumption.

CRediT authorship contribution statement

M. Castilla: Writing – review & editing, Supervision, Formal analysis. **J.L. Redondo:** Writing – review & editing, Writing – original draft, Supervision, Methodology. **A. Martínez:** Software, Resources, Formal analysis, Data curation. **J.D. Álvarez:** Writing – review & editing, Resources, Formal analysis, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This work has been financed by COMMIT4.0EB (ref. PID2021-126889OB-I00) and COMP4HEALTH (ref. PID2021-123278OB-I00) funded by MCIN/AEI/ 10.13039/ 501100011033 and by "ERDF A way of making Europe" and NTech4Build (ref. TED2021-131655B-I00) funded by AEI/10.13039/ 501100011033 and by "European Union Next GenerationEU".

References

Castilla, M., Álvarez, J.D., Rodríguez, F., Berenguel, M., 2014. Comfort Control in Buildings. Springer.

European Climate Foundation and the European Alliance to Save Energy, 2022. Building Europe's net-zero future. Summary report, European Climate Foundation, https://europeanclimate.org/wp-content/uploads/2022/03/ecf-building-emissions-problem-march2022.pdf.

Ghenai, C., Husein, L.A., Al Nahlawi, M., Hamid, A.K., Bettayeb, M., 2022. Recent trends of digital twin technologies in the energy sector: A comprehensive review. Sustain. Energy Technol. Assess. 54, 102837.

Kumar, V., Manikandan, V., Manavaalan, G., Elango, S., 2023. Developing digital twin design for enhanced productivity of an automated anodizing industry and process prediction using hybrid deep neural network. Eng. Appl. Artif. Intell. 122, 106086.

Kuts, V., Cherezova, N., Sarkans, M., Otto, T., 2020. Digital twin: Industrial robot kinematic model integration to the virtual reality environment. J. Mach. Eng. 20 (2), 53–64.

Li, X., He, B., Zhou, Y., Li, G., 2020. Multisource model-driven digital twin system of robotic assembly. IEEE Syst. J. 15 (1), 114–123.

Liu, M., Fang, S., Dong, H., Xu, C., 2021. Review of digital twin about concepts, technologies, and industrial applications. J. Manuf. Syst. 58, 346–361.

Machado, D.O., Chicaiza, W.D., Escaño, J.M., Gallego, A.J., de Andrade, G.A., Normey-Rico, J.E., Bordons, C., Camacho, E.F., 2023. Digital twin of a fresnel solar collector for solar cooling. Appl. Energy 339, 120944.

Pasamontes, M., Álvarez, J.D., Guzmán, J.L., Lemos, J.M., Berenguel, M., 2011. A switching control strategy applied to a solar collector field. Control Eng. Pract. 19 (2), 135–145.

Rachmawati, S.M., Putra, M.A.P., Lee, J.M., Kim, D.S., 2023. Digital twin-enabled 3D printer fault detection for smart additive manufacturing. Eng. Appl. Artif. Intell. 124, 106430.

Semeraro, C., Lezoche, M., Panetto, H., Dassisti, M., 2021. Digital twin paradigm: A systematic literature review. Comput. Ind. 130, 103469.

Semeraro, C., Olabi, A., Aljaghoub, H., Alami, A.H., Al Radi, M., Dassisti, M., Abdelkareem, M.A., 2023. Digital twin application in energy storage: Trends and challenges. J. Energy Storage 58, 106347.

Shafto, M., Conroy, M., Doyle, R., Glaessgen, E., Kemp, C., LeMoigne, J., Wang, L., 2010. Draft modeling, simulation, information technology & processing roadmap. Technol. Area 11, 1–32.

Sharma, N., Sharma, R., Jindal, N., 2021. Machine learning and deep learning applications-A vision. Glob. Transit. Proc. 2 (1), 24–28. http://dx.doi.org/10.1016/j.gltp.2021.01.004.

Soliman, A., Al-Ali, A., Mohamed, A., Gedawy, H., Izham, D., Bahri, M., Erbad, A., Guizani, M., 2023. AI-based UAV navigation framework with digital twin technology for mobile target visitation. Eng. Appl. Artif. Intell. 123, 106318.

Tao, F., Xiao, B., Qi, Q., Cheng, J., Ji, P., 2022. Digital twin modeling. J. Manuf. Syst. 64, 372–389.

Yan, K., Xu, W., Yao, B., Zhou, Z., Pham, D.T., 2018. Digital twin-based energy modeling of industrial robots. In: Methods and Applications for Modeling and Simulation of Complex Systems: 18th Asia Simulation Conference, AsiaSim 2018, Kyoto, Japan, October 27–29, 2018, Proceedings 18. Springer, pp. 333–348.